# Projection-Based Optimal Mode Scheduling

T. M. Caldwell and T. D. Murphey

***Abstract*— This paper develops an iterative optimization technique that can be applied to mode scheduling. The algorithm provides both a mode schedule and timing of that mode schedule with convergence guarantees. Moreover, the algorithm takes advantage of a line search, and the number of iterations in the line search is bounded. There are two key ingredients in the algorithm. First, a projection operation is used that takes arbitrary curves and maps them to feasible switching controls. Second, a descent direction that incorporates the projection is calculated using the mode insertion gradient. Similar to derivative-based finite dimensional optimization, the convergence guarantees and sufficient decrease criteria follow from a local approximation of the cost in the direction of the search direction, but this local approximation is not the standard quadratic approximation. An example demonstrates the steps to implement the optimization algorithm and illustrates convergence.**

## 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], [15]. 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 [14], the gradient projection method is reviewed for finite dimensional inequality constrained optimization. Furthermore, in [9], a projection operator is used for optimal control of trajectory functionals.

Optimization techniques based on differentiability locally approximate the cost function in order to calculate a new estimate of the optimum [14]. In finite dimensions, the descent direction is calculated from the gradient and Hessian which give the first- and second-order approximations of the cost [14]. Furthermore, tests such as descent direction and sufficient descent depend on the gradient [1], [14].

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 will result in sufficient descent guarantees for convergence. 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], [15], [19] 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 [7], [8], [18] 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 [7], [8], the insertion time and inserted mode are calculated directly from the mode insertion gradient, while in [18] 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 numerical 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 optimal mode scheduling, and the mode insertion gradient is in Section II. Section III

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

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 the sufficient descent and backtracking, Contributions C and D. Finally, examples are in Section VIII.

## II. Review

The following reviews switching control of switched systems [4], [5], the switching time gradient [3], [7], [10], [20], the max-projection operator for switched systems [4], [5], projection-based optimal mode scheduling [4], [5], and the mode insertion gradient [7], [8], [18].

### 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, throughout the paper, we will switch 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 $\Omega$.

*Definition 3:* The piecewise constant function of time $\sigma : [0, T] \to \{1, \ldots, N\}$ is an *active mode function* if $\sigma$ does not chatter—i.e. in the interval $[0, T]$, the number of times $\sigma$ 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 $e_{\sigma_i}$ is the $\sigma_i^{\text{th}}$ vector of the $N$ dimensional identity matrix; (switching control→active mode function) given a switching control $u \in \Omega$, the active mode function $\sigma$ 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.

A switched system is then the state and the switching control, $(x, u)$—alternatively, $(x, (\Sigma, \mathcal{T}))$ or $(x, \sigma)$—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 $t \in [0, T]$.[2]

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], [7], [10], [20]. 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 $\Sigma$. Supposing each mode, $f_i(x(t))$, and the running cost, $\ell(x(t))$, is $\mathcal{C}^1$, the $i^{\text{th}}$ switching time derivative of the cost is ([3], [7], [11], [10], [20])

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

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

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

where $\rho(T) = 0$.

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

[2] The integral is the Lebesgue integral.

### 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, $\mu_i$ and $\mu_j$ 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 $\Omega$. We refer to this subset as the *admissible subset of* $\mathcal{U}$. In [5], we give a sufficient condition for a form of $\mu$ 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

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

Notice the max-projection does not depend on $\alpha$. 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 $\alpha$.

### 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

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

Notice the cost is calculated on admissible state and switching control trajectories. Furthermore, Problem 1 is equivalent to the constrained problem $\arg\min_{(x,u)\in\mathcal{S}} J(x, u)$ [4], [5].

### E. Mode Insertion Gradient

For projection-based switched system optimization, the cost does not have a natural gradient. However, it does have a function with a similar role in the optimization as the gradient does for finite dimensional optimization. This function is referred to as the *mode insertion gradient* [7], [8], [18]. The mode insertion gradient calculates 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 $a \in \{1, \dots, N\}$ is

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

where $\rho$ is the solution to the adjoint equation Eq.(3) and $\sigma(t)$ is the active mode function [7], [8], [18]. Since the mode insertion gradient can be calculated for each $t \in [0, T]$ and mode $a \in \{1, \dots, N\}$, define $d : [0, T] \to \mathbb{R}^N$ to be the *mode insertion gradient* of $u$.[3] It is the list of the $N$ mode insertion gradients of each mode—i.e. $d(t) = \{d_1(t), \dots, d_N(t)\}$.

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 $j \in \{1, \dots, N\}$, $f_j(x(t))$ is $\mathcal{C}^2$ and $\|D^2 f_j(x(t))\| \leq K_2$ then there is an $L > 0$ such that for each $a \neq b \in \{1, \dots, N\}$ and $t_1, t_2 \in [0, T]$,

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

*Proof:* First, $d_{ab} = \rho(t)^T (f_a(x(t)) - f_b(x(t)))$ where the state and adjoint equations are in Eqs.(1) and (3). Consider each $t \in [0, T]$, $x(t) \in \mathbb{R}^n$ and $j \in \{1, \dots, N\}$. Since $\|D^2 f_j(x(t))\| < K_2$, there is a $K_0 > 0$ and $K_1 > 0$ such that $\|f_j(x(t))\| \leq K_0$ and $\|Df_j(x(t))\| \leq K_1$. Therefore, for all $u \in \Omega$, $\dot{x}(t) = F(x(t), u(t)) \leq K_0$ and

$$\ddot{x}(t) = D_1 F(x(t), u(t)) F(x(t), u(t)) \leq K_0 K_1.$$

By the assumptions on $F(x(t), u(t))$, for $u \in \Omega$, $F(x(t), u(t))$ is piecewise continuous with respect to $t$ and Lipschitz with respect to $x(t)$. Therefore, $x(t; u)$, defined as the solution to Eq.(1) for $u \in \Omega$, is unique.

Define $g(\rho(t)) := -D_1 F(x(t; u), u(t))^T \rho(t) - D\ell(x(t; u))$.

Since $\|D_1 F(x(t; u), u(t))\| \leq K_1$ and $D_1 F(x(t; u), u(t))$ is piecewise continuous in $t$ for all $u \in \Omega$,

$$\begin{aligned} &\|g(\rho_2(t)) - g(\rho_1(t))\| \\ &\quad \leq \|D_1 F(x(t; u), u(t))^T\| \|\rho_2(t) - \rho_1(t)\| \\ &\quad = K_1 \|\rho_2(t) - \rho_1(t)\|. \end{aligned} \tag{7}$$

Therefore, $\rho(t; u)$, defined as the solution to Eq.(3) for $u \in \Omega$, is unique. Thus, there is a $K_0' > 0$ such that for all $u \in \Omega$, $\rho(t; u) \leq K_0'$. Additionally, there is a $K_1' > 0$ such that for

[3] In this paper, the mode insertion gradient is defined as $d$, an $n$-dimensional list of curves, while in [7], [8], [18], the mode insertion gradient is $d_a(t)$, the evaluation of $d$ for the $a^{\text{th}}$ mode at time $t$.

all $u \in \Omega$, $\|\dot{\rho}(t;u)\| = \|g(\rho(t;u))\| \le K_1'$. It follows that for each $t_1, t_2 \in [0,T]$, $\|\rho(t_2;u) - \rho(t_1;u)\| < K_1'|t_2 - t_1|$. From Eq.(7), there is $L_1$ such that $\|\dot{\rho}(t_2;u) - \dot{\rho}(t_1;u)\| \le L_1|t_2 - t_1|$. Note,

$$\begin{aligned}\ddot{\rho}(t;u) = &-D_1^2 F(x(t;u),u(t)) \circ (\rho(t;u), F(x(t;u),u(t))) \\ &-D_1 F(x(t;u),u(t))^T \dot{\rho}(t;u) \\ &-D^2 \ell(x(t;u)) F(x(t;u),u(t)).\end{aligned}$$

By the bounds on $F(\cdot,\cdot)$, $DF(\cdot,\cdot)$, and $D^2F(\cdot,\cdot)$, and that $\rho(t;u)$ and $\dot{\rho}(t;u)$ are Lipschitz, there is $L_2 > 0$ such that $\|\dot{\rho}(t_2;u) - \dot{\rho}(t_1;u)\| \le L_2|t_2 - t_1|$. Finally, by these bounds and that $\rho(t;u)$, $\dot{\rho}(t;u)$ and $\ddot{\rho}(t;u)$ are Lipschitz, $\ddot{d}_{ab}$ is Lipschitz with some constant $L > 0$. ∎

## 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.

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 sufficient decrease in cost is achieved. The descent must be sufficient so that the sequence generated by the iterative optimization algorithm converges to a stationarity point—or at least for an optimality function to go to zero.

