

ABNORMAL SUB-RIEMANNIAN MINIMIZERS

By

Wensheng Liu

and

Héctor J. Sussmann

IMA Preprint Series # 1059

November 1992

ABNORMAL SUB-RIEMANNIAN MINIMIZERS

Wensheng Liu and Héctor J. Sussmann¹

Institute for Mathematics and its Applications²
University of Minnesota
Minneapolis, Minnesota 55455

E-mail: wslu@ima.umn.edu and sussmann@hamilton.rutgers.edu

October 22, 1992

ABSTRACT. We give a new, simple proof of optimality of abnormal extremals, for an example similar to those recently considered by R. Montgomery and I. Kupka. Our optimality proof proceeds by directly establishing the desired inequality by elementary means, without making a detailed study of the Pontryagin extremals or using geometric arguments.

§1. Introduction .

One of the most important questions in sub-Riemannian geometry is whether optimal abnormal extremals can exist (cf. [2], [7], [8]). In [7], a general statement was made that sub-Riemannian minimizers are necessarily smooth, and this was derived from the proposition that all minimizers are what we would now call “normal extremals.” Subsequently, it was noticed that the proof of this assertion involved an incorrect application of the Pontryagin Maximum Principle, and that a truly correct analysis based on this result implied the possibility that a minimizer might be “abnormal.” In [8], it was pointed out that the result of [7] remained valid for a very restrictive class of sub-Riemannian manifolds, namely, those that obey the “strong bracket-generating condition.” Until recently, it was an open question whether optimal abnormal extremals can in fact occur. This was settled in recent work by R. Montgomery [5], who showed that there exist sub-Riemannian structures in \mathbb{R}^3 , associated to a two-dimensional subbundle E of the tangent bundle, for which there exist strictly abnormal extremals that are uniquely optimal. (We call an abnormal extremal *strictly abnormal* if it is not also a normal extremal. We call an admissible trajectory *optimal* if it minimizes length among all admissible trajectories with the same initial and terminal points, and *uniquely optimal* if it is the only optimal trajectory joining these

¹ Sussmann’s work was supported in part by the National Science Foundation Grant DMS-8902994.

² Address valid until June 93 only. Sussmann’s address after June 1993: Department of Mathematics, Rutgers University, New Brunswick, NJ 08903

two points, up to reparametrization of the time interval.) Montgomery’s optimality proof is rather lengthy and involved, making it desirable to find simpler ways of establishing the result. I. Kupka has provided in [3] a different proof, based on a detailed analysis of the solutions of the differential equation defining the normal extremals.

In this note, we present another, much simpler proof, for an example similar to those of Montgomery and Kupka. The optimality proof relies on an elementary inequality, and does not require any analysis of Pontryagin extremals. Moreover, we actually prove a slightly stronger optimality property, namely, that our abnormal extremal is optimal within the class of trajectories of a larger system, in which the control constraint $u^2 + v^2 \leq 1$ is replaced by the weaker restriction $|u| \leq 1, |v| \leq 1$.

REMARK 1. Our example involves a two-dimensional “nonregular distribution” E in \mathbb{R}^3 . The sections of the subbundle E are globally spanned by two vector fields f and g , and the vectors $f(p), g(p), [f, g](p)$ and $[f, [f, g]](p)$ span \mathbb{R}^3 for every p . However, the span of $f(p), g(p)$ and $[f, g](p)$ is not always three-dimensional, since the dimension drops to 2 on the planes $x = 0$ and $x = 2$. So E is not a “regular distribution.” It is easy, however, to modify our example and produce a *regular* two-dimensional distribution \tilde{E} on \mathbb{R}^4 , with an associated metric, such that for the corresponding sub-Riemannian structure there exist uniquely optimal strictly abnormal extremals (cf. Remark 2 below).

§2. Sub-Riemannian manifolds and abnormal extremals .

