On an Optimization Problem for a Class of Impulsive Hybrid Systems

Sid Ahmed Attia · Vadim Azhmyakov · Joerg Raisch

Abstract This contribution addresses the problem of optimal control for a class of hybrid systems, where discrete transitions are accompanied by instantaneous changes in the continuous state variables, and where these changes can be considered as control variables. Based on a variational approach, necessary conditions of optimality are first established. The problem is then cast as a parametric optimization problem for which gradient information is derived. Finally, we discuss assumptions that guarantee convergence of a conceptual algorithm to a stationary solution. A brief discussion on the main implementation issues is also included.

1 Introduction

For a long time, discrete-continuous dynamical interactions have been recognized as a major challenge in the process control area. The emergence of a hybrid systems modelling framework has provided a new perspective on some important problems.

S. A. Attia (✉) · J. Raisch
TU-Berlin, Fachgebiet Regelungssysteme, EN11-Einsteinufer 17, 10587 Berlin, Germany
e-mail: attia@ieee.org

V. Azhmyakov
Departamento de Control Automatico, CINVESTAV,
Av. I.P.N. 2508 Col. San Pedro Zacatenco, 07360 Mexico, Mexico
e-mail: vazhmyakov@ctrl.cinvestav.mx

J. Raisch
Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1,
39106 Magdeburg, Germany
e-mail: Raisch@control.tu-berlin.de

The ability to operate hybrid systems in an optimal way remains a challenging task. Indeed, for the general setting of hybrid systems, one has to deal not only with the infinite dimensional optimization problems related to the continuous dynamics, but also with a potential combinatorial explosion related to the discrete part.

In this context and with focus on particular classes, many schemes have been proposed to tackle the problem. Some are based on a newly elaborated condition of optimality (see e.g., Shaikh and Caines 2003, 2007; Rantzer 2006), others are more related to semi-classical approaches, (see e.g., Attia et al. 2005; Alamir and Attia 2004; Hedlund and Rantzer 2002; Xu and Antsaklis 2003a). In the last few years, there has been a revived interest in gradient based methods, (see e.g., Lu et al. 1993; Cassandras et al. 1998; Azhmyakov and Raisch 2006; Egerstedt et al. 2006; Axelsson et al. 2006). This fact is due to their intuitive interpretation, reliability and the existence of well established convergence results. It is the aim of this contribution to extend this approach to a particular class of hybrid systems with autonomous switching. This class is characterized by the fact that the discrete transitions are accompanied by instantaneous changes (jumps) in the continuous states and that these state jumps can be considered as control variables. Discrete transitions, or switches, are called autonomous if they cannot be triggered by a discrete control command but depend on the evolution of the continuous states. Hybrid systems with jumps in the continuous states are after referred to as *impulsive hybrid systems*. The problem of interest here is formulated as a sequential problem i.e., for a particular execution the time axis is partitioned into subintervals, in each interval, the dynamics are characterized by a set of ODE's, with transitions being triggered internally (autonomous switches). This is the approach that has been considered since the initial formulation of the corresponding optimal control problem (Clarke and Vinter 1989) and can be seen as a natural way to tackle the problem. The class of impulsive hybrid systems discussed in this paper arises most frequently in the area of process control and may be a result of adding and processing material, (see e.g., Alamir 2006, for a typical benchmark problem). In particular, we have been motivated by a specific chemical process control application, namely a preferential crystallization process used to separate enantiomers (see Elsner et al. 2005 for the physical aspects and Raisch et al. 2005 for more details on the control problems). This process is often operated in cyclic batch mode, where after each cycle, material to be separated is added. This represents an instantaneous change of the system state. The amount of added material is a degree of freedom and can be considered as an additional control input.

Hybrid systems with autonomous switching and non-controlled state jumps have been considered in Xu and Antsaklis (2003b). The objective is to find the sequence of jump instants such that a cost functional is minimized. For that purpose, the authors develop a second order scheme that is further specialized to linear systems in Xu and Antsaklis (2003c). Recently, Verriest and coworkers (Verriest et al. 2004) have considered another class of hybrid systems. This consists of systems with controlled switching where some delay is present on the states and where both the jump magnitudes and instants of switching are considered as control variables. Based on variational arguments, necessary conditions of optimality are derived and used in a first order scheme, see Verriest et al. (2005) for an epidemic control application. Another contribution by the same authors is the class of hybrid systems introduced in Verriest (2006). The class includes systems with variable state space dimensions

where the transitions are time triggered and can thus be viewed as a general class of switched systems. The variation of the the state space dimension is captured by introducing fictitious reset maps. The tools developed can be as well applied to the case where the reset maps are real as exemplified in Mehta et al. (2007) where an optimal control problem is formulated and solved using a variational approach. In this contribution, we consider a related problem where no delay is present on the state and where the switching is autonomous.

The paper is organized as follows: in Section 2 the problem is stated formally. Section 3 is devoted to the statement of the necessary conditions of optimality. In Section 4, the gradient formulas are derived and a conceptual algorithm together with some convergence properties is stated. Finally, some conclusions and suggestions for future work are given in Section 5.

2 Problem formulation

We consider the following class of impulsive hybrid systems with autonomous switching (see e.g., Simic et al. 2009, for the general modelling framework).

Definition 1 An impulsive hybrid system with autonomous switching is a collection

$$\mathscr{H} = (\mathscr{Q}, \ \mathscr{E}, \ \mathscr{X}, \ \mathscr{F}, \ \mathscr{G}, \ \mathscr{R})$$

where

- $\mathscr{Q} = \{q_0, q_1 \ldots, q_Q\}$ is a finite set of locations
- $\mathscr{E} \subseteq \mathscr{Q} \times \mathscr{Q}$ is a set of edges
- $\mathscr{X} = \{\mathscr{X}_q\}_{q \in \mathscr{Q}}$ is a collection of state spaces where, for all $q \in \mathscr{Q}$, $\mathscr{X}_q$ is an open subset of $\mathbb{R}^n$.
- $\mathscr{F} = \{f_q\}_{q \in \mathscr{Q}}$ is a collection of vector fields. For all $q \in \mathscr{Q}$, $f_q : \mathscr{X}_q \to \mathbb{R}^n$
- $\mathscr{G} = \{\mathscr{G}_e\}_{e \in \mathscr{E}}$ is a collection of guards. For all possible transitions $e = (q_i, \ q_j) \in \mathscr{E}$, $\mathscr{G}_e \subset \mathscr{X}_{q_i}$
- $\mathscr{R} = \{\mathscr{R}_e\}_{e \in \mathscr{E}}$ is a collection of reset maps. For all $e = (q_i, \ q_j) \in \mathscr{E}$, $\mathscr{R}_e : \mathscr{G}_e \to 2^{\mathscr{X}_{q_j}}$ where $2^{\mathscr{X}_{q_j}}$ denotes the power set of $\mathscr{X}_{q_j}$

