Geometric aspects of the maximum principle in control theory

Bavo Langerock
Ghent University, Department of Mathematical Physics and Astronomy,
Krijgslaan 281 S9, 9000 Gent
email: bavo.langerock@rug.ac.be

Abstract— **In a previous paper [1] we have proven a geometric formulation of the maximum principle for time dependent optimal control problems with fixed endpoint conditions. As a side result we obtained a necessary and sufficient condition for a control to be a (strictly) abnormal extremal. As an application of that result we consider in this paper two known examples of optimal control systems having abnormal extremals.**

Keywords— **geometric control theory, abnormal extremals.**

I. Introduction

The development of optimal control theory in a differential geometric setting has been carried out, for instance, by H.J. Sussmann in [2], where a coordinate-free formulation of the maximum principle is given. Many problems that are encountered in differential geometry, can be related to control systems . For instance, the problem of characterising length minimising curves in sub-Riemannian geometry (see [3], [4], [5]) has become one of the standard examples in "geometric optimal control theory".

The relevance of creating a differential geometric framework for studying control theory becomes apparent when we consider, for instance, some work of H.J. Sussmann in [6], where he used the theory of integrability of generalised distributions in order to study some controllability problems.

In [1] we have given a proof of the coordinate-free maximum principle for (time-dependent) optimal control systems with fixed endpoint conditions, relying on the approach developed by L.S. Pontryagin et al. in [7]. As a side result of our approach, we were able to give some necessary and sufficient conditions for the existence of what are called *(strictly) abnormal extremals* (for an example of a strictly abnormal extremal, we refer to [8]). In this paper we wish to apply these conditions to two known examples in sub-Riemannian geometry.

In Section II we briefly recall the differential geometric framework formulated in [1] and sketch a proof of the maximum principle in this setting. The definitions and results from Section II should make the necessary and sufficient condition for a control to be an abnormal extremal, formulated in Section III, more accessible to the reader. These conditions are then applied in Section IV to two known examples. It turns out however, that these conditions are rather difficult to compute and in Section V we discuss how these conditions can be weakened to yield sufficient conditions that are much easier to compute.

II. A differential geometric framework for control theory

We now proceed towards the definition of a geometric control structure. It should be noted that we impose rather strong smoothness conditions. However, it occurs to us that there is a sufficiently large and relevant class of control problems that fit within the framework presented below (see [2] for a different approach).

Definition 1. A *geometric control structure* is a triple (τ, ν, ρ) consisting of: (i) a fibre bundle $\tau : M \to \mathbb{R}$ over the real line, where M is called the event space and with typical fibre Q, which is referred to as the configuration manifold, (ii) a fibre bundle $\nu : U \to M$, called the control space, and (iii) a bundle morphism $\rho : U \to J^1\tau$ over the identity on M, such that $\tau_{1,0} \circ \rho = \nu$.

In the above definition, the first jet space of τ is denoted by $J^1\tau$ and the projections $J^1\tau \to \mathbb{R}$ and $J^1\tau \to M$ are denoted by, respectively, τ_1 and $\tau_{1,0}$. The map ρ is called the anchor map and makes the following diagram commutative.

$$\begin{tikzcd} U \arrow[r, "\rho"] \arrow[d, "\nu"'] & J^1\tau \arrow[dl, "\tau_{1,0}"] \\ M \arrow[d, "\tau"'] & \\ \mathbb{R} & \end{tikzcd}$$

A section $c : I \to U$ of $\tau \circ \nu$ is called a *smooth control* if $\rho \circ c = j^1\tilde{c}$, where $\tilde{c} = \nu \circ c$ is called the *smooth base* of the control c.

In order to fix the ideas we will first elaborate on the notion of smooth controls. Fix a bundle adapted coordinate chart on M and let $(t, q^1, \ldots, q^n)$ (where $\dim Q = n$) denote the associated coordinate functions, i.e. the projection τ equals the function t. Similarly, we consider an adapted coordinate chart of the control space U, with coordinate functions $(t, q^1, \ldots, q^n, u^1, \ldots, u^k)$ (with $\dim U = 1 + n + k$). The map ρ is represented by $\rho(t, q, u) = \partial/\partial t + \gamma^i(t, q, u)\partial/\partial q^i$. A smooth control c, is a section of $\tau \circ \nu$ which is locally represented by n functions $q^i(t)$, $i = 1, \ldots, n$ and k functions $u^a(t)$, $a = 1, \ldots, k$, and has to satisfy, by definition, the following equation:

$$\dot{q}^i(t) = \gamma^i(t, q^i(t), u^a(t)).$$

The above equations are easily recognised (see [7]) as the "law of motion" that occurs in standard control theory.

It turns out, however, (see also [7]) that the class of smooth controls should be further extended to sections admitting (a finite number of) discontinuities in the form of certain 'jumps' in the fibres of ν, such that the corresponding base section is piecewise smooth. For instance, assume that $c_1 : [a,b] \to U$ and $c_2 : [b,c] \to U$ are two smooth controls with respective bases $\tilde{c}_1$ and $\tilde{c}_2$, such that $\tilde{c}_1(b) = \tilde{c}_2(b)$. The composite control $c_2 \cdot c_1 : [a,c] \to U$ of c_1 and c_2 is defined by:

$$c_2 \cdot c_1(t) = \begin{cases} c_1(t) & t \in [a,b], \\ c_2(t) & t \in]b,c]. \end{cases}$$

It is readily seen that $c_2 \cdot c_1$ is (in general) discontinuous at $t = b$, but that the base $\nu \circ (c_2 \cdot c_1)$ is continuous. This definition can easily be extended to any finite number of smooth controls, yielding what we shall call in general a *control* (a detailed definition can be found in [1]). We say that a control $c : [a,b] \to U$, with base $\tilde{c}$, takes x to y if $\tilde{c}(a) = x$ and $\tilde{c}(b) = y$, with $x, y \in M$. The set of points R_x, containing all points $y \in M$ such that there exists a control taking x to y is called the set of *reachable points from* x.