The problem of convergence of iterative optimization algorithms is considered for both smooth [1], [14], [16] and non-smooth problems [12], [13]. Polak and Wardi, in [17], 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. Indeed, in Wardi's recent work with Egerstedt, [18], on switched system optimization, Wardi and Egerstedt argue this point when comparing their iterative algorithm and convergence results to Gonzalez et al.'s similar results [8]. Wardi and Egerstedt give their convergence result with respect to an optimality function going to zero while Gonzalez et al. assume an accumulation point exists. We provide a similar result to Wardi and Egerstedt in Section VII-C.

The iterative method follows. Note, 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 $k = 0$.[4]
1) Calculate $-d^k$, Eq.(6).
2) Calculate step size $\gamma^k$ 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.

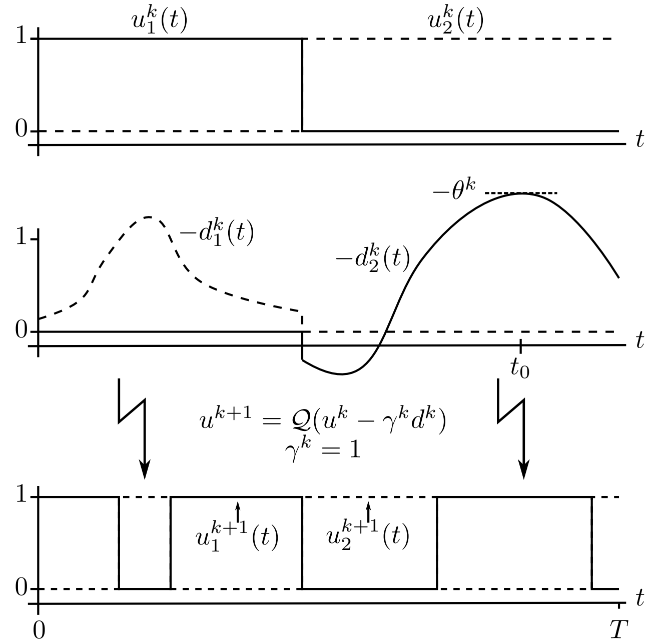


Fig. 1. Example curves $u^k = [u_1^k, u_2^k]^T \in \Omega$ and $-d^k = [-d_1^k, -d_2^k]^T$ as well as the updated curve $u^{k+1} = \mathcal{Q}(u^k - \gamma^k d^k)$ where $\gamma^k = 1$. The value $-\theta^k$ is given in Eq.(8) where $t_0$ is shown, the active mode is $\sigma^k(t_0) = 2$ and the inserted mode is $a_0 = 1$.

Calculating $\gamma^k$ correctly is critical for the sequence of $u^k$ generated by executing the algorithm to locally minimize the cost. Remarks:

1) The negative mode insertion gradient, $-d^k$ must be a descent direction in order to guarantee there is a $\gamma^k \in \mathbb{R}^+$ for which $J(u^{k+1}) < J(u^k)$. The definition of and proof that $-d^k$ is a descent direction are given in Section VI.
2) The step size $\gamma^k$ must be chosen so that a sufficient descent is achieved—i.e. so that if the algorithm calculates an infinite sequence $\{u^k\}$ then $\lim_{k\to\infty} J(u^k)$ has locally minimal value. The sufficient descent condition and proof of existence of a step size achieving sufficient descent is in Section VII-A. Furthermore, Section VII-B considers backtracking for calculating such a step size.

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 $\gamma^k$ were much smaller than 1 then $u^{k+1}$ would equal $u^k$. In other words, $\gamma^k$ must be large enough for $\mathcal{Q}(u^k - \gamma^k d^k) \ne u^k$.

*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. In other words,

- the state $x^k$—calculated as the solution to the state equations, Eq.(1), with switching control $u^k$—is such that $x^k(t) \in \mathbb{R}^n$ for each $t \in [0,T]$,
- for each $i \in \{1,\dots,N\}$, $f_i \in \mathcal{C}^2$, and
- there is $K > 0$ such that for each $t \in [0,T]$ and $i \in \{1,\dots,N\}$, $\|D^2 f_i(x^k(t))\| \le K$.

### A. Sufficiently Large Step Size for Differing Mode Schedules

As can be seen in Fig.1, if $\gamma^k$ 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 $\gamma_0^k$. 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 $\gamma$ must be greater than $-1/d^k_a(t)$. Consequently, there must be a $a \in \{1, \ldots, N\}$ 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 $\gamma$.

The lower bound on $\mathbb{R}^+$ for which $u^k \neq \mathcal{Q}(u^k - \gamma d^k)$, labelled $\gamma_0^k$, 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 $\theta^k \in \mathbb{R}$:

$$\theta^k := d^k_{a_0}(t_0). \tag{8}$$

This value is pictured in Fig.1. Finally,

$$\gamma_0^k := -\frac{1}{\theta^k}. \tag{9}$$

Note $\theta^k$ 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 $\theta^k$ as the optimality function for testing when to terminate Algorithm 1. Equivalent calculations to $\theta^k$ are used in the mode insertion gradient literature, [7], [8], [18].

## 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 $\gamma$. 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 $\gamma_0^k$. In order to do so we must first investigate the derivative of $J^k(\gamma)$.

Note, when the mode sequence is constant, only the switching times of $\mathcal{Q}(u^k - \gamma d^k)$ vary as $\gamma$ varies. Define $\Gamma^k$ as the $\gamma \in \mathbb{R}^+$ where the mode sequence changes.[5]

$$\begin{aligned}\Gamma^k := \{\gamma \in \mathbb{R}^+ | \forall \delta\gamma > 0, \exists \gamma' \in B_{\delta\gamma}(\gamma) \cap \mathbb{R}^+, \\ \text{where } \Sigma(\mathcal{Q}(u^k - \gamma d^k)) \neq \Sigma(\mathcal{Q}(u^k - \gamma' d^k))\}.\end{aligned}$$

The mode sequence is constant for all $\gamma \notin \Gamma_{u,v}$ and only the switching times vary. 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 $\gamma$, the derivative of the cost with respect to $\gamma$ is

$$DJ^k(\gamma) = D_2 J(\Sigma^k(\gamma), \mathcal{T}^k(\gamma)) \cdot D\mathcal{T}^k(\gamma) \tag{10}$$

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 $\gamma \notin \Gamma_{u,v}$—i.e. $\Sigma^k(\gamma)$ is constant—, then the $i^{\text{th}}$ 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

$$DT_i(\gamma) = -\frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}{\gamma^2 \dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}, \tag{11}$$

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.(10), 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. In general, the derivative will not exist everywhere. For example, $DJ^k(\gamma)$ goes unbounded for $\gamma$ where $\mathcal{Q}(u^k - \gamma d^k)$ has a switching time $T_i(\gamma)$ that approaches $t_{crit} \in [0, T]$ where $\dot{d}^k_{\sigma_i\sigma_{i+1}}(t_{crit}) = 0$—see Eq.(11). In fact, the derivative of the cost will likely go unbounded at $\gamma_0^k$ since $\dot{d}_{\sigma_i\sigma_{i+1}}(t) = 0$ at maximum and minimum points. When this is the case, the cost can not be approximated at $\gamma_0^k$ directly using a Taylor expansion. However, the cost may still be approximated, as we will see in the next section.

## V. Approximation of the Cost and Switching Times

Many of the algorithms and theory in optimization is designed from local approximations of the cost function. Indeed, the gradient and Hessian are the solutions to local quadratic models [21]. Also, there are derivative-free methods that make descent decisions based on local approximations [14].

For the projection-based optimal mode scheduling problem, the design variable $\mu$ 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 (Contribution B), proving sufficient descent (Contribution C) and designing a backtracking algorithm (Contribution D) as we will see in Sections VI and VII.