We assume that the vector fields f_q are smooth enough (see assumptions below), and that the sets $\mathscr{G}_e$ are nonempty for all $e \in \mathscr{E}$. An execution is as follows: starting from an initial condition (x_0, q_{i_0}) the continuous state evolves according to the autonomous differential equations

$$\dot{x}(t) = f_{q_{i_0}}(x(t)) \tag{1}$$

The discrete state $q(\cdot) = q_{i_0}$ remains constant as long as the trajectory does not reach a guard $\mathscr{G}_{(q_{i_0}, q_{i_1})}$. In our set up, once this guard is reached, $q(\cdot)$ will switch from q_{i_0} to q_{i_1},[1] at the same time the continuous state gets reset to some value according to the

---

[1]Note that this is different from the standard hybrid automaton framework, where switches are enforced by the continuous state violating invariants.

map $\mathscr{R}_{(q_{i_0}, q_{i_1})}$ and the whole process is repeated. Next, we suppose that the guards can be described by smooth $(n-1)$ dimensional surfaces in $\mathbb{R}^n$

$$\mathscr{G}_e = \{x \mid S_e(x) = 0\}, \qquad \text{for all } e \in \mathscr{E} \tag{2}$$

and that the reset maps are linear maps of the form, for all $e \in \mathscr{E}$,

$$\mathscr{R}_{(q_{i_{k-1}}, q_{i_k})}(x) = x + \theta_k \tag{3}$$

with θ_k belonging to Θ a compact subset of $\mathbb{R}^n$. In our setup, the variables θ_k are degrees of freedom that can be selected by the controller.

The dynamic optimization problem of interest can now be formulated as follows:

Problem 1 Under a fixed switching sequence of locations $\{q_{i_k}\}_k$, solve the following optimization problem

$$\min_{\theta} J[\theta] := \sum_{k=0}^{K} \int_{t_k}^{t_{k+1}} L(x(\tau)) d\tau \tag{4}$$

such that

$$\dot{x}(t) = f_{q(t)}(x(t)) \tag{5}$$

$$q(t) = q_{i_k}, \qquad t \in [t_k, t_{k+1}) \tag{6}$$

$$x(t_0^+) = x(t_0) + \theta_0 \tag{7}$$

$$x\left(t_{k+1}^+\right) = x(t_{k+1}) + \theta_{k+1}, \quad S_{(q_{i_k}, q_{i_{k+1}})}(x(t_{k+1})) = 0 \tag{8}$$

$$\dot{x}(t) = f_{q_{i_K}}(x(t)), \qquad t \in [t_K, t_{K+1}] \tag{9}$$

with $x(t_0) = x_0$, $K \in \mathbb{Z}^+$ is the total number of switches, the vector θ denotes the $n \times (K+1)$ dimensional vector $(\theta'_0 \ldots \theta'_K)'$, $L : \mathbb{R}^n \to \mathbb{R}^+$ is a supplied cost function, t_0 and $t_{K+1} = T$ are both finite and given.

Remark 1 Note that in the problem formulation, the sequence of locations is fixed. Although the sequence of locations that the model defined in Definition 1 will exhibit, is not fixed à priori and will depend on the chosen control inputs (parameters), we concentrate on finding the optimal parameters for a given sequence of locations. This implies that the solution developed in the reminder of this paper will preserve the specified sequence of locations. Additionally, a dynamic programming or a heuristic search approach can be added to generate possible sequences, but this is beyond the scope of the paper. Note also that the switching instants are *not* defined à priori. This follows from the degree of freedom allowed on the state jumps, i.e., variation of these quantities induces variations of the switching instants, see Fig. 1 for an example of an execution within the framework of Problem 1.

We assume the following to hold:

A1   For all q_i in $\mathscr{Q}$, $\mathscr{X}_{q_i} = \mathbb{R}^n$
A2   For all q_i in $\mathscr{Q}$, the functions f_{q_i} are continuously differentiable
A3 L is a twice continuously differentiable function
A4   There exists a constant M such that $\| f_{q_i}(x) \| < M$ for all $x \in \mathscr{X}_{q_i}$ and $q_i \in \mathscr{Q}$

Fig. 1 An example of
execution with $K = 4$ switches
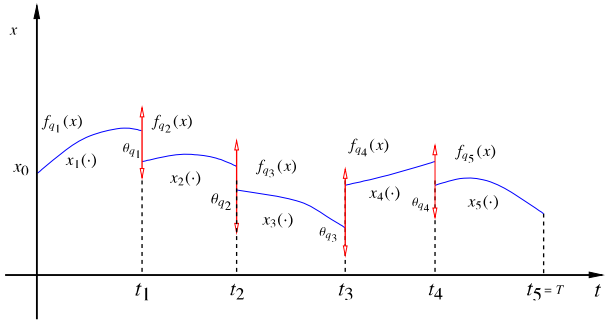
Remark 2 Since the number of switchings and the sequence of locations are fixed,
the problem is well posed. Indeed, sliding and Zeno behaviors are ruled out.

Problem 1 could be seen as a collection of initial value optimization subproblems.
Using classical approaches, the subproblems might be solved separately but nothing
guarantees that the trajectory obtained by concatenation of the different solutions is
the optimal one. The link between the different subproblems is provided by a set of
necessary conditions stated in the next section.

3 Necessary conditions of optimality

In this section, necessary conditions of optimality of a solution to Problem 1 are
derived. The arguments used are of variational type (see e.g., Gelfand and Fomin
(1963) and Bryson and Ho (1975) for basic material on the Euler-Lagrange theory).
Let us first define the Hamiltonian associated to location q_{i_k} as

$$H_{q_{i_k}}(x, \lambda) = L(x) + \lambda' f_{q_{i_k}}(x) \tag{10}$$

where λ denotes the adjoint variables. We then have the following result

Theorem 1 *If θ^* is an interior optimal solution to Problem* 1 *under Assumptions
A1–A4 and $x^*(t)$ is its corresponding state trajectory for $t \in [t_0, T]$, then there exists
a nontrivial adjoint $\lambda^*(t)$ and multipliers π_k^* such that the following equations hold*

$$\dot{\lambda}^*(t)' = -\nabla_x H_{q_{i_k}}^*, \qquad t \in [t_k, t_{k+1}) \tag{11}$$

