Projection-Based Optimal Mode Scheduling

T. M. Caldwell and T. D. Murphey

***Abstract*— Mode scheduling is a challenging problem: this paper furthers work on projection-based switched system optimization for calculating the optimal mode sequence and switching times of a switched system. The standard optimization algorithm of iteratively calculating a search direction, taking a step of size that satisfies convergence properties, and updating the controls is adapted to mode scheduling. Since the switching controls are constrained to the integers, a projection operator maps an unconstrained set to the set of valid switched system trajectories so that the search directions are Lebesgue integrable curves. For the specific search direction calculated from the negative mode insertion gradient, convergence guarantees are established such as sufficient descent and backtracking. Similar to derivative-based finite dimensional optimization, the convergence guarantees and the test for descent direction follow from the local approximation of the cost in the direction of the search direction. Finally, an example demonstrates the steps to implement the optimization algorithm.**

I. Introduction

This paper is concerned with the problem of switched system optimal control. Switched systems evolve over distinct dynamic modes, transitioning between the modes at discrete times. The problem is to schedule the modes—i.e. calculate the sequence of modes and the transition times—that optimize a performance index. As is common, we parameterize the mode schedule by a set of functions of time, $u(t)$, with values constrained to be either 0 or 1 [2], [3], [16]. While in general, optimization based on differentiability is not applicable to integer constrained problems, we use a projection-based technique so that the mode scheduling problem shares underlying principles, particularly absolute continuity of line search.

Projection operators are commonly employed to solve constrained optimization problems. For example, in [15], the gradient projection method is reviewed for finite dimensional inequality constrained optimization. Furthermore, in [10], a projection operator is used for optimal control of trajectory functionals.

In this paper, we continue our projection-based switched system work in [4], [5]. In [4] we showed equivalency between the projection-based switched system optimum with a hybrid maximum principle. In [5] we showed that the cost is absolutely continuous with respect to a search direction. With such a property one may expect a line search could be implemented so that sufficient descent is achieved. Indeed, this paper finds this expectation to be true.

For projection-based optimal mode scheduling, the state, x, and switching control, u, are unconstrained. In other words, x and u need not satisfy the dynamics and the value of u need not have integer value 0 or 1. However, unlike embedding methods [2], [16], [20] which embed $u(t)$ in the interval $[0, 1]$, the cost J is calculated on the projection $\mathcal{P}$ of (x, u) onto the set of non-chattering switched system trajectories. In comparison to insertion methods [8], [9], [19] since u is not constrained to the integers the local variations are curves in $\mathcal{L}_2[0, T]$ as opposed to necessarily being needle variations.

While the underlying strategy presented in this paper is fundamentally different to insertion methods, the high level algorithm is the same—i.e. to iteratively alter the mode schedule so that there are guarantees on convergence. Furthermore, both strategies base update decisions using the mode insertion gradient, defined in the insertion literature. In [8], [9], the insertion time and inserted mode are calculated directly from the mode insertion gradient, while in [19] the insertion duration is also calculated using an Armijo-like line search. In this paper, the negative mode insertion gradient is an $\mathcal{L}_2[0, T]$ variation and is a search direction similar to the negative gradient in derivative-based optimization.

In this paper, we propose an iterative mode schedule optimization algorithm. The contributions of this paper are: (A) Approximation of the cost function in the direction of the negative mode insertion gradient. (B) Showing the negative mode insertion gradient is a descent direction. (C) Testing for sufficient descent. (D) Showing that backtracking will calculate a step size that satisfies sufficient descent in a finite number of iterations. Similar to optimization techniques based on differentiability, we will find that Contributions B, C, and D follow largely from Contribution A. We show Contributions C and D for the descent direction calculated from the mode insertion gradient. We leave the results for general descent directions to future work.

This paper is organized as follows: Review of the projection operator, projection-based switched system optimization, and the mode insertion gradient is in Section II. Section III reviews the iterative optimization algorithm and discusses the challenges of calculating a step size for convergence guarantees. Section IV examines the derivative of the cost with respect to the switching times. Contribution A, the local approximation of the cost, is in Section V. Showing the negative mode insertion gradient is a descent direction, Contribution B, is in Section VI. Section VII presents both

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 sufficient descent and backtracking, Contributions C and D. Finally, an example is in Section VIII. *Due to page restrictions, the Lemma proofs are found in [7].*

II. Review

The following reviews switching control of switched systems [4], [5], the switching time gradient [3], [8], [11], [21], the max-projection operator for switched systems [4], [5], projection-based optimal mode scheduling [4], [5], and the mode insertion gradient [8], [9], [19].

A. Switched Systems

A switched system evolves according to one of N modes $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in \{1,\ldots,N\}$ at any time. The control problem is to determine the schedule over the time interval $[0,T]$ where final time $T > 0$. Note we will alternatively label the initial time $T_0 := 0$ and final time $T_M := T$. We consider three representations to parameterize a switched system: mode schedule, switching control, and active mode function. Each representation is equivalent in that a unique mapping exists between each. Depending on the material, one of the representations is often clearer for presentation than the others. For this reason, the paper switches between the representations. The three representations are:

Definition 1: The *mode schedule* is composed of the pair $\{\Sigma,\mathcal{T}\}$ where $\Sigma = \{\sigma_1,\ldots,\sigma_M\}$ is the mode sequence and $\mathcal{T} = \{T_1,\ldots,T_{M-1}\}$ is the strictly monotonically increasing set of switching times. Here, each mode is $\sigma_i \in \{1,\ldots,N\}$, each switching time is $T_i \in [0,T]$, and the total number of modes in the mode sequence is $M \in \mathbb{N}$.[1]

Definition 2: The curve $u = [u_1,\ldots,u_N]^T$ composed of N piecewise constant functions of time is a *switching control* if

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

Denote the set of all admissible switching controls as Ω.

Definition 3: The piecewise constant function of time $\sigma : [0,T] \to \{1,\ldots,N\}$ is an *active mode function* if σ does not chatter—i.e. in the interval $[0,T]$, the number of times σ switches between values $\{1,\ldots,N\}$ is finite.

