Discretized Switching Time Optimization Problems*

Kathrin Flaßkamp[1], Todd Murphey[2], and Sina Ober-Blöbaum[1]

Abstract— **A switched system is defined by a family of vector fields together with a switching law which chooses the active vector field at any time. Thus, the switching law encoding the switching times and the sequence of modes may serve as a design parameter. Switching time optimization (STO) focuses on the optimization of the switching times in order to govern the system evolution to a desired behavior described by some cost function. However, it is rare that a STO problem can be solved analytically leading to the use of numerical approximation methods. In this contribution, we directly start with applying integration schemes to approximate the system's state and adjoint trajectories and study the effect of this discretization. It turns out that in contrast to the continuous time problem, the discretized problem loses differentiability with respect to the optimization variables. The isolated nondifferentiable points can be precisely identified though. Nevertheless, to solve the STO problem, nonsmooth optimization techniques have to be applied which we illustrate using a hybrid double pendulum.**

I. INTRODUCTION

Switched systems consist of a family of vector fields, that define differential equations describing the system's continuous dynamics, together with some switching law taking into account discrete events, i.e. instantaneous switches between the different vector fields. The dynamics of a switched system is described by hybrid trajectories which are solutions to the currently active differential equation at every point of time.

Regarding control purposes, switched systems cannot only be controlled by feedforward or feedback controls, but also by the switching law itself. The switching law encodes the *switching times* and the sequence of active vector fields, the so called *modes*. Both provide additional design parameters, but in this contribution, we focus on the *switching time optimization* (STO) and assume the mode sequence to be fixed. Switching time optimization of hybrid dynamical systems has been studied in various settings and from different perspectives in the last years (cf. among others [1–8]).

As in ordinary optimal control, numerical techniques have to be applied to approximate solutions of STO problems. While most approaches compute necessary optimality conditions first and perform a discretization afterwards (often implicitly when solving the state / adjoint system by some black box numerical integrator), in our work, we start with a discretization of the time continuous STO problem. Analogous adjoint equations for the discretized problem can be derived and, furthermore, this approach enables us to study the effects of this discretization. Unlike the ordinary smooth optimal control problem, where differentiability of the discrete time control problem is inherited by the continuous time problem, the STO problem in discrete time will in general, although in a very structured way, be nondifferentiable.

This paper is organized as follows: starting with a short introduction to switching time optimization in continuous time in Section II, we briefly recall some important facts from [9] on discretized STO problems in Section III. This forms the basis for the major contributions of this paper: in Section IV we analyze the existence of smooth or nonsmooth optimal values despite nonsmooth cost functions and give examples for both cases. A subgradient method tailored to our particular nonsmooth optimization problem is presented in Section V. Finally, the optimization is performed for two examples and we conclude with an outlook to future work in Section VI.

II. SWITCHING TIME OPTIMIZATION IN CONTINUOUS TIME

In this contribution, we focus on switched autonomous dynamical systems, i.e. there are instantaneous switches between the autonomous vector fields. Switching is not assumed to be state-dependent, neither do we consider discrete jumps in the states. This naturally leads to hybrid trajectories which are continuous, but only piecewise smooth w.r.t. time. However, the STO problem is smooth, i.e. the cost function is differentiable w.r.t. the switching times under reasonable assumptions that are specified below. Even in the simplest case of two vector fields and one single switching point, nondifferentiability in the discretized setting can occur, therefore we restrict to the following basic problem setting in the remainder of this paper.

Problem 2.1: Let $\mathscr{X} \subset \mathbb{R}^n$ be a state space with $x_0 \in \mathscr{X}$. Let $T, \tau \in \mathbb{R}$ with $0 \leq \tau \leq T$, $f_1, f_2 \in C^1$ and $\ell \in C^1$, $m \in C^1$. Then we consider the following problem

$$\min_{\tau} J(\tau) = \int_0^T \ell(x(t),t)\,dt + m(x(T)) \tag{1}$$

w.r.t.

$$\dot{x}(t) = \begin{cases} f_1(x(t)) & t < \tau \\ f_2(x(t)) & t \geq \tau \end{cases} \quad \text{and } x(0) = x_0. \tag{2}$$

Here, the hybrid trajectory is in fact a function of the switching time τ as well. Its derivative w.r.t. τ for $t \in (\tau, T)$

*This contribution was partly developed and published in the course of the Collaborative Research Centre 614 "Self-Optimizing Concepts and Structures in Mechanical Engineering" funded by the German Research Foundation (DFG) under grant number SFB 614.

[1]K. Flaßkamp and S. Ober-Blöbaum are with the Department of Mathematics, University of Paderborn, Warburger Str. 100, 33098 Paderborn {kathrinf, sinaob}@math.uni-paderborn.de

[2]T. Murphey is with the Mechanical Engineering Department, Northwestern University, 2145 Sheridan Rd., Evanston, IL 60208, USA t-murphey@northwestern.edu

is given by (cf. [1])

$$\frac{dx(t)}{d\tau} = \Phi(t,\tau)(f_1(x(\tau)) - f_2(x(\tau))), \tag{3}$$

with $\Phi(t,\tau)$ being the state transition matrix of the autonomous linear system $\dot{z} = \frac{\partial f_2(x(t))}{\partial x} z$.

