Projection-Based Switched System Optimization: Absolute Continuity of the Line Search

T. M. Caldwell and T. D. Murphey

Abstract—The line search is considered for the problem of numerical switched system optimization using projection-based techniques. Switched system optimization may be formulated as an infinite dimensional optimal control problem where the switching control design variables are constrained to the integers. Projection-based techniques handle the integer constraint by considering an equivalent problem with unconstrained design variables but where the cost is dependent on the projection of the design variables to the constrained set of feasible switched system trajectories. This paper is concerned with the line search step of the projection-based optimization procedure. The main result provides sufficient conditions on the descent direction so that the update rule is absolutely continuous with respect to the step size.

I. Introduction

This paper considers the line search of projection-based switched system optimization. Switched systems evolve according to multiple distinct modes where only one mode is active at any time. The control is the scheduling of the mode sequence and the timing of the mode transitions. An equivalent control representation is realized by a set of functions of time, labeled the switching control, that dictate which mode is active at any time. However, the values of the switching control must be constrained to the integers, which makes efficient numerical optimization difficult. This paper furthers our work in [4] which handles the integer constraint using a projection operator.

Other switched system optimization methods include: switching time optimization [3], [5], [7], [14] which fixes the mode sequence and optimizes only over the switching times; mode injection methods [5], [6] which compute the timing for when an injected mode will result in a decrease to the cost; embedding methods [2], [11], [13] which relax the integer constraint on the switching control design variables and optimizes the relaxed cost.

In comparison, for projection-based methods, the design variables live in an unconstrained space but the cost is computed on the projection of the design variables to the set of feasible switched system trajectories. Suppose J is the cost, μ is the unconstrained switching control and \mathcal{P} projects μ onto the set of feasible switching controls where

This material is based upon work supported by the National Science Foundation under award IIS-1018167 as well as the Department of Energy Office of Science Graduate Fellowship Program (DOE SCGF), made possible in part by the American Recovery and Reinvestment Act of 2009, administered by ORISE-ORAU under contract no. DE-AC05-06OR23100

T. M. Caldwell and T. D. Murphey are with the Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road Evanston, IL 60208, USA E-mail: caldwelt@u.northwestern.edu; t-murphey@northwestern.edu

the integer constraint is in effect. The goal is to solve the problem

$$\arg\min_{\mu} J(\mathcal{P}(\mu)).$$

Numerical optimization algorithms begin with an estimate of the optimizer, μ , choose a search direction, v, take a step of size γ in that direction and iterates. At each iteration, two quantities need to be calculated: v and γ . This paper assumes v is fixed and is concerned with calculating the step size. The process of calculating γ is known as the line search and calls for finding the γ that minimizes $J(\mathcal{P}(\mu + \gamma v))$. However, calculating the minimum exactly is often less computationally desirable than finding a γ that satisfies Armijo and weak Wolfe conditions [1], [8], [10]. For nonsmooth optimization, [8] presents a line search algorithm and proves that it terminates in a finite number of steps to a step size satisfying Armijo and weak Wolfe conditions if J is locally Lipschitz and weakly-semismooth.

However, for projection-based switched system optimization, the derivative of the cost with respect to γ can go unbounded and therefore, J is not locally Lipschitz. We prove, though, that for the max-projection presented in this paper and for the reasonable assumptions on the search direction v, $J(\mathcal{P}(u+\gamma v))$ is absolutely continuous in γ . Absolute continuity is the main result of this paper since the line search is viable because of work by Lewis and Overton [9]. They consider a similar line search to the one presented in [8] and show that if J is absolutely continuous, the line search converges to a γ containing points satisfying the Armijo and weak Wolfe conditions.

This paper is organized as follows: Section II presents the numerical optimization algorithm for projection-based switched system optimization. Section III introduces switched systems and states the optimization problem. In Section IV, the max-projection is introduced and the update rule for numerical optimization is considered. Section V shows that under certain assumptions on the variations of the switching control, the line search is not differentiable every where and that it can go unbounded. However, the derivative is still Lebesgue integrable and we find the line search is absolutely continuous. Finally, Section VI discusses the viability of implementing an inexact line search.

II. ITERATIVE PROJECTION-BASED OPTIMIZATION

This paper is concerned with the line search step of iterative projection-based switched system optimization. Iterative optimization methods compute a new estimate of the optimum by taking a step in a search direction from the current estimate of the optimum so a decrease in cost is achieved. Even if the chosen direction descends the cost, the size of the step taken must ensure that a sufficient decrease is achieved for convergence. This section presents an iterative procedure in the context of projection-based switched system optimization:

Algorithm 1: Suppose η is the constrained design variable, $J(\eta)$ is the cost and \mathcal{P} maps the unconstrained space to the constrained space and is a projection. Then, the iterative projection-based optimization algorithm is as follows:

- 1) Set initial optimal estimate η^0 and set k=1.
- 2) Choose a search direction ζ^k
- 3) Solve for step size γ : $\arg\min_{\gamma\in\mathbb{R}^+}J(\mathcal{P}(\eta^{k-1}+\gamma\zeta^k))$
- 4) Update: $\eta^k = \mathcal{P}(\eta^{k-1} + \gamma \zeta^k)$.
- 5) If η^k satisfies a terminal condition, then exit, else, increment k and repeat from step 2.

Remarks:

- In step 3, $\eta^{k-1} + \gamma \zeta^k$ is calculated in the unconstrained space and \mathcal{P} projects the unconstrained result onto the constrained space.
- For smooth optimization problems, it is computationally desirable to approximate the minimizing step size with an inexact line search (see [1], [8]). The primary purpose of this paper is to validate the line search for projection-based switched system optimization. As we will see in Section V, the line search for the problem of switched systems is non-smooth. However, under certain conditions on the descent direction and with the projection proposed in Section IV, the line search is in fact absolutely continuous and thus a line search is viable [9].

In the following section we present switched systems and state the switched system optimization problem.

III. SWITCHED SYSTEM OPTIMIZATION

A switched system's evolution is dictated by multiple modes but where the instantaneous evolution is dictated by only one of those mode. The control is the scheduling of the modes. There are many useful switched system representations. One representation specifies the mode sequence and the timings for when mode switches occur. Another representation assigns a function of time to each mode where at any time only one function has value 1 and all others have value 0. A value of 1 implies that function's corresponding mode is currently active. We call the set of these functions the switching control.

A. Switched System

Let $\mathcal X$ and $\mathcal U$ be spaces of Lebesgue integrable functions from the time interval [0,T] to, respectively, \mathbb{R}^n and \mathbb{R}^N . Consider a switched system with n states $x = [x_1, \ldots, x_n]^T \in \mathcal{X}, \ N \text{ switching controls } u = [u_1, \ldots, u_N]^T \in \mathcal{U}, \text{ and } N \text{ modes } f_i(x), \ i \in \{1, \ldots, N\}$ which are $\mathcal{C}^r, r > 0$, on \mathcal{X} . The state equations are given by

$$\dot{x}(t) = F(x(t), u(t)) := \sum_{i=1}^{N} u_i(t) f_i(x(t)), \quad x(0) = x_0.$$

We say the pair (x, u) satisfies the state equations if

$$G(x, u, t) := x(t) - x(0) - \int_0^t F(x(\tau), u(\tau)) d\tau = 0$$
 (2)

for almost all time $t \in [0, T]$. The integral is understood to be the Lebesgue integral.

