Manuscript submitted to
AIMS' Journals
Volume **X**, Number **0X**, XX **200X**

Website: http://AIMsciences.org
pp. **X–XX**

LYAPUNOV–SCHMIDT REDUCTION FOR OPTIMAL CONTROL PROBLEMS

Heinz Schättler
Dept. of Electrical and Systems Engineering
Washington University
St. Louis, Mo 63130, USA

Urszula Ledzewicz
Dept. of Mathematics and Statistics
Southern Illinois University at Edwardsville
Edwardsville, Il 62026, USA

This paper is dedicated to Avner Friedman on the occasion of his 80th birthday.

(Communicated by Aim Sciences)

Abstract. In this paper, we use the method of characteristics to study singularities in the flow of a parameterized family of extremals for an optimal control problem. By means of the Lyapunov–Schmidt reduction a characterization of *fold* and *cusp* points is given. Examples illustrate the local behaviors of the flow near these singular points. Singularities of fold type correspond to the typical conjugate points as they arise for the classical problem of minimum surfaces of revolution in the calculus of variations and local optimality of trajectories ceases at fold points. Simple cusp points, on the other hand, generate a cut-locus that limits the optimality of close-by trajectories globally to times prior to the conjugate points.

1. **Introduction.** In optimization problems on function spaces, there still exist significant gaps between the theories of necessary and sufficient conditions for optimality. For a finite-dimensional deterministic optimal control problem, the main necessary conditions for optimality are given by the Pontryagin maximum principle (for example, see [12, 3, 15]), a Lagrange multiplier type result. Sufficient conditions, on the other hand, center around the value function and are based on the dynamic programming principle and solutions to the Hamilton–Jacobi–Bellman (HJB) equation, the combination of a first-order PDE with a finite-dimensional minimization problem. The connection between these two vastly different approaches is made through the method of characteristics: the necessary conditions for optimality given in the Pontryagin maximum principle also define the characteristic equations for the Hamilton–Jacobi–Bellman equation (for instance, see [1]). This allows to construct smooth solutions to the Hamilton–Jacobi–Bellman equation in regions in the state space that are covered diffeomorphically by a field of extremals—controlled trajectories that satisfy the necessary conditions for optimality and are generated through an analysis of the conditions of the maximum principle. In general, however, except for few exceptional cases like the linear–quadratic regulator

2000 *Mathematics Subject Classification.* Primary: 49J15; Secondary: 93C50.
Key words and phrases. optimal control, maximum principle, Lyapunov–Schmidt reduction, fold and simple cusp singularities.

problem, such a covering will not exist globally since the value function for an optimal control problem typically exhibits singularities. The difficulties in finding solutions to optimal control problems, or, equivalently, in finding solutions to the Hamilton–Jacobi–Bellman equation, precisely lie in analyzing these sets where the value function loses differentiability or is not even continuous. The latter is related to questions about local controllability of the system, the former to singularities in the flow of extremals.

In this paper, we outline how the method of characteristics can be used to study singularities in the flow of a parameterized family of extremals. The Lyapunov–Schmidt reduction procedure is systematically employed to give easily verifiable, explicit characterizations for the two most common (least degenerate) type of these singularities, *fold* and *simple cusp points*. These two cases are commonly encountered in low-dimensional optimal control problems (related to the fact that these are the only generic singularities in dimension 2 [16]) and often form determining structures in the optimal solutions. Singularities that are of fold type correspond to the typical picture of *conjugate points* as arises in the classical problem of determining the minimum surfaces of revolution in the calculus of variations and, as in that example, local optimality of trajectories ceases at fold points. However, a different geometric picture emerges near a simple cusp point. In this case, the local flow of extremals generates a *cut-locus* between the trajectories which creates a shock in the solution of the Hamilton–Jacobi–Bellman equation. This has global implications and trajectories lose optimality already at times prior to the conjugate point. These geometric properties have already been the topic of the paper [9] where the local structure of the value function near these singularities was analyzed for systems in normal form. Here we complement those technical constructions with simple examples that illustrate the local geometric properties of the parameterized flow of extremals and the resulting implications on the optimality of controlled trajectories will be discussed. For these examples, all required computations can readily be done explicitly. The use of the Lyapunov–Schmidt reduction for the optimal control problem introduced in this paper provides some of the fundamental underlying steps missing in the earlier constructions in [9]. We especially highlight the connections between the conditions of the maximum principle and the eigenvectors for eigenvalue 0 that arise in the Lyapunov–Schmidt characterizations of corank 1 singular points.

2. **Parameterized Families of Extremals.**

We briefly review the method of characteristics in optimal control. In order to simplify the presentation, and since it brings out the main features more clearly without unnecessary technical burden, we consider a fixed horizon optimal control problem without terminal constraints. The reader is referred to our text [15] for the general formulation.

[OC]: For a fixed terminal time T and given initial conditions $(t_0, x_0) \in \mathbb{R} \times \mathbb{R}^n$, minimize the functional

$$\mathcal{J}(u) = \int_{t_0}^{T} L(s, x(s), u(s))ds + \varphi(x(T)) \tag{1}$$

over all locally bounded, Lebesgue measurable functions u that take values in a prescribed control set U, $u : [t_0, T] \to U \subset \mathbb{R}^m$, $t \mapsto u(t)$, for which the solution x of the initial value problem

$$\dot{x}(t) = f(t, x(t), u(t)), \qquad x(t_0) = x_0, \tag{2}$$

exists over the full interval $[t_0, T]$.

We assume that all the functions defining the data of the problem formulation are continuous and r-times continuously differentiable in x and u with $r \geq 1$. In principle, the control set U can be an arbitrary subset of $\mathbb{R}^m$, but the situations we study here typically arise for controls that take values in the interior of the control set and in all our examples $U = \mathbb{R}^m$. We call a locally bounded, Lebesgue measurable function u with values in U (a.e.) an admissible control and the solution x of the differential equation (2) for u is the corresponding trajectory; the pair (x, u) is called a *controlled trajectory*. Also, H denotes the control Hamiltonian defined by,

$$H = H(t.\lambda_0, \lambda, x, u) = \lambda_0 L(t, x, u) + \lambda f(t, x, u). \tag{3}$$

with $\lambda_0 \in \mathbb{R}$ and λ an n-dimensional row vector, $\lambda \in (\mathbb{R}^n)^*$.

Necessary conditions for optimality are given by the Pontryagin maximum principle [12, 3, 15]. If (x_*, u_*) is an optimal controlled trajectory defined over the interval $[t_0, T]$, then there exist a constant $\lambda_0 \geq 0$ and a co-vector $\lambda : [t_0, T] \to (\mathbb{R}^n)^*$, the so-called *adjoint variable*, such that the following conditions are satisfied:

1. *Nontriviality* of the multipliers: $(\lambda_0, \lambda(t)) \neq 0$ for all $t \in [t_0, T]$;
2. *Adjoint equation:* the adjoint variable λ is a solution to the time-varying linear differential equation

$$\begin{aligned}\dot{\lambda}(t) &= -\frac{\partial H}{\partial x}(t, \lambda_0, \lambda(t), x_*(t), u_*(t)) \\ &= -\lambda_0 L_x(t, x_*(t), u_*(t)) - \lambda(t) f_x(t, x_*(t), u_*(t))\end{aligned} \tag{4}$$

3. *Minimum condition:* everywhere in $[t_0, T]$ we have that

$$H(t, \lambda_0, \lambda(t), x_*(t), u_*(t)) = \min_{v \in U} H(t, \lambda_0, \lambda(t), x_*(t), v). \tag{5}$$

4. *Transversality condition:*

$$\lambda(T) = \lambda_0 \nabla\varphi(x_*(T)). \tag{6}$$

A parameterized family of extremals is a collection of controlled trajectories and multipliers that satisfy these conditions. More specifically, we want that the parameterizations are smooth and we assume that with p a parameter, the controls $u = u(t, p)$ are continuous and for t fixed are r-times continuously differentiable in the parameter p with the partial derivatives continuous in (t, p). We write $u \in C^{0,r}$ for this class of functions. In principle, the parameter p could be anything, but for problem [OC] there is a canonical choice taking the value of the trajectory at the endpoint, $p = x(T)$, and then integrating the dynamics and adjoint equation backward from the terminal T while maintaining the minimum condition. This leads to the following formal definition:

Definition 2.1. ***(C^r-parameterized family of extremals for [OC])*** *Given an open subset P of $\mathbb{R}^n$ and an r-times continuously differentiable function $t_- : P \to (-\infty, T)$, $t \mapsto t_-(p)$, let $D = \{(t, p) : p \in P, t_-(p) \leq t \leq T\}$. A C^r-parameterized family $\mathcal{E}$ of extremals (or extremal lifts) with domain D consists of*

1. *a family of controlled trajectories $(x, u) : D \to \mathbb{R}^n \times U$, $(t, p) \mapsto (x(t, p), u(t, p))$, such that $u \in C^{0,r}(D)$ and*

$$\dot{x}(t, p) = f(t, x(t, p), u(t, p)), \qquad x(T, p) = p. \tag{7}$$

2. *a non-negative multiplier* $\lambda_0 \in C^{r-1}(P)$ *and co-state* $\lambda : D \to (\mathbb{R}^n)^*$, $\lambda = \lambda(t,p)$, *so that* $(\lambda_0(p), \lambda(t,p)) \neq (0,0)$ *for all* $(t,p) \in D$ *and the adjoint equation,*
$$\dot{\lambda}(t,p) = -\lambda_0(p)L_x(t,x(t,p),u(t,p)) - \lambda(t,p)f_x(t,x(t,p),u(t,p)), \tag{8}$$
is satisfied on the interval $[t_-(p), T]$ *with terminal condition*
$$\lambda(T,p) = \lambda_0(p)\frac{\partial \varphi}{\partial x}(p) \tag{9}$$
3. *the controls* $u = u(t,p)$ *solve the minimization problem*
$$H(t,\lambda_0(p),\lambda(t,p),x(t,p),u(t,p)) = \min_{v\in U} H(t,\lambda_0(p),\lambda(t,p),x(t,p),v); \tag{10}$$

This definition provides the framework for our constructions. It merely formalizes that all controlled trajectories in the family $\mathcal{E}$ satisfy the conditions of the maximum principle while some smoothness properties are satisfied by the parametrization. Analogous definitions can be given for general formulations of the optimal control problem (see, for instance, [13, 15]), but will not be needed for the purpose of this paper. A similar approach was pursued by L.C. Young under the notion of "lines of flights" and "concourse of flight" [17], but our formulations avoid the imprecise notion of a "descriptive" mapping. Note that it has not been assumed that the parametrization $\mathcal{E}$ of extremals covers the state-space injectively and our objective precisely is to use this framework to analyze the geometry of the flow of the associated controlled trajectories as injectivity becomes lost (e.g., at conjugate points).

