Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

A Provably Convergent Algorithm for Transition-Time Optimization in Switched Systems

Henrik Axelsson, Yorai Wardi, Magnus Egerstedt, and Erik Verriest

Abstract— **This paper concerns a mode-sequencing and switching-time optimization problem defined on autonomous switched-mode hybrid dynamical systems. The design parameter consists of two elements: (i) the sequence of dynamic-response functions associated with the modes, and (ii) the duration of each mode. The sequencing element is a discrete parameter which may render the problem of computing the optimal schedule exponentially complex. Therefore we are not seeking a global minimum, but rather a local solution in a suitable sense. To this end we endow the parameter space with a local continuous structure which allows us to apply gradient-descent techniques. With this structure, the problem is cast in the form of a nonlinear-programming problem defined on a sequence of nested Euclidean spaces with increasing dimensions. We characterize suboptimality in an appropriate sense, define a corresponding convergence criterion, and devise a provably-convergent optimization algorithm.**

Keywords. Hybrid Systems, Switching Modes, Optimal Control, Gradient Descent, Numerical Algorithms

I. INTRODUCTION

Consider the dynamical system defined by the following equation,

$$\dot{x} \in \{f_\alpha(x)\}_{\alpha\in A} \tag{1}$$

where $x \in \mathbb{R}^n$, A is a finite set, and for every $\alpha \in A$ $f_\alpha : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function. Let $x_0 := x(0)$ be the initial condition, and suppose that the system evolves over a given time-horizon $[0,T]$ for a fixed final time T. The functions f_α represent different modes of the system and hence are called *modal functions*, and the system is labelled a *switched-mode system.* Furthermore, assume that there is a finite number of switchings between the modes in $[0,T]$.

Let $L : \mathbb{R}^n \to \mathbb{R}$ be a cost function defined on the state of the system, $x(t)$, and let J be the aggregate cost functional, defined by

$$J = \int_0^T L(x)dt. \tag{2}$$

Let us view the scheduling of the modal functions in the Right-Hand Side (RHS) of (1) as a control variable. A particular schedule has two elements: the sequencing of the modal functions, and the switching times between successive modes. Suppose there are N switching times, denoted by τ_i, $i = 1,\ldots,N$, in an increasing order. We extend the notation to define $\tau_0 := 0$ and $\tau_{N+1} := T$. For all $i = 1,\ldots,N+1$, let us denote the modal function in the interval $[\tau_{i-1},\tau_i)$ by $f_{\alpha(i)}$ for some $\alpha(i) \in A$. Then the sequence of modal functions is $\{f_{\alpha(i)}\}_{i=1}^{N+1}$, and we will refer to it by the corresponding index-sequence σ, defined by $\sigma := \{\alpha(i)\}_{i=1}^{N+1}$, henceforth labelled the *modal sequence.* The switching-times vector is $(\tau_1,\ldots,\tau_N)^T \in \mathbb{R}^N$, and it will be denoted by $\bar{\tau}$. The control parameter is $(\sigma,\bar{\tau})$. Note that N, the length of the modal sequence, is not fixed, but rather it is defined by σ.

Consider the problem of minimizing the functional J as a function of the control parameter $(\sigma,\bar{\tau})$. Of the two variables σ and $\bar{\tau}$, the former is more problematic, since it is an integer variable that gives the problem an exponential complexity. On the other hand, for a fixed σ, optimizing J with respect to $\bar{\tau}$ is a continuous-parameter problem that can be solved by nonlinear-programming techniques.

More broadly, optimal control for switched-mode hybrid systems has been investigated in the past few years in the more general setting, where the state equation has an external control input, u. Ref. [1] defined a general framework for optimal control, and [2], [3] developed variants of the maximum principle. Refs. [4], [5], [6], [7] considered the special case of piecewise-linear or affine systems, and [7], [2], [8], [9], [10] addressed the general case of nonlinear systems. In particular, Refs. [8], [9], [10] focused on autonomous nonlinear systems, where the input $u(t)$ is absent. Likewise, this paper concerns only autonomous systems whose state equation is (1).

Ref. [10] derived a costate-based formula for the gradient $\nabla J(\bar{\tau})$, and extended it to obtain sensitivity (derivative) information about inserting new modes to a given schedule. This suggests a natural way to deploy gradient-descent algorithms to problems where σ is part of the variable parameter. That observation constitutes the starting point of the present paper, which aims at analyzing convergence of a suitable algorithm.

The algorithm that we later analyze is basically a nonlinear-programming algorithm that modifies the sequencing parameter by adding new modes to it at each iteration. Therefore, it is not defined over a single parameter space, but rather over a nested sequence of Euclidean spaces with increasing dimensions. The standard theory of nonlinear programming is not quite suitable to handle such problems, and therefore we first have to define a notion of local optimality and a suitable concept of convergence, and only then are we in a position to analyze convergence properties of our algorithm.[1]

The rest of the paper is organized as follows. Section II recalls prior results and sets the stage for the analysis that follows. Section III describes the algorithm, and Section IV analyzes its convergence. Finally, Section V concludes the

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 U.S.A., {henrik,ywardi,magnus,verriest}@ece.gatech.edu

[1] The algorithm has been presented in [11] without proofs.

paper.

II. Background

Let us fix a modal sequence σ having $N+1$ modal functions and let $\bar{\tau} := (\tau_1, \ldots, \tau_N)^T \in \mathbb{R}^N$ denote the switching-time vector. Consider the problem of minimizing J as a function of $\bar{\tau}$, henceforth denoted by P_σ. To simplify the notation we will use the term f_i for $f_{\alpha(i)}$, so that the state equation becomes

$$\dot{x} = f_i(x), \quad t \in [\tau_{i-1}, \tau_i) \quad i = 1, \ldots, N+1. \tag{3}$$

We make the following assumption regarding the modal functions.