Additionally, for the evolution of the state equation to be consistent with that of a switched system—i.e. for only one mode to be active at a time—the switching controls must belong to the following set of admissible switching controls:

Definition 3.1: The curve $u = [u_1, \dots, u_N]^T$ composed of N piecewise constant functions of time is an admissible switching control if

- for almost each $t \in [0,T]$ and each $i \in \{1,\ldots,N\}$, $u_i(t) \in \{0, 1\},\$
- for almost each $t \in [0,T]$, $\sum_{i=1}^N u_i(t)=1$, and for each $i \in \{1,\dots,N\}$: u_i does not chatter—i.e. in the time interval [0, T], the number of switches between values 0 and 1 is finite.

Denote the set of all admissible switching controls as Ω . According to the first two properties, u(t) is equal to one of the standard basis vectors of \mathbb{R}^N . In other words, define $E^N = \{e_1, \dots, e_N\}$ where e_i has value 1 at its i^{th} entry and 0 for every other entry. Then, $u(t) = e_i$ for some $i \in \{1, ..., N\}$. As such, the state, given by Eq.(1), evolves according to only one mode for almost all time. The third property disallows Zeno behavior, in which an infinite number of mode switches occur in finite time.

If the state, x, and switching control, u, satisfy the state equations, Eq.(2), and u is admissible, then (x, u) constitutes an admissible switched system. Formally, define the set of such (x, u) as:

Definition 3.2: The pair $(x, u) \in \mathcal{X} \times \mathcal{U}$ constitutes a valid switched system if

- 1) $u \in \Omega$ and
- 2) G(x, u, t) = 0 for almost all $t \in [0, T]$.

Denote the set of all such pairs of state and switching controls by S.

B. Switching Schedule

We relate the switching control given by Definition 3.1 with the equivalent representation of switching schedules. A switching schedule is the mode sequence as well as the times the switches occur. We define two mappings Σ and Ton Ω which return the mode sequence and switching times respectively. The switching times are the discontinuity points of $u \in \Omega$

$$\mathcal{T}(u) := \{ t \in [0, T] | u(t^+) \neq u(t^-) \}$$

As for the mode sequence, consider the switching control $u \in \Omega$. Let $\{T_1, \dots, T_{M-1}\} = \mathcal{T}(u)$ and suppose that $u(t) = e_{\sigma_i}$ for $t \in (T_{i-1}, T_i)$ where $\sigma_i \in \{1, \dots, N\}$ and $e_{\sigma_i} \in E^N$. The i^{th} mode in the mode sequence corresponding to u is σ_i and set $\Sigma(u) = {\sigma_1, \ldots, \sigma_M}$. The switching schedule corresponding to u is $(\Sigma(u), \mathcal{T}(u))$. When we are only concerned with the schedule on a connected interval $\mathcal{I} \subset [0,T]$, we denote the switching times and mode sequence for that interval by $\mathcal{T}^{\mathcal{I}}(u)$ and $\Sigma^{\mathcal{I}}(u)$.

The state equations given by the switching schedule representation is

$$\dot{x}(t) = f_{\sigma_i}(x(t)), \quad T_{i-1} < t < T_i, \quad \text{for } i \in \{1, \dots, M\}$$
where $x(0) = x_0, T_0 = 0$ and $T_M = T$.

C. Problem Statement

The objective is to find the switching control—or equivalently, the mode sequence and switching times—that optimizes the performance of the system. Define the usual cost function as

$$J(x,u) = \int_0^T \ell(x(\tau),u(\tau))d\tau$$

where the running cost, $\ell: \mathcal{X} \times \mathcal{U} \to \mathbb{R}$ is continuously differentiable with respect to both \mathcal{X} and \mathcal{U} . The problem of interest is to minimize J with respect to x and u under the constraint that x and u constitute an admissible switched system—i.e. $(x, u) \in \mathcal{S}$.

Problem 1 (Constrained Problem): Solve

$$\arg\min_{(x,u)\in\mathcal{S}}J(x,u).$$

Note the constraints $(x, u) \in \mathcal{S}$ include the integer constraint on u that for each $i \in \{1, \ldots, N\}$ and $t \in [0, T]$, $u_i(t) \in \{0, 1\}$.

This paper handles the integer constraint by furthering the projection-based method introduced in [4]. In [4], we consider an equivalent problem to the constrained problem where the design variables live in the unconstrained space $(\mathcal{X}, \mathcal{U})$ and the cost is evaluated on the projection of the design variables to the set of admissible switched system trajectories:

Problem 2 (Unconstrained Problem): Suppose $\mathcal{P}: \mathcal{X} \times \mathcal{U} \to \mathcal{S}$ is a projection—i.e. $\mathcal{P}(\mathcal{P}(\alpha, \mu)) = \mathcal{P}(\alpha, \mu)$. Solve

$$\arg\min_{(\alpha,\mu)\in\mathcal{X}\times\mathcal{U}}J(\mathcal{P}(\alpha,\mu)).$$

Observe that what we refer to here as the unconstrained problem differs from the procedure of finding the solution to a problem where the cost is minimized for unconstrained variables and the result is projected onto the constrained set of switched system trajectories. Instead, the cost for the unconstrained problem is calculated on the projected unconstrained design variables. Therefore, both the unconstrained problem and constrained problem are equivalent in that if (x^*, u^*) is a solution to the constrained problem then there is an $(\alpha^*, \mu^*) \in \mathcal{P}^{-1}(x^*, u^*)$ which is a solution to the unconstrained problem and if (α^*, μ^*) is a solution to the unconstrained problem, then $(x^*, u^*) = \mathcal{P}(\alpha^*, \mu^*)$ is a solution to the constrained problem.

The next section introduces the max-projection operator for solving the unconstrained problem.

IV. PROJECTION OPERATOR

The projection maps curves from an unconstrained space $\mathcal{X} \times \mathcal{U}$ to the set of switched systems, \mathcal{S} . The projection enriches the set of local variations. To elaborate, suppose η^{k-1} lives in a constrained space and the iterative optimization procedure computes the new estimate of the optimum, η^k , by adding a descent direction, ζ^k , scaled by a step size γ to the previous estimate, η^{k-1} —i.e. $\eta^k = \eta^{k-1} + \gamma \zeta^k$. For the case of switched systems, consider $\eta^k = (x^k, u^k) \in \mathcal{S}$ and $\zeta^k = (z^k, v^k)$. Since $u^{k-1} \in \Omega$ is constrained to specific integers, the only feasible variation of u^{k-1} is trivially $v^k = 0(t)$ since no other curve adds with u^{k-1} —under the usual sense of addition—for general $\gamma \in \mathbb{R}^+$ to a feasible curve in Ω . However, by computing $u^{k-1} + \gamma v^k$ in an unconstrained space and projecting the result to the constrained set Ω , the admissible variations are only limited by the choice of projection. In this section, we propose the max-projection.