[5]The ball $B_{\delta\gamma}(\gamma) = (\gamma - \delta\gamma, \gamma + \delta\gamma)$.

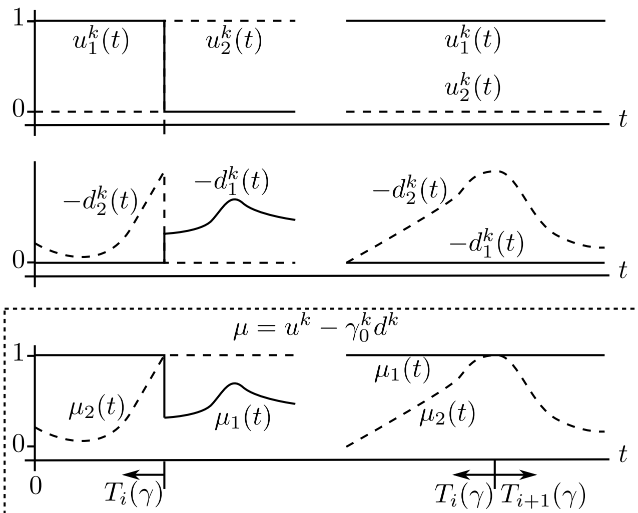


Fig. 2. Example curves of $u^k$, $-d^k$ and $u^k - \gamma_0^k d^k$ showing type-1 (left) and type-2 (right) switching times. The direction in time the switching times for $\gamma > \gamma_0^k$ vary are also shown.

### A. Approximation of the Switching Times

Recall Contribution A in which we wish to locally approximate $J^k(\gamma)$ in a neighborhood of $\gamma_0^k$ 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 $\gamma$. 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.(11), $DT_i(\gamma_0^{k^+})$ 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 $\gamma_0^k$ 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 $\sigma_i$ 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 will 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 were 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 even 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 $\sigma_i$ 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.

*Proof:* The proof follows from the two facts that $\ddot{d}^k_{\sigma_i\sigma_{i+1}}(\cdot)$ is Lipschitz and $DT_i(\gamma)$ exists when $T_i(\gamma)$ is not a critical point of $\mu^k(\cdot) := u^k_{\sigma_i\sigma_{i+1}}(\cdot) - \gamma d^k_{\sigma_i\sigma_{i+1}}(\cdot)$—see Lemmas 1 and 2 respectively.

If $m^k(T_i(\gamma_0^k)) = 1$, then $\dot{d}^k_{\sigma_i\sigma_{i+1}}(\gamma_0^k) \neq 0$. Since $\ddot{d}^k_{\sigma_i\sigma_{i+1}}(\cdot)$ is Lipschitz, $\dot{d}^k_{\sigma_i\sigma_{i+1}}(\cdot)$ is continuous. Therefore, there is $\delta t \in \mathbb{R}$ such that for all $T_i(\gamma) \in [T_i(\gamma_0^k), T_i(\gamma_0^k) + \delta t]$, $\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) \neq 0$. Since $DT_i(\gamma)$ exists when $\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) \neq 0$, 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.

Finally, If $m^k(T_i(\gamma_0^k)) = 2$, then $\dot{d}^k_{\sigma_i\sigma_{i+1}}(\gamma_0^k) = 0$ but $\ddot{d}^k_{\sigma_i\sigma_{i+1}}(\gamma_0^k) \neq 0$. Furthermore, since $\ddot{d}^k_{\sigma_i\sigma_{i+1}}(\cdot)$ is Lipschitz, there is $\delta t \in \mathbb{R}$ such that for all $T_i(\gamma) \in [T_i(\gamma_0^k), T_i(\gamma_0^k) + \delta t]$, $d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))$ and $\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))$ are strictly monotonic. Consequently, $\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) \neq 0$ for $T_i(\gamma) \in (T_i(\gamma_0^k), T_i(\gamma_0^k) + \delta t]$ and thus by Lemma 2, $DT_i(\gamma)$ exists. It follows that 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. All that remains is to prove $T_i(\gamma_0^k)$ is continuous from the right. Recall $u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) - \gamma d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) = 0$ for $T_i(\gamma)$ to be a switching time. Rewriting, $d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) = u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))/\gamma$. Since $d^k_{\sigma_i\sigma_{i+1}}(t)$ is strictly monotonic for $t \in [T_i(\gamma_0^k), T_i(\gamma_0^k) + T_i(\gamma_0^k + \overline{\delta\gamma})]$, $d^k_{\sigma_i\sigma_{i+1}}$ is bijective in this domain and the inverse is continuous. Thus, $T_i(\gamma) = d^{k^{-1}}_{\sigma_i\sigma_{i+1}}(u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))/\gamma)$ where $d^{k^{-1}}_{\sigma_i\sigma_{i+1}}(\cdot)$ is the inverse function of $d^k_{\sigma_i\sigma_{i+1}}(\cdot)$. There is $\epsilon > 0$, such that $|T_i(\gamma_0^k) - T_i(\gamma)| < \epsilon$. Therefore,

$$\left| d^{k^{-1}}_{ab} \left( \frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^k))}{\gamma_0^k} \right) - d^{k^{-1}}_{ab} \left( \frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}{\gamma} \right) \right| < \epsilon.$$

By the continuity of $d^{k^{-1}}_{ab}(\cdot)$, there exists some $\bar{\delta}(\epsilon)$ such that

$$\left| \frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^k))}{\gamma_0^k} - \frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma))}{\gamma} \right| = \left| \frac{\gamma - \gamma_0^k}{\gamma\gamma_0^k} \right| < \bar{\delta}(\epsilon).$$

Following,

$$|\gamma - \gamma_0^k| < \left|\bar{\delta}(\epsilon)\gamma\gamma_0^k\right| < \min\left\{ \left|\bar{\delta}(\epsilon)(\gamma_0^{k^2} + \gamma_0^k)\right|, 1 \right\} =: \delta(\epsilon),$$

which proves $\epsilon - \delta$ continuity. ■

The approximation of $T_i(\gamma)$ for $\gamma > \gamma_0^k$ near $\gamma_0^k$ 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 $\sigma_i$ and $\sigma_{i+1} \in \Sigma^k(\gamma)$. There is $\overline{\delta\gamma} \in (0, \delta\gamma]$

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

$$T_i(\gamma) = T_i(\gamma_0^{k+}) - \frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))^2}{\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}(\gamma-\gamma_0^k) + o(\gamma-\gamma_0^k) \tag{12}$$

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

$$T_i(\gamma) = T_i(\gamma_0^{k+}) \pm \Big[ -\frac{2u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))^2}{\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}(\gamma-\gamma_0^k) \tag{13}$$
$$hspace10pt + o(\gamma-\gamma_0^k)\Big]^{\frac{1}{2}}$$

*Proof:* For $T_i(\gamma)$ to be a switching time, $u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) - \gamma d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) = 0$. Define $\tau(\gamma) = T_i(\gamma) - T_i(\gamma_0^k)$. Begin with the case where $m^k(T_i(\gamma_0^k)) = 1$. Taylor expand $d^k_{\sigma_i\sigma_{i+1}}(\cdot)$ around $T_i(\gamma_0^{k^+})$ for a neighborhood of $T_i(\gamma_0^{k^+})$:

$$-\frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}{\gamma} + d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) + \dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\tau(\gamma) + o(\tau(\gamma)) = 0.$$

Since $\ddot{d}^k_{\sigma_i\sigma_{i+1}}(\cdot)$ is Lipschitz, there is a $\delta t \in \mathbb{R}$ such that this equation is valid for each $T_i(\gamma) \in [T_i(\gamma_0^k), T_i(\gamma_0^k) + \delta t)$. Further, due to Lemma 3, $T_i(\gamma)$ is continuous and therefore there exists a $\delta\gamma' \in (0, \delta\gamma]$ such that this equation is valid for $\gamma \in [\gamma_0^k, \gamma_0^k + \delta\gamma')$. Noting $d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) = u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))/\gamma_0^k$ and reordering,

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

Taylor expanding $\frac{\gamma-\gamma_0^k}{\gamma\gamma_0^k}$ around $\gamma_0^k$, concludes in Eq.(12).

