TRANSITION-TIME OPTIMIZATION FOR SWITCHED SYSTEMS

Henrik Axelsson, Yorai Wardi and Magnus Egerstedt

{henrik,ywardi,magnus}@ece.gatech.edu
Electrical and Computer Engineering,
Georgia Institute of Technology, Atlanta, GA 30332, USA

Abstract: This paper proposes an algorithmic framework for optimal mode switches in hybrid dynamical systems. The problem is cast in the setting of optimal control, whose variable parameter consists of the switching times, and whose associated cost criterion is a functional of the state trajectory. The number of switching times (and hence of switching modes) is also a variable which may be unbounded, and therefore the optimization problem is not defined on a single metric space. Rather, it is defined on a sequence of spaces of possibly increasing dimensions. The paper characterizes optimality in terms of sequences of optimality functions and proposes an algorithm that is demonstrably convergent in this context.

Keywords: Hybrid systems, switched dynamical systems, optimal control, optimization theory, gradient-descent algorithms.

1. INTRODUCTION

Switched dynamical systems are often described by differential inclusions of the form

$$\dot{x} \;=\in\; \{f_\alpha(x(t), u(t)\}_{\alpha \in A}, \tag{1}$$

where $x(t) \in R^n$ is the state variable, $u(t) \in R^k$ is the input (control), $t \in [0, T]$ for a given $T > 0$, and $\{f_\alpha : R^{n \times k} \rightarrow R^n\}_{\alpha \in A}$ is a collection of functions parameterized by α belonging to some set A. Let $\{\tau_i\}_{i=1}^N \subset [0, T]$ be a monotone-increasing set of switching times, namely times at which the parameters α are changed. Thus, denoting by $\alpha(i)$ the *ith* consecutive value of α, Eq. (1) assumes the following form,

$$\dot{x} = f_{\alpha(i)}(x, u), \text{ for all } i \in [\tau_{i-1}, \tau_i), \; i = 1, \ldots, N, \tag{2}$$

where we define $\tau_0 := 0$ and $\tau_{N+1} := T$. Such hybrid systems arise in a variety of application domains including robotics (Arkin, 1998), manufacturing (Boccadoro and Valigi, 2003; Cassandras *et al.*, 2001), power converters (Flieller *et al.*, 1998), and scheduling of medical treatment (Verriest, 2003). More generally, they characterize situations where a controller has to switch attention among various subsystems, or collect data sequentially from a number of sensory sources.

Recently there has been a mounting interest in the optimal control of such systems, where the variable parameter consists of the switching law as well as the input function, and the cost criterion is comprised of a functional defined on the state trajectory and the input function. Ref. (Branicky *et al.*, 1998) defined a general framework for optimal control, and (Sussmann, 2000; Shaikh and Caines, 2002) developed variants of the maximum principle. Refs. (Bemporad and Morari, 2000; Guia *et al.*, 2001; Rantzer and Johansson, 2000) considered piecewise-linear or affine systems, while Refs. (Xu and Antsaklis, 2002a; Xu and Antsaklis, 2002b; Egerstedt *et al.*, 2003a) focused on autonomous nonlinear systems without an external input, and investigated numerical algorithms for their optimization. In these optimal control problems, the parameter consists of the switching times while the sequence of switching modes (i.e., the functions in the right-hand side of Eq. (2)) are given.

The optimal control problem becomes more complicated when the modal sequencing, namely the sequence of functions $f_{\alpha(i)}$ in the right-hand side of Eq. (2), constitutes part of the design variable. What we then have is a discrete-variable problem, in fact we have a scheduling problem. With this problem in mind, Ref. (Egerstedt *et al.*, 2003b) derived a variational formula for inserting a switching mode at a given time $t \in [0, T]$. Specifically, consider an autonomous system of the form

$$\dot{x} \ = \ f(x, t) \tag{3}$$

where

$$f(x, t) = f_{\alpha(i)}(x), t \in [\tau_{i-1}, \tau_i), i = 1, \ldots, N, \tag{4}$$