Derivatives of the cost function w.r.t. the switching time can be computed by means of costate (also called adjoint) differential equations, as it has been proven in several works (cf. e.g. [1, 3, 7]). We recall from [7]:

Lemma 2.1: Let f_1, f_2 and ℓ be as in Problem 2.1. Define the costate by

$$\begin{aligned}\dot{\rho}(t) &= -\left(\frac{df_2}{dx}(x(t))\right)^T \rho(t) - \left(\frac{d\ell}{dx}(x(t))\right)^T \\ \rho(T) &= \frac{dm}{dx}(x(T))^T.\end{aligned} \tag{4}$$

Then, $J'(\tau)$ has the following form,

$$J'(\tau) = \rho(\tau)^T [f_1(x(\tau)) - f_2(x(\tau))]. \tag{5}$$

This result is extended to several switching times and different vector fields in [1]. In particular it follows that the cost function is differentiable for any set of disjoint switching times. In [3], the special case of coinciding switching times is studied, to which we will refer later in our analysis.

Formulas for the second order derivative have been derived in [7] and [8]. Lemma 2.1 can be used to develop gradient-based optimization techniques for the computation of an optimal switching time $\tau_{opt} = \mathrm{argmin}_\tau J(\tau)$. Such numerical techniques are based on solving the state equation (2) and the adjoint equation (4) in alternation to generate a descent direction for the next iteration. A feasible step size can be generated e.g. by the Armijo rule (cf. [3] or [7]). In case of multiple switches, the algorithm presented in [3] also deals with coinciding switching points.

III. DISCRETIZED SWITCHED SYSTEMS

In most applications, e.g. when facing highly nonlinear dynamics, it is impossible to solve either the state (2) or the adjoint equation (4) analytically. Thus, numerical methods have to be applied to approximate a solution. While the approaches cited in Section II implicitly use a numerical integration method in the optimization algorithms, in this contribution, we directly start with a discretization of Problem 2.1. Then we can analyze the influence of the discretization regarding differentiability of the cost function.

Problem 3.1: Let $\Delta t = \{t_0, t_1, \ldots, t_N\}$ be a discrete time grid with $t_0 = 0, t_N = T$ and $\tau \in [t_i, t_{i+1}]$ for some $i \in \{0, \ldots, N\}$. Let $\mathscr{X} \subset \mathbb{R}^n$ be the state space with $x_0 \in \mathscr{X}$, $f_1, f_2 \in C^1$ and $\ell \in C^1$. Then we consider the following problem,

$$\min_\tau J_d(\tau) = \sum_{k=0}^{N} \Psi_k(x_k) \approx \int_0^T \ell(x(t))\,dt + m(x(T)) \tag{6}$$

w.r.t.

$$\mathbf{K}\left(\{t_k\}_{k=0}^N, \tau, \{x_k\}_{k=0}^N\right) = 0, \tag{7}$$

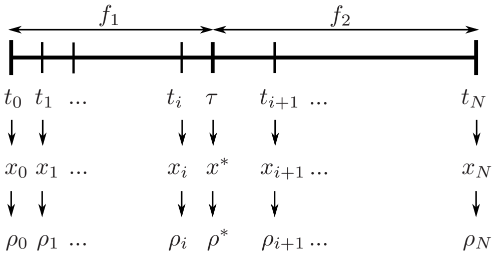

Fig. 1. Notation for discretization as used in the integration schemes and for the definition of discrete adjoints.

a system of algebraic equations resulting from the discretization of (2).

The discretized trajectory $x_d = \{x_k\}_{k=0}^N$ is an approximation of the exact solution, i.e. $x_k \approx x(t_k)$ for $k = 0, \ldots, N$, and it also depends on the switching time τ. Note that τ is still allowed to vary continuously in $[0, T]$. If existent, the derivative of (6) is given by

$$J_d{}'(\tau) = \frac{d}{d\tau} J_d(\tau) = \sum_{k=0}^{N} D\Psi_k(x_k) \cdot \frac{d}{d\tau} x_k. \tag{8}$$

Assuming continuously differentiable functions Ψ_k for $k = 0, \ldots, N$, the crucial part is the derivative of the discrete trajectory. For simplicity, we restrict to explicit one-step integration schemes in the following, although it has to be emphasized that qualitatively identical results arise for implicit schemes.

We discretize (2) by an explicit one-step scheme, i.e. (7) has the following form

$$\mathbf{F} = \begin{cases} x_{k+1} - F_1(x_k, t_k, t_{k+1}) = 0 & k = 0, \ldots, i-1, \\ x^* - F_1(x_i, t_i, \tau) = 0 & \text{and} \\ x_{i+1} - F_2(x^*, \tau, t_{i+1}) = 0, & \\ x_{k+1} - F_2(x_k, t_k, t_{k+1}) = 0 & k = i+1, \ldots, N-1. \end{cases} \tag{9}$$

F_1 and F_2 denote the schemes for the different vector fields f_1 and f_2 that switch at time τ (cf. Fig. 1). Thus, it holds $\tau \in [t_i, t_{i+1}]$ for some $i \in \{0, \ldots, N\}$. It can be seen that $\{x_k\}_{k=0}^N$ is continuous w.r.t. τ. For the derivative w.r.t. τ, the following holds

$$\frac{d}{d\tau} x_{k+1} = \begin{cases} 0 & \text{for } k = 0, \ldots, i-1, \\ D_1F_2(x^*, \tau, t_{i+1}) \cdot D_3F_1(x_i, t_i, \tau) & \\ \; + D_2F_2(x^*, \tau, t_{i+1}), & \\ D_1F_2(x_k, t_k, t_{k+1}) \cdot \frac{d}{d\tau} x_k & \text{for } k = i+1, \ldots, N-1. \end{cases} \tag{10}$$