The degree r in the definition denotes the smoothness of the parametrization of the controls in the parameter p, $u \in C^{0,r}$. The control u extends as a $C^{0,r}$-function onto an open neighborhood of D and it thus follows from classical results about solutions to ODEs that the trajectories $x(t,p)$ and their time-derivatives $\dot{x}(t,p)$ are r-times continuously differentiable in p and that these derivatives are continuous jointly in (t,p) in an open neighborhood of D, i.e., $x \in C^{1,r}(D)$. For the adjoint variable there is a loss of differentiability since the Lagrangian L and the dynamics f are differentiated in x. The condition $\lambda_0 \in C^{r-1}(P)$ implies that the boundary values $\lambda(T,p)$ also lie in $C^{r-1}(P)$ and thus the multipliers lie in $\lambda \in C^{1,r-1}$. In particular, for a C^1-parameterized family of extremals only continuity in p is required. If the data defining the problem [OC] possess an additional degree of differentiability in x and if the the multiplier λ_0 is r-times continuously differentiable with respect to p, then it follows that $\lambda \in C^{1,r}$ as well. In such a case, we call $\mathcal{E}$ a *nicely* C^r-parameterized family of extremals. Also, if $\lambda_0(p) > 0$ for all $p \in P$, then all extremals are normal and by diving by $\lambda_0(p)$ we may assume that $\lambda_0(p) \equiv 1$ and we call such a family *normal.*

Definition 2.2. ***(flow of controlled trajectories)*** *Let* $\mathcal{E}$ *be a* C^r*-parameterized family of extremals. The flow associated with the controlled trajectories* (x,u) *is the mapping*
$$\digamma : D \to \mathbb{R}\times\mathbb{R}^n, \quad (t,p) \longmapsto \digamma(t,p) = \begin{pmatrix} t \\ x(t,p) \end{pmatrix},$$
i.e., is defined in terms of the graphs of the corresponding trajectories. We say the flow $\digamma$ *is a* $C^{1,r}$*-mapping on an open set* $Q \subset D$ *if the restriction of* $\digamma$ *to* Q *is continuously differentiable in* (t,p) *and* r *times differentiable in* p *with derivatives that are jointly continuous in* (t,p)*. If* $\digamma \in C^{1,r}(Q)$ *is injective and the Jacobian matrix*

$D\mathcal{F}(t,p)$ is nonsingular everywhere on Q, then we say $\mathcal{F}$ is a $C^{1,r}$-diffeomorphism onto its image $\mathcal{F}(Q)$.

Definition 2.3. ***(parameterized cost or cost-to-go function)*** *Given a C^r-parameterized family $\mathcal{E}$ of extremals with domain D, the corresponding parameterized cost or cost-to-go function is defined as*

$$C : D \to \mathbb{R}, \qquad (t,p) \mapsto C(t,p) = \int_t^T L(s, x(s,p), u(s,p))\, ds + \varphi(p).$$

It represents the value of the objective $J(u)$ for the control $u = u(\cdot, p)$ if the initial condition at time t is given by $x(t,p)$.

The results given below are all classical. For proofs that employ the framework presented here we refer the interested reader to our text [15] or the paper [10]. Essential in the constructions is the following relation between the multiplier λ and the cost C.

Lemma 2.1. ***(shadow price lemma)*** [10, 15] *Let $\mathcal{E}$ be a C^1-parameterized family of extremal lifts with domain D. Then for all $(t,p) \in D$*

$$\lambda_0(p) \frac{\partial C}{\partial p}(t,p) = \lambda(t,p) \frac{\partial x}{\partial p}(t,p) \tag{11}$$

For normal extremals, the shadow price lemma implies that the cost-to-go function is a classical solution to the Hamilton–Jacobi–Bellman equation on a region G of the state space that is covered injectively by the corresponding flow of trajectories. If $\mathcal{F}$ is an injective $C^{1,r}$-map from some open set $Q \subset D$ onto a region $G \subset \mathbb{R} \times \mathbb{R}^n$, then $\mathcal{F}$ is a $C^{1,r}$-diffeomorphism if and only if the Jacobian matrix $\frac{\partial x}{\partial p}$ is non-singular on D. In this case, for t fixed, the map $\mathcal{F}(t,\cdot) : p \mapsto x(t,p)$ is a C^r-diffeomorphism and its inverse $\mathcal{F}^{-1}(t,\cdot) : x \mapsto \mathcal{F}^{-1}(t,x)$ also is r-times continuously differentiable in x. Then the value $V^{\mathcal{E}} = V$ corresponding to the parameterized family can be defined in the state space as

$$V^{\mathcal{E}} = V : G \to \mathbb{R}, \quad V = C \circ \mathcal{F}^{-1},$$

and is continuously differentiable. The shadow price lemma allows to identify the multiplier $\lambda(t,p)$ with the gradient $\frac{\partial V^{\mathcal{E}}}{\partial x}(t, x(t,p))$ and the minimum condition (10) then becomes the Hamilton–Jacobi–Bellman equation.

Theorem 2.1. [10] *Let $\mathcal{E}$ be a C^r-parameterized family of normal extremals for problem $[OC]$ and suppose the restriction of the flow $\mathcal{F}$ to some open set $Q \subset D$ is a $C^{1,r}$-diffeomorphism onto an open subset $G \subset \mathbb{R} \times \mathbb{R}^n$ of the (t,x)-space. Then the function*

$$V^{\mathcal{E}} = V : G \to \mathbb{R}, \quad V = C \circ \mathcal{F}^{-1},$$

is continuously differentiable in (t,x) and r-times continuously differentiable in x for fixed t. The function

$$u_* : G \to \mathbb{R}, \quad u_* = u \circ \mathcal{F}^{-1},$$

is an admissible feedback control that is continuous and r-times continuously differentiable in x for fixed t. Together, the pair (V, u_*) is a classical solution of the Hamilton–Jacobi–Bellman equation*,*

$$\frac{\partial V}{\partial t}(t,x) + \min_{u \in U} \left\{ \frac{\partial V}{\partial x}(t,x) f(t,x,u) + L(t,x,u) \right\} \equiv 0, \tag{12}$$

on G. Furthermore, the following identities hold in the parameter space on Q:

$$\frac{\partial V}{\partial t}(t,x(t,p)) = -H(t,\lambda(t,p),x(t,p),u(t,p)) \tag{13}$$

$$\frac{\partial V}{\partial x}(t,x(t,p)) = \lambda(t,p). \tag{14}$$

Definition 2.4. ***(local field of extremals)*** *A C^r-parameterized local field of extremals $\mathcal{F}$ for problem [OC] is a C^r-parameterized family of normal extremals such that the associated flow $\digamma : D \to \mathbb{R}\times\mathbb{R}^n$, $(t,p)\mapsto \digamma(t,p)$, is a $C^{1,r}$-diffeomorphism.*

Since we do not impose terminal constraints at the final time, the flow $\digamma$ extends as a $C^{1,r}$-diffeomorphism onto an open neighborhood of the domain D. We just remark that this definition and the entire construction can easily be modified to allow for terminal constraints given by a smooth embedded submanifold, even if the terminal time T is defined implicitly as considered in [4].

Corollary 2.1. [10] *Let $\mathcal{F}$ be a C^r-parameterized local field of extremals for problem [OC] and suppose the associated flow $\digamma$ covers a domain G. Then, given any initial condition $(t_0,x_0)\in G$, $x_0 = x(t_0,p_0)$, the open-loop control $\bar{u}(t) = u(t,p_0)$, $t_0\le t\le T$, is optimal when compared with any other admissible controlled trajectory (x,u) with the same initial condition for which the graph of x lies in G.*

Hence the local optimality of a controlled trajectory $(x(\cdot,p_0),u(\cdot,p_0))$ is connected with the local injectivity of the flow $\digamma$ along this trajectory and thus closely related to the regularity of the differential of the flow, $D\digamma(\cdot,p_0)$. The following observation follows from a standard compactness argument.

Lemma 2.2. *If $D\digamma(t,p_0)$, or, equivalently, $\frac{\partial x}{\partial p}(t,p_0)$, is nonsingular over an interval $[\tau,T]$, then there exists a neighborhood P of p_0 so that the restriction of the flow $\digamma$ to $[\tau,T]\times P$ is a $C^{1,r}$-diffeomorphism.*

Proof. It follows from the inverse function theorem that for every time $s\in[\tau,T]$ there exists a neighborhood D_s of (s,p_0) such that the restriction of $\digamma$ to D_s is a $C^{1,r}$-diffeomorphism. Without loss of generality we may take D_s in the form $D_s = I_s\times P_s$ where I_s is an open interval and P_s an open neighborhood of p_0. The sets $\{D_s : s\in[\tau,T]\}$ form an open cover of the compact line segment $[\tau,T]\times\{p_0\}$ and thus there exists a finite subcover $\{D_{s_i} : s_i\in[\tau,T],\ i=1,\dots,r\}$. Let $P = \bigcap_{i=1}^r P_{s_i}$. Then the map $\digamma$ is a $C^{1,r}$-diffeomorphism on $D = \{(t,p) :\ p\in P, \tau\le t\le T\}$. For, if $\digamma(s_1,p_1) = \digamma(s_2,p_2)$, then, since the flow map $\digamma$ is defined in terms of the graphs of the trajectories, we have $s_1 = s_2$ and this time lies in one of the interval I_{s_i}. But $\digamma \restriction I_{s_i}\times P$ is a $C^{1,r}$-diffeomorphism and since both p_1 and p_2 lie in P, we have $p_1 = p_2$ as well. Thus $\digamma$ is injective on D. Furthermore, $\digamma$ has a differentiable inverse by the inverse function theorem. □

In this case, the controlled reference trajectory $(\bar{x},\bar{u}) = (x(\cdot,p_0),u(\cdot,p_0))$ thus gives a strong local minimum for the problem $[OC]$ in the following classical form: there exists an $\epsilon>0$ such that for any other admissible controlled trajectory (x,u) with the same initial condition, $x(\tau) = \bar{x}(\tau)$, that satisfies $\|x(t)-\bar{x}(t)\|<\epsilon$ for all $t\in[\tau,T]$, we have that $\mathcal{J}(\bar{u})\le\mathcal{J}(u)$. The regularity of the flow $\digamma$ along the reference controlled trajectory $(\bar{x},\bar{u})$ thus is the optimal control version of the *strengthened Jacobi condition* and, as in the calculus of variations, it can equivalently be formulated in terms of the Jacobi equation or solutions to equivalent Riccati differential equations (see, for example, [4, Chapter 6] or [15]).