We now introduce the notion of optimality. Assume that *a cost function* $L \in C^\infty(U)$ is given. With any control $c : [a,b] \to M$ we are now able to define its *cost* $\mathcal{J}(c)$:

$$\mathcal{J}(c) = \int_a^b L(c(t))dt.$$

A control $c : I \to U$ taking x to y is said to be *optimal* if, given any other control $c' : I \to U$ taking x to y, then

$$\mathcal{J}(c) \leq \mathcal{J}(c').$$

The problem of finding the optimal controls taking x to y is called an *optimal control problem*.

In order to obtain necessary conditions for a control $c : I = [a,b] \to U$ taking x to y to be optimal, we introduced in [1] the notion of *variations of* c and, subsequently, the notion of a *tangent vector to such a variation of* c. A tangent vector to a variation is a tangent vector in $T_{\tilde{c}(b)}M$. It turns out that the set of tangent vectors to variations of the given control c carries the structure of a convex cone in the linear space $T_{\tilde{c}(b)}M$. The main theorem in [1] says that a curve in M, through y, whose tangent vector at y belongs to the interior of this cone, is locally contained in the set of reachable points from x. We will briefly recall that result. To fix the above ideas, we first give some additional definitions.

The definition of a variation of a control is based on the fact that the base curve $\tilde{c}$ of any control $c : [a,b] \to U$ can be written as a *concatenation of integral curves* of vector fields of the form $\mathbf{T} \circ \rho \circ \sigma$, where σ is a section of ν, i.e. $\sigma \in \Gamma(\nu)$, and where $\mathbf{T} : J^1\tau \to TM$ denotes the total time derivative, defined by $\mathbf{T}(j^1_t \tilde{c}') = \dot{\tilde{c}}'(t)$ for any $c' \in \Gamma(\tau)$. Assume that $(\sigma_\ell, \ldots, \sigma_1)$ denotes an ordered family of sections of ν and let $\{\phi^i_t\}$ denote the flow of the vector field $X_i = \mathbf{T} \circ \rho \circ \sigma_i$. Let $(t_\ell, \ldots, t_1) \in \mathbb{R}^\ell$ be given such that $t_i \geq 0$, and define $\tilde{c}_1 : [a_0, a_1] \to M$ as the integral curve of X_1 through a given point $x = \tilde{c}(a_0)$, where $a_1 - a_0 = t_1$. Then, define $\tilde{c}_i : [a_{i-1}, a_i] \to M$ recursively as the integral curve of X_i through $\tilde{c}_{i-1}(a_{i-1})$, for $i = 1, \ldots, \ell$ and $a_i - a_{i-1} = t_i$. It is easily seen that, by definition, the curves $c_i(t) = \sigma_i(\tilde{c}_i(t))$ are smooth controls, and that the composition $c_\ell \cdot \ldots \cdot c_1$ is also a control whose base curve is the concatenation of integral curves of the vector fields X_i, $i = 1, \ldots, \ell$. On the other hand, using the fact that the base curve of a control is an immersion, i.e. $\dot{\tilde{c}}(t) \neq 0$, one can prove, using similar arguments as in [9], that the converse also holds, i.e. given any control c taking x to y, then there exists an ordered family of sections of ν, say $(\sigma_\ell, \ldots, \sigma_1)$ and some $(t_\ell, \ldots, t_1) \in \mathbb{R}^\ell_+$ such that $c = c_\ell \cdot \ldots \cdot c_1$ holds.

We now define the *variation cone* associated to a control $c : [a,b] \to U$ taking x to y. Assume that for a given control $c : [a,b] \to U$ a family of sections $(\sigma_\ell, \ldots, \sigma_1)$ of ν and an ℓ-tuple $(t_\ell, \ldots, t_1)$ are held fixed such that, with the previous conventions, $c = c_\ell \cdot \ldots \cdot c_1$ holds. Then, the variation cone $C_y R_x$ in $T_y M$ is the convex cone containing all linear combinations with positive coefficients of tangent vectors of the form

$$T\phi^\ell_{t_\ell} \circ \cdots \circ T\phi^i_{t_i - \epsilon}\big(\mathbf{T}(\rho(u))\big),$$

and

$$T\phi^\ell_{t_\ell} \circ \cdots \circ T\phi^i_{t_i - \epsilon}\Big(-\mathbf{T}\big(\rho(c(a_{i-1} + \epsilon))\big)\Big)$$

with $0 < \epsilon \leq t_i$, $i = 1, \ldots, \ell$ and $u \in U_{\tilde{c}(a_{i-1}+\epsilon)}$. The necessary and sufficient conditions for a control to be an abnormal extremal will be expressed in terms of a subcone of $C_y R_x$, called the *vertical variation cone* $V_y R_x$, which is defined as the convex cone containing all linear combinations with positive coefficients of tangent vectors of the form

$$T\phi^\ell_{t_\ell} \circ \cdots \circ T\phi^i_{t_i - \epsilon}\Big(\mathbf{T}(\rho(u)) - \mathbf{T}\big(\rho(c(a_{i-1} + \epsilon))\big)\Big),$$

with $0 < \epsilon \leq t_i$, $i = 1, \ldots, \ell$ and $u \in U_{\tilde{c}(a_{i-1}+\epsilon)}$. This cone is contained in $V_y\tau = \ker T_y\tau$ since $T\tau(V_y R_x) = 0$, justifying the above definition.

