Math Meth Oper Res (2011) 73:363–380
DOI 10.1007/s00186-011-0355-4

ORIGINAL ARTICLE

Combinatorial integral approximation

Sebastian Sager · Michael Jung · Christian Kirches

Received: 22 February 2010 / Accepted: 18 August 2010 / Published online: 16 April 2011

Abstract We are interested in structures and efficient methods for mixed-integer nonlinear programs (MINLP) that arise from a *first discretize, then optimize* approach to time-dependent mixed-integer optimal control problems (MIOCPs). In this study we focus on combinatorial constraints, in particular on restrictions on the number of switches on a fixed time grid. We propose a novel approach that is based on a decomposition of the MINLP into a NLP and a MILP. We discuss the relation of the MILP solution to the MINLP solution and formulate bounds for the gap between the two, depending on Lipschitz constants and the control discretization grid size. The MILP solution can also be used for an efficient initialization of the MINLP solution process. The speedup of the solution of the MILP compared to the MINLP solution is considerable already for general purpose MILP solvers. We analyze the structure of the MILP that takes switching constraints into account and propose a tailored Branch and Bound strategy that outperforms *cplex* on a numerical case study and hence further improves efficiency of our novel method.

Keywords MINLP · MIOCP · MILP · Optimal control · Integer programming

Mathematics Subject Classification (2000) 90C11 · 90C30 · 49J15 · 90C57

1 Introduction

The main motivation for this paper are mixed-integer optimal control problems (MIOCPs) in ordinary differential equations (ODE) of the following form. We assume that

S. Sager (✉) · M. Jung · C. Kirches
Interdisciplinary Center for Scientific Computing, University of Heidelberg, INF 368,
69120 Heidelberg, Germany
e-mail: sebastian.sager@iwr.uni-heidelberg.de

one of the controls needs to take binary values and can only change these values on a prefixed time grid

$$0 = t_1 < \cdots < t_{n_t+1} = t_f, \tag{1}$$

which we will use for a discretization of the control in a first discretize, then optimize approach. For the sake of notational simplicity we consider a problem with linearly entering piecewise constant binary control functions,

$$\omega_k(t) = p_{k,i}, \quad t \in [t_i, t_{i+1}], \quad k = 1 \ldots n_\omega, \; i = 1 \ldots n_t \tag{2}$$

with $p_{k,i} \in \{0, 1\}$. We want to minimize a Mayer term

$$\min_{x,p} \Phi(x(t_f)) \tag{3a}$$

over the differential states $x(\cdot)$ and the discretized binary control p subject to the n_x-dimensional ODE system

$$\dot{x}(t) = f_0(x(t)) + \sum_{k=1}^{n_\omega} f_k(x(t)) \; p_{k,i}, \quad t \in [t_i, t_{i+1}], \tag{3b}$$

fixed initial values

$$x(0) = x_0, \tag{3c}$$

integrality of the control function $\omega(\cdot)$

$$p_{k,i} \in \{0, 1\}, \quad k = 1 \ldots n_\omega, \; i = 1 \ldots n_t, \tag{3d}$$

and switching constraints

$$\sum_{i=1}^{n_t-1} |p_{k,i+1} - p_{k,i}| \le \sigma_{k,\max}, \quad k = 1 \ldots n_\omega. \tag{3e}$$

Note that the generalization towards the more general case in which $\omega(\cdot)$ enters in a nonlinear way into the right-hand side can be achieved by means of an SOS1 constraint. Also additional continuous controls, path constraints, or multi-stage formulations can be included, compare the results in Sager et al. (2009) and Sager (2009). For the sake of notational simplicity, however, we concentrate on the special case stated above.

Although in practice we will use a simultaneous approach, e.g., collocation (Kameswaran and Biegler 2006) or direct multiple shooting (Leineweber et al. 2003), we will consider the differential states as dependent variables in the theoretical part that can be determined uniquely, whenever the controls are fixed. This transforms (3) into a MINLP with finitely many degrees of freedom. The difference to MIOCPs as they are defined, e.g., in Sager et al. (2009) and Sager (2009) are the additional switching restriction (3e) and the fixed time grid (1) which do not allow the usage of a switching time optimization. More remotely related is the question of the maximum number of switches for equivalent reachable sets. For a special case of a switched system it is shown in Sharon and Margaliot (2007) that 4 switches are enough. A counterexample based on Fuller's phenomenon is given in Margaliot (2007). However, these approaches are based on continuous time, not on fixed switching grids. Therefore we focus on combinatorial approaches, i.e., integer programming, in this paper.