Here, we use the slot derivative notation, i.e. $D_1F_2(\cdot,\cdot,\cdot)$ is the partial derivative of F_2 w.r.t. its first argument, $D_2F_2(\cdot,\cdot,\cdot)$ is the derivative w.r.t. the argument in the second slot and so forth. For $\tau \in (t_i, t_{i+1})$, $\frac{d}{d\tau} x_{k+1}$ for $k = 0, \ldots, N-1$ is continuous, if F_1 and F_2 are continuously differentiable, which is a reasonable requirement on an explicit integration scheme. Now we study the case when τ coincides with a grid point, say $\tau = t_{i+1}$. Therefore, we look at the left and right limits: while we have $\lim_{\substack{\tau \to t_{i+1} \\ \tau > t_{i+1}}} \frac{d}{d\tau} x_{i+1} = 0$, because switching happens afterwards, in general, we have

$$\lim_{\substack{\tau \to t_{i+1} \\ \tau < t_{i+1}}} \frac{d}{d\tau} x_{i+1} = \lim_{\substack{\tau \to t_{i+1} \\ \tau < t_{i+1}}} D_1 F_2(x^*, \tau, t_{i+1}) \cdot D_3 F_1(x_i, t_i, \tau) + D_2 F_2(x^*, \tau, t_{i+1}) \neq 0 \tag{11}$$

and thus, $\frac{d}{d\tau} x_{i+1}$ and therefore all $\frac{d}{d\tau} x_{k+1}$ for $k > i$ are nondifferentiable for $\tau = t_{i+1}$. Although (11) has to be checked for each integration scheme and each system individually, most likely the nondifferentiability of x_k, $(k = i+1, \ldots, N)$ at time grid points is existent for a system with arbitrary switching vector fields. As we saw in (8), $\frac{d}{d\tau} x_k$ is part of the discrete cost function derivative and thus, nondifferentiability of the discrete trajectory generally leads to nondifferentiability of J_d. The iterative relation of the derivatives at neighboring trajectory points gives rise to a transition operator

$$\Phi(k+1, k) := D_1 F_2(x_k, t_k, t_{k+1}) \tag{12}$$

for $k \in \{i+1, \ldots, N-1\}$. We define $\Phi(k,k) := 1$ and for $l > k+1$, $\Phi(l,k) := \Phi(l, l-1) \cdot \ldots \cdot \Phi(k+2, k+1) \cdot \Phi(k+1, k)$. Thus, for $k \in \{i+1, \ldots, N-1\}$ one receives the propagation scheme

$$\frac{d}{d\tau} x_{k+1} = \Phi(k+1, i+1) \cdot \frac{d}{d\tau} x_{i+1},$$

which is reminiscent of (3) in the continuous time setting.

We define the discrete adjoints recursively by

$$\rho_k = D\Psi_k(x_k) + \rho_{k+1} \cdot \Phi(k+1, k)$$

for $k = N, \ldots, i+1$, with boundary value $\rho_N = D\Psi_N(x_N)$. Thus, the adjoints are continuous w.r.t. τ, if the $D\Psi_k$ and the transition operator are continuous, which is reasonable to assume. With the help of the adjoints, the discrete cost function derivative can be written as

$$\begin{aligned} J_d{}'(\tau) &= \sum_{k=0}^{N} D\Psi_k(x_k) \frac{d}{d\tau} x_k \\ &= \sum_{k=i+1}^{N} D\Psi_k(x_k) \cdot \Phi(k, i+1) \cdot \frac{d}{d\tau} x_{i+1} = \rho_{i+1} \cdot \frac{d}{d\tau} x_{i+1}. \end{aligned}$$

So it can be nicely seen that although the adjoint itself is continuous, its argument, i.e. $\frac{d}{d\tau} x_{i+1}$ leads to nondifferentiability of J_d. In fact, if $\tau = t_i$ for $i \in \{0, \ldots, N\}$, $\frac{d}{d\tau} x_{i+1}$ and therefore $J_d{}'$ can only be defined by either the left or the right limit as defined in (11).

Example 3.1 (Explicit Euler): The explicit Euler scheme for a switched system for $k \in \{0, \ldots, N-1\}$ is given by

$$F_j(x_k, t_k, t_{k+1}) = x_k + (t_{k+1} - t_k) \cdot f_j(x_k), \quad j = \{1, 2\}$$

and on the switching interval with x^* and τ in the appropriate arguments. For $\tau \in (t_i, t_{i+1})$ we receive

$$\frac{d}{d\tau} x_{i+1} = f_1(x_i) + \frac{d}{dx} f_2(x^*) \cdot (t_{i+1} - \tau) \cdot f_1(x_i) - f_2(x^*)$$

with $x^* = x_i + f_1(x_i) \cdot (\tau - t_i)$. Thus, at $\tau = t_{i+1}$, $\frac{d}{d\tau} x_{i+1}$ switches from zero to $f_1(x_i) - f_2(x_{i+1})$. What is the effect on the next node x_{i+1} (and thus on all the following nodes)? From (10) we know that $\frac{d}{d\tau} x_{i+2} = (1 + (t_{i+2} - t_{i+1}) \frac{\partial}{\partial x} f_2(x_{i+1})) \frac{d}{d\tau} x_{i+1}$ and hence,