Assumption 2.1. (i). The functions f_α, $\alpha \in A$, and L, are twice continuously differentiable on $\mathbb{R}^n$. (ii). There exists a constant $K > 0$ such that, for every $x \in \mathbb{R}^n$, and for every $\alpha \in A$,

$$||f_\alpha(x)|| \ \leq \ K(||x|| + 1). \tag{4}$$

This assumption guarantees that, with the given initial condition x_0, the differential equation (3) has a unique solution $x(t)$ on the interval $[0, T]$, which is confined to a bounded set in $\mathbb{R}^n$ that does not depend on the values of the switching times (τ_i), their number (N), or the order of the switching functions $(f_i(x))$. Moreover, that assumption guarantees that J is continuously differentiable in the switching times. Let us define the function $f(x,t)$ by the right-hand side of (3), namely,

$$f(x,t) = f_i(x), \ t \in [\tau_{i-1}, \tau_i) \ i = 1, \ldots, N+1. \tag{5}$$

Then the state trajectory $x(t)$ is continuous in t and we define the notation $x_i := x(\tau_i)$. Next, we define the costate $p(t) \in \mathbb{R}^n$ by the following differential equation,

$$\dot{p}(t) \ = \ -\left(\frac{\partial f}{\partial x}(x,t)\right)^T p(t) \ - \ \left(\frac{\partial L}{\partial x}(x)\right)^T, \tag{6}$$

with the boundary condition $p(T) = 0$. Then the costate trajectory $p(t)$ is continuous in t, and we define the notation $p_i := p(\tau_i)$. The partial derivative $\partial J / \partial \tau_i$ is continuous in $\bar{\tau}$ (since σ is fixed) and is expressed by (see [10]):

$$\frac{\partial J}{\partial \tau_i} \ = \ p_i^T \left(f_i(x_i) - f_{i+1}(x_i)\right). \tag{7}$$

Consider now the problem P_σ. By definition, it involves constrains related to the particular sequence σ, and these constraints can be expressed by the following inequalities, $0 \ = \ \tau_0 \ \leq \ \tau_1 \ \ldots \ \leq \tau_N \ \leq \ \tau_{N+1} \ = \ T$. It is natural to solve P_σ by a constrained gradient-descent algorithm. For instance, [11] used a gradient-projection algorithm that reduces the value of J at each iteration while maintaining feasibility. By "solving the problem P_σ" we mean computing a point satisfying the Kuhn-Tucker first-order optimality condition.

Now suppose that the problem P_σ has been solved to the extent of computing a Kuhn-Tucker point $\bar{\tau}$. It may be possible, of course, to further reduce the value of J by altering the sequence σ. An incremental approach, proposed in [10], is based on the following result. Let g be a modal function, namely $g = f_\alpha$ for some $\alpha \in A$, and fix $\tau \in [0, T]$. Consider inserting the function g at the time τ for the duration of λ seconds, where $\lambda > 0$. By this insertion we are introducing two new switching points, one at $\tau - \lambda/2$ and the other at $\tau + \lambda/2$, and the modal function g between them.[2] Let us denote by $J_{g,\tau}(\lambda)$ the effect of λ on the cost functional J, where we note the dependence on the modal function g and the insertion time τ. Assume that $\tau \in (\tau_i, \tau_{i+1})$ for some $i = 1, \ldots, N$. Then the following formula characterizes the one sided derivative of J at $\lambda = 0$:

$$\frac{dJ_{g,\tau}(0)}{d\lambda^+} \ = \ p(\tau)^T \Big(g(x(\tau)) - f_{i+1}(x(\tau))\Big). \tag{8}$$

Moreover, $\frac{dJ_{g,\tau}(0)}{d\lambda^+}$ is continuous in τ throughout the interval $(0, T)$ at the Kuhn-Tucker points $\bar{\tau}$ for P_σ (see [10]). Eq. (8) suggests the following algorithm for minimizing J.

Algorithm 2.1.

Given: A modal sequence σ having N switching points.

Step 1. Use a feasible gradient-descent algorithm to compute a Kuhn-Tucker point $\bar{\tau}_N$ for P_σ.

Step 2. Compute the number Θ_N defined by

$$\Theta_N := \min\left\{ \frac{dJ_{g,\tau}(0)}{d\lambda^+} | g = f_\alpha, \alpha \in A; \tau \in [0, T] \right\}. \tag{9}$$

Step 3. If $\Theta_N = 0$ then stop and exit. If $\Theta_N < 0$ then, with the pair (g, τ) comprising the *argmin* in (9), append to σ two switching points at the time τ with the modal function g between them, and goto Step 1.

Note that the condition $\Theta_N = 0$ defines a stopping rule in Step 3. This motivates us to define a parameter $(\sigma, \bar{\tau})$ as a local minimum of J if $\bar{\tau}$ is a Kuhn-Tucker point for P_σ and $\Theta_N = 0$. If the algorithm stops after a finite number of iterations then it has computed such a local minimum. Otherwise, it computes an infinite sequence of iteration points in Step 1. Let us denote by $\bar{\tau}_{N(m)}$ the value of $\bar{\tau}$ computed at the *mth* iteration of Step 1 where $N(m)$ denotes the dimension of the vector $\bar{\tau}_{N(m)}$. Suppose that the algorithm does not stop after a finite number of iterations, so that it computes a sequence $\{\bar{\tau}_{N(m)}\}_{m=1}^\infty$. In the theory of nonlinear programming, an algorithm often is said to be convergent if every accumulation point of a sequence it computes satisfies an appropriate optimality condition (see [12]). In our case, the concept of *convergent algorithm* has to be different, since the sequence $\{\bar{\tau}_{N(m)}\}_{m=1}^\infty$ cannot converge or diverge: its elements have different dimensions and hence are in different spaces. However, and following the notion of convergent algorithms proposed in [13], we say that the algorithm is convergent if every sequence $\{\bar{\tau}_{N(m)}\}_{m=1}^\infty$ it computes has the property that $\lim_{m\to\infty} \Theta_{N(m)} \ = \ 0$. Whether Algorithm 2.1 is convergent in this sense or not depends on the specific details of the algorithm used in Step 1. In particular, we next will show that the following property guarantees convergence.

Definition 2.1: Algorithm 2.1 has *sufficient descent* if for every $\varepsilon > 0$ there exists $\eta > 0$ such that, for every $m = 1, 2, \ldots$, if $|\Theta_{N(m)}| > \varepsilon$ then

$$J(\bar{\tau}_{N(m+1)}) - J(\bar{\tau}_{N(m)}) \ \leq \ -\eta. \tag{10}$$

Proposition 2.1: If Algorithm 2.1 has sufficient descent then it is convergent.

Proof: By Assumption 2.1 the state trajectory $x(t)$, $t \in [0, T]$, is confined to a compact set in $\mathbb{R}^n$. Consequently, and

[2] If $\tau \in \{0, T\}$ then the insertion is made in the intervals $[0, \lambda]$ and $[T - \lambda, T]$, respectively.

by the assumption that A is finite, (2), and the fact that L is continuously differentiable, $|J(\bar{\tau}_{N(m)})|$ is bounded from above by a constant that is independent of m. If the algorithm has sufficient descent and it computes an infinite sequence of iteration points, then (10) immediately follows by the fact that Step 1 deploys a gradient-descent technique. ■

The next section describes an algorithm for Step 1 of Algorithm 2.1 which guarantees the sufficient descent property.