Progress in mixed-integer linear programming (MILP) started with the fundamental work of Dantzig and coworkers on the Traveling Salesman problem in the 1950s. Since then, enormous progress has been made in areas such as *linear programming* (and especially in the *dual simplex* method that is the core of almost all MILP solvers because of its restart capabilities), in the understanding of *branching rules* and more powerful selection criteria such as *strong branching*, the derivation of tight *cutting planes*, novel *preprocessing* and *bound tightening procedures*, and of course the computational advances roughly following Moore's law. For specific problem classes problems with millions of integer variables can now be routinely solved (Applegate et al. 2009). Also generic problems can often be solved very efficiently in practice, despite the known exponential complexity from a theoretical point of view (Bixby et al. 2004).

The situation is different in the field of Mixed-Integer Nonlinear Programming (MINLP). Only at first sight many properties of MILP seem to carry over to the nonlinear case. Restarting nonlinear continuous relaxations within branching trees is essentially more difficult than restarting linear relaxations (which some global solvers also use for nonlinear problems), as no dual algorithm comparable to the dual simplex is available in the general case. Nonconvexities lead to local minima and do not allow for easy calculation of subtrees, which is important to avoid an explicit enumeration. Additionally, nonlinear solvers are slower and less robust than LP solvers. However, the last decade saw great progress triggered by cross-disciplinary work of integer and nonlinear optimizers, resulting in generic MINLP solvers, e.g., Abhishek et al. (2006) and Bonami et al. (2009), or efficient heuristics such as the Feasibility Pump Bonami et al. (2009). Most of them, however, still require the underlying functions to be convex. Comprehensive surveys on algorithms and software for convex MINLPs are given in Grossmann (2002) and Bonami et al. (2009). Recent progress in the solution of nonconvex MINLPs is in most cases based on methods from global optimization, in particular convex under- and overestimation. See, e.g., Belotti et al. (2009), Floudas et al. (2005), Tawarmalani and Sahinidis (2002) for references on general under—and overestimation of functions and sets. In our study we use the solver *Bonmin* Bonami et al. (2009) for comparison and show how important it is to exploit problem-class specific structures.

The basic idea of our new approach to solve problem (3) consists of a decomposition of the MINLP into an NLP and an MILP, which we can both solve comparatively efficiently. This idea is related to ideas of Burgschweiger et al. (2008). The authors reformulate the MIOCP as a large-scale, structured nonlinear program (NLP) and solve a small scale linear integer program on a second level to approximate the calculated continuous aggregated output of all pumps in a water works. However, their decomposition is tailored to the special structure of the water network application, while our approach targets generic problems of the form (3).

To guarantee error bounds on the obtained solution compared to the MINLP solution, we revise some theoretical results in Sect. 2. In Sect. 3 we will discuss our new method that is based on a combinatorial approximation of the integral over control deviations. In Sect. 4 we analyze the structure of the MILP and provide a structure exploiting Branch and Bound algorithm. In Sect. 5 we present results for a numerical benchmark example. Finally, we will conclude and give an outlook in Sect. 6.

2 Approximation results

We revise some results from Sager et al. (2011). The following theorem states how the difference of solutions to the initial value problem (3b–3c) depends on the integrated difference between control functions and the difference between the initial values. From now on we will often leave the argument (t) away for the sake of notational simplicity. In the following $\|\cdot\|$ will denote the maximum norm.