$$\lim_{\substack{\tau \to t_{i+1} \\ \tau < t_{i+1}}} \frac{d}{d\tau} x_{i+2} = \left(1 + (t_{i+2} - t_{i+1}) \frac{\partial}{\partial x} f_2(x_{i+1}) \right) \cdot [f_1(x_i) - f_2(x_{i+1})],$$

but for the limit from the right we receive by shifting the index in (9)

$$\lim_{\substack{\tau \to t_{i+1} \\ \tau > t_{i+1}}} \frac{d}{d\tau} x_{i+2} = \left(1 + (t_{i+2} - t_{i+1}) \frac{\partial}{\partial x} f_2(x_{i+1}) \right) \cdot f_1(x_{i+1}) - f_2(x_{i+1}).$$

In general, these two limits do not coincide. Thus, $\{x_k\}_{k=0}^N$ is nondifferentiable at $\tau = t_{i+1}$. However, when reducing the time steps, i.e. in particular $t_{i+1} - t_i \to 0$, the discrete case ($\frac{d}{d\tau} x_{i+1}$) matches the continuous, in which the limit is

$$\lim_{\substack{t \to \tau \\ t > \tau}} \frac{d}{d\tau} x(\tau) = f_1(x(\tau)) - f_2(x(\tau)).$$

Example 3.2 (Hybrid double pendulum): We discretize the STO problem for a hybrid double pendulum by an explicit Euler scheme and study the effects on the smoothness of the original problem. The model of the pendulum consists of two mass points m_1, m_2 on massless rods of length l_1, l_2. The motion of the pendula are described by two angles, φ_1 and φ_2 (cf. Fig. 2). The standard double pendulum is turned into a hybrid system by introducing two different modes: **M1**: The outer pendulum is locked w.r.t. the inner pendulum with angle θ, i.e. the system behaves like a single pendulum with a special inertia tensor. **M2**: Both pendula can move freely as in the standard case.

In M1, the following energy terms for the kinetic energy K and potential energy V are valid

$$\begin{aligned} K_1(\varphi_1, \dot{\varphi}_1) &= \frac{1}{2}(m_1 l_1^2 + m_2 r^2) \cdot \dot{\varphi}_1^2 \\ V_1(\varphi_1) &= (m_1 + m_2) g l_1 \cos(\varphi_1) + m_2 g l_2 \cos(\varphi_1 + \theta - \pi) \end{aligned}$$

with distance r of outer mass to origin $r^2 = l_1^2 + l_2^2 - 2 l_1 l_2 \cos(\theta)$. The position of the outer mass can be updated according to $\varphi_2 = \varphi_1 + \theta - \pi$ and it naturally follows that $\dot{\varphi}_1 = \dot{\varphi}_2$. In M2, the system is defined by

$$\begin{aligned} &K_2(\varphi_1, \varphi_2, \dot{\varphi}_1, \dot{\varphi}_2) = \frac{1}{2} \begin{pmatrix} \dot{\varphi}_1 \\ \dot{\varphi}_2 \end{pmatrix}^T \cdot \\ &\begin{pmatrix} (m_1 + m_2) l_1^2 & m_2 l_1 l_2 \cos(\varphi_1 - \varphi_2) \\ m_2 l_1 l_2 \cos(\varphi_1 - \varphi_2) & m_2 l_2^2 \end{pmatrix} \cdot \begin{pmatrix} \dot{\varphi}_1 \\ \dot{\varphi}_2 \end{pmatrix} \\ &V_2(\varphi_1, \varphi_2) = m_1 g l_1 \cos(\varphi_1) + m_2 g (l_1 \cos(\varphi_1) + l_2 \cos(\varphi_2)). \end{aligned}$$

In both cases, the equations of motion are derived by the Euler-Lagrange equations $\frac{d}{dt} \frac{\partial L_i}{\partial \dot{q}} - \frac{\partial L_i}{\partial q} = 0$ for $L_i(q, \dot{q}) = K_i(q, \dot{q}) - V_i(q, \dot{q})$ $(i = 1, 2)$ with $q = (\varphi_1, \varphi_2)$ being the configurations and $\dot{q} = (\dot{\varphi}_1, \dot{\varphi}_2)$ the corresponding velocities. We focus on the scenario, when the system switches a single time from M1 to M2. One can check that the energies of M1 and M2 coincide in a switching point $x_\tau = (\varphi_1, \varphi_1 + \theta - \pi, \dot{\varphi}_1, \dot{\varphi}_1)$ and thus we will have energy conservation along the entire hybrid trajectory. We assume that the velocities

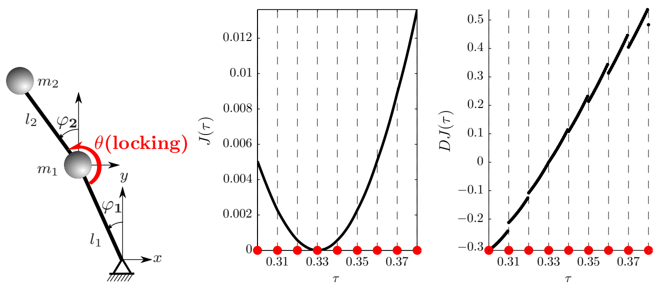

Fig. 2. *Left:* Sketch of the locked double pendulum: in mode 1, the outer pendulum is locked w.r.t. the inner pendulum with angle θ. In mode 2, the system is a normal planar double pendulum. *Right:* Cost function evaluations and its derivative for a switched trajectory of the pendulum: while J is continuous w.r.t. switching time τ, nondifferentiable points occur when τ coincides with a node of the discrete time grid (red dots). This is caused by the approx. trajectory, which is nondifferentiable w.r.t. τ at those points.

