# OPTIMAL CONTROL OF NONSMOOTH SYSTEMS WITH CLASSICALLY DIFFERENTIABLE FLOW MAPS

**Héctor J. Sussmann** [*,1]

[*] *Department of Mathematics, Rutgers University, Piscataway, NJ 08854, U.S.A.*

Abstract: We present a version of the Pontryagin Maximum Principle for control dynamics with a possibly non smooth, nonlipschitz and even discontinuous right-hand side. The usual adjoint equation, where state derivatives occur, is replaced by an integrated form, containing only differentials of the reference flow maps. The resulting "integrated adjoint equation" leads to "adjoint vectors" that need not be absolutely continuous, and could be discontinuous and unbounded. We illustrate this with the "reflected brachistochrone problem," for which the adjoint vectors have a singularity at an interior point of the interval of definition of the reference trajectory.

Keywords: Optimal control, minimum-time control

## 1. INTRODUCTION

We consider autonomous *Lagrangian optimization problems* in which it is desired to minimize an integral $J = \int_{\tau_-(\xi)}^{\tau_+(\xi)} f_0(\xi(t),\eta(t))\,dt$, subject to a dynamical constraint $\dot\xi(t) = f(\xi(t),\eta(t))$ and an endpoint condition $\partial\xi \in S$. Here the state $x$ takes values in an open subset $\Omega$ of $\mathbb{R}^n$, the control $u$ has values in a set $U$, $S$ (the "endpoint constraint set") is a given subset of $\Omega\times\Omega$, $(\xi,\eta)$ is a trajectory-control pair, i.e., a pair consisting of an open-loop control $\eta$ and a corresponding trajectory $\xi$, $\tau_-(\xi)$, $\tau_+(\xi)$ are the initial and terminal times of the trajectory $\xi$, and we write $\partial_-\xi \stackrel{\text{def}}{=} \xi(\tau_-(\xi))\,,\ \partial_+\xi \stackrel{\text{def}}{=} \xi(\tau_+(\xi))\,,\ \partial\xi \stackrel{\text{def}}{=} (\partial_-\xi,\partial_+\xi)\,.$ Under suitable smoothness conditions on the map $\Omega\times U \ni (x,u) \mapsto F(x,u) \in \mathbb{R}^{n+1} \sim \mathbb{R}\times\mathbb{R}^n$ (where $F(x,u) \stackrel{\text{def}}{=} (f_0(x,u), f(x,u)))$ and the set $S$ (for example, if $f_0$ and $f$ are of class $C^1$ with respect to $x$ for each $u$, some extra technical conditions are satisfied for the dependence on $u$, and $S$ is a smooth submanifold, or a closed convex set, or, more generally, a set locally equivalent to a closed convex set by means of a diffeomorphism of class $C^1$) one can write the "adjoint equation"

$$\dot\pi = -\pi\cdot\frac{\partial f}{\partial x} + \pi_0\frac{\partial f_0}{\partial x}\ \Big(\ = -\frac{\partial H}{\partial x}\Big) \qquad (1)$$

for a row-vector-valued function $t \mapsto \pi(t) \in \mathbb{R}_n$ and a nonnegative real constant $\pi_0$. If we let $\mathcal{M}_H(x,p_0,p) = \max\{H(x,u,p_0,p) : u \in U\}$, then the classical Pontryagin Maximum Principle says that if $(\xi,\eta)$ is optimal then there exists a nontrivial solution $(\pi_0,\pi)$ of (1) for which the "Hamiltonian maximization condition"

$$H(\xi(t),\eta(t),\pi_0,\pi(t)) = \mathcal{M}_H(\xi(t),\pi_0,\pi(t)) \quad (2)$$

holds, as well as the"transversality condition"

$$(v,w) \in C \Longrightarrow -\langle\partial_-\pi, v\rangle + \langle\partial_+\pi, w\rangle \geq 0\,. \quad (3)$$

Here $H$ is the Hamiltonian, defined by

$$H(x,u,p_0,p) = p\cdot f(x,u) - p_0 f_0(x,u)\,, \qquad (4)$$

[1] Partially supported by NSF Grant DMS01-03901

and $C$ is the Bouligand tangent cone to $S$ at $\partial\xi$. (Recall that the *Bouligand tangent cone* to a subset $A$ of $\mathbb{R}^m$ at a point $a \in A$ is the set of all vectors $v \in \mathbb{R}^m$ such that $v = \lim_{j\to 0} \frac{a_j - a}{h_j}$ for some sequence $\{a_j\}$ of points of $A$ that converges to $a$ and some sequence $\{h_j\}$ of positive reals that converges to 0.)

A similar necessary condition can be derived if $f$ and $f_0$ are just Lipschitz with respect to $x$ (with appropriate conditions on the $u$-dependence). All that is needed is to replace the adjoint differential equation (1) by the adjoint differential inclusion $\dot{\pi}(t) \in -\partial_x H(\xi(t), \eta(t), \pi_0, \pi(t))$, where we use $\partial_x H(\xi(t), \eta(t), \pi_0, \pi(t))$ to denote the Clarke generalized gradient at the point $\xi(t)$ of the Lipschitz function $x \mapsto H(x, \eta(t), \pi_0, \pi(t))$. In this "nonsmooth maximum principle" the adjoint vector $\pi$ is still absolutely continuous, as in the classical case.

The purpose of this note is to present, and illustrate with a very classical example, another "nonsmooth" version of the Maximum Principle in which $f$ and $f_0$ are allowed to be even less smooth than Lipschitz (and could even be discontinuous), and **the adjoint vector can fail to be absolutely continuous and can be discontinuous and even unbounded**, but is a solution of a perfectly well defined "integrated" version of the adjoint equation.

## 2. THE MAIN THEOREM

To state our generalization of the maximum principle, we will need some definitions.

If $E$ is a totally ordered set, with ordering $\preceq$, we use $E^{\preceq,2}$ to denote the set of all ordered pairs $(s,t) \in E \times E$ such that $s \preceq t$, and write $E^{\preceq,3}$ to denote the set of all ordered triples $(r,s,t) \in E \times E \times E$ such that $r \preceq s \preceq t$.

If $S$ is a set, then $\mathbb{I}_S$ will denote the identity map of $S$. If $A$, $B$ are sets, then the notations $f : A \hookrightarrow B$, $f : A \mapsto B$ will indicate, respectively, that $f$ is a possibly partially defined (abbr. "ppd") map from $A$ to $B$ and that $f$ is an everywhere defined map from $A$ to $B$.