Theorem 1 *Let $x(\cdot)$ and $y(\cdot)$ be solutions of the initial value problems*

$$\dot{x}(t) = A(t, x(t)) \cdot \alpha(t), \quad x(0) = x_0, \tag{4a}$$
$$\dot{y}(t) = A(t, y(t)) \cdot \omega(t), \quad y(0) = y_0, \tag{4b}$$

with $t \in [0, t_f]$, for given measurable functions $\alpha, \omega : [0, t_f] \to [0, 1]^{n_\omega}$ and a differentiable $A : \mathbb{R}^{n_x+1} \mapsto \mathbb{R}^{n_x \times n_\omega}$. If positive numbers $C, L \in \mathbb{R}^+$ exist such that for $t \in [0, t_f]$ almost everywhere it holds that

$$\left\| \frac{\mathrm{d}}{\mathrm{d}t} A(t, x(t)) \right\| \le C, \tag{4c}$$
$$\| A(t, y(t)) - A(t, x(t)) \| \le L \| y(t) - x(t) \|, \tag{4d}$$

and $A(\cdot, x(\cdot))$ is essentially bounded by $M \in \mathbb{R}^+$ on $[0, t_f]$, and it exists $\epsilon \in \mathbb{R}^+$ such that for all $t \in [0, t_f]$

$$\left\| \int_0^t \alpha(\tau) - \omega(\tau) \, \mathrm{d}\tau \right\| \le \epsilon \tag{4e}$$

then it also holds

$$\| y(t) - x(t) \| \le (\| x_0 - y_0 \| + (M + Ct)\epsilon) \, e^{Lt} \tag{4f}$$

for all $t \in [0, t_f]$.

As a corollary to Theorem 4, with $A = (f_1 \; f_2 \; \dots \; f_{n_\omega} \; f_0) \in \mathbb{R}^{n_x \times (n_\omega+1)}$ and artificial entries $\alpha_{n_\omega+1} = \omega_{n_\omega+1} = 1$, one obtains that for a differentiable function

$\Phi(\cdot)$ it holds

$$\Phi(x(t_\mathrm{f})) - \Phi(y(t_\mathrm{f})) \le \bar{C}\,\epsilon \tag{5}$$

with problem-dependent, but constant $\bar{C}$. Assume we have solved a relaxed problem for $\omega_k(t) = q_{k,i}, k = 1 \dots n_\omega, t \in [t_i, t_{i+1}]$ for $i = 1 \dots n_\mathrm{t}$ and $q_{k,i} \in [0, 1]$ and obtained an optimal trajectory $(x^*(\cdot), \alpha(\cdot))$ with

$$\alpha_k(t) = q_{k,i}, \quad k = 1 \dots n_\omega,\ t \in [t_i, t_{i+1}]. \tag{6}$$

We can use q to construct a binary control $\omega(\cdot) \in \{0, 1\}^{n_\omega}$ and get an estimate that we can use for assumption (4e). We write $\Delta t_i := t_{i+1} - t_i$ and Δt for the maximum distance between two time points, $\Delta t := \max_{i=1\dots n_\mathrm{t}} \Delta t_i = \max_{i=1\dots n_\mathrm{t}}\{t_{i+1} - t_i\}$. Let then a function $\omega(\cdot) : [0, t_f] \mapsto \{0, 1\}^{n_\omega}$ be defined by

$$\omega_k(t) = p_{k,i}^{\mathrm{SUR}}, \quad k = 1 \dots n_\omega,\ t \in [t_i, t_{i+1}] \tag{7}$$

where the $p_{k,i}^{\mathrm{SUR}}$ are binary values given for $k = 1 \dots n_\omega$ by

$$p_{k,i}^{\mathrm{SUR}} = \begin{cases} 1 & \text{if } \sum_{j=1}^{i} q_{k,j}\Delta t_j - \sum_{j=1}^{i-1} p_{k,j}^{\mathrm{SUR}}\Delta t_j \ge 0.5\Delta t_i \\ 0 & \text{else} \end{cases}. \tag{8}$$

We refer to p^{SUR} as the *Sum Up Rounding* solution (Sager 2005) and define σ_k^{SUR} to be the minimal number for which inequality (3e) holds for p_k^{SUR}. We then have the following estimate on the integral over the difference between the control functions $\alpha(\cdot)$ and $\omega(\cdot)$.