3. **Singular Points and the Lyapunov–Schmidt Reduction.** Loss of local optimality often is in some way, directly or indirectly, related to points where the differential $D\mathcal{F}(t,p_0)$ becomes singular.

Definition 3.1. ***(singular points)*** *Let $\mathcal{E}$ be a C^1-parameterized family of normal extremals. A point $(t_0,p_0)\in D$, $t_0<T$, is a corank $\ell>0$ singular point if the matrix*

$$D\mathcal{F}(t_0,p_0)=\begin{pmatrix} 1 & 0\\ \frac{\partial x}{\partial t}(t_0,p_0) & \frac{\partial x}{\partial p}(t_0,p_0)\end{pmatrix},$$

or equivalently, $\frac{\partial x}{\partial p}(t_0,p_0)$, has rank $n-\ell$. We call a singular point (t_0,p_0) t-regular if for $\Delta(t,p)=\det\left(D\mathcal{F}(t,p)\right)$ we have that $\frac{\partial\Delta}{\partial t}(t_0,p_0)\neq 0$.

For a t-regular singular point, by the implicit function theorem, there exists an open neighborhood $W=(t_0-\varepsilon,t_0+\varepsilon)\times B_\delta(p_0)$ of (t_0,p_0), with the property that the equation $\Delta(t,p)=0$ has a unique solution on W given in the form $t=\sigma(p)$ with a continuously differentiable function $\sigma:B_\delta(p_0)\to(t_0-\varepsilon,t_0+\varepsilon)$, $p\mapsto\sigma(p)$. This is the situation we are interested in for an optimal control problem. Since the rank of a matrix is a lower semicontinuous function, if (t_0,p_0) is a t-regular corank 1 singular point, then for δ sufficiently small all points in the singular set will be of corank 1. He henceforth **assume** that this is the case, i.e.,

- **(A):** The singular set S is a codimension 1 embedded submanifold of D that entirely consists of corank 1 singular points and can be described as the graph of a continuously differentiable function $\sigma:P\to D$, $p\mapsto\sigma(p)$,

$$S=\{(t,p)\in\operatorname{int}(D):\ t=\sigma(p)\}=\mathfrak{gr}(\sigma).$$

Lemma 3.1. *Under assumption (A), there exist a nonzero C^1 right-eigenvector field $v:P\to\mathbb{R}^n$, $v=v(p)$, and a nonzero C^1 left-eigenvector field $w:P\to(\mathbb{R}^n)^*$, $w=w(p)$, for the eigenvalue 0:*

$$\left(\frac{\partial x}{\partial p}(\sigma(p),p)\right)v(p)=0\qquad\text{and}\qquad w(p)\left(\frac{\partial x}{\partial p}(\sigma(p),p)\right)=0.$$

This immediately follows from the implicit function theorem and v and w are simply smooth selections of left- and right-eigenvectors for the eigenvalue 0. For the optimal control problem there exists a remarkable relation between these left- and right-eigenvectors that follows from the useful fact below, itself a corollary of the shadow price lemma, Lemma 2.1.

Proposition 3.1. *For any C^2-parameterized family of normal extremals, the matrix*

$$\Xi(t,p)=\frac{\partial\lambda}{\partial p}(t,p)\frac{\partial x}{\partial p}(t,p)\tag{15}$$

is symmetric.

Proof. By the shadow price lemma, the partial derivative $\frac{\partial C}{\partial p_j}(t,p)$ of the parameterized cost with respect to the parameter p_j is given by

$$\frac{\partial C}{\partial p_j}(t,p)=\sum_{k=1}^{n}\lambda_k(t,p)\frac{\partial x_k}{\partial p_j}(t,p)=\lambda(t,p)\frac{\partial x}{\partial p_j}(t,p),\qquad j=1,\dots,n,$$

where $\frac{\partial x}{\partial p_j}(t,p)$ is the column vector of the partial derivatives of the components of x with respect to p_j. Differentiating this equation with respect to p_i gives

$$\frac{\partial^2 C}{\partial p_i \partial p_j}(t,p) = \sum_{k=1}^{n} \left(\frac{\partial \lambda_k}{\partial p_i}(t,p) \frac{\partial x_k}{\partial p_j}(t,p) + \lambda_k(t,p) \frac{\partial^2 x_k}{\partial p_i \partial p_j}(t,p) \right).$$

For a C^2-parameterized family of extremals, the mixed second partial derivatives of the functions C and x_k are equal and thus we have that

$$\sum_{k=1}^{n} \left(\frac{\partial \lambda_k}{\partial p_i}(t,p) \frac{\partial x_k}{\partial p_j}(t,p) \right) = \sum_{k=1}^{n} \left(\frac{\partial \lambda_k}{\partial p_j}(t,p) \frac{\partial x_k}{\partial p_i}(t,p) \right).$$

But these terms, respectively, are the (i,j) and (j,i) entries of the matrix Ξ. □

Corollary 3.1. *Let*

$$w(p) = v^T(p) \frac{\partial \lambda}{\partial p}(\sigma(p), p). \tag{16}$$

If nonzero, then w is a left-eigenvector field for the eigenvalue 0 of $\frac{\partial x}{\partial p}(\sigma(p),p)$ on P.

Proof. We have that

$$\begin{aligned} w(p)\frac{\partial x}{\partial p}(\sigma(p),p) &= v^T(p)\frac{\partial \lambda}{\partial p}(\sigma(p),p)\frac{\partial x}{\partial p}(\sigma(p),p) = v^T(p)\Xi(t,p) \\ &= (\Xi(t,p)v(p))^T = \left(\frac{\partial \lambda}{\partial p}(\sigma(p),p)\frac{\partial x}{\partial p}(\sigma(p),p)v(p)\right)^T = 0. \end{aligned}$$

□

For the optimal control problem [OC] it can be shown that w is always nonzero for a C^1-parameterized family of extremals if the strengthened Legendre condition is satisfied (i.e., if $\frac{\partial H}{\partial u} \equiv 0$ and $\frac{\partial^2 H}{\partial u^2}$ is positive definite along the reference controlled extremal) [14, 15].

In terms of the full differential of the flow map $\digamma$, $D\digamma(t,p)$, the right-eigenvector vector field is given by

$$V: \ P \to \mathbb{R}^{n+1}, \ p \mapsto V(p) = \begin{pmatrix} 0 \\ v(p) \end{pmatrix}$$

and corank 1 singularities are broadly classified as fold or cusp points depending on whether this vector field V is transversal to the tangent space $T_{(t_0,p_0)}S$ of the singular set S at (t_0,p_0) or not [7].

Definition 3.2. ***(fold and cusp points)*** *A corank 1 singular point is called a fold point if*

$$T_{(t_0,p_0)}S \ \oplus \ \text{lin span } \{V(p_0)\} = \mathbb{R}^{n+1};$$

it is called a cusp point if

$$V(p_0) \in T_{(t_0,p_0)}S.$$

In our setup, we have the following simple criterion:

Lemma 3.2. *The point (t_0,p_0), $t_0 = \sigma(p_0)$, is a fold point if and only if the Lie derivative of σ along the vector field v does not vanish at p_0, i.e.,*

$$L_v\sigma(p_0) = \nabla\sigma(p_0)v(p_0) \neq 0.$$

Proof. Since S is the graph of the function σ, $S = \{(t,p) : t-\sigma(p) = 0\}$, the tangent space to S at (t_0, p_0) consists of all vectors that are orthogonal to $(1, -\nabla\sigma(p_0))$. □

Singular points can be characterized in terms of the left- and right-eigenvectors v and w of the matrix $\frac{\partial x}{\partial p}(t,p)$. This gives rise to a more convenient description of the singular set S as the zero set of a scalar function known as the *Lyapunov–Schmidt reduction* [8]. Let

$$\zeta : D \to \mathbb{R}, \qquad (t,p) \mapsto \zeta(t,p) = w(p)\frac{\partial x}{\partial p}(t,p)v(p). \tag{17}$$

Note that with

$$w(p) = v(t)^T \frac{\partial \lambda}{\partial p}(t,p),$$

this function is of the form

$$\zeta(t,p) = w(p)\frac{\partial x}{\partial p}(t,p)v(p) = v(p)^T \Xi(t,p)v(p) = \langle v(p), \Xi(t,p)v(p)\rangle$$

so that ζ is a symmetric quadratic form. Denote the zero set of ζ in D by Z. Clearly, Z contains the singular set S. If the gradient of ζ, $\nabla\zeta$, does not vanish at a singular point (t_0, p_0), $t_0 = \sigma(p_0)$, then Z is an embedded n-dimensional manifold near (t_0, p_0) that contains S, under our assumptions itself an embedded n-dimensional manifold. Hence in a sufficiently small neighborhood, S and Z are equal. The gradient of ζ at a point $(\sigma(p), p) \in S$ is easily calculated: dropping the arguments, we have that

$$\zeta = w\frac{\partial x}{\partial p}v = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i \frac{\partial x_i}{\partial p_j} v_j$$

and thus the partial derivative with respect to t is simply given by the quadratic form

$$\frac{\partial \zeta}{\partial t} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i \frac{\partial^2 x_i}{\partial t \partial p_j} v_j = w\frac{\partial^2 x}{\partial t \partial p}v.$$