At the switching instant t_{k+1}, the following jump conditions are satisfied

$$\lambda^*|_{t_{k+1}^+} = \lambda^*|_{t_{k+1}} - \pi_k^* \nabla_x S'_{(q_{i_k}, q_{i_{k+1}})}\Big|_{t_{k+1}} \tag{12}$$

$$H_{q_{i_{k+1}}}^*|_{t_{k+1}^+} = H_{q_{i_k}}^*|_{t_{k+1}} \tag{13}$$

for $k = 0, \ldots K - 1$, with

$$\lambda^*|_{t_{K+1}} = \mathbf{0} \tag{14}$$

and

$$\nabla_\theta J[\theta^*] = \boldsymbol{0} \tag{15}$$

Proof The augmented Lagrangian can be written as

$$\mathscr{L} = \sum_{k=0}^{K} \left[\int_{t_k}^{t_{k+1}} \left(H_{q_{i_k}}(x, \lambda) - \lambda' \dot{x} \right) d\tau + \pi_k S_{(q_{i_k}, q_{i_{k+1}})} |_{t_{k+1}} \right] \tag{16}$$

where for simplicity time dependence is dropped. The increment of $\mathscr{L}$ with respect to x can be written as

$$\Delta_x \mathscr{L} = \mathscr{L}(x+h) - \mathscr{L}(x) \tag{17}$$

where h is a continuously differentiable function of time. Development of the these terms lead to the following

$$\Delta_x \mathscr{L} = \sum_{k=0}^{K} \left[\int_{t_k+dt_k}^{t_{k+1}+dt_{k+1}} \left(H_{q_{i_k}}(x+h, \lambda) - \lambda' \left(\dot{x} + \dot{h} \right) \right) d\tau \right.$$
$$\left. - \int_{t_k}^{t_{k+1}} \left(H_{q_{i_k}}(x, \lambda) - \lambda' \dot{x} \right) d\tau + \pi_k \Delta_x S_{(q_{i_k}, q_{i_{k+1}})} |_{t_{k+1}} \right) \right] \tag{18}$$

with dt_k a small time increment (the existence of which follows from the smoothness assumptions). After integration by part of the second term under the integral sign and rearrangement, Eq. 18 can be written as

$$\Delta_x \mathscr{L} = \sum_{k=0}^{K} \left[\int_{t_k}^{t_{k+1}} \left(H_{q_{i_k}}(x+h, \lambda) - H_{q_{i_k}}(x, \lambda) + \dot{\lambda}' h \right) d\tau \right.$$
$$- \left(H_{q_{i_k}}(x, \lambda) - \lambda' \dot{x} \right) |_{t_k^+} dt_k + \left(H_{q_{i_k}}(x, \lambda) - \lambda' \dot{x} \right) |_{t_{k+1}^+} dt_{k+1}$$
$$\left. - \lambda' h |_{t_k^+}^{t_{k+1}} + \pi_k \Delta_x S_{(q_{i_k}, q_{i_{k+1}})} |_{t_{k+1}} \right] \tag{19}$$

Now, using Taylor's theorem, we obtain (up to first order) the following expression

$$\Delta_x \mathscr{L} = \sum_{k=0}^{K} \left[\int_{t_k}^{t_{k+1}} \left(\nabla_x H_{q_{i_k}} + \dot{\lambda}' \right) h \, d\tau - \left(H_{q_{i_k}}(x, \lambda) - \lambda' \dot{x} \right) |_{t_k^+} dt_k \right.$$
$$\left. + \left(H_{q_{i_k}}(x, \lambda) - \lambda' \dot{x} \right) |_{t_{k+1}} dt_{k+1} - \lambda' h |_{t_k^+}^{t_{k+1}} + \pi_k \left. \frac{\partial S_{(q_{i_k}, q_{i_{k+1}})}}{\partial x} \right|_{t_{k+1}} dx(t_{k+1}) \right] \tag{20}$$

where $dx(t_{k+1})$ is the exact state variation at the instant t_{k+1}. Following simple geometrical arguments, it can be approximated to the first order by the following expression

$$dx(t_{k+1}) = h(t_{k+1}) + \dot{x}(t_{k+1}) dt_{k+1} \tag{21}$$

Using Eq. 21, the first order variation of the augmented Lagrangian $\mathscr{L}$ can be written as

$$\delta_x \mathscr{L} = \sum_{k=0}^{K} \left[\int_{t_k}^{t_{k+1}} \left(\nabla_x H_{q_{i_k}} + \dot{\lambda}' \right) h d\tau - \left(H_{q_{i_k}}|_{t_{k+1}} dt_{k+1} - H_{q_{i_k}}|_{t_k^+} dt_k \right) \right.$$

$$\left. + \lambda|_{t_k^+} dx(t_k) + \left(\pi_k \nabla_x S_{(q_{i_k}, q_{i_{k+1}})} \Big|_{t_{k+1}} - \lambda|_{t_{k+1}} \right) dx(t_{k+1}) \right] \qquad (22)$$

Along the optimal pair (x^*, θ^*) the following is satisfied

$$\delta_x \mathscr{L} = 0 \qquad (23)$$

After rearrangement and using the fact that the optimal problem is without terminal constraints but with fixed initial and final time, the desired result (11)–(14) follows. The necessity of Eq. 15 can be shown analogously. $\square$

Remark 3 Instead of affine reset maps, it is straightforward to derive necessary conditions of optimality for the case of a nonlinear parametrized map of the form $x(t_k^+) = \psi(x(t_k), \theta)$, provided that some smoothness requirements are satisfied. In this case, the gradient of ψ would have appeared in the above conditions. On the other hand, the gradient formulas obtained next will be much harder to derive for this case, which will therefore not be pursued further in this contribution.

Remark 4 The proof techniques used here are of the variational type meaning that the derived necessary conditions of optimality are valid for smooth hybrid impulsive systems under the aforementioned assumptions. The characterized minimum is of the weak type in contrast to a strong minimum that has been recently derived for other classes of hybrid systems (see e.g., Sussmann 1999; Shaikh and Caines 2007).

Remark 5 The conditions stated above characterize a local minimum in the sense that the optimal trajectory is compared only to trajectories that have the same switching sequence of locations (Fig. 2).

Fig. 2 A figure showing how the optimal solution to problem 1, under Assumptions A1–A4, looks like. The implementable Algorithm 1 in Section 4 shows how to find the optimal parameters, and thus the optimal trajectory, iteratively by solving a set of initial value problems
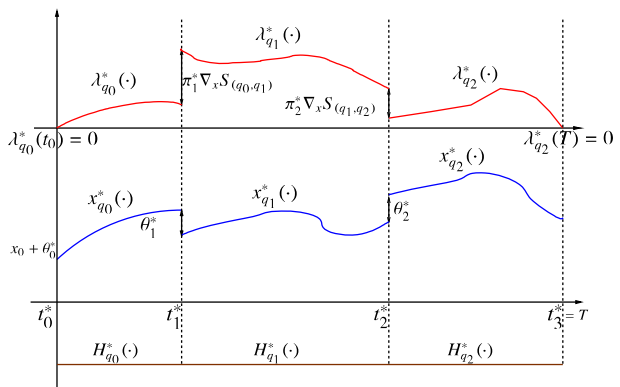
4 A gradient based approach

The basic idea is to cast Problem 1 into a parameter optimization framework and then to compute, as required by the necessary conditions of optimality, the gradient of the cost functional. Gradient descent techniques can then be used to compute the optimal parameter values. Before the gradient formula is stated in a theorem, a lemma concerning the sensitivity of the states is given.