Theorem 2 *Let the functions $\alpha : [0, t_f] \mapsto [0, 1]^{n_\omega}$ and $\omega : [0, t_f] \mapsto \{0, 1\}^{n_\omega}$ be given by* (6) *and* (7, 8)*, respectively. Then it holds*

$$\left\| \int_0^t \alpha(\tau) - \omega(\tau)\, \mathrm{d}\tau \right\| \le \eta$$

with $\eta = 0.5\ \Delta t$.

Note that for the more general case in which the integer control functions enter in a nonlinear way into the differential equations the SUR strategy can be modified to incorporate the SOS1 constraint (Sager et al. 2011). Theorem 2 still holds with a constant η which is a function of n_ω.

3 Approximating the integral over the controls by MILP techniques

The results of Sect. 2 have been used in several ways. Most importantly they imply that, if the control discretization grid is fine enough, no integer gap exists (Sager et al.

2009), because Δt can be chosen arbitrarily small and the estimation carries over to continuous objective and constraint functions. Also, the specific way of constructing a binary solution (7, 8) can be used, e.g., in the adaptive algorithm MINTOC (Sager et al. 2009; Sager 2009). However, both uses require that the constructed binary control is feasible for the original problem. This is not a problem if only constraints on the differential states are present when $\Delta t \to 0$, but constraints of the type (3e) will typically be violated if Δt is small.

Therefore we propose to change the point of view: while before it was argued that the difference between integer and relaxed solution will become arbitrarily small if $\Delta t \to 0$, we now consider Δt to be fixed and allow a larger constant to obtain a feasible solution.

To be able to include constraint (3e) we determine p not by (7, 8), but as the solution of the MILP

$$\begin{aligned}
&\min_{p} \max_{k=1\dots n_\omega} \max_{i=1\dots n_\mathrm{t}} \left| \sum_{j=1}^{i} (q_{k,j} - p_{k,j}) \Delta t_j \right| \\
&\text{subject to} \\
&\sigma_{k,\max} \geq \sum_{i=1}^{n_\mathrm{t}-1} \left| p_{k,i} - p_{k,i+1} \right|, \quad k = 1 \dots n_\omega, \\
&p_{k,i} \in \{0, 1\}, \quad k = 1 \dots n_\omega, \ i = 1 \dots n_\mathrm{t}.
\end{aligned} \tag{9}$$

To get rid of the min max formulation and the absolute values, we introduce slack variables $\eta \in \mathbb{R}$ and $s \in [0, 1]^{n_\omega \times (n_\mathrm{t}-1)}$ and obtain

$$\begin{aligned}
&\min_{\eta, s, p} \quad \eta \\
&\text{subject to} \\
&\eta \geq \sum_{j=1}^{i} (q_{k,j} - p_{k,j}) \Delta t_j, \quad k = 1 \dots n_\omega, \ i = 1 \dots n_\mathrm{t}, \\
&\eta \geq -\sum_{j=1}^{i} (q_{k,j} - p_{k,j}) \Delta t_j, \quad k = 1 \dots n_\omega, \ i = 1 \dots n_\mathrm{t}, \\
&s_{k,i} \geq p_{k,i} - p_{k,i+1}, \quad k = 1 \dots n_\omega, \ i = 1 \dots n_\mathrm{t} - 1, \\
&s_{k,i} \geq -p_{k,i} + p_{k,i+1}, \quad k = 1 \dots n_\omega, \ i = 1 \dots n_\mathrm{t} - 1, \\
&\sigma_{k,\max} \geq \sum_{i=1}^{n_\mathrm{t}-1} s_{k,i}, \quad k = 1 \dots n_\omega, \\
&p_{k,i} \in \{0, 1\}, \quad k = 1 \dots n_\omega, \ i = 1 \dots n_\mathrm{t},
\end{aligned} \tag{10}$$

for fixed control values q that stem from the solution of the relaxed problem (3) and given upper bounds on the number of switches, $\sigma_{k,\max}$.

Although problem (10) is a MILP and thus typically hard to solve, for certain values $\sigma_{k,\max}$ the solution can be calculated in polynomial time using the Sum Up Rounding strategy (7, 8). This is the content of the following theorem. In analogy to the maximal interval length $\Delta t := \max_{i=1\ldots n_t} \Delta t_i$ we also define the minimal one, $\delta t := \min_{i=1\ldots n_t} \Delta t_i$.