directly before and after the switch are the same, i.e. $\dot{\varphi}_1^- = \dot{\varphi}_1^+ = \dot{\varphi}_2^+$. As a cost function we choose $J(\tau) = m(x(T)) = \left\| \begin{pmatrix} \varphi_1(T) \\ \varphi_2(T) \end{pmatrix} - q_{\text{final}} \right\|^2$ to minimize the distance to a given final point. This is an algebraic cost function as considered in Problem 3.1. The final point $q_{\text{final}} = (-1.5487, -1.9733)$ is chosen such that the optimal value is $\tau^* = 0.33$. We approximate the switching time derivative $J_d{}'(\tau)$ by evaluating the corresponding formula for $\frac{d}{d\tau}x_{i+1}$ and the appropriate discrete adjoints. In Fig. 2 (right) the nondifferentiable points of $J_d{}'(\tau)$, i.e. points in which the left hand and right hand side derivatives do not coincide, can be clearly seen.

IV. ANALYSIS OF NONSMOOTHNESS

Fig. 2 (right) shows a typical behavior for a large class of systems: nondifferentiable points occur at all time grid points except for the optimal switching time point, at which the derivative DJ_d smoothly crosses zero (in Example 3.2, this happens at $\tau = 0.33$). We will first explain this effect for scalar and higher dimensional systems, but afterwards introduce an example which has a nondifferentiable optimal point, i.e. a kink at the minimum of J_d.

Let us first consider the one-dimensional case. The following lemma will be helpful.

Lemma 4.1: Let f be at least $C^1(\mathbb{R})$ and convex, while g is a $C(\mathbb{R})$ function only, i.e. there exist isolated nondifferentiable points. We assume that at such points – one of them be x_0 – a left hand side and a right hand side limit of the difference quotient exist but do not coincide. Then, $f \circ g$ is differentiable in $x_0 \in \mathbb{R}$ if and only if f has an (unconstrained) extremum in $g(x_0)$.

Proof: Assume first that f is extremal in $y_0 := g(x_0)$, i.e. $f'(g(x_0)) = 0$. Then we consider the one sided difference quotient of $f \circ g$ to which we are allowed to apply the chain rule, since one sided limits of both functions exists.

$$\lim_{x \uparrow x_0} \frac{(f \circ g)(x) - (f \circ g)(x_0)}{x - x_0} = f'(g(x_0)) \cdot \lim_{x \uparrow x_0} \frac{g(x) - g(x_0)}{x - x_0} = 0$$

Analogously, this holds for the limit from the right. Thus both directional derivatives coincide and so they define the derivative of $f \circ g$ at x_0 to be zero. If we now assume this last statement to be true, the only solution of this one dimensional equation

$$f'(g(x_0)) \cdot \lim_{x \uparrow x_0} \frac{g(x) - g(x_0)}{x - x_0} = f'(g(x_0)) \cdot \lim_{x \downarrow x_0} \frac{g(x) - g(x_0)}{x - x_0}$$

is $f'(g(x_0)) = 0$. So $g(x_0)$ is a critical point of f and because f is assumed to be convex, it is an extremum. ■

Back to discretized STO problems, we see that the interplay of the discretized cost functions Ψ_k with the discrete trajectory as a C^0-function of τ may or may not cause nondifferentiable optimal points. Whenever there is an admissible $\tau \in [0, T]$ that generates a discrete trajectory which minimizes $\Psi(x_d) = \sum_{k=0}^{N} \Psi_k(x_k)$ as an unconstrained optimum, this will be the minimizer of $J_d = \Psi(x_d(\tau))$ as well and it will be smooth regardless of a possible nonsmoothness of $x_d(\tau)$ at that point.

In higher dimensions (but the same situation as in Lemma 4.1), it still holds that if $g(x_0) \in \mathbb{R}^n$ is an unconstrained extremum of $f : \mathbb{R}^n \to \mathbb{R}$ on $\mathbb{R}^n$, then $f \circ g$ is differentiable in x_0 with $D(f \circ g)(x_0) = 0$. In addition, there are further chances of differentiability in x_0 despite nondifferentiability of g in this point, since $D(f \circ g)(x_0) = Df(g(x_0)) \cdot Dg(x_0) = 0$ may be also achieved if the derivatives are orthogonal to each other. However, this does not, of course, generally exclude the existence of nonsmooth optima in higher dimensional switched systems.

In the following we present a simple one-dimensional example which does not fulfill the requirements of Lemma 4.1 such that the optimum is nonsmooth. Although this is a cooked up example, it shows that nonsmoothness at optima may occur in more complex control systems as well and has to be accounted for in the optimization.