A unique mapping exists between each representation: (mode schedule→switching control) given a mode schedule, $\{\Sigma,\mathcal{T}\}$, the switching control u is $u(t) = e_{\sigma_i}$ for $t \in [T_{i-1},T_i)$, $i = 1,\ldots,M$ where e_{σ_i} is the σ_i^{th} vector of the N dimensional identity matrix; (switching control→active mode function) given a switching control $u \in \Omega$, the active mode function σ for each time $t \in [0,T]$ is the $\sigma(t) \in \{1,\ldots,N\}$ for which $e_{\sigma(t)} = u(t)$; (active mode schedule→mode schedule) given an active mode function $\sigma(t)$, the mode schedule is $(\Sigma,\mathcal{T}) = (\{\sigma_1,\ldots,\sigma_M\},\{T_1,\ldots,T_{M-1}\})$ where $\mathcal{T} = \{t \in [0,T] | \sigma(t^+) \neq \sigma(t^-)\}$ and $\sigma_i = \sigma(t)$ for $t \in [T_{i-1},T_i)$, $i = 1,\ldots,M$. We will write $(\Sigma(u),\mathcal{T}(u))$, when it is necessary to be explicit the switching control the mode schedule corresponds to.

[1] We define the naturals $\mathbb{N}$ as the positive integers $\{1,2,\ldots\}$.

A switched system is then the state and the switching control, (x,u)—alternatively, $(x,(\Sigma,\mathcal{T}))$ or (x,σ)—that satisfies the state equations. Let $\mathcal{X}$ and $\mathcal{U}$ be sets 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 $x = [x_1,\ldots,x_n]^T \in \mathcal{X}$, and N switching controls $u = [u_1,\ldots,u_N]^T \in \mathcal{U}$. The switched system state equations are given by

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

Formally, define a switched system as:

Definition 4: The pair $(x,u) \in \mathcal{X} \times \mathcal{U}$ is a *non-chattering switched system* if

- $u \in \Omega$ and
- $x(t) - x(0) - \int_0^t F(x(\tau),u(\tau))d\tau = 0$ for almost all $t \in [0,T]$.

Denote the set of all such pairs of state and switching controls by $\mathcal{S}$.

B. Switching Time Gradient

The problem of optimizing the switching times when the mode sequence is fixed is considered in [3], [8], [11], [21]. Consider the problem

$$\min_{\mathcal{T}} J(\mathcal{T}) := \int_0^T \ell(x(\tau))d\tau$$

constrained to the state equation Eq.(1) with fixed Σ. Supposing each mode, $f_i(x(t))$, and the running cost, $\ell(x(t))$, is $\mathcal{C}^1$, the i^{th} switching time derivative of the cost is ([3], [8], [12], [11], [21])

$$D_{T_i} J(\mathcal{T}) = \rho^T(T_i)(f_{\sigma_i}(x(T_i)) - f_{\sigma_{i+1}}(x(T_i))) \tag{2}$$

where x is the solution to the state equations, Eq.(1), and ρ is the solution to the following adjoint equation

$$\begin{array}{c} \dot{\rho}(t) = -Df_{\sigma_i}(x(t))^T\rho(t) - D\ell(x(t))^T, \\ T_{i-1} < t < T_i \quad \text{for } i \in \{1\ldots,M\} \end{array} \tag{3}$$

where $\rho(T) = 0$.

C. Projection Operator

In [4], [5], we propose the max-projection operator. The projection maps curves from the unconstrained set $\mathcal{X} \times \mathcal{U}$ to the set of non-chattering switched systems, $\mathcal{S}$. In order to define the max-projection, we first define the mapping $\mathcal{Q} : \mathcal{U} \to \Omega$. Suppose $\mu \in \mathcal{U}$, then

$$\mathcal{Q}_i(\mu(t)) := \prod_{j\neq i}^N 1(\mu_i(t) - \mu_j(t)). \tag{4}$$

where $1 : \mathbb{R} \to \{0,1\}$ is the step function—i.e. $1(\mu_i(t) - \mu_j(t)) = 0$ if $\mu_i(t) - \mu_j(t) < 0$ and $1(\mu_i(t) - \mu_j(t)) = 1$ if $\mu_i(t) - \mu_j(t) \geq 0$. Note $\mathcal{Q}$ is not well defined for all curves in $\mathcal{U}$. For example, μ_i and μ_j may have equal greatest value for a connected interval of time. For this reason, let us only consider a subset $\mathcal{R} \subset \mathcal{U}$ for which $\mathcal{Q}$ is well defined and maps to Ω. We refer to this subset as the *admissible subset of* $\mathcal{U}$. In [5], we give a sufficient condition for a form of μ to be an element of $\mathcal{R}$.

Now, define the max-projection as:

Definition 5: Take $\mu \in \mathcal{R}$. The *max-projection*, $\mathcal{P} : \mathcal{X} \times \mathcal{R} \to \mathcal{S}$, at time $t \in [0, T]$ is

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

Notice the max-projection does not depend on α. The unconstrained state is included in the left hand side of the definition in order for $\mathcal{P}$ to be a projection. Other projections proposed in [4] do depend on α.

D. Projection-Based Optimal Mode Scheduling

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 a feasible switched system—i.e. $(x, u) \in \mathcal{S}$.

This paper furthers our work in [4], [5], in which we consider an equivalent problem to the constrained problem where the design variables are elements of an unconstrained set $(\mathcal{X},\mathcal{U})$ and the cost is evaluated on the projection of the design variables to the set of feasible switched system trajectories:

Problem 1: 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

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

E. Mode Insertion Gradient

In [8], [9], [19], the mode insertion gradient is used to decide the timing and mode inserted at each step of a descent algorithm. The mode insertion gradient has a similar role in this paper. It is the calculation of the change to the cost from inserting a mode at some time t for an infinitesimal interval. The mode insertion gradient at time $t \in [0, T]$ and mode $a \in \{1, \ldots, N\}$ is

$$d_a(t) := \rho(t)^T (f_a(x(t)) - f_{\sigma(t)}(x(t))) \tag{6}$$

where ρ is the solution to the adjoint equation Eq.(3) and $\sigma(t)$ is the active mode function [8], [9], [19]. Since the mode insertion gradient can be calculated for each $t \in [0, T]$ and mode $a \in \{1, \ldots, N\}$, define $d : [0, T] \to \mathbb{R}^N$ to be the *mode insertion gradient* of u.[2] It is the list of the N mode insertion gradients of each mode—i.e. $d(t) = \{d_1(t), \ldots, d_N(t)\}$.

[2]In this paper, the mode insertion gradient is defined as d, an n-dimensional list of curves, while in [8], [9], [19], the mode insertion gradient is $d_a(t)$, the evaluation of d for the a^{th} mode at time t.

In Section VII-A, the proof of sufficient descent relies on the assumption that $\ddot{d}_{ab}(t) := \ddot{d}_a(t) - \ddot{d}_b(t)$ is Lipschitz continuous. The following Lemma gives the conditions on f_a and f_b to ensure this assumption is valid.