Now consider the case $m^k(T_i(\gamma_0^k)) = 2$ where $\dot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) = 0$. Taylor expand $d^k_{\sigma_i\sigma_{i+1}}(\cdot)$ around $T_i(\gamma_0^{k^+})$ in a neighborhood of $T_i(\gamma_0^k)$ and recall $u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) - \gamma d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma)) = 0$:

$$-\frac{u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}{\gamma} + d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) + \tfrac{1}{2}\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\tau(\gamma)^2 + o(\tau(\gamma)^2) = 0.$$

Since $\ddot{d}^k_{\sigma_i\sigma_{i+1}}(\cdot)$ is Lipschitz, there is a $\delta t \in \mathbb{R}$ such that this equation is valid for each $T_i(\gamma) \in [T_i(\gamma_0^k), T_i(\gamma_0^k) + \delta t)$. Further, due to Lemma 3, $T_i(\gamma)$ is continuous and therefore there is a $\delta\gamma' \in (0, \delta\gamma]$ such that this equation is valid for $\gamma \in [\gamma_0^k, \gamma_0^k + \delta\gamma')$. Note $d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) = u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))/\gamma_0^k$. Reordering,

$$\tfrac{1}{2}\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\tau(\gamma)^2 = -u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\tfrac{\gamma-\gamma_0^k}{\gamma\gamma_0^k} + o(\tau(\gamma)^2).$$

Taylor expanding $\frac{\gamma-\gamma_0^k}{\gamma\gamma_0^k}$ around $\gamma_0^k$,

$$\tfrac{1}{2}\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\tau(\gamma)^2 = -u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\tfrac{\gamma-\gamma_0^k}{{\gamma_0^k}^2} + o(\gamma-\gamma_0^k) + o(\tau(\gamma)^2), \tag{14}$$

By the Taylor expansion of $d^k_{\sigma_i\sigma_{i+1}}(\cdot)$ around $T_i(\gamma_0^{k^+})$, $o(\tau(\gamma)^2)$ is of lesser order than $\frac{1}{2}\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\tau(\gamma)^2$. In order for the equality of Eq. (14) to be true, $o(\tau(\gamma)^2)$ must also be of lesser order than $\gamma - \gamma_0^k$. Therefore, $o(\tau(\gamma)^2) = o(\gamma-\gamma_0^k)$. Recall $\gamma_0^k = u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))/d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))$ and reorder:

$$\tau(\gamma)^2 = -u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\frac{2d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))^2}{\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}(\gamma-\gamma_0^k) + o(\gamma-\gamma_0^k). \tag{15}$$

There is $\overline{\delta\gamma} \in (0, \delta\gamma']$ such that for each $\gamma \in [\gamma_0^k, \gamma_0^k + \overline{\delta\gamma})$,

$$\left| \frac{2d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))^2}{\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}(\gamma-\gamma_0^k) \right| > o(\gamma-\gamma_0^k).$$

Since $m^k(T_i(\gamma_0^{k^+})) = 2$, $u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))$ and $\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))$ must have opposite signs. Therefore,

$$-u^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))\frac{2d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))^2}{\ddot{d}^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+}))}(\gamma-\gamma_0^k) > 0$$

As such, the right side of Eq.(15) has a single positive real valued square root and a single negative real valued square root for each $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma]$, completing the proof. ■

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

$$\tilde{J}^k(\gamma) = J^k(0) + D_2J^k(\Sigma^k(\gamma_0^{k^+}), \mathcal{T}^k(\gamma_0^{k^+})) \cdot (\mathcal{T}^k(\gamma) - \mathcal{T}^k(\gamma_0^{k^+}))$$

The term $D_2J(\Sigma^k(\gamma_0^{k^+}), \mathcal{T}^k(\gamma_0^{k^+}))$ is the switching time gradient Eq.(2) and thus $\tilde{J}^k(\gamma)$ becomes

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

By the definition of $\gamma_0^k$, there is at least one $T_i(\gamma) \in \mathcal{T}^k(\gamma)$ that is not constant for $\gamma > \gamma_0^k$ near $\gamma_0^k$. If increasing, notice the active mode function at time $T_i(\gamma)$ is $\sigma^k(T_i(\gamma)) = \sigma_{i+1}$. Alternatively, if decreasing, notice $\sigma^k(T_i(\gamma)) = \sigma_i$. Thus, if $T_i(\gamma)$ is increasing, 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 $\sigma_i$ just after $T_i(\gamma_0^k)$ and is also the optimality value, Eq.(8), of $u^k$. Similarly, if decreasing, then

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

Set $\omega_i = 0$ if $T_i(\gamma)$ is increasing or constant in value with $\gamma$ 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

$$\tilde{J}^k(\gamma) = J^k(0) \\ + \textstyle\sum_{i=1}^{M-1}(-1)^{\omega_i}\theta^k(T_i(\gamma) - T_i(\gamma_0^k)). \tag{16}$$

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 $\gamma_0^k$ and $I_2^k$ as the set of indexes of type-2 switching times at $\gamma_0^k$. 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

$$m^k := max(\{m^k(T_i(\gamma_0^{k^+}))\}_{i=1}^{M-1}) \tag{17}$$

to have the value of greatest type of switching time at $\gamma_0^k$. The approximation of the switching times for $m^k(T_i(\gamma_0^{k^+})) = 1$ and 2 is given in Lemma 4. We see that 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

$$\hat{J}^k(1;\gamma) = J^k(0) + \sum_{i \in I_1^k}(-1)^{\omega_i}\frac{(\theta^k)^3}{\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))}(\gamma - \gamma_0^k), \tag{18}$$

while if $m^k = 2$, then

$$\hat{J}^k(2;\gamma) = J^k(0) - \sum_{i \in I_2^k}\frac{\sqrt{2}(\theta^k)^2}{\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}}(\gamma - \gamma_0^k)^{\frac{1}{2}}, \tag{19}$$

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 $\gamma_0^k$. In other words, $\hat{J}^k(m^k;\gamma)$ is a valid approximation of $J^k(\gamma)$ near $\gamma_0^k$.

*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 $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, $|\hat{J}^k(m^k;\gamma) - J^k(0)| \geq |R(\gamma)|$.

*Proof:* The first order approximation of $J^k(\gamma)$ with respect to $\tau(\gamma) := \mathcal{T}^k(\gamma) - \mathcal{T}^k(\gamma_0^k)$ is $\tilde{J}^k(\gamma)$, see Eq.(16). Thus,

$$J^k(\gamma) = \tilde{J}^k(\gamma) + o(|\tau(\gamma)|).$$

The approximation $\hat{J}^k(m^k;\gamma)$ is a further approximation from $\tilde{J}^k(\gamma)$, which includes the approximation of $\tau(\gamma)_i := T_i(\gamma) - T_i(\gamma_0^k)$ using Lemma 4. Consider $m^k = 1$ first. Set $H = (I_1^k)^c$ as the complement of $I_1^k$. By the definition of $m^k$, for each $h \in H$, $\tau(\gamma)_h = 0$. Therefore, using Eq.(12), $\tau(\gamma)$ varies linearly with $\gamma - \gamma_0^k$ and thus,

$$J^k(\gamma) = \hat{J}^k(1;\gamma) + o(\gamma - \gamma_0^k).$$

Therefore, $R(\gamma) = o(\gamma - \gamma_0^k)$ and $|\hat{J}^k(1;\gamma) - J^k(0)| \geq |R(\gamma)|$.