and let $g(x) : R^n \to R^n$ be a function. Fix a time $\tau \in (0, T)$, and consider modifying the function f by inserting the function g in an interval of length λ centered at τ, $(\tau - \lambda/2, \tau + \lambda/2)$, for $\lambda > 0$. In other words, g is a "new" modal function inserted at time τ. Let us view the cost functional as a function of λ. Then, a formula for the one-sided derivative of the cost functional with respect to λ, at the point $\lambda = 0$, was developed in (Egerstedt *et al.*, 2003b). This formula allowed us to extend the gradient-descent algorithms developed in (Egerstedt *et al.*, 2003a) from the case where the number of switching times was a given constant to the case where the number of switching times, and to an extent the switching sequence itself, were part of the variable. Of course we cannot claim to have solved the modal-sequencing (scheduling) problem to the extent of computing a global minimum, but we have managed to overlay a continuous structure upon the discrete problem, that facilitates the deployment of gradient-descent algorithms for computing local minima. The theoretical underpinnings of this approach is the main concern of this paper.

The main difficulty in analyzing an algorithm that adds switching times is that the optimization problem is not defined on a single vector space, but rather, the parameter seems to belong to a sequence of Euclidean spaces with growing dimensions. Such problems have not had a systematic framework for analysis, and they pose theoretical and practical challenges. In the theory of optimization, the concept of *optimality functions* has provided a measure of how far or close a point is from satisfying given optimality conditions, and this measure being "close to 0" provides a practical stopping rule for an algorithm (see (Polak, 1997)). For example, the magnitude of the gradient in unconstrained problems qualifies as an optimality function, and a practical rule for stopping an algorithm is when the gradient of the function is small. This, however, assumes a single parameter space containing all the iteration points of an algorithm. No such concept exists for situations involving nested spaces with growing dimensions, and this paper provides such a framework and uses it to characterize and prove convergence for the problem at hand. As a last remark we mention that the results derived in this paper works for the case when the cost function, $L(x(t))$, is time variant and when we have constraints on the order of the modal sequence.

Section 2 presents our algorithmic approach and the challenges it poses. Section 3 provides a numerical example, and Section 4 concludes the paper. The algorithm's convergence analysis is highly technical and hence, and due to space constraints, we relegate it to a Technical Memorandum (Axelsson *et al.*, 2004) that can be downloaded from the web.

2. PROBLEM FORMULATION AND AN ALGORITHMIC APPROACH

Consider the hybrid dynamical system described by Eqs. (3) and (4). Defining, for simplicity, the notation $\alpha(i) = i$, and recalling that $\tau_0 := 0$ and $\tau_{N+1} := T$, the system described by Eqs. (3) and (4) can be described by the following equation,

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

where we henceforth assume that the initial condition $x(0) = x_0$ is given and fixed. We also assume that the modal set A is finite but N may be arbitrarily large, so that a particular modal function f_α, $\alpha \in A$, may appear multiple times in the modal sequence $\{f_i\}_{i=1}^{N+1}$. Let $L : R^n \to R$ be a cost function defined on the state variable, and define the cost functional J on the switching times $\tau_1, \ldots, \tau_N$ by

$$J \ = \ \int_0^T L(x(t))dt. \tag{6}$$

The dependence of J on the switching times is apparent from Eq. (5). We make the following assumption.

Assumption 2.1. (i). The functions f_α, $\alpha \in A$, and L, are twice continuously differentiable on R^n. (ii). For every compact set $\Gamma \subset R^n$ there exists a constant $K > 0$ such that, for every $x \in \Gamma$, and for every $\alpha \in A$,

$$||f_\alpha(x)|| \ \leq \ K(||x|| + 1). \tag{7}$$

This assumption guarantees that, with the given initial condition x_0, the differential equation (5) has a unique solution $x(t)$ on the interval $[0, T]$, which is confined to a bounded set in 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, the 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 (5), namely,

$$f(x,t) = f_i(x), t \in [\tau_{i-1}, \tau_i) \ i = 1, \ldots, N+1. \quad (8)$$

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

$$\dot{p}(t) \ = \ -\left(\frac{\partial f}{\partial x}(x,t)\right)^T p(t) \ - \ \left(\frac{\partial L}{\partial x}(x)\right)^T, \quad (9)$$

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 following formula for the partial derivative $dJ/d\tau_i$ was derived in (Egerstedt *et al.*, 2003a):

$$\frac{dJ}{d\tau_i} \ = \ p_i^T\big(f_i(x_i) - f_{i+1}(x_i)\big). \quad (10)$$

Let us denote by σ the given sequence of modal functions $\{f_i\}_{i=1}^N$, and denote by P_σ the problem of minimizing J for the given σ. Note that σ specifies N and the functions f_i but not their switching times, which constitute the parameter for P_σ. The following result, concerning a necessary optimality condition, was derived in (Egerstedt *et al.*, 2003b).