Lemma 1 (Lipschitz condition for $\ddot{d}_{ab}(t)$): Suppose d is the mode insertion gradient for some $u \in \Omega$. If there exists $K_2 > 0$ such that for each $t \in [0, T]$, $x(t) \in \mathbb{R}^n$ and for each $j \in \{1, \ldots, N\}$, $f_j(x(t))$ is $\mathcal{C}^2$ and $\|D^2 f_j(x(t))\| \leq K_2$ then there is an $L > 0$ such that for each $a \neq b \in \{1, \ldots, N\}$ and $t_1, t_2 \in [0, T]$,

$$|\ddot{d}_{ab}(t_2) - \ddot{d}_{ab}(t_1)| \leq L|t_2 - t_1|$$

III. Iterative Optimization

This paper pursues the problem of calculating the switching control u and switched system state x that optimize the performance metric $J(x, u)$ using projection-based techniques. Similar to derivative-based algorithms for optimization, an iterative algorithm is proposed.

The iterative method follows. In the algorithm and for the rest of the paper, a variable with the superscript k implies that the variable depends directly on u^k.

Algorithm 1: Choose u^0 and set $k = 0$.[3]

1) Calculate $-d^k$, Eq.(6).
2) Calculate step size γ^k by backtracking, Section VII-B.
3) Update: $u^{k+1} = \mathcal{Q}(u^k - \gamma^k d^k)$—Eq. (4).
4) If u^{k+1} satisfies a terminating condition, then exit, else, increment k and repeat from step 1.

An example of one iteration of Algorithm 1 is in Fig.1. Notice in the example, the number of modes in the mode sequence of $\mathcal{Q}(u^k - \gamma^k d^k)$ increases by 4 compared with the mode sequence of u^k. Also notice if γ^k were much smaller than 1 then u^{k+1} would equal u^k. In other words, γ^k must be large enough for $\mathcal{Q}(u^k - \gamma^k d^k) \neq u^k$.

Calculating γ^k correctly is critical for the sequence of u^k generated by executing the algorithm to converge to a local minimizer. The negative mode insertion gradient, $-d^k$ must be a descent direction and the step size γ^k must be chosen so that sufficient descent is achieved.

The convergence analysis in Section VII of Algorithm 1 builds upon concepts from the literature. Convergence of iterative optimization algorithms has been thoroughly studied for both smooth [1], [15], [17] and non-smooth problems [13], [14]. Polak and Wardi, in [18], consider the case where the cost minimizing sequence is not guaranteed—or even likely—to have an accumulation point. In the context of this paper, the set of control inputs is infinite-dimensional and incomplete and therefore, the sequence of control inputs calculated by the iterative algorithm to minimize the cost might not have an accumulation point. Instead, we will provide convergence results with respect to an optimality function going to zero.

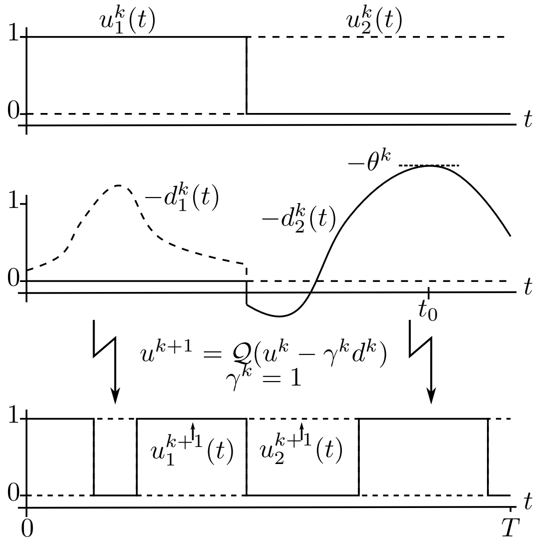

Fig. 1. Example curves $u^k = [u_1^k, u_2^k]^T \in \Omega$ and $-d^k = [-d_1^k, -d_2^k]^T$ as well as the updated curve $u^{k+1} = \mathcal{Q}(u^k - \gamma^k d^k)$ where $\gamma^k = 1$. The value $-\theta^k$ is given in Eq.(7) where t_0 is shown, the active mode is $\sigma^k(t_0) = 2$ and the inserted mode is $a_0 = 1$.

Note for the rest of the paper: Since the search direction is the negative mode insertion gradient, $-d^k$, calculated from u^k, we assume the conditions in Lemma 1 are true.

A. Sufficiently Large Step Size for Differing Mode Schedules

As can be seen in Fig.1, if γ^k is small enough, then $\mathcal{Q}(u^k - \gamma^k d^k)$ equals u^k and the updated mode schedule does not differ from the previous mode schedule. In other words, there is $\gamma_0^k > 0$ such that for every $\gamma \in [0, \gamma_0^k)$,

$$u^k = \mathcal{Q}(u^k - \gamma d^k).$$

We wish to calculate γ_0^k. Define $\sigma^k(t) \in \{1, \ldots, N\}$ as the active mode of u^k at time t. By Eq.(4), for $\mathcal{Q}(u^k - \gamma d^k)$ to differ from u^k, there must be a time $t \in [0, T]$ and mode $a \in \{1, \ldots, N\}$, $a \neq \sigma^k(t)$, for which $u^k_{a\sigma^k(t)}(t) - \gamma d^k_{a\sigma^k(t)}(t) > 0$. Note, $u^k_{a\sigma^k(t)}(t) := u^k_a(t) - u^k_{\sigma^k(t)}(t) = -1$ and $d^k_{a\sigma^k(t)}(t) = d^k_a(t)$. Therefore, this γ must be greater than $-1/d^k_a(t)$. Consequently, there must be an $a \in \{1, \ldots, N\}$ and $t \in [0, T]$ for which $d^k_a(t)$ is negative valued in order for the mode schedule of $\mathcal{Q}(u^k - \gamma d^k)$ to change for any γ. The lower bound on $\mathbb{R}^+$ for which $u^k \neq \mathcal{Q}(u^k - \gamma d^k)$, labelled γ_0^k, is calculated from the pair (a_0, t_0):

$$(a_0, t_0) = \arg \min_{a \in \{1, \ldots, N\}, t \in [0,T]} d^k_a(t)$$

Define $\theta^k \in \mathbb{R}$:

$$\theta^k := d^k_{a_0}(t_0). \tag{7}$$

This value is pictured in Fig.1. Finally,

$$\gamma_0^k := -\frac{1}{\theta^k}. \tag{8}$$

Note θ^k is always non-positive since $d^k_{\sigma^k(t)}(t) = 0$ for all $t \in [0, T]$. Since as $\theta^k \to 0$, $\gamma_0^k \to \infty$, we use θ^k as the optimality function for testing when to terminate Algorithm 1. Equivalent calculations to θ^k are used in the mode insertion gradient literature, [8], [9], [19].