If M is a C^∞ manifold, and $p \in M$, we use T_pM to denote the tangent space of M at p , and TM to denote the tangent bundle of M . If E is any C^∞ vector bundle over M , then $\Gamma(E)$ denotes the set of all C^∞ sections of E . A subbundle E of TM is sometimes called a *distribution* on M . A *nonholonomic subbundle* (also known as a *bracket-generating distribution*) is a subbundle E of TM such that the Lie algebra $L(\Gamma(E))$ of vector fields generated by the sections of E has the *full rank property*, i.e. satisfies $\{X(p) : X \in L(\Gamma(E))\} = T_pM$ for all $p \in M$. An *E -admissible arc* is an absolutely continuous curve γ on M , defined on some compact interval $[a, b]$, such that $\dot{\gamma}(t) \in E(\gamma(t))$ for almost all $t \in [a, b]$. If E is nonholonomic and M is connected, then any two points in M can be joined by an E -admissible arc.

A C^∞ *Riemannian metric* on E is a C^∞ section $p \rightarrow G_p$ of the bundle $E^* \otimes E^*$ such that for each $p \in M$ the bilinear form $T_pM \times T_pM \ni (v, w) \rightarrow G_p(v, w) \in \mathbb{R}$ is symmetric and positive definite. A *sub-Riemannian structure* on a manifold M is a pair (E, G) where E is a nonholonomic C^∞ subbundle of TM and G is a C^∞ Riemannian metric on E . A *sub-Riemannian manifold* is a triple (M, E, G) such that M is a C^∞ manifold and (E, G) is a sub-Riemannian structure on M . One can always construct a Riemannian metric on any subbundle E of TM by just taking a Riemannian metric on TM and restricting it to E . If $p \in M, v \in E(p)$, then the *length* $\|v\|_G$ of v is the number $G_p(v, v)^{1/2}$. The *length* $\|\gamma\|_G$ of an E -admissible arc $\gamma : [a, b] \rightarrow M$ is the integral $\int_a^b \|\dot{\gamma}(t)\|_G dt$. If $p, q \in M$, then the infimum of the lengths of all the E -admissible curves γ that go from p to q is the *distance*

from p to q , and is denoted by $d_G(p, q)$. If M is connected and E is nonholonomic, then $d_G(p, q) < \infty$ for all p, q , and $d_G : M \times M \rightarrow \mathbb{R}$ is a metric whose associated topology is the one of M . An E -admissible curve $\gamma : [a, b] \rightarrow M$ such that $d_G(\gamma(a), \gamma(b)) = \|\gamma\|_G$ is called a *minimizer*.

An E -admissible curve γ is *parametrized by arc length* if $\|\dot{\gamma}(t)\|_G = 1$ for almost all t in the domain of γ . If $\gamma : [a, b] \rightarrow M$ is E -admissible, then we can define $\tau(t) = \int_a^t \|\dot{\gamma}(s)\|_G ds$, so τ is a monotonically nondecreasing function on $[a, b]$ with range $[0, \|\gamma\|_G]$. Moreover, if $t_1 < t_2$ but $\tau(t_1) = \tau(t_2)$, then $\gamma(t_2) = \gamma(t_1)$. So we can define $\tilde{\gamma} : [0, \|\gamma\|_G] \rightarrow M$ by letting $\tilde{\gamma}(s) = \gamma(t)$ if $\tau(t) = s$. Then, if $s_1 < s_2$, and $s_i = \tau(t_i)$ for $i = 1, 2$, the points $\tilde{\gamma}(s_1)$ and $\tilde{\gamma}(s_2)$ can be joined by the restriction of γ to the interval $[t_1, t_2]$, whose G -length is $s_2 - s_1$. So $d_G(\tilde{\gamma}(s_1), \tilde{\gamma}(s_2)) \leq s_2 - s_1$. If \hat{G} is a Riemannian metric on M (i.e. a metric defined on the whole tangent bundle TM) that extends G , then the \hat{G} -distance $d_{\hat{G}}(\tilde{\gamma}(s_1), \tilde{\gamma}(s_2))$ is *a fortiori* $\leq s_2 - s_1$. So $\tilde{\gamma}$ is Lipschitz as a map into $(M, d_{\hat{G}})$. Clearly, $\gamma = \tilde{\gamma} \circ \tau$. Since $\tilde{\gamma}$ is Lipschitz and τ is integrable, we have $\int_0^s \|\dot{\tilde{\gamma}}(\sigma)\|_G d\sigma = \int_0^t \|\dot{\gamma}(\theta)\|_G d\theta = s$, if $s = \tau(t)$. So $\|\dot{\tilde{\gamma}}(s)\| = 1$ for almost all s . Therefore $\tilde{\gamma}$ is parametrized by arc length.