*Definition 2.1.* Let $E$ be a totally ordered set with ordering $\preceq$, and let $\Omega$ be a set. A *flow* on $\Omega$ with time set $E$ is a family $\Phi = \{\Phi_{t,s}\}_{(s,t)\in E^{\preceq,2}}$ of ppd maps from $\Omega$ to $\Omega$ such that

(F1) $\Phi_{t,s} \circ \Phi_{s,r} = \Phi_{t,r}$ *whenever* $(r,s,t) \in E^{\preceq,3}$,
(F2) $\Phi_{t,t} = \mathbb{I}_\Omega$ *whenever* $t \in E$.

If $\Phi$ is a flow on $\Omega$ with time set $E$, a *real augmentation* of $\Phi$ is a family $c = \{c_{t,s}\}_{(s,t)\in E^{\preceq,2}}$ of ppd functions from $\Omega$ to $\mathbb{R}$ such that

(RA) $c_{t,r}(x) = c_{s,r}(x) + c_{t,s}(\Phi_{s,r}(x))$ *whenever* $x \in \Omega$ *and* $(r,s,t) \in E^{\preceq,3}$.

A *real-augmented flow* on $\Omega$ with time set $E$ is a pair $(\Phi, c)$ such that $\Phi$ is a flow on $\Omega$ with time set $E$ and $c$ is a real augmentation of $\Phi$. $\diamondsuit$

To any real-augmented flow $(\Phi, c)$ on a set $\Omega$ with time set $E$ we can associate a family of mappings $\Phi^c_{t,s} : \Omega^{aug} \hookrightarrow \Omega^{aug}$—where $\Omega^{aug} = \mathbb{R} \times \Omega$—by letting $\Phi^c_{t,s}(x_0, x) = (x_0 + c_{t,s}(x), \Phi_{t,s}(x))$ for each $(s,t) \in E^{\preceq,2}$. Then $\Phi^c = \{\Phi^c_{t,s}\}_{(s,t)\in E^{\preceq,2}}$ is a flow on $\Omega^{aug}$.

*Definition 2.2.* If $U$ is a set, a *$U$-control* is a mapping $\eta : [a,b] \mapsto U$ defined on some nonempty compact subinterval $[a,b] \stackrel{\text{def}}{=} \text{Dom}(\eta)$ of $\mathbb{R}$.

Suppose that $\eta, \tilde{\eta}$ are $U$-controls, $\text{Dom}(\eta) = [a,b]$, $h \in \mathbb{R}$, and $h \geq 0$. We say that *$\tilde{\eta}$ is obtained from $\eta$ by an equal time constant control interval replacement of length $h$* if (a) $\text{Dom}(\eta) = \text{Dom}(\tilde{\eta})$, and (b) there exist a subinterval $J$ of $[a,b]$ of length $h$ and a $u \in U$ such that $\tilde{\eta}(t) = \eta(t)$ whenever $t \in [a.b]\backslash J$ and $\tilde{\eta}(t) = u$ whenever $t \in J$. We say that *$\tilde{\eta}$ is obtained from $\eta$ by an interval deletion of length $h$* if $\text{Dom}(\tilde{\eta}) = [a, b-h]$ and there exists $\alpha$ such that (a) $a \leq \alpha \leq b-h$, (b) $\tilde{\eta}(t) = \eta(t)$ whenever $a \leq t < \alpha$, and (c) $\tilde{\eta}(t) = \eta(t+h)$ whenever $\alpha \leq t \leq b-h$. Finally, we will say that *$\tilde{\eta}$ is obtained from $\eta$ by a constant control interval insertion of length $h$* if $\text{Dom}(\tilde{\eta}) = [a, b+h]$ and there exist $\alpha$, $u$ such that $a \leq \alpha \leq b$ and $u \in U$, for which (a) $\tilde{\eta}(t) = \eta(t)$ whenever $a \leq t < \alpha$, (b) $\tilde{\eta}(t) = \eta(t-h)$ whenever $\alpha + h < t \leq b+h$, and (c) $\tilde{\eta}(t) = u$ whenever $\alpha \leq t \leq \alpha + h$. If $m \in \mathbb{Z}$, $h \in \mathbb{R}$, $m \geq 0$, and $h \geq 0$, we say that *$\tilde{\eta}$ is obtained from $\eta$ by $m$ variable time constant control interval operations of total length $h$* if there exist $U$-controls $\tilde{\eta}_0, \ldots, \tilde{\eta}_m$ and nonnegative numbers $h_1, \ldots, h_m$ such that $h = h_1 + \cdots + h_m$, $\tilde{\eta}_0 = \eta$, and $\tilde{\eta}_m = \tilde{\eta}$, for which $\tilde{\eta}_j$ is obtained from $\tilde{\eta}_{j-1}$, for $j = 1, \ldots, m$, by either (a) an equal time constant control interval replacement of length $h_j$, or (b) an interval deletion of length $h_j$, or (c) a constant control interval insertion of length $h_j$. If in addition $\tilde{\eta}$ can be obtained in this way using only the operation of Item (a), then we say that *$\tilde{\eta}$ is obtained from $\eta$ by $m$ equal time constant control interval operations of total length $h$.*

An *equal time variational neighborhood* of a $U$-control $\eta$ is a set $\mathcal{U}$ of $U$-controls having the property that whenever $m \in \mathbb{Z}$ and $m \geq 0$ there exists a positive $\bar{h}(m) \in \mathbb{R}$ such that, whenever a $U$-control $\tilde{\eta}$ is obtained from $\eta$ by $m$ equal time constant control interval replacements of total length $\leq \bar{h}(m)$ it follows that $\tilde{\eta} \in \mathcal{U}$. If, in addition, $\bar{h}(m)$ can be chosen, for every $m$, so that

that $\tilde{\eta} \in \mathcal{U}$ whenever $\tilde{\eta}$ is obtained from $\eta$ by $m$ variable time constant control interval operations of total length $\leq \bar{h}(m)$, then we call $\mathcal{U}$ a *variable time variational neighborhood* of $\eta$. ♢

We now consider the optimal control problem described in the introduction, and assume:

(A1) *$n \in \mathbb{Z}$, $n \geq 0$, $\Omega$ is open in $\mathbb{R}^n$, $U$ is a set, $a_*, b_* \in \mathbb{R}$, and $a_* \leq b_*$,*
(A2) *$\eta_*$ (the "reference control") is a $U$-control with domain $[a_*, b_*]$,*
(A3) *$f$, $f_0$, $F$ are maps on $\Omega \times U$ with values in $\mathbb{R}^n$, $\mathbb{R}$, $\mathbb{R}^{n+1}$ respectively, such that $F(x,u) = (f_0(x,u), f(x,u))$ if $x \in \Omega$, $u \in U$,*
(A4) *$\xi_* : [a_*, b_*] \mapsto \Omega$ is a trajectory of the system $\dot{x} = f(x,u)$ corresponding to $\eta_*$ (i.e., $\xi_*$ is absolutely continuous and satisfies $\dot{\xi}_*(t) = f(\xi_*(t), \eta_*(t))$ for a.e. $t \in [a_*, b_*]$).*
(A5) *$[a_*, b_*] \ni t \mapsto f_0(\xi_*(t), \eta_*(t)) \in \mathbb{R}$ is a Lebesgue integrable function.*
(A6) *The map $t \mapsto F(x, \eta_*(t))$ is measurable for each $x$.*
(A7) *$E$ is a relatively open subset of $[a_*, b_*]$ such that $a_* \in E$ and $b_* \in E$, and $(\Phi, c)$ is a real-augmented flow on $\Omega$ with time set $E$, for which*
(A7.a) *$(\Phi, c)$ is the restriction to $E$ of the augmented flow of the reference control (that is, if $x, \tilde{x} \in \Omega$, $y \in \mathbb{R}$, $s, t$ $, inE,$ and $s \leq t$, then $(\tilde{x}, y) = (\Phi_{t,s}(x), c_{t,s}(x))$ iff there exists an absolutely continuous curve $\xi : [s,t] \mapsto \Omega$ such that $\dot{\xi}(r) = f(\xi(r), \eta_*(r))$ for almost all $r \in [s,t]$, $\xi(s) = x$, $\xi(t) = \tilde{x}$, and $y = \int_s^t f_0(\xi(r), \eta_*(r))\, dr$),*
(A7.b) *if $\bar{s}, \bar{t} \in E$ and $\bar{s} \leq \bar{t}$, then the maps $(s,t,x) \hookrightarrow \Phi_{t,s}(x)$ and $(s,t,x) \hookrightarrow c_{t,s}(x)$ are defined and continuous on a neighborhood of $(\bar{s}, \bar{t}, \xi_*(\bar{s}))$ and are differentiable at $(\bar{s}, \bar{t}, \xi_*(\bar{s}))$,*
(A7.c) *the map $x \mapsto F(x,u)$ is continuous near $\xi_*(t)$ whenever $(u,t) \in U \times E$,*
(A8) *The set $S \subseteq \Omega \times \Omega$ coincides, near the point $q_* = (\xi_*(a_*), \xi_*(b_*))$, with the image of a closed convex set by a diffeomorphism of class $C^1$, and $C$ is the Bouligand tangent cone to $S$ at $q_*$.*
(A9) *The class $\mathcal{U}$ of admissible controls is an equal time variational neighborhood of $\eta_*$.*

*Theorem 2.3.* If (A1,...,9) hold, and $(\xi_*, \eta_*)$ is a solution of the optimal control problem, then there exist a map $E \ni t \mapsto \pi(t) \in \mathbb{R}_n$ and a $\pi_0 \in \mathbb{R}$ such that $\pi_0 \geq 0$ and

(C1) $(\pi(t), \pi_0) \neq (0,0)$ for all $t \in E$,
(C2) $\pi(s) = \pi(t) D\Phi_{t,s}(\xi_*(s)) - \pi_0 \nabla c_{t,s}(\xi_*(s))$ whenever $s, t \in E$ and $s \leq t$,
(C3) $H(\xi_*(t), \eta_*(t), \pi_0, \pi(t)) \geq H(\xi_*(t), u, \pi_0, \pi(t))$ whenever $t \in E_{diff}$, $u \in U$, where $H$ is the Hamiltonian, defined by (4), and $E_{diff}$ is the set of those $t \in E$ such that $\dot{\xi}_*(t) = f(\xi_*(t), \eta_*(t))$,
(C4) the transversality condition (3) holds.

If $\mathcal{U}$ is a variable time variational neighborhood of $\eta_*$, then $\pi$ and $\pi_0$ can be chosen so that (C1,2,3,4) hold and in addition

(C5) *The equality $H(\xi_*(t), \eta_*(t), \pi_0, \pi(t)) = 0$ holds for every $t \in E_{diff}$.* ♢

*Remark 2.4.* The five conclusions (C1), (C2), (C3), (C4), and (C5) are known, respectively, as the *nontriviality condition*, the *integrated adjoint equation*, the *Hamiltonian maximization condition*, the *transversality condition*, and the *vanishing Hamiltonian condition*, and will be referred to by the corresponding acronyms "NTC," "IAE," "HMC," "TC" and "VHC." ♢

*Remark 2.5.* Theorem 2.3 contains the classical "smooth" maximum principle because, when $f$ and $f_0$ are of class $C^1$ with respect to $x$ (and satisfy some extra regularity conditions for the dependence on $u$) then all our hypotheses hold, and the resulting IAE reduces to the classical adjoint equation by differentiating both sides with respect to $s$. On the other hand, our result, as stated, clearly does not contain the nonsmooth maximum principle for Lipschitz righ-hand sides. There is, however, a more general version, in which suitable "generalized differentials" are used instead of the ordinary differential. This theorem does contain the non-smooth result, which turns out to be a special case corresponding to the choice of the "Warga derivate containers" as the concept of generalized differential. ♢

## 3. THE REFLECTED BRACHISTOCHRONE

As an example of a nontrivial application of Theorem 2.3, we let $\mathcal{P}$ be the minimum time problem for the dynamical law

$$\dot{x} = u\sqrt{|y|}\,, \qquad \dot{y} = v\sqrt{|y|}\,,$$

with state $(x,y) \in \mathbb{R}^2$ and control $(u,v) \in \mathbb{R}^2$ subject to the control constraint $u^2 + v^2 \leq 1$. Given points $A, B \in \mathbb{R}^2$, we want to characterize the minimum-time trajectory from $A$ to $B$.