The partial derivatives with respect to p simplify at a singular point since w and v are the left- and right-eigenvectors of $\frac{\partial x}{\partial p}(\sigma(p), p)$. Generally,

$$\begin{aligned}
\frac{\partial}{\partial p_k}\left(w\frac{\partial x}{\partial p}v\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial w_i}{\partial p_k}\frac{\partial x_i}{\partial p_j}v_j + w_i\frac{\partial^2 x_i}{\partial p_k \partial p_j}v_j + w_i\frac{\partial x_i}{\partial p_j}\frac{\partial v_j}{\partial p_k}\right) \\
&= \sum_{i=1}^{n}\frac{\partial w_i}{\partial p_k}\left(\sum_{j=1}^{n}\frac{\partial x_i}{\partial p_j}v_j\right) + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i\frac{\partial^2 x_i}{\partial p_k \partial p_j}v_j \\
&\quad + \sum_{j=1}^{n}\left(\sum_{i=1}^{n} w_i\frac{\partial x_i}{\partial p_j}\right)\frac{\partial v_j}{\partial p_k}
\end{aligned}$$

and, upon evaluating at a singular point, we obtain that

$$\frac{\partial}{\partial p_k}\left(w\frac{\partial x}{\partial p}v\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i\frac{\partial^2 x_i}{\partial p_k \partial p_j}v_j$$

since both $\frac{\partial x}{\partial p}v$ and $w\frac{\partial x}{\partial p}$ vanish. These partial derivatives can be expressed in a more convenient and compact form if we consider the directional derivative in the direction of a tangent vector z in p-space. For then we have that

$$\begin{aligned}\langle \nabla_p \zeta, z\rangle &= \sum_{k=1}^{n} \frac{\partial}{\partial p_k}\left(w\frac{\partial x}{\partial p}v\right) z_k = \sum_{k=1}^{n}\sum_{i=1}^{n} w_i \left(\sum_{j=1}^{n} \frac{\partial^2 x_i}{\partial p_k \partial p_j} v_j\right) z_k \\ &= \sum_{i=1}^{n} w_i \left(\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\partial^2 x_i}{\partial p_k \partial p_j} v_j z_k\right) = \left\langle w, \frac{\partial^2 x}{\partial p^2}(v,z)\right\rangle = w\frac{\partial^2 x}{\partial p^2}(v,z).\end{aligned}$$

Thus, this directional derivative is the inner product of the left eigenvector w with the column vector $\frac{\partial^2 x}{\partial p^2}(v,z)$ whose entries are the quadratic forms of the second derivatives of the components x_i of the state with respect to p acting on the vectors v and z. Summarizing this calculation, at a singular point $(\sigma(p),p)$, $p\in P$, the directional derivative in direction of the vector $\mathfrak{z} = (\tau, z)$ can be expressed as

$$\langle \nabla\zeta(\sigma(p),p), \mathfrak{z}\rangle = \left\langle w(p), \frac{\partial^2 x}{\partial t \partial p}(\sigma(p),p)v(p)\tau + \frac{\partial^2 x}{\partial p^2}(\sigma(p),p)\Big(v(p), z\Big)\right\rangle. \quad (18)$$

We thus have the following statement:

Theorem 3.1. *Suppose (t_0,p_0) is a t-regular corank 1 singular point for which the gradient $\nabla\zeta(\sigma(p_0),p_0)$ does not vanish. Then there exist an open neighborhood $W = (t_0-\varepsilon, t_0+\varepsilon)\times P$ of (t_0,p_0) and a C^1-function σ defined on P, $\sigma : P \to (t_0-\varepsilon, t_0+\varepsilon)$ so that the singular set S is given by $S = (t,p)\in W:\ t=\sigma(p)\}$. Given a C^1 right-eigenvector field $v : P \to \mathbb{R}^n$, $\frac{\partial x}{\partial p}(\sigma(p),p)v(p)\equiv 0$, and a C^1 left-eigenvector field $w : P \to (\mathbb{R}^n)^*$, $w(p)\frac{\partial x}{\partial p}(\sigma(p),p)\equiv 0$, the set S can be described as*

$$S = \{(t,p)\in W:\ \zeta(t,p) = w(p)\frac{\partial x}{\partial p}(t,p)v(p) = 0\} \quad (19)$$

and the tangent space to S at $(\sigma(p),p)$ is given by

$$T_{(\sigma(p),p)}S = \Bigg\{\mathfrak{z} = \begin{pmatrix}\tau \\ z\end{pmatrix} \in \mathbb{R}^{n+1}: \quad (20)$$
$$w(p)\left(\frac{\partial^2 x}{\partial t\partial p}(\sigma(p),p)v(p)\cdot\tau + \frac{\partial^2 x}{\partial p^2}(\sigma(p),p)\,(v(p),z)\right) = 0.\Bigg\}$$

Corollary 3.2. *Under the same assumptions, (t_0,p_0) is a fold singularity if and only if with $v_0 = v(p_0)$ and $w_0 = w(p_0)$ we have that*

$$w_0\frac{\partial^2 x}{\partial p^2}(t_0,p_0)\,(v_0,v_0) \neq 0. \quad (21)$$

Proof. The point (t_0,p_0) is a fold if and only if the vector $V_0 = (0,v_0)^T$ is transversal to S at (t_0,p_0). Assuming that the gradient $\nabla\zeta(t_0,p_0)$ does not vanish, this is equivalent to $w_0\frac{\partial^2 x}{\partial p^2}(t_0,p_0)\,(v_0,v_0)\neq 0$. Note that this condition by itself guarantees that $\nabla\zeta(t_0,p_0)$ does not vanish. □

A singular point (t_0,p_0) of a C^2-parameterized family $\mathcal{E}$ of normal extremals thus is a *cusp point* if

$$w_0\frac{\partial^2 x}{\partial p^2}(t_0,p_0)\,(v_0,v_0) = 0. \quad (22)$$

This extra equality constraint typically restricts the set of all cusp points to be a lower-dimensional subset of S. In the least degenerate case, i.e., when no additional equality relations are satisfied, the set C of cusp points is a codimension 1 submanifold of S, hence a codimension 2 submanifold of D.

Definition 3.3. ***(simple cusp points)*** *A cusp point (t_0, p_0), $t_0 = \sigma(p_0)$, is called simple if the vector $V_0 = (0, v_0)^T$ is transversal to the submanifold C of cusp points in $T_{(t_0,p_0)}S$, i.e.,*

$$T_{(t_0,p_0)}C \ \oplus \ \text{lin span } \{V(p_0)\} = T_{(t_0,p_0)}S.$$

By Lemma 3.2, the set of cusp points is also given by

$$C = \{(t,p) \in D : \ t = \sigma(p), \quad L_v\sigma(p) = \nabla\sigma(p)v(p) = 0\},$$

i.e., it is the set of points where the Lie derivative of the function σ in direction of the zero eigenvector v vanishes. If the gradient of this function in p does not vanish at p_0, then C is a codimension 1 embedded submanifold of S and $(t_0, p_0) \in C$, $t_0 = \sigma(p_0)$, is simple if and only if the second Lie derivative $L_v^2\sigma$ of σ in direction of v does not vanish at p_0. (In this case $V(p_0)$ is not tangent to the submanifold defined by the equation $L_v\sigma(p) = 0$, but it is a tangent vector to S at (t_0, p_0).) Note that

$$L_v^2\sigma(p_0) = \left(\frac{\partial}{\partial p}_{|p_0} \nabla\sigma(p)v(p)\right) \cdot v_0 = \frac{\partial^2\sigma}{\partial p^2}(p_0)(v_0, v_0) + \nabla\sigma(p_0)\frac{\partial v}{\partial p}(p_0)v_0$$

with $\frac{\partial^2\sigma}{\partial p^2}(p_0)(v_0, v_0)$ denoting the quadratic form defined by the Hessian matrix of σ acting on the eigenvector v_0 and $\frac{\partial v}{\partial p}(p)$ the Jacobian matrix of the vector field v. In terms of the Lyapunov–Schmidt reduction, $\zeta(t,p) = w(p)\frac{\partial x}{\partial p}(t,p)v(p)$, if the gradient $\nabla\zeta(\sigma(p_0), p_0)$ does not vanish, then the set C of cusp points is given by

$$C = \left\{(t,p) \in W : \ w(p)\frac{\partial x}{\partial p}(t,p)v(p) = 0, \ w(p)\frac{\partial^2 x}{\partial p^2}(t,p)\left(v(p), v(p)\right) = 0.\right\}$$

For example, the condition that $\nabla\zeta(\sigma(p_0), p_0) \neq 0$ can simply be guaranteed by

$$\frac{\partial\zeta}{\partial t}(t_0, p_0) = w_0\frac{\partial^2 x}{\partial t\partial p}(t_0, p_0)v_0 \neq 0.$$

Making this assumption, simple cusp points can be described as follows:

Proposition 3.2. *Suppose (t_0, p_0) is a t-regular corank 1 singular point for which $\frac{\partial\zeta}{\partial t}(t_0, p_0) = w_0\frac{\partial^2 x}{\partial t\partial p}(t_0, p_0)v_0 \neq 0$. Then (t_0, p_0) is a simple cusp if and only if*

$$w_0\frac{\partial^2 x}{\partial p^2}(t_0, p_0)\left(v_0, v_0\right) = 0$$

and

$$w_0\left(\frac{\partial^3 x}{\partial p^3}(t_0, p_0)\left(v_0, v_0, v_0\right) + 3\frac{\partial^2 x}{\partial p^2}(t_0, p_0)\left(\frac{\partial v}{\partial p}(p_0)v_0, v_0\right)\right) \neq 0 \tag{23}$$

Proof. For all $p \in P$ we have that

$$\frac{\partial x}{\partial p}(\sigma(p), p)v(p) = 0.$$

Differentiating, and letting the derivative act upon $v(p)$ gives

$$\frac{\partial^2 x}{\partial t\partial p}(\sigma(p), p)v(p) \cdot L_v\sigma(p) + \frac{\partial^2 x}{\partial p^2}(\sigma(p), p)\left(v(p), v(p)\right) + \frac{\partial x}{\partial p}(\sigma(p), p)\frac{\partial v}{\partial p}(p)v(p) = 0.$$

