ELSEVIER

Journal of Geometry and Physics 46 (2003) 203–230

JOURNAL OF GEOMETRY AND PHYSICS

www.elsevier.com/locate/jgp

A connection theoretic approach to sub-Riemannian geometry

B. Langerock

Department of Mathematical Physics and Astronomy, Ghent University, Krijgslaan 281 S9, B-9000 Gent, Belgium

Received 3 December 2001

Abstract

We use the notion of generalized connection over a bundle map in order to present an alternative approach to sub-Riemannian geometry. Known concepts, such as normal and abnormal extremals, will be studied in terms of this new formalism. In particular, some necessary and sufficient conditions for the existence of abnormal extremals will be derived. We also treat the problem of characterizing those curves that verify both the nonholonomic equations and the so-called vakonomic equations for a "free" particle submitted to some kinematical constraints.

MSC: 53C05; 53C17; 58E25

Subj. Class.: Differential geometry

Keywords: Sub-Riemannian geometry; Connections; Nonholonomic mechanics; Vakonomic dynamics

1. Introduction

A sub-Riemannian structure on a manifold is a generalization of a Riemannian structure in that a metric is only defined on a proper vector sub-bundle of the tangent bundle to the manifold (i.e. on a regular distribution), rather than on the whole tangent bundle. As a result, in sub-Riemannian geometry a notion of length can only be assigned to a certain privileged set of curves, namely curves that are tangent to the given regular distribution on which the metric is defined. The problem then arises to find those curves that minimize length, among all curves connecting two given points. The characterization of these length minimizing curves is one of the main research topics in sub-Riemannian geometry, which has also interesting links to control theory and to vakonomic dynamics (for the latter, see for instance [7]).

The connection with control theory will be touched upon in Section 2 where, without entering into the details, we will present a formulation of the Maximum Principle, following

0393-0440/02/$ – see front matter © 2002 Published by Elsevier Science B.V.
PII: S0393-0440(02)00026-8

the work of Strichartz [18,19] and Sussmann [20]. This will lead, among others, to the definition of *normal* and *abnormal extremals*. The connection with vakonomic dynamics will be explored in Section 6.

The main goal of this paper is to give an application to sub-Riemannian geometry of the theory of generalized connections over a bundle map, developed in a previous paper in collaboration with Cantrijn [4]. In Section 3, we consider some aspects of this theory in the framework of sub-Riemannian geometry. Then, normal extremals will appear as "geodesics" and abnormal extremals as "base curves of parallel transported sections" with respect to a suitable generalized connection associated to the sub-Riemannian structure. Apart from shedding some new light on certain elements of sub-Riemannian geometry, this formulation also allows us to prove some known results in an elegant way.

The main subtlety in studying length minimizing curves of a sub-Riemannian structure lies in the existence of "abnormal minimizers", i.e. length minimizing abnormal extremals. Montgomery [15] was the first to construct an explicit example of such abnormal curves. Since then, many other examples were found, for instance by Liu and Sussmann in [14]. We will deal with this topic in Section 4, where necessary and sufficient conditions for the existence of abnormal extremals are given.

In this paper, we only consider real, Hausdorff, second countable smooth manifolds, and by smooth we will always mean C^∞. The set of (real-valued) smooth functions on a manifold M will be denoted by $\mathcal{F}(M)$, the set of smooth vector fields by $\mathcal{X}(M)$ and the set of smooth 1-forms by $\mathcal{X}^*(M)$. Let V be a real vector space, and W a subspace, then the annihilator space of W is given by

$$W^0 = \{\beta \in V^* | \langle \beta, w\rangle = 0 \quad \forall w \in W\}.$$

If E is a vector bundle over a manifold M and F any vector sub-bundle, then the annihilator bundle F^0 of F is the sub-bundle of the dual bundle E^* of E over M whose fiber over a point $x \in M$ is the annihilator space of the subspace F_x of E_x. The set of smooth (local) sections of an arbitrary bundle E over M is denoted by $\Gamma(E)$. In this paper, the domain of a curve will usually be taken to be a closed (compact) interval in $\mathbb{R}$. Whenever we say that such a curve, defined on an interval $[a, b]$, is an integral curve of a vector field, we simply mean that it is the restriction of a maximal integral curve defined on an open interval containing $[a, b]$.

2. General definitions

In this section, we first give a brief review of some natural objects associated to a sub-Riemannian structure and we recall the necessary conditions, derived from the Maximum Principle, for a curve to be length minimizing. Next, we discuss some general aspects of the theory of connections over a bundle map.

2.1. Sub-Riemannian structures: preliminary definitions

Suppose that M is a smooth manifold of dimension n, equipped with a regular distribution $Q \subset TM$ (i.e. Q is a smooth distribution of constant rank, say of rank k). In view of the regularity, Q can alternatively be regarded as a vector sub-bundle of TM over M. The natural injection $i : Q \hookrightarrow TM$ is then a linear bundle mapping fibered over the identity. A regular distribution is also completely characterized by its annihilator, i.e. giving Q is equivalent to specifying the sub-bundle Q^0 of the cotangent bundle T^*M whose fiber over $x \in M$ consists of all co-vectors at x which annihilate all vectors in the subspace Q_x of T_xM.

A smooth Riemannian bundle metric h on Q is a smooth section of the tensor bundle $Q^* \otimes Q^* \to M$ such that it is symmetric and positive definite, i.e. for all $X_x, Y_x \in Q_x$ one has:

$$h(x)(X_x, Y_x) = h(x)(Y_x, X_x),$$
$$h(x)(X_x, X_x) \geq 0, \quad \text{and the equality holds iff} \quad X_x = 0.$$

With a Riemannian bundle metric one can associate a smooth linear bundle isomorphism $\flat_h : Q \to Q^*$, $X_x \mapsto h(x)(X_x, \cdot)$, fibered over the identity on M, with inverse denoted by $\sharp_h := \flat_h^{-1} : Q^* \to Q$.

Definition 1. A sub-Riemannian structure (M, Q, h) is a triple, where M is a smooth manifold, Q a smooth regular distribution on M, and h a Riemannian bundle metric on Q.

Although it is not explicitly mentioned in the definition, it will always be tacitly assumed, as it is customary in sub-Riemannian geometry, that Q is a nonintegrable distribution and, therefore, does not induce a foliation of M. A manifold M equipped with a sub-Riemannian structure, will be called a *sub-Riemannian manifold*. With a sub-Riemannian structure (M, Q, h) one can associate a smooth mapping $g : T^*M \to TM$ defined by

$$g(\alpha_x) = i(\sharp_h(i^*(\alpha_x))) \in TM,$$

where $i^* : T^*M \to Q^*$ is the adjoint mapping of i, i.e. for any $\alpha_x \in T_x^*M$, $i^*(\alpha_x)$ is determined by $\langle i^*(\alpha_x), X_x\rangle = \langle \alpha_x, i(X_x)\rangle$ for all $X_x \in Q_x$. Clearly, g is a linear bundle mapping whose image set is precisely the sub-bundle Q of TM and whose kernel is the annihilator Q^0 of Q. To simplify notations we shall often identify an arbitrary vector in Q with its image in TM under i and smooth sections of Q (i.e. elements of $\Gamma(Q)$) will often be regarded as vector fields on M.

With g we can further associate a section $\bar{g}$ of $TM \otimes TM \to M$ according to

$$\bar{g}(x)(\alpha_x, \beta_x) = \langle g(\alpha_x), \beta_x\rangle$$

for all $x \in M$ and $\alpha_x, \beta_x \in T_x^*M$. From

$$\begin{aligned}\bar{g}(x)(\alpha_x, \beta_x) &:= \langle g(\alpha_x), \beta_x\rangle = \langle \sharp_h(i^*\alpha_x), i^*(\beta_x)\rangle = h(x)(\sharp_h(i^*\alpha_x), \sharp_h(i^*\beta_x)) \\ &= h(x)(g(\alpha_x), g(\beta_x)),\end{aligned}$$

we conclude that $\bar{g}$ is symmetric.

Let G be a Riemannian metric on M. It is easily seen that, given a regular distribution Q on M, we can associate with the metric G a sub-Riemannian structure (M, Q, h_G), where h_G is the restriction of G to the sub-bundle Q, i.e. $h_G(x)(X_x, Y_x) := G(x)(X_x, Y_x)$ for any $x \in M$ and $X_x, Y_x \in Q_x$. Given a sub-Riemannian structure (M, Q, h) and a Riemannian metric G on M, we say that the Riemannian metric restricts to h if $h_G = h$. Now, every sub-Riemannian structure can be seen as being determined (in a nonunique way) by the restriction of a Riemannian metric. Indeed, let h be a Riemannian bundle metric on a vector sub-bundle Q of TM, and let $\{U_\alpha\}$ be an open covering of M such that, on each U_α, there exists an orthogonal basis $\{X_1, \dots, X_k\}$ of local sections of Q with respect to h. Extend this to a basis of vector fields $\{X_1, \dots, X_n\}$ on U_α and define a Riemannian metric on U_α by

$$G_\alpha(x)(X_x, Y_x) = \sum_{i,j=1}^{k} a^i b^j h(x)(X_i(x), X_j(x)) + \sum_{i=k+1}^{n} a^i b^i,$$

where $X_x = a^i X_i(x)$ and $Y_x = b^i X_i(x)$, with $a^i, b^i \in \mathbb{R}$. One can then glue these metrics together, using a partition of unity sub-ordinate to the given covering $\{U_\alpha\}$. This procedure, which is similar to the one adopted for constructing a Riemannian metric on an arbitrary smooth manifold (see for instance [2, Proposition 9.4.1]), produces a Riemannian metric on M which, by construction, restricts to h.

In the sequel, we will repeatedly make use of a Riemannian metric G which restricts to a given sub-Riemannian metric h. In that connection we now introduce some further notations and prove some useful relations associated to G and h. The natural bundle isomorphism between TM and T^*M induced by G will be denoted by $\sharp_G$, with inverse $\flat_G = \sharp_G^{-1}$. Let $x \in M$ and let $X_x, Y_x \in Q_x$, then one has:

$$\langle i^*\flat_G(i(X_x)), Y_x\rangle = \langle \flat_G(i(X_x)), i(Y_x)\rangle = \langle \flat_h(X_x), Y_x\rangle,$$

which implies that $\flat_h = i^* \circ \flat_G \circ i$. Inserting this into $g \circ \flat_G \circ i$ and taking into account the definition of g, we conclude that

$$g \circ \flat_G \circ i = i \quad \text{or} \quad g \circ \flat_G|_Q = \mathrm{id}_Q,$$

where id_Q is the identity mapping on Q. The orthogonal projections of TM onto Q and onto its G-orthogonal complement $Q^\perp$ will be denoted by π and $\pi^\perp$, respectively. Now, T^*M can be written as the direct sum of $(Q^\perp)^0$ and Q^0 and the corresponding projections will be denoted by τ and $\tau^\perp$, respectively. It is easily proven that $(Q^\perp)^0 \cong \flat_G(Q)$ and that

$$\tau^\perp = \flat_G \circ \pi^\perp \circ \sharp_G, \qquad \tau = \flat_G \circ \pi \circ \sharp_G.$$

Using the fact that $g \circ \flat_G|_Q = \mathrm{id}_Q$ and $\ker g = Q^0$, we also have: $g = g \circ \tau = \pi \circ \sharp_G$.

To any regular distribution Q on M one can associate a natural tensor field acting on $Q^0 \otimes Q \otimes Q$. Indeed, let $\eta \in \Gamma(Q^0)$, $X, Y \in \Gamma(Q)$ and let $[X, Y]$ denote the Lie bracket of X and Y, regarded as vector field on M. Then it is easily proven that the expression $\langle \eta, [X, Y]\rangle$ is $\mathcal{F}(M)$-linear in all three arguments and, therefore, determines a tensorial object. Now, Q is involutive if and only if this tensor is identically zero. Next, assume that $Y \in \mathcal{X}(M)$, with η and X as before, then $\langle \eta, [X, Y]\rangle$ is still $\mathcal{F}(M)$-linear in η and X (but not in Y). This justifies the following notation, which will be used later on in our discussion of the Maximum Principle: for any $x \in M$, $\eta_x \in Q^0_x$, $X_x \in Q_x$ and arbitrary $Y \in \mathcal{X}(M)$, put

$$\langle \eta_x, [X_x, Y]\rangle := \langle \eta, [X, Y]\rangle(x), \tag{1}$$

where η (resp. X) may be any section of Q^0 (resp. Q) such that $\eta(x) = \eta_x$ (resp. $X(x) = X_x$).