Proposition 2.1. Suppose that $\bar{\tau}_N := (\tau_i, \ldots, \tau_N)^T$ is an optimal point for the problem P_σ. Then:

(1) If $\tau_{k-1} < \tau_k = \tau_n < \tau_{n+1}$ for a pair of integers k and n satisfying the inequalities $1 \leq k \leq n \leq N$, then,

$$\sum_{j=k}^{n} \frac{dJ}{d\tau_j} \ = \ 0, \quad (11)$$

and for every $i = k, \ldots, n$,

$$\sum_{j=k}^{i} \frac{dJ}{d\tau_j} \ \leq \ 0. \quad (12)$$

(2) If $0 = \tau_n < \tau_{n+1}$ for some integer $n \in \{1, \ldots, N\}$, then for every $i = 1, \ldots, n$,

$$\sum_{j=i}^{n} \frac{dJ}{d\tau_j} \ \geq \ 0. \quad (13)$$

Similarly, if $\tau_{k-1} < \tau_k = T$ for some integer $k \in \{1, \ldots, N\}$, then for every $i = k, \ldots, N$,

$$\sum_{j=k}^{i} \frac{dJ}{d\tau_j} \ \leq \ 0. \quad (14)$$

Proof. See (Egerstedt *et al.*, 2003b). □

Next, suppose that the problem P_σ has been solved, for a given modal sequence σ having N switching points, to the extent of computing a vector $\bar{\tau}_N$ satisfying the above optimality condition. Even if this is a global minimum, it may be possible, of course, to further reduce the value of J by altering the sequence σ. An incremental approach, proposed in (Egerstedt *et al.*, 2003b), 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. 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 switching time τ. Suppose now that $\tau \notin \{\tau_1, \ldots, \tau_N\}$, namely it is not one of the switching points, and suppose that $\tau \in (\tau_i, \tau_{i+1})$ for some $i = 1, \ldots, N$. Recall the definition of the costate, p, as given by Eq. (9). Then the following formula characterizes the one sided derivative of J at $\lambda = 0$:

$$\frac{dJ_{g,\tau}(0)}{d\lambda^+} \ = \ p(\tau)^T\big(g(x(\tau)) - f_{i+1}(x(\tau))\big). \quad (15)$$

We point out that similar insertions of multiple modal functions at time τ, according to precedence rules mandated by the system, are possible as well (see (Egerstedt *et al.*, 2003a)), but to keep the discussion simple we confine the discussion to the case where the insertions are made one-at-a-time in the manner described above, and where the sequence σ has no precedence constraints.

The following algorithm can be used to reduce the value of J.

Algorithm 2.1.
Given: A modal sequence σ having N switching points.
Step 1. Use a gradient-descent algorithm to compute a feasible vector $\bar{\tau}_N$ satisfying the optimality condition for P_σ.
Step 2. Compute the number Θ_N defined by

$$\Theta_N \ := \ \min\{\frac{dJ_{g,\tau}(0)}{d\lambda^+} \mid g = f_\alpha; \alpha \in A, \ \tau \in (0, T)\}. \quad (16)$$

Step 3. If $\Theta_N = 0$ then stop and exit. If $\Theta_N < 0$ then, with the pair (g, τ) comprising the argmin in (16), append to σ two switching points at the time τ with the modal function g between them, and goto Step 1.

We mention that when the algorithm goes to Step 1 from Step 3, it will separate the two switching times inserted in Step 3. The dimension of the problem the algorithm solves at Step 1 increases by 2 from one iteration to the next.

The condition $\Theta_N = 0$ constitutes a necessary local-optimality condition. Generally, $\Theta_N \leq 0$ since inserting a function $g = f_\alpha$ during an interval where f_α acts as the dynamic response will give 0 for the min term in Eq. (16). Consequently, we view Θ_N as an optimality function defined on the points computed by Step 1 of the algorithm.

From theoretical and practical standpoints there are two apparent problems associated with this algorithm. First, the parameter space grows (possibly) without bound, and hence we have to define an appropriate notion of convergence. Second, the one-sided derivative $dJ_{g,\tau}(0)/d\lambda^+$ is discontinuous in τ when τ passes through a switching time τ_i; see Eq. (15). To address these difficulties we characterize convergence in an appropriate way, and we use the special structure of the problem to construct a convergent algorithm that is based on Algorithm 2.1.