Multiplying on the left with the left-eigenvector $w(p)$ annihilates the last term. Furthermore, for a cusp point $L_v\sigma(p) = 0$, and thus $w_0 \frac{\partial^2 x}{\partial p^2}(t_0, p_0)(v_0, v_0) = 0$ follows. Now differentiate this relation once more, evaluate the result at the cusp point, and multiply with $v(p)$ on the right and with $w(p)$ on the left. When doing so, all terms that include $L_v\sigma(p_0)$, $w_0 \frac{\partial x}{\partial p}(t_0, p_0)$ or $w_0 \frac{\partial^2 x}{\partial p^2}(t_0, p_0)(v_0, v_0)$ vanish. This leaves us with

$$\left(w_0 \frac{\partial^2 x}{\partial t \partial p}(t_0, p_0) v_0 \right) L_v^2 \sigma(p_0) + \\ w_0 \frac{\partial^3 x}{\partial p^3}(t_0, p_0)(v_0, v_0, v_0) + 3 w_0 \frac{\partial^2 x}{\partial p^2}(t_0, p_0) \left(\frac{\partial v}{\partial p}(p_0) v_0, v_0 \right) = 0.$$

By assumption, $w_0 \frac{\partial^2 x}{\partial t \partial p}(t_0, p_0) v_0$ is nonzero and thus $L_v^2 \sigma(p_0) \neq 0$ if and only if (23) holds. □

Corollary 3.3. *The tangent space to the submanifold C of cusp points at $(\sigma(p), p)$ is given by*

$$T_{(\sigma(p),p)} C = \left\{ \mathfrak{z} = \begin{pmatrix} \tau \\ z \end{pmatrix} \in \mathbb{R}^{n+1} : \; w(p) \left(\frac{\partial^2 x}{\partial t \partial p}(\sigma(p), p) v(p) \cdot \tau \right. \right. \\ \left. \left. + \frac{\partial^3 x}{\partial p^3}(\sigma(p), p)(v(p), v(p), z) + 3 \frac{\partial^2 x}{\partial p^2}(\sigma(p), p) \left(\frac{\partial v}{\partial p}(p) v(p), z \right) \right) = 0 \right\}. \tag{24}$$

4. **Fold Singularities and Conjugate Points.** The name for the fold singularity has its origin in the geometric properties of the mapping at such a point that resemble those of a quadratic function. These mapping properties are exactly the same as in the classical problem for the minimum surfaces of revolution in the calculus of variations if only smooth extremals are considered. In the language of optimal control, this problem can be formulated as to minimize an objective of the form

$$J(u) = \int_0^T x \sqrt{1 + u^2} dt \tag{25}$$

over all controls $u : [0, T] \to \mathbb{R}$ subject to the trivial dynamics $\dot{x} = u$ and fixed boundary conditions. If we fix the initial point to be $x(0) = 1$, then it is well-known that the extremals are catenaries [2] and we can parameterize the full family in the form

$$x(t, p) = \frac{\cosh(p + t \cosh p)}{\cosh p}, \quad p \in \mathbb{R}, \quad t > 0. \tag{26}$$

Figure 1 on the left depicts this family and on the right the corresponding parameterized value $V^{\mathcal{E}}$ is shown as the graph of a multivalued function defined over (t, x)-space clearly showing the fold property along the envelope of the family.

We give another simple one-dimensional example which, like the catenaries, encompasses all the features of fold singularities. It fits the model discussed here and has the advantage that in contrast to the minimum surface problem all computations are readily done explicitly. The normal form of the fold is built into the penalty term φ in the objective.

[Fold]: For a fixed terminal time T, minimize the objective

$$J(u) = \frac{1}{2} \int_{t_0}^T u^2 dt + \frac{1}{3} x(T)^3$$

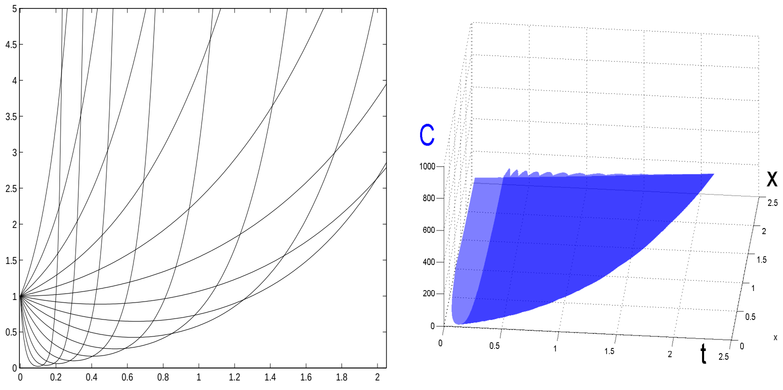

FIGURE 1. The family of catenaries (left) for the problem of finding minimum surfaces of revolution in the calculus of variations and the corresponding multivalued function $V^{\mathcal{E}} = C \circ \digamma^{-1}$ (right).

over all piecewise continuous functions $u : [t_0, T] \to \mathbb{R}$ subject to the dynamics $\dot{x} = u$.

It is straightforward to define a real-analytic parameterized family $\mathcal{E}$ of extremals for this problem: The Hamiltonian is given by $H = \frac{1}{2}\lambda_0 u^2 + \lambda u$ and the adjoint equation is $\dot{\lambda} \equiv 0$ with terminal condition $\lambda(T) = \lambda_0 x(T)^2$. The nontriviality condition on the multipliers implies that extremals are normal and we set $\lambda_0 \equiv 1$. Furthermore, multipliers and controls are constant. We choose as domain D for the parameterization $D = \{(t, p) : \ t \leq T, \ p \in \mathbb{R}\}$ with p denoting the terminal point, $p = x(T)$. All extremals can then be described in the form

$$x(t, p) = p + (T - t)p^2, \quad u(t, p) = -p^2, \quad \lambda(t, p) = p^2,$$

and the parameterized cost is given by

$$C(t, p) = \frac{1}{2}(T - t)p^4 + \frac{1}{3}p^3.$$

This defines a real-analytic parameterized family of extremals.

For this 1-dimensional problem the singular set is given by the solutions to $\frac{\partial x}{\partial p}(t, p) = 0$ and formally left- and right-eigenvector fields are given by the constants $v(t, p) \equiv 1$ and $w(t, p) \equiv 1$, but we also could take $w(t, p) = v(t, p)^T \frac{\partial \lambda}{\partial p}(t, p) = 2p \neq 0$. Thus

$$S = \{(t, p) \in D : t = \sigma(p) = T + \frac{1}{2p}, \ p < 0\}$$

and since $\frac{\partial^2 x}{\partial p^2}(t, p) = 2(T - t) > 0$ all singular points are fold points. Figure 2 shows the flow of extremals for $p < 0$.

Figure 3 shows four slices of the value function in the state space for $t = 0.6$, $t = 0.8$, $t = 1$, and $t = 1.6$. For $t = 1.6$ the fold point does not lie in the range shown and thus the corresponding value function $V^{\mathcal{E}} = C \circ \digamma^{-1}$ is single-valued. For the other three time slices, the value function is multivalued and shows the characteristic behavior near a fold singularity.

These mapping properties can be established in general near a fold point. Let $M = \digamma(S)$ be the image of the singular manifold S under the flow $\digamma$. It follows from the chain rule that the tangent space to M at the point $\digamma(\sigma(p), p) =$

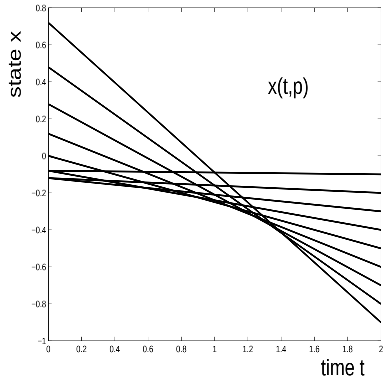

FIGURE 2. The flow of the parameterized family of extremals $\mathcal{E}$ over the interval $[0, T]$ for $T = 2$ and $p < 0$.

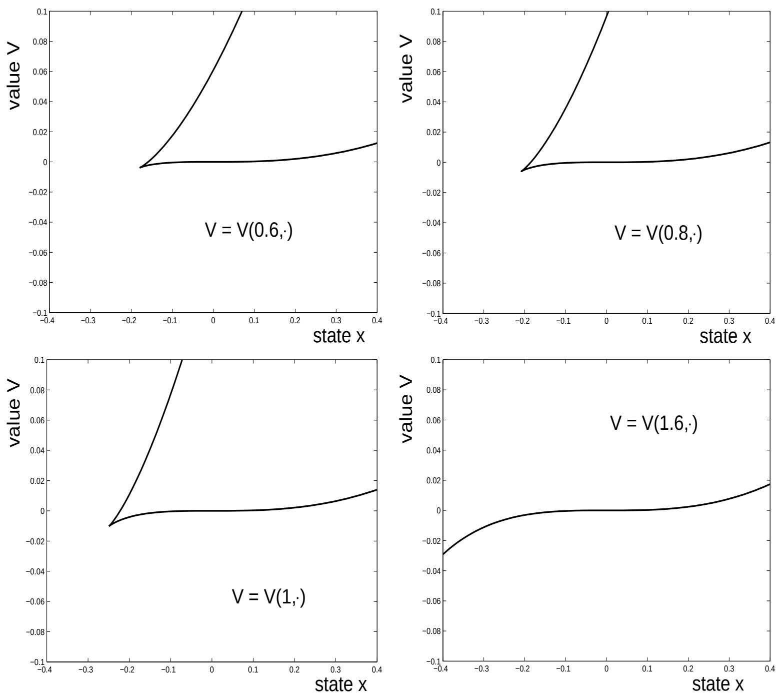

FIGURE 3. Slices of the value corresponding to the parameterized family $\mathcal{E}$ of extremals for $t = const$ for the problem [Fold].