2.2. Necessary conditions for length minimizing curves

For the further discussion in this paper it is important that we give a precise description of the class of curves we will be dealing with. First of all, by a *curve* in an arbitrary manifold P we shall always mean a smooth mapping (in the C^∞ sense) $c : I \to P$, with $I \subset \mathbb{R}$ a closed interval, and such that c admits a smooth extension to an open interval containing I. A mapping $c : [a, b] \to P$ will be called a *piecewise curve* in P if there exists a finite subdivision $a_1 := a < a_2 < \cdots < a_k < a_{k+1} := b$ such that the following conditions are fulfilled:

(1) c is left continuous at each point a_i for $i = 2, \ldots, k+1$, i.e. $\lim_{t\to a_i^-} c(t)$ exists and equals $c(a_i)$;
(2) $\lim_{t\to a_i^+} c(t)$ is defined for all $i = 1, \ldots, k$ and $\lim_{t\to a_1^+} c(t) = c(a_1)$ (i.e. c is right continuous at $a_1 = a$);
(3) for each $i = 1, \ldots, k$, the mapping $c^i : [a_i, a_{i+1}] \to P$, defined by $c^i(t) = c(t)$ for $t \in]a_i, a_{i+1}]$ and $c^i(a_i) = \lim_{t\to a_i^+} c(t)$, is smooth (i.e. is a curve in P).

A piecewise curve which is continuous everywhere, will simply be called a *continuous piecewise curve* (and corresponds to what is often called in the literature, a piecewise smooth curve.)

In the sequel, whenever we are dealing with a (continuous) piecewise curve $c : [a, b] \to P$, the notation c^i will always refer to the curve defined on the ith subinterval of $[a, b]$, bounded by points where c fails to be smooth.

Consider now a sub-Riemannian structure (M, Q, h), with associated bundle map $g : T^*M \to TM$. A curve (resp. piecewise curve) $c : [a, b] \to M$ is said to be *tangent to* Q if $\dot c(t) \in Q_{c(t)}$ for all $t \in [a, b]$ (resp. for all t where the derivative exists). Next, let $\alpha : I \to T^*M$ be a curve in T^*M and put $c = \pi_M \circ \alpha$, with $\pi_M : T^*M \to M$ the natural cotangent bundle projection. Then, we say that α is *g-admissible* if

$$g(\alpha(t)) = \dot c(t) \quad \forall t \in I.$$

The projected curve c will be called the *base curve* of α. If $\alpha : I = [a, b] \to T^*M$ is a piecewise curve, then α will be called g-admissible if its projection $c = \pi_m \circ \alpha$ is a continuous piecewise curve such that, in addition, $g(\alpha^i(t)) = \dot c^i(t)$ for $t \in [a_i, a_{i+1}]$ (where we have used the notational conventions introduced above). We now prove the following result which will be of use later on.

Lemma 2. *Given a sub-Riemannian structure (M, Q, h) and any curve (resp. continuous piecewise curve) c in M, tangent to Q. Then, there always exists a g-admissible curve (resp. piecewise curve) in T^*M which projects onto c.*

Proof. Take a Riemannian metric G which restricts to h on Q. If $c : [a, b] \to M$ is a curve tangent to Q, one can simply put $\alpha(t) = \flat_G(\dot c(t))$ for all $t \in [a, b]$. Clearly, α then defines a g-admissible curve in T^*M with base curve c.

Next, assume $c : [a, b] \to M$ is a continuous piecewise curve, tangent to Q. We can then define a piecewise curve α in T^*M as follows: put $\alpha(t) = \flat_G(\dot c(t))$ for all t, where $\dot c(t)$

is defined and, using the notational conventions introduced above, $\alpha(a_{i+1}) = \flat_G(\dot{c}^i(a_{i+1}))$ for $i = 1, \ldots, k$. It is easy to check that the mapping $\alpha : [a, b] \to T^*M$ thus constructed, is a g-admissible piecewise curve, projecting onto c. □

We will now introduce the notion of length of curves, and of continuous piecewise curves, tangent to Q.

Definition 3. Given a sub-Riemannian structure (M, Q, h), then the length of a curve $c : [a, b] \to M$, tangent to Q, is given by

$$L(c) := \int_a^b \sqrt{h(c(t))(\dot{c}(t), \dot{c}(t))}\, \mathrm{d}t.$$

Given any g-admissible curve α in T^*M with base curve c, and a Riemannian metric G which restricts to h, then the length of c still equals

$$L(c) = \int_a^b \sqrt{\bar{g}(c(t))(\alpha(t), \alpha(t))}\, \mathrm{d}t = \int_a^b \sqrt{G(c(t))(\dot{c}(t), \dot{c}(t))}\, \mathrm{d}t.$$

In particular, the value of these integrals do not depend on the specific choice of α, resp. G.

The above notion of length can be easily extended to the class of continuous piecewise curves c, tangent to Q, by putting $L(c) = \sum_{i=1}^k L(c^i)$.

For the following discussion, which is partially inspired on Sussmann's [20] approach to a coordinate-free version of the Pontryagin Maximum Principle, we make two additional assumptions. First, we assume that M is pathwise connected, and secondly, we take the distribution Q of the given sub-Riemannian structure (M, Q, h) to be bracket generating, i.e. if $L(Q)$ denotes the Lie algebra generated by sections of Q, regarded as vector fields on M, then we assume that at each point $x \in M$, $T_xM = \{X(x) |\text{ for all } X \in L(Q)\}$.

Both assumptions imply, in particular, that any two points of M can be joined by a continuous piecewise curve tangent to Q, as follows from a well-known theorem of Chow [6]. Therefore, under these assumptions it makes sense to talk about the length minimizing curve connecting two given points. More precisely, given a continuous piecewise curve $c : [a, b] \to M$ tangent to Q, connecting two points x_0 and x_1 (i.e. $c(a) = x_0$, $c(b) = x_1$), then c is called *length minimizing* if $L(c) \leq L(\tilde{c})$ for any other continuous piecewise curve $\tilde{c} : [a, b] \to M$ tangent to Q, with $\tilde{c}(a) = x_0$ and $\tilde{c}(b) = x_1$.

Note that, given a continuous piecewise curve c, connecting two points x_0 and x_1, one can always determine a parameterization of c such that $c : [0, 1] \to M$, with $c(0) = x_0$, $c(1) = x_1$, and for which there exists a nonzero constant k such that $h(c(t))(\dot{c}(t), \dot{c}(t)) = k$ for all t, where $\dot{c}(t)$ is defined. Following Sussmann, we will call this a *parameterization by constant times arc-length*.

We now arrive at the following weak version of the Maximum principle.

Theorem 4. *Consider a sub-Riemannian structure (M, Q, h) with M connected and Q bracket generating. Let $c : [0, 1] \to M$ be a continuous piecewise curve which is length minimizing, and parameterized by constant times arc-length. Then, there exists a continuous* *piecewise curve* $\psi : [0, 1] \to T^*M$ *along c, i.e.* $\pi_M(\psi(t)) = c(t)$, *which does not intersect the zero section and such that at least one of the following two conditions is satisfied:*

(i) $\psi(t)$ *is an integral curve of the Hamiltonian vector field* X_H on T^*M, *with Hamiltonian given by the smooth function* $H(\alpha_x) = (1/2)\bar{g}(x)(\alpha_x, \alpha_x)$ *for* $\alpha_x \in T^*_x M$, *which, in particular, implies that both* ψ *and c are smooth;*

(ii) $\psi(t) \in Q^0$ *for all* $t \in I$, *and for any piecewise g-admissible curve* α *with base curve c, the following equation holds:*

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_t \langle \psi(t), X(c(t))\rangle = \langle \psi(t), [g(\alpha(t)), X]\rangle,$$

for all $X \in \mathcal{X}(M)$ *and all* $t \in [0, 1]$, *where* $\dot{c}(t)$ *is well defined.*

Note that on the right-hand side of the equation in (ii) we have used the notation introduced in (1). For a derivation of this weak version of the Maximum principle in terms of the more general class of absolutely continuous curves, we refer to [20]. Inspired on the (local) analysis presented in [17, p. 79], the proof of the above theorem follows by making some minor adjustments to the one given in [20].

Definition 5. A continuous piecewise curve c tangent to Q is called a normal (resp. abnormal) extremal if there exists a continuous piecewise curve ψ in T^*M along c, which does not intersect the zero section of T^*M, satisfying condition (i) (resp. (ii)) of Theorem 4.

Note that, according to this definition, normal or abnormal extremals do not have to be length minimizing and that c can be simultaneously a normal and an abnormal extremal.

2.3. Connections over a bundle map: general setting

Inspired by some recent work of Fernandes [8] on "contravariant connections" in Poisson geometry and, more generally, connections associated with Lie algebroids (see [9]), we have recently embarked on the study of a general notion of connection, namely connections defined over a vector bundle map. This concept covers, besides the standard notions of linear and nonlinear connections, various generalizations such as partial connections and pseudo-connections, as well as the Lie algebroid connections considered by Fernandes. For a detailed treatment we refer to a forthcoming paper, written in collaboration with Cantrijn and Langerock [4]. After briefly sketching the main idea underlying the notion of a generalized connection over a vector bundle map, we shall apply this notion of connection to a sub-Riemannian structure.

Let M be a manifold and $\nu : N \to M$ a vector bundle over M. Assume, in addition, that a linear bundle map $\rho : N \to TM$ is given such that $\tau_M \circ \rho = \nu$, where $\tau_M : TM \to M$ denotes the natural tangent bundle projection. Note that we do not require ρ to be of constant rank. Hence, the image set Im ρ need not be a vector sub-bundle of TM, but rather determines a generalized distribution as defined by Stefan and Sussmann (see, e.g. [13, Appendix 3]). It follows that ρ induces a mapping of sections, $\Gamma(N) \to \mathcal{X}(M) : s \mapsto \rho \circ s$, also denoted by ρ. Next, let $\pi : E \to M$ be an arbitrary fiber bundle over M. We may then consider the pull-back bundle $\tilde{\pi}_1 : \pi^* N \to E$, which is a vector bundle over E. Note that $\pi^* N$ may also be regarded as a fiber bundle over N, with projection denoted by $\tilde{\pi}_2 : \pi^* N \to N$.

Definition 6. A generalized connection on E over the bundle map ρ (or, shortly, a ρ-connection on E) is then defined as a linear bundle map $h : \pi^* N \to TE$ from $\tilde{\pi}_1$ to τ_E, over the identity on E, such that, in addition, the following diagram is commutative

$$\begin{tikzcd} \pi^* N \arrow[r, "h"] \arrow[d, "\tilde{\pi}_2"'] & TE \arrow[d, "T\pi"] \\ N \arrow[r, "\rho"'] & TM \end{tikzcd}$$

where $T\pi$ denotes the tangent map of π.

The image set $\operatorname{Im} h$ determines a generalized distribution on E which projects onto $\operatorname{Im} \rho$. It is important to note that $\operatorname{Im} h$ may have nonzero intersection with the bundle VE of π-vertical tangent vectors to E. The standard notion of a connection on E is recovered when putting $N = TM$, $\nu = \tau_M$, and ρ the identity map. In case P is a principal G-bundle over M, with right action $R : P \times G \to P$, $(e, g) \mapsto R(e, g) = R_g(e)(= eg)$, a ρ-connection h on P will be called a *principal ρ-connection* if, in addition, it satisfies

$$TR_g(h(e, n)) = h(eg, n)$$

for all $g \in G$ and $(e, n) \in \pi^* N$. Slightly modifying the construction described by Kobayashi and Nomizu [11], given a principal ρ-connection on P, one can construct a ρ-connection on any associated fiber bundle E.

Assume E is a vector bundle and let $\{\phi_t\}$ denote the flow of the canonical dilation vector field on E. A ρ-connection h on E is then called *a linear ρ-connection* if

$$T\phi_t(h(e, n)) = h(\phi_t(e), n)$$

for all $(e, n) \in \pi^* N$. In [4], it is shown that such a linear ρ-connection can be characterized by a mapping $\nabla : \Gamma(N) \times \Gamma(E) \to \Gamma(E)$, $(s, \sigma) \mapsto \nabla_s \sigma$ such that the following properties hold:

(1) ∇ is $\mathbb{R}$-linear in both arguments;
(2) ∇ is $\mathcal{F}(M)$-linear in s;
(3) for any $f \in \mathcal{F}(M)$ and for all $s \in \Gamma(N)$ and $\sigma \in \Gamma(E)$ one has: $\nabla_s(f\sigma) = f\nabla_s\sigma + (\rho \circ s)(f)\sigma$.