In particular, every minimizer γ is equivalent modulo reparametrization to an arc γ^* which is parametrized by arc length and is optimal for the *minimum time problem* of joining two points p, q by means of an E -admissible arc γ that satisfies $\|\dot{\gamma}(t)\| \leq 1$ for almost all t and minimizes time among all such arcs. (It is clear that, if γ is a minimizer, then the arc $\tilde{\gamma}$ constructed above is a solution of the minimum time problem.) Conversely, it is easy to see that, if γ is a solution of the minimum time problem, then γ is a minimizer parametrized by arc length. So the class of solutions of the minimum time problem coincides with the class of minimizers that are parametrized by arc length.

The solutions of the minimum time problem satisfy a necessary condition for optimality given by the Pontryagin Maximum Principle (cf. [1], [4], [6]). A trajectory that satisfies this condition is called a *Pontryagin extremal*. For our sub-Riemannian situation, the condition defining the Pontryagin extremals can be stated most succinctly as follows. We associate to our sub-Riemannian manifold (M, E, G) a real-valued function \mathcal{H} on the cotangent bundle T^*M defined by $\mathcal{H}(\lambda) = -\|\lambda\|_{E, G}$, where the norm $\|\lambda\|_{E, G}$ of a covector $\lambda \in T_p^*M$ is the norm of the linear functional λ_E obtained by restricting λ to $E(p)$, and $E(p)$ is endowed with the norm arising from the quadratic form G_p . The function \mathcal{H} is smooth on the set $\{\lambda \in T^*M : \mathcal{H}(\lambda) \neq 0\}$. However, \mathcal{H} is not C^1 on the set of zeros of \mathcal{H} . Since \mathcal{H} is locally Lipschitz, it has a well defined *generalized gradient* $\partial\mathcal{H}$, which is a set-valued function that assigns to each $\lambda \in T^*M$ a convex compact subset $\partial\mathcal{H}(\lambda)$ of the cotangent space $T_\lambda^*(T^*M)$. (By definition, $\partial\mathcal{H}(\lambda)$ is the closed convex hull of the set of all covectors $\Lambda \in T_\lambda^*(T^*M)$ such that Λ is the limit of a sequence $\{\Lambda_j\}$ of covectors at points $\lambda_j \in T^*M$ such that \mathcal{H} is differentiable at λ_j and its differential at λ_j is Λ_j .) The symplectic structure of T^*M gives rise to linear maps $J_\lambda : T_\lambda^*(T^*M) \rightarrow T_\lambda(T^*M)$, for $\lambda \in T^*M$, depending smoothly on λ . So we can define a set-valued mapping $J\partial\mathcal{H}$, assigning to each λ a compact convex subset of $T_\lambda T^*M$. A *Pontryagin extremal* is then, simply, an E -admissible arc $\gamma : [a, b] \rightarrow M$ which is the projection of a nontrivial solution

$\Gamma : t \rightarrow (\gamma(t), \lambda(t))$ of the differential inclusion $\dot{\lambda} \in J\partial\mathcal{H}(\lambda)$. (This means that Γ is absolutely continuous, $\dot{\Gamma}(t) \in J\partial\mathcal{H}(\Gamma(t))$ for almost all t , and $\lambda(t) \neq 0$ for all t .) As long as $\mathcal{H} \neq 0$, the set-valued map $J\partial\mathcal{H}$ is in fact single-valued, and is exactly the Hamilton vector field on T^*M associated with \mathcal{H} . It follows in particular that, as long as $\mathcal{H} \neq 0$, a solution Γ of the differential inclusion is in fact a solution in the ordinary Hamilton differential equations associated with \mathcal{H} . So \mathcal{H} is constant along Γ . Therefore \mathcal{H} is constant along *all* solutions of the differential inclusion.