IV. Derivative of the Cost with Respect to the Step Size

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. Fix $u^k \in \Omega$. We consider the cost as a function of only the step size γ. Define

$$J^k(\gamma) := J(\mathcal{P}(u^k - \gamma d^k)),$$

which is only variable on $\gamma \in \mathbb{R}^+$. Recall Contribution A of the paper. We wish to approximate $J^k(\gamma)$ near γ_0^k. In order to do so we must first investigate the derivative of $J^k(\gamma)$.

When the mode sequence is constant only the switching times of $\mathcal{Q}(u^k - \gamma d^k)$ vary as γ varies. For this reason, we use the mode schedule representation. Define $\Sigma^k(\gamma) := \Sigma(\mathcal{Q}(u^k - \gamma d^k)) = \{\sigma_1, \ldots, \sigma_M\}$ and $\mathcal{T}^k(\gamma) := \mathcal{T}(\mathcal{Q}(u - \gamma d^k)) = \{T_1(\gamma), \ldots, T_{M-1}(\gamma)\}$. The cost parameterized by the mode schedule is

$$J(\Sigma^k(\gamma), \mathcal{T}^k(\gamma)) := J^k(\gamma)$$

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

$$DJ^k(\gamma) = D_2 J(\Sigma^k(\gamma), \mathcal{T}^k(\gamma)) \cdot D\mathcal{T}^k(\gamma) \tag{9}$$

where $D_2 J(\Sigma^k(\gamma), \mathcal{T}^k(\gamma))$ is the switching time gradient, Eq.(2), and $D\mathcal{T}^k(\gamma)$ is the derivative of the switching times with respect to the step size and is given in the following lemma. The proof is in [5].

Lemma 2 (Derivative of switching times): If $u^k \in \Omega$ and $\Sigma^k(\gamma)$ is constant then the i^{th} element of the derivative of $\mathcal{T}^k(\gamma)$, $D\mathcal{T}^k(\gamma)_i = DT_i(\gamma)$, is given for the following two cases:

1) If $T_i(\gamma)$ is not a critical time of $\mu^k_{\sigma_i\sigma_{i+1}} := u^k_{\sigma_i\sigma_{i+1}} - \gamma d^k_{\sigma_i\sigma_{i+1}}$, then

$$DT_i(\gamma) = -\frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}{\gamma^2 \dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}, \tag{10}$$

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

As follows from Eq.(9), the derivative of the cost, $DJ^k(\gamma)$ is given by the dot product of the result in Lemma 2 with the switching time gradient, Eq.(2), as long as the mode sequence is constant and each switching time satisfies the conditions for either case 1 or case 2. Using this Lemma, the approximation for the cost with respect to γ is in the next section.

V. Approximation of the Cost and Switching Times

Many of the algorithms and much of the theory in optimization are designed from local approximations of the cost function—e.g. from the gradient. For the projection-based optimal mode scheduling problem, the design variable μ is infinite dimensional and $\mathcal{U}$ does not form a Hilbert space. Therefore, the gradient of the cost is not expected to exist. However, the cost may still be approximated (Contribution A of the paper). This approximation will be useful for testing a candidate descent direction (Cont. B), proving sufficient descent (Cont. C) and designing a backtracking algorithm (Cont. D) as we will see in Sections VI and VII.

A. Approximation of the Switching Times

Recall Contribution A in which we wish to locally approximate $J^k(\gamma)$ in a neighborhood of γ_0^k for $\gamma > \gamma_0^k$. There exists some $\delta\gamma > 0$ for which $\Sigma^k(\gamma)$ is constant for $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, which follows from Lemma 3 of [5]. Consequently, only $\mathcal{T}^k(\gamma)$ varies for $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$ and the approximation of $J^k(\gamma)$ depends directly on the approximation of $\mathcal{T}^k(\gamma)$.

For the mode schedule to vary for $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, at least one switching time of $\mathcal{T}^k(\gamma_0^{k^+})$ must vary with γ. Suppose $T_i(\gamma)$ is this switching time separating modes $\sigma_i \in \{1, \dots, N\}$ and $\sigma_{i+1} \in \{1, \dots, N\}$ in the mode schedule. Often, a function approximation is made from its Taylor expansion. For $T_i(\gamma)$, however, $DT_i(\gamma_0^{k^+})$ may be unbounded and in which case, $T_i(\gamma)$ would not have a first-order Taylor expansion. Referring to Eq.(10), $DT_i(\gamma_0^{k^+})$ is unbounded when $\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) = 0$ and since $\dot{d}^k_{\sigma_i\sigma_{i+1}}(\cdot) = 0$ at extremums, it is likely for $DT_i(\gamma_0^{k^+})$ to be unbounded. We will shortly present an alternative approximation for when $\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) = 0$, but since the approximation depends on whether $\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))$ is zero or not, we label the switching times at γ_0^k with a *type*.

Definition 6: Suppose $u^k \in \Omega$, $\delta\gamma > 0$, and $T_i(\gamma) \in \mathcal{T}^k(\gamma)$ is the switching time between modes σ_i and $\sigma_{i+1} \in \Sigma^k(\gamma)$ for $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$. The *type of switching time* $T_i(\gamma)$ is $m^k(T_i(\gamma_0^k)) \in \mathbb{N}$ where $m^k(T_i(\gamma_0^k)) = \min\{m \in \mathbb{N} | d^{k(m)}_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^k)) \neq 0\}$

For the purposes of this paper, we only consider type-1 and type-2 switching times since we do not foresee a reason to expect type-3 or greater. In both examples in Section VIII, only type-1 and type-2 switching times are encountered. Fig.2 shows two example sets of curves $-d^k$ for which type-1 (pictured left) and type-2 (pictured right) switching times occur. Before considering an approximation of $T_i(\gamma)$ for either type, we first show $T_i(\gamma)$ is continuous in a neighborhood of $T_i(\gamma_0^{k^+})$.