The importance of the above definition of the variation cone $C_y R_x$ lies in the fact that its interior "generates" the set of reachable points in a neighbourhood of y. The following theorem, taken from [1], shows how this is accomplished.

Theorem 1. *Let $\zeta(t) : [0,1] \to M$ denote a curve with $\zeta(0) = y$ and $\dot{\zeta}(0)$ in the interior of $C_y R_x \subset T_y M$. Then there exists an $\epsilon \in]0,1]$ such that $\zeta(t) \in R_x$ for all $0 \leq t \leq \epsilon$.*

The key observation made by L.S. Pontryagin [7], was that with a given geometric control structure (τ, ν, ρ) and a cost function L, one can associate an extended geometric control structure $(\overline{\tau}, \overline{\nu}, \overline{\rho})$ such that the cost function L is incorporated in the new anchor map $\overline{\rho}$. Assume that $\overline{M} = M \times \mathbb{R}$ and that $\overline{U} = U \times \mathbb{R}$. The extended control structure is defined as follows:

1. $\overline{\tau} : \overline{M} \to \mathbb{R} : (x, J) \to \tau(x)$,
2. $\overline{\nu} : \overline{U} \to \overline{M} : (u, J) \to (\nu(u), J)$,
3. $\overline{\rho} : \overline{U} \to J^1\overline{\tau} : (u, J) \to (\rho(u), J, L(u))$,

where we used the natural identification of $J^1\overline{\tau}$ with $J^1\tau \times \mathbb{R}^2$. Controls in both geometric control structure are in a one-to-one relation with each other. However, a control in the extended setting provides us with some more information since it will keep track of the cost of the corresponding control. This can be shown as follows. Assume that $c : [a, b] \to U$ is a control taking x to y and consider the following function:

$$J(t) = \int_a^t L(c(t'))dt'.$$

Next, with $c(t)$ and $J(t)$ we define a control $\overline{c} : [a, b] \to \overline{U}$ in the extended setting: $\overline{c}(t) = (c(t), J(t))$. It is easily seen that $\overline{c}$ takes $(x, 0) \in \overline{M}$ to $(y, \mathcal{J}(c)) \in \overline{M}$. The converse also holds. Assume that $\overline{c}$ is a control in the extended setting, taking $(x, 0)$ to (y, J_y). Then, the projection $c(t)$ of $\overline{c}(t) = (c(t), J(t))$ onto U is a control in (τ, ν, ρ) with cost $\mathcal{J}(c) = J(b)$. Indeed, it is easily seen that $\dot{J}(t) = L(c(t))$ and $j^1_t\tilde{c} = \rho(c(t))$ hold. The variation cone of a control in the extended setting taking $\overline{x} \in \overline{M}$ to $\overline{y} \in \overline{M}$ will be denoted by $C_{\overline{y}}R_{\overline{x}}$ (or $V_{\overline{y}}R_{\overline{x}}$ for the vertical variation cone).

We can now briefly sketch how one can derive necessary conditions for an optimal control $c(t)$ using Theorem 1. Assume that $c : [a, b] \to U$ is an optimal control, taking x to y. Consider the associated control $\overline{c}(t) = (c(t), J(t))$ taking $\overline{x} = (x, 0)$ to $\overline{y} = (y, \mathcal{J}(c))$ in the extended setting, as defined in the previous paragraph. Let $\overline{\zeta} : [0, 1] \to \overline{M}$ denote the curve through $\overline{y} = (y, \mathcal{J}(c))$ at $t = 0$, defined by $\overline{\zeta}(t) = (y, \mathcal{J}(c) - t)$. The tangent vector to $\overline{\zeta}(t)$ at $t = 0$ then equals $-\partial/\partial J$. If we apply Theorem 1 in the extended setting $(\overline{\tau}, \overline{\nu}, \overline{\rho})$, then we can conclude that $-\partial/\partial J$ can not be contained in the interior of the variation cone $C_{\overline{y}}R_{\overline{x}}$. Indeed, assume that the converse holds, then from Theorem 1 we conclude that there exists an $\epsilon > 0$ such that $\overline{\zeta}(\epsilon) \in R_{\overline{x}}$. This implies that there exists a control $c'(t)$ such that $\mathcal{J}(c') = J(c) - \epsilon$, which leads to a contradiction, since c is assumed to be optimal.