Lemma 1 *The sensitivity* $\Delta_{q_{i_k}}^{\theta_k} x(\cdot)$ *of the state trajectory* $x(\cdot)$, *corresponding to the dynamics* (5)–(9) *under Assumptions* A1–A4, *w.r.t. the vectors* θ_k *can be computed as a solution to the following variational equation*

$$\Delta_{q_{i_k}}^{\theta_k} \dot{x}(t) = \frac{\partial f_{q_{i_k}}}{\partial x} \Delta_{q_{i_k}}^{\theta_k} x(t), \qquad t \in [t_k, t_{k+1}) \tag{24}$$

under the following initial conditions

$$\Delta_{q_{i_0}}^{\theta_0} x(t_0) = I_{n \times n} \tag{25}$$

$$\Delta_{q_{i_{k+1}}}^{\theta_{k+1}} x\left(t_{k+1}^+\right) = \Delta_{q_{i_k}}^{\theta_k} x(t_{k+1}) + I_{n \times n} + \left(f_{q_{i_k}}(x(t_{k+1})) - f_{q_{i_{k+1}}}\left(x\left(t_{k+1}^+\right)\right) \right) \Delta^{\theta_{k+1}} t_{k+1} \tag{26}$$

with

$$\Delta^{\theta_{k+1}} t_{k+1} = - \frac{\left. \nabla_x S_{(q_{i_k}, q_{i_{k+1}})} \right|_{t_{k+1}}}{\left. \nabla_x S_{(q_{i_k}, q_{i_{k+1}})} \right|_{t_{k+1}} f_{q_{i_k}}(x(t_{k+1}))} \Delta_{q_{i_k}}^{\theta_k} x(t_{k+1}) \tag{27}$$

Proof See Appendix 1.   □

Equations 24–27 allow for the computation of the sensitivity trajectories. The first three Eq. 24–26 reflect the sensitivity of the states with respect to the parameters θ and are obtained by introducing smooth parametric variations in the initial conditions corresponding to each location. The perturbations are then carefully propagated through the dynamics of the system. Equation 27 is the sensitivity of the switching instants with respect to the parameters θ. It is obtained using a first order Taylor series approximation. If the switching instants were fixed or freely controllable, the above conditions would simplify. Specifically, Eq. 27 would be zero.

The result in Lemma 1 is used to establish the following theorem.

Theorem 2 *The gradient of the cost functional* J *corresponding to Problem* 1, *under Assumptions* A1–A4, *can be computed as follows for* $k = 0, \ldots K - 1$

$$\nabla_{\theta_0} J = \lambda'|_{t_0} \tag{28}$$

$$\nabla_{\theta_{k+1}} J = \lambda'|_{t_{k+1}} - \pi_{k+1}^* \nabla_x S_{(q_{i_k}, q_{i_{k+1}})}$$

$$\times \left(\Delta_{q_{i_{k+1}}}^{\theta_{k+1}} x\left(t_{k+1}^+\right) + f_{q_{i_{k+1}}}\left(x\left(t_{k+1}^+\right)\right) \Delta^{\theta_{k+1}} t_{k+1} \right) \tag{29}$$

Proof See Appendix 2.   □

Equation 28 is the gradient of the cost functional with respect to the parameter vector θ_0. The same type of equation can be obtained if one considers a classical variational problem (in a non-hybrid setting) where only the initial conditions are allowed to vary. Equation 28 can readily be used to study the impact that the initial condition can have on a classical variational problem and can also be used on its own in a gradient algorithm to derive optimal initial conditions. The theory developed here recovers this important result. Equation 29 is the gradient of the cost with respect to parameter θ_{k+1} and is one of the main contributions of the paper. It depends on the different sensitivity components already derived in Lemma 1. The idea behind the proof is again to introduce variations in the parameters and study their impact on the cost. Using the necessary conditions of optimality obtained in Theorem 1, Eqs. 28 and 29 follow. As will be described in more detail in the forthcoming paragraphs, the computational complexity in this type of problems is considerable. Indeed, the difficulty consists in solving a boundary value problem of a special type. We will now state a conceptual algorithm and show that, under some additional assumptions, it converges to the infimum of the optimization Problem 1 under Assumptions A1–A4.

Conceptual Algorithm 1

1. *Choose an admissible parameter vector $\theta^{(0)} = \left(\theta_0^{(0)}, \theta_1^{(0)}, \ldots, \theta_K^{(0)}\right)$ and set $l = 0$*
2. *Compute the trajectory $x^{(l)}(\cdot)$ and the corresponding adjoint $\lambda^{(l)}(\cdot)$ such that the conditions (12)–(14) are satisfied,*
3. *Update the parameter vector $\theta^{(l)}$ using the gradient information Eq. 28 and 29 in a gradient projection algorithm, set $l := l + 1$ and go to step 2*

We can now state the following result

Proposition 1 *If θ^* is an accumulation point of the sequence $\{\theta^{(l)}\}_l$ generated by the Conceptual Algorithm 1, then it is a stationary point i.e., $\nabla J[\theta^*] = 0$.*

Proof The proof of Proposition 1 follows using descent properties (see e.g., Bertsekas 1995; Pshenichny and Danilin 1982).   □

In the following paragraph, some important implementation issues are discussed.
Denote by P_Θ the projection operator on the set Θ. An implementable version of the preceding Conceptual Algorithm can be stated as follows

Algorithm 1

Step 0   *Choose parameters β, μ as positive real numbers from the set $(0, 1)$, a small positive real number ε and an admissible parameter vector $\theta^{(0)}$. Set $l = 0$*

Step 1   *Compute the trajectory $x^{(l)}(\cdot)$ by forward integrating the state Eqs. 5–9 under the specified initial condition. Let $t^{(l)}$ be the resulting sequence of switching instants $\{t_k\}_k$*

Step 2   *Compute the sensitivity trajectories using Eq. 24 and the corresponding initial conditions (25) and (26) and Eq. 27*

Step 3   *Backward integrate the adjoint $\lambda^{(l)}(\cdot)$ using Eq. 11 with the terminal condition (14). At the switching instants $t^{(l)}$ compute the multipliers $\pi_k^{(l)}$ in Eq. 12 such that Eq. 13 is enforced. Update the adjoint at the switching instants $t^{(l)}$ using Eq. 12 and the so far computed multipliers $\pi^{(l)}$.*