The solutions Γ of the differential inclusion $\dot{\lambda} \in J\partial\mathcal{H}(\lambda)$ along which $\mathcal{H} \neq 0$ (resp. $\mathcal{H} = 0$) are called *normal* (resp. *abnormal*) *solutions*, and a Pontryagin extremal that is the projection of a nontrivial normal (resp. abnormal) solution is called a *normal* (resp. *abnormal*) *extremal*. Since a curve γ in M can in principle be the projection of more than one solution of the differential inclusion, an extremal can be normal and abnormal at the same time. An extremal is *strictly abnormal* if it is abnormal and not normal. A normal solution Γ of the differential inclusion is also a solution (up to a reparametrization $t \rightarrow \kappa t$ of time) of the Hamilton equations for the Hamiltonian $-\mathcal{H}^2$, which is everywhere smooth. In particular, a normal extremal is smooth. This probably explains the erroneous but widespread belief that all the optimal trajectories are projections of Hamilton trajectories of $-\mathcal{H}^2$. What makes the belief erroneous is precisely the abnormal extremals, for which no simple characterizations exist.

§3. The example .

We now construct our example of a sub-Riemannian manifold for which there exist uniquely optimal strictly abnormal extremals. We let $M = \mathbb{R}^3$, with the usual coordinates x, y, z , and take E to be the kernel of the 1-form

$$\omega = x^2 dy - (1 - x) dz .$$

Since ω never vanishes, it is clear that E is a smooth two-dimensional subbundle of the tangent bundle TM . The vector fields f, g , given by

$$f = \frac{\partial}{\partial x} , \quad g = (1 - x) \frac{\partial}{\partial y} + x^2 \frac{\partial}{\partial z} ,$$

form a global basis of sections of E . The Lie brackets $[f, g], [f, [f, g]]$ are easily computed, and turn out to be

$$[f, g] = -\frac{\partial}{\partial y} + 2x \frac{\partial}{\partial z} , \quad [f, [f, g]] = 2 \frac{\partial}{\partial z} .$$

Moreover,

$$[g, [f, g]] = [f, [f, [f, g]]] = [g, [f, [f, g]]] = 0 .$$

This implies in particular that *the Lie algebra of vector fields generated by f and g is nilpotent*. Moreover, f, g and $[f, g]$ are linearly independent everywhere except where

$x = 0$ or $x = 2$, and f, g and $[f, [f, g]]$ are independent everywhere except where $x = 1$. It follows that for every $p \in \mathbb{R}^3$ the values at p of the four vector fields $f, g, [f, g]$ and $[f, [f, g]]$ span \mathbb{R}^3 . Therefore *the subbundle E is nonholonomic* or, equivalently, *the “distribution” E is bracket-generating*.

We now define a metric G on E by

$$G = dx^2 + h(x)(dy^2 + dz^2), \quad \text{where } h(x) = \frac{1}{(1-x)^2 + x^4}.$$

Then the vector fields f and g form an orthonormal basis of sections of E .

The E -admissible arcs are the absolutely continuous curves $t \rightarrow (x(t), y(t), z(t)) = \gamma(t)$ that satisfy the differential equation $x^2 \dot{y} = (1-x)\dot{z}$. If $\gamma : [a, b] \rightarrow M$ is such an arc, and $\gamma(t) = (x(t), y(t), z(t))$, then we let

$$u(t) = \dot{x}(t), \quad v(t) = \frac{\dot{y}(t)}{1-x(t)} = \frac{\dot{z}(t)}{x(t)^2}.$$

(Notice that $v(t)$ is well defined and belongs to L^1 , because the functions $1-x$ and x^2 never vanish simultaneously, and the two expressions defining v agree when both $1-x$ and x^2 are $\neq 0$.) Equivalently, we can define u and v by expressing $\dot{\gamma}(t)$ as a linear combination $\dot{\gamma}(t) = u(t)f(\gamma(t)) + v(t)g(\gamma(t))$.