The condition that $-\partial/\partial J$ is not contained in the interior of $C_{\overline{y}}R_{\overline{x}}$ can be reformulated using the notion of the dual cone of $C_{\overline{y}}R_{\overline{x}}$. We first recall some properties and terminology regarding linear spaces and convex cones in a linear space. Let $\mathcal{V}$ be an arbitrary (finite dimensional) linear space. A hyperplane in $\mathcal{V}$ (i.e. a linear subspace of co-dimension one) can always be defined as the set of all vectors $v \in \mathcal{V}$ satisfying $\langle \alpha, v \rangle = 0$ for some (non-zero) co-vector $\alpha \in \mathcal{V}^*$. Such a hyperplane divides $\mathcal{V}$ into two 'half-spaces' which are given by the set of all v satisfying $\langle \alpha, v \rangle \leq 0$, resp. $\langle \alpha, v \rangle \geq 0$, and which are called the 'negative' half-space and the 'positive' half-space, respectively. Let C denote a convex cone in $\mathcal{V}$ (we always assume that the vertex of C is taken in the origin of $\mathcal{V}$). The set of all α for which $\langle \alpha, v \rangle \leq 0$, $\forall v \in C$, is called the *dual cone* of C and is denoted by C^*. It is readily seen that C^* is again a convex cone. Using the above definitions, we can say that the dual cone C^* represents all hyperplanes in $\mathcal{V}$ such that the cone C is contained in the negative half-space. Let us now return to optimal control theory.

The condition that $-\partial/\partial J$ is not contained in the interior of $C_{\overline{y}}R_{\overline{x}}$, is equivalent to the condition that the tangent vector $-\partial/\partial J$ can be separated from $C_{\overline{y}}R_{\overline{x}}$, i.e. there exists a co-vector $\overline{\alpha} \in (C_{\overline{y}}R_{\overline{x}})^* \subset T^*\overline{M}$ such that, $-\partial/\partial J$ is contained in the positive half-space determined by $\overline{\alpha}$, i.e. $\langle \overline{\alpha}, -\partial/\partial J \rangle \geq 0$. In [1] we have proven that the existence of such a co-vector leads to the necessary conditions for optimal controls, from the maximum principle. In order to arrive at a formulation of the maximum principle within the present framework, we first introduce some additional concepts.

Let $\lambda \in \mathbb{R}$ and consider the one-parameter family of sections σ_λ of the bundle $\nu^*T^*M \to \nu^*V^*\tau$ defined by: $\sigma_\lambda(u, \eta) = (u, \alpha)$, where α is completely determined by the conditions that α projects onto η and that

$$\langle \alpha, \mathbf{T}(\rho(u)) \rangle + \lambda L(u) = 0.$$

Fix a coordinate system (t, q^i, u^a, p_0, p_i) on ν^*T^*M, and consider the corresponding coordinate system (t, q^i, u^a, p_i) on $\nu^*V^*\tau$. We then locally have

$$\begin{aligned}\sigma_\lambda&(t, q^i, u^a, p_i) \\ &= (t, q^i, u^a, -\gamma^i(t, q^i, u^a)p_i - \lambda L(t, q^i, u^a), p_i).\end{aligned}$$

Using these sections, we can pull-back the canonical symplectic form ω on T^*M to a presymplectic form ω_λ on $\nu^*V^*\tau$. Let $pr_2 : \nu^*T^*M \to T^*M$ denote the standard projection onto the second factor, then ω_λ is defined by $\omega_\lambda = \sigma^*_\lambda(pr^*_2\omega)$. In terms of the coordinates introduced above, one obtains

$$\omega_\lambda = d\big(-\gamma^i(t, q^i, u^a)p_i - \lambda L(t, q^i, u^a)\big) \wedge dt + dp_i \wedge dq^i.$$

We are now ready to state the maximum principle in the following form.

Theorem 2. *Let $c : [a, b] \to U$ denote an optimal control. Then there exists a piecewise smooth section $\eta(t)$ of $V^*\tau$ along the base curve of $\tilde{c}$ and a real number $\lambda \in \mathbb{R}$, such that*
1. *$i_{(\dot{c}(t),\dot{\eta}(t))}\omega_\lambda = 0$ for all t where $c(t)$ is smooth;*
2. *$(\eta(t), \lambda) \neq 0$ for all t;*
3. *$\lambda \leq 0$;*
4. *$\langle \sigma_\lambda(c(t), \eta(t)), \mathbf{T}(\rho(u)) \rangle \geq 0$ for all $u \in U_{\tilde{c}(t)}$ and $t \in [a, b]$.*

Note that Condition 4. says that the function $u \mapsto \langle \sigma_\lambda(c(t), \eta(t)), \mathbf{T}(\rho(u)) \rangle$ on $U_{\tilde{c}(t)}$ attains a maximum at $u = c(t)$ (this follows from the fact that, by definition, $\langle \sigma_\lambda(c(t), \eta(t)), \mathbf{T}(\rho(c(t))) \rangle = 0$). Let c denote an arbitrary control taking x to y. A pair $(\eta(t), \lambda)$ with $\eta(t)$ a section of $V^*\tau$ along $\tilde{c}$ and $\lambda \in \mathbb{R}$, is called a *multiplier* if conditions 1., 2. and 4. are satisfied. The relation between the maximum principle and the variation cone is expressed by the following result (where $\overline{x} = (x, 0)$ and $\overline{y} = (y, \mathcal{J}(c))$ and where we used the identification of $T^*\overline{M}$ with $T^*M \times \mathbb{R} \times \mathbb{R}$).

Theorem 3. *A pair $(\eta(t), \lambda)$ is a multiplier for a control $c : [a, b] \to U$ iff $(\eta(b), \mathcal{J}(c), \lambda) \in \big(V_{\overline{y}}R_{\overline{x}}\big)^*$ (or equivalently, $(\sigma_\lambda(c(b), \eta(b), \mathcal{J}(c), \lambda)) \in \big(C_{\overline{y}}R_{\overline{x}}\big)^*$).*

Using this theorem, it is not difficult to see that the conditions of the maximum principle are equivalent to saying that there exists a co-vector $(\sigma_\lambda(c(b), \eta(b), \mathcal{J}(c), \lambda))$ in $\big(C_{\overline{y}}R_{\overline{x}}\big)^*$ such that the tangent vector $-\partial/\partial J$ lies in the corresponding positive half-space.