Lemma 3 (Continuity of switching times): Suppose $u^k \in \Omega$ and there exists $\delta\gamma > 0$ such that for $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, $T_i(\gamma) \in \mathcal{T}^k(\gamma)$ is the switching time between modes σ_i and $\sigma_{i+1} \in \Sigma^k(\gamma)$. If $m^k(T_i(\gamma_0^k)) = 1$ or 2, then there is $\overline{\delta\gamma} \in (0, \delta\gamma]$ such that for all $\gamma \in [\gamma_0^k, \gamma_0^k + \overline{\delta\gamma}]$, $T_i(\gamma)$ is continuous.

The approximation of $T_i(\gamma)$ for $\gamma > \gamma_0^k$ near γ_0^k when $m^k(T_i(\gamma_0^k)) = 1$ or 2 is given in the following lemma.

Lemma 4 (Approximation of switching times): Consider $u^k \in \Omega$. Suppose there exists $\delta\gamma > 0$ such that for $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, $T_i(\gamma) \in \mathcal{T}^k(\gamma)$ is the switching time between modes σ_i and $\sigma_{i+1} \in \Sigma^k(\gamma)$. There is $\overline{\delta\gamma} \in (0, \delta\gamma]$

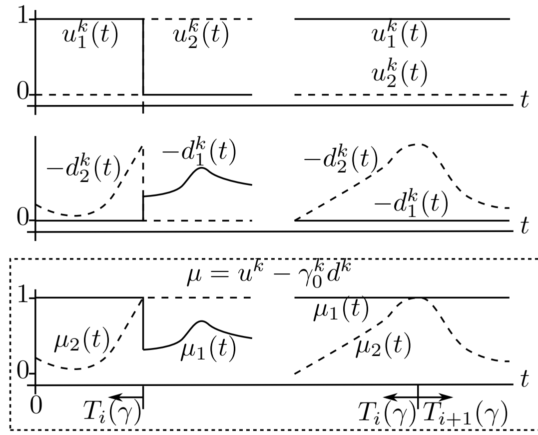

Fig. 2. Example curves of u^k, $-d^k$ and $u^k - \gamma_0^k d^k$ showing type-1 (left) and type-2 (right) switching times. The direction in time the switching times for $\gamma > \gamma_0^k$ vary are also shown.

such that for all $\gamma \in [\gamma_0^k, \gamma_0^k + \overline{\delta\gamma})$, $m^k(T_i(\gamma_0^k)) = 1$ implies

$$T_i(\gamma) = T_i(\gamma_0^{k+}) - \frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))^2}{\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}(\gamma - \gamma_0^k) + o(\gamma - \gamma_0^k) \quad (11)$$

and $m^k(T_i(\gamma_0^k)) = 2$ implies

$$T_i(\gamma) = T_i(\gamma_0^{k+}) \pm \Big[- \frac{2u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))^2}{\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}(\gamma - \gamma_0^k) + o(\gamma - \gamma_0^k)\Big]^{\frac{1}{2}} \quad (12)$$

B. Approximation of the Cost

For smooth finite dimensional optimization, the first order term of the approximation of the cost is the gradient composed with the search direction. We find that similar to the finite dimensional gradient, the mode insertion gradient, Eq.(6), has a similar role for approximating the projection-based switched system cost.

Let $\Sigma^k(\gamma) = \{\sigma_1, \dots, \sigma_M\}$ and $\mathcal{T}^k(\gamma) = \{T_1(\gamma), \dots, T_{M-1}(\gamma)\}$ be the mode schedule for $\gamma > \gamma_0^k$ near γ_0^k. Let $\tilde{J}^k(\gamma)$ be the first-order Taylor expansion of $J^k(\gamma) := J(\Sigma^k(\gamma), \mathcal{T}^k(\gamma))$, around $\mathcal{T}^k(\gamma_0^{k^+})$:

$$\tilde{J}^k(\gamma) = J^k(0) + D_2 J^k(\Sigma^k(\gamma_0^{k^+}), \mathcal{T}^k(\gamma_0^{k^+})) \cdot (\mathcal{T}^k(\gamma) - \mathcal{T}^k(\gamma_0^{k^+}))$$

The term $D_2 J(\Sigma^k(\gamma_0^{k^+}), \mathcal{T}^k(\gamma_0^{k^+}))$ is the switching time gradient Eq.(2) and thus $\tilde{J}^k(\gamma)$ becomes

$$\tilde{J}^k(\gamma) = J^k(0) + \sum_{i=1}^{M-1} \rho(T_i(\gamma_0^{k^+}))^T \cdot [f_{\sigma_i}(x(T_i(\gamma_0^{k^+}))) - f_{\sigma_{i+1}}(x(T_i(\gamma_0^{k^+})))](T_i(\gamma) - T_i(\gamma_0^{k^+})).$$

Each $T_i(\gamma)$ may be increasing in value or decreasing in value with γ. If increasing, notice $\sigma^k(T_i(\gamma)) = \sigma_{i+1}$ and if decreasing, notice $\sigma^k(T_i(\gamma)) = \sigma_i$. Thus, if $T_i(\gamma)$ is increasing in value, then

$$\begin{aligned}&\rho(T_i(\gamma_0^{k^+}))^T [f_{\sigma_i}(x(T_i(\gamma_0^{k^+}))) - f_{\sigma_{i+1}}(x(T_i(\gamma_0^{k^+})))] \\ &= \rho(T_i(\gamma_0^{k^+}))^T [f_{\sigma_i}(x(T_i(\gamma_0^{k^+}))) - f_{\sigma^k(T_i(\gamma_0^{k^+}))}(x(T_i(\gamma_0^{k^+})))] \\ &= d^k_{\sigma_i}(T_i(\gamma_0^{k^+})) = \theta^k,\end{aligned}$$

which is the mode insertion gradient of σ_i just after $T_i(\gamma_0^k)$ and is also the optimality value, Eq.(7), of u^k. Similarly, if decreasing, then

$$\begin{aligned}\rho(T_i(\gamma_0^{k^+}))^T[f_{\sigma_i}(x(T_i(\gamma_0^{k^+}))) - f_{\sigma_{i+1}}(x(T_i(\gamma_0^{k^+})))] \\ = -d^k_{\sigma_{i+1}}(T_i(\gamma_0^{k^+})) = -\theta^k.\end{aligned}$$

Set $\omega_i = 0$ if $T_i(\gamma)$ is increasing or constant in value with γ and $\omega_i = 1$ if decreasing—i.e. $\omega_i = 0$ (alt. $\omega_i = 1$) implies there is $\delta\gamma > 0$ such that for each $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, $T_i(\gamma) \geq T_i(\gamma_0^k)$ (alt. $T_i(\gamma) < T_i(\gamma_0^k)$). Then, $\tilde{J}^k(\gamma)$ is

$$\tilde{J}^k(\gamma) = J^k(0) + \textstyle\sum_{i=1}^{M-1}(-1)^{\omega_i}\theta^k(T_i(\gamma) - T_i(\gamma_0^k)). \tag{13}$$