Now for the case where $m^k = 2$. First, the approximations of $\tau(\gamma)_i = T_i(\gamma) - T_i(\gamma_0^k)$ for $i \in \{1,\dots,M-1\}$ are at least of order $(\gamma - \gamma_0^k)^{\frac{1}{2}}$ and thus $o(|\tau(\gamma)|) = o((\gamma - \gamma_0^k)^{\frac{1}{2}})$. Second, set $H = (I_2^k)^c$ is the complement of $I_2^k$. For each $h \in H$, $\tau(\gamma)_h$ is at least order $(\gamma - \gamma_0^k)$. Thus, for each $h \in H$,

$$(-1)^{\omega_h}\theta^k\tau(\gamma)_h = o(\gamma - \gamma_0^k).$$

Finally, plugging Eq.(13) in for each $i \in I_2^k$,

$$(-1)^{\omega_i}\theta^k\tau(\gamma)_i = \\ = -\frac{\sqrt{2}(\theta^k)^2}{\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}}(\gamma - \gamma_0^k)^{\frac{1}{2}} + o((\gamma - \gamma_0^k)^{\frac{1}{2}}).$$

Therefore, the remainder term is

$$R(\gamma) = \sum_{i \in I^k(m^k)} o((\gamma - \gamma_0^k)^{\frac{1}{2}}) + \sum_{h \in H} o((\gamma - \gamma_0^k)) \\ + o((\gamma - \gamma_0^k)^{\frac{1}{2}}) = o((\gamma - \gamma_0^k)^{\frac{1}{2}}).$$

Since $\hat{J}^k(2;\gamma) - J^k(0)$ is not $o((\gamma - \gamma_0^k)^{\frac{1}{2}})$, the lemma is proven. ∎

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

## VI. Descent Direction

In order to show sufficient descent (Contribution C) and for backtracking to be applicable (Contribution D), $-d^k$ must be a descent direction (Contribution B). In this section we prove $-d^k$ is a descent direction directly from the approximation of the cost (Contribution 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 $\gamma \in (\gamma_0^k, \gamma_0^k + \delta\gamma)$, $J^k(\gamma) < J^k(0)$.

*Proof:* First, note $\theta_k \leq d_a^k(t) < 0$. The proof follows from showing $\hat{J}^k(m^k;\gamma) - J^k(0) < 0$, $m^k = 1$ or 2, for $\gamma > \gamma_0^k$ and invoking Lemma 5 to argue $\hat{J}^k(m^k;\gamma)$ dominates the remainder for $\gamma$ near $\gamma_0^k$. Refer to Eqs.(12) and (13) for $\hat{J}^k(m^k;\gamma)$ and consider $m^k = 1$ first. Clearly $\hat{J}^k(1;\gamma) - J^k(0) < 0$ if for each $i \in I_1^k$,

$$(-1)^{\omega_i}\frac{(\theta^k)^3}{\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))} < 0. \tag{20}$$

Recall for $T_i(\gamma_0^{k^+})$ to be a switching time, $u_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) - \gamma_0^k d^k_{\sigma_i\sigma_{i+1}}(T_i(\gamma_0^{k^+})) = 0$. Using the $\omega_i$ notation, $u_{\sigma_{i+\omega_i}\sigma^k(T_i(\gamma_0^{k^+}))}(T_i(\gamma_0^{k^+})) -$

$\gamma_0^k d^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})) = -1 - \gamma_0^k d^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})) = 0$. The derivative with respect to $\gamma_0^k$ must be zero:

$$d^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})) + \gamma_0^k \dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))\dot{T}_i(\gamma_0^{k^+}) = 0$$

Rearranging and noting $d_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})) = \theta^k < 0$,

$$\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})) = -\frac{\theta^k}{\gamma_0^k \dot{T}_i(\gamma_0^{k^+})}.$$

Recall $\omega_i = 0$ implies $\dot{T}_i(\gamma_0^{k^+}) > 0$ and thus $\dot{d}^k_{\sigma_i}(T_i(\gamma_0^{k^+})) > 0$. Similarly, $\omega_i = 1$ implies $\dot{T}_i(\gamma_0^{k^+}) < 0$ and thus $\dot{d}^k_{\sigma_{i+1}}(T_i(\gamma_0^{k^+})) < 0$. Therefore, Eq.(20) is true.

Now for $m^k = 2$. Since $\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})) > 0$,

$$-\sum_{i\in I_2^k} \frac{\sqrt{2}(\theta^k)^2}{\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}}(\gamma - \gamma_0^k)^{\frac{1}{2}} < 0$$

and thus $\hat{J}^k(2;\gamma) - J^k(0) < 0$. Since $\hat{J}^k(m^k;\gamma) - J^k(0) < 0$, $m^k = 1$ or 2, for all $\gamma > \gamma_0^k$ and by Lemma 5, $\hat{J}^k(m^k;\gamma) - J^k(0)$ dominates the remainder for $\gamma$ near $\gamma_0^k$, the Lemma is proved. ∎

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 $\gamma_0^k$ such that $J^k(\gamma^k) < J^k(0)$. Therefore, by choosing such a $\gamma^k$, 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 $\gamma^k$ 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 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 $\theta^k$, which is calculated from Eq.(8).[6] If $\theta^k = 0$, then $\gamma_0^k = -1/\theta^k = \infty$ 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 (Contribution C), show that a step size $\gamma^k$ that satisfies the sufficient descent condition can be calculated in a finite number of backtracking iterations (Contribution D) and finally that executing Algorithm 1 for such a $\gamma^k$ results in $\lim_{k\to\infty}\theta^k = 0$. Each of these contributions follows from the approximation of the cost (Contribution A).

[6]The optimality function $\theta^k$ has the same role in [16], [8], [18].

### A. Type 2 Sufficient Descent Condition

The sufficient descent condition (Contribution C) follows directly from the approximation of the cost $\hat{J}^k(m^k;\gamma)$, Eqs.(18) and (19) (Contribution 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 $\gamma_0^k$ 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 $\gamma_0^k$, 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], [7], [11], [20] for switching time optimization. For these reasons, only the type-2 sufficient descent is considered in this paper.

*Definition 7:* Set

$$s_2^k = -\sum_{i\in I_2^k} \frac{\sqrt{2}(\theta^k)^2}{\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}} \tag{21}$$

The *type 2 sufficient descent* condition is

$$J^k(\gamma) - J^k(0) < \alpha s_2^k(\gamma - \gamma_0^k)^{\frac{1}{2}} \tag{22}$$

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 $\gamma_1^k$, $\gamma_2^k$ and $\gamma_3^k$, each given in the lemma. The first, $\gamma_1^k$, is the step size where for each $\gamma \in (\gamma_0^k, \gamma_1^k)$, $J^k(\gamma)$ is differentiable. In other words, $\gamma_1^k$ is an upper bound on where the derivative-based approximation is valid. The second, $\gamma_2^k$, 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, $\gamma_3^k$, is a constant scaling away from $\gamma_0^k$—i.e. $\gamma_3^k = \gamma_0^k\kappa$ where depending on $\alpha \in (0,1)$, $\kappa$ 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.(22), is true for each $\gamma \in (\gamma_0^k, \hat{\gamma}^k]$.

*Proof:* Recall from Eqs.(8) and (9), $\theta^k = -1/\gamma_0^k = d^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})) < 0$ for each $i \in I_2^k$. Also, since $\ddot{d}^k(\cdot)$ is

Lipschitz and $-1-\gamma d^k_{\sigma_i+\omega_i}(T_i(\gamma)) = 0$, for each $i \in I_2^k$, there is a neighborhood of $\gamma_0^k$ for which $d^k_{\sigma_i+\omega_i}(T_i(\gamma)) > 0$, $(-1)^{\omega_i}\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma)) > 0$ and $\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma)) > 0$. Set

$$H(\gamma) := -\alpha\sqrt{2}\text{card}(I_2^k)\frac{(\theta^k)^2}{\nu^{\frac{1}{2}}}(\gamma-\gamma_0^k)^{\frac{1}{2}}$$

Notice the right hand side of the type-2 sufficient descent condition, Eq.(22), is greater than $H(\gamma)$ for all $\gamma > \gamma_0^k$. The proof follows by finding the $\gamma \in (\gamma_0^k, \gamma_1^k)$ for which the derivative of left hand side of Eq.(22) is more negative than the derivative of the right hand side. The derivative of the left hand side is

$$DJ^k(\gamma) = \sum_{i\in I_2^k}(-1)^{\omega_i}\frac{d^k_{\sigma_i+\omega_i}(T_i(\gamma))^3}{\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma))}$$

which is negative valued. The derivative of the right hand side is bounded below by $DH(\gamma)$:

$$DH(\gamma) := -\alpha\frac{\sqrt{2}}{2}\text{card}(I_2^k)\frac{(\theta^k)^2}{\nu^{\frac{1}{2}}}(\gamma-\gamma_0^k)^{-\frac{1}{2}} \tag{23}$$

The rest of the proof shows $DJ^k(\gamma) < DH(\gamma)$ for all $\gamma \in (\gamma_0^k, \hat{\gamma}^k)$.