It immediately follows that $\nabla_s\sigma(m)$ only depends on the value of s at m, and therefore we may also write it as $\nabla_{s(m)}\sigma$. Clearly, ∇ plays the role of the covariant derivative operator in the case of an ordinary linear connection. Henceforth, we will also refer to the operator ∇ as a linear ρ-connection. Let k and ℓ denote the fiber dimensions of N and E, respectively, and let $\{s^\alpha : \alpha = 1, \ldots, k\}$, resp. $\{\sigma^A : A = 1, \ldots, \ell\}$, be a local basis of sections of ν, resp. π, defined on a common open neighborhood $U \subset M$. We then have $\nabla_{s^\alpha}\sigma^A = \Gamma_B^{\alpha A}\sigma^B$, for some functions $\Gamma_B^{\alpha A} \in \mathcal{F}(U)$, called the connection coefficients of the given ρ-connection. A ρ-connection ∇ can be extended to an operator, also denoted by ∇, acting on sections of any tensor product bundle of E. This can be achieved by applying standard arguments, and the details are left to the reader. We just like to mention here that the action on $\mathcal{F}(M)$ and $\Gamma(E^*)$ is determined by the following relations: for $s \in \Gamma(N)$, $f \in \mathcal{F}(M)$, $\sigma \in \Gamma(E)$ and $\zeta \in \Gamma(E^*)$,

$$\nabla_s f := (\rho \circ s)(f), \qquad \nabla_s \langle \sigma, \zeta \rangle = \rho(s)\langle \sigma, \zeta \rangle = \langle \nabla_s \sigma, \zeta \rangle + \langle \sigma, \nabla_s \zeta \rangle.$$

In order to associate a notion of parallel transport to a linear ρ-connection, we first need to introduce a special class of curves in N. A curve $\tilde{c} : I = [a, b] \to N$ is called *ρ-admissible* if for all $t \in I$, one has $\dot{c}(t) = (\rho \circ \tilde{c})(t)$, where $c = \nu \circ \tilde{c}$ is the projected curve on M. Curves in M that are projections of ρ-admissible curves in N are called *base curves*. (We will see that this terminology is in agreement with the one introduced in the previous section.) Note that, in principle, a base curve may reduce to a point.

As in standard connection theory, with any linear ρ-connection ∇ on a vector bundle $\pi : E \to M$, and any ρ-admissible curve $\tilde{c} : [a, b] \to N$, one can associate an operator $\nabla_{\tilde{c}}$, acting on sections of π defined along the base curve $c = \nu \circ \tilde{c}$. The operator $\nabla_{\tilde{c}}$ is completely determined by the following prescriptions. For arbitrary sections σ of π along c (i.e. curves $\sigma : [a, b] \to E$, satisfying $\pi \circ \sigma = c$) and for arbitrary $f \in \mathcal{F}([a, b])$:

(1) $\nabla_{\tilde{c}}$ is $\mathbb{R}$ linear;
(2) $\nabla_{\tilde{c}} f\sigma = \dot{f}\sigma + f\nabla_{\tilde{c}}\sigma$;
(3) $\nabla_{\tilde{c}}\sigma(t) = \nabla_{\tilde{c}(t)}\bar{\sigma}$, for $\bar{\sigma} \in \Gamma(\pi)$ such that $\bar{\sigma}(c(t)) = \sigma(t)$ for all $t \in [a, b]$.

Definition 7. A section σ of π, defined along the base curve of a ρ-admissible curve $\tilde{c} : [a, b] \to N$, will be called parallel along $\tilde{c}$ if $\nabla_{\tilde{c}}\sigma(t) = 0$ for all $t \in [a, b]$.

Taking again $\{s^\alpha\}$, resp. $\{\sigma^A\}$, to be a local basis of sections of ν, resp. π, and putting $\sigma(t) = \sigma_A(t)\sigma^A(c(t))$ and $\tilde{c}(t) = \tilde{c}_\alpha(t)s^\alpha(c(t))$, we find that σ is parallel along $\tilde{c}$ if

$$\nabla_{\tilde{c}}\sigma(t) = (\dot{\alpha}_A(t) + \Gamma_A^{\alpha B}(\tilde{c}(t))\sigma_B(t)\tilde{c}_\alpha(t))\sigma^A(c(t)) = 0,$$

which gives a system of linear differential equations for the components of σ. Again using standard arguments, one can show that this leads to a notion of parallel transport on E along ρ-admissible curves in N (cf. [4] for more details).

Suppose we are given two ρ-admissible curves $\tilde{c}^i : [a_i, b_i] \to N$, $i = 1, 2$ with $\tilde{c}^1(b_1)$ and $\tilde{c}^2(a_2)$ belonging to the same fiber of ν, i.e. $c^1(b_1) = c^2(a_2)$, where c^i is the base curve of $\tilde{c}^i$. Given any point in $E_{c^1(a_1)}$ one can construct a unique parallel section along $\tilde{c}^1$, starting from that point. The endpoint of this curve (at $t = b_1$) lies in the fiber $E_{c^2(a_2)}$ and, therefore, can be taken as the initial point of a unique parallel curve along $\tilde{c}^2$. This construction can now be easily extended to the class of piecewise ρ-admissible curves defined below.

Recalling the definition of a piecewise curve, given in the previous section, and using the notational conventions introduced there, a *piecewise ρ-admissible curve* $\tilde{c}$ is defined as a piecewise curve in N such that: (i) for each $i = 2, \ldots, k$, $\tilde{c}^i(a_i)$ and $\tilde{c}^{i+1}(a_i)$ ($= \lim_{t \to a_i^+} c(t)$) belong to the same fiber of ν or, equivalently, the projection $c = \nu \circ \tilde{c}$ is a continuous piecewise curve, (ii) $\rho(\tilde{c}^i(t)) = \dot{c}^i(t)$ for all $i = 1, \ldots, k$ and $t \in [a_i, a_{i+1}]$.

Extending the above construction in the case of two ρ-admissible curves $\tilde{c}^1, \tilde{c}^2$, it is now clear how to determine the notion of parallel transport along a piecewise ρ-admissible curve.

The following class of linear ρ-connections will play an important role in the further analysis.

Definition 8. A linear ρ-connection on a vector bundle E is called partial if for any $\sigma \in \Gamma(E)$ and $n \in \ker(\rho)$, we have $\nabla_n \sigma = 0$.

It is instructive to know that the condition for a connection to be partial is equivalent to the property that no (nonzero) vertical tangent vectors to E exist that are also contained in $\operatorname{Im} h$, as stated in the following proposition. For the proof, which is quite technical, we refer to [4].

Proposition 9. *Let ∇ be a linear ρ-connection. Then ∇ is partial if and only if* $\operatorname{Im} h \cap VE = \{0\}$.

3. Connections on a sub-Riemannian structure

Fix a sub-Riemannian structure (M, Q, h) and consider the associated bundle map $g : T^*M \to TM$. In this section, we will be interested in generalized connections on T^*M over g. Our main goal is the characterization of normal and abnormal extremals of the sub-Riemannian structure in terms of such generalized connections. Let U be the domain of a coordinate chart in M. We will always denote coordinates on U by x^i, $i = 1, \ldots, n$. The coordinates on the corresponding bundle chart of T^*M are denoted by (x^i, p_i), $i = 1, \ldots, n$.

Definition 10. A g-connection on (M, Q, h) is a linear generalized connection on T^*M over the bundle map $g : T^*M \to TM$.

Comparing with the notations from the previous section, we see that a g-connection on a sub-Riemannian manifold is a linear ρ-connection with $N = E = T^*M$ and $\rho = g$. Note that with these identifications, the definition of a g-admissible curve, as given in the context of sub-Riemannian geometry, agrees with the notion of a ρ-admissible curve.

Definition 11. A g-admissible curve $\alpha : I \to T^*M$ is said to be an *auto-parallel curve* with respect to a g-connection ∇ if it satisfies $\nabla_\alpha \alpha(t) = 0$ for all $t \in I$. Its base curve $c = \pi \circ \alpha$ is then called a geodesic of ∇.

In coordinates, an auto-parallel curve $\alpha(t) = (x^i(t), p_i(t))$ satisfies the equations

$$\dot{x}^i(t) = g^{ij}(x(t)) p_j(t), \qquad \dot{p}_j(t) = -\Gamma_j^{ik}(x(t)) p_i(t) p_k(t),$$

where g^{ij} and $\Gamma_j^{ik} \in \mathcal{F}(U)$ are the local components of the contravariant tensor field $\bar{g}$ associated to the sub-Riemannian structure (cf. Section 2.1) and the connection coefficients of ∇, respectively. In fact, given a linear g-connection ∇ one can always define a smooth vector field Γ^∇ on T^*M whose integral curves are auto-parallel curves with respect to ∇. In canonical coordinates, this vector field reads:

$$\Gamma^\nabla(x,p) = g^{ij}(x)p_j\frac{\partial}{\partial x^i} - \Gamma_j^{ik}(x)p_ip_k\frac{\partial}{\partial p_j}.$$

A proof of this property follows by standard arguments, and is left to the reader. This implies, in particular, that given any $\alpha_0 \in T^*M$, there exists an auto-parallel curve α passing through α_0. Note that it may happen that two different auto-parallel curves correspond to the same base curve (i.e. may project onto the same geodesic).

Now, we would like to find a g-connection on a sub-Riemannian manifold whose geodesics are precisely the normal extremals. Recalling the definition of a normal extremal (Definition 5), it follows that we will have to look for a g-connection ∇ for which $\Gamma^\nabla = X_H$, where X_H denotes the Hamiltonian vector field corresponding to $H(\alpha_x) = (1/2)\bar{g}(\alpha_x,\alpha_x) \in \mathcal{F}(T^*M)$. A first step in that direction is the construction of a symmetric product associated with a given g-connection, which fully characterizes the geodesics of the g-connection under consideration.

Two linear g-connections ∇ and $\bar\nabla$ have the same geodesics if and only the tensor field $D : \mathcal{X}^*(M)\otimes\mathcal{X}^*(M) \to \mathcal{X}^*(M)$, $(\alpha,\beta) \mapsto \nabla_\alpha\beta - \bar\nabla_\alpha\beta$ is skew-symmetric, or equivalently $D(\alpha,\alpha) \equiv 0$. In local coordinates, the components of D are given by $D_k^{ij} = \Gamma_k^{ij} - \bar\Gamma_k^{ij}$, where Γ_k^{ij} and $\bar\Gamma_k^{ij}$ are the connection coefficients of ∇ and $\bar\nabla$, respectively. We immediately see that D is skew-symmetric iff $\Gamma^\nabla = \Gamma^{\bar\nabla}$, proving the previous statement. Define the *symmetric product* of a connection ∇ as

$$\langle\alpha:\beta\rangle_\nabla := \nabla_\alpha\beta + \nabla_\beta\alpha \quad \text{for } \alpha,\beta \in \mathcal{X}^*(M).$$

Observe that this is not a tensorial quantity, i.e. $\langle\alpha:\beta\rangle_\nabla$ is not $\mathcal{F}(M)$-linear in its arguments. By replacing α by $\alpha+\beta$ in $D(\alpha,\alpha)$ the following lemma is easily proven.

Lemma 12. *The geodesics of a linear g-connection ∇ are completely determined by the symmetric product $\langle\alpha:\beta\rangle_\nabla$ in the sense that, given two g-connections ∇ and $\bar\nabla$, then both have the same geodesics if and only if $\langle\alpha:\beta\rangle_\nabla = \langle\alpha:\beta\rangle_{\bar\nabla}$ for all $\alpha,\beta \in \mathcal{X}^*(M)$.*

In the following we shall construct a symmetric bracket of 1-forms, associated to a sub-Riemannian structure (M,Q,h), which coincides with the symmetric product of a g-connection ∇ on T^*M iff $\Gamma^\nabla = X_H$.

Before proceeding, we first recall that the Levi–Civita connection ∇^G associated to an arbitrary Riemannian metric G is completely determined by the relation:

$$\begin{aligned}2G(\nabla_X^G Y,Z) = {} & X(G(Y,Z)) + Y(G(X,Z)) - Z(G(X,Y)) + G([X,Y],Z)\\ & - G([X,Z],Y) - G(X,[Y,Z])\end{aligned}$$

for all $X,Y,Z \in \mathcal{X}(M)$. This can still be rewritten as

$$2\flat_G(\nabla_X^G Y) = \mathcal{L}_X\flat_G(Y) + \mathcal{L}_Y\flat_G(X) + \flat_G([X,Y]) - \mathrm{d}(G(X,Y)),$$

and the symmetric product of two vector fields X, Y, defined by $\langle X : Y\rangle_{\nabla^G} = \nabla^G_X Y + \nabla^G_Y X$, then satisfies

$$\flat_G(\langle X : Y\rangle_{\nabla^G}) = \mathcal{L}_X \flat_G(Y) + \mathcal{L}_Y \flat_G(X) - \mathrm{d}(G(X, Y)).$$

The right-hand side of this equation now inspires us to propose the following definition of a symmetric bracket of 1-forms on a sub-Riemannian manifold.