Consider the *mth* iteration of Algorithm 2.1. Step 1 concerns a modal sequence $\sigma(m)$ having $N(m)$ switching times, and it computes, by a gradient-descent algorithm, a vector $\bar{\tau}_{N(m)}$ satisfying the optimality condition for the problem $P_{\sigma(m)}$. The descent algorithm generally computes a sequence of iteration points, denoted by $\{\bar{\tau}(\ell)\}$, $\ell = 1, 2, \ldots$, such that $J(\bar{\tau}(\ell+1)) \leq J(\bar{\tau}(\ell))$ and $J(\bar{\tau}(\ell)) \to J(\bar{\tau}_{N(m)})$ as $\ell \to \infty$. We assume in the present paper that the limit point $\bar{\tau}_{N(m)}$ is computed exactly, although an implementation of the algorithm will only compute an approximation to it. Let us denote the starting point of this procedure by $\bar{\tau}(0) := (\tau_1(0), \ldots, \tau_{N(m)}(0))$, and observe that, for $\ell > 1$, $\bar{\tau}(0)$ was the point with which Algorithm 2.1 returned from Step 3 to Step 1 in its previous iteration. Now the next iteration point of the procedure is $\bar{\tau}(1)$, and it has the following coordinates, $\bar{\tau}(1) = (\tau_1(1), \ldots, \tau_{N(m)}(1))$.

The specific algorithm deployed in Step 1 of Algorithm 2.1 will determine whether Algorithm 2.1 converges or not. We require certain properties of the first iteration of that algorithm, namely the computation of $\bar{\tau}(1)$ from $\bar{\tau}(0)$; regarding subsequent iterations, we require nothing other that the algorithm be of a gradient-descent type. From the first iteration, we require the following property of *sufficient descent.*

Property of sufficient descent. For every $\epsilon > 0$ there exists $\eta > 0$ such that, in Step 1 of Algorithm 2.1, if $\Theta_N > \epsilon$, then $J(\bar{\tau}(1)) - J(\bar{\tau}(0)) \leq -\eta$.

Similar sufficient-descent properties have been used to prove convergence of various algorithms; see (Polak, 1997) for the general case, and (Axelsson *et al.*, 2004) for our particular algorithm.

The computation of $\bar{\tau}(1)$ from $\bar{\tau}(0)$ involves an Armijo step size along a *descent curve*. Traditionally, the Armijo step size (providing an approximate-line minimization) is deployed along a *descent direction* (Armijo, 1966; Polak, 1997), but we have been unable to determine such a direction that would yield convergence of Algorithm 2.1. Instead, we use a descent curve that consists of a concatenation of multiple linear segments. This curve, parameterized by $\lambda \geq 0$, is denoted by $c(\lambda)$, and its segments are denoted by $c_\nu(\lambda)$, $\nu = 0, 1, 2, \ldots$. The initial point and end point of $c_\nu(\lambda)$ are denoted by $\bar{t}_\nu$ and $\bar{t}_{\nu+1}$, respectively. c_ν is linearly parameterized by $\lambda \in [\lambda_\nu, \lambda_{\nu+1}]$ for some end-points λ_ν and $\lambda_{\nu+1} \geq \lambda_\nu$, so that $c_\nu(\lambda) = \bar{t}_\nu + (\lambda - \lambda_\nu)(\bar{t}_{\nu+1} - \bar{t}_\nu)/(\lambda_{\nu+1} - \lambda_\nu)$. Defining $\bar{h}_\nu$ by $\bar{h}_\nu = (\bar{t}_{\nu+1} - \bar{t}_\nu)/(\lambda_{\nu+1} - \lambda_\nu)$, we have that

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

with $c_\nu(\lambda_{\nu+1}) = \bar{t}_{\nu+1}$. Certainly $c_\nu(\lambda_{\nu+1}) = c_{\nu+1}(\lambda_{\nu+1})$ so that $c(\lambda)$ is a continuous curve comprised of the concatenation of the segments $c_\nu(\lambda)$, $\nu = 0, 1, \ldots$.