*Remark 3.1.* For minimum time problems such as $\mathcal{P}$, $\mathcal{U}$ is a variable time variational neighborhood of $\eta_*$, and $f_0(x,u) \equiv 1$. Then $c_{t,s}(x) = t - s$, so $c_{t,s}$ is independent of $x$, and then the IAE becomes the simpler statement that $\pi(s) = \pi(t) D\Phi_{t,s}(\xi_*(s))$ whenever $s, t \in E$ and $s \leq t$. Furthermore, if we define $h(x,u,p) = \langle p, f(x,u) \rangle$, then the HMC just says that $h(\xi_*(t), \eta_*(t), \pi(t)) \geq h(\xi_*(t), u, \pi(t))$ whenever $t \in E$, $u \in U$. The last conclusion of the theorem then says that the function $E \ni t \mapsto h(\xi_*(t), \eta_*(t), \pi(t))$ is constant and has a nonnegative value. ♢

To solve $\mathcal{P}$, we use Theorem 2.3 together with the classical (1696-7) results about the solutions of the *brachistochrone problem* (abbr. "BP") of Johann Bernouli. Define closed half-planes $H^+$, $H^-$, by $H^+ = \{(x,y) : y \geq 0\}$, $H^- = \{(x,y) : y \leq 0\}$. Let $\mathcal{P}^+$, resp. $\mathcal{P}^-$, be the minimum time problems for curves entirely contained in $H^+$ (resp. $H^-$) with endpoints in $H^+$ (resp. $H^-$). Define a "v-cycloid" to be an arc which is entirely contained in $H^+$ or $H^-$ and is either (a) a vertical line segment or (b) a cycloid generated by a point $P$ on a circle $\Gamma$ that is tangent to the $x$ axis and rolls without slipping. (In particular, if $H = H^+$ or $H = H^-$, and $\xi_* : [0,T] \mapsto H$ is a v-cycloid, then $\xi_*(t) \notin H^+ \cap H^-$ whenever $0 < t < T$.) Then it is well known that the solutions of $\mathcal{P}^+$ and $\mathcal{P}^-$ are v-cycloids.

We now solve $\mathcal{P}$. Let $\xi_* : [0,T] \mapsto \mathbb{R}^2$ be a solution of $\mathcal{P}$ with endpoints $A$, $B$. If $\xi_*$ is entirely contained in $H^+$ or $H^-$, then $\xi_*$ is a solution of $\mathcal{P}^+$ or of $\mathcal{P}^-$, so $\xi_*$ is a v-cycloid. So all we need is to determine the minimum-time trajectories $\xi_*$ that are not entirely contained in $H^+$ or $H^-$. Fix one such $\xi_*$. Then there must exist a time $\tau$ such that $0 \leq \tau \leq T$ and $\xi_*(\tau) \in H^+ \cap H^-$. It is then easy to show that $\tau$ is unique. (If $\tau$ was not unique, let $\tau_1$ be the smallest $t$ such that $\xi_*(t) \in H^+ \cap H^-$, and let $\tau_2$ be the largest. Then $0 \leq \tau_1 < \tau_2 \leq T$, $\xi_*(\tau_i) \in H^+ \cap H^-$ for $i = 1,2$, and $\xi_*(t) \notin H^+ \cap H^-$ for $0 \leq t < \tau_1$ or $\tau_2 < t \leq T$. Assume, without loss of generality, that $\xi_*(t) \in H^+$ for $0 \leq t \leq \tau_1$. Then the set $S = \{t \in [\tau_1, \tau_2] : \xi_*(t) \notin H^+ \cap H^-\}$ is open, so it is a union of a finite or countable set $\mathcal{I}$ of pairwise disjoint open intervals, each one of which is of the form $]\,\alpha, \beta\,[$, with $\tau_1 \leq \alpha < \beta \leq \tau_2$, $\xi_*(\alpha) \in H^+ \cap H^-$, and $\xi_*(\beta) \in H^+ \cap H^-$. If $I$ is one of those intervals, then either $\xi_*(t) \in H^+ \backslash H^-$ for all $t \in I$ or $\xi_*(t) \in H^- \backslash H^+$ for all $t \in I$. In the latter case, we may replace the restriction of $\xi_*$ to $I$ by its reflection with respect to the $x$ axis without changing the time. If we do this for all $I \in \mathcal{I}$, we obtain a new trajectory $\tilde{\xi}_*$ that goes from $A$ to $B$ in the same time as $\xi_*$ and is such that $\tilde{\xi}_*(t) \in H^+ \backslash H^-$ for all $t \in I$ for all $I \in \mathcal{I}$. Then the restriction $\hat{\xi}_*$ of $\tilde{\xi}_*$ to the interval $[0, \tau_2]$ is a time-optimal trajectory that goes from $A$ to $\xi_*(\tau_2)$ and is entirely contained in $H^+$. Hence $\hat{\xi}_*$ is a v-cycloid, and $\hat{\xi}_*(t)$ can only belong to the $x$ axis when $t$ is one of the endpoints of $[0, \tau_2]$. Since $\hat{\xi}_*(\tau_1) \in H^+ \cap H^-$, and $\tau_1 < \tau_2$, it follows that $\tau_1 = 0$. A similar argument shows that $\tau_2 = T$. Hence both $A$ and $B$ belong to $H^+ \cap H^-$. It then follows that $\tilde{\xi}$ is a solution of $\mathcal{P}^+$ with endpoints $A$, $B$. So $\tilde{\xi}(t) \notin H^+ \cap H^-$ whenever $0 < t < T$, and this implies, given our construction of $\tilde{\xi}$ from $\xi$ by reflections, that $\tilde{\xi}$ is either $\xi$ itself or its reflection with respect to the $x$ axis. In either case, $\xi$ is entirely contained in one of the half-planes $H^+$, $H^-$, which is a contradiction.)