Definition 13. The symmetric bracket associated to a sub-Riemannian structure (M, Q, h) is the mapping $\{\cdot, \cdot\} : \mathcal{X}^*(M) \times \mathcal{X}^*(M) \to \mathcal{X}^*(M)$ defined by

$$\{\alpha, \beta\} = \mathcal{L}_{g(\alpha)}\beta + \mathcal{L}_{g(\beta)}\alpha - \mathrm{d}(\bar{g}(\alpha, \beta)).$$

In the following proposition we list some properties of this bracket, the first of which justifies the denomination “symmetric bracket”. The proofs of these properties are straightforward and immediately follow from the above definition.

Proposition 14. *The symmetric bracket satisfies the following properties: for any* α, $\beta \in \mathcal{X}^*(M)$

(1) $\{\alpha, \beta\} = \{\beta, \alpha\}$;
(2) *the bracket is* $\mathbb{R}$-*bilinear*;
(3) $\{f\alpha, \beta\} = g(\beta)(f)\alpha + f\{\alpha, \beta\}$ *with* $f \in \mathcal{F}(M)$;
(4) $\{\alpha, \eta\} = \mathcal{L}_{g(\alpha)}\eta$, *for any* $\eta \in \Gamma(Q^0)$, *and* $\{\alpha, \eta\} = 0$ *if both* α *and* η *belong to* $\Gamma(Q^0)$.

The first three properties justify the following definition.

Definition 15. A g-connection ∇ is said to be normal if the associated symmetric product equals the symmetric bracket, i.e. if $\langle\alpha : \beta\rangle_\nabla = \{\alpha, \beta\}$ holds for all $\alpha, \beta \in \mathcal{X}^*(M)$.

The connection coefficients of a normal g-connection satisfy the relations

$$\Gamma^{ij}_k + \Gamma^{ji}_k = \frac{\partial g^{ij}}{\partial x^k} \quad \forall i, j, k = 1, \ldots, n.$$

We are now going to introduce a special operator, determined by the given distribution Q, which will play an important role later on.

For that purpose, we first recall that, given a regular involutive distribution D on a manifold M, there exists a canonical connection ∇^B on the bundle $D^0 \to M$ over the natural injection $i : D \to TM$, sometimes called the “Bott connection”, defined by: $\nabla^B_X \eta = i_X\,\mathrm{d}\eta$, where $X \in \Gamma(D)$ and $\eta \in \Gamma(D^0)$. Indeed, under the hypothesis that D is involutive, the image of ∇^B is again an element of $\Gamma(D^0)$. This connection was used by Bott et al. [3] to prove, among others, that certain Pontryagin classes of the bundle $D^0 \to M$ are identically zero. However, in the setting of a sub-Riemannian structure (M, Q, h), the distribution Q is assumed not to be involutive and, hence, the 1-form $i_X\,\mathrm{d}\eta$ in general will not belong to $\Gamma(Q^0)$. Nevertheless, this mapping naturally pops up in our approach to characterize normal and abnormal extremals and, therefore, deserves some special attention. More specifically, with any sub-Riemannian structure (M, Q, h) we can associate a mapping δ^B according to

$$\delta^B : \Gamma(Q) \times \Gamma(Q^0) \to \mathcal{X}^*(M), (X, \eta) \mapsto \delta^B_X \eta = i_X \, d\eta.$$

The superscript B is kept to remind us of the fact that this map reduces to the Bott connection in the case of involutive distributions.

Definition 16. Given a sub-Riemannian structure (M, Q, h), a g-connection ∇ is said to be adapted to the bundle Q (shortly Q-adapted) if $\nabla_\alpha \eta = \delta^B_{g(\alpha)} \eta$ for all $\alpha \in \mathcal{X}^*(M)$ and $\eta \in \Gamma(Q^0)$.

For the following theorem, recall the notation introduced in Section 2.1 for the projection operators associated with a Riemannian metric G restricting to h, namely $\tau : T^*M \to \flat_G(Q)$, $\tau^\perp : T^*M \to Q^0$.

Theorem 17. *Let ∇ be a g-connection, then the following statements are equivalent:*

(1) *∇ is a normal g-connection;*
(2) *for all $\alpha \in \mathcal{X}^*(M) : \nabla_\alpha \alpha = (1/2)\{\alpha, \alpha\}$;*
(3) *$\langle \nabla_\alpha X, \beta \rangle + \langle \nabla_\beta X, \alpha \rangle = \langle [g(\alpha), X], \beta \rangle + \langle [g(\beta), X], \alpha \rangle + X(g(\alpha, \beta))$ for all $\alpha, \beta \in \mathcal{X}^*(M)$ and $X \in \mathcal{X}(M)$;*
(4) *$\Gamma^\nabla = X_H$ or, equivalently, every geodesic of ∇ is a normal extremal and vice versa;*
(5) *let G be a Riemannian metric restricting to h and let ∇^G be its Levi–Civita connection, then for all $\alpha \in \mathcal{X}^*(M)$, ∇ satisfies:*

$$\nabla_\alpha \alpha = \nabla^G_{g(\alpha)} \tau(\alpha) + \delta^B_{g(\alpha)} \tau^\perp(\alpha).$$

Note that the right-hand side of (3) agrees with the definition of the symmetrized covariant derivative considered in [18].

Proof. The equivalence of (1) and (2) follows directly from the definition of a normal g-connection, and the equivalence of (1) and (3) follows from $\langle \nabla_\alpha \beta, X \rangle = g(\alpha)(\langle \beta, X \rangle) - \langle \beta, \nabla_\alpha X \rangle$ after some tedious but straightforward calculations.

(2) $\Leftrightarrow$ (4). Choose an arbitrary $\alpha_0 \in T^*M$. Let U be a coordinate neighborhood of $x_0 = \pi_M(\alpha_0)$ and put $\alpha_0 = (x^i_0, p^0_j)$. Then, $\nabla_\alpha \alpha = (1/2)\{\alpha, \alpha\}$ implies, in particular, that the connection coefficients Γ^{ij}_k of ∇ on U satisfy

$$\Gamma^{ij}_k(x_0) p^0_i p^0_j = \frac{1}{2} \frac{\partial g^{ij}}{\partial x^k}(x_0) p^0_i p^0_j.$$

The coordinate expression for the Hamiltonian vector field X_H at α_0 equals:

$$X_H(\alpha_0) = g^{ij}(x_0) p^0_j \left. \frac{\partial}{\partial x^i} \right|_{\alpha_0} - \frac{1}{2} \frac{\partial g^{ij}}{\partial x^k} p^0_i p^0_j \left. \frac{\partial}{\partial p_k} \right|_{\alpha_0}.$$

Recalling the definition of Γ^∇ it is easy to see that $\Gamma^\nabla(\alpha_0) = X_H(\alpha_0)$ for any $\alpha_0 \in T^*M$ if and only if $\nabla_\alpha \alpha = (1/2)\{\alpha, \alpha\}$ for each $\alpha \in \mathcal{X}^*(M)$.

(2) ⇔ (5). Let G be a Riemannian metric restricting to h. Recall the following property of the Levi–Civita connection ∇^G:

$$\flat_G(\langle X : Y\rangle_{\nabla^G} = \mathcal{L}_X \flat_G(Y) + \mathcal{L}_Y \flat_G(X) - \mathrm{d}(G(X, Y)).$$

Putting $X = Y = g(\alpha)$, this equation becomes

$$\flat_G(\nabla^G_{g(\alpha)} g(\alpha)) = \mathcal{L}_{g(\alpha)} \flat_G(g(\alpha)) - \tfrac{1}{2}\,\mathrm{d}(\bar{g}(\alpha, \alpha)).$$

Using the identity $\flat_G(g(\alpha)) = \tau(\alpha)$ derived in the Section 2.1, and taking into account that ∇^G preserves the metric G, i.e. $\nabla^G \circ \flat_G = \flat_G \circ \nabla^G$, we obtain

$$\nabla^G_{g(\alpha)} \tau(\alpha) = \mathcal{L}_{g(\alpha)} \tau(\alpha) - \tfrac{1}{2}\,\mathrm{d}(\bar{g}(\alpha, \alpha)) = \tfrac{1}{2}\{\alpha, \alpha\} - \mathcal{L}_{g(\alpha)} \tau^{\perp}(\alpha).$$

Since $\tau^{\perp}(\alpha) \in \Gamma(Q^0)$ and $g(\alpha) \in \Gamma(Q)$, the last term on the right-hand side reduces to $\delta^B_{g(\alpha)} \tau^{\perp}(\alpha)$, which completes the proof. □

Theorem 17 implies, in particular, that normal g-connections exist. For instance, the mapping ∇ defined by $\nabla_\alpha \beta = \nabla^G_{g(\alpha)} \tau(\beta) + \delta^B_{g(\alpha)} \tau^{\perp}(\beta)$ is a linear g-connection and it is normal, in view of the equivalence of (1) and (5). Moreover, for $\beta \in \Gamma(Q^0)$ we find that $\nabla_\alpha \beta = \delta^B_{g(\alpha)} \beta$, i.e. the connection under consideration is also Q-adapted. Summarizing, we have shown the following result.

Proposition 18. *Given a sub-Riemannian structure (M, Q, h), one can always construct a normal and a Q-adapted g-connection.*

Furthermore, the g-connection constructed gives us a relation between a normal g-connection, the Levi–Civita connection ∇^G of any Riemannian metric restricting to h and the operator δ^B. This relation will be very useful when we study the relation between vakonomic dynamics and nonholonomic mechanics (see Section 6).

In the following theorem, we shall characterize an abnormal extremal in terms of a Q-adapted g-connection. According to Definition 5, a continuous piecewise curve c tangent to Q is an abnormal extremal if there exists a continuous piecewise section ψ of Q^0 along c such that

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_t \langle \psi(t), X(c(t))\rangle = \langle \psi(t), [g(\alpha(t)), X]\rangle \tag{2}$$

holds for an arbitrary chosen piecewise g-admissible curve α projecting onto c, for any $X \in \mathcal{X}(M)$ and for all t, where $\dot{c}(t)$ is defined. We can now state the following interesting result.

Theorem 19. *Given a continuous piecewise curve $c : I \to M$, tangent to Q. There exists a continuous piecewise section of Q^0 along c which is parallel with respect to a Q-adapted g-connection if and only if c is an abnormal extremal.*

Proof. Recall from Section 2.2 that a continuous piecewise curve on $I = [a, b]$ is defined as a continuous map which can be regarded as a concatenation of a finite number of curves

c^i $(i = 1, \dots, k)$, with domain, say $[a_i, a_{i+1}] \subset I$ for $a_1 = a < a_2 < \cdots < a_k < a_{k+1} = b$ and such that $c^i(a_{i+1}) = c^{i+1}(a_{i+1})$.

Let c be an abnormal extremal such that (2) holds. We shall denote the curves associated to α and ψ on the subinterval $[a_i, a_{i+1}]$, by α^i and ψ^i, respectively. Since ψ is continuous, we have $\psi^i(a_{i+1}) = \psi^{i+1}(a_{i+1})$. Then (2) can equivalently be rewritten as

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_t \langle \psi^i(t), X(c^i(t))\rangle = \langle \psi^i(t), [g(\alpha^i(t)), X]\rangle \quad \forall t \in [a_i, a_{i+1}], \quad i = 1, \dots, k.$$

Now, take a Q-adapted g-connection ∇ (which always exists in view of Proposition 18). By definition, ∇ satisfies $\nabla_\beta \eta = \delta^B_{g(\beta)}\eta$ for all $\beta \in \mathcal{X}^*(M)$ and $\eta \in \Gamma(Q^0)$. Now, assume $\nabla_\beta \eta = 0$. This is clearly equivalent to the condition $\langle \nabla_\beta \eta, X\rangle = 0$ for any $X \in \mathcal{X}(M)$ which, in view of the fact that ∇ is Q-adapted, can be rewritten as $\langle \mathcal{L}_{g(\beta)}\eta, X\rangle = 0$ or $g(\beta)(\langle \eta, X\rangle) = \langle \eta, [g(\beta), X]\rangle$. Herewith, we have proven that $\nabla_\beta \eta = 0$ iff $g(\beta)(\langle \eta, X\rangle) = \langle \eta, [g(\beta), X]\rangle$ for any $X \in \mathcal{X}(M)$. This equivalence can be restated in the following way. Given a g-admissible curve α^i, with base curve c^i and ψ^i a section of Q^0 along c^i, then $\nabla_{\alpha^i}\psi^i(t) = 0$ if and only if

$$\frac{\mathrm{d}}{\mathrm{d}t}(\langle \psi^i(t), X(c^i(t))\rangle) = \langle \psi^i(t), [g(\alpha^i(t)), X]\rangle \quad \forall X \in \mathcal{X}(M).$$

Now, $\nabla_{\alpha^i}\psi^i(t) = 0$ for all $t \in [a_i, a_{i+1}]$ and $i = 1, \dots, k$, with $\psi^i(a_{i+1}) = \psi^{i+1}(a_{i+1})$ implies, by definition, that the continuous piecewise section ψ of Q^0 is parallel with respect to the Q-adapted g-connection ∇ (see Section 2.3). This already proves one half of the theorem.