Example 4.1: Consider the switched linear system

$$\dot{x} = \begin{cases} Ax & t \leq \tau \\ Bx & t > \tau \end{cases}$$

with $A = 0.2$ and $B = -1.5$ and initial point $x_0 = 10.1$. We take a rough discretization of the time interval $[0, 2]$ by $\Delta t = \{0, 1, 2\}$. The discrete trajectory $x_d = \{x_0, x_1, x_2\}$ together with x_τ at the switching point is generated by an explicit Euler scheme (Fig. 3). Fig. 4 shows that the minimum of $x_N(\tau)$ is at $\tau = t_1 = 1$. For the cost function, only a final point cost is considered, such that (6) reduces to

$$J_d(\tau) = \Psi_N(x_N(\tau)).$$

We choose $\Psi_N(x) = (x + 10)^2$ which is smooth and convex, but its extremum at $x = -10$ is not in the image of $x_N(\tau)$ for $\tau \in [0, 2]$ (see Fig. 4). The resulting $J_d(\tau)$ is given in Fig. 5. Obviously, it has a minimum in $\tau^* = 1$ but it is nonsmooth there, so $J_d'(\tau^*) = 0$, the usual optimality condition, does not hold.

V. OPTIMIZATION

To solve a switching time optimization problem as Problem 2.1, we apply numerical optimization techniques to the discretized Problem 3.1. In the previous section we showed

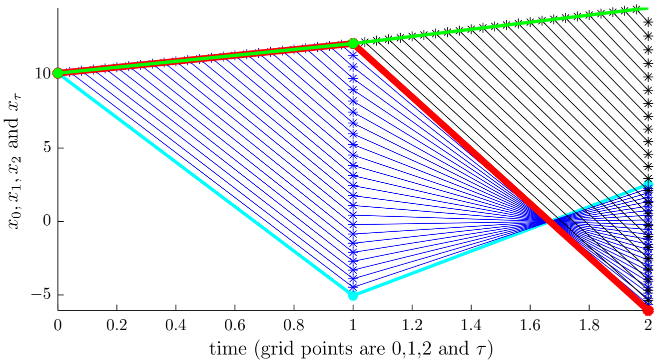

Fig. 3. Trajectories for varying $\tau \in [0,2]$ of Example 4.1. The first trajectory for $\tau = 0$ is plotted in cyan, the following ones in blue, the trajectory corresponding to $\tau = 1$ in red (note that this one gives the minimal value for x_{N+1}). Trajectories for $\tau > 1$ are plotted in black except for the last one ($\tau = 2$) which is green.

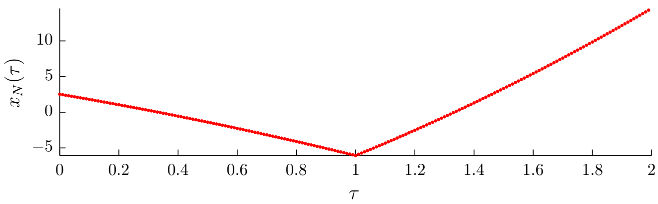

Fig. 4. The final point of the discrete trajectory, x_N, plotted as a function of τ for $\tau \in [0,2]$, is nonsmooth at the grid point $\tau = 1$.

that the discretized problem is nonsmooth. Therefore we cannot apply ordinary methods for nonlinear optimization, since those are known to fail for nonsmooth problems (cf. e.g. [10]): convergence to wrong points, failure of the stopping criteria or extremely inaccurate gradient approximations may occur.

A simple method of smooth optimization is the method of gradient descent [11]. For nonsmooth functions the gradient, which does not exist everywhere, can be replaced by subgradients. In the following we assume that the cost function is convex, such that we can use the theory of ordinary subgradients. However, convexity cannot be assumed for discretized STO problems in general, so it might be necessary to use generalized subgradients as proposed e.g. in [12].

Definition 5.1 (Subgradient and Subdifferential, cf. [10]): Let $f : X \to \mathbb{R}$ be a convex function on the convex open set X. A vector $g \in \mathbb{R}^n$ is called a *subgradient* of f in $x \in X$, if

$$f(y) - f(x) \geq g^T \cdot (y - x) \quad \forall y \in X.$$

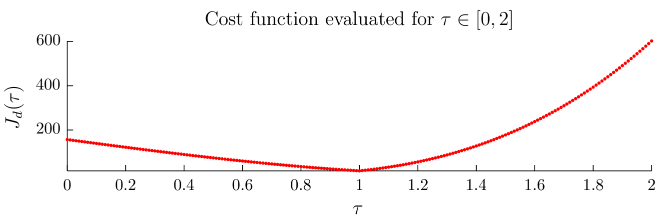

Fig. 5. $J_d = \Psi(x_N(\tau))$ has a nonsmooth point at the optimum, although $\Psi(x) = (x+10)^2$ itself is quadratic, thus smooth and convex.

The set $\partial f(x) \subset \mathbb{R}^n$,

$$\partial f(x) = \{g \in \mathbb{R}^n \,|\, g^T(y-x) \leq f(y) - f(x) \;\forall y \in X\}$$

is called the *subdifferential* of f in $x \in X$.

The necessary and sufficient optimality condition for convex nonsmooth functions is: $x^* \in X$ is a minimum of f if and only if $0 \in \partial f$. For the directional derivative $f'(x;d) = \inf_{t>0} \frac{f(x+td)-f(x)}{t}$ with $d \in \mathbb{R}^n$ it holds that $f'(x;d) = \max_{g \in \partial f(x)} g^T d$ for all $d \in \mathbb{R}^n$. If f is differentiable in x, the subdifferential reduces to $\nabla f(x)$ [12].