III. Algorithm with Sufficient Descent

Recall that the algorithm computes a vector $\bar{\tau}_{N(m)} \in \mathbb{R}^{N(m)}$ at its *mth* iteration in Step 1. This computation is done by a feasible gradient-descent algorithm (henceforth labelled the *inner algorithm*), and therefore the sequence $\{J(\bar{\tau}_{N(m)}\}$ is monotone non-increasing. To guarantee the property of sufficient descent, we specify the first step of the inner algorithm, while subsequent steps need not be specified as long as they give a descent in J. The first step of the inner algorithm uses the Armijo step-size (described below) along a descent curve. The Armijo step-size is typically deployed in a descent direction (see [14], [12]), but we were unable to find such a direction guaranteeing sufficient descent. Therefore, we have adopted a piecewise-linear curve which will endow the algorithm with the sufficient descent property. This curve, and the first step of the inner algorithm, will next be described.

Consider the $m+1$ iteration of Algorithm 2.1, and let us denote by $\bar{t}(0) := (t_1(0), \ldots, t_{N(m+1)}(0))^T$ the point with which the algorithm enters Step 1 from Step 3 of its *mth* iteration. In that iteration, the algorithm inserted a modal function g between two identical switching times. Thus, the point $\bar{t}(0) = (t_1(0), t_2(0), \ldots, t_{N(m+1)}(0))^T$ has the following features:

1) There exists $i \in \{1, \ldots, N(m+1)\}$ such that the last insertion times at Step 3 were $t_{i-1}(0) = t_i(0)$, and the function inserted between them was $g = f_i$.
2) If $t_{i-2}(0) < t_{i-1}(0)$ and $t_i(0) < t_{i+1}(0)$ then the modal function f_{i-1} and f_{i+1} are identical, namely $f_{i-1} = f_{i+1}$, since the function f_i was inserted during the course of the mode defined by that function.

Suppose for a moment that the inserted switching points $t_{i-1}(0) = t_i(0)$ were not equal to any one of the existing switching times of the previous iteration. Then, by (7), (8), and (9), and by the fact that $f_{i-1} = f_{i+1}$, it follows that

$$\frac{\partial J}{\partial \tau_i}(\bar{t}(0)) \;=\; -\frac{\partial J}{\partial \tau_{i-1}}(\bar{t}(0)) \;=\; \Theta_{N(m)}, \tag{11}$$

whereby the term "$\partial J/\partial \tau_i$" we mean the partial derivative with respect to the *ith* variable, i.e., $t_i(0)$ in our case. On the other hand, if that new insertion time $t_i(0)$ is identical to a group of one or more of the "old" insertion times, then we have that $t_{i-1}(0) = t_{i-2}(0)$ or $t_i(0) = t_{i+1}(0)$. In this case, define the integers $k_{i-1}(0)$ and $n_i(0)$ as follows, $k_{i-1}(0) := \min\{j \leq i-1 : t_j(0) = t_{i-1}(0)\}$, and $n_i(0) := \max\{j \geq i : t_j(0) = t_i(0)\}$. Then we have the following extension of (11) (see [10] for a proof):

$$\sum_{j=i}^{n_i(0)} \frac{\partial J}{\partial \tau_j}(\bar{t}(0)) \;=\; -\sum_{j=k_{i-1}(0)}^{i-1} \frac{\partial J}{\partial \tau_j}(\bar{t}(0)) \;=\; \Theta_{N(m)}. \tag{12}$$

We next define the curve along which the inner algorithm applies the Armijo stepsize, and denote it by $\{\bar{t}(\lambda)\}_{\lambda \geq 0}$. It is defined by reducing $t_{i-1}(0)$ to 0 and increasing $t_i(0)$ to T at the rate of -Θ_N. If $t_{i-1}(0)$ ($t_i(0)$, resp.) "bumps" into other switching times on the way, then it "drags" them along with it so that the order of the modal sequence is maintained. Such a "bump" causes a change in the direction of $\bar{t}(\lambda)$, and hence this curve consists of successive linear segments.

To formalize matters, we denote the break points of the curve $\{\bar{t}(\lambda)\}_{\lambda \geq 0}$, namely the points at which the curve changes directions, by λ_ν, $\nu = 1, 2, \ldots$, and we define, as a matter of convention, $\lambda_0 = 0$, while recalling that $\bar{t}(0)$ is the starting point of the curve. We next define the curve in a recursive manner on the segments $[\lambda_\nu, \lambda_{\nu+1}]$. Recall that the starting point is $\bar{t}(0) = \bar{t}(\lambda_0)$. Given $\nu \in \{0, 1, \ldots\}$, suppose that $\bar{t}(\lambda)$ has been defined for all $\lambda \in [0, \lambda_\nu]$; we now define the next segment of the curve, namely the end-point $\lambda_{\nu+1}$ and the curve $\bar{t}(\lambda)$ for all $\lambda \in [\lambda_\nu, \lambda_{\nu+1}]$. First, define the integers $k_{i-1}(\lambda_\nu)$ and $n_i(\lambda_\nu)$ by

$$k_{i-1}(\lambda_\nu) \;:=\; \min\{j \leq i-1 \;:\; t_j(\lambda_\nu) = t_{i-1}(\lambda_\nu)\}, \tag{13}$$

$$n_i(\lambda_\nu) \;:=\; \max\{j \geq i \;:\; t_j(\lambda_\nu) = t_i(\lambda_\nu)\}, \tag{14}$$

furthermore, for the sake of notation, define $t_{-1}(\lambda_\nu) := -\infty$ and $t_{N(m+1)+2}(\lambda_\nu) := \infty$. Next, define the direction $\bar{h}(\lambda_\nu) := (h_1(\lambda_\nu), \ldots, h_{N(m+1)}(\lambda_\nu))^T$ by

$$h_j(\lambda_\nu) = \begin{cases} \Theta_{N(m)}, & \forall\ j \in \{k_{i-1}(\lambda_\nu), \ldots, i-1\}, \\ & \text{if } t_{i-1}(\lambda_\nu) > 0 \\ 0, & \forall\ j \in \{k_{i-1}(\lambda_\nu), \ldots, i-1\}, \\ & \text{if } t_{i-1}(\lambda_\nu) = 0 \\ -\Theta_{N(m)}, & \text{for all } j \in \{i, \ldots, n_i(\lambda_\nu)\}, \\ & \text{if } t_i(\lambda_\nu) < T \\ 0, & \forall\ j \in \{i, \ldots, n_i(\lambda_\nu)\}, \\ & \text{if } t_i(\lambda_\nu) = T \\ 0, & \forall \text{ other } j \in \{1, \ldots, N(m+1)\}. \end{cases} \tag{15}$$