$(\sigma(p), x(\sigma(p), p))$ is given by the image of the tangent space to S, $T_{(\sigma(p),p)}S$, under the differential $D\mathcal{F}$. At a corank 1 singular point the differential $D\mathcal{F}(t, p)$ has rank n with null-space spanned by $V(p) = (0, v(p))^T$. For a fold point this vector does not lie in the tangent space to S at $(\sigma(p), p)$ and therefore the image $T_{\mathcal{F}(\sigma(p),p)}M = D\mathcal{F}\left(T_{(\sigma(p),p)}S\right)$ is n-dimensional. In other words, the fold condition guarantees that the restriction of the mapping $\mathcal{F}$ to its singular set S is a diffeomorphism and thus M also is an n-dimensional manifold. Furthermore,

$$\left(-w(p)\frac{\partial x}{\partial t}(\sigma(p),p),\ w(p)\right)D\mathcal{F}(\sigma(p),p)=(0,0)$$

and thus $\mathfrak{n}=\left(-w(p)\frac{\partial x}{\partial t}(\sigma(p),p),\ w(p)\right)$ is a normal vector to $T_{\mathcal{F}(\sigma(p),p)}M$.

Let $\gamma_\pm:[0,\varepsilon]\to D$, $s\mapsto(\sigma(p),p\pm\sqrt{s}v(p)$, be a small line-segment in the parameter space starting at the point $(\sigma(p),p)\in S$. Since $(0,v(p))^T$ is transversal to S, by making the neighborhood P of p_0 smaller, if necessary, we can assume that $\gamma_\pm(s)\notin S$ for all $0<s\leq\varepsilon$ and $p\in P$. Let $\phi_\pm=\mathcal{F}\circ\gamma_\pm$ be the images of the curves $\gamma_\pm$ under the flow $\mathcal{F}$. Since $\frac{\partial x}{\partial p}(\sigma(p),p)v(p)=0$, by Taylor's theorem we have that

$$x(\sigma(p),p\pm\sqrt{s}v(p))=x(\sigma(p),p)+\frac{1}{2}s\frac{\partial^2x}{\partial p^2}(\sigma(p),p)\,(v(p),v(p))+o(s)$$

and thus

$$\phi_\pm(s)=\mathcal{F}\left(\gamma_\pm(s)\right)=\begin{pmatrix}\sigma(p)\\ x(\sigma(p),p)\end{pmatrix}+\frac{1}{2}s\begin{pmatrix}0\\ \frac{\partial^2x}{\partial p^2}(\sigma(p),p)\,(v(p),v(p))\end{pmatrix}+o(s).$$

Taking the inner product of the tangent vector

$$\dot{\phi}_\pm=\dot{\phi}_\pm(0)=\frac{1}{2}\begin{pmatrix}0\\ \frac{\partial^2x}{\partial p^2}(\sigma(p),p)\,(v(p),v(p))\end{pmatrix}$$

with the normal vector $\mathfrak{n}$ to $T_{\mathcal{F}(\sigma(p),p)}M$, we get that

$$\left\langle\mathfrak{n},\dot{\phi}_\pm\right\rangle=\frac{1}{2}w(p)\frac{\partial^2x}{\partial p^2}(\sigma(p),p)\,(v(p),v(p))$$

which has constant nonzero sign on P. It thus follows that both curves ϕ_+ and ϕ_- point to the same side of the tangent space to M at $\mathcal{F}(\sigma(p),p)$. Thus, if we define the curves $\gamma:(-\varepsilon,\varepsilon)\to D$, $s\mapsto(\sigma(p),p+sv(p)$, and $\phi=\mathcal{F}\circ\gamma$, then ϕ has order 1 contact with M at $\phi(0)$, i.e., ϕ is tangent to M at $\phi(0)$, but lies to one side of M. In other words, the image touches M in the point $\phi(0)\in M$, but then "folds" back. Since this holds for all controlled trajectories in a neighborhood of p_0, it follows that the flow $\mathcal{F}$ is $2:1$ near the singular set S. While the restriction $\mathcal{F}\restriction S$ is $1:1$ and maps S diffeomorphically onto M, away from M the mapping is $2:1$. There exist open neighborhoods V of (t_0,p_0) and G of $(t_0,x_0)=(t_0,x(t_0,p_0))$ such that S splits V into two connected components V_+ and V_-, $V=V_-\cup S\cup V_+$, with the property that the flow $\mathcal{F}$ restricted to V_+ or V_- is a $C^{1,2}$ diffeomorphism and the restrictions map V_+, respectively V_-, onto a region $G_+\subset G$, $\mathcal{F}(V_+)=G_+=\mathcal{F}(V_-)$. These geometric properties imply that fold points of the flow $\mathcal{F}$ are *conjugate points* in the sense of the calculus of variations and local optimality will be lost at the fold point. This can be shown in great generality, for example, by using a geometric argument based on envelopes that mimics the classical reasoning for the family of catenaries [14, 15].

5. **Simple Cusp Singularities and Cut-Loci.** The geometry of the flow near cusp singularities is more intricate and so are its implications on local optimality of the controlled trajectories. We again illustrate these features for a one-dimensional example that incorporates the normal form of a simple cusp into the penalty term φ in the objective.

[Simple Cusp]: [11] For a fixed terminal time T, minimize the objective

$$J(u) = \frac{1}{2}\int_{t_0}^{T} u^2 dt + \frac{1}{2}\left(x(T)^4 - x(T)^2\right) \tag{27}$$

over all piecewise continuous functions $u : [t_0, T] \to \mathbb{R}$ subject to the dynamics $\dot{x} = u$.

This simple regulator problem over a finite interval is related to Burgers's equation, a fundamental partial differential equation in fluid mechanics that is a prototype for equations that develop shock waves (e.g., see [5, 6]). The penalty term is a multi-modal function with a local maximum at $x = 0$ and global minima at $x = \pm\frac{1}{2}\sqrt{2}$. As we shall see, this has significant implications.

Again it is straightforward to construct a real-analytic parameterized family $\mathcal{E}$ of extremals. As in the example above, extremals are normal and the multipliers and controls are constant. A simple application of the conditions of the maximum principle shows that all extremals can be parameterized over $D = \{(t,p) : \ t \leq T, \ p \in \mathbb{R}\}$ with p denoting the terminal point, $p = x(T)$, in the form

$$x(t,p) = p + (t-T)(p-2p^3), \quad u(t,p) = p - 2p^3, \quad \lambda(t,p) = 2p^3 - p,$$

and the parameterized cost becomes

$$C(t,p) = \frac{1}{2}(T-t)(p-2p^3)^2 + \frac{1}{2}(p^4 - p^2). \tag{28}$$

All functions are polynomial and this allows explicit computations in the analysis of the flow $\digamma(t,p) = (t, x(t,p))$. Since the parameter set is one-dimensional, the singular set S is given by the solutions to the equation $\frac{\partial x}{\partial p}(t,p) = 0$ and solving for t gives

$$t = \sigma(p) = T - \frac{1}{1-6p^2} \tag{29}$$

which is finite and less than T for $|p| < \frac{1}{\sqrt{6}}$. Since

$$\frac{\partial^2 x}{\partial p^2}(t,p) = -12p(t-T).$$

the map has fold singularities at $(\sigma(p), p)$ for $p \in (-\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}})$, $p \neq 0$ and $(t,p) = (T-1, 0)$ is a simple cusp point. For, $\frac{\partial^2 x}{\partial t \partial p}(t,p) = 1 - 6p^2$ is positive on the singular set and $\frac{\partial^3 x}{\partial p^3}(T-1,0) = 12$. Since $\frac{\partial v}{\partial p}(p) \equiv 0$, this is equivalent to the transversality condition (23).

A detailed analysis of the mapping properties of the flow $\digamma$ for this example has been carried out in [11] and we briefly recall the important features. For $\alpha \geq 0$, define curves

$$\Phi_\alpha : \left(-\frac{1}{\sqrt{\alpha}}, \frac{1}{\sqrt{\alpha}}\right) \to \mathbb{R}^2, \qquad p \mapsto (t_\alpha(p), p), \quad t_\alpha(p) = T - \frac{1}{1-\alpha p^2},$$

and denote the half-curves for positive and negative values of p by Φ_α^+ and Φ_α^-, respectively. Note that $\sigma = t_6$ defines the singular set. Except for the trivial intersection in the point $(T-1, 0)$ which all curves have in common, the curves Φ_α define a foliation of the subset $\widetilde{D} = \{(t,p) \in D : t \leq T-1\}$ and the image of $\widetilde{D}$

under $\mathcal{F}$ is the subset $\widetilde{G} = (-\infty, T-1] \times \mathbb{R}$ of G. But there exist nontrivial overlaps in the map. An explicit calculation verifies that

$$x(t_\alpha(p), p) = \frac{(2-\alpha)p^3}{1-\alpha p^2} \tag{30}$$

and using $p^2 = \frac{1}{\alpha}\left(1 - \frac{1}{T-t_\alpha}\right)$, the image of the curve Φ_α is given by

$$x = \pm \frac{2-\alpha}{\alpha} \sqrt{\frac{(T-t-1)^3}{\alpha\,(T-t)}}, \qquad t \leq T-1.$$

In particular, the image of the singular set S under the flow $\mathcal{F}$ is given by the curve

$$G_c = \mathcal{F}(S) = \left\{ (t, \pm \sqrt{\frac{2}{27} \frac{(T-t-1)^3}{(T-t)}} : t \leq T-1 \right\},$$

a cusp, which gives the singularity its name. It follows from (30) that

$$\mathcal{F}(t_6(p), p) = \mathcal{F}(t_{1.5}(-2p), -2p), \quad |p| < \frac{1}{\sqrt{6}}, \tag{31}$$

and thus the curve $\Phi_{1.5}$ is also mapped onto G_C. But note that the curve

$$G_+ = \left\{ (t, \sqrt{\frac{2}{27} \frac{(T-t-1)^3}{(T-t)}} : t < T-1 \right\} = \mathcal{F}\left(\Phi_6^+\right) = \mathcal{F}\left(\Phi_{1.5}^-\right)$$

is the image of the semi-curves Φ_6^+ and $\Phi_{1.5}^-$ while

$$G_- = \left\{ (t, -\sqrt{\frac{2}{27} \frac{(T-t-1)^3}{(T-t)}} : t < T-1 \right\} = \mathcal{F}\left(\Phi_6^-\right) = \mathcal{F}\left(\Phi_{1.5}^+\right)$$

is the image of Φ_6^- and $\Phi_{1.5}^+$.