Step 4   *Compute the gradient using Eqs. 28 and 29. Update the parameter vector $\theta^{(l+1)}$*

$$\theta^{(l+1)} = P_\Theta \left(\theta^{(l)} - \gamma^{(l)} \nabla_\theta J \right) \qquad (30)$$

with $\gamma^{(l)} = \mu^{j_l}$ and j_l as the smallest nonnegative integer j satisfying the following inequality- **Armijo step size rule**

$$J\left[\theta^{(l)} - \mu^j \nabla_\theta J\right] - J\left[\theta^{(l)}\right] \leq -\beta\mu^j \|\nabla_\theta J\|^2 \qquad (31)$$

Step 5   *If $J[\theta^{(l)}] - J[\theta^{(l+1)}] \leq \varepsilon$ Then STOP Else set $l := l+1$ and go to Step 1*

A good choice of the algorithm parameters β, μ and ε depends on the problem at hand. Numerical experience has shown the universality of some values (see e.g., Polak 1971) for indications. The computation in Step 1 involves the solution of $(K+1)$ Initial Value Problems (IVP). Particular attention should be paid to the detection of the switching instants. This can be done using the event location capabilities of Matlab IVP solvers (Shampine and Thompson 2000). Step 2 involves the solution of a linear time varying system that should pose no major difficulties. In Step 3, the multipliers are computed such that the Hamiltonian continuity condition is enforced. Recall here that a closed form solution of the multipliers can be found by combining Eqs. 12 and 13. Step 4 is the costliest part of the algorithm. Indeed, evaluation of the cost functional in the right-hand side of inequality (31) makes internal calls to Step 1–Step 3. However, the number of such calls is finite. In Figs. 3 and 4, some iterations of Algorithm 1 applied to a first-order system are shown schematically. The rate of decay of the cost is provided by Eq. 31. The rate is large whenever the gradient is large and this happens usually during the first iterations of the algorithm (Figs. 5 and 6).


Fig. 3 Iteration 0 of the algorithm under $K = 2$ switches for a generic example
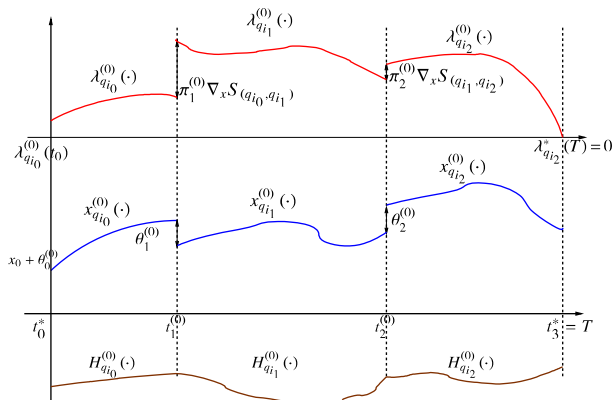
Fig. 4 Iteration 1 of the algorithm under $K = 2$ switches for a generic example. Note the change of the switching instants compared to those in Fig. 3. See Remark 1 for a thorough discussion
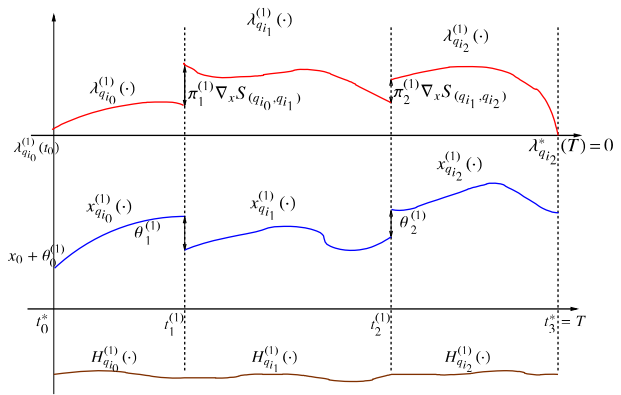
Fig. 5 Iteration m of the algorithm under $K = 2$ switches for a generic example. Note that the initial value of the adjoint is converging to the point **0**, this is a direct consequence of the necessary conditions of optimality (see Eqs. 15 and 28)
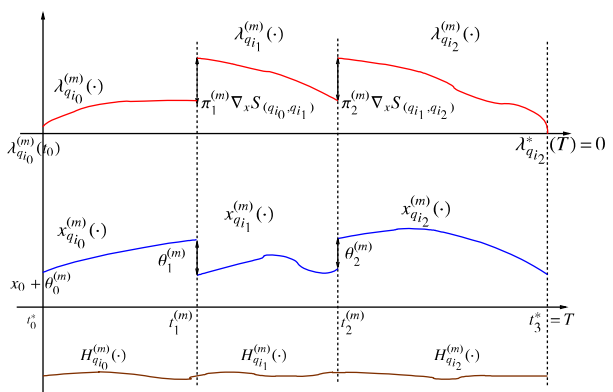
Fig. 6 Optimal solution to problem 1. For the last iteration, all the necessary conditions of optimality (see Eqs. 11–15) are satisfied
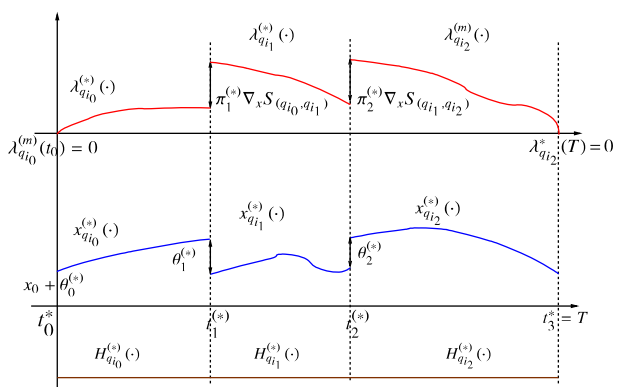
5 Conclusions

This paper addresses optimization problems for a class of hybrid systems arising fre-
quently in the process industries. Using variational principles, necessary conditions
of optimality and a gradient formula are derived. A conceptual algorithm is then
presented together with a convergence analysis and a discussion of implementation
issues. A topic of current research is the study of the cyclic operation of the system
i.e., the switching sequence of locations is periodic. We expect that a cycle-to-cycle
improvement is possible and could be used to further reduce the complexity of the
developed approach.