Set $\tau_i(\gamma) = T_i(\gamma) - T_i(\gamma_0^k)$. Since $\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma))$ is Lipschitz, by the mean value theorem,

$$(-1)^{\omega_i}\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma)) \le \ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^k))\tau(\gamma) - L\tau(\gamma)^2.$$

Therefore, for $\tau_i(\gamma) \le \tau_{i,max} := \frac{\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^k))}{2L}$

$$(-1)^{\omega_i}\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma)) \le \frac{3}{2}\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^k))\tau_i(\gamma). \tag{24}$$

By Lipschitz, a lower bound of $\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma))$ for $\tau_i(\gamma) \le \tau_{i,max}$ is

$$\begin{aligned}\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma)) &\ge \ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})) + L\tau_i(\gamma)\\ &\ge \tfrac{1}{2}\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+})).\end{aligned}$$

By the Taylor expansion of $-1-\gamma d^k_{\sigma_i+\omega_i}(T_i(\gamma))$ around $T_i(\gamma)$, with remainder $r(T_i(\gamma))$,

$$\frac{-1}{\gamma} + \frac{1}{\gamma_0^k} + \frac{1}{2}r(T_i(\gamma))\tau_i(\gamma)^2 = 0.$$

For $\tau(\gamma) < \tau_{i,max}$ the lower bound of $\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma))$ is also the lower bound of the remainder term. In other words, $r(T_i(\gamma)) > \frac{1}{2}\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))$ and thus for $\tau_i(\gamma) < \tau_{i,max}$,

$$\tau_i(\gamma) \ge \frac{-2\theta^k}{\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}}(\gamma-\gamma_0^k)^{\frac{1}{2}}. \tag{25}$$

Indeed, for each $i \in I_2^k$ and $\gamma \in (\gamma_0^k, \min\{\gamma_1^k, \gamma_2^k\}]$, the right hand side of Eq.(25) is less than or equal to $\tau_{i,max}$. Plugging $\gamma_2^k$ into the right hand side of Eq.(25) reduces to,

$$\frac{\nu^{\frac{3}{2}}}{2L\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}} \le \frac{\nu}{2L} \le \tau_{i,max}.$$

Therefore, Eqs (24) and (25) are true for every $\gamma \in (\gamma_0^k, \min\{\gamma_1^k, \gamma_2^k\}]$. For these $\gamma$, an upper bound on $(-1)^{\omega_i}\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma))$ is

$$(-1)^{\omega_i}\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma)) \le -3\theta^k\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))^{\frac{1}{2}}(\gamma-\gamma_0^k)^{\frac{1}{2}}.$$

Let $\overline{\nu} = \max_{i\in I_2^k}\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))$ and $\psi = \overline{\nu}/\nu > 1$. Thus, for each $i \in I_2^k$,

$$(-1)^{\omega_i}\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma)) \le -3\theta^k(\nu\psi)^{\frac{1}{2}}(\gamma-\gamma_0^k)^{\frac{1}{2}}. \tag{26}$$

To find a upper bound on $d^k_{\sigma_i+\omega_i}(T_i(\gamma))$, integrate Eq.(24) with respect to $\tau_i(\gamma)$.

$$\begin{aligned}d^k_{\sigma_i+\omega_i}(T_i(\gamma)) &< \theta^k + \int_0^{\tau_i(\gamma)}\tfrac{3}{2}\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^k))s\,ds\\ &= \theta^k + \tfrac{3}{4}\ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^k))\tau_i(\gamma)^2\end{aligned}$$

Using the bound in Eq.(25) and by setting $\beta(\gamma) = 1 + 3\theta^k(\gamma-\gamma_0^k)$,

$$d^k_{\sigma_i+\omega_i}(T_i(\gamma)) \le \theta^k\beta(\gamma) \tag{27}$$

With the bounds on $d^k_{\sigma_i+\omega_i}(T_i(\gamma))$, Eq.(27), and $(-1)^{\omega_i}\dot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma))$, Eq.(26), $DJ^k(\gamma)$ is bounded above by

$$DJ^k(\gamma) \le -\text{card}(I_2^k)\frac{\beta(\gamma)^3}{3}\frac{(\theta^k)^2}{(\nu\psi)^{\frac{1}{2}}}(\gamma-\gamma_0^k)^{-\frac{1}{2}}. \tag{28}$$

Comparing Eqs. (23) and (28),

$$\beta(\gamma)^3 \ge \alpha\frac{3\sqrt{2}}{2}\psi^{\frac{1}{2}} \ge \alpha\frac{3\sqrt{2}}{2},$$

implies $DJ^k(\gamma) < DH(\gamma)$, which is valid for every $\gamma \in \min\{\gamma_1^k, \gamma_2^k, \gamma_3^k\} = \hat{\gamma}^k$. It follows that each $\gamma \in (\gamma_0^k, \hat{\gamma}^k]$ satisfies the sufficient descent condition. ■

### *B. Backtracking*

Calculating $\hat{\gamma}^k = \min\{\gamma_1^k, \gamma_2^k, \gamma_3^k\}$ directly is computationally inefficient due to $\gamma_2^k$. Calculating $\gamma_1^k$ and $\gamma_3^k$ is possible though: $\gamma_1^k$ is the nearest $\gamma > \gamma_0^k$ to $\gamma_0^k$ for which $J^k(\gamma)$ is not differentiable and therefore, $\gamma_1^k$ can be calculated from knowledge of the critical times of $u^k$ and $d^k$; $\gamma_3^k$ is a constant scaling from $\gamma_0^k$. Conversely, $\gamma_k^2$ requires calculating the Lipschitz constant $L$ a priori. Similar to smooth finite dimensional optimization [1], [12], it is more efficient to calculate a step size that satisfies the sufficient descent criteria using a backtracking method than it is to calculate $\gamma_2^k$ 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

$$j^k := \min\{j = \mathbb{N} | J^k(\gamma^k(j)) - J^k(0) < \alpha s_2^k(\gamma^k(j) - \gamma_0^k)^{\frac{1}{2}}\}. \tag{29}$$

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 $\gamma_0^k$. Depending on $\alpha$, $\gamma_3^k = \gamma_0^k\kappa$ where $\kappa$ is a number between approximately 1.5717 and 2. The following algorithm calculates $\gamma^k$ 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.(21).

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.

*Proof:* The proof follows from Lemmas 3 and 7. According to Lemma 7, $\hat{\gamma}^k = \min\{\gamma_1^k, \gamma_2^k, \gamma_3^k\}$ satisfies the sufficient descent condition. From Lemma 3, $\gamma_1^k$ is bounded from $\gamma_0^k$. Furthermore, by the bounds on $\theta^k$ and $\ddot{v}^k_{\sigma_i+\omega_i}(T_i(\gamma))$, $\gamma_2^k$ and $\gamma_3^k$ are bounded from $\gamma_0^k$. Let $b_3 > 0$ bound $\hat{\gamma}^k$ from $\gamma_0^k$—i.e. $\hat{\gamma}^k - \gamma_0^k > b_3$. Then,

$$j^k = \text{ceil}\left(\log_\beta \frac{b_3}{\gamma_3^k - \gamma_0^k}\right)$$

which is finite.[7] ■

### C. Locally Minimizing Sequence

For the type-2 sufficient descent condition, we have shown backtracking will find a $\gamma^k$ 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 $\theta^k$ 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.(17)),
   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.(29)).

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

*Proof:* Since each $J(u^k)$ is strictly monotonically decreasing and bounded below, $\lim_{k\to\infty} J(u^k) - J(u^{k+1}) = 0$. Consider $u^{k+1} \in S_2$ which was calculated from $u^k$ using backtracking so that $u^k - \gamma^k d^k$ satisfies the type 2 sufficient descent condition, Eq.(22) and set $\nu^k := \min_{i\in I_2^k} \ddot{d}^k_{\sigma_i+\omega_i}(T_i(\gamma_0^{k^+}))$. The cost difference between $u^{k+1}$ and $u^k$ is

$$J(u^k) - J(u^{k+1}) > \alpha\sqrt{2}\text{card}(I_2^k)\frac{(\theta^k)^2}{(\nu^k)^{\frac{1}{2}}}(\gamma^k - \gamma_0^k)^{\frac{1}{2}}. \quad (30)$$

Since $S_2$ has infinite cardinality, it is the case that as $k \to 0$, the right hand side of Eq.(30) goes to zero. By Lemma 7 and the assumption on $\gamma_1^k$, $\gamma_2^k$, and $\gamma_3^k$, any $\gamma \in (\gamma_0^k, \min\{\gamma_2^k, \gamma_3^k\}]$, defined in Lemma 7, satisfies the type-2 sufficient descent condition. Let $L$ be the Lipschitz constant of $\ddot{d}^k_a(\cdot)$ for each $a \in \{1, \ldots, N\}$ and every $u^k \in \mathcal{S}^2$. This constant exists due to the assumptions made in Section III. Recall $\gamma^k = (\gamma_3^k - \gamma_0^k)\beta^{j^k} + \gamma_0^k$ is calculated by backtracking and therefore, if $\gamma_3^k \leq \gamma_2^k$, then $\beta^{j^k} = 0$ and $\gamma^k = \gamma_3^k$.