The proof of the converse statement, namely that the existence of a continuous piecewise section ψ of Q^0 along c, satisfying the appropriate conditions, implies that c is an abnormal extremal, simply follows by reversing the above arguments. □

To conclude this section we make some further remarks on normal and Q-adapted g-connections. It is well known that the Levi–Civita connection ∇^G, associated with a Riemannian metric G, is uniquely determined by the properties that it preserves the metric, i.e. $\nabla^G G = 0$, and that its torsion is zero. We would like to consider now metric g-connections ∇ on a sub-Riemannian manifold, i.e. $\nabla \bar{g} = 0$, where $\bar{g}$ is the symmetric contravariant 2-tensor field defined in Section 2.1. From above we know that normal extremals of a sub-Riemannian structure, resp. abnormal extremals, can be characterized as geodesics of a normal g-connection, resp. as parallel transported sections of Q^0 for a Q-adapted g-connection (see Theorem 17, resp. Theorem 19). Therefore, it is natural to look for g-connections that are simultaneously normal and Q-adapted. It has been shown above that such a g-connection always exists, namely $\nabla_\alpha \beta = \nabla^G_{g(\alpha)}\tau(\beta) + \delta^B_{g(\alpha)}\tau^0(\beta)$, with G any Riemannian metric restricting to h. We will prove, however, that no metric g-connection can be found that is also Q-adapted. First we prove an interesting result relating the notion of partial g-connection (see Definition 8) with that of a Q-adapted normal g-connection.

Proposition 20. *Let ∇ be a normal g-connection. Then ∇ is partial if and only if ∇ is Q-adapted.*

Proof. Let ∇ be a normal g-connection, i.e. $\nabla_\alpha\beta + \nabla_\beta\alpha = \{\alpha, \beta\}$ for all $\alpha, \beta \in \mathcal{X}^*(M)$. Suppose ∇ is partial, then for $\beta \in \Gamma(Q^0)$ the previous relation becomes

$$\nabla_\alpha\beta = \{\alpha, \beta\} = \mathcal{L}_{g(\alpha)}\beta = \delta^B_{g(\alpha)}\beta,$$

i.e. ∇ is Q-adapted. Conversely, suppose ∇ is normal and Q-adapted, then $\nabla_\alpha\beta = \{\alpha, \beta\} - \nabla_\beta\alpha$. Let $\alpha \in \Gamma(Q^0)$, then the right-hand side of this equation is zero, and thus $\nabla_\alpha\beta = 0$ for all $\alpha \in \Gamma(Q^0)$ and $\beta \in \mathcal{X}^*(M)$. This proves the proposition. □

We will now describe a general method for constructing normal g-connections.

Let $[\cdot, \cdot] : \mathcal{X}^*(M) \times \mathcal{X}^*(M) \to \mathcal{X}^*(M)$ denote a skew-symmetric bracket that is $\mathbb{R}$-linear in both arguments and satisfies, for any $f \in \mathcal{F}(M)$, $[\alpha, f\beta] = g(\alpha)(f)\beta + f[\alpha, \beta]$. Given such a bracket on $\mathcal{X}^*(M)$, one can define a unique normal g-connection ∇ for which $[\alpha, \beta] = \nabla_\alpha\beta - \nabla_\beta\alpha$, namely:

$$\nabla_\alpha\beta = \tfrac{1}{2}([\alpha, \beta] + \{\alpha, \beta\}).$$

Conversely, given a normal g-connection ∇, one can define a skew-symmetric bracket with the desired properties by putting $[\alpha, \beta] = \nabla_\alpha\beta - \nabla_\beta\alpha$. Henceforth, we shall denote the bracket associated with a normal g-connection ∇ by $[\alpha, \beta]_\nabla$.

As can be easily verified, for a g-connection ∇ which is both normal and Q-adapted, the skew-symmetric bracket satisfies: $[\alpha, \eta]_\nabla = \delta^B_{g(\alpha)}\eta$ for all $\eta \in \Gamma(Q^0)$ and $\alpha \in \mathcal{X}^*(M)$. Therefore, if a Riemannian metric G is chosen, with projections τ and $\tau^\perp$ on $\flat_G(Q)$ and Q^0, respectively, and which restricts to h, this bracket takes the form:

$$[\alpha, \beta]_\nabla = [\tau(\alpha), \tau(\beta)]_\nabla + \delta^B_{g(\alpha)}\tau^\perp(\beta) - \delta^B_{g(\beta)}\tau^\perp(\alpha).$$

We only have to know the value of the bracket acting on sections of $\flat_G(Q) \cong Q$. For example, for the g-connection given by $\nabla_\alpha\beta = \nabla^G_{g(\alpha)}\tau(\beta) + \delta^B_{g(\alpha)}\tau^\perp(\beta)$, the associated bracket becomes

$$[\alpha, \beta]_\nabla = \flat_G([g(\alpha), g(\beta)]) + \delta^B_{g(\alpha)}\tau^\perp(\beta) - \delta^B_{g(\beta)}\tau^\perp(\alpha),$$

where $[g(\alpha), g(\beta)] = \mathcal{L}_{g(\alpha)}g(\beta)$ is the usual Lie bracket on vector fields. Note, however, that there does not seem to exist a "natural" skew-symmetric bracket on $\mathcal{X}^*(M)$, independent of the chosen Riemannian extension G of h, which could be used to identify a "standard" g-connection which is both normal and Q-adapted. One might think of imposing a metric condition in order to completely determine such a ∇, but the following result tells us that it is impossible to find a Q-adapted g-connection which is also metric.

Proposition 21. *A Q-adapted g-connection is not metric.*

Proof. Let ∇ be Q-adapted g-connection. Suppose that ∇ leaves $\bar{g}$ invariant. This can be equivalently rewritten as $g(\nabla_\alpha\beta) = \nabla_\alpha(g(\beta))$ for all $\alpha, \eta \in \mathcal{X}^*(M)$. Let $\eta \in \Gamma(Q^0)$, then, since ∇ is Q-adapted this equation becomes $g(\delta^B_{g(\alpha)}\eta) = 0$ for all $\alpha \in \mathcal{X}^*(M)$ and $\eta \in \Gamma(Q^0)$. However, this is equivalent to saying that Q is involutive. Indeed, from

$g(\delta^B_{g(\alpha)}\eta) = 0$ we have

$$0 = \langle \beta, g(\delta^B_{g(\alpha)}\eta)\rangle = \langle \delta^B_{g(\alpha)}\eta, g(\beta)\rangle = -\langle \eta, [g(\alpha), g(\beta)]\rangle$$

for arbitrary $\alpha, \beta \in \mathcal{X}^*(M)$ and $\eta \in \Gamma(Q^0)$, hence $[g(\alpha), g(\beta)] \in \Gamma(Q)$. □

4. Abnormal extremals

For the remainder of this paper we will always restrict ourselves to curve c that are immersions, i.e. $\dot{c}(t) \neq 0$ for all $t \in \mathrm{Dom}(c)$. Such curves can always, at least locally, be seen as (part of) an integral curve of a smooth vector field (see, e.g. [10, p. 28]). Bearing this in mind, we will establish in this section a geometrical characterization of abnormal extremals on a manifold with a regular, nonintegrable distribution Q. First, we will restrict ourselves to curves that are integral curves of a vector field. Next, we will extend the analysis to general continuous piecewise curves tangent to Q, whose smooth parts are immersions such that they can be regarded as a concatenation of integral curves of vector fields belonging to $\Gamma(Q)$.

Consider a manifold M equipped with a regular distribution Q. Choose an arbitrary sub-Riemannian metric h (e.g. the restriction of some Riemannian metric on M) and let ∇ be a fixed Q-adapted g-connection associated to the sub-Riemannian structure (M, Q, h). (From the previous section we know that such a g-connection can always be found.) Suppose that $c : I \to M$ is a curve tangent to Q, which is (part of) an integral curve of a vector field $X \in \Gamma(Q)$, defined on a neighborhood of $\mathrm{Im}(c)$. In particular, we have that $\dot{c}(t) = X(c(t))$ for all $t \in I$. Then we know that c is an abnormal extremal if there exists a section η of Q^0 along c such that $\nabla_\alpha \eta(t) = 0$ for all $t \in I$, with α a g-admissible curve with base curve c. Let $\{\phi_s\}$ denote the (local) flow of X such that for any fixed $t \in [a, b]$, $\phi_s(c(t)) = c(t + s)$ for all s for which the right-hand side is defined. We denote the dual of the tangent map $T\phi_s$ of ϕ_s by $T^*\phi_s$, i.e. for $\alpha \in T^*_{\alpha_s(x)}M$, $T^*\phi_s(\alpha)$ is the co-vector at x defined by $T^*\phi_s(\alpha)(Y_x) = \alpha(T\phi_s(Y_x))$ for all $Y_x \in T_xM$ (with $x \in \mathrm{Dom}(\phi_s)$). We can now prove the following lemma.

Lemma 22. *Let $c : I \to M$ be an integral curve of $X \in \Gamma(Q)$ and let η be an arbitrary section of Q^0 along c. Then, for any g-admissible curve α with base curve c, the following equation holds:*

$$\nabla_\alpha \eta(t) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} (T^*\phi_s(\eta(t+s))) \quad \forall t \in I.$$

Proof. Fix an arbitrary $t \in I$ and choose a local coordinate neighborhood of M containing the point $c(t)$. Since $\nabla_\alpha \eta(t)$ is independent of the g-admissible curve α projecting onto c, we can choose $\alpha(t) = \bar{\alpha}(c(t))$, where $\bar{\alpha} = \flat_G(X)$ and G is any Riemannian metric restricting to h. From Section 2.1 we know that $g(\bar{\alpha}) = X$, which implies indeed that $\alpha(t) = \bar{\alpha}(c(t))$ is a g-admissible curve with base curve c. In coordinates, $\nabla_\alpha \eta(t)$ reads:

$$\nabla_\alpha \eta(t) = \left(\dot{\eta}_i(t) + \frac{\partial g^{jk}}{\partial x^i}(c(t))\alpha_k(t)\eta_j(t)\right) \mathrm{d}x^i|_{c(t)}.$$

Since, for fixed t and for sufficiently small s, the mapping $s \mapsto T^*\phi_s(\eta(t+s))$ defines a curve in the fiber $T^*_{c(t)}M$, the derivative at $s=0$ is well defined and can be identified with an element of $T^*_{c(t)}M$. In coordinates this curve is given by

$$T^*\phi_s(\eta(t+s)) = \frac{\partial \phi_s^j}{\partial x^i}(c(t))\eta_j(t+s)\,\mathrm{d}x^i|_{c(t)},$$

and its derivative at $s=0$ equals

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(T^*\phi_s(\eta(t+s))) = \left(\dot\eta_i(t) + \frac{\partial X^j}{\partial x^i}(c(t))\eta_j(t)\right)\mathrm{d}x^i|_{c(t)}.$$

Using the fact that $g(\bar\alpha) = X$ this leads to the desired result, since

$$\frac{\partial X^j}{\partial x^i}(c(t))\eta_j(t) = \frac{\partial (g^{jk}\bar\alpha_k)}{\partial x^i}(c(t))\eta_j(t) = \frac{\partial g^{jk}}{\partial x^i}(c(t))\alpha_j(c(t))\eta_j(t),$$

where the second equality follows from $\eta \in \Gamma(Q^0)$. □

Herewith, we derive the following characterization of an abnormal extremal.