Finally, define $\lambda_{\nu+1}$ as follows for the case where $t_{i-1}(\lambda_\nu) > 0$ or $t_{N(m+1)}(\lambda_\nu) < T$:

$$\lambda_{\nu+1} \;= \min\big\{\lambda > \lambda_\nu : \text{either } \; t_{k_{i-1}(\lambda_\nu)}(\lambda_\nu) + (\lambda - \lambda_\nu) h_{k_{i-1}(\lambda_\nu)}(\lambda_\nu) = t_{k_{i-1}(\lambda_\nu)-1}(\lambda_\nu), \;\text{ or }\; t_{n_i(\lambda_\nu)}(\lambda_\nu) + (\lambda - \lambda_\nu) h_{n_i(\lambda_\nu)}(\lambda_\nu) = t_{n_i(\lambda_\nu)+1}(\lambda_\nu)\big\} \tag{16}$$

(for the case where $t_{i-1}(\lambda_\nu) = 0$ and $t_i(\lambda_\nu) = T$, there is no need to define $\lambda_{\nu+1}$). At last, we define $\bar{t}(\lambda)$ for $\lambda \in [\lambda_\nu, \lambda_{\nu+1}]$ by

$$\bar{t}(\lambda) \;= \bar{t}(\lambda_\nu) + (\lambda - \lambda_\nu)\bar{h}(\lambda_\nu). \tag{17}$$

Recall that the above curve is used to compute the first iteration of the inner algorithm, whose starting point is $\bar{t}(0)$. Let us denote the resulting computed vector by $\bar{t}_{next}$. Now this point, $\bar{t}_{next}$, lies on the curve, namely $\bar{t}_{next} = \bar{t}(\lambda)$ for some $\lambda > 0$, and we denote that value of λ by λ_{next}, so that $\bar{t}_{next} = \bar{t}(\lambda_{next})$. The point λ_{next} is computed by the Armijo stepsize rule along the curve $\{\bar{t}(\lambda)\}$. That is (see [12]), given $\alpha \in (0,1)$ and given a monotone-decreasing sequence $\lambda(\ell)$ such that $\lambda(\ell) \to 0$ as $\ell \to \infty$, λ_{next} is defined

$$\lambda_{next} \;:=\; \max\big\{\lambda(\ell) \;:\; \ell = 0, 1, \ldots, \text{ such that } J(\bar{t}(\lambda(\ell))) - J(\bar{t}(0)) \leq -\alpha\lambda(\ell)\Theta^2_{N(m)}\big\}. \tag{18}$$

It follows that the inner algorithm has the following form.

Algorithm 3.1:

Algorithm 2.1 enters Step 1, from Step 3 of its previous iteration, with a point $\bar{t}(0)$.

Step 1. Compute λ_{next} according to (18), and set

$$\bar{t}_{next} := \bar{t}(\lambda_{next}). \qquad (19)$$

Step 2. Stating at $\bar{t}_{next}$, use any feasible gradient descent algorithm to compute $\bar{\tau}_{N(m+1)}$.

The next section carries out and analysis of Algorithm 3.1.

IV. Convergence Analysis

Gradient-descent algorithms with Armijo stepsizes generally have a sufficient-descent property as long as the function they attempt to minimize is continuously differentiable. However, in the setting of our problem and Algorithm 3.1, there is no continuity of the gradient. As a matter of fact, each time the curve $\{\bar{t}(\lambda)\}_{\lambda\geq 0}$ has a break point λ_ν, the gradient of J is discontinuous. Another problem is that Algorithm 3.1 acts on spaces of increasing dimensions each time it enters Step 1 of Algorithm 2.1, which makes it harder to guarantee that an inequality like the one in Eq. (10) is satisfied. For these reasons the convergence analysis, whose centerpiece is a proof of sufficient descent, is quite complicated. That proof is based on the fact that the starting point of Algorithm 3.1 is a Kuhn-Tucker point for P_σ. Applied to any other point the algorithm will not guarantee sufficient descent, and this is the reason for the specification in Step 1 of the algorithm, made only at its starting point, while requiring only a gradient descent in the later iterations.

The rest of this section carries out the convergence analysis. First, a number of preliminary results are presented without proofs.

Lemma 4.1. There exists a constant $K_1 > 0$ such that, for every modal sequence σ having any number of switching times N, and for every $t \in [0,T]$, $||x(t)|| \leq K_1$ and $||p(t)|| \leq K_1$.

Lemma 4.2. There exists a constant $K_2 > 0$ such that, for every N, $|\Theta_N| \leq K_2$, where Θ_N is defined in Step 2 of Algorithm 2.1.

Lemma 4.3. There exists a constant $K_3 > 0$ such that, for every modal sequence σ having any number of switching times N, and for every $t_1 \in [0,T]$ and $t_2 \in [t_1,T]$, $||x(t_2) - x(t_1)|| \leq K_3(t_2 - t_1)$ and $||p(t_2) - p(t_1)|| \leq K_3(t_2 - t_1)$.

Next, consider two modal sequences $\bar{\sigma}$ and $\tilde{\sigma}$ having possibly different numbers of switching times, $\bar{N}$ and $\tilde{N}$, and let $\bar{f}(x,t)$ and $\tilde{f}(x,t)$ be the associated dynamic response functions as defined by (5). For all $t \in [0,T]$, define $\bar{\alpha}(t) := \{\alpha \in A : \bar{f}(x,t) = f_\alpha(x)\}$, and similarly define $\tilde{\alpha}(t) := \{\alpha \in A : \tilde{f}(x,t) = f_\alpha(x)\}$. Denote by $\bar{x}(t)$ and $\bar{p}(t)$ ($\tilde{x}(t)$ and $\tilde{p}(t)$, resp.) the state trajectory and costate trajectory associated with $\bar{\sigma}$ ($\tilde{\sigma}$, resp.) via (3), (5) and (6). Let $I \subset [0,T]$ be defined by $I = \{t \in [0,T] : \bar{\alpha}(t) \neq \tilde{\alpha}(t)\}$, and denote by $|I|$ the Lebesgue measure of I.

Lemma 4.4: There exists a constant $K_4 > 0$ such that, for every pair of modal sequences $\bar{\sigma}$ and $\tilde{\sigma}$; for every $t \in [0,T]$, $||\bar{x}(t) - \tilde{x}(t)|| < K_4|I|$ and $||\bar{p}(t) - \tilde{p}(t)|| < K_4|I|$.

The next lemma concerns variational bounds on the gradient of J along the curve $\{\bar{t}(\lambda)\}$. First, we establish some notation. For every $\lambda \geq 0$, we define $k_{i-1}(\lambda)$ and $n_i(\lambda)$ by

$$k_{i-1}(\lambda) := \min\{j \leq i-1 \ : \ t_j(\lambda) = t_{i-1}(\lambda)\}, \qquad (20)$$

$$n_i(\lambda) := \max\{j \geq i \ : \ t_j(\lambda) = t_i(\lambda)\}. \qquad (21)$$

Lemma 4.5: There exists a constant $K > 0$ such that, for every point $\bar{t}(0)$ with which Algorithm 2.1 enters Step 1 from Step 3 of its previous iteration; for every $\nu = 0,1,2,\ldots$; and for every $\lambda \in [\lambda_\nu, \lambda_{\nu+1}]$,

$$| < \nabla J(\bar{t}(\lambda)) - \nabla J(\bar{t}(0)) \ , \ \bar{h}(\lambda_\nu) > | \ \leq \ K\lambda. \qquad (22)$$

Moreover, the constant K is independent of $N(m)$.