III. Necessary and sufficient conditions for abnormal and strictly abnormal extremals

The following definitions are standard in control theory. A control $c(t)$ for which there exists a multiplier $(\eta(t), \lambda)$ with $\lambda \leq 0$ is called an *extremal* (since it satisfies the necessary conditions for a control to be optimal). A control $c(t)$ for which there exists a multiplier $(\eta(t), \lambda)$ with $\lambda < 0$ is called a *normal extremal*. Note that a control $c(t)$ for which there exists a multiplier with $\lambda = 0$ satisfies conditions that do not depend on the cost function L, which is exceptional, since these conditions are necessary for $c(t)$ to be optimal with respect to the cost function L. This is why a control for which there exists a multiplier $(\eta(t), \lambda)$ with $\lambda = 0$ is called an *abnormal extremal*. These definitions are misleading, since an extremal can be simultaneously normal and abnormal (a control can admit more then one multiplier !). If an extremal is abnormal, but not normal, then it is called a *strictly abnormal extremal*.

The following necessary and sufficient conditions for a control to be an abnormal or strictly abnormal extremal were proven in [1].

Theorem 4. *Let $c(t)$ denote a control taking x to y. Then $c(t)$ is an abnormal extremal iff $V_yR_x \neq T_yM$. Let $\overline{x} = (x, 0)$ and $\overline{y} = (y, \mathcal{J}(c))$. Then $c(t)$ is a strictly abnormal extremal iff $-\partial/\partial J$ is not contained in the boundary of $V_{\overline{y}}R_{\overline{x}}$.*

Remark 1. It should be remarked that the above theorem makes use of the variation cone of a control $c(t)$. By definition of the variation cone, one needs to fix an ordered family of sections of ν such that these sections generate $c(t)$ (cf. Section II). In [1] we have proven that the interior (or closure) of a variation cone does not depend on a specific choice of such an ordered family of sections, but only on the control $c(t)$.

Intuitively, the above characterisation says that a control is abnormal if there do not exist enough variations for their tangent vectors to span the entire tangent space at the endpoint. This characterisation does not depend on the cost function, which justifies the terminology.

In the following section we will illustrate the applicability of Theorem 4 by means of two examples.

IV. Examples of strictly abnormal extremals

In this section we consider two examples of a sub-Riemannian structure on $\mathbb{R}^3$ and use the necessary and sufficient conditions for the existence of (strictly) abnormal extremals to prove that, for the structures under consideration, strictly abnormal extremals do exist. Before proceeding, we first recall what is meant by a sub-Riemannian structure, and how it is related to optimal control theory. A sub-Riemannian structure consists of a Q is a manifold, equipped with a regular smooth distribution $\pi : D \to Q$ and a Riemannian bundle metric h on D. Using the bundle metric we can define a notion of length only for a privileged set of curves, i.e. those curves that are *tangent to the distribution D*. Assume that $\tilde{c} : [a, b] \to Q$ is a curve in Q, such that $\dot{\tilde{c}}(t) \in D_{\tilde{c}(t)}$. The *length* $\ell(\tilde{c})$ of $\tilde{c}$ with respect to the sub-Riemannian structure is defined by

$$\ell(\tilde{c}) = \int_a^b \sqrt{h\big(\dot{\tilde{c}}(t), \dot{\tilde{c}}(t)\big)}dt.$$

A curve everywhere tangent to the distribution D is called an *admissible curve*. An admissible curve $\tilde{c} : [a, b] \to Q$ with $q_i = \tilde{c}(a)$ and $q_f = \tilde{c}(b)$, is called *length minimising* if, given any other admissible curve $\tilde{c}' : [a, b] \to Q$ with $q_i = \tilde{c}'(a)$ and $q_f = \tilde{c}'(b)$, then $\ell(\tilde{c}) \leq \ell(\tilde{c}')$. The main problem in sub-Riemannian geometry is the characterisation of length minimising curves. This problem can be reformulated as an optimal control problem in the following way.

Consider the trivial bundle $\tau : M := \mathbb{R} \times Q \to \mathbb{R}$ and define $\nu : U := \mathbb{R} \times D \to M : \nu(t, v) \mapsto (t, \pi(v))$ and $\rho : U \to J^1\tau \cong \mathbb{R} \times TQ : \rho(t, v) \mapsto (t, v)$. It is easily seen that the set of controls $c(t) = (t, \dot{\tilde{c}}(t))$ is in a one-to-one correspondence with the set of (piecewise smooth) admissible curves. Therefore, if we consider the cost function $L(u) = \sqrt{h(v, v)}$, with $u = (t, v)$, $u \in U$ and $v \in D$, then $\mathcal{J}(c) = \ell(\tilde{c})$, where $(t, \tilde{c}(t))$ is the base curve of $c(t)$. Therefore, if a control $c(t)$ is optimal with respect to L, then the corresponding admissible curve is length minimising (and vice versa). In [3] we have reformulated the maximum principle for this optimal control structure in terms of geometric structures that live on Q. However, for brevity we will not mention these conditions explicitly (see e.g. [4], [5]) and in the following we only concentrate on the existence of abnormal extremal curves in sub-Riemannian geometry.