We apply the following algorithm, similar to subgradient methods proposed in [10] or [13], to the discretized switching time optimization problem[1] with cost function $J_d : [0,T] \subset \mathbb{R} \to \mathbb{R}$, $J_d(\tau) = \sum_{k=0}^{N} \Psi_k(x_k(\tau))$.

Algorithm 5.1 (Subgradient descent with projections): Take an initial point $\tau^{(0)}$, choose small values $\mathrm{tol}_g, \mathrm{tol}_\tau$ and set $k := 0$.

1) Compute a subgradient $g^{(k)} \in \partial J_d(\tau^k)$
2) Stopping criteria: if $\|g^{(k)}\| \leq \mathrm{tol}_g$ or $\|\tau^{(k-1)} - \tau^{(k)}\| \leq \mathrm{tol}_\tau \longrightarrow$ stop!
3) Let $d^{(k)} = -g^{(k)}/\|g^{(k)}\|$. Choose some appropriate $s_k \geq 0$ and define $\tau^{(k+1)} = P_{[0,T]}(\tau^{(k)} + s_k d^{(k)})$.
4) Set $k := k+1$ and return to 1.

Note that in the previous section, we identified J_d to be "piecewise-C^1" ([10]), i.e. a gradient exists almost everywhere (and a directional derivative can be always given). Thus, the probability that we have to compute a real subgradient in step 1 of Algorithm 5.1 is zero, otherwise we would take a directional derivative. However, in step 2, the first stopping criterion ($\|g^{(k)}\| \leq \mathrm{tol}_g$), which is common in smooth optimization, does not take effect if the minimum is a kink as in Example 4.1. Therefore, we add the second, very simple stopping criterion. Advanced nonsmooth optimization techniques such as bundle methods (cf. [10] for an early reference; much research on these methods followed since then) allow more sophisticated stopping criteria. In step 3, $P_{[0,T]}$ denotes a projection onto the feasible (convex) set (cf. [13]), i.e. the interval $[0,T]$ in our case.

Convergence of the algorithm, even though with a very low rate ([10, 13]), is assured if the step sizes fulfill $\lim_{k\to\infty} s_k = 0$ and $\sum_{k=0}^{\infty} s_k = \infty$. A simple choice of step sizes that meet this conditions is $s_k = 1/(k+1)$. In case the optimal value, J_d^* is known (e.g. if the distance to a reference trajectory has to be minimized, which is admissible for some $\tau \in [0,T]$) an optimal choice of step sizes is given by

$$s_k = \|g^{(k)}\|^{-1} \cdot (J_d(\tau^{(k)}) - J_d^*) \tag{13}$$

(see e.g. [10]).

Example 5.1 (Nonsmooth optimum): We consider again Example 4.1. As seen before (cf. Fig. 5) the cost function is nonsmooth at the optimum. So $\|g_k\|$ will never come close to zero, although the optimal solution is obviously given by

[1] The general statements hold for STO problems with multiple switches, of course. However, since we apply the algorithm to the previously introduced one-dimensional optimization problems, we omit to introduce a vector of switching times here.

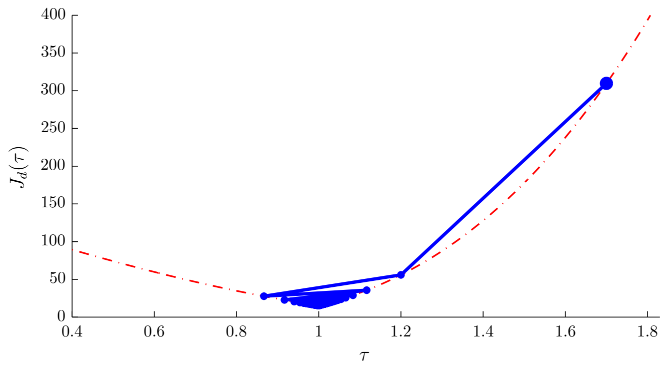

Fig. 6. Algorithm 5.1 shows, as expected, a slow convergence when applied to Example 5.1 with $\tau^{(0)} = 1.7$ and $\mathrm{tol}_\tau = 5 \cdot 10^{-4}$.

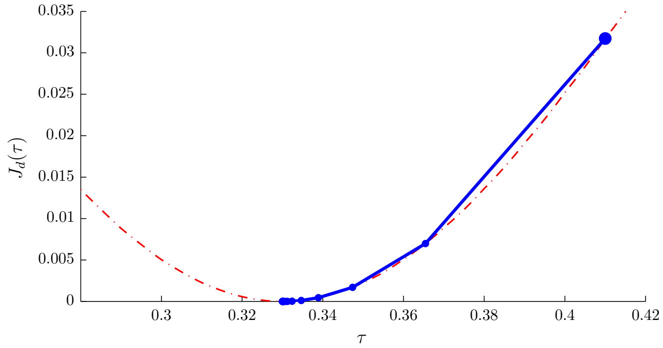

Fig. 7. Steps of the subgradient algorithm applied to Example 5.2 with the optimal choice of step sizes and starting point $\tau^{(0)} = 0.41$.