Let $\bar{\tau}$ be the unique $\tau$ such that $0 \leq \tau \leq T$ and $\xi_*(\tau) \in H^+ \cap H^-$. Then $0 < \bar{\tau} < T$, and the points $A$ and $B$ belong to different sides of the $x$ axis. (Indeed, if $\bar{\tau} = 0$ then $\xi_*(t)$ would belong to one of $H^+$, $H^-$ whenever $0 < t \leq T$, so $\xi_*$ would be entirely contained in $H^+$ or $H^-$. A similar contradiction would arise if $\bar{\tau} = T$. So $0 < \bar{\tau} < T$. If $A$ and $B$ were both in $H^+$, then $\xi_*(t) \in H^+$ for $0 \leq t < \bar{\tau}$ and also for $\bar{\tau} < t \leq T$, so once again $\xi_*$ would be entirely contained in $H^+$. A similar contradiction arises if $A \in H^-$ and $B \in H^-$.) So without loss of generality we may assume that $A \in H^+ \backslash H^-$ and $B \in H^- \backslash H^+$. Then $\xi_*(t) \in H^+ \backslash H^-$ whenever $0 \leq t < \bar{\tau}$ and $\xi_*(t) \in H^- \backslash H^+$ whenever $\bar{\tau} < t \leq T$. So $\xi_*$ is the concatenation of two time-optimal curves $\xi_*^+ : [0, \bar{\tau}] \mapsto H^+$, $\xi_*^- : [\bar{\tau}, T] \mapsto H^-$. Then $\xi_*^+$ and $\xi_*^-$ are v-cycloids contained in $H^+$ and $H^-$.

Let us assume that $\xi_*^+$ and $\xi_*^-$, are both arcs of cycloids. Let $C_*$ be the point where $\xi_*$ crosses the $x$ axis, so $C_* = \xi_*(\bar{\tau})$. Then the necessary conditions of the classical maximum principle do not determine $C_*$, because they only apply on the intervals $\{t : 0 \leq t < \bar{\tau}\}$, $\{t : \bar{\tau} < t \leq T\}$, and say nothing about what happens at time $\bar{\tau}$, where our controlled dynamics is not of class $C^1$. We will now show how Theorem 2.3 yields an extra condition that determines $C_*$.

Our first step is to embed $\xi_*$ in a flow arising from a feedback control law. The arcs $\xi_*^+$, $\xi_*^-$, are parts of full cycloid arcs $\Xi_*^+$, $\Xi_*^-$, such that $\Xi_*^+$ goes from a point $Q^+$ on the $x$ axis to the point $C_*$ and has the property that all the other points of $\Xi_*^+$ belong to $H^+ \backslash H^-$, while $\Xi_*^-$ goes from $C_*$ to a point $Q^-$ on the $x$ axis and is such that all the other points of $\Xi_*^-$ belong to $H^- \backslash H^+$. Write $Q^+ = (\alpha^+, 0)$, $Q^- = (\alpha^-, 0)$, $C_* = (\alpha^0, 0)$.

The arcs $\Xi_*^+$, $\Xi_*^-$, are the loci of points $P^+$, $P^-$, attached to rolling circles $\Gamma^+$, $\Gamma^-$, of radii $R^+$, $R^-$, and then $|\alpha^0 - \alpha^+| = 2\pi R^+$ and $|\alpha^0 - \alpha^-| = 2\pi R^-$. Parametric equations for $\Xi_*^+$ can be written using as parameter the abscissa $\alpha$ of the point where the rolling circle $\Gamma^+$ intersects the $x$ axis $H^+ \cap H^-$. Then $\alpha$ takes values in the interval $I^+ = [\min(\alpha^0, \alpha^+), \max(\alpha^0, \alpha^+)]$, which has length $2\pi R^+$. If we let $\theta \stackrel{\text{def}}{=} (R^+)^{-1}(\alpha - \alpha_+)$, then the position of $P^+$ for a given value of $\alpha$ is $\Xi_*^+(\alpha) = (\alpha - R^+ \sin\theta\,,\, R^+(1 - \cos\theta))$. (The circle $\Gamma^+$ rolls from left to right if $\alpha^+ < \alpha^0$, and from right to left if $\alpha^0 < \alpha^+$.)

The midpoint $\mu^+$ of the interval $I^+$ is given by $\mu^+ = \frac{1}{2}(\alpha^+ + \alpha^0)$. We let $\hat{Q}^+$ be the point where $\Gamma^+$ intersects the $x$ axis when $\alpha = \mu^+$, so that $\hat{Q}^+ = (\mu^+, 0)$. We define parametrized trajectories $\Xi_*^{+,\sigma}$, for each $\sigma$ in a neighborhood

$N^+ = [1-\varepsilon_1^+, 1+\varepsilon_2^+]$ of 1 (where $\varepsilon_1^+, \varepsilon_2^+$ are chosen so that $0 < \varepsilon_1^+ < 1$ and $0 < \varepsilon_2^+$), by letting $\Xi_*^{+,\sigma}(\alpha) = \hat{Q}^+ + \sigma(\Xi_*^+(\alpha) - \hat{Q}^+)$ whenever $\alpha \in I^+$. Then each $\Xi_*^{+,\sigma}$ is an arc of cycloid, generated exactly like $\Xi_*^+$, with $R^+$ replaced by $\sigma R^+$, and having contact points $Q^{+,\sigma}$, $C^{+,\sigma}$ with the $x$ axis, where $Q^{+,\sigma} = \hat{Q}^+ + \sigma(\Xi_*^+(\alpha^+) - \hat{Q}^+)$ and $C^{+,\sigma} = \hat{Q}^+ + \sigma(\Xi_*^+(\alpha^0) - \hat{Q}^+)$, so that $Q^{+,\sigma}$ and $C^{+,\sigma}$ are given by $Q^{+,\sigma} = \Big((1-\sigma)\mu^+ + \sigma\alpha^+, 0\Big)$,

$$C^{+,\sigma} = \Big((1-\sigma)\mu^+ + \sigma\alpha^0, 0\Big). \tag{5}$$

Then, if we let $\mathcal{S}^+ = \{\Xi_*^{+,\sigma}(\alpha) : \sigma \in N^+, \alpha \in I^+\}$, the set $\mathcal{S}^+$ is clearly the homeomorphic image of the rectangle $\mathcal{R}^+ \stackrel{\text{def}}{=} N^+ \times I^+$ under the map $\Psi^+ : \mathcal{R}^+ \mapsto H^+$ given by $\Psi^+(\sigma, \alpha) \stackrel{\text{def}}{=} \Xi_*^{+,\sigma}(\alpha)$. Furthermore, the two images $\Psi^+(N^+ \times \{\alpha^+\})$, $\Psi^+(N^+ \times \{\alpha^0\})$, are subintervals of the $x$ axis, while the images of all the points of $\mathcal{R}^+$ that do not belong to $N^+ \times \{\alpha^+, \alpha^0\}$ lie in the open half-plane $H^+ \backslash H^-$. The map $\Psi^+$ is real analytic, and the partial derivatives $\frac{\partial \Psi^+}{\partial \alpha}$, $\frac{\partial \Psi^+}{\partial \sigma}$, are given by the formulas $\frac{\partial \Psi^+}{\partial \alpha} = \sigma \Xi_*^{+\prime}(\alpha)$, $\frac{\partial \Psi^+}{\partial \sigma} = \Xi_*^+(\alpha) - \hat{Q}^+$, where $\Xi_*^{+\prime}(\alpha) = (1 - \cos\theta, \sin\theta)$.