(i) In the first example we study a sub-Riemannian structure in which we search for strictly abnormal extremals using Theorem 4. This example has been used by R. Montgomery [8] in order to prove that strictly abnormal extremals do exist and, in addition, that such extremals can be length minimising.

Let $Q = \mathbb{R}^3$ (we use cylindrical coordinates (r, θ, z) on $\mathbb{R}^3$), and consider two vector fields $X_1 = \partial/\partial r$ and $X_2 = \partial/\partial\theta - p(r)\partial/\partial z$, where $p(r)$ is a function on $\mathbb{R}$ with a single non degenerate maximum at $r = 1$, i.e. p satisfies:

$$\frac{d}{dr}p(r)\bigg|_{r=1} = 0 \quad \text{and} \quad \frac{d^2}{dr^2}p(r)\bigg|_{r=1} < 0.$$

The distribution thus defined is everywhere of rank two, and is differentiable by construction. The flows of X_1, X_2 are denoted by $\{\phi_s\}$, $\{\psi_s\}$, respectively. In particular, we have $\phi_t(r, \theta, z) = (t+r, \theta, z)$, $\psi_t(r, \theta, z) = (r, \theta+t, z-p(r)t)$. Let $\tilde{c} : [0, 1] \to Q$ be an integral curve of X_1 through $q_0 = (r_0, \theta_0, z_0)$ at $t = 0$, with endpoint q_1. In the following we will prove that the integral curves of X_1 are not abnormal. Note that these conditions do not depend on the cost function, i.e. the Riemannian metric on D, which allows us to say that the results below hold for any metric h on D. The vertical variation cone equals

$$V_{(1,q_1)}R_{(0,q_0)} = \text{span}\Big\{ X_1(q_1), X_2(q_1), \frac{\partial}{\partial\theta}\bigg|_{q_1} - p(r_0 + t)\frac{\partial}{\partial z}\bigg|_{q_1} \quad \Big| \quad \forall t \in [0, 1] \Big\}.$$

Since the bundle τ is trivial, we have identified $V_{(1,q_1)}\tau$ with $T_{q_1}Q$. The variation cone coincides with the whole tangent space, i.e. $V_{(1,q_1)}R_{(0,q_0)} = T_{q_1}Q$, by observing that for an arbitrary $(v_r, v_\theta, v_z) \in T_{q_1}Q$:

$$\begin{aligned} v_r \left.\frac{\partial}{\partial r}\right|_{q_1} + v_\theta \left.\frac{\partial}{\partial \theta}\right|_{q_1} + v_z \left.\frac{\partial}{\partial z}\right|_{q_1} = \\ v_r X_1(q_1) + v_\theta X_2(q_1) \\ + \frac{v_z + v_\theta p(r_0)}{p(r_0+t) - p(r_0)}(X_2 - \phi^*_{-t}X_2)(q_1), \end{aligned}$$

where t is chosen such that $p(r_0+t) \neq p(r_0)$. Consequently, in view of Theorem 4, one can conclude that an integral curve of X_1 can not be abnormal with respect to any metric on $D \to Q$.

We now repeat the above computations for the integral curves of X_2. Let $\tilde{c}' : [0,1] \to Q$ be an integral curve of X_2, with $\tilde{c}'(0) = q_0 = (r_0, \theta_0, z_0)$ and endpoint q_1. The vertical variation cone now becomes

$$\begin{aligned} V_{(1,q_1)}R_{(0,q_0)} = \operatorname{span}\Big\{ X_1(q_1), X_2(q_1), \\ \left.\frac{\partial}{\partial r}\right|_{q_1} + p'(r_0)t \left.\frac{\partial}{\partial z}\right|_{q_1} \;\Big|\; \forall t \in [0,1] \Big\}. \end{aligned}$$

If q_0 is a point on the cylinder defined by $r = 1$, then one easily sees that $V_{(1,q_1)}R_{(0,q_0)} \neq T_{q_1}Q$, since $p'(1) = 0$. Therefore, every helix $\tilde{c}' : [0,1] \to Q : t \mapsto (1, \theta+t, z-p(1)t)$ is an abnormal extremal.

The following step in our treatment consists of proving that $\tilde{c}'$ is in fact strictly abnormal. For that purpose, we need to work in the extended setting (i.e. on $\mathbb{R} \times Q \times \mathbb{R}$) and compute the extended vertical cone of variations. It is now necessary to fix a sub-Riemannian metric h, determining the energy cost E. In the example constructed by R. Montgomery [8], the metric h on D is given by $h_{11} = 1$, $h_{12} = h_{21} = 0$ and $h_{22} = r^2$, when expressed with respect to the basis $\{X_1, X_2\}$ of D. It is easily seen, with this choice for h, that the sub-Riemannian length of a curve tangent to D is precisely the length of its projection on the (x,y)-plane with respect to the standard Riemannian metric on $\mathbb{R}^2$. In order to compute the extended vertical variation cone, we first recall the definition of the extended geometric control structure. The anchor map in the extended setting $\overline{\rho} : (\mathbb{R} \times Q \times \mathbb{R}) \times \mathbb{R}^2 \to T(\mathbb{R} \times Q \times \mathbb{R})$ is defined by:

$$\begin{aligned} \overline{\rho}\big((q,J),(u^1,u^2)\big) \quad = \\ u^1\frac{\partial}{\partial r} + u^2\left(\frac{\partial}{\partial \theta} - p(r)\frac{\partial}{\partial z}\right) + \frac{1}{2}\big((u^1)^2 + (ru^2)^2\big)\frac{\partial}{\partial J}, \end{aligned}$$