Theorem 3 *Assume p^{SUR} to be the solution obtained by Sum Up Rounding (7, 8). The following claims hold for the optimal solution (η^*, s^*, p^*) of the MILP (10):*

$$
\begin{aligned}
&(a)\ \eta^* < 0.5\ \delta t = 0.5 \min_{i=1\ldots n_t} \Delta t_i \\
&\quad \Rightarrow (b)\ p^* = p^{SUR} \\
&\quad \Rightarrow (c)\ \sigma_{k,\max} \geq \sigma_k^{SUR} \quad \forall\, k = 1 \ldots n_\omega \\
&\quad \Rightarrow (d)\ \eta^* \leq 0.5\ \Delta t = 0.5 \max_{i=1\ldots n_t} \Delta t_i
\end{aligned}
$$

where the solution $p^* = p^{SUR}$ in (b) is unique.

Proof *"(a) $\Rightarrow$ (b)". Assume first $\eta^* < 0.5\ \delta t$ and $p^* \neq p^{SUR}$. Then there must exist indices $k \in \{1, \ldots, n_\omega\}$ and $i \in \{1, \ldots, n_t\}$ such that $p^*_{k,j} = p^{SUR}_{k,j}$ for all $j < i$ and $p^*_{k,i} \neq p^{SUR}_{k,i}$.*

We have two cases for the binary variables $p^*_{k,i} \neq p^{SUR}_{k,i}$. If $p^*_{k,i} = 0$ and $p^{SUR}_{k,i} = 1$, then from (8) it follows that

$$\sum_{j=1}^{i} q_{k,j} \Delta t_j - \sum_{j=1}^{i-1} p^{SUR}_{k,j} \Delta t_j \geq 0.5\ \Delta t_i$$

and hence

$$\sum_{j=1}^{i} (q_{k,j} - p^*_{k,j}) \Delta t_j = \sum_{j=1}^{i} q_{k,j} \Delta t_j - \sum_{j=1}^{i-1} p^{SUR}_{k,j} \Delta t_j \geq 0.5\ \Delta t_i. \tag{11}$$

Equivalently, if $p^*_{k,i} = 1$ and $p^{SUR}_{k,i} = 0$ then

$$\sum_{j=1}^{i} q_{k,j} \Delta t_j - \sum_{j=1}^{i-1} p^{SUR}_{k,j} \Delta t_j < 0.5\ \Delta t_i$$

and therefore

$$\sum_{j=1}^{i} (q_{k,j} - p^*_{k,j}) \Delta t_j = -\Delta t_i + \sum_{j=1}^{i} q_{k,j} \Delta t_j - \sum_{j=1}^{i-1} p^{SUR}_{k,j} \Delta t_j < -0.5\ \Delta t_i. \tag{12}$$

As (η^*, s^*, p^*) *is a feasible solution of* (10)*, with* (11) *and* (12) *we have the contradiction*

$$\eta^* \geq \left| \sum_{j=1}^{i} (q_{k,j} - p^*_{k,j}) \Delta t_j \right| \geq 0.5 \; \Delta t_i \geq 0.5 \quad \delta t \tag{13}$$

to the assumption $\eta^* < 0.5\ \delta t$*. Therefore* $p^* = p^{SUR}$*.*

"(b) ⇒ (c)". Assume now $p^* = p^{SUR}$*. As* (η^*, s^*, p^*) *is a feasible solution of* (10)*, the number of switches of* p^{SUR} *given by* $\sigma^{SUR}_{\max}$ *is necessarily at least* $\sigma_{\max}$*, componentwise.*

"(c) ⇒ (d)". If it holds that $\sigma_{k,\max} \geq \sigma^{SUR}_k$ *for all* $k = 1 \dots n_\omega$*, then the vector given by*

$$\begin{aligned} \eta &= 0.5 \Delta t, \\ p &= p^{SUR}, \\ s_{k,i} &= |p_{k,i} - p_{k,i+1}|, \quad k = 1 \dots n_\omega, \; i = 1 \dots n_t - 1 \end{aligned}$$

is a feasible solution of (10) *as follows from Theorem* 2*, and yields hence an upper bound on the objective function value* η^**.* □