Approximations of the switching times are given in Section V-A. Recall the different *types* of switching times. Partition $\{1, \dots, M-1\}$ into sets of equivalent type of switching time. Define I_1^k as the set of indexes of the type-1 switching times at γ_0^k and I_2^k as the set of indexes of type-2 switching times at γ_0^k. In other words, for $j = 1, 2$,

$$I_j^k = \{i \in \{1, \dots, M-1\} | m^k(T_i(\gamma_0^{k^+})) = j\}.$$

Further, define

$$m^k := max(\{m^k(T_i(\gamma_0^{k^+}))\}_{i=1}^{M-1}) \tag{14}$$

to have the value of greatest type of switching time at γ_0^k. The approximation of the switching times for $m^k(T_i(\gamma_0^{k^+})) = 1$ and 2 is given in Lemma 4. The switching times with the greatest type will dominate the approximation of the cost—e.g. type-1 switching times vary linearly with $\gamma - \gamma_0^k$ while type-2 switching times vary with $(\gamma - \gamma_0^k)^{\frac{1}{2}}$. Label the approximation of the cost with the approximation of the switching times as $\hat{J}^k(m^k;\gamma)$. If $m^k = 1$, then

$$\hat{J}^k(1;\gamma) = J^k(0) + \sum_{i \in I_1^k}(-1)^{\omega_i}\frac{(\theta^k)^3}{\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))}(\gamma - \gamma_0^k), \tag{15}$$

while if $m^k = 2$, then

$$\hat{J}^k(2;\gamma) = J^k(0) - \sum_{i \in I_2^k}\frac{\sqrt{2}(\theta^k)^2}{\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}}(\gamma - \gamma_0^k)^{\frac{1}{2}}, \tag{16}$$

The following lemma states that $\hat{J}^k(m^k;\gamma)$ dominates the remaining terms of $J^k(\gamma)$ for $\gamma > \gamma_0^k$ near γ_0^k. In other words, $\hat{J}^k(m^k;\gamma)$ is a valid approximation of $J^k(\gamma)$ near γ_0^k.

Lemma 5 (Approximation of the Cost): Set $J^k(\gamma) = \hat{J}^k(m^k;\gamma) + R(\gamma)$ where $R(\gamma)$ is the remainder. If $m^k = 1$ or 2, then there exists $\delta\gamma > 0$ such that for all $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, $|\hat{J}^k(m^k;\gamma) - J^k(0)| \geq |R(\gamma)|$.

As we show next, the negative mode insertion gradient is a descent direction.

VI. Descent Direction

In order to show sufficient descent (Cont. C) and for backtracking to be applicable (Cont. D), $-d^k$ must be a descent direction (Cont. B). In this section we prove $-d^k$ is a descent direction directly from the approximation of the cost (Cont. A).

The search direction $-d^k$ is a descent direction if there is a $\delta\gamma > 0$ such that for each $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, $J^k(\gamma) < J^k(0)$. The following lemma states that $-d^k$ is a descent direction.

Lemma 6 (Descent Direction): If $m^k = 1$ or 2 and there exists an $a \in \{1, \dots, N\}$ and a $t \in [0, T]$ for which $d_a^k(t) < 0$, then there exists $\delta\gamma > 0$ such that for each $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, $J^k(\gamma) < J^k(0)$.

The following section gives a condition on the step size for sufficient descent.

VII. Sufficient Descent

Since $-d^k$ is a descent direction, there is a $\gamma^k > \gamma_0^k$ in the neighborhood of γ_0^k such that $J^k(\gamma^k) < J^k(0)$. Therefore, by choosing such a γ^k, each execution of the loop in Algorithm 1 will result in a cost decrease from the previous iteration. Supposing $J(\cdot)$ is bounded below by $\underline{J} \in \mathbb{R}$, the algorithm will converge to a cost $H \geq \underline{J}$. However, it is unclear whether H is the cost at a local minimum unless each γ^k satisfies a sufficient descent condition and is calculated from backtracking.

It can be unclear, though, what it means for H to be a local minimum. In finite dimensional derivative-based optimization, the optimization algorithm converges to a stationarity point where the norm of the gradient of the cost is zero. Since the set $\mathcal{U}$ is infinite dimensional and not a Hilbert space, there is no reason to expect a gradient of $J(\mathcal{P}(\cdot))$ to exist. Instead of the normed gradient, we choose a different optimality function on $\mathcal{U}$ and give conditions for which it goes to zero. This optimality function is θ^k, which is calculated from Eq.(7). When $\theta^k = 0$, $\gamma_0^k = -1/\theta^k = \infty$ which implies that $-d^k$ has zero utility to reduce $J(\mathcal{P}(u^k))$ further. In that respect, u^k is a stationarity point for the descent direction $-d^k$.

In this section, we give the sufficient descent condition (Cont. C), show that a step size γ^k that satisfies the sufficient descent condition can be calculated in a finite number of backtracking iterations (Cont. D) and finally that executing Algorithm 1 for such a γ^k results in $\lim_{k\to\infty}\theta^k = 0$. Each of these contributions follows from the approximation of the cost (Cont. A).

A. Type 2 Sufficient Descent Condition

The sufficient descent condition (Cont. C) follows directly from the approximation of the cost $\hat{J}^k(m^k;\gamma)$, Eqs.(15) and (16) (Cont. A). Set $\alpha \in (0, 1)$. The type-m^k sufficient descent condition is

$$J^k(\gamma) - J^k(0) < \alpha(\hat{J}^k(m^k;\gamma) - J^k(0)).$$

We study the type-2 sufficient descent condition since the greatest type of switching time at γ_0^k is usually $m^k = 2$. In fact, in the example in Section VIII, each of the 50 iterations of Algorithm 1 inserted type-2 switching times. Except by design, m^k is rarely greater than 2. However, $m^k = 1$ is common. At γ_0^k, type-1 switching times occur at switching times of u^k or at the boundary times. Since the approximation of type-1 switching times is linear in $(\gamma - \gamma_0^k)$, for $m^k = 1$, sufficient descent and backtracking for projection-based switched system optimization and switching time optimization are equivalent—see [3], [8], [12], [21] for switching time optimization. For these reasons, only the type-2 sufficient descent is considered in this paper.