The segments $c_\nu(\lambda)$ are defined recursively in the following manner. For $\nu = 0$, let $\bar{t}_0 = \bar{\tau}_{N(m)}$, and let $\lambda_0 = 0$. Next, for $\nu \geq 0$, suppose we are given $\bar{t}_\nu$ and λ_ν; we will define $\bar{h}_\nu$ and $\lambda_{\nu+1}$, so that $c_\nu(\lambda)$ if defined for all $\lambda \in [\lambda_\nu, \lambda_{\nu+1}]$ by (17), and $\bar{t}_{\nu+1} = c_\nu(\lambda_{\nu+1})$.

Let us return to the initial segment, $c_0(\lambda)$, whose starting point is $\bar{t}_0 = \bar{\tau}(0)$, and let $\bar{t}_0 = (t_{0,1}, t_{0,2}, \ldots, t_{0,N(m)})$. Recall that Algorithm 2.1 enters Step 1 with this point from Step 3 of its previous iteration. In the latter iteration, the algorithm inserted a modal function g between two identical switching times, denoted by τ. Thus, the point $\bar{t}_0 = (t_{0,1}, t_{0,2}, \ldots, t_{0,N(m)})$ with which the algorithm enters Step 1, has the following features:

(1) There exists $i \in \{1, \ldots, N(m)\}$ such that $\tau = t_{0,i-1} = t_{0,i}$ and $g = f_i$.
(2) 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 this function.
(3) If we take out the switching times $t_{0,i-1}$ and $t_{0,i}$ (which are identical) and the modal function f_i between them, we obtain the point $\bar{\tau}_{N-2}$ computed by Step 1 of Algorithm 2.1 in its previous iteration, and hence satisfying the optimality condition for the problem $P_{\sigma(m-1)}$.

The inserted switching point $t_{0,i}$ may or may not have been equal to any one of the existing switching times of the previous iteration. In any case, we define $k(i) := \min\{j \leq i-1 : t_{0,j} = t_{0,i-1}\}$ and we define $n(i) := \max\{j \geq i : t_{0,j} = t_{0,i}\}$. Then we have the following equation (see (Axelsson *et al.*, 2004) for a proof):

$$\sum_{j=i}^{n(i)} \frac{dJ}{dt_{0,j}} \;=\; -\sum_{j=k(i)}^{i-1} \frac{dJ}{dt_{0,j}} \;=\; \Theta_{N(m)}. \tag{18}$$

We define the curve $c_\nu(\lambda)$ by reducing $t_{0,i-1}$ to 0 and increasing $t_{0,i}$ to T at the rate of -$\Theta_{N(m)}$. If $t_{0,i-1}$ ($t_{0,i}$, 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. Each such a "bump" causes a change in the direction of $c(\lambda)$, and hence this curve consists of multiple linear segments. Formally, consider $\nu \geq 0$. Given the initial point $\bar{t}_\nu$ and the initial parameter λ_ν of the segment $c_\nu(\lambda)$, compute the rest of the segment as follows. Define: $k_\nu(i-1) := \min\{j \leq i-1 : t_{\nu,j} = t_{\nu,i-1}\}$; $n_\nu(i) := \max\{j \geq i : t_{\nu,j} = t_{\nu,i}\}$; $t_{\nu,-1} := -\infty$; and $t_{\nu;N(m)+2} := \infty$. Define the direction $\bar{h}_\nu := (h_{\nu,1}, \ldots, h_{\nu,N(m)})$ by

$$h_{\nu,r} = \begin{cases} \Theta_N, & \text{for all } r \in \{k_\nu(i-1), \ldots, i-1\}, \\ & \text{if } t_{\nu,i-1} > 0 \\ 0, & \text{for all } r \in \{k_\nu(i-1), \ldots, i-1\}, \\ & \text{if } t_{\nu,i-1} = 0 \\ -\Theta_N, & \text{for all } r \in \{i, \ldots, n_\nu(i)\}, \\ & \text{if } t_{\nu,i} < T \\ 0, & \text{for all } \; r \in \{i, \ldots, n_\nu(i)\}, \\ & \text{if } t_{\nu,i} = T \\ 0, & \text{for all other } r \in \{1, \ldots, N(m)\}. \end{cases} \tag{19}$$