Remark 1 The asymmetry in Theorem 3 even for an equidistant grid with $\delta t = \Delta t = \Delta t_i$ is due to the degenerate case where $\eta^* = 0.5 \Delta t$. While p^{SUR} always yields a solution with $\eta^{\text{SUR}} \leq 0.5 \Delta t$, this solution might switch more often than another control resulting in $\eta^* = 0.5 \Delta t$. The easiest example is $q_k = (0.5, 0, \dots, 0)$, which results in $p^{\text{SUR}}_k = (1, 0, \dots, 0)$ with one switch. The same value of $\eta^* = 0.5 \Delta t$ is obtained by $p_k = (0, 0, \dots, 0)$. This is also the optimal solution of the MILP instance with q_k and $\sigma_{k,\max} = 0$ for which p^{SUR} is infeasible, but still $\eta^* = 0.5 \Delta t$.

Remark 2 It holds $\eta^{\text{SUR}} \leq 0.5 \Delta t$, and therefore also the optimal objective function values η^* of MILP (10) decrease, as Δt is decreased. However, this is not necessarily strictly monotonic, as the amount of reduction depends heavily on the values of q and Δt_i.

Theorem 3 is particularly interesting, as we know from (5) and Theorem 2 that if $n_{\text{t}} \to \infty$, then $\Phi(x^{\text{SUR}}) \to \Phi^*$, i.e., the solution obtained with Sum Up Rounding, p^{SUR}, will yield an objective function value that converges against the lower bound Φ^* obtained by solving the relaxed version of (3).

However, the solution p^{SUR} may violate the switching constraint (3e). Hence, solving one of the MILPs yields a compromise between the approximation of the control integral, which has been shown to imply convergence towards the objective's lower bound if the control discretization is refined, and the incorporation of switching constraints—and possibly all other types of linear constraints on p—by means of a mixed-integer linear program.

4 Solving the MILP

The mixed-integer linear program (10) can be solved with standard solvers, such as *cplex*. However, as the structure is generic for all MIOCPs with switching restrictions, we have a closer look at the facets of the convex hull of all feasible points in Sect. 4.1. To speed up computational runtimes we also propose a tailored Branch and Bound strategy in Sect. 4.2.

4.1 Facet defining inequalities

Important insight can be gained by investigating the feasibility polytope. An investigation of min down/up polytopes, for example, can be found in Lee et al. (2004).

The MILP (10) has a specific structure, partly independent of the values of q and $\sigma_{\max}$. To identify the structure—especially the facets—of the convex hull of all feasible points of MILP (10) we use the software-package *PORTA* (Christof and Reinelt 1996; Christof and Löbel). The following constraints define facets of this polytope,

$$\begin{aligned}
s_{k,i} &\geq p_{k,i} - p_{k,i+1}, \quad k = 1 \ldots n_\omega,\ i = 1 \ldots n_\mathrm{t} - 1,\\
s_{k,i} &\geq -p_{k,i} + p_{k,i+1}, \quad k = 1 \ldots n_\omega,\ i = 1 \ldots n_\mathrm{t} - 1,\\
s_{k,i} &\leq p_{k,i} + p_{k,i+1}, \quad k = 1 \ldots n_\omega,\ i = 1 \ldots n_\mathrm{t} - 1,\\
s_{k,i} &\leq 2 - p_{k,i} - p_{k,i+1}, \quad k = 1 \ldots n_\omega,\ i = 1 \ldots n_\mathrm{t} - 1.
\end{aligned} \tag{14}$$

Depending on whether the $\sigma_{k,\max}$ are fixed to a certain value or not, the corresponding facets are different. If $\sigma_{k,\max}$ is free, they read as

$$\sigma_{k,\max} \geq \sum_{i=1}^{n_\mathrm{t}-1} s_{k,i}, \quad k = 1 \ldots n_\omega. \tag{15a}$$