So γ is a trajectory of the control system Σ given by $\dot{p} = uf(p) + vg(p)$, i.e. by

$$\Sigma : \quad \dot{x} = u, \quad \dot{y} = (1-x)v, \quad \dot{z} = x^2v,$$

for a pair (u, v) of integrable functions of t . Conversely, any absolutely continuous curve $\gamma : [a, b] \rightarrow M$ which is a solution of Σ for some pair (u, v) of integrable real-valued functions on $[a, b]$ is necessarily E -admissible. The length of γ is then given by

$$\|\gamma\|_G = \int_a^b (u(t)^2 + v(t)^2)^{1/2} dt.$$

If γ is parametrized by arc-length, i.e. $\|\dot{\gamma}(t)\| = 1$ for almost all t , then the length of γ is just the *duration* of γ , i.e. the number $b - a$.

The minimum time problem in our case is the problem of minimizing time in the class of trajectories of Σ generated by control functions that satisfy the constraint $u^2 + v^2 \leq 1$, i.e. the class of trajectories of the system

$$\Sigma_1 : \quad \dot{x} = u, \quad \dot{y} = (1-x)v, \quad \dot{z} = x^2v, \quad u^2 + v^2 \leq 1.$$

We now describe the Pontryagin extremals for this problem, and in particular determine the abnormal ones. In doing so, we revert to the terminology of optimal control theory, and use the control-theoretic Hamiltonian H , which is a smooth function on T^*M times

the set U where the controls take values (i.e., in our case, the closed unit disc in the (u, v) plane). The Hamiltonian H of Σ_1 is given by

$$H(x, y, z, \xi, \eta, \zeta, u, v) = \xi u + \eta(1 - x)v + \zeta x^2 v .$$

The adjoint equations say:

$$\dot{\xi} = (\eta - 2\zeta x)v , \quad \dot{\eta} = \dot{\zeta} = 0 .$$

If $\gamma : [a, b] \rightarrow M$ is a time-optimal trajectory for Σ_1 , whose components are functions $t \rightarrow x(t)$, $t \rightarrow y(t)$, $t \rightarrow z(t)$, such that γ is generated by a control $t \rightarrow (u(t), v(t))$, then the Pontryagin Maximum Principle tells us that γ is a Pontryagin extremal, that is, there exist three absolutely continuous functions $t \rightarrow \xi(t)$, $t \rightarrow \eta(t)$, $t \rightarrow \zeta(t)$ that satisfy the adjoint equations as well as the *nontriviality condition* (namely, $(\xi(t), \eta(t), \zeta(t)) \neq (0, 0, 0)$ for all t), and the *minimization property*, which in our case simply says that, if the vector $V(t) = \left(\xi(t), \eta(t)(1 - x(t)) + \zeta(t)x(t)^2 \right)$ is $\neq (0, 0)$, then the control vector $(u(t), v(t))$ is equal to $-\frac{V(t)}{\|V(t)\|}$. The adjoint equations then imply that η and ζ are actually constant, and that the length of the vector $V(t)$ is constant as well.

An abnormal extremal is an extremal for which the two functions $t \rightarrow \xi(t)$ and $t \rightarrow \eta((1 - x(t)) + \zeta x(t)^2)$ vanish identically. The nontriviality condition then implies that $(\eta, \zeta) \neq (0, 0)$. So, if $P(x)$ is the polynomial $\eta(1 - x) + \zeta x^2$, then P has at most two zeros, and therefore the function $t \rightarrow x(t)$ —whose values are zeros of P — is constant (because it is continuous). Therefore the derivative $u(t) = \dot{x}(t)$ vanishes identically. Since $\dot{\xi} = (\eta - 2\zeta x)v$, the constant value of x must satisfy $\eta = 2\zeta x$, unless $v \equiv 0$. If $v \equiv 0$, then γ is a constant trajectory, and so γ is not time-optimal, unless $b = a$. So, if $b > a$, then x satisfies the two equations $\eta(1 - x) + \zeta x^2 = 0$, $-\eta + 2\zeta x = 0$. Therefore $\zeta \neq 0$ (for otherwise η would be 0 as well, contradicting $(\eta, \zeta) \neq (0, 0)$), and x is a double root of P . So the discriminant of P vanishes, that is, $\eta^2 = 4\eta\zeta$. Then either $\eta = 0$ or $\eta = 4\zeta$. In the former case, the equation $\eta = 2\zeta x$ yields $x = 0$. In the latter case, we get $x = 2$. We will only be interested in the case $x = 0$, which corresponds to $\eta = 0$. Clearly, when $x \equiv 0$, z is constant, and y satisfies $\dot{y} = v$. It is clear that such a trajectory cannot be optimal unless $v \equiv 1$ or $v \equiv -1$.