A. Max-Projection

Let the set $\overline{\mathcal{R}}$ to be the admissible subset of $\mathcal{X} \times \mathcal{U}$ which the max-projection maps to \mathcal{S} —i.e. maps to solutions satisfying Definition 3.2. In order to define the max-projection, we first define the following *reproducing* mapping—i.e. a mapping $\mathcal{Q}: \overline{\mathcal{R}} \to \Omega$ where for all $(x, u) \in \mathcal{S}$, $u = \mathcal{Q}(x, u)$:

Definition 4.1: Take $(\alpha, \mu) \in \overline{\mathcal{R}}$. The i^{th} : $i \in \{1, \dots, N\}$ element of the max-reproducing mapping, $\mathcal{Q} : \overline{\mathcal{R}} \to \Omega$, at time $t \in [0, T]$ is

$$Q_i(\alpha(t), \mu(t)) := \begin{cases} 1 & \mu_i(t) = \max\{\mu_1(t), \dots, \mu_N(t)\} \\ 0 & \text{else.} \end{cases}$$
(4)

Now, define the max-projection as:

Definition 4.2: Take $(\alpha, \mu) \in \mathcal{R}$. The max-projection, \mathcal{P} : $\overline{\mathcal{R}} \to \mathcal{S}$, at time $t \in [0, T]$ is

$$\mathcal{P}(\alpha(t), \mu(t)) := \begin{cases} \dot{x}(t) = F(x(t), u(t)), & x(0) = x_0 \\ u(t) = \mathcal{Q}(\alpha(t), \mu(t)). \end{cases}$$
(5)

Remarks:

- 1) Notice the max-reproducing mapping and the max-projection do not depend on \mathcal{X} . Therefore, for the remainder of the paper, we write $\mathcal{P}(\mu)$ and $\mathcal{Q}(\mu)$. Likewise, we denote the admissible domain of $\mathcal{P}(\mu)$ as $\mathcal{R} \subset \mathcal{U}$ with the understanding that $\overline{\mathcal{R}} = (\mathcal{X}, \mathcal{R})$.
- 2) Other projections depend on \mathcal{X} (see the feedback projection in [4]).
- 3) The max-projection is so named because it maps the element of $\mu(t)$ with greatest value to 1 and all other elements to 0. Furthermore, since the maximal element of $\mu(t)$ maps to the maximal value of 1, it is clear that $\mathcal{P}(\mathcal{P}(\mu)) = \mathcal{P}(\mu)$ —i.e. \mathcal{P} is a projection.

As in [4], the reproducing condition, Eq. (4), can be written using the step function. Define $1:\mathbb{R}\to\{0,1\}$ to be the step function where for $a\in\mathbb{R}$,

$$1(a) := \left\{ \begin{array}{ll} 0 & a < 0 \\ 1 & a \ge 0. \end{array} \right.$$

The max-reproducing mapping is given by the product of step functions as follows:

$$Q_i(\mu(t)) = \prod_{j \neq i}^{N} 1(\mu_i(t) - \mu_j(t)).$$
 (6)

For brevity, we write $\mu_{ij} := \mu_i - \mu_j$. For t to be a switching time, there must be $i \neq j \in \{1, \dots, N\}$ such that μ_{ij} crosses 0 at t. Since μ_{ij} may cross 0 at a point of discontinuity, we define a zero crossing and the number of zero crosses as follows:

Definition 4.3: A function crosses 0 at time $t \in [0,T]$ if

- $y(t^-) < 0$ and $y(t^+) > 0$ or
- $y(t^-) > 0$ and $y(t^+) < 0$.

Further, the number of times y crosses 0 in the interval [0,T] is $\mathcal{N}(y)$.

B. Update Rule

In the optimization procedure Algorithm 1, a new estimate of the optimum is obtained by varying from the current estimate and projecting the result to the set of feasible switched system trajectories. Suppose $u \in \Omega$, $\gamma \in \mathbb{R}^+$ is the step size and $v \in \mathcal{V}$ is a variation of u. The update rule is given as $\mathcal{P}(u+\gamma v)$. The set \mathcal{V} is the admissible set of variations—i.e. if $v \in \mathcal{V}$, then $u+\gamma v \in \mathcal{R}$ for all $\gamma \in \mathbb{R}^+$. A sufficient condition on the curve v to be admissible is given by the following assumption:

Assumption 1: Assume the curve $v = [v_1, \dots, v_N]^T$ is the N piecewise \mathcal{C}^0 functions on [0,T] where [0,T] may be partitioned into the disjoint sets, \mathcal{I} and $\mathcal{J} \subset [0,T]$ where $\mathcal{I} \cup \mathcal{J} = [0,T]$ and

- for each $i \neq j \in \{1, ..., N\}$, $v_i v_j$ has a finite number of critical points¹ in \mathcal{I} or
- for each $t \in \mathcal{J}$, $v_1(t) = v_2(t) = \cdots = v_N(t)$.

Recall the calculation of the max-projection as given by the max-reproducing mapping in Eq.(6) and let $u_{ij}=u_i-u_j$ and $v_{ij}=v_i-v_j$. Then, $\mathcal{Q}_i(u(t)+\gamma v(t))=\prod_{j\neq i}^N 1(u_{ij}(t)+\gamma v_{ij}(t))$. By supposing v satisfies Assumption 1, $u_{ij}(t)+\gamma v_{ij}(t)$ may be partitioned into disjoint time intervals where either the number of critical points of v_{ij} is finite—and therefore the number of critical points of $u_{ij}+\gamma v_{ij}$ is also finite—or $v_{ij}=0$. Since critical points separate strictly monotonic intervals, the number of times $u_{ij}+\gamma v_{ij}$ crosses zero is finite and thus the number of times $1(u_{ij}+\gamma v_{ij})$ switches between values 1 and 0 is as well finite. As such, $\mathcal{Q}(u+\gamma v)$ does not chatter. We state this conclusion formally in the following lemma:

Lemma 1: Suppose $u \in \Omega$, $\gamma \in \mathbb{R}^+$ and v satisfies Assumption 1. Then, $u + \gamma v \in \mathcal{R}$.

Proof: Set $\mu = u + \gamma v$ and use notation $\mu_{ij} = \mu_i - \mu_j$, $u_{ij} = u_i - u_j$ and $v_{ij} = v_i - v_j$. In order for $\mu \in \mathcal{R}$, it must be the case that $\mathcal{Q}(\mu) \in \Omega$. Consider each property of Definition 3.1. First, according to Eq.(4), $\mathcal{Q}_i(\mu(t)) \in \{0,1\}$ for each $t \in [0,T]$. Second, property 2 is satisfied as long as

for almost all times $t \in [0,T]$, there exists a single greatest element of $\mu(t)$, which is clear since the greatest element of $\mu(t)$ maps to value 1 and all other elements map to value 0. Finally, the rest of the proof shows that the number of times $\mathcal{Q}(\mu)$ changes values is finite.

We first consider all intervals $\mathcal{J} \subset [0,T]$ where $v_1 = v_2 = \cdots = v_N$. For these intervals, the greatest element at any time $t \in \mathcal{J}$ of $\mu(t) = (u + \gamma v)(t)$ is given by u(t) and thus $\mathcal{Q}(\mu(t)) = \mathcal{Q}(u(t)) = u(t)$. Since $u \in \Omega$, it must be the case that u switches values a finite number of times in \mathcal{J} .