Appendix 1: Proof of Lemma 1

Proof Let ρ_k be an arbitrary but a fixed vector in $\mathbb{R}^n$ with $k \in \mathcal{K} = \{0, \ldots, K\}$.
Define

$$\theta_k(\varepsilon) = \theta_k + \varepsilon\rho_k \tag{32}$$

Where ε is an arbitrary small real number. Since the trajectory is absolutely continu-
ous between any two consecutive switching instants, we have

$$x(t) = x(t_{k+1}) + \theta_{k+1} + \int_{t_{k+1}}^{t} f_{q_{i_{k+1}}}(x(\tau))d\tau \tag{33}$$

Now, the introduction of the perturbation (32) in the jump map propagates through
the whole trajectory and the following equation holds

$$x(t; \varepsilon) = x(t_{k+1}(\varepsilon); \varepsilon) + \theta_{k+1} + \varepsilon\rho_{k+1} + \int_{t_{k+1}(\varepsilon)}^{t} f_{q_{i_{k+1}}}(x(\tau; \varepsilon))d\tau \tag{34}$$

By differentiating the preceding Eq. 34 with respect to ε, we obtain the following
expression

$$\frac{\partial x(t; \varepsilon)}{\partial \varepsilon} = \frac{\partial x}{\partial \varepsilon}(t_{k+1}(\varepsilon); \varepsilon) + \rho_{k+1} - f_{q_{i_{k+1}}}\left(x\left(t_{k+1}^{+}(\varepsilon); \varepsilon\right)\right)\frac{d}{d\varepsilon}t_{k+1}(\varepsilon)$$
$$+ \int_{t_{k+1}(\varepsilon)}^{t} \frac{\partial f_{q_{i_{k+1}}}}{\partial x}(x(\tau; \varepsilon))\frac{\partial x}{\partial \varepsilon}(\tau; \varepsilon)d\tau \tag{35}$$

From the preceding switching instant, we have

$$\frac{\partial x}{\partial \varepsilon}(t_{k+1}(\varepsilon); \varepsilon) = \frac{\partial x}{\partial \varepsilon}(t_k(\varepsilon); \varepsilon) + \rho_k + f_{q_{i_k}}(x(t_{k+1}(\varepsilon); \varepsilon))\frac{d}{d\varepsilon}t_{k+1}(\varepsilon)$$
$$- f_{q_{i_k}}(x(t_k^{+}(\varepsilon); \varepsilon))\frac{d}{d\varepsilon}t_k(\varepsilon) + \int_{t_k(\varepsilon)}^{t_{k+1}(\varepsilon)} \frac{\partial f_{q_{i_k}}}{\partial x}(x(\tau; \varepsilon))\frac{\partial x}{\partial \varepsilon}(\tau; \varepsilon)d\tau \tag{36}$$

And

$$\left.\frac{\partial x}{\partial \varepsilon}(t; \varepsilon)\right|_{t=t_{k+1}(\varepsilon)} = \frac{\partial x}{\partial \varepsilon}(t_k(\varepsilon); \varepsilon) + \rho_k - f_{q_{i_k}}(x(t_k^+(\varepsilon); \varepsilon)) \frac{d}{d\varepsilon} t_k(\varepsilon)$$

$$+ \int_{t_k(\varepsilon)}^{t_{k+1}(\varepsilon)} \frac{\partial f_{q_{i_k}}}{\partial x}(x(\tau; \varepsilon)) \frac{\partial x}{\partial \varepsilon}(\tau; \varepsilon) d\tau \qquad (37)$$

By using Eqs. 36 and 37, Eq. 35 can be written as follows

$$\frac{\partial x}{\partial \varepsilon}(t; \varepsilon) = \left.\frac{\partial x}{\partial \varepsilon}(t; \varepsilon)\right|_{t=t_{k+1}(\varepsilon)} + \rho_{k+1} + f_{q_{i_k}}(x(t_{k+1}(\varepsilon); \varepsilon)) \frac{d}{d\varepsilon} t_{k+1}(\varepsilon)$$

$$- f_{q_{i_{k+1}}}\left(x\left(t_{k+1}^+(\varepsilon); \varepsilon\right)\right) \frac{d}{d\varepsilon} t_{k+1}(\varepsilon) + \int_{t_{k+1}(\varepsilon)}^{t} \frac{\partial f_{q_{i_{k+1}}}}{\partial x}(x(\tau; \varepsilon)) \frac{\partial x}{\partial \varepsilon}(\tau; \varepsilon) d\tau$$

$$(38)$$

Evaluation of the last expression (38) at $\varepsilon = 0$ and using the following notation for any $t \in [t_k, t_{k+1}]$ and $k \in \mathcal{K}$

$$\Delta_{q_{i_k}}^{\varepsilon} x(t) = \Delta_{q_{i_k}}^{\theta_k} x(t) \rho_k = \left.\frac{\partial x(t; \varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0}, \quad \Delta_{q_{i_k}} \dot{x}(t) = \frac{\partial f_{q_{i_k}}}{\partial x} \Delta_{q_{i_k}}^{\varepsilon} x(t) \qquad (39)$$

and

$$\Delta^{\varepsilon} t_k = \Delta^{\theta_k} t_k \rho_k = \left.\frac{d}{d\varepsilon} t_k(\varepsilon)\right|_{\varepsilon=0} \qquad (40)$$

We then have

$$\Delta_{q_{i_{k+1}}}^{\varepsilon} x(t) = \Delta_{q_{i_k}}^{\varepsilon} x(t_{k+1}) + \rho_{k+1} + \left(f_{q_{i_k}}(x(t_{k+1})) - f_{q_{i_{k+1}}}\left(x\left(t_{k+1}^+\right)\right) \right) \Delta^{\varepsilon} t_{k+1}$$

$$+ \int_{t_{k+1}}^{t} \frac{\partial f_{q_{i_{k+1}}}}{\partial x}(x(\tau; \varepsilon)) \Delta_{q_{i_{k+1}}}^{\varepsilon} x(\tau) d\tau \qquad (41)$$

With

$$\Delta_{q_{i_0}}^{\varepsilon} x(t) = \rho_0 + \int_{t_0}^{t} \frac{\partial f_{q_{i_0}}}{\partial x}(x(\tau; \varepsilon)) \Delta_{q_{i_0}}^{\varepsilon} x(\tau) d\tau \qquad (42)$$

By using Eq. 39 and the arbitrariness of ρ_k, the desired result (24)–(26) follows and ends the first part of the proof. The second part of the result consists in proving Eq. 27, this follows from arguments of differentiability of the switching surfaces. Since the perturbed trajectory will, by assumption, hit the surface, we then have

$$S_{(q_{i_k}, q_{i_{k+1}})}(x(t_{k+1}(\varepsilon); \varepsilon)) = 0 \qquad (43)$$

Expansion around the pertubed switching instant and by neglecting higher order terms gives

$$S_{(q_{i_k}, q_{i_{k+1}})}(x(t_{k+1}(\varepsilon); \varepsilon)) = \nabla_x S_{(q_{i_k}, q_{i_{k+1}})} f_{q_{i_k}} \frac{d}{d\varepsilon} t_{k+1}(\varepsilon)$$

$$+ \nabla_x S_{(q_{i_k}, q_{i_{k+1}})} \frac{\partial x}{\partial \varepsilon} = 0 \qquad (44)$$