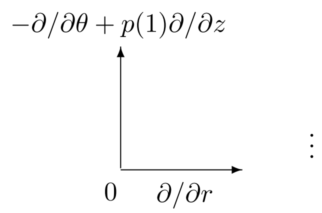
where we have used the basis $\{X_1, X_2\}$ of D to fix a coordinate system (u^1, u^2) on the anchored bundle $\tau_Q|_D$. Note that we have omitted explicit reference to the time coordinate, since eventually (when computing the vertical variation cone) it will be canceled out. The vector field on $Q \times \mathbb{R}$ defined by $\overline{X}_2(q,J) = \partial/\partial\theta - p(r)\partial/\partial z + \frac{1}{2}r^2\partial/\partial J$, with flow $\{\overline{\psi}_t\}$ given by $\overline{\psi}_t(r,\theta,z,J) = (r, \theta+t, z-p(r)t, J+\frac{1}{2}r^2t)$, satisfies the property that its integral curve through $(q_0, 0)$, with $q_0 = (1,0,0)$, is precisely the curve

$$t \mapsto \left(c'(t), \int_a^t \sqrt{h(\dot{\tilde{c}}'(t'), \dot{\tilde{c}}'(t')dt'}\right).$$

The extended vertical variational cone is generated by tangent vectors of the form:

$$\begin{aligned} T\overline{\psi}_{1-t}\Big(\overline{\rho}\big(\overline{\psi}_t(q_0,0),(u^1,u^2)\big)\Big) - \overline{X}_2(q_1) = \\ u^1\left(\frac{\partial}{\partial r} + (1-t)\frac{\partial}{\partial J}\right) + (u^2-1)\left(\frac{\partial}{\partial \theta} - p(1)\frac{\partial}{\partial z}\right) \\ + \frac{1}{2}\left((u^1)^2 + (u^2)^2 - 1\right)\frac{\partial}{\partial J}, \end{aligned}$$

where $u^1, u^2 \in \mathbb{R}$ and $t \in]0,1]$ are arbitrary. In order to prove that $-\partial/\partial J$ is contained in the boundary of extended variational cone, we introduce new control coordinates (s,ϕ): $u^1 = s\cos\phi$ and $u^2 = s\sin\phi$. By replacing u^1 and u^2 in the above family of tangent vectors by these new coordinates, we now construct two circles in the extended variation cone. Let $t = 1$, then it is easily seen that the circle in $T_{\overline{\psi}_1(q_0,0)}(Q \times \mathbb{R})$ with centre at the point corresponding to the tangent vector $-(\partial/\partial\theta - p(1)\partial/\partial z)$ and determined by the tangent vectors parameterised by $s = 1$ and $\phi \in [0,2\pi]$, is entirely contained in the extended vertical variational cone. The tangent line to this circle at the origin is spanned by $\partial/\partial r$. Therefore, both vectors $\partial/\partial r$ and $-\partial/\partial r$ are contained in the closure of this cone (see Remark 2 and the picture below).



By taking $t = 0$, $s = 1$ and $\phi = [0,2\pi]$ and repeating the above reasoning, we obtain that the straight line spanned by $-\partial/\partial r - \partial/\partial J$ is contained in the closure of the cone. Adding $\partial/\partial r$ to this vector, shows that $-\partial/\partial J$ is contained in the closure of the extended vertical variational cone. Therefore $\tilde{c}'$ is strictly abnormal, which agrees with the results of R. Montgomery.

Remark 2. Let C denote a cone in a finite dimensional vector space $\mathcal{V}$, with vertex at the origin. Let $c : [0,1] \to \mathcal{V}$ denote a curve through 0 at $t = 0$ such that $c([0,1]) \subset C$. The tangent vector at 0 is defined by:

$$\dot{c}(0) = \lim_{h \to 0^+} \frac{1}{h}c(h).$$

Since $c(h)/h$ is contained in C for any $h > 0$, this implies that the limit itself is contained in the closure $\operatorname{cl} C$ of C. We thus conclude that $\dot{c}(0) \in \operatorname{cl} C$. A similar argument can be applied if we consider a curve $c : [-1,0] \to \mathcal{V}$ such that

$c([-1,0]) \subset C$ and $c(0)=0$. Consider again the tangent vector at $t=0$:

$$\lim_{h\to 0^-}\frac{1}{h}c(h) = -\lim_{(-h)\to 0^+}\frac{1}{(-h)}c(h).$$

We obtain that $-\dot c \in \mathrm{cl}\, C$.

(ii) The second example of length minimising strictly abnormal extremals is taken from W. Liu and H.J. Sussmann [10].