The regions and submanifolds defined below form a partition (in fact, a stratification [11]) of the domain of the parameterized family of extremals into embedded analytic submanifolds,

$$\mathcal{D} = \{D_T, D_0, D_1^-, D_1^0, D_1^+, D_-^{cp}, D_+^{cp}, D_-, D_+, D_c\}$$

(see Figure 4), where $D_T = \{T\} \times \mathbb{R}$, $D_c = \{(T-1, 0)\}$,

$$\begin{aligned} D_0 &= \{(t,p) \in D : t_{1.5}(p) < t < T\}, \\ D_1^+ &= \{(t,p) \in D : t_6(p) < t < t_{1.5}(p),\ p > 0\}, \\ D_1^- &= \{(t,p) \in D : t_6(p) < t < t_{1.5}(p),\ p < 0\}, \\ D_1^0 &= \{(t,p) \in D : t < t_6(p),\ p \in \mathbb{R}\}, \end{aligned}$$

and

$$D_+ = \left\{ (t,p) \in D : t = t_{1.5}(p),\ 0 < p < \sqrt{\frac{2}{3}} \right\},$$

$$D_- = \left\{ (t,p) \in D : t = t_{1.5}(p),\ -\sqrt{\frac{2}{3}} < p < 0 \right\},$$

$$D_-^{cp} = \left\{ (t,p) \in D : t = t_6(p),\ 0 < p < \frac{1}{\sqrt{6}} \right\},$$

$$D_+^{cp} = \left\{ (t,p) \in D : t = t_6(p),\ -\frac{1}{\sqrt{6}} < p < 0 \right\}.$$

Accordingly, partition the range of $\mathcal{F}$, $G = \mathcal{F}(D)$, into the open sets

$$G_0 = \left\{ (t,x) \in G : |x| > \sqrt{\frac{2}{27} \frac{(T-t-1)^3}{(T-t)}},\ t \le T-1 \right\} \cup (T-1,T) \times \mathbb{R},$$

$$G_1 = \left\{ (t,x) \in G : |x| < \sqrt{\frac{2}{27} \frac{(T-t-1)^3}{(T-t)}},\ t < T-1 \right\},$$

the curves G_- and G_+ defined above, the cusp point $G_c = \{(T-1,0)\}$, and the terminal manifold $G_T = \{T\} \times \mathbb{R}$, (see Figure 4),

$$\mathcal{G} = \{G_T, G_0, G_-, G_+, G_c, G_1\}.$$

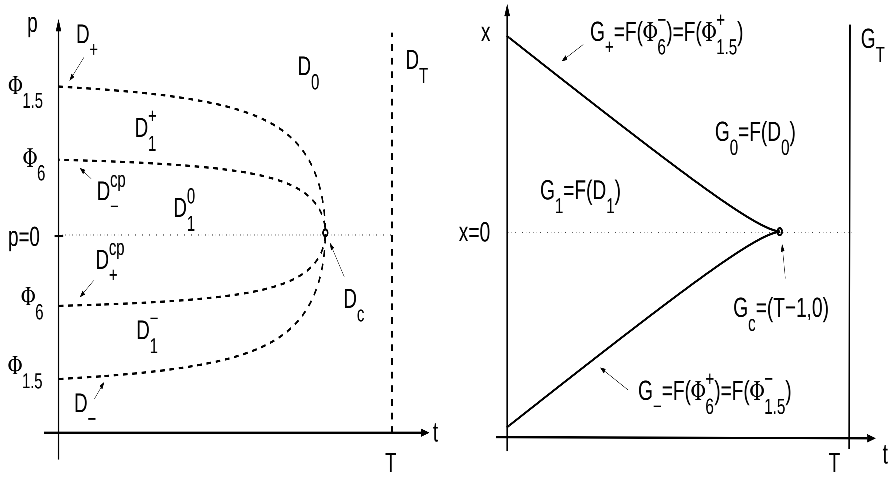

FIGURE 4. Stratifications of the domain D (left) and the range G (right) that are compatible with the map $\mathcal{F}$.

The embedded submanifolds in the collections $\mathcal{D}$ and $\mathcal{G}$ give a natural decomposition of the domain and the range of the flow of extremals into embedded submanifolds and $\mathcal{F}$ maps each submanifold in $\mathcal{D}$ diffeomorphically onto exactly one of the submanifolds in $\mathcal{G}$. We say that these decompositions are compatible with the flow map $\mathcal{F}$. Specifically,

$$\begin{aligned} &G_T = \mathcal{F}(D_T), \quad G_0 = \mathcal{F}(D_0), \quad G_c = \mathcal{F}(D_c), \\ &G_- = \mathcal{F}(D_-) = \mathcal{F}(D_-^{cp}), \quad G_+ = \mathcal{F}(D_+) = \mathcal{F}(D_+^{cp}), \\ &G_1 = \mathcal{F}(D_1^-) = \mathcal{F}(D_1^0) = \mathcal{F}(D_1^+). \end{aligned}$$

Thus the flow $\mathcal{F}$ is $3:1$ onto G_1, $2:1$ onto the branches G_- and G_+ of fold points, and $1:1$ otherwise. In particular, for each initial point in G_0 there only exists one extremal and this one is optimal. But we need to see which of the three values is optimal over G_1. The parameterized cost $C(t,p)$ for the family is easily evaluated. The value C along the curves Φ_α is given by

$$C(t_\alpha(p), p) = \frac{(4-\alpha)p^2 + (\alpha-3)}{1-\alpha p^2} p^4.$$

For $\alpha = 2$ it follows that $C(t_2(p), p) = -p^4$ and thus the values for the two trajectories corresponding to $\pm p$ are equal. Furthermore, by (30) the curves

$$D_-^{cl} = \left\{(t,p) \in D : t = t_2(p), -\frac{1}{\sqrt{2}} < \ p < 0\right\}$$

and

$$D_+^{cl} = \left\{(t,p) \in D : t = t_2(p), 0 < \ p < \frac{1}{\sqrt{2}}\right\}$$

both are mapped diffeomorphically onto the half-line

$$\widehat{G} = \{(t,0) \in G_1 : t < T-1\}.$$

Thus, given an initial condition $(t,0) \in \widehat{G}$, the two trajectories which are determined by the pre-images $(t, \pm p)$ in D_-^{cl} and D_+^{cl} have the same value for the objective. Define the value $V^{\mathcal{E}} = V$ of the parameterized family $\mathcal{E}$ of extremals as $V^{\mathcal{E}} = C \circ \mathcal{F}^-$, where $\mathcal{F}^-$ now denotes the multivalued map which assigns to a point (t,x) all possible parameter values (t,p) that satisfy $\mathcal{F}(t,p) = (t,x)$. Thus the map $\mathcal{F}^-$ is single-valued on $G_T \cup G_0 \cup G_c$, has two values on G_- and G_+ and three values on G_1. If we define π_i, $i = -, 0, +$, as the inverse parameter maps for the restrictions of $\mathcal{F}$ to D_1^i, then the associated value $V^{\mathcal{E}}$ has three sections over the set G_1 which we denote by

$$V_i : G_1 \to \mathbb{R}, \qquad V_i(t,x) = C(t, \pi_i(t,x)).$$

We have shown that

$$V_+(t,0) \equiv V_-(t,0) \qquad \text{for } (t,0) \in \widehat{G} \tag{32}$$

and V_0 will never be minimal. In fact,

$$V_0(t,x) > V_\pm(t,x) \qquad \text{for all } (t,x) \in G_1.$$

Analyzing the map $\mathcal{F}$ further, it can be seen that $\mathcal{F}$ maps each of the regions

$$D_{opt}^+ = \left\{(t,p) \in D_1^+ : t_2(p) < t < t_{1.5}(p)\right\}$$

and

$$D_{not}^- = \{(t,p) \in D_2^- : t_6(p) < t < t_2(p)\}$$

diffeomorphically onto $\widehat{G}_+ = G_1 \cap \{(t,x) \in G : x > 0\}$. Analogously, the regions

$$D_{opt}^- = \left\{(t,p) \in D_1^- : t_2(p) < t < t_{1.5}(p)\right\}$$

and

$$D_{not}^+ = \{(t,p) \in D_1^+ : t_6(p) < t < t_2(p)\}$$

are mapped diffeomorphically onto $\widehat{G}_- = G_1 \cap \{(t,x) \in G : x < 0\}$. Computing the values one obtains that

$$V_+(t,x) < V_-(t,x) \qquad \text{for all } (t,x) \in \widehat{G}_+$$

and

$$V_+(t,x) > V_-(t,x) \qquad \text{for all } (t,x) \in \widehat{G}_-$$

This gives the following result:

Theorem 5.1. *A parametrization of the globally optimal trajectories for the optimal control problem [Simple Cusp] is obtained if the domain is restricted to*

$$D_{opt} = \{(t,p) \in D : t_2(p) \leq t \leq T\}.$$

The curve Φ_2 gives a parametrization of the cut-locus Γ of the two branches corresponding to the parameterizations over D_1^+ and D_1^-, respectively, and for initial points on Γ there exist two optimal trajectories in the family. In the interior of D_{opt} the parametrization is an analytic diffeomorphism.