Proposition 23. *Let $c : I = [a,b] \to M$ be a curve tangent to Q, such that it is an integral curve of a vector field $X \in \Gamma(Q)$ with flow $\{\phi_s\}$. Then, c is an abnormal extremal if and only if there exists a section η of Q^0, defined along c, such that $\eta(t) = T^*\phi_{-(t-a)}(\eta(a))$ for all $t \in I$.*

Proof. According to Theorem 19, c is an abnormal extremal iff there exists a section of Q^0 along c such that $\nabla_\alpha\eta(t) = 0$, with α a g-admissible curve. Using the preceding lemma, this is still equivalent to

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0}(T^*\phi_s(\eta(t+s))) = 0 \quad \forall t \in I.$$

Acting with the map $T^*\phi_{(t-a)} : T^*_{c(t)}M \to T^*_{c(a)}M$ on both sides of this equation, we obtain the equivalent condition:

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t}(T^*\phi_{(t-a)}(\eta(t))) = 0 \quad \forall t \in I$$

from which it follows that $T^*\phi_{(t-a)}(\eta(t)) = \eta(a)$. □

This characterization of abnormal extremals that are integral curves of a vector field leads us to the following construction. Let c be a curve tangent to Q, with domain $I = [a,b]$, such that it is an integral curve of a vector field $X \in \Gamma(Q)$. For each $t \in I$ consider the subset c_t^*Q of the tangent space $T_{c(t)}M$, given by

$$c_t^*Q = \mathrm{Span}\{T\phi_{-s}(Y_{c(t+s)})| \quad \forall Y \in Q_{c(t+s)}, \quad s \in [a-t, b-t]\}.$$

It is immediately verified that c_t^*Q is in fact a linear subspace of $T_{c(t)}M$. Moreover, it is also easily seen that, for each $t \in I$ and $s \in [a-t, b-t] : T\phi_s(c_t^*Q) = c_{t+s}^*Q$. Therefore, the dimension of the linear space c_t^*Q is independent of t. As an aside of the following theorem it will follow that c_t^*Q only depends on the set $\{\dot{c}(s) = X(c(s))|s \in [a, b]\}$.

Theorem 24. *Let c be a curve tangent to Q with domain $I = [a, b]$, such that it is an integral curve of $X \in \Gamma(Q)$. Then c is an abnormal extremal if and only if $c_a^*Q \neq T_{c(a)}M$.*

Proof. Suppose that $c_a^*Q \neq T_{c(a)}M$, i.e. there exists a nonzero $\eta_a \in (c_a^*Q)^0 \subset T^*_{c(a)}M$. Define a curve η in T^*M along c by $\eta(t) = T^*\phi_{-(t-a)}(\eta_a)$. Note that $\eta(t) \neq 0$ for all t. We now prove that $\eta(t) \in Q^0$ and, hence, $c(t)$ is an abnormal extremal (see Proposition 23). For any $Y_{c(t)} \in Q_{c(t)}$, we have to show that $\langle \eta(t), Y_{c(t)}\rangle = 0$. By definition of $\eta(t)$ this is indeed the case, since

$$\langle \eta(t), Y_{c(t)}\rangle = \langle \eta_a, T\phi_{-(t-a)}(Y_{c(t)})\rangle \quad \text{and} \quad T\phi_{-(t-a)}(Y_{c(t)}) \in c_a^*Q.$$

Conversely, suppose that $c(t)$ is an abnormal extremal, then, again in view of Proposition 23, there exists a section η of Q^0 along c, which does not intersect the zero section, such that $\eta(t) = T^*\phi_{-(t-a)}(\eta(a))$. Since $\eta(t) \in Q^0$, we then have that $\langle \eta(t), Y_{c(t)}\rangle = 0$ for all t and for arbitrary $Y_{c(t)} \in Q_{c(t)}$. This relation can be rewritten as follows:

$$\langle \eta(t), Y_{c(t)}\rangle = \langle \eta(a), T\phi_{-(t-a)}(Y_{c(t)})\rangle = 0,$$

and, hence, we conclude that $0 \neq \eta(a) \in (c_a^*Q)^0$, which completes the proof. □

From the above proof it follows that each element of $(c_a^*Q)^0$ determines a unique section of Q^0 along c by parallel transport with respect to a Q-adapted g-connection, and vice versa. Since for a parallel section η of Q^0 along c the equation $\nabla_\alpha\eta(t) = 0$ only depends on the tangent vector to the base curve c, one may indeed conclude that the space c_a^*Q only depends on $\{\dot{c}(t)|; t \in [a, b]\}$.

Remark 25. Given a vector field $X \in \Gamma(Q)$ such that $X(c(t)) = \dot{c}(t)$, consider the subspace of $T_{c(a)}M$ spanned by $Q_{c(a)}$ and by all tangent vectors of the form $[X, [X, \ldots [X, Y] \ldots](c(a))$ for arbitrary $Y \in \Gamma(Q)$, and let us denote this space by $D_{c(a)}$. It is not difficult to prove that the space spanned by $D_{c(a)}$ is contained in (but, in general differs from) c_a^*Q.

A well-known result concerning abnormal extremals (see, for instance [18]) states that if Q is "strongly bracket generating", i.e. if $T_xM = \mathrm{Span}\{Y(x) + [X, Y'](x)|Y, Y' \in \Gamma(Q)\}$ for every $x \in M$ and $X \in \Gamma(Q)$, then there are no abnormal extremals. Since $\mathrm{Span}\{Y(x) + [X, Y'](x)|Y, Y' \in \Gamma(Q)\} \subset D_{c(a)}$, the previous remark shows that this result is compatible with Theorem 24. At least for the class of curves we are considering here, this result can even be generalized in the following sense. If for some $X \in \Gamma(Q)$ we have that at every point $x \in \mathrm{Dom}(X)$ we have $T_xM = D_x$, then no integral curve of X passing trough the point x can be an abnormal extremal.

So far, we have only characterized those abnormal extremals that can be regarded as integral curves of a vector field tangent to Q. We shall now extend Theorem 24 to the class of abnormal extremals that may be continuous piecewise curves.

Given any curve $c : I = [a, b] \to M$ tangent to Q which is an immersion, then there exists a finite subdivision of I, such that the restriction of c to each subinterval is an integral curve of a vector field tangent to Q (cf. [10, p. 28]). This further implies that, given any continuous piecewise curve $c : I = [a, b] \to M$ tangent to Q, we can apply this property to each smooth part $c^i : [a_i, a_{i+1}] \to M$ of c (for $i = 1, \dots, k$), where we are using the conventions of Section 2.2. More precisely, each sub-curve c^i can be regarded by itself as a concatenation of integral curves of (local) vector fields belonging to $\Gamma(Q)$. For the sake of clarity, we will now consider the simple case of a continuous piecewise curve consisting of a concatenation of two integral curves of vector fields tangent to Q. This will suffice to show how to proceed in the general case of continuous piecewise curves.

Let $c : [a, b] \to M$ be a continuous piecewise curve consisting of two smooth sub-curves $c^1 : [a_1, a_2] \to M$ and $c^2 : [a_2, a_3] \to M$, where $a_1 = a < a_2 < a_3 = b$ and $c^i(t) = c(t)$ for $t \in]a_i, a_{i+1}]$, and whereby we assume that both c^1 and c^2 are integral curves of vector fields $X^1 \in \Gamma(Q)$ and $X^2 \in \Gamma(Q)$, respectively. Denote the local flow of X^i by $\{\phi^i_s\}$, $i = 1, 2$. Since $\dot{c}^i(t) = X^i(c^i(t))$ we have: $c^i(t) = \phi^i_{(t-a_i)}(c^i(a_i))$, $i = 1, 2$. Consider the subspace $c^*_a Q$ of $T_{c(a)}M$ given by

$$c^*_a Q = (c^1)^*_a Q + T\phi^1_{-(a_2-a_1)}((c^2)^*_{a_2} Q),$$

where the spaces $(c^i)^*_{a_i} Q$ are defined as above. Assume that $\eta_a \in (c^*_a Q)^0$. Then the continuous piecewise curve in Q^0 is defined by

$$\eta(t) = \begin{cases} T^*\phi^1_{-(t-a_1)}(\eta_a) & \forall t \in [a_1, a_2], \\ T^*\phi^2_{-(t-a_2)}(T^*\phi^1_{-(a_2-a_1)}(\eta_a)) & \forall t \in [a_2, a_3] \end{cases}$$

is a parallel transported section of Q^0 with respect to a Q-adapted connection (apply Proposition 23 to c^1 and c^2). This proves that if $c^*_a Q \neq T_{c(a)}M$ then c is an abnormal extremal. Conversely, assume that c is an abnormal extremal.

By definition there exist parallel transported sections η^1 and η^2 of Q^0 along c^1 and c^2, respectively, such that $\eta^1(a_2) = \eta^2(a_2)$. Theorem 24 implies that $0 \neq \eta^1(a_1) \in ((c^1)^*_{a_1} Q)^0$ and $0 \neq \eta^2(a_2) \in ((c^2)^*_{a_2} Q)^0$. Since $\eta^2(a_2) = \eta^1(a_2) = T^*\phi^1_{-(a_2-a_1)}(\eta^1(a_1))$, we conclude that

$$0 \neq \eta^1(a_1) \in ((c^1)^*_{a_1} Q)^0 \cap (T\phi^1_{-(a_2-a_1)}((c^2)^*_{a_2} Q))^0 = (c^*_a Q)^0.$$

This reasoning can now be easily extended to the case, where c is a general continuous piecewise curve tangent to Q (for which $\dot{c}(t) \neq 0$ at all points where the derivative is defined). Summarizing, we have derived the following characterization of abnormal extremals within the class of continuous piecewise curves.

Theorem 26. *Let $c : I = [a, b] \to M$ be a continuous piecewise curve tangent to Q, with $\dot{c}(t) \neq 0$ at each point where the derivative exists. Then, there always exists a finite subdivision of I, with endpoints $a_1 = a < a_2 < \cdots < a_\ell < a_{\ell+1} = b$, such that c is a concatenation of integral curves $c^i : [a_i, a_{i+1}] \to M$ of vector fields X^i tangent to Q, with* *flow* $\{\phi^i_s\}$, $i = 1, \ldots, \ell$. *We then have that c is an abnormal extremal if and only if*

$$T_{c(a)}M \neq c^*_a Q := (c^1)^*_{a_1} Q + \sum_{i=2}^{\ell} T\phi^1_{-(a_2-a_1)} \cdots T\phi^{i-1}_{-(a_i-a_i-1)}((c^i)^*_{a_i} Q).$$

Note that, although we have used the theory of g-connections associated to a sub-Riemannian structure for its derivation, the above characterization of abnormal extremals is independent of the choice of a sub-Riemannian metric, but only depends on the geometry of the given distribution Q. This is indeed in full agreement with the notion of abnormal extremal.

Remark 27. While finalizing this paper, we have come across a recent paper by Piccione and Tausk [16], in which, following a different approach, a similar characterization for abnormal extremals was obtained.

We shall now give two examples to illustrate the previous results.

Example 28. Here we consider an example of abnormal extremals, constructed by Montgomery [15]. Let $M = \mathbb{R}^3 - \{0\}$ and let Q be the two-dimensional distribution spanned by the vector fields (expressed in cylindrical coordinates): $X_1 = \partial/\partial r$, $X_2 = \partial/\partial\theta - F(r)(\partial/\partial z)$, where $F(r)$ is a function on M with a single nondegenerate maximum at $r = 1$, i.e. F satisfies:

$$\left.\frac{\mathrm{d}}{\mathrm{d}r} F(r)\right|_{r=1} = 0 \quad \text{and} \quad \left.\frac{\mathrm{d}^2}{\mathrm{d}r^2} F(r)\right|_{r=1} < 0.$$

Such a function can always be constructed (take, for instance, $F(r) = (1/2)r^2 - (1/4)r^4$). The distribution thus defined is everywhere of rank 2, and is differentiable by definition. The flows of X_1, X_2 are denoted by $\{\phi_s\}$, $\{\psi_s\}$, respectively. In particular, we have $\phi_t(r, \theta, z) = (t + r, \theta, z)$, $\psi_t(r, \theta, z) = (r, \theta + t, z - F(r)t)$. Let $c : [0, 1] \to M$ be an integral curve of X_1 through $x_0 = (r_0, \theta_0, z_0)$ at $t = 0$. The subspace

$$c^*_0 Q = \mathrm{Span}\left\{ X_1(x_0), X_2(x_0), \left.\frac{\partial}{\partial\theta}\right|_{x_0} - F(r+t) \left.\frac{\partial}{\partial z}\right|_{x_0} |\forall t \in [0, 1] \right\}.$$

This subspace coincides with the whole tangent space at x, as can be seen from

$$v_r \left.\frac{\partial}{\partial r}\right|_{x_0} + v_\theta \left.\frac{\partial}{\partial\theta}\right|_{x_0} + v_z \left.\frac{\partial}{\partial z}\right|_{x_0} = v_r X_1(x_0) + v_\theta X_2(x_0) + \frac{v_z + v_\theta F(r_0)}{F(r_0 + t) - F(r_0)} (X_2 - \phi^*_t X_2)(x_0),$$

where t is chosen such that $F(r_0 + t) \neq F(r_0)$. So, in view of Theorem 24, one can conclude that an integral curve of X_1 cannot be an abnormal extremal. Let $c' : [0, 1] \to M$ be an integral curve of X_2, with $c'(0) = x_0 = (r_0, \theta_0, z_0)$. Then we have

$$c'^*_0 Q = \mathrm{Span}\left\{ X_1(x_0), X_2(x_0), \left.\frac{\partial}{\partial r}\right|_{x_0} + F'(r_0)t \left.\frac{\partial}{\partial z}\right|_{x_0} |\forall t \in [0, 1] \right\}.$$