Let $M = \mathbb{R}^3$ and D 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 set $\{X_1, X_2\}$ forms a basis for D and it determines a coordinate system (u^1,u^2) on the control space. This allows us to define a metric h on D and the cost function on D then becomes

$$L((x,y,z),u^1,u^2) = \sqrt{(u^1)^2+(u^2)^2},$$

We investigate the abnormality of the integral curves of X_2. The flows $\{\phi_t\}$ of X_1 and $\{\psi_t\}$ 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 $x=2$. 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 $x=2$, is an abnormal extremal (i.e. the vertical variational cone equals D). We now compute that the integral curves of X_2 are strictly abnormal. The anchor map $\overline{\rho}$ in the extended setting $\mathbb{R}^4$ becomes (again we leave out the time coordinate) in coordinates, with $\overline{q} = (x,y,z,J)$:

$$\begin{aligned}\overline{\rho}(q,u^1,u^2) &= u^1\left.\frac{\partial}{\partial x}\right|_{\overline{q}} + u^2\left((1-x)\left.\frac{\partial}{\partial y}\right|_{\overline{q}} + x^2\left.\frac{\partial}{\partial z}\right|_{\overline{q}}\right)\\ &+\frac{1}{2}\left((u^1)^2+(u^2)^2\right)\left.\frac{\partial}{\partial J}\right|_{\overline{q}}.\end{aligned}$$

Consider the vector field $\overline{X}_2$, defined on $\mathbb{R}^4$:

$$\overline{X}_2(\overline{q}) = \overline{\rho}(\overline{q},0,1) = (1-x)\left.\frac{\partial}{\partial y}\right|_{\overline{q}} + x^2\left.\frac{\partial}{\partial z}\right|_{\overline{q}} + \frac{1}{2}\left.\frac{\partial}{\partial J}\right|_{\overline{q}},$$

with flow $\{\overline{\psi}_t\}$ defined by $\overline{\psi}_t(x,y,z,J) = (\psi_t(x,y,z), J + \frac{1}{2}t)$. By definition, the extended vertical variational cone is generated by tangent vectors of the form:

$$\begin{aligned}&T\overline{\psi}_{b-t}\Big(\overline{\rho}(\overline{\psi}_{t-a}(\overline{q}),u^1,u^2)\Big) - \overline{X}_2(\overline{\psi}_b(\overline{q})) = \qquad (1)\\ &u^1\left(\frac{\partial}{\partial x} - (b-t)\frac{\partial}{\partial y} + 2(b-t)x\frac{\partial}{\partial z}\right)\\ &+ \quad (u^2-1)\left((1-x)\frac{\partial}{\partial y} + x^2\frac{\partial}{\partial z}\right)\\ &+ \quad \frac{1}{2}\Big((u^1)^2+(u^2)^2-1\Big)\frac{\partial}{\partial J},\end{aligned}$$

with $u^1,u^2 \in \mathbb{R}$ and $t \in]a,b]$. Consider the following parametrisation: let $u^1 = s\cos\theta$ and $u^2 = s\sin\theta$, for $s>0$ and $\theta\in[0,2\pi[$. If we assume that $s=1$, the coefficient of $\partial/\partial J$ becomes zero. If θ varies, we obtain a curve in the cone trough the origin. We know from Remark 2 that the tangent ray at the origin, i.e. the straight line spanned by

$$\frac{\partial}{\partial x} - (b-t)\frac{\partial}{\partial y} + 2(b-t)x\frac{\partial}{\partial z},$$

lies in the *closure* of the cone. If we substitute $b=t$, then we have that $\pm Y_1 = \pm\partial/\partial x$ is contained in the closure of the cone. If, on the other hand, $b>t$, then we obtain that $\pm Y_2 = \pm(\partial/\partial y - 2x\partial/\partial z)$ is contained in the closure of the cone. Now assume that $x=0$ or $x=2$. It is now easily seen that, given u^1,u^2 such that $(u^1)^2+(u^2)^2<1$, then a linear combination (with positive coefficients) of the tangent vector in (1) and $\pm Y_1, \pm Y_2$ can be found which is proportional to $-\partial/\partial J$, up to a positive multiple.

V. Discussion

The necessary and sufficient conditions for a control to be a (strictly) abnormal extremal are rather difficult to compute. In my phd-dissertation new sufficient conditions for a control to not be abnormal have been found, which are easier to compute, and which are stronger in comparison with known sufficient conditions. These new conditions are stated in terms of iterated Lie brackets of vector fields on M. A proof of these conditions would be out of the scope of this paper.

Acknowledgements

I am indebted to F. Cantrijn for his unconditional support, the useful discussions and the careful reading of this paper.

References

[1] B. Langerock, "Geometric aspects of the maximum principle and lifts over a bundle map," *Acta Appl. Math.*, 2003, to appear (math.DG/0212055).

[2] H.J. Sussmann, "An introduction to the coordinate-free maximum principle," in *Geometry of Feedback and Optimal Control*, B. Jakubczyk and W. Respondek, Eds. 1997, pp. 463–557, Marcel Dekker, New York.

[3] B. Langerock, "A connection theoretic approach to sub-Riemannian geometry," *J. Geom. Phys.*, 2002, to appear (math.DG/0210004).

[4] R.S. Strichartz, "Sub-Riemannian geometry," *J. Diff. Geom.*, vol. 24, pp. 221–263, 1986.

[5] R.S. Strichartz, "Corrections to "sub-Riemannian geometry"," *J. Diff. Geom.*, vol. 30, pp. 595–596, 1989.

[6] H.J. Sussmann, "Orbits of families of vector fields and integrability of distributions," *Trans. Amer. Math. Soc.*, vol. 180, pp. 171–188, 1973.

[7] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamklelidze, and E.F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Wiley, Interscience, 1962.

[8] R. Montgomery, "Abnormal minimizers," *SIAM J. Control Optim.*, vol. 32, pp. 1605–1620, 1994.

[9] S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Academic Press, New York, 1978.

[10] W. Liu and H.J. Sussmann, "Shortest paths for sub riemannian metrics on rank two distributions," *Memoirs AMS*, vol. 118, 1995.