Where all the right hand side terms are evaluated at t_{k+1}. From the last equation (Eq. 44), one can extract the desired expression which gives when evaluated at $\varepsilon = 0$

$$\Delta^{\varepsilon} t_{k+1} = - \frac{\nabla_x S_{(q_{i_k}, q_{i_{k+1}})} \big|_{t_{k+1}}}{\nabla_x S_{(q_{i_k}, q_{i_{k+1}})} \big|_{t_{k+1}} f_{q_{i_k}} (x(t_{k+1}))} \Delta^{\varepsilon}_{q_{i_k}} x(t_{k+1}) \tag{45}$$

After simplification, the desired result follows and ends the proof.   □

Appendix 2: Proof of Theorem 2

Proof Let again ρ_k be an arbitrary but a fixed vector in $\mathbb{R}^n$ with $k \in \mathcal{K} = \{0, \ldots, K\}$. Define

$$\theta_k(\varepsilon) = \theta_k + \varepsilon \rho_k \tag{46}$$

Where ε is an arbitrary small real number. The pertubed cost $J[\theta(\varepsilon)]$ can be decomposed and written as follows

$$J[\theta(\varepsilon)] = \sum_{k=0}^{K} \left[\int_{t_k(\varepsilon)}^{t_{k+1}(\varepsilon)} \left[H_{q_{i_k}} (x(\tau; \varepsilon), \lambda(\tau)) - \lambda'(\tau) f_{q_{i_k}} (x(\tau; \varepsilon)) \right] d\tau \right] \tag{47}$$

Computing the derivative of Eq. 47 with respect to ε, evaluating it at $\varepsilon = 0$ and using Theorem 1 (Eq. 11) we obtain after simplification and rearrangement (using $\lambda|_T = \mathbf{0}$ (see Eq. 14 from Theorem 1), t_0 and T are fixed), we obtain

$$\frac{d}{d\varepsilon} J[\theta(\varepsilon)] \bigg|_{\varepsilon=0} = \lambda'|_{t_0^+} \Delta^{\varepsilon}_{q_{i_0}} x(t_0^+)$$

$$+ \sum_{k=0}^{K-1} \left[\left(H_{q_{i_k}} |_{t_{k+1}} - H_{q_{i_{k+1}}} |_{t_{k+1}^+} \right. \right.$$

$$\left. + \lambda'|_{t_{k+1}^+} f_{q_{i_{k+1}}} \left(x \left(t_{k+1}^+ \right) \right) - \lambda'|_{t_{k+1}} f_{q_{i_k}} (x(t_{k+1})) \right) \Delta^{\varepsilon} t_{k+1}$$

$$\left. + \lambda'|_{t_{k+1}^+} \Delta^{\varepsilon}_{q_{i_{k+1}}} x \left(t_{k+1}^+ \right) - \lambda'|_{t_{k+1}} \Delta^{\varepsilon}_{q_{i_k}} x(t_{k+1}) \right] \tag{48}$$

Simplification of the Eq. 48 can be now carried out using Eq. 41 from Lemma 1 that is written as follows

$$\Delta^{\varepsilon}_{q_{i_{k+1}}} x \left(t_{k+1}^+ \right) = \Delta^{\varepsilon}_{q_{i_k}} x(t_{k+1}) + \rho_{k+1}$$

$$- \left(f_{q_{i_{k+1}}} \left(x \left(t_{k+1}^+ \right) \right) - f_{q_{i_k}} \left(x \left(t_{k+1} \right) \right) \right) \Delta^{\varepsilon} t_{k+1} \tag{49}$$

Extracting the term $\Delta^{\varepsilon}_{q_{i_k}} x(t_{k+1})$ from the preceding Eq. 49 and using it in Eq. 48. This gives after simplification

$$
\left. \frac{d}{d\varepsilon} J[\theta(\varepsilon)] \right|_{\varepsilon=0} = \lambda'|_{t_0^+} \Delta^{\varepsilon}_{q_{i_0}} x\left(t_0^+\right) + \sum_{k=0}^{K-1} H_{q_{i_k}}|_{t_{k+1}} - H_{q_{i_{k+1}}}|_{t_{k+1}^+} + \left(\lambda'|_{t_{k+1}^+} - \lambda'|_{t_{k+1}}\right)
$$

$$
\times \left(\Delta^{\varepsilon}_{q_{i_{k+1}}} x\left(t_{k+1}^+\right) + f_{q_{i_{k+1}}}\left(x\left(t_{k+1}^+\right)\right) \Delta^{\varepsilon} t_{k+1} \right) + \lambda'|_{t_{k+1}} \rho_{k+1} \qquad (50)
$$

Now using Theorem 1 the terms involving the Hamiltonians can be eliminated and along the optimal trajectory the gradient can be written as

$$
\left. \frac{d}{d\varepsilon} J[\theta(\varepsilon)] \right|_{\varepsilon=0} = \nabla_{\theta} J[\theta] \rho = \lambda'\left(t_0^+\right) \Delta^{\theta_0}_{q_{i_0}} x(t_0^+) \rho_0
$$

$$
- \sum_{k=0}^{K-1} \pi_k^* \nabla_x S_{(q_{i_k}, q_{i_{k+1}})} \left(\Delta^{\theta_{k+1}}_{q_{i_{k+1}}} x\left(t_{k+1}^+\right) + f_{q_{i_{k+1}}}\left(x\left(t_{k+1}^+\right)\right) \Delta^{\theta_{k+1}} t_{k+1} \right) \rho_{k+1}
$$

$$
+ \lambda'|_{t_{k+1}} \rho_{k+1}
$$

With $\rho = \left(\rho_0' \, \rho_1' \ldots \rho_K'\right)'$. Direct decomposition and simplification since the arbitrariness of ρ gives the desired result and end the proof. $\qquad\square$

References

Alamir M (2006) A benchmark for optimal control solvers for hybrid nonlinear systems. Automatica 42:1593–1598

Alamir M, Attia SA (2004) An efficient algorithm to solve optimal control problems for nonlinear switched hybrid systems. In: IFAC NOLCOS. Stuttgart, 1–3 September 2004

Attia SA, Alamir M, Canudas de Wit C (2005) Suboptimal control of switched nonlinear systems under location and switching constraints. In: 16th IFAC World Congress. Prague, 3–8 July 2005

Axelsson H, Wardi Y, Egerstedt M (2006) Convergence of gradient-descent algorithms for mode-scheduling problems in hybrid systems. In: Proceedings of the 17th international symposium on mathematical theory of networks and systems. Kyoto, July 2006, pp 625–627