Definition 7: Set

$$s_2^k = -\sum_{i \in I_2^k} \frac{\sqrt{2}(\theta^k)^2}{\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}} \tag{17}$$

The *type 2 sufficient descent* condition is

$$J^k(\gamma) - J^k(0) < \alpha s_2^k (\gamma - \gamma_0^k)^{\frac{1}{2}} \tag{18}$$

The following Lemma finds that there exists a $\hat{\gamma} > \gamma_0^k$ for which each $\gamma \in (\gamma_0^k, \hat{\gamma}]$ satisfies the type-2 sufficient descent condition. The step size $\hat{\gamma}$ is the minimum of γ_1^k, γ_2^k and γ_3^k, each given in the lemma. The first, γ_1^k, is the step size where for each $\gamma \in (\gamma_0^k, \gamma_1^k)$, $J^k(\gamma)$ is differentiable. In other words, γ_1^k is an upper bound on where the derivative-based approximation is valid. The second, γ_2^k, depends on the constant L that satisfies the Lipschitz condition on the second time derivative of d^k, which exists based on the assumptions made in Section III and due to Lemma 1. The third, γ_3^k, is a constant scaling away from γ_0^k—i.e. $\gamma_3^k = \gamma_0^k \kappa$ where depending on $\alpha \in (0,1)$, κ is between $2 - \frac{\sqrt[3]{\alpha \frac{3\sqrt{2}}{2}}}{3} \approx 1.5717$ and 2.

In the following Lemma, set $\nu := \min_{i \in I_2^k} \ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))$.

Lemma 7: Suppose $m^k = 2$ and there exists $\gamma_1^k > \gamma_0^k$ such that for each $i \in I_2^k$ and $\gamma \in (\gamma_0^k, \gamma_1^k)$, $T_i(\gamma)$ exists. Set

$$\gamma_2^k = \gamma_0^k \left(1 - \frac{\nu^3}{\theta^k 16 L^2}\right).$$

and

$$\gamma_3^k := \gamma_0^k \left(2 - \frac{\sqrt[3]{\alpha \frac{3\sqrt{2}}{2}}}{3}\right).$$

Then, defining $\hat{\gamma}^k := \min\{\gamma_1^k, \gamma_2^k, \gamma_3^k\}$, the type-2 sufficient descent condition, Eq.(18), is true for each $\gamma \in (\gamma_0^k, \hat{\gamma}^k]$.

B. Backtracking

Calculating $\hat{\gamma}^k = \min\{\gamma_1^k, \gamma_2^k, \gamma_3^k\}$ directly is computationally inefficient due to γ_2^k. Calculating γ_1^k and γ_3^k is possible though: γ_1^k is the nearest $\gamma > \gamma_0^k$ to γ_0^k for which $J^k(\gamma)$ is not differentiable and therefore, γ_1^k can be calculated from knowledge of the critical times of u^k and d^k; γ_3^k is a constant scaling from γ_0^k. Conversely, γ_k^2 requires calculating the Lipschitz constant L a priori. Similar to smooth finite dimensional optimization [1], [13], it is more efficient to calculate a step size that satisfies the sufficient descent criteria using a backtracking method than it is to calculate γ_2^k and thus $\hat{\gamma}^k$ directly. Define $\gamma^k(j)$ as

$$\gamma^k(j) = (\gamma_3^k - \gamma_0^k)\beta^j + \gamma_0^k.$$

Now, define $j^k \in \{0, 1, \ldots\}$ for $\beta \in (0,1)$ as

$$j^k := \min\{j = \mathbb{N} | J^k(\gamma^k(j)) - J^k(0) < \alpha s_2^k (\gamma^k(j) - \gamma_0^k)^{\frac{1}{2}}\}. \tag{19}$$

Then, $\gamma^k := \gamma^k(j^k)$ satisfies the sufficient descent condition. Note, if $j^k = 0$, then $\gamma^k = \gamma_3^k$, which is a constant scaling from γ_0^k. Depending on α, $\gamma_3^k = \gamma_0^k \kappa$ where κ is a number between approximately 1.5717 and 2. The following algorithm calculates γ^k using backtracking. It should be implemented as an inner loop of Algorithm 1 at step 2.

Algorithm 2: Set $j = 0$ and calculate s_2^k from Eq.(17).

1) If $J^k(\gamma^k(j)) - J^k(0) < \alpha s_2^k(\gamma^k(j) - \gamma_0^k)^{\frac{1}{2}}$ then return $\gamma^k = \gamma^k(j)$ and terminate.
2) Increment j and repeat from Step 1.

Lemma 8 (Backtracking): If there exists $b_1 > 0$ and $b_2 > 0$ such that $\theta^k < -b_1$ and for each of the $i \in I_2^k$, $\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma)) > b_2$, then j^k is finite.

C. Locally Minimizing Sequence

For the type-2 sufficient descent condition, we have shown backtracking will find a γ^k for which the condition is satisfied. In the following lemma, we find that if $\{u^k\}$ is the sequence calculated using Algorithm 1 from u^0 where there is an infinite subsequence of $\{u^k\}$ for which $m^k = 2$, then the optimality function θ^k goes to zero.

Lemma 9: Suppose $u^0 \in \Omega$ and $S = \{u^k\}$ is an infinite sequence where

1) $J(u^0) = \overline{J} < \infty$,
2) $J(u)$ is bounded below for all $u \in \Omega$,
3) $J(u^{k+1}) < J(u^k)$, and
4) $S_2 \subset S$ is an infinite subsequence where each $u^{k+1} \in S_2$ is calculated from $u^{k+1} = \mathcal{Q}(u^k - \gamma^k d^k)$ and
   a) $m^k = 2$ (see Eq.(14)),
   b) $\gamma_2^k < \gamma_1^k$ or $\gamma_3^k < \gamma_1^k$ (see Lemma 7),
   c) there is $K_2 > 0$ such that for each $i \in I_2^k$, $\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^k)) \geq K_2$, and
   d) $\gamma^k = (\gamma_3^k - \gamma_0^k)\beta^{j^k} + \gamma_0^k$ (see Eq.(19)).

then, $\lim_{k \to \infty} \theta^k = 0$.

The restrictive assumption in Lemma 9 is assumption 4b. If the greatest γ for which the derivative-based approximation of the cost is valid goes to zero at a faster rate than θ^k goes to zero, then the minimizing sequence is not be guaranteed to converge to $\theta^k = 0$.