[7]The function ceil$(\cdot) : \mathbb{R} \to \mathbb{Z}$ rounds to the nearest integer of greater value.

Conversely, suppose $\gamma_2^k < \gamma_3^k$. Due to backtracking, it is possible for $\gamma^k = (\gamma_3^k - \gamma_0^k)\beta^{j^k} + \gamma_0^k < \gamma_2^k$. If this is the case, then $(\gamma_3^k - \gamma_0^k)\beta^{j^k-1} + \gamma_0^k > \gamma_2^k$. Therefore, $\gamma^k$ is in the interval

$$\gamma^k \in [(\gamma_2^k - \gamma_0^k)\beta + \gamma_0^k, \gamma_2^k]$$

and thus

$$\gamma^k = \gamma_0^k + \psi^k \frac{(\nu^k)^3}{(\theta^k)^2 16L^2} \quad (31)$$

where $\psi^k \in [\beta, 1]$.

By assumptions, it must be the case that there are an infinite number of $u^{k+1}$ calculated from $u^k$ where either 1) $\gamma^k = \gamma_3^k$ or 2) $\gamma^k$ is given by Eq.(31). Since $\lim_{k\to\infty} J(u^k) - J(u^{k+1}) = 0$, the limit of the right hand side of Eq.(30) goes to zero. If case 1), then

$$\lim_{k\to\infty} \alpha\sqrt{2}\text{card}(I_2^k)\left(1 - \frac{\sqrt[3]{\alpha\frac{3\sqrt{2}}{2}}}{3}\right)^{\frac{1}{2}} \frac{(\theta^k)^{\frac{3}{2}}}{(\nu^k)^{\frac{1}{2}}} = 0.$$

Since $\nu^k \leq LT$, $\lim_{k\to\infty} \theta^k = 0$. Now, if 2), then

$$\lim_{k\to\infty} \frac{\alpha\sqrt{2\psi^k}\text{card}(I_2^k)}{4L}\theta^k\nu^k = 0.$$

Since $\nu^k \geq K_2$ and $\psi^k \geq \beta > 0$, once again, $\lim_{k\to\infty} \theta^k = 0$ and the proof is complete. ■

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

## VIII. Example

We consider two examples. The first is an instructive example from the literature. The switched system is composed of two linear modes. The second example considers a simple aircraft model with three flight modes. The airplane will either fly straight, bank right, or bank left, all at fixed velocities and at a fixed altitude. The goal is to schedule the flight pattern that best approximates an infeasible desired turning maneuver.

### A. Scheduling a Linear Two-Mode Switched System

Consider the linear time-invariant switched system example in [6] and [7]. 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 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 $u^0$

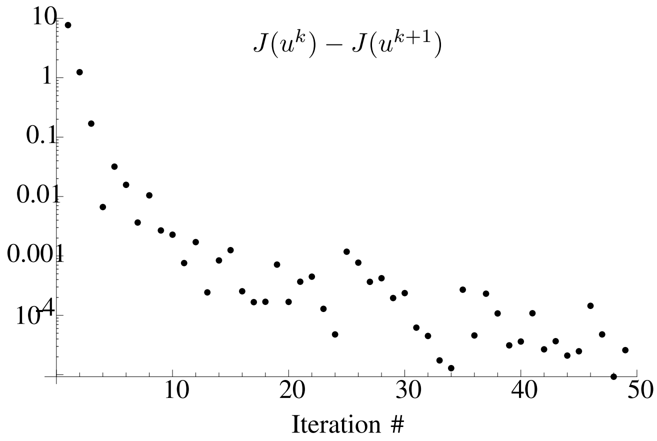


Fig. 3. Plot of $J(u^k) - J(u^{k+1})$ for 50 iterations. Notice the difference between successive costs decreases with iteration.

and $x^0$, the adjoint for the first iterate, $\rho^0$, 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 $\gamma$, $J^0(\gamma)$, is integrated from $\ell(x', u')$. The optimality function, $\theta^0$, and the smallest step size for which there is change to the cost, $\gamma_0^k$, are calculated from Eqs.(8) and (9) respectively. The greatest switching time type at $\gamma_0^k$, $m^0$, is calculated next from Eq.(17). 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 $\gamma^0$. 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$^{\text{th}}$ iteration, the cost reduced from 11.8372 down to 2.2122. See Fig.(3) for the difference of current and previous cost at each iteration. The switching control calculated at the 50$^{\text{th}}$ iterations is given in Fig.(4). Even though $u^0$ has zero mode transitions, $u^{10}$ has 16 mode transitions and $u^{50}$ has 98. Finally, Fig.(5) shows that the optimality condition, $\theta^k$, trends toward 0 where $\theta^{50} = -0.04854$.

### B. Scheduling Flight Pattern for Simple Aircraft Model

For the second example, consider an aircraft with three modes of flight at a fixed altitude. The plane's state is given by its position in $\mathbb{R}^2$, $(X, Y)$ and its orientation $\Theta$—i.e. $x(t) = (X, Y, \Theta)(t)$. It flies straight, banks right, or banks left at fixed velocities. The three modes are

$$f_i(x(t)) = (4\cos(\Theta(t)), 4\sin(\Theta(t)), \omega_i)^T$$

where $\omega_1 = 0$, $\omega_2 = 1$, and $\omega_3 = -1$. The goal is to schedule the flight pattern that best approximates the turning maneuver pictured by the dashed lines in Fig.(6). This desired trajectory is

$$x_{des}(t) = \begin{pmatrix} X_{des} \\ Y_{des} \\ \Theta_{des} \end{pmatrix}(t) = \begin{cases} \begin{pmatrix} 2\sqrt{3}t \\ 2t \\ -\pi/3 \end{pmatrix} & t < 10 \\ \begin{pmatrix} 20\sqrt{3} - 20 + 2t \\ 20\sqrt{3} + 20 - 2\sqrt{3}t \\ \pi/6 \end{pmatrix} & t \geq 10 \end{cases}.$$

The desired flight trajectory is to fly along a straight path for 10s, conduct an infeasible $\pi/2$ radian point turn, and fly straight once again. At best, the aircraft can approximate the turn. Furthermore, the aircraft is not initially positioned or oriented with the desired trajectory. The initial condition is $x_0 = (5, 2, 5\pi/18)^T$.

The objective is to calculate the flight schedule that best approximates the desired trajectory. In other words, the goal is to calculate the switching control that minimizes

$$J(u) = \frac{1}{2}\int_0^{15} (x(\tau) - x_{des}(\tau))^T (x(\tau) - x_{des}(\tau)) d\tau.$$

Starting with an initial guess of $u^0(t) = (1, 0, 0)^T(t)$, we executed Algorithm 1 for 70 iterations. We choose sufficient descent parameter $\alpha = 0.4$ and backtracking parameter $\beta = 0.4$. Note, *switching time optimization is not conducted between each iteration.* The switching control and state

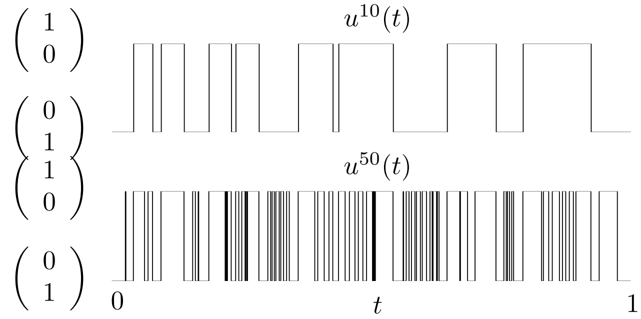


Fig. 4. Results of $u^{10}$ and $u^{50}$.