$\tau^* = 1.0$ in accordance with the nonsmooth optimality conditions: zero (interpreted as a horizontal straight line supporting the graph of $J_d(\tau)$ in $\tau = \tau^*$) is in the subgradient of $J_d(1.0)$. We apply Algorithm 5.1 with the simple choice of step sizes, $s_k = 1/(k+1)$ and choose tolerances $\mathrm{tol}_g = \mathrm{tol}_\tau = 5 \cdot 10^{-4}$. Starting with $\tau^{(0)} = 1.7$, the algorithm terminates after almost 2000 steps because the change in $\tau^{(k)}$ is less than tol_τ. The best, i.e. minimal $J_d^{(k)}$ has already occurred at step 981. This example shows the bad convergence of this simple algorithm (also cf. Fig. 6) and the need for improvement, especially in step size control. Assuming we knew the optimal value of J, the optimal step size strategy (13) needs only six steps to find the optimal switching time.

Example 5.2 (Hybrid double pendulum): As it has been already seen in Example 3.2, the cost function chosen for the hybrid double pendulum is smooth at the optimum (cf. again Fig. 2), so the (normal) stopping criterion $\|g_k\| \leq \mathrm{tol}_g$ of Algorithm 5.1 may apply now. Recall that we had $J_d(\tau) = \Psi_N(x_N(\tau)) = \left\|(\varphi_{1,N}, \varphi_{2,N})^T - q_{\mathrm{final}}\right\|^2$ with q_{final} generated by the discretized trajectory with $\tau^* = 0.33$. Thus, we have an admissible optimal solution and therefore, the optimal costs are known to be $J_d^* = 0$. We choose $\tau^{(0)} = 0.41$, $\mathrm{tol}_\tau = 10^{-12}$, and $\mathrm{tol}_g = 10^{-8}$. Taking optimal step sizes, as explained above, the algorithm terminates after 27 steps with $\|DJ_d\| = 6.01 \cdot 10^{-9}$ and the optimal solution $\tau^* = 0.33$ up to machine precision (cf. Fig. 7).

VI. CONCLUSIONS

The numerical treatment of STO problems requires a discretization of the switched system's dynamics. It has been shown that this discretization destroys the problem's smoothness. However, the discretized problem remains differentiable almost everywhere and gradient formula based on discrete adjoints have been proposed for the discretized problem. Nevertheless, the nondifferentiable points forbid an application of ordinary optimization techniques, since even the optimal value may be at a kink of the discretized cost function. We therefore proposed a simple subgradient method which has been successfully applied to an example with a nonsmooth optimum and to the STO of a hybrid pendulum.

In future work, more sophisticated nonsmooth optimization techniques such as bundle methods have to replace Algorithm 5.1 to improve convergence even for higher dimensional problems. On the one hand, the generalization to several switching times $\bar{\tau} = (\tau_1, \ldots, \tau_m)$, $m > 0$ is straightforward since the formula derived for the derivatives directly apply to the partial derivatives $\frac{\partial J_d}{\partial \tau_i}$ for $i = 1, \ldots, m$. On the other hand, if we do not forbid coinciding switching points $\tau_i = \tau_{i+1}$, the nondifferentiability caused by these events (cf. [3]), has to be taken into account as well. Thus, it would be interesting to combine the algorithm proposed by Egerstedt et al. [3] with our optimization method for discretized STO problems.

REFERENCES

[1] M. Egerstedt, Y. Wardi, and F. Delmotte, "Optimal control of switching times in switched dynamical systems," in *IEEE Conference on Decision and Control*, 2003, pp. 2138–2143.

[2] M. Alamir and S. A. Attia, "On solving optimal control problems for switched hybrid nonlinear systems by strong variations algorithms," in *Proceedings of 6th IFAC Symposium on Nonlinear Control Systems*, 2004, pp. 558–563.

[3] M. Egerstedt, Y. Wardi, and H. Axelsson, "Transition-time optimization for switched-mode dynamical systems," *IEEE Transactions on Automatic Control*, vol. 51, no. 1, pp. 110–115, 2006.

[4] A. Schild, X. C. Ding, M. Egerstedt, and J. Lunze, "Design of optimal switching surfaces for switched autonomous systems," in *IEEE Conference on Decision and Control*, 2009, pp. 5293–5298.

[5] X. Xu and P. Antsaklis, "Optimal control of switched autonomous systems," in *IEEE Conference on Decision and Control*, 2002, pp. 4401–4406.

[6] ——, "Optimal control of switched systems based on parameterization of the switching instants," *IEEE Transactions on Automatic Control*, vol. 49, no. 1, pp. 2–16, 2004.

[7] T. Caldwell and T. Murphey, "Switching mode generation and optimal estimation with application to skid-steering," *Automatica*, vol. 47, no. 1, pp. 50–64, 2011.

[8] E. Johnson and T. Murphey, "Second-order switching time optimization for nonlinear time-varying dynamic systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1953–1957, 2011.

[9] K. Flaßkamp, T. Murphey, and S. Ober-Blöbaum, "Switching time optimization in discretized hybrid dynamical systems," in *Proccedings of the 51th IEEE Conference on Decision and Control*, 2012, pp. 707 – 712.

[10] C. Lemaréchal, "Nondifferentiable optimization," in *Handbook in OR & MS*, G. L. N. et al., Ed. North-Holland: Elsevier, 1989, vol. 1, pp. 529–572.

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

[12] F. H. Clarke, *Optimization and nonsmooth analysis*. New York: Wiley, 1983.

[13] C. Geiger and C. Kanzow, *Theorie und Numerik restringierter Optimierungsaufgaben*. Berlin: Springer, 2002.