If x_0 is a point on the cylinder defined by $r = 1$, then one easily sees that $c_0'^{*}Q \neq T_xM$ since $F'(1) = 0$. Therefore, every helix $c' : [0, 1] \to \mathbb{R}^3 : t \mapsto (1, \theta + t, z - F(1)t)$ is an abnormal extremal, i.e. there exists a section of Q^0 along the curve c' through $x_0 = (1, 0, 0)$ such that

$$\eta(t) := T^*\psi_{-t}(F(1)\,\mathrm{d}\theta|_x + \mathrm{d}z|_x) = F(1)\,\mathrm{d}\theta|_{(1,t,-F(1)t)} + \mathrm{d}z|_{(1,t,-F(1)t)}.$$

Example 29. We now treat an example that was constructed by Liu and Sussmann [14]. Let $M = \mathbb{R}^3$ and Q spanned by $X_1 = \partial/\partial x$, $X_2 = (1 - x)(\partial/\partial y) + x^2(\partial/\partial z)$, where we use Cartesian coordinates x, y, z. The flows $\{\phi_s\}$ of X_1 and $\{\psi_s\}$ of X_2 are given by $\phi_t(x, y, z) = (x + t, y, z)$ and $\psi_t(x, y, z) = (x, (1 - x)t + y, x^2t + z)$. The pull-back of X_1 under ψ_t equals $\psi_t^*X_1 = \partial/\partial x + t(\partial/\partial y) - 2xt(\partial/\partial z)$, and this vector field can be written as a linear combination of X_1, X_2 for any value of t and at all points for which $x = 0$ or 2. Indeed, if $x = 0$, then $\psi_t^*X_1(0, y, z) = X_1(0, y, z) + tX_2(0, y, z)$. If $x = 2$, then $\psi_t^*X_1(2, y, z) = X_1(2, y, z) - tX_2(2, y, z)$. Therefore, each curve defined by $c : I \to M : t \mapsto (x, (1 - x)t + y, x^2t + z)$ for any given point (x, y, z) with $x = 0$ or 2, is an abnormal extremal.

To end this section, we present a construction for the tangent vector to certain variations of a given curve $c : [a, b] \to M$ tangent to Q, that have been used in a derivation of the Maximum Principle in [17]. We shall see that the set of all such tangent vectors determines a subspace of the tangent space T_bM that equals c_b^*Q. Suppose that $c : [a, b] \to M$ is a curve tangent to Q, which is an integral curve of a vector field with flow $\{\phi_t\}$, such that $c(a + t) = \phi_t(c(a))$. The type of variations of c we have in mind here are specified by a triple $(Y, \tau, \delta t)$ with $Y \in \Gamma(Q)$, $\tau \in [a, b]$ and $\delta t \geq 0 \in \mathbb{R}$. Denote the flow of Y by $\{\psi_s\}$. The variation $\tilde{c} : [a, b] \times \mathbb{R} \to M$, associated to the triple $(Y, \tau, \delta t)$ for $\tau \in]a, b]$, is then defined by

$$\tilde{c}(t, \epsilon) = \begin{cases} c(t), & a \leq t \leq \tau - \epsilon\delta t, \\ \psi_{t-(\tau-\epsilon\delta t)}(c(\tau - \epsilon\delta t)), & \tau - \epsilon\delta t \leq t \leq \tau, \\ \phi_{t-\tau}(\psi_{\epsilon\delta t}(c(\tau - \epsilon\delta t))), & \tau \leq t \leq b, \end{cases}$$

which is well defined for ϵ small enough. For $\tau = a$, a slightly different definition for $\tilde{c} : [a, b] \to M$ is needed: $\tilde{c}(t, \epsilon) = \phi_{t-a}(\psi_{\epsilon\delta t}(\phi_{-\epsilon\delta t}(c(a))))$. The tangent vector to any variation $\tilde{c}$ at $(t, \epsilon) = (b, 0)$ equals:

$$V(Y, \tau, \delta t) = T\phi_{b-\tau}(\delta t Y(c(\tau)) - \delta t\dot{c}(\tau)).$$

Since $Y(c(\tau)) - \dot{c}(\tau) \in Q_{c(\tau)}$, the vector $V(Y, \tau, \delta t)$ belongs to c_b^*Q. Even more, the space spanned by all $V(Y, \tau, \delta t)$ with $Y \in \Gamma(Q)$, $\tau \in [a, b]$ and $\delta t \in \mathbb{R}$, equals c_b^*Q. Therefore, the necessary and sufficient condition from Theorem 24 measures the dimensionality of the space spanned by tangent vectors to variations. A more detailed discussion will be presented in a forthcoming paper in which we will construct a natural connection over a bundle map associated with a control problem, which will lead to a weaker version of the Maximum Principle.

5. Normal extremals

In this section, we will make use of Theorem 17 to recover some known results about normal extremals. Consider a sub-Riemannian structure (M, Q, h) and let G be an arbitrary Riemannian metric on M restricting to h. Theorem 17 then says $\nabla_\alpha\alpha(t) = \nabla^G_{\dot c}\tau(\alpha)(t) + \delta^B_{\dot c}\tau^\perp(\alpha)(t)$, where α is a g-admissible curve with base curve c, and ∇ is any normal g-connection. This immediately leads to the following result.

Proposition 30. *Let $c : I \to M$ be a curve tangent to Q that is a geodesic with respect to a Riemannian metric G restricting to h, then c is a normal extremal.*

Proof. The curve c is a normal extremal if there exists a g-admissible curve α with base curve c, which is auto-parallel with respect to a normal g-connection ∇. Since $c : I \to M$ is a geodesic with respect to G, i.e. $\nabla^G_{\dot c}\dot c(t) = 0\ \forall t \in I$, we know from Section 2.1 that $\alpha = \flat_G(c)$ is a g-admissible curve with base curve c for which $\tau(\alpha) = \alpha$ or $\tau^\perp(\alpha) = 0$. It then follows that $\nabla_\alpha\alpha(t) = 0$ since $\nabla_\alpha\alpha(t) = \nabla^G_{\dot c}\tau(\alpha)(t) = \flat_G(\nabla^G_{\dot c}\dot c(t)) = 0$. □

Let $c : I = [a, b] \to M$ be a normal extremal. Then there exists a g-admissible curve α which is auto-parallel with respect to a normal g-connection. Given any $t_0 \in I$, then one can always find a 1-form $\bar\alpha$ and a compact subinterval J of I containing t_0, such that $\bar\alpha(c(t)) = \alpha(t)$ for all $t \in J$ and $c(J)$ is contained in a coordinate neighborhood U. We will now construct a local Riemannian metric G restricting to h on Q such that $c|_J$ is a geodesic with respect to this Riemannian metric.

Since $g(\bar\alpha) \neq 0$, one can construct a local basis of $\mathcal{X}^*(U)$, namely $\{\bar\alpha = \beta^1, \beta^2, \dots, \beta^n\}$, such that $\beta^{k+1}, \dots, \beta^n$ determine a local basis for $\Gamma(Q^0)$, defined on U. Let $\{X_1, \dots, X_n\}$ denote the dual basis of $\mathcal{X}(U)$. Then the vector fields X_j, for $j = 1, \dots, k$, form a local basis for $\Gamma(Q)$, since $\langle\beta^i, X_j\rangle \equiv 0$ for $i = k+1, \dots, b$. We can now define a Riemannian metric G on U, restricting to h, as in Section 2.1, i.e. for arbitrary vector fields Y and Z on U,

$$G(x)(Y, Z) = \sum_{r,s=1}^{k} Y^r Z^s h(x)(X_r(x), X_s(x)) + \sum_{r=k+1}^{n} Y^r Z^r,$$

where we have put $Y(x) = Y^r X_r(x)$ and $Z(x) = Z^r X_r(x)$ for some $Y^r, Z^r \in \mathbb{R}$ $(r = 1, \dots, n)$. From the definition of G we can derive that $Q^\perp$ is spanned by $\{X_{k+1}, \dots, X_n\}$ or $\tau^\perp(\bar\alpha) = 0$, implying that $\tau^\perp(\alpha(t)) = 0$ or $\flat_G(\dot c(t)) = \alpha(t)$. From $\nabla_\alpha\alpha(t) = 0$ and $\tau^\perp(\alpha(t)) = 0$ we obtain $\nabla^G_{\dot c}\dot c(t) = 0$ for any $t \in J$.

Proposition 31. *Let $c : I \to M$ be a normal extremal. Then for any $t \in I$ there exists a compact neighborhood J of t such that c restricted to J is a geodesic with respect to some Riemannian metric restricting to h on Q.*

This proves, in particular, that a normal extremal is locally length minimizing.

Let c be a normal extremal and let ∇ be a normal and Q-adapted g-connection (recall that such a ∇ always exists). Suppose that c is degenerate in the following sense: there exist two g-admissible curves α, β with base curve c, such that $\nabla_\alpha\alpha(t) = \nabla_\beta\beta(t) = 0$.

We will now see that c is then also an abnormal extremal. We have proven before that a normal and Q-adapted connection is partial, i.e. $\nabla_\alpha = \nabla_\beta$ if $g(\alpha) = g(\beta)$. Therefore, one obtains that $\nabla_\alpha(\alpha - \beta)(t) = 0$. Since $g(\alpha(t) - \beta(t)) = 0$, or $\eta(t) = (\alpha - \beta)(t) \in Q^0$ for all t, η is a parallel transported section along α, lying entirely in Q^0 and, hence, c is an abnormal extremal. Conversely, assume that c is a normal extremal, i.e. c is the base curve of an auto-parallel curve α with respect to ∇, and that c is also an abnormal extremal. Let η denote a parallel transported section along α lying in Q^0. Then, using the same arguments as before, $\alpha + \eta$ is also an auto-parallel curve with base curve c. We can conclude that curves that are both normal and abnormal are degenerate in the sense that they admit more than one g-admissible curve that is auto-parallel.

6. Vakonomic dynamics and nonholonomic mechanics

As a natural consequence of the approach to sub-Riemannian structures in terms of generalized connections, we will see how to establish coordinate independent conditions for the motions of a free mechanical system subjected to linear nonholonomic constraints to be normal extremals with respect to the associated sub-Riemannian structure, and vice versa. We first give a definition of what we understand under a free mechanical systems subjected to linear nonholonomic constraints (shortly free nonholonomic mechanical system) and the associated sub-Riemannian structure.

Assume that a manifold M is equipped with a nonintegrable regular distribution Q on M and a Riemannian metric G. A free mechanical system with linear nonholonomic constraint Q consists of a free particle with Lagrangian $L(v) = (1/2)G(v, v) \in \mathcal{F}(TM)$, subjected to the constraint $v \in Q$. (“Free” refers here to the absence of external forces.) The problem of determining the *dynamics of the free nonholonomic mechanical system* then consists in finding the solutions of the following equation (see [1,5])

$$\pi(\nabla^G_{\dot c}\dot c(t)) = 0 \quad \text{and} \quad \dot c(t) \in Q \quad \forall t,$$

where π is the orthogonal projection of TM onto Q with respect to G and ∇^G the Levi–Civita connection associated with G. The associated sub-Riemannian structure is given by (M, Q, h_G), with h_G the restriction of G to Q.

In [12], we have constructed a unique generalized connection ∇^{nh} over the bundle map $i : Q \hookrightarrow TM$ on the linear bundle Q, namely: $\nabla^{nh}_X Y = \pi(\nabla^G_X Y)$ (we have identified $X \in \Gamma(Q)$) with $i \circ X \in \mathcal{X}(M)$). The i-connection ∇^{nh} preserves the sub-Riemannian metric h_G on Q, i.e. $\nabla^{nh}_X h_G = 0$ for any $X \in \Gamma(Q)$, and satisfies $\nabla^{nh}_X Y - \nabla^{nh}_Y X - \pi[X, Y] = 0$ for all $X, Y \in \Gamma(Q)$. One can prove that ∇^{nh} is completely determined by these two properties. In this setting, the i-admissible curves are precisely curves tangent to Q. Therefore, a motion c of the free nonholonomic mechanical system is characterized by the condition that $\nabla^{nh}_{\dot c}\dot c(t) = 0$ for all t.

The *vakonomic dynamical problem*, associated with the free particle with linear nonholonomic constraints, consists in finding normal extremals with respect to the associated sub-Riemannian structure (M, Q, h_G). It is interesting to compare the solutions o the nonholonomic mechanical problem with the solutions of the vakonomic dynamical problem, because the equations of motion for the mechanical problem are derived by means of d'Alembert's principle, whereas the normal extremals are derived from a variational principle. This has been discussed for more general Lagrangian systems by Cortés et al. [7]. For the free particle case, we shall present here an alternative (coordinate free) approach.