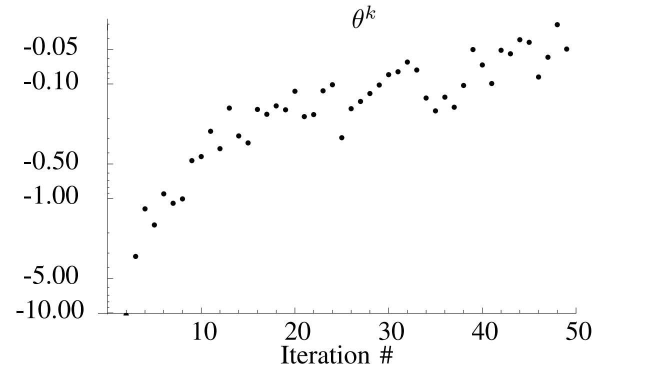


Fig. 5. Plot of optimality condition $\theta^k$ on a log scale for 50 iterations. Notice the optimality condition decreases with iteration.

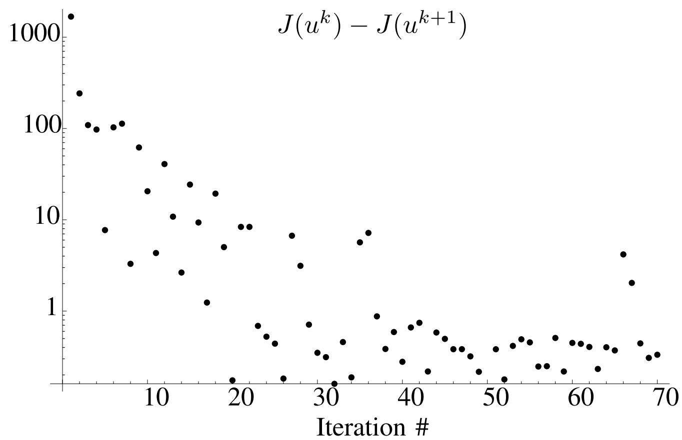


Fig. 7. Plot of $J(u^k) - J(u^{k+1})$ for 70 iterations.

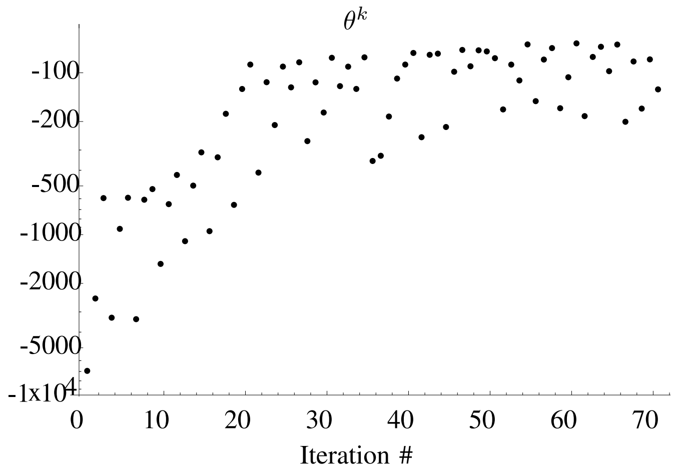


Fig. 8. Plot of optimality function, $\theta^k$, on a log scale for 70 iterations.

of iterations 5, 20 and 70 are pictured in Fig.(6). The desired trajectory is also pictured for comparison. Notice, that despite a poor initial guess—i.e. fly straight for the time interval—by the 70$^{\text{th}}$ execution of the algorithm loop, the aircraft approximates the desired trajectory. The difference of the cost between successive iterations is given in Fig.(7). Furthermore, the value of the optimality function for each iteration is given in Fig.(8).

## 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.

## X. Acknowledgments

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.

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

[8] 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.

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

[10] 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.

[11] 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.

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

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

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

[15] 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.

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

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

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

[19] 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.

[20] 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.

[21] E. Zeidler. *Applied Functional Analysis: Applications to Mathematical Physics*. Springer, 1995.

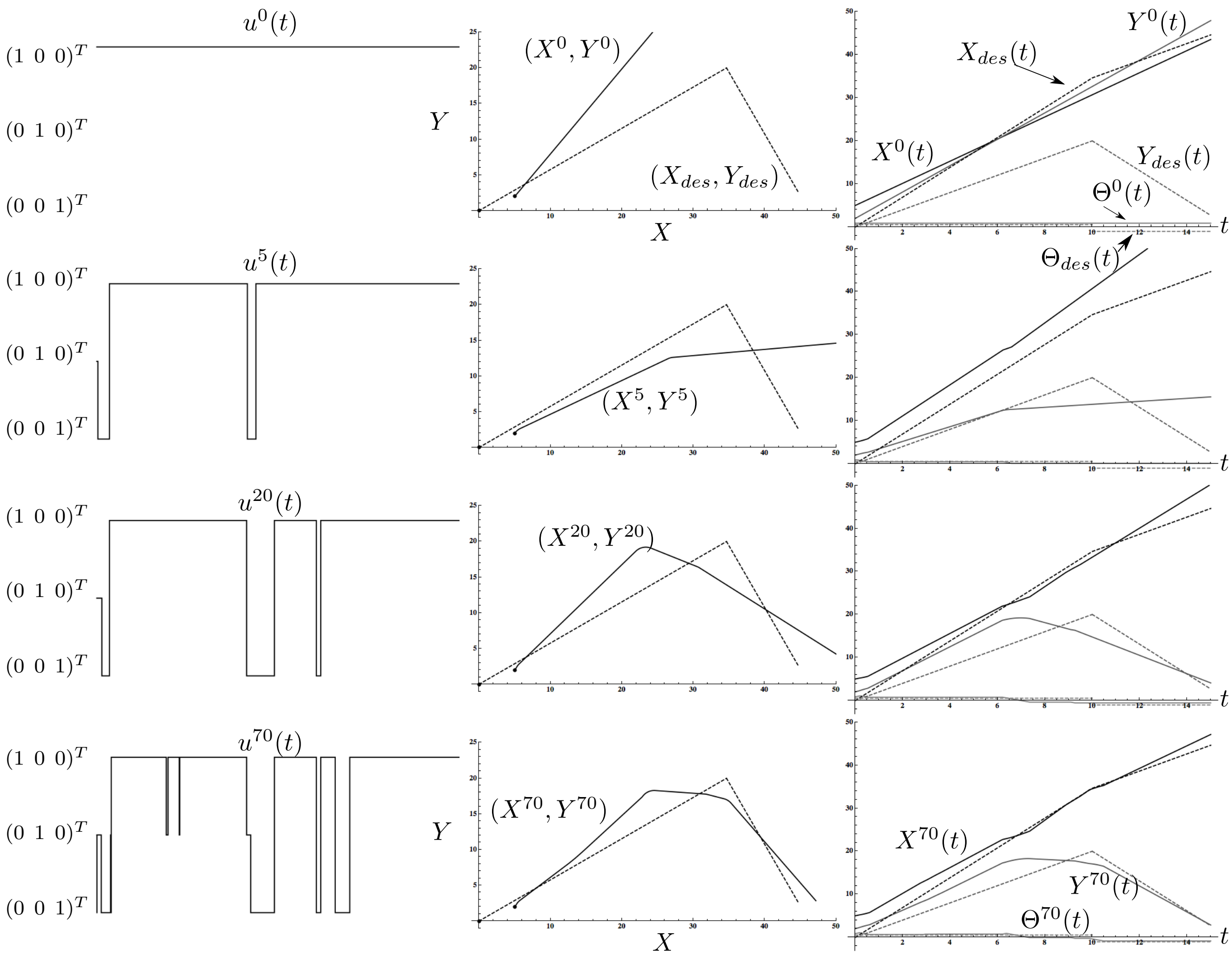


Fig. 6. Results for iterations 0, 5, 20 and 70. The calculated switching control is shown left, the $(X, Y)$ trajectory is shown middle, and the state $x = (X, Y, \Theta)$ versus time is shown right. The desired trajectories are shown with dashed lines.