So we are left with the trajectories γ along which $x \equiv 0$ (and therefore $u \equiv 0$) and $v \equiv 1$ or $v \equiv -1$. It is easy to see —by taking $\xi \equiv \eta \equiv 0$ and $\zeta \equiv 1$ — that any such trajectory γ is an abnormal extremal. We now show that γ is strictly abnormal, i.e. that it is not a normal extremal. Indeed, for γ to be a normal extremal we would have to be able to choose a function $t \rightarrow \xi(t)$ and constants η, ζ that satisfy the adjoint equations as well as the minimization and nontriviality conditions, and are such that the value of the Hamiltonian is $\neq 0$. We would still have to have $\xi \equiv 0$, because $u \equiv 0$. Then $(\eta, \zeta) \neq 0$ by nontriviality. The Hamiltonian would then have the value $\eta v(t)$, i.e. $-|\eta|$. But the x -derivative of H at $x = 0$ is $-\eta v$, i.e. $|\eta|$. So $-|\eta| = \dot{\xi} = 0$, and then the value of the Hamiltonian would be 0, which is a contradiction.

REMARK 2. Recall that a subbundle E is *regular* if, for each integer $k > 0$, the dimension of the subspace $E^{(k)}(p)$ is independent of p . (Here $E^{(k)}(p)$ is the linear span of all the vectors $X(p)$, for all vector fields X that are Lie brackets of k or fewer local sections of E defined near p .) Clearly, the subbundle E of our example is not regular. One can, however, construct a regular example by adding one more variable w , with the equation $\dot{w} = (x+y^2)v$. In other words, we let \tilde{E} be the two-dimensional subbundle of $T\mathbb{R}^4$ spanned by the two vector fields $\tilde{f} = \frac{\partial}{\partial x}$, $\tilde{g} = (1-x)\frac{\partial}{\partial y} + x^2\frac{\partial}{\partial z} + (x+y^2)\frac{\partial}{\partial w}$. The metric on \tilde{E} is defined by stipulating that \tilde{f} and \tilde{g} form an orthonormal basis of sections. The four vector fields \tilde{f} , \tilde{g} , $[\tilde{f}, \tilde{g}]$ and $[\tilde{f}, [\tilde{f}, \tilde{g}]]$ are linearly independent everywhere, so \tilde{E} is regular. Let $\gamma^* : [0, T] \rightarrow \mathbb{R}^3$ be one of our abnormal extremals for our three-dimensional problem, given by $\gamma^*(t) = (0, \bar{y} + t, \bar{z})$, and let $\tilde{\gamma}^* : [0, T] \rightarrow \mathbb{R}^4$ be a trajectory of the new system that projects down to γ^* . It is then easily shown that $\tilde{\gamma}^*$ is a strictly abnormal extremal. (Indeed, the Hamiltonian for our new problem is $\tilde{H} = \xi u + (\eta(1-x) + \zeta x^2 + \mu(x+y^2))v$, where μ is a new adjoint variable dual to w . The adjoint equations say that ζ and μ are constant, and $\dot{\xi} = \eta - \mu - 2\zeta x$, $\dot{\eta} = -2\mu y$. Since $x(t) \equiv 0$ along $\tilde{\gamma}^*$ we can satisfy the adjoint equations by taking $\xi \equiv \eta \equiv \mu \equiv 0$, $\zeta = 1$. With this choice, the switching functions ξ and $\eta(1-x) + \zeta x^2 + \mu y^2$ vanish identically, so $\tilde{\gamma}^*$ is an abnormal extremal. On the other hand, if $\tilde{\gamma}^*$ was a normal extremal as well, then there would exist a nontrivial solution of the adjoint equations that gives rise to a nonzero value of \tilde{H} and satisfies the minimization condition. The latter condition implies that $\xi \equiv 0$. Since $\dot{\xi} = \eta - \mu - 2\zeta x = \eta - \mu$, we have $\eta = \mu$, so η is a constant as well. Since $\dot{\eta} = -2\mu y$, and y does not vanish identically, we have $\mu = 0$, and then $\eta = 0$. So the value of the Hamiltonian is $\zeta x^2 v$. i.e. 0. This is a contradiction.) Finally, it is obvious that if γ^* is optimal then $\tilde{\gamma}^*$ is optimal as well. ■