Azhmyakov A, Raisch J (2006) A gradient-based approach to a class of hybrid optimal control problems. In: Proceedings of the conference on analysis and design of hybrid systems ADHS. Alghero, June 2006, pp 89–94

Bertsekas DP (1995) Nonlinear programming. Athena Scientific, Clermont-Ferrand

Bryson AE, Ho Y-C (1975) Applied optimal control: optimization, estimation and control. Hemisphere, Washington, DC

Cassandras CG, Pepyne DL, Wardi Y (1998) Generalized gradient algorithms for hybrid system models of manufacturing systems. In Proceedings of the IEEE conference on decision and control. Tampa, December 1998, pp 2627–2632

Clarke F, Vinter R (1989) Optimal multiprocesses. SIAM J Control Optim 27:1072–1090

Egerstedt M, Wardi Y, Axelsson H (2006) Transition-time optimization for switched-mode dynamical systems. IEEE Trans Automat Contr 51(1):110–115

Elsner MP, Mendez DF, Muslera AE, Seidel-Morgenstern A (2005) Experimental study and simplified mathematical description of preferential crystallisation. Chirality (17):183–195

Gelfand IM, Fomin, SV (1963) Calculus of variations. Prentice-Hall, Englewood Cliffs

Hedlund S, Rantzer A (2002) Convex dynamic programming for hybrid systems. IEEE Trans Automat Contr 47(9):1536–1540

Lu J, Liao LZ, Nerode A, Taylor JH (1993) Optimal control of systems with continuous and discrete states. In: Proceedings of the IEEE conference on decision and control. San Antonio, 15–17 December 1993, pp 2292–2297

Mehta TR, Yeung D, Verriest EI, Egerstedt M (2007) Optimal control of multi-dimensional, hybrid ice-skater model. In: American control conference. New York, 11–13 July 2007, pp 2787–2792

Polak E (1971) Computational methods in optimization: a unified approach. Academic, London

Pshenichny BN, Danilin YM (1982) Numerical methods in extremal problems. Mir, Grand Forks

Raisch R, Vollmer U, Angelov I (2005) Control problems in batch crystallization of enantiomers. In: Computer Methods and Systems, CMS. Krakow, November 2005

Rantzer, A (2006) On relaxed dynamic programming in switching systems. IEE Proc Control Theory Appl 153(5):567–574

Shaikh MS, Caines, PE (2003) On the optimal control of hybrid systems: optimization of trajectories, switching times, and location schedules. In: HSCC 2003. Prague, 3–5 April 2003

Shaikh MS, Caines, PE (2007) On the hybrid optimal control problem: theory and algorithms. IEEE Trans Automat Contr 52(9):1587–1603

Shampine LF, Thompson S (2000) Event location for ordinary differential equations. Comp Math Appl 39(5–6):43–54

Simic SN, Johansson KH, Sastry S, Lygeros J (2009) Towards a geometric theory of hybrid systems. In: Lynch N, Krogh B (eds) HSCC 2000. Lecture notes in computer science. Springer, Berlin Heidelberg New York, pp 421–436

Sussmann HJ (1999) A nonsmooth hybrid maximum principle. In: Aeyels D, Lamnabhi-Lagarrigue F, van der Schaft AJ (eds) Stability and stabilization of nonlinear systems. Lecture notes in control and information sciences, vol 246. Springer, Berlin Heidelberg New York, pp 325–354

Verriest EI (2006) Multi-mode multi-dimensional systems. In: Seventeenth International Symposium on MTNS, MTNS2006. Kyoto, 24–28 July 2006

Verriest E, Delmotte F, Egerstedt M (2004) Optimal impulsive control of point delay systems with refractory period. In: Proceedings of the 5th IFAC workshop on time delay systems. Leuven, September 2004

Verriest E, Delmotte F, Egerstedt M (2005) Control of epidemics by vaccination. In: Proceedings of the American control conference. Portland, June 2005, pp 985–990

Xu X, Antsaklis PJ (2003a) Results and perspectives on computational methods for optimal control of switched systems. In: Maler O, Pnueli A (eds) HSCC 2003. Lecture notes in computer science. Springer, Berlin Heidelberg New York, pp 540–555

Xu A, Antsaklis P (2003b) Optimal control of hybrid autonomous systems with state jumps. In Proceedings of the American control conference. Denver, June 2003, pp 5191–5196

Xu X, Antsaklis P (2003c) Quadratic optimal control problems for hybrid linear autonomous systems with state jumps. In: Proceedings of the American control conference. Denver, June 2003, pp 3393–3398

Sid Ahmed Attia received his Bsc in Electrical Engineering from the University of Sciences and Technology-Oran (Algeria), in 1999. He got the Msc and PhD degrees both in control engineering from the Institut National Polytechnique de Grenoble (France) respectively in 2002 and 2005. Since 2006 he is with the Control Systems Group at Technische Universitaet Berlin (Germany). His research interests are in the field of computational optimal control, hybrid systems and control of and over networks.

Vadim Azhmyakov received his M.S. degree in Control and Aeromechanics from Faculty for Aerospace Engineering the Moscow Technical University "N. E. Bauman" (Russia) in 1989. His Ph.D. degree in Control Engineering has been completed in 1994 at the Institute of Control Sciences (Moscow, Russia). In addition, V. Azhmyakov received his Habilitation in Applied Mathematics from the University of Greifswald (Germany) in 2006. From 1996 to 1999, he was a Lecturer at the Faculty of Mathematical Statistics, University of Bayreuth (Germany). From 1999 until 2005 he was an Assistant Professor at the Faculty for Numerical Analysis, University of Greifswald (Germany). From 2005 until 2006 he was a Senior Researcher at the Department of Electrical Engineering, University of Magdeburg (Germany) and from 2006 until 2007 an Assistant Professor at the Department of Electrical Engineering, Technische Universitaet Berlin (Germany). Since 2007 he is with the Department of Control Automation of the CINVESTAV, Mexico City (Mexico). His research interests include hybrid and conventional optimal control, computational methods in control and optimization, nonlinear and convex analysis.

Joerg Raisch is a professor at Technische Universitaet Berlin (Germany), where he heads the Control Systems Group within the Department of Electrical Engineering and Computer Science. He is also head of the Systems and Control Theory Group at the Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg (Germany). He studied Engineering Cybernetics and Control Systems at Stuttgart University and UMIST, Manchester (UK). He got his Ph.D. and "Habilitation", both from Stuttgart University, in 1991 and 1998, respectively. From 1991–1993 he was a postdoc in the Systems Control Group at the University of Toronto (Canada). From 2000–2006 he was a professor at the Otto-von-Guericke University Magdeburg, where he headed the automatic control lab. His research interests are in hybrid systems and hierarchical control and include biomedical control and chemical process control applications.