Second, we consider all other intervals $\mathcal{I} = \mathcal{J}^c$ where $\forall i \neq j \in \{1, \dots, N\}, v_{ij} \text{ has a finite number of critical}$ points. Further partition \mathcal{I} into pairwise disjoint intervals $\{\mathcal{I}_k\}$ where $\cup \mathcal{I}_k = \mathcal{I}$. Let $\{s_{k,1},\ldots,s_{k,\ell_k}\} \in \mathcal{I}_k$ be the collection of all of the critical points for each $i \neq j \in$ $\{1,\ldots,N\}$ in the k^{th} interval. Further set $s_{k,0}=\inf \mathcal{I}_k$ and $s_{k,\ell_k+1} = \sup \mathcal{I}_k$ and note ℓ_k is finite. Clearly each v_{ij} is strictly monotonic over each interval $(s_a, s_{a+1}), a \in$ $\{0,\ldots,\ell_k\}$ and thus, v_{ij} can cross any value at most once in each (s_a, s_{a+1}) . This implies that for a constant $u, u_{ij} + \gamma v_{ij}$ can cross zero a single time in the interval (s_a, s_{a+1}) . However, u_{ij} can have values 0, 1, or -1 and thus in each interval $(s_a, s_{a+1}), u_{ij} + \gamma v_{ij}$ can at most cross zero three times in addition to the finite number of times u_{ij} changes values. Thus for each i and j, μ_{ij} crosses zero a finite number of times for each of the $\ell_k + 1$ intervals of each of the finite number of intervals \mathcal{I}_k . Therefore, μ_{ij} crosses zero a finite number of times. Finally, by inspection of Eq.(6), $Q(\mu)$ can switch values only at a time when there is an $i \neq j \in \{1, \dots, N\}$ in which μ_{ij} crosses zero. Thus, $\mathcal{Q}(\mu)$ switches values a finite number of times and property 3 is satisfied, proving $u + \gamma v \in \mathcal{R}$.

We will find in the next section that for v satisfying Assumption 1 with an additional assumption that bounds the flatness of v that $J(\mathcal{P}(u+\gamma v))$ is absolutely continuous with respect to γ . This result is important for the line search—step 3 of Algorithm 1—since absolute continuity is assumed for the non-smooth line search in [9].

V. ABSOLUTE CONTINUITY OF LINE SEARCH

Assume $u \in \Omega$ and $v \in \mathcal{V}$ are fixed and only γ varies. Define

$$J(\gamma) := J(\mathcal{P}(u + \gamma v)).$$

We find in this section that the derivative of the cost with respect to γ —i.e. $DJ(\gamma)$ —exists almost everywhere and we provide sufficient conditions for $DJ(\gamma)$ to be Lebesgue integrable. Consequently, $J(\gamma)$ is then absolutely continuous.

A. Dependence of the Mode Sequence on γ

When varying γ the mode sequence of $\mathcal{Q}(u + \gamma v)$ only changes when the local mode order changes at some time $t \in [0, T]$. Define constant local mode order as follows:

Definition 5.1: Suppose $u \in \Omega$, $v \in \mathcal{V}$, $\gamma \in \mathbb{R}^+$ and $t \in (0,T)$. If there is $\delta t > 0$ such that $\mathcal{I}_{\delta t} = (t - \delta t, t + \delta t) \subset$

¹A critical point is a point t_0 of a real valued function y in which either $\dot{y}(t_0) = 0$ or y is not differentiable at t_0 .

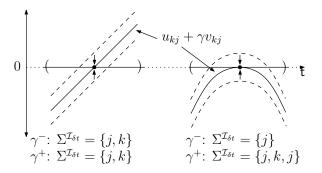


Fig. 1. An example where the mode order local to γ and t is constant (left) and an example where it changes (right).

[0,T] and the left and right limits of the local mode sequence to γ are equal—i.e.

$$\lim_{\gamma' \to \gamma^-} \Sigma^{\mathcal{I}_{\delta t}}(\mathcal{Q}(u + \gamma' v)) = \lim_{\gamma'' \to \gamma^+} \Sigma^{\mathcal{I}_{\delta t}}(\mathcal{Q}(u + \gamma'' v))$$

then the mode order local to γ and t of $\mathcal{Q}(u(t)+\gamma v(t))$ is constant.

The γ where the mode sequence changes are non-differentiable points of $J(\gamma)$. Figure 1 shows an example where the local mode order is constant and an example where the local mode order changes. The first example is not at a critical point of $u_{kj} + \gamma v_{kj}$ while the second example is. In fact, as shown in the following lemma, the local mode order is constant for all $\gamma \in \mathbb{R}^+$ if $t \in (0,T)$ is a non-critical time.

Lemma 2: Suppose $u \in \Omega$, $v \in \mathcal{V}$, $\gamma \in \mathbb{R}^+$ and $t \in (0,T)$. If for each $i \neq j \in \{1,\ldots,N\}$, t is not a critical point of $\mu_{ij} = u_{ij} + \gamma v_{ij}$, then the mode order local to γ and t of $\mathcal{Q}(u(t) + \gamma v(t))$ is constant.

Proof: Since each $\mu_{ij}(t)$ is not critical, $\dot{\mu}_{ij}(t)$ exists and is non-zero. As such, $\mu_{ij}(t)$ is not a maximum or minimum. Separately consider the case where a single element of $\mu(t)$ has greatest value and the case where multiple elements of $\mu(t)$ have equal greatest value. First, suppose there is a $k \in \{1,\ldots,N\}$ where $\mu_{ki}(t)>0$ for each $i\neq k$. By the continuity of each $\mu_{ki}(t)$, there is a $\delta t>0$ such that for each $t'\in(t-\delta t,t+\delta t)=:\mathcal{I}_{\delta t}$ and each $i\neq k$, $\mu_{ki}(t')>0$. Furthermore, by the continuity of each $\mu_{ki}=u_{ki}+\gamma v_{ki}$ with respect to γ , there is an open interval around γ such that for each γ' in this open interval, $u_{ki}(t')+\gamma' v(t')>0$ and thus

$$\lim_{\gamma' \to \gamma^{-}} \Sigma^{\mathcal{I}_{\delta t}}(\mathcal{Q}(u+\gamma'v)) = \lim_{\gamma'' \to \gamma^{+}} \Sigma^{\mathcal{I}_{\delta t}}(\mathcal{Q}(u+\gamma''v)) = \{k\}.$$

Now, consider the case where multiple elements of $\mu(t)$ have equal greatest value. By the assumption that each μ_{ij} are not critical at t, a single element of $\mu(t)$ can have greatest value just prior to time t and another just after time t. Suppose these two elements have index k and j and thus $\mu_{kj}(t)=0$, $\mu_{ki}(t)>0$ and $\mu_{ji}(t)>0$ for each $i\in\{1,\ldots,N\}$ not equal to k or j. This case is depicted in Figure 1. Since $\mu_{kj}(t)$ is not critical it is strictly monotonic. Assume the monotonicity is increasing and if not swap k and j. Thus, there is a $\delta t>0$ such that $\mu_{kj}(t')>0$ for each $t'\in(t,t+\delta t)$ and $\mu_{kj}(t')<0$

for each $t' \in (t - \delta t, t)$. Setting $\mathcal{I}_{\delta t} = (t - \delta t, t + \delta t)$, the local mode sequence is $\Sigma^{\mathcal{I}_{\delta t}}(\mathcal{Q}(u + \gamma v)) = \{j, k\}$. By the continuity of μ_{kj} with respect to γ , the left and right limits of $\mu_{kj}(t')$ do not change the strict inequalities for t' < t and t' > t and thus,

$$\lim_{\gamma' \to \gamma^{-}} \Sigma^{\mathcal{I}_{\delta t}}(\mathcal{Q}(u+\gamma'v)) = \lim_{\gamma'' \to \gamma^{+}} \Sigma^{\mathcal{I}_{\delta t}}(\mathcal{Q}(u+\gamma'v)) = \{j, k\}$$

completing the proof.