Proof. Consider a point $\bar{t}(0)$ as in the statement of the lemma. Fix $\nu \geq 0$, and fix $\lambda \in [\lambda_\nu, \lambda_{\nu+1})$. Assume, without loss of generality, that $t_{i-1}(\lambda_{\nu+1}) > 0$ and $t_i(\lambda_{\nu+1}) < T$. Consider first the term $< \nabla J(\bar{t}(\lambda)), \bar{h}(\lambda_\nu) >$. For ease of notation, we define $\widetilde{\Theta} = \Theta_{N(m)}$. By (15),

$$< \nabla J(\bar{t}(\lambda)), \bar{h}(\lambda_\nu) > = \widetilde{\Theta}\textstyle\sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} \frac{dJ}{d\tau_j}(\bar{t}(\lambda)) - \widetilde{\Theta}\sum_{j=i}^{n_i(\lambda_\nu)} \frac{dJ}{d\tau_j}(\bar{t}(\lambda)). \qquad (23)$$

Let us denote by $\{x_\lambda(t)\}$ and $\{p_\lambda(t)\}$ the state trajectory and costate trajectory associated with the point $\bar{t}(\lambda)$ via (3) and (6), respectively. Then, by (7) and (23),

$$< \nabla J(\bar{t}(\lambda)), \bar{h}(\lambda_\nu) > = \widetilde{\Theta}\textstyle\sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} p_\lambda(t_j(\lambda))^T \times \{f_j(x_\lambda(t_j(\lambda))) - f_{j+1}(x_\lambda(t_j(\lambda)))\} - \widetilde{\Theta}\displaystyle\sum_{j=i}^{n_i(\lambda_\nu)} p_\lambda(t_j(\lambda))^T \times\{f_j(x_\lambda(t_j(\lambda))) - f_{j+1}(x_\lambda(t_j(\lambda)))\}. \qquad (24)$$

By the definition of $k_{i-1}(\lambda)$ (20), we have that $t_j(\lambda) = t_{i-1}(\lambda)$ for all $j \in \{k_{i-1}(\lambda), \ldots, i-1\}$, and moreover, by (16), $k_{i-1}(\lambda) = k_{i-1}(\lambda_\nu)$; hence, $t_j(\lambda) = t_{i-1}(\lambda)$ for all $j \in \{k_{i-1}(\lambda_\nu), \ldots, i-1\}$. In a similar way (by (21) and (16)), $t_j(\lambda) = t_i(\lambda)$ for all $j \in \{i, \ldots, n_i(\lambda_\nu)\}$. Consequently, (24) implies that,

$$< \nabla J(\bar{t}(\lambda)), \bar{h}(\lambda_\nu) > = \widetilde{\Theta} p_\lambda(t_{i-1}(\lambda))^T \times \Big(f_{k_{i-1}(\lambda_\nu)}\big(x_\lambda(t_{i-1}(\lambda))\big) - f_i\big(x_\lambda(t_{i-1}(\lambda))\big)\Big) - \widetilde{\Theta} p_\lambda(t_i(\lambda))^T \Big(f_i\big(x_\lambda(t_i(\lambda))\big) - f_{n_i(\lambda_\nu)+1}\big(x_\lambda(t_i(\lambda))\big)\Big). \qquad (25)$$

Next, consider the second term in the left-hand side of (22), namely, $< \nabla J(\bar{t}(0)), \bar{h}(\lambda_\nu) >$. By (15), we obtain, that

$$< \nabla J(\bar{t}(0)), \bar{h}(\lambda_\nu) > = \widetilde{\Theta} \sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} \frac{dJ}{d\tau_j}(\bar{t}(0)) - \widetilde{\Theta} \sum_{j=i}^{n_i(\lambda_\nu)} \frac{dJ}{d\tau_j}(\bar{t}(0)). \qquad (26)$$

If we denote by $\{x_0(t)\}$ and $\{p_0(t)\}$ the state- and costate-trajectories associated with the point $\bar{t}(0)$ via (3) and (6), respectively. Then, with some algebra, and (26), we get that

$$< \nabla J(\bar{t}(0)), \bar{h}(\lambda_\nu) > = \widetilde{\Theta}\{p_0(t_{k_{i-1}(\lambda_\nu)}(0))^T f_{k_{i-1}(\lambda_\nu)}\big(x_0(t_{k_{i-1}(\lambda_\nu)}(0))\big) - p_0(t_i(0))^T f_i\big(x_0(t_i(0))\big)\} + \widetilde{\Theta} \sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} f_{j+1}\big(x_0(t_{j+1}(0))\big) \times$$

$$\{p_0(t_{j+1}(0))^T - p_0(t_j(0))^T\} + \widetilde{\Theta} \sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} p_0(t_j(0))^T \times$$
$$\{f_{j+1}(x_0(t_{j+1}(0))) - f_{j+1}\big(x_0(t_j(0))\big)\} - \widetilde{\Theta}\Big(p_0(t_{i-1}(0))^T \times$$
$$f_i(x_0(t_{i-1}(0))) - p_0(t_{n_i(\lambda_\nu)}(0))^T f_{n_i(\lambda_\nu)+1}\big(x_0(t_{n_i(\lambda_\nu)}(0))\big)\Big)$$
$$-\widetilde{\Theta} \sum_{j=i}^{n_i(\lambda\nu)} \{p_0(t_j(0))^T - p_0(t_{j-1}(0))^T\} f_j\big(x_0(t_{j-1}(0))\big)$$
$$-\widetilde{\Theta} \sum_{j=i}^{n_i(\lambda_\nu)} p_0(t_j(0))^T \{f_j\big(x_0(t_j(0))\big) - f_j\big(x_0(t_{j-1}(0))\big)\}. \quad (27)$$