If $J^+$ is the Jacobian determinant of $\Psi^+$ with respect to $\sigma$ and $\alpha$ then a simple calculation shows that $J^+ = 0$ iff $\alpha = \alpha^+$ or $\alpha = \alpha^0$. So

($*$) $\Psi^+$ *is a real analytic diffeomorphism on the set* $N^+ \times \text{Interior}(I^+)$.

We now analyze the time parameter along the curves $\Xi_*^{+,\sigma}$. Let $\delta^+ = +1$ if $\alpha^+ < \alpha^0$ (i.e., if $\Gamma^+$ rolls from left to right, so time increases as $\alpha$ increases, i.e., $dt/d\alpha > 0$), and $\delta^+ = -1$ if $\alpha^+ > \alpha^0$ (i.e., if $\Gamma^+$ rolls from right to left, in which case $dt/d\alpha < 0$). If $\Xi_*^{+,\sigma}(\alpha) = (x(\alpha), y(\alpha))$, then it is easy to see that $dt = 2\sigma(R^+)^{-\frac{1}{2}}\delta^+ \, d\alpha$. It follows that

(#) *the time along the curve* $\Xi_*^{+,\sigma}$ *from* $\Xi_*^{+,\sigma}(\alpha_1)$ *to* $\Xi_*^{+,\sigma}(\alpha_2)$ *is* $2\sigma\delta^+(R^+)^{-\frac{1}{2}}(\alpha_2 - \alpha_1)$.

A similar construction works for $\Xi_*^-$. In this case, the parametric equations turn out to be $\Xi_*^-(\alpha) = (\alpha - R^- \sin\theta, -R^-(1-\cos\theta))$, where the variable $\alpha$ now takes values in the interval $I^- = [\min(\alpha^0, \alpha^-), \max(\alpha^0, \alpha^-)]$ (which has length $2\pi R^-$), and $\theta = \frac{\alpha - \alpha^-}{R^-}$.

The circle $\Gamma^-$ rolls from left to right if $\alpha^0 < \alpha^-$, and from right to left if $\alpha^- < \alpha^0$. (Notice that $\Gamma^-$ rolls from left to right iff it rotates *counterclockwise*, whereas $\Gamma^+$ rolls from left to right iff it rotates clockwise.)

The midpoint of $I^-$ is $\mu^- = \frac{1}{2}(\alpha^- + \alpha^0)$. We let $\hat{Q}^- = (\mu^-, 0)$. We then define parametrized arcs $\Xi_*^{-,\sigma}$, for $\sigma$ in a neighborhood $N^- = [1-\varepsilon_1^-, 1+\varepsilon_2^-]$ of 1 (where $0 < \varepsilon_1^- < 1$ and $0 < \varepsilon_2^-$), by letting $\Xi_*^{-,\sigma}(\alpha) = \hat{Q}^- + \sigma(\Xi_*^-(\alpha) - \hat{Q}^-)$ for $\alpha \in I^-$. Then each $\Xi_*^{-,\sigma}$ is an arc of cycloid, having contact points $Q^{-,\sigma}$, $C^{-,\sigma}$ with the $x$ axis, where $Q^{-,\sigma} = \hat{Q}^- + \sigma(\Xi_*^-(\alpha^-) - \hat{Q}^-)$ and $C^{-,\sigma} = \hat{Q}^- + \sigma(\Xi_*^-(\alpha^0) - \hat{Q}^-)$, so that $Q^{-,\sigma}$ and $C^{-,\sigma}$ are given by $Q^{-,\sigma} = \Big((1-\sigma)\mu^- + \sigma\alpha^-, 0\Big)$,

$$C^{-,\sigma} = \Big((1-\sigma)\mu^- + \sigma\alpha^0, 0\Big). \tag{6}$$

Then, if we let $\mathcal{S}^- = \{\Xi_*^{-,\sigma}(\alpha) : \sigma \in N^-, \alpha \in I^-\}$, it is clear that the set $\mathcal{S}^-$ is the homeomorphic image of $\mathcal{R}^- \stackrel{\text{def}}{=} N^- \times I^-$ under the smooth map $\Psi^- : \mathcal{R}^- \mapsto H^-$ given by $\Psi^-(\sigma, \alpha) \stackrel{\text{def}}{=} \Xi_*^{-,\sigma}(\alpha)$. The Jacobian determinant of $\Psi^-$ vanishes iff $\alpha = \alpha^-$ or $\alpha = \alpha^0$. Hence

($**$) $\Psi^-$ *is a diffeomorphism on* $N^- \times \text{Interior}(I^-)$.

If we let $\delta^- = +1$ if $\alpha^0 < \alpha^-$ (i.e., if $\Gamma^-$ rolls from left to right, in which case $dt/d\alpha > 0$), and $\delta^- = -1$ if $\alpha^- < \alpha^0$ (i.e., if $\Gamma^-$ rolls from right to left, in which case $dt/d\alpha < 0$), then $dt = 2\sigma\delta^-(R^-)^{-\frac{1}{2}} \, d\alpha$, from which it follows that

(##) *the time along* $\Xi_*^{-,\sigma}$ *from* $\Xi_*^{-,\sigma}(\alpha_1)$ *to* $\Xi_*^{-,\sigma}(\alpha_2)$ *is equal to* $2\sigma\delta^-(R^-)^{-\frac{1}{2}}(\alpha_2 - \alpha_1)$.

We now combine the two constructions by letting $\Xi_*^\sigma$ be, for each $\sigma \in I^+$, the concatenation of $\Xi_*^{+,\sigma}$ and $\Xi_*^{-,\hat{\sigma}}$ where $\hat{\sigma}$ is chosen so that $C^{-,\hat{\sigma}} = C^{+,\sigma}$. In view of (5) and (6), it follows that $\hat{\sigma}$ is given in terms of $\sigma$ by $\hat{\sigma} = \zeta(\sigma)$, where

$$\zeta(\sigma) \stackrel{\text{def}}{=} (\alpha^0 - \mu^-)^{-1}(\mu^+ - \mu^- + \sigma(\alpha^0 - \mu^+)). \tag{7}$$

(We guarantee that the map $I^+ \ni \sigma \mapsto \hat{\sigma} \in I^-$ is bijective by choosing the $\varepsilon_j^\pm$ so that $\zeta(I^+) = I^-$.)