If $\sigma_{k,\max}$ is fixed to an even value, then as

$$\begin{aligned}
\sigma_{k,\max} &\geq p_{k,1} - p_{k,n_\mathrm{t}} + \sum_{i=1}^{n_\mathrm{t}-1} s_{k,i}, \quad k = 1 \ldots n_\omega,\\
\sigma_{k,\max} &\geq p_{k,n_\mathrm{t}} - p_{k,1} + \sum_{i=1}^{n_\mathrm{t}-1} s_{k,i}, \quad k = 1 \ldots n_\omega,
\end{aligned} \tag{15b}$$

and alternatively if $\sigma_{k,\max}$ is fixed to an odd value as

$$\begin{aligned}
\sigma_{k,\max} &\geq 1 - p_{k,1} - p_{k,n_\mathrm{t}} + \sum_{i=1}^{n_\mathrm{t}-1} s_{k,i}, \quad k = 1 \ldots n_\omega,\\
\sigma_{k,\max} &\geq p_{k,1} + p_{k,n_\mathrm{t}} - 1 + \sum_{i=1}^{n_\mathrm{t}-1} s_{k,i}, \quad k = 1 \ldots n_\omega.
\end{aligned} \tag{15c}$$

Table 1 Number of all facets for problems with only one control and all the given $q_{1,j}$ fixed to a certain value, all the $\sigma_{k,\max}$ are free

n_t	Control values $q_{1,j}$ fixed to								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
4	16	18	23	33	29	33	23	18	16
5	21	31	87	189	54	189	87	31	21
6	30	60	745	612	248	612	745	60	30
7	47	150	4838	4840	922	4840	4838	150	47
8	83	899	37470	29884	4212	29884	37470	899	83

Unfortunately, the facets arising from the approximation inequalities

$$\eta \geq \sum_{j=1}^{i} (q_{k,j} - p_{k,j}) \Delta t_j, \quad k = 1 \dots n_\omega, \; i = 1 \dots n_t,$$
$$\eta \geq -\sum_{j=1}^{i} (q_{k,j} - p_{k,j}) \Delta t_j, \quad k = 1 \dots n_\omega, \; i = 1 \dots n_t. \tag{16}$$

cannot be expressed as easily, as far as we know. They mainly depend on the values of q and generally are dense in both $p_{k,i}$ and $s_{k,i}$. Additionally, their number strongly increases with the size of the problem, as can be seen in Table 1. Therefore it is hard to identify structures in the corresponding facets which would possibly enable cutting plane methods.

4.2 Solving the MILPs efficiently

As an alternative to cutting planes we implemented a structure exploiting pure Branch and Bound algorithm. It uses the structure of the approximation inequalities (16) that model the min max formulation.

We branch on controls p and determine s as dependent variables. We branch in increasing order of the time index i in $p_{k,i}$. This way 2 n_ω inequalities are fixed for each i, i.e., all variables $s_{k,j}$ and $p_{k,j}$ with $j \leq i$ are fixed and we can give a new bound on η using constraints (16). Because of this lower bound it is not necessary to solve an LP relaxation.

We will present a short outline of the algorithm. Each node of the branching tree contains the four components

- depth d of the node, i.e., the number of timesteps for which the controls are fixed,
- the fixed control variables $p_{k,j}$ for $j \leq d$,
- the fixed slack variables $s_{k,j}$ for $j \leq d$,
- the corresponding lower bound on η.

Algorithm 1 Combinatorial Branch and Bound

Input : Relaxed controls q, time grid $\{t_i\}$, $i = 1 \dots n_t$, max. numbers of switches $\sigma_{k,\max}$, $k = 1 \dots n_\omega$.
Result: Optimal solution (η^*, s^*, p^*) of (10).
begin
  Create empty priority queue Q ordered by $a.\eta$ (non-decreasing), if equal by $a.d$ (non-increasing). Push an empty node (0, { }, { }, 0.0) into the queue. **while** Q is not empty **do**
    a = top node of Q and remove the node from Q. /*1st solution found is optimal since best-first search is used */
    if $a.d = n_t$ **then**
      Return optimal solution $(a.\eta, a.s, a.p)$.
    /*Create child nodes, use strong branching. */
    else
      forall possible permutations ϕ of $\{0, 1\}^{n_\omega}$ **do**
        Create new node n with $n.d = d + 1$, $n.p = a.p$, $n.s = a.s$. Set $n.p_{k,d+1} = \phi_k$, calculate $n.s_{k,d+1}$. **if** $n.s$ fulfills switching constraint (3e) until time $d + 1$ **then**
          $n.\eta = \max \left\{a.\eta, \max_{k=1}^{n_\omega}\{\pm \sum_{j=1}^{d+1}(q_{k,j} - p_{k,j})\Delta t_j\}\right\}$ Push n into Q.
end