Definition 32. Given a Riemannian metric G and a regular distribution Q on a manifold M. We can then define the following two tensorial operators:

$$\Pi^G : \Gamma(Q) \otimes \Gamma(Q) \to \Gamma(Q^\perp), (X, Y) \mapsto \pi^\perp(\nabla^G_X Y),$$
$$\Pi^B : \Gamma(Q) \otimes \Gamma(Q^0) \to \Gamma((Q^\perp)^0), (X, \eta) \mapsto \tau(\delta^B_X \eta).$$

It is indeed easily seen that both Π^G and Π^B are $\mathcal{F}(M)$-bilinear in their arguments and, hence, their action can be defined pointwise, with expressions like $\Pi^G(X_x, Y_x)$ and $\Pi^B(X_x, \eta_x)$, for $X_x, Y_x \in Q_x$ and $\eta_x \in Q^0$, having an obvious and unambiguous meaning.

The operator Π^B is related to the "curvature" of the distribution Q as follows: let $X, Y \in \Gamma(Q)$, then one has:

$$\langle \Pi^B(X, \eta), Y\rangle = \langle \delta^B_X \eta, Y\rangle = -\langle \eta, [X, Y]\rangle \quad \text{for any } \eta \in \Gamma(Q^0).$$

Thus, $\Pi^B \equiv 0$ if and only if Q is involutive. The following lemma shows the importance of these tensors. First, define a linear connection $\tilde{\nabla}^B$ over $i : Q \hookrightarrow TM$ on the bundle Q^0 by the prescription $\tilde{\nabla}^B_X \eta = \tau^\perp(\delta^B_X \eta)$ with $X \in \Gamma(Q)$ and $\eta \in \Gamma(Q^0)$.

Lemma 33. *Given a Riemannian metric G and a regular distribution Q on a manifold M. Assume that $c : I = [a, b] \to M$ is a curve tangent to Q and let ∇ be a Q-adapted g-connection with respect to the associated sub-Riemannian structure (M, Q, h_G). Then, the following properties hold:*

(1) *Given $Y_a \in Q_{c(a)}$, denote the parallel transported curves along c, with initial point Y_a, with respect to ∇^{nh}, resp. ∇^G, by $\tilde{Y}(t)$, by $Y(t)$. Then $\tilde{Y}(t) = Y(t)$ for all t, if and only if $\Pi^G(\dot{c}(t), \tilde{Y}(t)) = 0$ for all $t \in I$.*
(2) *Given $\eta_a \in Q^0_{c(a)}$, denote the parallel transported curves along c, with initial point η_a, with respect to $\tilde{\nabla}^B$, resp. ∇ by $\tilde{\eta}(t)$, resp. $\eta(t)$. Then $\tilde{\eta}(t) = \eta(t)$ if and only if $\Pi^B(\dot{c}(t), \tilde{\eta}(t)) = 0$.*

Proof. (1) From the definition of Π^G it follows that, given any section $\tilde{Z}(t)$ of Q along c, the following equation holds: $\nabla^{nh}_{\dot{c}} \tilde{Z}(t) = \nabla^G_{\dot{c}} \tilde{Z}(t) - \Pi^G(\dot{c}(t), \tilde{Z}(t))$. Assume that $\tilde{Z}(t) = \tilde{Y}(t) = Y(t)$, then we have $\Pi^G(\dot{c}(t), \tilde{Y}(t)) = 0$. This already proves the statement in direction. The converse follows from the fact that parallel transported curves with respect to any connection are uniquely determined by their initial conditions.

The proof of (2) follows from similar arguments. □

Note that property (2) of the previous lemma gives necessary and sufficient conditions for the existence of curves that are abnormal extremals, i.e.: c is an abnormal extremal if and only if there exists a parallel transported section $\tilde{\eta}$ of Q^0 along c with respect to $\tilde{\nabla}^B$

such that, in addition, $\Pi^B(\dot{c}(t), \tilde{\eta}(t)) = 0$ for all t. We shall now investigate some further properties of the operators Π^B and Π^G.

Definition 34. For $x \in M$, let X_x be a nonzero element of Q_x. Define a subspace of T_xM as follows:

$$Q_x + [X, Q_x] = \mathrm{Span}\{Y(x) + [\tilde{X}, Y'](x) | Y, Y' \in \Gamma(Q); \tilde{X} \in \Gamma(Q) \\ \text{with } \tilde{X}(x) = X\}.$$

As a side result of the following lemma, it will be seen that the space $Q_x + [X, Q_x]$ is independent of the extension $\tilde{x}$ of X_x used in its definition and, hence, also justifies the notation.

Lemma 35. *Let $\eta_x \in Q_x^0$ and $X_x \in Q_x$ for some $x \in M$. Then $\Pi^B(X_x, \eta_x) = 0$ if and only if $\eta \in (Q_x + [X, Q_x])^0$.*

Proof. Let $\Pi^B(X, \eta) = 0$. This is equivalent to $\langle \eta, [\tilde{X}, Y'](x)\rangle = 0$ for any $\tilde{X}, Y' \in \Gamma(Q)$ with $\tilde{X}(x) = X_x$. Since $\eta_x \in Q_x^0$, we may conclude that $\eta_x \in (Q_x + [X, Q_x])^0$. The converse follows by reversing the previous arguments. □

Another useful property is given by the following lemma.

Lemma 36. *Let M be a manifold with a Riemannian metric G and a regular nonintegrable distribution Q, and consider the associated sub-Riemannian structure (M, Q, h_G). Let ∇ be a normal g-connection. We then have for $\alpha \in \mathcal{X}^*(M)$ that $\nabla_\alpha \alpha = 0$ if and only if*

$$\flat_G(\nabla^{nh}_{g(\alpha)} g(\alpha)) = -\Pi^B(g(\alpha), \tau^\perp(\alpha)) \quad \text{and} \quad \tilde{\nabla}^B_{g(\alpha)} \tau^\perp(\alpha) = -\flat_G(\Pi^G(g(\alpha), g(\alpha))).$$

Proof. From Theorem 17 one has that $\nabla_\alpha \alpha = 0$ if and only if $\nabla^G_{g(\alpha)} \tau(\alpha) + \delta^B_{g(\alpha)} \tau^\perp(\alpha) = 0$. Using the following relations:

$$\begin{aligned} &\tau(\alpha) = \flat_G(g(\alpha)), \qquad \nabla^G \circ \flat_G = \flat_G \circ \nabla^G, \\ &\nabla^G_{g(\alpha)} g(\alpha) = \nabla^{nh}_{g(\alpha)} g(\alpha) + \Pi^G(g(\alpha), g(\alpha)), \\ &\delta^B_{g(\alpha)} \tau^0(\alpha) = \tilde{\nabla}^B_{g(\alpha)} \tau^0(\alpha) + \Pi^B(g(\alpha), \tau^0(\alpha)) \end{aligned}$$

together with the fact that $T^*M = \flat_G(Q) \oplus Q^0$ and $Q^0 \cong \flat_G(Q^\perp)$, the equivalence is immediately proven. □

The previous lemmas can now be used to derive necessary and sufficient conditions for a motion of a free nonholonomic mechanical system to be normal extremals and vice versa. Let M again be a manifold with a Riemannian metric G and a regular nonintegrable distribution Q.

Proposition 37. *A solution $c : [a, b] \to M$ of a free nonholonomic system determined by the triple (M, Q, G) is a solution of the corresponding vakonomic problem, and vice versa,* *if and only if there exists a section η of Q^0 along c such that*

$$\tilde{\nabla}^B_{\dot{c}}\eta(t) = -\flat_G(\Pi^G(\dot{c}(t), \dot{c}(t))), \tag{3}$$

and such that, in addition $\eta(t) \in (Q_{c(t)} + [\dot{c}(t), Q_{c(t)}])^0$ for all t.

Proof. The condition for any g-admissible curve $\alpha(t) = \flat_G(\dot{c}(t)) + \eta(t)$ with base curve c (where $\eta(t)$ is any section of Q^0 along c) to be parallel transported with respect to a normal g-connection is that $\nabla_\alpha \alpha(t) = 0$. This can equivalently be written as

$$\flat_G(\nabla^{nh}_{\dot{c}}\dot{c}(t)) = -\Pi^B(\dot{c}(t), \eta(t)) \quad \text{and} \quad \tilde{\nabla}^B_{\dot{c}}\eta(t) = -\flat_G(\Pi^G(\dot{c}(t), \dot{c}(t))).$$

Thus, $\nabla^{nh}_{\dot{c}}\dot{c}(t) = 0$ if and only if $\Pi^B(\dot{c}(t), \eta(t)) = 0$, where $\eta(t)$ is a solution of $\tilde{\nabla}^B_{\dot{c}}\eta(t) = -\flat_G(\Pi^G(\dot{c}(t), \dot{c}(t)))$. □

Remark 38. Given any η_0 in $(Q_{c(a)} + [\dot{c}(a), Q_{c(a)}])^0$ then (3) always admits a solution, $\eta(t)$ with initial condition $\eta(a) = \eta_0$. The obstruction for c to be simultaneously a motion of the nonholonomic mechanical system and a solution to the vakonomic dynamical problem, lies in the fact that $\eta(t)$ should belong to $(Q_{c(t)} + [\dot{c}(t), Q_{c(t)}])^0$ for all t, this is not guaranteed by the fact that $\eta(t)$ is a solution of (3). The search for geometric conditions for solutions $\eta(t)$ of this equation to remain in $(Q_{c(t)} + [\dot{c}(t), Q_{c(t)}])^0$ for all t, is left for future work.

Acknowledgements

This work has been supported by a grant from the “Bijzonder onderzoeks-fonds” of Ghent University. Special thanks goes to F. Cantrijn for useful discussions and support.

References

[1] A.M. Bloch, P.E. Crouch, Newton’s law and integrability of nonholonomic systems, SIAM J. Contr. Optim. 36 (1998) 2020–2039.
[2] F. Brickell, R.S. Clark, Differentiable Manifolds: An Introduction, Van Nostrand Reinhold, London, 1970.
[3] R. Bott, S. Gitler, I.M. James, Lectures on Algebraic and Differential Topolgy, Lecture Notes in Mathematics, Vol. 279, Springer, Berlin, 1972.
[4] F. Cantrijn, B. Langerock, Generalised connections over a vector bundle map, Diff. Geom. Appl. (2003), in press.
[5] F. Cantrijn, J. Cortés, M. de Léon, M. Martín de Diego, On the geometry of generalized Chaplygin system, Math. Proc. Camb. Phil. Soc. 132 (2002) 323–351.
[6] W.L. Chow, Über systeme von linearen partiellen differentialgleichungen erseter ordnung, Math. Ann. 117 (1939) 98–105.
[7] J. Cortés, M. de Léon, D. Martín de Diego, S. Martínez, Geometric description of vakonomic and nonholonomic dynamics. Comparison of solutions, SIAM J. Control Optim. 41 (5) (2003) 1389–1412.
[8] R.L. Fernandes, Connections in Poisson geometry. I. Holonomy and invariants, J. Diff. Geom. 54 (2) (2000) 303–365.
[9] R.L. Fernandes, Lie Algebroids, Holonomy and Characteristic Classes, Adv. Math. 170 (1) (2002) 119–179.
[10] S. Helgason, Differential Geometry, Lie groups, and Symmetric Spaces, Academic Press, New York, 1978.
[11] S. Kobayashi, K. Nomizu, Foundations of Differential Geometry, Vols. I and II, Interscience, London, 1963.

[12] B. Langerock, Nonholonomic mechanics and connections over a bundle map, J. Phys. A 34 (2001) L609–L615.
[13] P. Libermann, C.-M. Marle, Symplectic Geometry and Analytical Mechanics, Reidel, Dortrecht, 1987.
[14] W. Liu, H.J. Sussmann, Shortest paths for sub Riemannian metrics on rank two distributions, Memoirs AMS 118 (1995).
[15] R. Montogomery, Abnormal minimizers, SIAM J. Contr. Optim. 32 (1994) 1605–1620.
[16] P. Piccione, D.V. Tausk, Variational aspects of the geodesics problem in sub-Riemannian geometry, J. Geom. Phys. 39 (2001) 183–206.
[17] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkelidze, E.F. Mischenko, The Mathematical Theory of Optimal Processes, Wiley/Interscience, New York, 1962.
[18] R.S. Strichartz, Sub-Riemannian geometry, J. Diff. Geom. 24 (1986) 221–263.
[19] R.S. Strichartz, Corrections to sub-Riemannian geometry, J. Diff. Geom. 30 (1989) 595–596.
[20] H.J. Sussmann, An introduction to the coordinate-free maximum principle, in: B. Jakubcyzk, W. Respondek (Eds.), Geometry of Feedback and Optimal Control, Marcel Dekker, New York, 1997, pp. 463–557.