When v satisfies Assumption 1, the number of critical points of each v_{ij} is finite. Therefore, the total number of γ for which the mode sequence of $\mathcal{Q}(u+\gamma v)$ changes is finite. Let $\Gamma(u,v)$ be the increasingly ordered set of $\gamma \in \mathbb{R}^+$ for which the mode sequence of $\mathcal{Q}(u+\gamma v)$ changes:

$$\Gamma_{u,v} := \{ \gamma \in \mathbb{R}^+ | \forall \delta \gamma > 0, \exists \gamma' \in (\gamma - \delta \gamma, \gamma + \delta \gamma) \cap \mathbb{R}^+,$$
 where $\Sigma(\mathcal{Q}(u + \gamma v)) \neq \Sigma(\mathcal{Q}(u + \gamma' v)) \}.$

For example, if $\{\gamma_k\} \in \Gamma(u, v)$ and ordered, then $\Sigma(\mathcal{Q}(u + \gamma v))$ is constant for each $\gamma \in (\gamma_k, \gamma_{k+1})$.

Lemma 3: Suppose $u \in \Omega$ and v satisfies Assumption 1. Then, the dimension of $\Gamma(u,v)$ is finite.

Proof: The mode sequence of $\mathcal{Q}(u+\gamma v)$ changes when there is a time $t\in(0,T)$ for which the local mode orde changes. According to Lemma 2, the only times t where there can be a non-constant local mode order is if there is an $i\neq j\in\{1,\ldots,N\}$ such that μ_{ij} is a critical point.

Take \mathcal{J} to be the subset of [0,T] for which $v_{ij}=0$ for each $i \neq j \in \{1, ..., N\}$. Here, $\mu = u$ and clearly the local mode order is trivially constant for all $\gamma \in \mathbb{R}^+$. Now, consider all other intervals, $\mathcal{I} = \mathcal{J}^c$, of [0,T] for which each μ_{ij} has a finite number of critical points. Let $\{s_{ij1},\ldots,s_{ij\ell_{ij}}\}$ be the collection of critical points of μ_{ij} and note ℓ_{ij} is finite. The local mode order can only change at a time s_{ijk} , $k \in \{1, \dots, \ell_{ij}\}$, if μ_{ij} crosses zero at s_{ijk} . If $\mu_{ij}(s_{ijk})$ is continuous, then $\mu_{ij}(s_{ijk})$ crosses zero for only a single $\gamma \in \mathbb{R}^+$. However, if $\mu_{ij}(s_{ijk})$ is discontinuous, then $\mu_{ij}(s_{ijk})$ crosses zero for a continuous interval of γ . For γ in the interior of this interval, the local mode order remains constant, and thus the only γ where the local mode order may not be constant are the two interval bounds. Thus, each critical point may contribute one or two values of γ toward the count of non-constant mode sequence points. Since there are a finite number of critical points for each of the μ_{ij} , the number of $\gamma \in \mathbb{R}^+$ for which the mode sequence of $Q(u + \gamma v)$ changes is finite—i.e. the dimension of $\Gamma(u, v)$ is finite.

B. Derivative of Line Search

When the mode sequence is constant, only the switching times of $\mathcal{Q}(u+\gamma v)$ vary as γ varies. For this reason, it is convenient to use the switching schedule representation instead of the switching control representation. Define $\Sigma_{u,v}(\gamma) := \Sigma(\mathcal{Q}(u+\gamma v))$ and $\mathcal{T}_{u,v}(\gamma) := \mathcal{T}(\mathcal{Q}(u+\gamma v))$. The cost parameterized by the switching schedule is $J(\Sigma_{u,v}(\gamma),\mathcal{T}_{u,v}(\gamma)) := J(\mathcal{P}(u+\gamma v)) = J(\gamma)$. Consider $\{\gamma_k\} = \Gamma(u,v)$ where for $\gamma \in (\gamma_k,\gamma_{k+1})$ the mode sequence

is constant. Assuming the cost is differentiable at γ , the derivative of the cost with respect to γ is

$$DJ(\gamma) = D_2 J(\Sigma_{u,v}(\gamma), \mathcal{T}_{u,v}(\gamma)) \cdot D\mathcal{T}_{u,v}(\gamma) \tag{7}$$

where $D_2J(\Sigma_{u,v}(\gamma), \mathcal{T}_{u,v}(\gamma))$ is the switching time gradient and $D\mathcal{T}_{u,v}(\gamma)$ is the derivative of the switching times with respect to the step size. Calculations for these derivative are given in the following two Lemmas. The switching time gradient is from the literature: [3], [5], [7]:

Lemma 4: Let $\{\sigma_1, \ldots, \sigma_M\}$ be the constant mode sequence of the switched system and $\tau = \{T_1, \ldots, T_{M-1}\}$ be the variable switching times. Suppose each mode, $f_i(x(t))$, and the running cost, $\ell(x(t))$, is \mathcal{C}^1 . Then, the i^{th} switching time derivative is

$$D_{T_i}J(\tau) = \rho^T(T_i)(f_{\sigma_i}(x(T_i)) - f_{\sigma_{i+1}}(x(T_i)))$$
 (8)

where x is the solution to the state equations, Eq.(3), and ρ is the solution to the following adjoint equation

$$\dot{\rho}(t) = -Df_{\sigma_i}(x(t))^T \rho(t) - D\ell(x(t))^T,$$

$$T_{i-1} < t < T_i \quad \text{for } i \in \{1 \dots, M\}$$

$$(9)$$

where $\rho(T) = 0$, $T_0 = 0$ and $T_M = T$.