We next derive upper bounds on the various sum terms in the right-hand side of (27). Consider the first sum term. By Lemma 4.1, Assumption 2.1, and the fact that the modal set A is finite, there exists a constant $K_5 > 0$ such that, for all $j \in \{0,\ldots,N(m+1)\}$, $||f_{j+1}\big(x_0(t_{j+1}(0))\big)|| \le K_5$. By Lemma 4.3, for all $j \in \{0,\ldots,N(m+1)\}$, $||p_0(t_{j+1}(0)) - p_0(t_j(0))|| \le K_4(t_{j+1}(0) - t_j(0))$. By Lemma 4.2, $|\widetilde{\Theta}|| \le K_2$. Consequently, we have that

$$||\widetilde{\Theta}\textstyle\sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} f_{j+1}\big(x_0(t_{j+1}(0))\big)\{p_0(t_{j+1}(0))^T$$
$$-p_0(t_j(0))^T\}|| \ \le\ K_2K_4K_5(t_i(0) - t_{k_{i-1}(\lambda_\nu)}(0)). \qquad (28)$$

By definition $t_i(0) = t_{i-1}(0)$ since this is the double switching times appended at Step 3 of the previous iteration of Algorithm 2.1. We have seen (by (16)) that $t_{k_{i-1}(\lambda_\nu)}(0) \ge t_{k_{i-1}(\lambda_\nu)}(\lambda_\nu)$, and by (20), $t_{k_{i-1}(\lambda_\nu)}(\lambda_\nu) = t_{i-1}(\lambda_\nu)$. Certainly $t_{i-1}(0) \ge t_{k_{i-1}(\lambda_\nu)}(0)$ since $i-1 \ge k_{i-1}(\lambda_\nu)$. Therefore, we have that

$$0 \le t_i(0) - t_{k_{i-1}(\lambda_\nu)}(0) \le t_{i-1}(0) - t_{k_{i-1}(\lambda_\nu)}(\lambda_\nu)$$
$$= t_{i-1}(0) - t_{i-1}(\lambda_\nu). \qquad (29)$$

By (15)-(17), $t_{i-1}(0) - t_{i-1}(\lambda_\nu) \le \lambda_\nu|\widetilde{\Theta}|$, and by Lemma 4.2 and the fact that $\lambda_\nu \le \lambda$ (by assumption), we have that $t_{i-1}(0) - t_{i-1}(\lambda_\nu) \le \lambda K_2$. Combining this inequality with (28) and (29) we obtain,$||\widetilde{\Theta}\sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} f_{j+1}\big(x_0(t_{j+1}(0))\big)\{p_0(t_{j+1}(0))^T - p_0(t_j(0))^T\}|| \le K_2^2K_4K_5\lambda$. This inequality provides an upper bound on the first sum term in the right-hand side of (27). In the same way, similar inequalities can be derived for the other three sum terms in the right-hand side of (27). Thus, there exists a constant $K_6 \ge 0$ such that, $||\widetilde{\Theta}\sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} f_{j+1}\big(x_0(t_{j+1}(0))\big)\{p_0(t_{j+1}(0))^T - p_0(t_j(0))^T\}|| + ||\widetilde{\Theta}\sum_{j=k_{i-1}(\lambda_\nu)}^{i-1} p_0(t_j(0))^T\{f_{j+1}\big(x_0(t_{j+1}(0))\big) - f_{j+1}\big(x_0(t_j(0))\big)\}|| + ||\widetilde{\Theta}\sum_{j=i}^{n_i(\lambda\nu)} f_j\big(x_0(t_{j-1}(0))\big)\{p_0(t_j(0))^T - p_0(t_{j-1}(0))^T\}|| + ||\widetilde{\Theta}\sum_{j=i}^{n_i(\lambda_\nu)} p_0(t_j(0))^T \times \{f_j\big(x_0(t_j(0))\big) - f_j\big(x_0(t_{j-1}(0))\big)\} \le \lambda K_6$. We next derive an upper bound on the term $| < \nabla J(\bar{t}(\lambda)) - \nabla J(\bar{t}(0)), \bar{h}(\lambda_\nu) > |$ in the Left-Hand Side (LHS) of (22) by considering the differences between the analogous terms in the RHS of (25) and (27), and using the above derived inequality for the remaining terms of (27). This yields the following inequality,

$$| < \nabla J(\bar{t}(\lambda)) - \nabla J(\bar{t}(0)), \bar{h}(\lambda_\nu) > |$$
$$\le |\widetilde{\Theta}\Big(p_\lambda(t_{i-1}(\lambda))^T f_{k_{i-1}(\lambda_\nu)}\big(x_\lambda(t_{i-1}(\lambda))\big)$$
$$-p_0(t_{k_{i-1}(\lambda_\nu)}(0))^T f_{k_{i-1}(\lambda_\nu)}\big(x_0(t_{k_{i-1}(\lambda_\nu)}(0))\big)\Big)| +$$
$$|\widetilde{\Theta}\{p_\lambda(t_{i-1}(\lambda))^T(f_i\big(x_\lambda(t_{i-1}(\lambda))\big) - p_0(t_i(0))^T f_i\big(x_0(t_i(0))\big)\}|$$
$$+|\widetilde{\Theta}\{p_\lambda(t_i(\lambda))^T f_i\big(x_\lambda(t_i(\lambda))\big)$$
$$-p_0(t_{i-1}(0))^T f_i(x_0(t_{i-1}(0)))\}|$$
$$+|\widetilde{\Theta}\Big(p_\lambda(t_i(\lambda))^T f_{n_i\left(\lambda_\nu\right)+1}(x_\lambda(t_i(\lambda)))$$
$$-p_0(t_{n_i(\lambda_\nu)}(0))^T f_{n_i(\lambda_\nu)+1}\big(x_0(t_{n_i(\lambda_\nu)}(0))\big)\Big)| + \lambda K_6. \qquad (30)$$

Now consider the various terms in the RHS of (30). Regarding the first term, we make the following observations. By (20) and the fact that $k_{i-1}(\lambda) = k_{i-1}(\lambda_\nu)$, we obtain that $t_{i-1}(\lambda) = t_{k_{i-1}(\lambda_\nu)}(\lambda)$, and hence, $p_\lambda(t_{i-1}(\lambda)) = p_\lambda(t_{k_{i-1}(\lambda_\nu)}(\lambda)$. Consequently, and by subtracting and adding $p_\lambda(t_{k_{i-1}(\lambda_\nu)}(0))$, we obtain,

$$p_\lambda(t_{i-1}(\lambda)) - p_0(t_{k_{i-1}(\lambda_\nu)}(0)) = p_\lambda(t_{k_{i-1}(\lambda_\nu)}(\lambda)$$
$$-p_\lambda(t_{k_{i-1}(\lambda_\nu)}(0)) + p_\lambda(t_{k_{i-1}(\lambda_\nu)}(0)) - p_0(t_{k_{i-1}(\lambda_\nu)}(0)). \quad (31)$$