§4. The optimality proof .

We now show that *every sufficiently short abnormal extremal along which $x \equiv 0$ is uniquely optimal*. Precisely, we will show that an abnormal extremal γ^* of length T along which $x \equiv 0$ and v is either $\equiv 1$ or $\equiv -1$ is uniquely optimal, if $T \leq \frac{2}{3}$. We will just study the case $v \equiv 1$. (The other case is similar.) And we will always choose the starting time to be zero, so γ^* is, for some $T > 0$, a function from $[0, T]$ to \mathbb{R}^3 given by $\gamma^*(t) = (0, \bar{y} + t, \bar{z})$. The optimality conclusion follows trivially from the following inequality:

LEMMA: Let $\bar{\tau} = \frac{2}{3}$. Let $0 < \tau \leq \bar{\tau}$, and let $u, v : [0, \tau] \rightarrow \mathbb{R}$ be two measurable functions such that $|u(t)| \leq 1$ and $|v(t)| \leq 1$ for almost all t . Define $x : [0, \tau] \rightarrow \mathbb{R}$ by $x(t) = \int_0^t u(s) ds$. Assume that $x(\tau) = 0$ and $\int_0^\tau x(t)^2 v(t) dt = 0$. Then $\int_0^\tau (1-x(t))v(t) \leq \tau$, and equality holds if and only if $u(t) = 0$ and $v(t) = 1$ for almost all $t \in [0, \tau]$.

To see that this lemma implies the optimality result, let $\gamma : t \rightarrow (x(t), y(t), z(t))$ be any trajectory of our system, defined on an interval $[0, \tau]$, that goes from $(0, \bar{y}, \bar{z})$ to $(0, \bar{y} + T, \bar{z})$, and corresponds to controls $u(\cdot), v(\cdot)$. Then the functions $u(\cdot), v(\cdot), x(\cdot)$ satisfy the hypotheses of the lemma, and therefore the integral $\int_0^\tau (1 - x(t))v(t)dt$, which is equal to $y(\tau) - y(0)$, i.e. to T , is $\leq \tau$, provided that $\tau \leq \frac{2}{3}$. Therefore $T \leq \tau$, if $T \leq \frac{2}{3}$. Moreover, the equality $T = \tau$ can only hold if $v \equiv 1, u \equiv 0$, i.e. if γ coincides with γ^* . Therefore *the restriction of γ^* to any interval of length $\leq \frac{2}{3}$ is uniquely optimal.*

PROOF OF THE LEMMA. Write $A = \int_0^\tau (1 - x(t))v(t)dt$. Let $h(t) = \int_0^t v(s)ds$. Then $-\tau \leq h(\tau) \leq \tau$. Let $\alpha = \tau - h(\tau)$, $\beta = \sup\{|x(t)| : t \in [0, \tau]\}$. Since $|v(t)| \leq 1$, we have

$$\left| \int_0^\tau x(s)v(s)ds \right| \leq \beta\tau ,$$

so that

$$A \leq \int_0^\tau v(s)ds + \left| \int_0^\tau x(s)v(s)ds \right| \leq h(\tau) + \beta\tau = h(\tau) - \tau + \tau + \beta\tau = \tau - \alpha + \beta\tau.$$

Our conclusion will follow if we show that $\beta\tau \leq \alpha$. If we prove that $\beta \leq \frac{3}{2}\alpha$, then $\beta\tau \leq \frac{3}{2}\alpha\tau$, which implies $\beta\tau \leq \alpha$, since $\tau \leq \frac{2}{3}$.