As for the second term in Eq.(7), the derivative of the switching times with respect to γ is given as:

Lemma 5: Suppose $u \in \Omega$, $v \in \mathcal{V}$, $\mu = u + \gamma v$ and $\gamma \in \mathbb{R}^+$ is such that $\Sigma(\mathcal{Q}(u + \gamma v))$ is constant. Let $\{\sigma_1, \ldots, \sigma_M\} = \Sigma_{u,v}(\gamma)$ and $\{T_1(\gamma), \ldots, T_{M-1}(\gamma)\} = \mathcal{T}_{u,v}(\gamma)$. Then the i^{th} derivative of $\mathcal{T}_{u,v}(\gamma)$, $DT_i(\gamma)$, is given for the following two cases:

1) If $T_i(\gamma)$ is not a critical point of $\mu_{\sigma_i\sigma_{i+1}}$, then

$$DT_i(\gamma) = \frac{u_{\sigma_i \sigma_{i+1}}(T_i(\gamma))}{\gamma^2 \dot{v}_{\sigma_i \sigma_{i+1}}(T_i(\gamma))},$$
 (10)

2) or if $T_i(\gamma)$ is a discontinuity point of $\mu_{\sigma_i\sigma_{i+1}}$ and $0 \in (\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^-), \mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^+))$, then

$$DT_i(\gamma) = 0.$$

Proof: For the immediate mode to switch from σ_i to $\sigma_{i+1},\ \mu_{\sigma_i\sigma_{i+1}}=u_{\sigma_i\sigma_{i+1}}+\gamma v_{\sigma_i\sigma_{i+1}}$ must cross 0 at time $T_i(\gamma)$, which is clear from Eq.(6). First, consider case 1 where $T_i(\gamma)$ is not a critical point of $\mu_{\sigma_i,\sigma_{i+1}}$. As such, $u_{\sigma_i\sigma_{i+1}}(T_i(\gamma))$ is constant, $v_{\sigma_i\sigma_{i+1}}(T_i(\gamma))$ is continuous and $\dot{v}_{\sigma_i\sigma_{i+1}}(T_i(\gamma))\neq 0$. For the continuous $\mu_{\sigma_i\sigma_{i+1}}$ to cross 0 at $T_i(\gamma)$,

$$u_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) + \gamma v_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) = 0.$$

Take the derivative with respect to γ :

$$v_{\sigma_i \sigma_{i+1}}(T_i(\gamma)) + \gamma \dot{v}_{\sigma_i \sigma_{i+1}}(T_i(\gamma)) DT_i(\gamma) = 0.$$

Solve for $DT_i(\gamma)$:

$$DT_i(\gamma) = \frac{-v_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}{\gamma \dot{v}_{\sigma_i\sigma_{i+1}}(T_i(\gamma))} = \frac{u_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}{\gamma^2 \dot{v}_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}.$$

As for the second case where $T_i(\gamma)$ is a discontinuity point of $\mu_{\sigma_i\sigma_{i+1}}$, for a zero cross to occur, the value 0 must be contained in the left and right limits of $\mu_{\sigma_i\sigma_{i+1}}$ at time $T_i(\gamma)$. The lemma calls for 0 to be in the interior of the

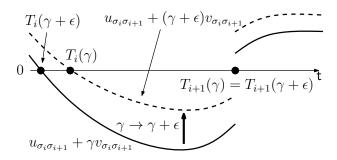


Fig. 2. Shows the transition from σ_i to σ_{i+1} at a continuous zero crossing, $T_i(\gamma)$, as well as the transition back to σ_{i+1} at a discontinuous zero crossing, $T_{i+1}(\gamma)$. When γ increases to $\gamma + \epsilon$, $u_{\sigma_i \sigma_{i+1}} + \gamma v_{\sigma_i \sigma_{i+1}}$ slides upward. Since $u_{\sigma_i \sigma_{i+1}}(T_i(\gamma)) + \gamma u_{\sigma_i \sigma_{i+1}}(T_i(\gamma))$ is decreasing, the i^{th} switching time, $T_i(\gamma)$ moves right as γ increases and so $DT_i(\gamma)$ is positive and given by case 1 of Lemma 5. However, the $i+1^{\text{th}}$ switching time occurs at a discontinuity point of $u_{\sigma_i \sigma_{i+1}} + \gamma v_{\sigma_i \sigma_{i+1}}$ and the derivative follows case 2 of Lemma 5 and is $DT_{i+1}(\gamma) = 0$.

limit values—i.e. $0 \in (\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^-), \mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^+))$. Consider the perturbation to γ resulting in the perturbation ν of $\mu_{\sigma_i\sigma_{i+1}}$. For small enough $\epsilon>0$, it is the case that $(\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^-)+\epsilon\nu(T_i(\gamma)^-), \mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^+)+\epsilon\nu(T_i(\gamma)^+))$ still contains 0 and thus the perturbed switching time remains at $T_i(\gamma)$. Therefore, $DT_i(\gamma)=0$.

Figure 2 shows two switching times, one for each of the two cases in Lemma 5. There may be other cases where the switching times are differentiable, however we will find that if v satisfies Assumption 1, there are only a finite number of $\gamma \in \mathbb{R}^+$ for which $\mathcal{Q}(u+\gamma v)$ has a switching time with derivative not given by one of the two cases.

As follows from Eq.(7), the derivative of the cost $DJ(\gamma)$ is given by the dot product of the result in Lemma 4 with the result in Lemma 5 if the mode sequence is constant and each switching time satisfies the conditions for either case 1 or case 2. In general, the derivative will not exist everywhere. For example, $DJ(\gamma)$ goes unbounded for γ where $Q(u+\gamma v)$ has a switching time $T_i(\gamma)$ near a critical point of $v_{\sigma_i\sigma_{i+1}}$ where $\dot{v}_{\sigma_i\sigma_{i+1}}(\cdot)=0$ —see Eq.(10). However, with the following additional assumption on v, the derivative of the switching time, $DT_i(\gamma)$, is still Lebesgue integrable around such γ .

Assumption 2: Take $v \in \mathcal{V}$ and consider every subset $\mathcal{I} \subset [0,T]$ where for each $i \neq j \in \{1,\ldots,N\}$, v_{ij} has a finite number of critical points in \mathcal{I} . Assume that if for any $i \neq j \in \{1,\ldots,N\}$ and $t \in \mathcal{I}$ it is the case that $\dot{v}_{ij}(t) = 0$, then there is an integer k: $k \geq 2$ where the k^{th} derivative of v_{ij} exists and is non-zero in a neighborhood of t.

Lemma 6: Suppose $u \in \Omega$, v satisfies Assumption 1 and Assumption 2 and $\gamma_1 \in \mathbb{R}^+$ is so that as $\gamma \to \gamma_1$, the i^{th} switching time $T_i(\gamma) \to T_{\gamma_1}$ where $\dot{v}_{\sigma_i \sigma_{i+1}}(T_{\gamma_1}) = 0$. Then, $DT_i(\gamma)$, Eq.(10) is Lebesgue integrable near γ_1 .