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

$$\begin{aligned} \lambda_{\nu+1} = \min \big\{ \lambda > \lambda_\nu : \\ \text{either } t_{\nu,k(i)} + (\lambda - \lambda_\nu) h_{\nu,k(i)} = t_{\nu,k(i)-1}, \\ \text{or } t_{\nu,n(i)} + (\lambda - \lambda_\nu) h_{\nu,n(i)} = t_{\nu,n(i)+1} \big\}. \end{aligned} \tag{20}$$

For the case where $t_{\nu,i-1} = 0$ and $t_{\nu,N(m)} = T$, there is no need to define $\lambda_{\nu+1}$. Now the segment $c_\nu(\lambda)$ is defined by Eq. (20) for all $\lambda \in [\lambda_\nu, \lambda_{\nu+1}]$, and $\bar{t}_{\nu+1}$ is set to $\bar{t}_{\nu+1} = c_\nu(\lambda_{\nu+1})$.

The computation of $\bar{\tau}(1)$ from $\bar{\tau}(0)$ involves the Armijo step size along the curve $c(\lambda)$ in the following manner (see (Armijo, 1966)). Let us fix $\alpha \in (0, 1)$ and a monotone-decreasing sequence $\{\lambda_s\}_{s=0}^{\infty}$ convergent to 0. Now, define λ_{arm} by

$$\begin{aligned} \lambda_{arm} = \max\{\lambda_s : s = 0, 1, \ldots, \text{ such that} \\ J(c(\lambda_s)) - J(c(0)) \leq -\alpha \lambda_s \Theta_N^2\}, \end{aligned} \tag{21}$$

and define $\bar{\tau}(1)$ by $\bar{\tau}(1) = c_\nu(\lambda_{arm})$.

Thus, the first iteration of the algorithm deployed by Step 1 of Algorithm 2.1 computes $\bar{\tau}(1)$ from $\bar{\tau}(0)$ by using the Armijo step size along the curve $c(\lambda)$, while the subsequent iterations have to use a descent algorithm. We mention that a curve like $c(\lambda)$ may not constitute a descent curve in any but the first iteration! We now state the following convergence result.

Proposition 2.1. Suppose that Algorithm 2.1 computes a sequence of iteration points, $\bar{\tau}_{N(m)}$, $m = 1, 2, \ldots,$. Then,

$$\lim_{m \to \infty} \Theta_{N(m)} = 0. \tag{22}$$

Proof. The proof is based on establishing the sufficient-descent property; please see (Axelsson *et al.*, 2004). □

3. NUMERICAL EXAMPLE

As an illustration, consider the problem of switching between the dynamics $\dot{x} = A_i x$, $i = 1, 2, 3$, given the system matrices

$$A_1 = \begin{pmatrix} -1.2 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}, \; A_2 = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & -1.1 & 0 \\ 0 & 0 & 0.5 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Here, mode 1 stabilizes x_1 while x_2 and x_3 are destabilized. Similarly, mode 2 stabilizes x_2, while mode 3 stabilizes x_3 at slightly different rates. (x_1 is driven to 0 faster in mode 1 than x_2 in mode 2, while x_2 goes to 0 faster in mode 2 than x_3 in mode 3.) We consider the problem where $t \in [0, 1]$, and where we start with three modes (mode 1, 2, and 3), with switches occurring at $\tau_1 = 0.2$, $\tau_2 = 0.8$. After these switch-times have been optimized using an Armijo steepest descent algorithm, a new mode and two new switch times are inserted.

In Figure 1 the result of running this algorithm is shown. Note how, towards the end of the simulation, very little is gained as new modes are inserted. The modal structure and the number of switch times aggregated at each switch point is shown in Figure 2.

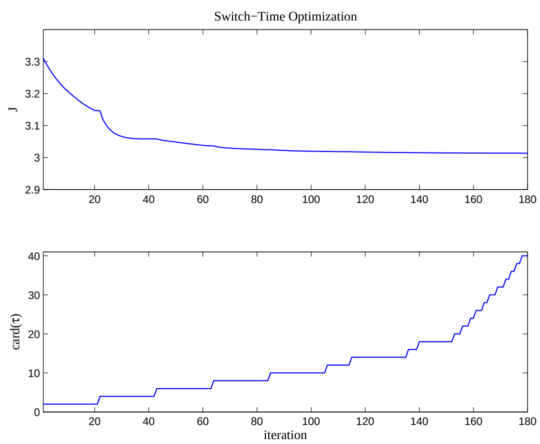

Fig. 1. J is depicted as a function of the iteration in the top figure. The lower figure shows how the length of $\bar{\tau}$ increases as new switch times are inserted.