By (15)-(17), $0 \le t_j(0) - t_j(\lambda) \le \lambda|\widetilde{\Theta}|$ for all $j \in \{0,\ldots,i-1\}$. Apply this with $j = k_{i-1}(\lambda_\nu)$, and use Lemma 4.3 and Lemma 4.2 to obtain,

$$||p_\lambda(t_{k_{i-1}(\lambda_\nu)}(\lambda)) - p_\lambda(t_{k_{i-1}(\lambda_\nu)}(0))|| \ \le\ K_3K_2\lambda. \qquad (32)$$

Next, denote by $f_\lambda(x,t)$ the dynamic response function defined by (5) with the switching times given by $\bar{t}(\lambda)$, and denote by $f_0(x,t)$ the dynamic response function defined by (5) with the switching times given by $\bar{t}(0)$. Furthermore, define $\alpha_\lambda(t) := \{\alpha \in A : f_\lambda(x,t) = f_i(x)\}$, define $\alpha_0(t) := \{\alpha \in A : f_0(x,t) = f_i(x)\}$, and define $I := \{t \in [0,T] : \alpha_\lambda(t) \ne \alpha_0(t)\}$. Let $|I|$ denote the Lebesgue measure of I. Then, by (15)-(17) and Lemma 4.2, we have that $|I| \le \lambda K_2$. Consequently, and by Lemma 4.4,

$$||p_\lambda(t_{k_{i-1}(\lambda_\nu)}(0)) - p_0(t_{k_{i-1}(\lambda_\nu)}(0))|| \ \le\ K_4K_2\lambda. \qquad (33)$$

It now follows from (31)-(33) that $||p_\lambda(t_{i-1}(\lambda)) - p_0(t_{k_{i-1}(\lambda_\nu)}(0)|| \ \le\ (K_3+K_4)K_2\lambda$. Defining $K_7 := (K_3 + K_2)K_4$, we obtain that

$$||p_\lambda(t_{i-1}(\lambda)) - p_0(t_{k_{i-1}(\lambda_\nu)}(0)|| \ \le\ K_7\lambda. \qquad (34)$$

By Lemma 4.1, Assumption 2.1, and the fact that A is a finite set, there exists a common Lipschitz constant K' for all the functions f_α, $\alpha \in A$, in a given compact set Γ containing the state trajectories $\{x_\lambda\}$ for all $\lambda > 0$. Then, similarly to (34), there exists a constant $K_8 > 0$ such that,

$$||f_{k_{i-1}(\lambda_\nu)}\big(x_\lambda(t_{i-1}(\lambda))\big) - f_{k_{i-1}(\lambda_\nu)}\big(x_0(t_{k_{i-1}(\lambda_\nu)}(0))\big)||$$
$$\le K_8\lambda. \quad (35)$$

By the boundedness of every term in the RHS of (30), and by (34) and (35), there exists a constant $K_9 > 0$ such that,

$$|\widetilde{\Theta}\Big(p_\lambda(t_{i-1}(\lambda))^T f_{k_{i-1}(\lambda_\nu)}\big(x_\lambda(t_{i-1}(\lambda))\big)$$
$$-p_0(t_{k_{i-1}(\lambda_\nu)}(0))^T f_{k_{i-1}(\lambda_\nu)}\big(x_0(t_{k_{i-1}(\lambda_\nu)}(0))\big)\Big)| \le K_9\lambda. \quad (36)$$

This establishes an upper bound on the first term in the right-hand side of (30). Using similar arguments, it is apparent that the next three terms in the right-hand side of (30) have similar upper bounds, that is, there are positive constants K_{10}, K_{11}, and K_{12} such that,

$$|\widetilde{\Theta}\{p_\lambda(t_{i-1}(\lambda))^T[f_i(x_\lambda(t_{i-1}(\lambda))) \\ -p_0(t_i(0))^T f_i(x_0(t_i(0)))]\}| \leq K_{10}\lambda, \quad (37)$$

$$|\widetilde{\Theta}\{p_\lambda(t_i(\lambda))^T f_i(x_\lambda(t_i(\lambda))) \\ -p_0(t_{i-1}(0))^T f_i(x_0(t_{i-1}(0)))\}| \leq K_{11}\lambda, \quad (38)$$

$$|\widetilde{\Theta}\{p_\lambda(t_i(\lambda))^T f_{n_i(\lambda_\nu)+1}(x_\lambda(t_i(\lambda))) - \\ p_0(t_{n_i(\lambda_\nu)}(0))^T f_{n_i(\lambda_\nu)+1}\big(x_0(t_{n_i(\lambda_\nu)}(0))\big)\}| \leq K_{12}\lambda. \quad (39)$$

Finally, defining $K := K_6 + K_9 + K_{10} + K_{11} + K_{12}$, Eq. (22) follows from (30) and (36)-(39). This completes the proof. ■

We can now establish the property of sufficient descent.

Proposition 4.1. For every $\varepsilon > 0$ there exists $\eta > 0$ such that, if $|\Theta_{N(m)}| \geq \varepsilon$ then

$$J(\bar{t}_{next}) - J(\bar{t}(0)) \ \leq \ -\eta. \quad (40)$$

Proof. Consider a point $\bar{t}(0)$ with which Algorithm 2.1 enters Step 1. According to (9), unless $\Theta_{N(m)} = 0$, the curve $\{\bar{t}(\lambda)\}$ separates the points $t_{i-1}(0)$ and $t_{i+1}(0)$ by reducing the former point and increasing the latter one. By (16), the lower branch of the curve stalls whenever $t_{i-1}(\lambda) = 0$ for some $\lambda \geq 0$, and the upper branch stalls whenever $t_i(\lambda) = T$ for some $\lambda \geq 0$. Since this proposition concerns a local result, we can assume, without loss of generality, that neither branch of the curve stalls.

Fix $\nu = 0, 1, 2, \ldots$, and consider $\lambda \in [\lambda_\nu, \lambda_{\nu+1}]$. By the mean value theorem and (17) there exists $\tilde{\lambda} \in [\lambda_\nu, \lambda]$ such that, $J(\bar{t}(\lambda)) - J(\bar{t}(\lambda_\nu)) = (\lambda - \lambda_\nu) < \nabla J(\bar{t}(\tilde{\lambda})), \bar{h}(\lambda_\nu) >$. By Lemma 4.5 and the fact that $\tilde{\lambda} - \lambda_\nu \leq \lambda - \lambda_\nu$, we obtain that

$$J(\bar{t}(\lambda)) - J(\bar{t}(\lambda_\nu)) \leq (\lambda - \lambda_\nu)(\lambda K + < \nabla J(\bar{t}(0)), \bar{h}(\lambda_\nu) >). \quad (41)$$