Before proving the lemma we first consider the following differential equation: Suppose $h: \mathbb{R} \to \mathbb{R}$ is continuously differentiable, d_1 and d_2 are constants and $k \geq 1$ is an

integer. The solution to

$$h'(z) = \frac{d_1}{(d_2 - h(z))^k} \tag{11}$$

is

$$h(z) = d_2 - ((k+1)(c-d_1z))^{\frac{1}{k+1}}$$

for some constant c. Now to prove Lemma 6 Proof: By the assumptions on $v_{\sigma_i\sigma_{i+1}}$ there is an integer k>2 and an $\epsilon\in\mathbb{R}$ such that $v_{\sigma_i\sigma_{i+1}}^{(k)}(T_i(\gamma))$ is non-zero for $\gamma\in(\gamma_1,\gamma_1+\epsilon]$ and thus $\dot{v}_{\sigma_i\sigma_{i+1}}$ can be expanded around T_{γ_1} as:

$$\dot{v}_{\sigma_i \sigma_{i+1}}(T_i(\gamma)) = \sum_{j=1}^k \frac{v_{\sigma_i \sigma_{i+1}}^{(j)}(T_{\gamma_1})}{(j-1)!} (T_{\gamma_1} - T_i(\gamma))^{j-1} + O((T_{\gamma_1} - T_i(\gamma))^j).$$

Without loss of generality, assume k is the least order derivative of $v_{\sigma_i\sigma_{i+1}}(T_{\gamma_1})$ that is non-zero. Then, for ϵ small,— γ near γ_1 and thus $T_i(\gamma)$ near T_{γ_1} ,

$$\dot{v}_{\sigma_i \sigma_{i+1}}(T_i(\gamma)) \approx \frac{v_{\sigma_i \sigma_{i+1}}^{(k)}(T_{\gamma_1})}{(k-1)!} (T_{\gamma_1} - T(\gamma))^{k-1}$$

Therefore, $DT_i(\gamma)$ is approximately

$$DT_{i}(\gamma) \approx \frac{u_{\sigma_{i}\sigma_{i+1}}(T_{\gamma_{1}})}{\gamma_{1}^{2} \frac{v_{\sigma_{i}\sigma_{i+1}}^{(k)}(T_{\gamma_{1}})}{(k-1)!} (T_{\gamma_{1}} - T(\gamma))^{k-1}},$$
 (12)

Set

$$d_1 = \frac{u_{\sigma_i \sigma_{i+1}}(T_{\gamma_1})(k-1)!}{\gamma_1^2 v_{\sigma_i \sigma_{i+1}}^{(k)}(T_{\gamma_1})}$$

Eq.(12) has the same form as Eq.(11) and thus

$$T_i(\gamma) \approx T_{\gamma_1} - (k(c - d_1 \gamma))^{\frac{1}{k}}$$

where c is such that $T_i(\gamma) \to T_{\gamma_1}$ as $\gamma \to \gamma_1$ —i.e. $c = d_1 \gamma_1$. Plugging $T_i(\gamma)$ into Eq.(12), $DT_i(\gamma)$ is approximately

$$DT_i(\gamma) \approx \frac{d_1}{(kd_1(\gamma_1 - \gamma))^{\frac{k-1}{k}}}$$

which is Lebesgue integrable as $\gamma_1 - \gamma$ goes to zero.

C. Absolute Continuity

In order for $J(\gamma) := J(\mathcal{P}(u+\gamma v))$ to be absolutely continuous, $DJ(\gamma)$ must exist almost everywhere and $J(\gamma)$ must be the indefinite integral of $DJ(\gamma)$ plus a constant term [12]. For the indefinite integral of $DJ(\gamma)$ to exist, $DJ(\gamma)$ must be Lebesgue integrable. First, we count the number of non-differentiable points of $J(\gamma)$ for v satisfying Assumption 1. We find the count is finite:

Lemma 7: Suppose $u \in \Omega$, v satisfies Assumption 1 and $DJ(\mathcal{T}_{u,v}(\gamma))$, Eq.(8), exists for $\gamma \in [0, \gamma_{max}]$ where $\gamma_{max} \in \mathbb{R}^+$. Then, $DJ(\gamma) = DJ(\mathcal{T}_{u,v}(\gamma)) \cdot \mathcal{T}_{u,v}(\gamma)$ —given by Eq.(7) and Lemmas 4 and 5—exists for all but a finite number of $\gamma \in [0, \gamma_{max}]$.

Proof: It is assumed that the switching time gradient, $DJ(\mathcal{T}_{u,v}(\gamma))$ exists. The only $\gamma \in [0, \gamma_{max}]$ for which $DJ(\gamma)$ does not exist are γ for which $\mathcal{T}(\mathcal{Q}(u+\gamma v))$ does not exist. These γ for which the cost is non-differentiable are such that either $\Sigma(\mathcal{Q}(u+\gamma v))$ is not constant or there

is a switching time that is not satisfied by either case 1 or 2 of Lemma 5. First, according to Lemma 3, the number of γ for which the mode sequence is not constant is finite. Second, if the derivative of the switching time $T_i(\gamma)$ is not given by case 1 or 2 of Lemma 5, then either 1) $\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma))$ is continuous and $T_i(\gamma)$ is a critical point of $\mu_{\sigma_i\sigma_{i+1}}$ or 2) $\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma))$ is discontinuous but zero is not in $(\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^-), \mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^+))$. Consider 1) first. For $T_i(\gamma)$ to be a continuity point and a switching time, it must be the case that $\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) = 0$. By the linearity of $\mu_{\sigma_i\sigma_{i+1}}$ with respect to γ , there can only be one γ for which $\mu_{\sigma_i\sigma_{i+1}}$ is zero at time $T_i(\gamma)$. Since $u \in \Omega$ and vsatisfies Assumption 1, there are a finite number of critical points of $\mu_{\sigma_i\sigma_{i+1}}$. Therefore, there are a finite number of γ with a switching time $T_i(\gamma)$ where $u_{\sigma_i\sigma_{i+1}}$ is continuous but are not included in case 1).

As for case 2), for $T_i(\gamma)$ to be a discontinuity point of $\mu_{\sigma_i\sigma_{i+1}}$ and a switching time, by definition of zero crossing, it is possible for $0 \in [\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^-), \mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^+)]$. In other words, it is possible for $\mu_{\sigma_i\sigma_{i+1}}$ to additionally cross zero at the boundaries of $(\mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^-), \mu_{\sigma_i\sigma_{i+1}}(T_i(\gamma)^+))$ and not just the interior, in which $DT_i(\gamma)$ exists and is given in case 2). There can only be a single γ for which $\mu_{\sigma_i\sigma_{i+1}}=0$ at each of the bounds of the interval. Thus, there are a finite number of γ with discontinuous $\mu_{\sigma_i\sigma_{i+1}}$ which are not included by case 2). It follows that $DJ(\gamma)$ exists except at the finite number of γ considered in the proof.

Now, using this lemma, Lemma 7, as well as Lemma 6, we give sufficient conditions for $J(\gamma)$ to be absolutely continuous:

Theorem 5.2: Suppose $u \in \Omega$, v satisfies Assumption 1 and $DJ(\mathcal{T}_{u,v}(\gamma))$, Eq.(8), exists for $\gamma \in [0, \gamma_{max}]$ where $\gamma_{max} \in \mathbb{R}^+$. Then, $DJ(\gamma)$ is absolutely continuous on the interval $[0, \gamma_{max}]$.

Proof: According to Lemma 7, there are a finite number of $\gamma \in \mathbb{R}^+$ for which $DJ(\gamma)$ does not exist. Since $DJ(\gamma)$ can go unbounded only for the γ considered in Lemma 6 and that the term of $DJ(\gamma)$ that goes unbounded is still Lebesgue integrable, $DJ(\gamma)$ is Lebesgue integrable. Therefore, we can define

 $H(\gamma) := \int_{0}^{\gamma} DJ(\gamma')d\gamma',$

which is absolutely continuous. Finally, by Theorem 37 of [12] (chapter 6), since $DH(\gamma)=DJ(\gamma)$ for almost every γ , H differs from J by a constant and therefore, $J(\gamma)$ is absolutely continuous.