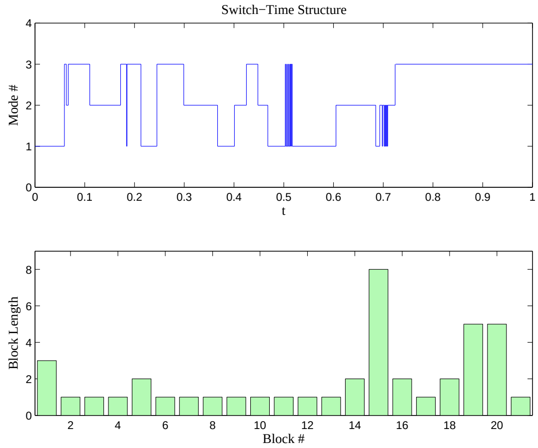

Fig. 2. The top figure shows the modal structure of the final solution. Depicted is the amount of time spent in what mode. The lower figure shows how the switching times aggregate and the "block length", i.e. the number of switching times aggregated together at each switching point.

4. CONCLUSIONS

We have proposed a gradient-descent algorithm for optimal control problems defined on switched-mode systems. The variable parameter consists of the switching times as well as certain aspects of modes' scheduling. We proved convergence of the algorithm in an appropriate sense by using optimality functions and techniques that had been developed in its context.

REFERENCES

Arkin, R.C. (1998). *Behavior Based Robotics*. The MIT Press, Cambridge, MA.

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

Axelsson, H., Y. Wardi, and M. Egerstedt (2004). Transition-Time Optimization in Switched Systems with Scheduling Constraints. Technical Memorandum, School of Electrical and Computer Engineering, Georgia Institute of Technology, {http://users.ece.gatech.edu/∼magnus/Papers/OptSwitchTechReport.pdf}.

Bemporad A., F. Borrelli and M. Morari (2000). Piecewise Linear Optimal Controllers for Hybrid Systems. In: *Proceedings of the American Control Conference*, pp. 1190-1194, Chicago, Illinois.

Boccadoro M. and P. Valigi (2003). A Modelling Approach for the Dynamic Scheduling Problem of Manufacturing Systems with Non-Negligible Setup Times and Finite Buffers. In: *42nd IEEE Conference on Decision and Control*, Maui, Hawaii, USA.

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

Cassandras C.G., D.L. Pepyne, and Y. Wardi (2001). Optimal Control of a Class of Hybrid Systems. In: *IEEE Trans. on Automatic Control*, Vol. 46, No. 3, pp. 398-415.

Egerstedt M., Y. Wardi, and F. Delmotte (2003a). Optimal Control of Switching Times in Switched Dynamical Systems. In: *Proceedings of the IEEE Conference on Decision and Control*, Maui, Hawaii.

Egerstedt M., Y. Wardi, and H. Axelsson (2003b). Transition-Time Optimization for Switched Systems. Submitted to the *IEEE Transactions on Automatic Control.*

Flieller D., J.P. Louis, and J. Barrenscheen (1998). General Sampled Data Modeling of Power Systems Supplied by Static Converter with Digital and Analog Controller. *Mathematics and Computer in Simulation*, Vol. 46, pp. 373-385.

Guia A., C. Seatzu, and C. Van der Mee (2001). Optimal Control of Switched Autonomous Linear Systems. In: *Proceedings of the 40th Conference on Decision and Control*, pp. 1816-1821, Phoenix, Arizona.

Polak E. (1997). *Optimization Algorithms and Consistent Approximations*. Springer-Verlag, New York.

Rantzer A. and M. Johansson (2000). Piecewise Linear Quadratic Optimal Control. In: *IEEE Transactions on Automatic Control*, Vol. 54, pp. 629-637.

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

Sussmann H.J. (2000). Set-Valued Differentials and the Hybrid Maximum Principle. *IEEE Conference on Decision and Control*, Vol. 1, pp. 558-563.

Verriest E. (2003). Regularization Method for Optimally Switched and Impulsive Systems with Biomedical Applications (I). In: *Proceedings of the IEEE 42nd Conference on Decision and Control*, Maui, Hawaii.

Xu X. and P. Antsaklis (2002a). Optimal Control of Switched Autonomous Systems. *IEEE Conference on Decision and Control*, Las Vegas, NV.

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