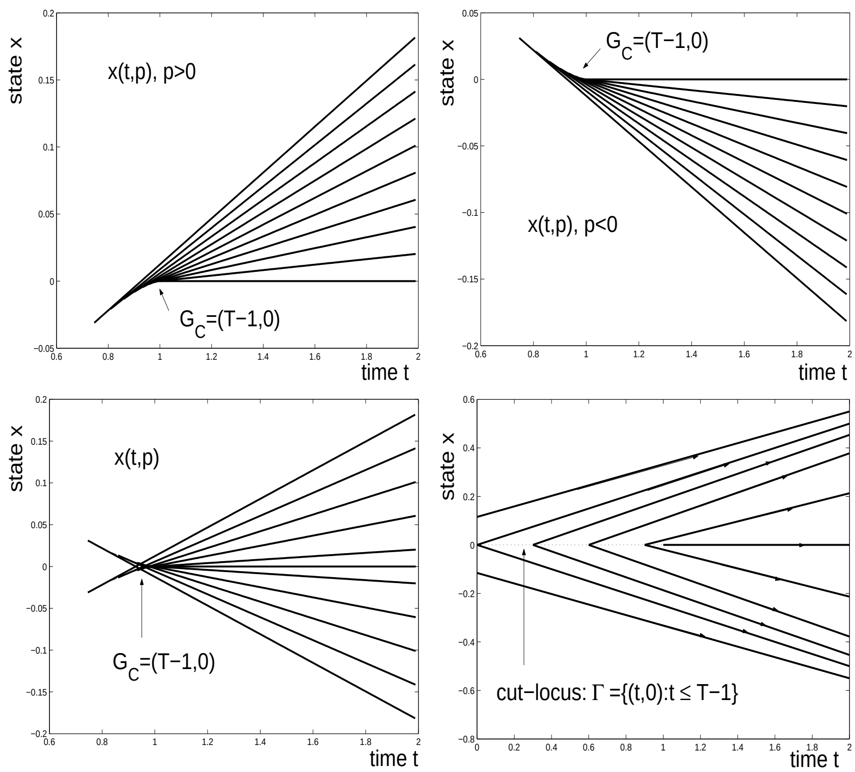

Figure 5. Flow of the parameterized family of extremals $\mathcal{E}$ for the problem [Simple Cusp]: the top row shows the flow over the intervals $[t_{cp}, T]$ for $p > 0$ (left) and $p < 0$ (right); the bottom row gives a combination of these two flows (left) and the resulting optimal synthesis of controlled trajectories (right).

Figure 5 illustrates the flow of the parameterized family of extremals: in the top row it shows the two fields of locally optimal trajectories defined for $p > 0$ and

$p < 0$. While the controlled trajectories $(x(\cdot,p), u(\cdot,p))$ for $p \neq 0$ are strong local minima over the interval $[t_0, T]$ if $t_0 > \sigma(p) = t_6(p)$, they are globally only optimal if $t_0 \geq t_2(p) > t_6(p)$. The simple cusp point at $(T-1, 0)$ generates a *cut-locus* (intersection) between the two branches of the parameterized family for $p > 0$ and $p < 0$ that limits the global optimality of trajectories at time $t_2(p)$ and thus the controlled trajectories for $p \neq 0$ already lose global optimality prior to the conjugate point. The optimal synthesis is shown in the bottom right of Figure 5.

Figure 6 shows various time-slices of the associated multivalued cost function $V^{\mathcal{E}} = C \circ F^{-1}$ in the (t,x)-space that illustrate the changes in the cost as the simple cusp point is passed. Interestingly, taken together, the graph of this multivalued function is qualitatively identical with the singular set of the swallow tail singularity, the next singularity in the order of cusps. The same observation holds for the fold singularities in which case the corresponding multi-valued value took the form of the singular set of a simple cusp. These relations also are a consequence of the fundamental relation for optimality expressed in the shadow price lemma.

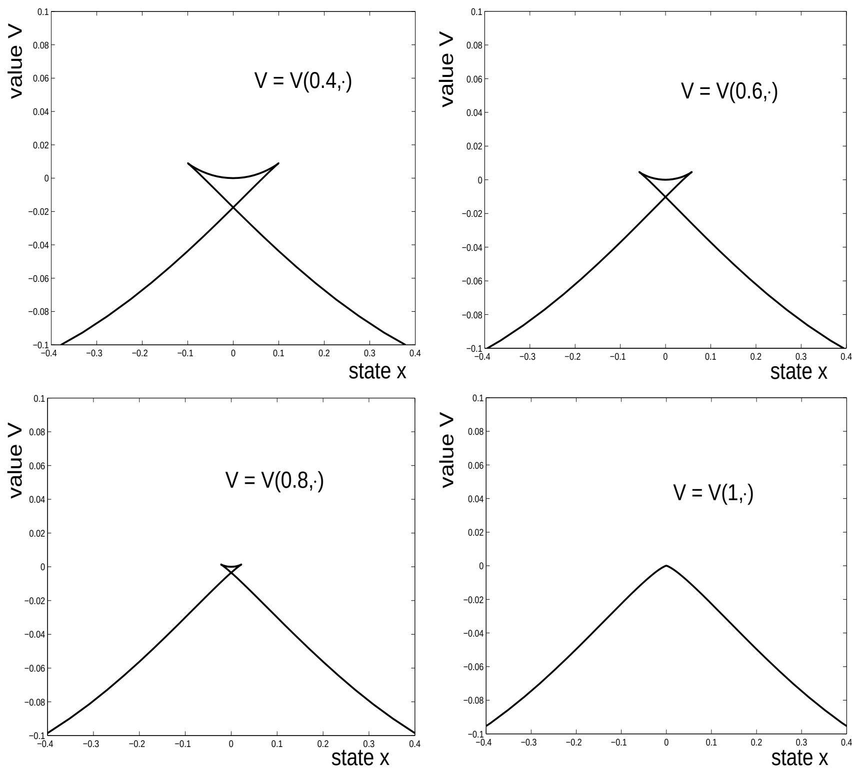

FIGURE 6. Slices of the multivalued function $V^{\mathcal{E}}$ corresponding to the parameterized family $\mathcal{E}$ of extremals for $t = const$ for the problem [Simple Cusp].

Note that we have $\sigma(0) = t_2(0) = T - 1$ for the extremal $(x(\cdot,0), u(\cdot,0))$ that passes through the simple cusp point and that this controlled trajectory remains globally optimal over the compact interval $[T-1, T]$. Thus this is an example where the trajectory only loses optimality after the conjugate point has been crossed (we are thinking of integrating the field backwards). On the contrary, for fold singularities, trajectories lose optimality at the conjugate point, i.e., are no longer optimal on the intervals $[t_{cp}(p), T]$ if $t_{cp}(p)$ denotes the corresponding conjugate time.

We close this section with the remark that these geometric properties are valid in general near simple cusp points. This can be shown by transforming the system into normal form and then analyzing the resulting system. The computations, however, are quite technical and we refer the reader to [9].

6. **Conclusion.** We have illustrated the geometric structure of the flow $\mathcal{F}$ of a parameterized family of extremals near fold and simple cusp singularities and its implications on the optimality of the trajectories in the family. Fold points generate the typical behavior of trajectories that are strong local minima and lose optimality at the fold point. However, the corresponding behavior of optimal trajectories is rarely, if ever seen in a regular synthesis of optimal trajectories. This is explained by the behavior of extremals near a simple cusp point. The simple cusp point generates a region in the state space which is covered $3:1$. In this region there exist two locally minimizing and one locally maximizing branch. The changes from the locally minimizing to the maximizing branch occur at the fold-loci and there trajectories lose strong local optimality. However, the two minimizing branches intersect and generate a cut-locus that limits the optimality of the close-by trajectories and indeed eliminates these trajectories from optimality near the cusp point prior to the conjugate point. In the language of PDE, *the simple cusp* point *generates a shock in the solution to the Hamilton–Jacobi–Bellman equation* and thus has significant implications on the structure of globally optimal controlled trajectories.

Acknowledgements. This material is based upon research supported by the National Science Foundation (NSF) under collaborative research grants DMS 1008209/ 1008221. U. Ledzewicz's research also was partially supported by an SIUE "STEP" grant in Summer 2011. Material from this paper will also be used in the forthcoming text "Geometric Optimal Control—Theory, Methods and Examples" by the authors to be published by Springer-Verlag.

REFERENCES

[1] L. Berkovitz, "Optimal Control Theory", Springer-Verlag, New York, 1974.
[2] G. Bliss, "Calculus of Variations", The Mathematical Association of America, 1925.
[3] A. Bressan and B. Piccoli, "Introduction to the Mathematical Theory of Control", American Institute of Mathematical Sciences (AIMS), 2007.
[4] A.E. Bryson, Jr. and Y.C. Ho, "Applied Optimal Control", Revised Printing, Hemisphere Publishing Company, New York, 1975.
[5] C.I. Byrnes and H. Frankowska, *Unicité des solutions optimales et absence de chocs pour les équations d'Hamilton–Jacobi–Bellman et de Riccati,* C.R. Acad. Sci. Paris, Série I, **315**, (1992), 427–431.
[6] C.I. Byrnes and A. Jhemi, *Shock waves for Riccati partial differential equations arising in nonlinear optimal control,* in "Systems, Models and Feedback: Theory and Applications", (eds. A. Isidori and T.J. Tarn), Birkhäuser, (1992), 211–225.
[7] M. Golubitsky and V. Guillemin, "Stable Mappings and their Singularities", Springer-Verlag, New York, 1973.
[8] M. Golubitsky and D.G. Schaefer, "Singularities and Groups in Bifurcation Theory, Vol. I", Applied Mathematical Sciences, vol. 51, Springer-Verlag, New York, 1985.

[9] M. Kiefer and H. Schättler, *Parametrized families of extremals and singularities in solutions to the Hamilton–Jacobi–Bellman equation,* SIAM J. on Control and Optimization, **37**, (1999), 1346–1371.
[10] J. Noble and H. Schättler, *Sufficient conditions for relative minima of broken extremals,* J. of Mathematical Analysis and Applications, **269**, (2002), 98–128.
[11] U. Ledzewicz, A. Nowakowski and H. Schättler, *Stratifiable families of extremals and sufficient conditions for optimality in optimal control problems,* J. of Optimization Theory and Applications (JOTA), **122**, (2004), 105–130.
[12] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mishchenko, "The Mathematical Theory of Optimal Processes", MacMillan, New York, 1964.
[13] H. Schättler and U. Ledzewicz, *Synthesis of optimal controlled trajectories with chattering arcs,* Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications & Algorithms, **19**, (2012) 161–186.
[14] H. Schättler and U. Ledzewicz, *Perturbation Feedback Control—A Geometric Interpretation,* Numerical Algebra, Control and Applications, (2012), to appear.
[15] H. Schättler and U. Ledzewicz, "Geometric Optimal Control—Theory, Methods and Examples", Springer-Verlag, New York, 2012.
[16] H. Whitney, *Elementary structure of real algebraic varieties,* Ann. Math., **66**, (1957), 545-556.
[17] L.C. Young, "Lectures on the Calculus of Variations and Optimal Control Theory", W.B. Saunders, Philadelphia, 1969.

E-mail address: `uledzew@siue.edu; hms@wustl.edu`