VI. IMPLEMENTATION LINE SEARCH

As seen in Algorithm 1, numerical optimization algorithms iteratively choose a search direction v^k and take a step in that direction from the current estimate of the optimizer u^{k-1} . An option for choosing the size of the step taken is to calculate the γ that minimizes the cost—i.e. $\arg\min_{\gamma\in\mathbb{R}^+}J(\mathcal{P}(u+\gamma v))$. This process is called the line search. Another option is to approximate the minimizer by a step size that satisfies Armijo and weak Wolfe conditions so

that a sufficient reduction to the cost and that a reasonable step is taken [10].

Observe for the projection-based switched system optimization problem that the step size must be sufficiently large for the new cost to differ from the current cost. To demonstrate, suppose $u^{k-1} \in \Omega$ is the current estimate of the optimizer, $v^k \in \mathcal{V}$ is the direction and $\gamma \in \mathbb{R}^+$ is the step size. For $J(\mathcal{P}(u^{k-1}+\gamma v^k))$ to differ from $J(\mathcal{P}(u^{k-1}))$, there must be a time $t \in [0,T]$ where $\mathcal{Q}(u^{k-1}(t)+\gamma v^k(t)) \neq u^{k-1}(t)$. Suppose $i \in \{1,\ldots,N\}$ is the active mode of u^{k-1} at time t—i.e. $u^{k-1}(t)=e_i$. Then, there must be a $j \in \{1,\ldots,N\}: j \neq i$ where $u^{k-1}_{ij}(t)+\gamma v^k_{ij}(t)<0$ for the new active mode at t to not be mode i. The inequality may be rewritten as $\gamma v^k_{ij}(t)<-1$ since $u^{k-1}_{ij}(t)=1$. Therefore, if $|v^k_{ij}|$ were bounded by $0 < L < \infty$ for each $i \neq j \in \{1,\ldots,N\}$, then γ must be greater than 1/L for $\mathcal{Q}(u^{k-1}+\gamma v^k)$ to differ from u^{k-1} .

Label γ_0^k as the lower bound on the step size for which $\mathcal{Q}(u^{k-1}(t)+\gamma v^k(t))$ differs from u^{k-1} . In order to calculate γ_0^k , let $i(t) \in \{1,\dots,N\}$ be the curve of active modes of u^{k-1} —i.e. $u^{k-1}(t) = e_{i(t)}$. We wish to find the least value of γ for which there is a $t \in [0,T]$ and $j \neq i(t)$ such that $u^{k-1}_{i(t)j}(t) + \gamma v^k_{i(t)j}(t) = 0$ —i.e. $\gamma = -1/v^k_{i(t)j}(t)$. Let $j(t) = \arg\min_{j \neq i(t)} v^k_{ij}(t)$. If there is a time t where $v^k_{i(t)j(t)}(t) < 0$, then

$$\gamma_0^k = -\frac{1}{\min_{t \in [0,T]} v_{i(t)j(t)}^k(t)}.$$

If $v^k_{i(t)j(t)}(t)$ is never negative, then there is not a γ for which $\mathcal{Q}(u^{k-1}+\gamma v^k)$ differs from u^{k-1} . Set

$$h(\kappa) = J(\mathcal{P}(u^{k-1} + (\gamma_0^k + \kappa)v^k)) - J(\mathcal{P}(u^{k-1}))$$

and note $J(\mathcal{P}(u^{k-1})) = J(\mathcal{P}(u^{k-1} + \gamma_0^k v^k)).$

The non-smooth line search in [9] assumes absolute continuity. Suppose the assumptions in Theorem 5.2 hold and that in addition.

$$s = \limsup_{\kappa \to 0} \frac{h(\kappa)}{\kappa} < 0.$$

Then, $h(\kappa)$ is absolutely continuous and according to Theorem 2.7 of [9], the line search algorithm given in [9], Algorithm 2.6, either

1) terminates to a step size satisfying the weak Wolfe and Armijo conditions, or

 it eventually generates a nested sequence of finite intervals which contain a set of nonzero measure of step sizes that satisfy the weak Wolfe and Armijo conditions.

VII. CONCLUSION

This paper considers the viability of the line search for projection-based switched system optimization. The cost is shown to be absolutely continuous with respect to the step size when the search direction satisfies reasonable assumptions. While assuming absolute continuity, [9] present a line search algorithm and shows that it either terminates to a step size satisfying Armijo and weak Wolfe conditions or it generates a nested sequence of finite intervals which contain step sizes satisfying the Armijo and weak Wolfe conditions. Future work is to find search directions with good convergence properties.

REFERENCES

- [1] L. Armijo. Minimization of functions having lipschitz continuous first-partial derivatives. *Pacific Journal of Mathematics*, 16:1–3, 1966.
- [2] S. C. Bengea and R. A. DeCarlo. Optimal control of switching systems. *Automatica*, 41:11–27, 2005.
- [3] T. M. Caldwell and T. D. Murphey. Switching mode generation and optimal estimation with application to skid-steering. *Automatica*, 47:50-64, 2011.
- [4] T. M. Caldwell and T. D. Murphey. Projection-based switched system optimization. *American Control Conference*, 2012. (Accepted).
- [5] M. Egerstedt, Y. Wardi, and H. Axelsson. Transition-time optimization for switched-mode dynamical systems. *IEEE Transactions on Automatic Control*, 51:110–115, 2006.
- [6] H. Gonzalez, R. Vasudevan, M. Kamgarpour, S. S. Sastry, R. Bajcsy, and C. J. Tomlin. A descent algorithm for the optimal control of constrained nonlinear switched dynamical systems. *Hybrid Systems: Computation and Control*, 13:51–60, 2010.
- [7] E. R. Johnson and T. D. Murphey. Second-order switching time optimization for non-linear time-varying dynamic systems. *IEEE Transactions on Automatic Control*, 2011. (In Press).
- [8] C. Lemarechal. A view of line-searches. Optimization and Optimal Control, 30:59–78, 1981.
- [9] A. S. Lewis and M. L. Overton. Nonsmooth optimization via BFGS. SIAM Journal of Optimization, 2009.
- [10] J. Nocedal and S. J. Wright. Numerical Optimization. Springer, 2006.
- [11] J. Le Ny, E. Feron, and G. J. Pappas. Resource constrained LQR control under fast sampling. *Hybrid Systems: Computation and Control*, 14:271–279, 2011.
- [12] C. C. Pugh. Real Mathematical Analysis. Springer, 2002.
- [13] S. Wei, K. Uthaichana, M. Zefran, R. DeCarlo, and S. Bengea. Applications of numerical optimal control to nonlinear hybrid systems. *Nonlinear Analysis: Hybrid Systems*, 1:264–279, 2007.
- [14] X. Xu and P. J. Antsaklis. Optimal control of switched systems via non-linear optimization based on direct differentiations of value functions. *International Journal of Control*, 75:1406 – 1426, 2002.