We now study the flow maps $\Phi_{t,s}$ associated to this family of trajectories. Let $\mathcal{S} = \mathcal{S}^+ \cup \mathcal{S}^-$. Given any point $q \in \mathcal{S}$, $q$ belongs to the curve $\Xi_*^\sigma$ for a unique $\sigma \in I^+$. If $s, t \in \mathbb{R}$, and $t \geq s$, then we can follow $\Xi_*^\sigma$ in the direction of increasing time, starting at $q$ at time $s$, until we exit $\mathcal{S}$. If $t$ does not exceed the exiting time from $\mathcal{S}$, then $\Phi_{t,s}(q)$ is defined, and equal to the point of $\Xi_*^\sigma$ attained in this way at time $t$. We also define the augmentations $c_{t,s}$ by letting $c_{t,s}(q) = t - s$.

In order to apply Theorem 2.3, we take $E$ to be the set $[0, T] \backslash \{\bar{\tau}\}$. In addition, it will also be convenient to embed our reference trajectory $\xi_*$ in the "extended reference trajectory" $\Xi_* = \Xi_*^1$, that we parametrize by time in such a way that $\Xi_*(\bar{\tau}) = C_*$, so that $\Xi_*(t) = \xi_*(t)$ for $t \in [0, T]$, and $\Xi_*$ is defined on the interval $[\tau_1, \tau_2]$, where $\tau_1 = \bar{\tau} - 4\pi\sqrt{R^+}$ and $\tau_2 = \bar{\tau} + 4\pi\sqrt{R^-}$. Then

(&) *If* $\tau_1 < s \leq t < \tau_2$, *and* $s \neq \bar{\tau} \neq t$, *then* $\Phi_{t,s}$ *is a real analytic diffeomorphism near* $\Xi_*(s)$.

To prove (&), we consider separately three cases, namely, (i) $t < \bar{\tau}$, (ii) $s > \bar{\tau}$, and (iii) $s < \bar{\tau} < t$. In Case (i), $\Xi_*(s)$ clearly belongs to the set $\Psi^+(N^+ \times \mathrm{Interior}(I^+))$, so for $q$ near $\Xi_*(s)$ we can find $\Phi_{t,s}(q)$ by inverting the diffeomorphism $\Psi^+$, letting $(\sigma, \alpha) = (\Psi^+)^{-1}(q)$, defining $\tilde{\alpha}$ by $\tilde{\alpha} = \alpha + \frac{1}{2}\delta^+\sqrt{R^+}(t-s)$, and, finally, writing $\Phi_{t,s}(q) = \Psi^+(\sigma, \tilde{\alpha})$. The conclusion then follows because the map $(\sigma, \alpha) \mapsto (\sigma, \alpha + \frac{\delta^+}{2\sigma}\sqrt{R^+}(t-s))$ is a diffeomorphism. The proof for Case (ii) is similar. Finally, in Case (iii) we can find $\Phi_{s,t}(q)$ by first inverting $\Psi^+$ near $q$ to find $(\sigma, \alpha)$, as in Case (i), and then going from time $s$ to time $t$ by letting $\nu^+ = \delta^+\sqrt{R^+}$, $\nu^- = \delta^-\sqrt{R^-}$, writing

$$\tilde{\sigma} = \zeta(\sigma)\,,\ \ \tilde{\alpha} = \alpha_0 + \frac{\nu^-}{2\tilde{\sigma}}(t-s) - \frac{\sigma\nu^-}{\tilde{\sigma}\nu^+}(\alpha_0 - \alpha)\,, \quad (8)$$

and then defining $\Phi_{t,s}(q) = \Psi^-(\tilde{\sigma}, \tilde{\alpha})$. The map $(\sigma, \alpha) \mapsto (\tilde{\sigma}, \tilde{\alpha})$ defined by (8) is a diffeomorphism, since $\frac{\partial\tilde{\sigma}}{\partial\sigma} = \frac{d\zeta}{d\sigma} \neq 0$ (in view of (7)), $\frac{\partial\tilde{\sigma}}{\partial\alpha} = 0$, and $\frac{\partial\tilde{\alpha}}{\partial\alpha} = \frac{\sigma\nu^-}{\tilde{\sigma}\nu^+} \neq 0$ (because of (8)). So the conclusion follows in Case (iii) as well, and (&) is proved.