To prove that $\beta \leq \frac{3}{2}\alpha$, we show that

$$\frac{2\beta^3}{3} \leq \int_0^\tau x(t)^2 dt \leq \beta^2 \alpha .$$

The upper bound for $\int_0^\tau x(t)^2 dt$ follows by observing that, since $\int_0^\tau x(t)^2 v(t)dt = 0$, we have

$$\int_0^\tau x(t)^2 dt = \int_0^\tau x(t)^2(1 - v(t))dt \leq \beta^2 \int_0^\tau (1 - v(t))dt = \beta^2(\tau - h(\tau)) = \beta^2 \alpha .$$

To prove the lower bound, we pick a point $a \in [0, \tau]$ such that $|x(a)| = \beta$. Then clearly $a \geq \beta$ and $\tau - a \geq \beta$, since $x(0) = 0, x(\tau) = 0$, and $|\dot{x}(t)| \leq 1$. So the intervals $I_1 = [a - \beta, a]$, $I_2 = [a, a + \beta]$ are entirely contained in $[0, \tau]$. On each of the two intervals I_j , the function $|x|$ is bounded below by the linear function φ_j that takes the value β at a and is 0 at the other endpoint. Clearly, $\int_{I_j} \varphi_j^2 = \frac{\beta^3}{3}$. So $\int_{I_j} x(t)^2 dt \geq \frac{\beta^3}{3}$. Therefore

$$\int_0^\tau x(t)^2 dt \geq \int_{I_1} x(t)^2 dt + \int_{I_2} x(t)^2 dt \geq \frac{2\beta^3}{3} .$$

This concludes the proof. ■

REMARK 3. The functions $u(\cdot)$, $v(\cdot)$ that correspond to a trajectory of Σ_1 satisfy $u^2 + v^2 \leq 1$. However, the only properties of these functions that are used in our optimality proof are the inequalities $|u| \leq 1$, $|v| \leq 1$. So we have in fact proved a stronger result, namely, that γ^* is time-optimal —if its length is $\leq \frac{2}{3}$ — within a larger class of trajectories, namely, those of the control problem Σ_2 obtained from Σ by imposing the new control constraint $|u| \leq 1$, $|v| \leq 1$.

REMARK 4. The condition $x(\tau) = 0$ was only used to conclude that $a + \beta \leq \tau$, forcing the interval I_2 to be contained in $[0, \tau]$. If we do not assume that $x(\tau) = 0$, then all the other steps of the proof remain valid, except that now we can only bound $\int_0^\tau x^2$ from below by $\int_{I_1} x^2$, which is $\geq \frac{\beta^3}{3}$. So the lemma is still valid even without the hypothesis $x(\tau) = 0$, provided that in this case we take $\bar{\tau} = \frac{1}{3}$ rather than $\bar{\tau} = \frac{2}{3}$.

REFERENCES

- [1] L.D. Berkovitz, *Optimal Control Theory*, Springer-Verlag, New York, 1974.
- [2] R.W. Brockett, “Control Theory and Singular Riemannian Geometry,” in *New Directions in Applied Mathematics* (P.J. Hilton and G.S. Young, eds.), Springer-Verlag, (1981).
- [3] I. Kupka, “Abnormal extremals,” preprint, 1992.
- [4] E.B. Lee and L. Markus, *Foundations of Optimal Control Theory*, Wiley, New York, 1968.
- [5] R. Montgomery, “Geodesics which do not Satisfy the Geodesic Equations,” 1991 preprint.
- [6] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mischenko, *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.
- [7] R. Strichartz, “Sub-Riemannian Geometry,” *J. Diff. Geom.* 24, 221-263, (1983).
- [8] R. Strichartz, “Corrections to ‘Sub-Riemannian Geometry’,” *J. Diff. Geom.* 30, no. 2, 595-596, (1989).