By (8)-(9) and the definition of $\bar{t}(0)$ (see Algorithm 2.1), $dJ(\bar{t}(0))/d\tau_j = -\Theta_{N(m)}$ for all $j \in \{k_{i-1}(0), \ldots, i-1\}$ and $dJ(\bar{t}(0))/d\tau_j = \Theta_{N(m)}$ for all $j \in \{i, \ldots, n_i(0)\}$, and hence, and by (15), $< \nabla J(\bar{t}(0)), \bar{h}(\lambda_\nu) >= -\big(n_i(0) - k_{i-1}(0) + 1\big)(\Theta_{N(m)})^2$. Since $n_i(0) - k_{i-1}(0) + 1 \geq 1$, we obtain that

$$< \nabla J(\bar{t}(0)), \bar{h}(\lambda_\nu) > \ \leq \ -\Theta^2_{N(m)}. \quad (42)$$

Thus, by (41) and (42), we have that

$$J(\bar{t}(\lambda)) - J(\bar{t}(\lambda_\nu)) \leq (\lambda - \lambda_\nu)(\lambda K - \Theta^2_{N(m)}). \quad (43)$$

Let us apply (43) to λ_j and λ_{j+1} in lieu of λ_ν and λ, respectively, to obtain,

$$J(\bar{t}(\lambda_{j+1})) - J(\bar{t}(\lambda_j)) \leq (\lambda_{j+1} - \lambda_j)(\lambda_{j+1}K - \Theta^2_{N(m)}). \quad (44)$$

Summing up (44) for $j = 0, \ldots, \nu - 1$, adding (43), noting that $\lambda_{j+1} \leq \lambda_\nu \leq \lambda$, and recalling that $\lambda_0 = 0$, we obtain,

$$J(\bar{t}(\lambda)) - J(\bar{t}(0) \ \leq \ -\lambda\big((\Theta_{N(m)})^2 - \lambda K\big). \quad (45)$$

Fix $\varepsilon > 0$ and suppose that $|\Theta_{N(m)}| > \varepsilon$. Define $\bar{\lambda} := (1 - \alpha)\varepsilon^2/K$. Then, for all $\lambda \in [0, \bar{\lambda}]$, $\Theta^2_{N(m)} - \lambda K \geq \Theta^2_{N(m)} - \bar{\lambda}K = \Theta^2_{N(m)} - (1-\alpha)\varepsilon^2 \geq \Theta^2_{N(m)} - (1-\alpha)\Theta^2_{N(m)} = \alpha\Theta^2_{N(m)}$. Therefore, and by (45), $J(\bar{t}(\lambda)) - J(\bar{t}(0)) \leq -\lambda\alpha(\Theta_{N(m)})^2$. By (18), we conclude that $\lambda_{next} \geq \bar{\lambda}$, and hence, $J(\bar{t}_{next}) - J(\bar{t}(0)) \leq -\bar{\lambda}\alpha(\Theta_{N(m)})^2$. By the definition of $\bar{\lambda}$ and the assumption that $|\Theta_{N(m)}| > \varepsilon$, we have that

$$J(\bar{t}_{next}) - J(\bar{t}(0)) \ \leq \ -\frac{(1-\alpha)\alpha\varepsilon^4}{K}. \quad (46)$$

Defining η to be the right-hand side of (46), (40) follows. ■

This leads us to the main result of the paper.

Theorem 4.1: If Algorithm 2.1 computes a sequence of iteration points $\{\bar{\tau}_{N(m)}\}_{m=1}^{\infty}$, then

$$\lim_{m\to\infty} \Theta_{N(m)} \ = \ 0. \quad (47)$$

Proof: Follows immediately from Proposition 4.1 and Proposition 2.1. ■

We point out that numerical results with the algorithm have been reported on in [11].

V. Conclusions

An algorithm for transition-mode optimization in autonomous hybrid dynamical systems was analyzed. The algorithm considered inserting new modes into the current modal structure by utilizing a variational formula. Based on the derivation of a gradient descent curve, proofs of sufficient descent and convergence of the algorithm were proved.

References

[1] M.S. Branicky, V.S. Borkar, and S.K. Mitter. A Unified Framework for Hybrid Control: Model and Optimal Control Theory. *IEEE Transactions on Automatic Control*, Vol. 43, pp. 31-45, 1998.

[2] M.S. Shaikh and P. Caines. On Trajectory Optimization for Hybrid Systems: Theory and Algorithms for Fixed Schedules. *IEEE Conference on Decision and Control*, Las Vegas, NV, Dec. 2002.

[3] H.J. Sussmann. A Maximum Principle for Hybrid Optimal Control Problems. *Proceedings of the 38th IEEE Conference on Decision and Control*, pp. 425-430, Phoenix, AZ, Dec. 1999.

[4] A. Bemporad, A. Giua, and C. Seatzu. A Master-Slave Algorithm for the Optimal Control of Continuous-Time Switched Affine Systems. *Proceedings of the 41st IEEE Conference on Decision and Control*, pp. 1976-1981, Las Vegas, Nevada, December 2002.

[5] A. Giua, C. Seatzu, and C. Van der Mee. Optimal Control of Switched Autonomous Linear Systems. In *Proceedings of the 40th Conference on Decision and Control*, pp. 1816-1821, Phoenix, Arizona, December 1999.

[6] A. Rantzer and M. Johansson. Piecewise Linear Quadratic Optimal Control. *IEEE Transactions on Automatic Control*, Vol. 54, pp. 629-637, 2000.

[7] P. Riedinger, F. Kratz, C. Iung, and C. Zanne. Linear Quadratic Optimization for Hybrid Systems. *Proceedings of the 38th IEEE Conference on Decision and Control*, pp. 3059-3064, Phoenix, Arizona, pp. 3059-3064, December 1999.

[8] X. Xu and P. Antsaklis. Optimal Control of Switched Autonomous Systems. *IEEE Conference on Decision and Control*, Las Vegas, NV, Dec. 2002.

[9] X. Xu and P.J. Antsaklis. Optimal Control of Switched Systems via Nonlinear Optimization Based on Direct Differentiations of Value Functions. *International Journal of Control*, Vol. 75, pp. 1406-1426, 2002.

[10] M. Egerstedt, Y. Wardi, and H. Axelsson. Transition-Time Optimization for Switched Systems. To appear in *IEEE Transactions on Automatic Control*, 2005.

[11] H. Axelsson, Y. Wardi, and M. Egerstedt, Transition-Time Optimization for Switched Systems. *Proc. IFAC World Congress*, Prague, The Czech Republic, July 2005.

[12] E. Polak. *Optimization Algorithms and Consistent Approximations*. Springer-Verlag, New York, New York, 1997.

[13] E. Polak and Y. Wardi. A Study of Minimizing Sequences. *SIAM Journal on Control and Optimization*, Vol. 22, No. 4, pp. 599-609, 1984.

[14] L. Armijo. Minimization of Functions Having Lipschitz Continuous First-Partial Derivatives. *Pacific Journal of Mathematics*, Vol. 16, pp. 1-3, 1966.