For $\tau_1 < s \le t < \tau_2$, $s \neq \bar{\tau} \neq t$, let $D_{t,s}$ be the Jacobian matrix of $\Phi_{t,s}$ at $\Xi_*(s)$. Then we can apply Theorem 2.3 to each of the three curves $\Xi_{*,i} : J^i \mapsto \mathbb{R}^2$, $i = 1, 2, 3$, where $\Xi_{*,1}$ and $\Xi_{*,2}$ are the restrictions of $\Xi_*$ to intervals $J_1$, $J_2$ of the form $[\tau_1 + \delta, \bar{\tau} - \delta]$ and $[\bar{\tau} + \delta, \tau_2 - \delta]$, for some small $\delta$, and $\Xi_{*,3} = \xi_*$, so $J_3 = [0, T]$. (We are assuming that $\Xi_{*,3}$ is time-optimal, and the curves $\Xi_{*,1}$ and $\Xi_{*,2}$ are also optimal because they are solutions of the classical BP.) If we let $\hat{J}_1 = J_1$, $\hat{J}_2 = J_2$, $\hat{J}_3 = E$, then our theorem implies that there exists nontrivial solutions $\hat{J}_i \ni t \mapsto \tilde{\pi}_i(t) \in \mathbb{R}_2$, of the IAE $\pi(s) = \pi(t).D_{t,s}$ such that the HMC holds for every $t \in \hat{J}^i$ and the values of the maximized Hamiltonian are nonnegative constants $\tilde{\pi}_{0,i}$. The HMC at $\Xi_{*,i}(t)$ says that (a) the control $\eta_*(t) = (u_*(t), v_*(t))$ must maximize the product $\tilde{\pi}_i(t) \cdot (u, v)^\dagger$ (where "†" denotes transpose) for $(u, v)$ in the unit disc of $\mathbb{R}_2$, from which it follows that $\|\tilde{\pi}_i(t)\|\eta_*(t) = \tilde{\pi}_i(t)$, and in addition (b) the maximum value $\tilde{\pi}_{0,i}$ of the Hamiltonian is $\sqrt{|y_*(t)|}\,\|\tilde{\pi}_i(t)\|$, if we write $\Xi_{*,i}(t) = (x_*(t), y_*(t))$. This implies, in particular, that $\tilde{\pi}_{0,i} > 0$, because if $\tilde{\pi}_{0,i} = 0$ then $\|\tilde{\pi}_i(t)\| = 0$ (since $\Xi_{*,i}(t) \notin H_+ \cap H_-$ whenever $t \in \hat{J}_i$), and then $\tilde{\pi}_i(t) = 0$, contradicting the NTC. It then follows that $\dot{\Xi}_{*,i}(t)^\dagger = \sqrt{|y_*(t)|}\frac{\tilde{\pi}_i(t)}{\|\tilde{\pi}_i(t)\|}$ for $t \in \hat{J}_i$, so $\langle\tilde{\pi}(t) \cdot w\rangle = 0$ whenever $w \in \mathbb{R}^2$ is orthogonal to $\dot{\Xi}_{*,i}(t)$. Now, if $t, s \in \hat{J}_i$, $s \le t$, $w \in \mathbb{R}^2$, and $\hat{w} = D_{t,s}w$, then $\tilde{\pi}_i(t) \cdot \hat{w} = \tilde{\pi}_i(t) \cdot D_{t,s}(w) = (\tilde{\pi}_i(t) \circ D_{t,s})(w)\,,$ so $\tilde{\pi}_i(t) \cdot \hat{w} = \tilde{\pi}_i(s) \cdot w$. Therefore $\langle\dot{\Xi}_{*,i}(t), D_{t,s}w\rangle = 0$ iff $\langle\dot{\Xi}_{*,i}(s), w\rangle = 0$, and we have obtained the geometric condition

(G1) *If $i \in \{1, 2, 3\}$, $s, t \in \hat{J}_i$, and $s \le t$, the linear map $D_{t,s}$ is such that a vector $w \in \mathbb{R}^2$ is orthogonal to $\dot{\Xi}_{*,i}(s)$ iff $D_{t,s}w$ is orthogonal to $\dot{\Xi}_{*,i}(t)$.*

Now, for $i = 1$, we can let $s_1$ be the time corresponding to the midpoint of the $\alpha$-interval $I^+$, so $s_1 = \bar{\tau} - \frac{\delta^+}{\sqrt{R^+}}(\alpha^0 - \alpha^+)$. Then we can choose $t_1$ to be any point in $\hat{J}_1 \cap \hat{J}_3$, i.e., in $J_1 \cap E$. Similarly, we can choose $t_2 = \bar{\tau} + \frac{\delta^-}{\sqrt{R^-}}(\alpha^- - \alpha^0)$, i.e. the time corresponding to the midpoint of the $\alpha$-interval $I^-$. Then we can choose $s_2$ to be any point in $\hat{J}_2 \cap \hat{J}_3$, i.e., in $J_2 \cap E$. If we then apply (G1) successively with $s = s_1$ and $t = t_1$, with $s = t_1$ and $t = s_2$, and with $s = s_2$ and $t = t_2$, and observe that the vectors orthogonal to $\dot{\Xi}_{*,1}(s_1)$ and those orthogonal to $\dot{\Xi}_{*,3}(t_2)$ are just the vertical vectors, we find

(G2) *The linear map $D_{t_2,s_1}$ sends vertical vectors to vertical vectors.*

The segments $\sigma \mapsto \Xi_*^{+,\sigma}(\mu^+)$, $\sigma \mapsto \Xi_*^{-,\zeta(\sigma)}(\mu^-)$ are vertical and go through $\Xi_*^+(\mu^+)$ and $\Xi_*^-(\mu^-)$, respectively, when $\sigma = 1$. Since it is clear that $\Phi_{t_2,s_1}(\Xi_*^+(\mu^+)) = \Xi_*^-(\mu^-)$, (G2) holds iff $\rho(\sigma) = t_2 - s_1 + o(|\sigma - 1|)$ as $\sigma \to 1$, where $\rho(\sigma)$ is the time to go from $\Xi_*^{+,\sigma}(\mu^+)$ to $\Xi_*^{-,\zeta(\sigma)}(\mu^-)$ along the curve $\Xi_*^\sigma$. On the other hand, (#) and (##) easily imply that $\rho(\sigma)$ is equal to $2\sigma\delta^+(R^+)^{\frac{1}{2}}(\alpha^0 - \mu^+) + 2\zeta(\sigma)\delta^-(R^-)^{\frac{1}{2}}(\mu^- - \alpha^0)$. Then (G2) holds iff $\rho'(1) = 0$. But $\rho'(1)$ equals $2\delta^+(R^+)^{\frac{1}{2}}(\alpha^0 - \mu^+) + 2\zeta'(1)\delta^-(R^-)^{\frac{1}{2}}(\mu^- - \alpha^0)\,,$ and (7) implies $\zeta'(\sigma) = \frac{\alpha^0 - \mu^+}{\alpha^0 - \mu^-}$. So $\rho'(1)$ equals $2\delta^+(R^+)^{\frac{1}{2}}(\alpha^0 - \mu^+) - 2\delta^-(R^-)^{\frac{1}{2}}(\alpha^0 - \mu^+)$, and then (G2) holds iff $\delta^+(R^+)^{\frac{1}{2}} = \delta^-(R^-)^{\frac{1}{2}}$, i.e., iff

(G3) $\delta_+ = \delta_-$ *and* $R_+ = R_-$.

In other words, **the additional necessary condition for optimality is that the rolling circles that generate the upper and lower parts of $\xi_*$ should roll in the same direction (i.e., both from left to right or both from right to left) and have equal radii.**

*Remark 3.2.* The above result has been proved, of course, under the assumption that both $\xi_*^+$ and $\xi_*^-$ are cycloid arcs. There remain to consider the degenerate cases when one or both are vertical segments. If both are vertical segments, then it is easy to see that $\xi_*$ is optimal. Finally, if one of $\xi_*^+$, $\xi_*^-$ is a cycloid arc, and the other one is a vertical segment, then an argument similar to the one we used for the case of two cycloid arcs (but much simpler) shows that $\xi_*$ is not optimal, concluding the analysis of all possible cases. ♢