VIII. Example

Consider the linear time-invariant switched system example in [6] and [8]. Suppose $x_0 = (1,1)^T$ and $f_1(x(t)) = A_1 x(t)$ and $f_2(x(t)) = A_2 x(t)$ where

$$A_1 = \begin{pmatrix} -1 & 0 \\ 1 & 2 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}.$$

We wish to solve Problem 1—i.e. to find the switching control inputs that minimize $J(x,u) = \int_0^1 \frac{1}{2} x(\tau)^T Q x(\tau) d\tau$ where Q is the identity matrix. We executed Algorithm 1

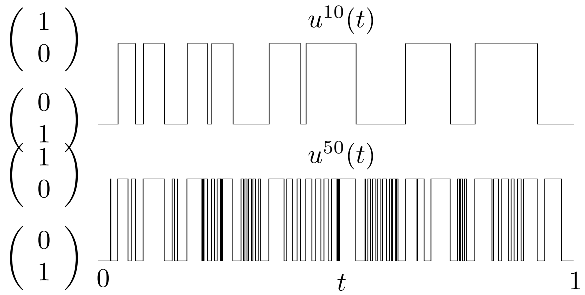

Fig. 3. Results of u^{10} and u^{50}.

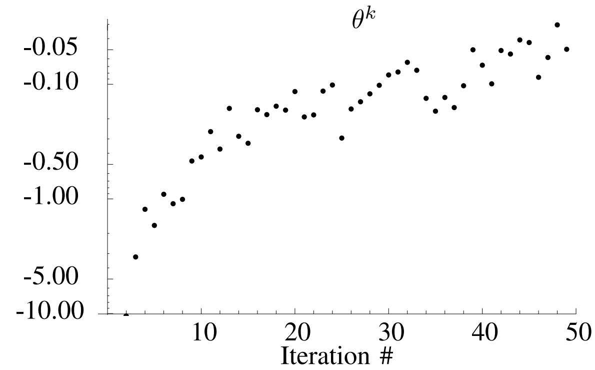

Fig. 4. Plot of optimality condition, θ^k, for 50 iterations. Notice the optimality condition decreases with iteration.

for 50 iterations starting with mode sequence $\Sigma^0 = \{1\}$ and switching time $\mathcal{T}^0 = \{\}$. This initial mode schedule is the switching control $u^0 = (1,0)^T(t)$.

The procedure for one iteration of Algorithm 1 follows. The initial state, x^0, is calculated from Eq.(1). With u^0 and x^0, the adjoint for the first iterate, ρ^0, is calculated from Eq.(3). The negative mode insertion gradient, $-d^0$ is calculated next from Eq.(6), which is guaranteed to be a descent direction due to Lemma 6. Now, for any $\gamma > 0$ the update, $\mathcal{Q}(u^0 - \gamma d^0)$ is calculated from Eq.(4) and thus the projection, $(x', u') = \mathcal{P}(u^0 - \gamma d^0)$, is calculated from Eq.(5). Notice the cost of the update at γ, $J^0(\gamma)$, is integrated from $\ell(x', u')$. The optimality function, θ^0, and the smallest step size for which there is change to the cost, γ_0^k, are calculated from Eqs.(7) and (8) respectively. The greatest switching time type at γ_0^k, m^0, is calculated next from Eq.(14). If $m^0 = 1$, then backtracking using the type-1 sufficient descent condition may be done to calculate u^1. In this example, however, $m^0 = 2$ and thus type-2 backtracking, Algorithm 2, is executed to calculate γ^0. We used $\alpha = 10^{-4}$ and $\beta = 10^{-1}$. The new estimate of the optimum is found from $u^1 = \mathcal{Q}(u^0 - \gamma^0 d^0)$.

We repeated this process for 50 iterations. In accordance with Lemma 8, each j^k was finite and in fact the greatest number of back stepping iterations needed was 3. Note we did not optimize the switching times between each iteration.

After the 50^{th} iteration, the cost reduced from 11.8372 down to 2.2122. The switching control calculated at the 50^{th} iterations is given in Fig.(3). Even though u^0 has zero mode transitions, u^{10} has 16 mode transitions and u^{50} has 98. Finally, Fig.(4) shows that the optimality condition, θ^k, trends toward 0 where $\theta^{50} = -0.04854$.

IX. Conclusion

An algorithm for optimal mode scheduling of switched systems is given. The algorithm has guarantees on convergence such as sufficient descent and backtracking for the search direction given by calculating the negative mode insertion gradient. The mode insertion gradient is a single example of a descent direction. In future work, we will study sufficient descent and backtracking for general descent directions.

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.

[5] T. M. Caldwell and T. D. Murphey. Projection-based switched system optimization: Absolute continuity of the line search. *IEEE Conference on Decision and Control*, 2012.

[6] T. M. Caldwell and T. D. Murphey. Single integration optimization of linear time-varying switched systems. *IEEE Transactions on Automatic Control*, 57:1592–1597, 2012.

[7] T. M. Caldwell and T. D. Murphey. Projection-based optimal mode scheduling. `http://nxr.northwestern.edu/sites/default/files/publications/caldwellmurphey-2013CDCSubmission.pdf`, 2013.

[8] M. Egerstedt, Y. Wardi, and H. Axelsson. Transition-time optimization for switched-mode dynamical systems. *IEEE Transactions on Automatic Control*, 51:110–115, 2006.

[9] 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.

[10] J. Hauser. A projection operator approach to the optimization of trajectory functionals. *IFAC World Congress*, 2002.

[11] E. R. Johnson and T. D. Murphey. Second-order switching time optimization for non-linear time-varying dynamic systems. *IEEE Transactions on Automatic Control*, 56(8):1953–1957, 2011.

[12] K. Flaßkamp, T. D. Murphey, and S. Ober-Blöbaum. Switching time optimization in discretized hybrid dynamical systems. *IEEE Conference on Decision and Control*, pages 707–712, 2012.

[13] C. Lemarechal. A view of line-searches. *Optimization and Optimal Control*, 30:59–78, 1981.

[14] A. S. Lewis and M. L. Overton. Nonsmooth optimization via BFGS. *SIAM Journal of Optimization*, 2009.

[15] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2006.

[16] 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.

[17] E. Polak. *Optimization: Algorithms and Consistent Approximations*. Springer, 1997.

[18] E. Polak and Y. Wardi. A study of minimizing sequences. *SIAM Journal on Control and Optimization*, pages 599–609, 1984.

[19] Y. Wardi and M. Egerstedt. Algorithm for optimal mode scheduling in switched systems. *American Control Conference*, pages 4546–4551, 2012.

[20] 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.

[21] 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.