The priority queue in Algorithm 1 models the search strategy, in our case a best-first search (if two nodes have the same objective value, the deeper one is preferred). Note that the algorithm does not solve any relaxed linear programs, but is purely based on efficient branching and constraint/objective evaluation.

5 Numerical results

An open online benchmark library for the problem class of MIOCPs is available at (sages in MIOCP benchmark site, http://mintoc.de). Here we present numerical results for the Lotka-Volterra benchmark fishing problem (Sager (MIOCP benchmark site, http://mintoc.de); Sager 2005) extended with an additional switching constraint (3e),

$$\min_{x,w} \quad x_2(t_f) \tag{17a}$$

$$\text{subject to} \quad \dot{x}_0(t) = x_0(t) - x_0(t)x_1(t) - c_0 x_0(t)\, w(t), \tag{17b}$$

$$\dot{x}_1(t) = -x_1(t) + x_0(t)x_1(t) - c_1 x_1(t)\, w(t), \tag{17c}$$

$$\dot{x}_2(t) = (x_0(t) - 1)^2 + (x_1(t) - 1)^2, \tag{17d}$$

$$x(0) = (0.5, 0.7, 0)^T, \tag{17e}$$

$$w(t) = p_i \in \{0, 1\}, \; t \in [t_i, t_{i+1}], \tag{17f}$$

$$\sigma_{\max} \geq \sum_{i=1}^{n_t-1} |p_{i+1} - p_i|, \tag{17g}$$

with $c_0 = 0.4, c_1 = 0.2, t_f = 12$, and different equidistant grids $\{t_1, \dots, t_{n_t+1}\}$. This problem is particularly suited for our study, because the optimal relaxed solution contains a singular arc (Sager 2005).

The differential equations have been discretized with an implicit Euler method and 10,000 equidistant time steps, independent of the control discretization.

All computational times refer to a two core Intel CPU with 3 GHz and 8GB RAM run under Ubuntu 9.10. We used *Bonmin* 1.2 trunk revision 1601[1] and *cplex* 8.1 with standard options, respectively.

Numerical results are shown in Tables 2 and 3. Here τ is the computing time in seconds, Φ denotes the objective function value. The number of switches of a solution is given by σ, while η is the maximum deviation of the integrated difference between relaxed and integer control over the time horizon. Note, however, that the values of η have been scaled by $\frac{t_\mathrm{f}}{n_\mathrm{t}}$ for better comparability. The upper script *rel* refers to the relaxed version of the OCP (3), *milp* to the solution of the MILP (10) obtained with either *cplex* 8.1 or with our own code (*bb*) as described in Sect. 4.2, and *minlp* to the solution of the MINLP resulting from a discretization of (17) and solution with *Bonmin* 1.2.

We used an upper time limit of 1,800 s, indicated by an * in Table 2 whenever active. If no feasible solution could be found within this upper time limit, this is indicated by an *, otherwise the value of the upper bound feasible solution is listed. In Table 3 a value for $\sigma_{\max}$ and a * indicate that no better solution than the one from the MILP initialization could be found.

5.1 MILP and MINLP solutions

Numerical results for the solution of problem (17) with different upper limits on the number of switchings of p_i between 0 and 1 and different equidistant discretizations (1) are shown in Table 2.

The first rows show the behavior of the solution of the relaxed MINLP (17). As predicted by theory, compare Sect. 2, the objective function values of the relaxed problem Φ^{rel} and the Sum Up Rounding solutions Φ^{SUR} converge towards a Φ^* that is the solution of the non-discretized, relaxed optimal control problem. However, the number of switches $\sigma_{\max}$ of the SUR solution increases significantly. All values of $\frac{1}{\Delta t}\ \eta^{\mathrm{SUR}} = \frac{n_\mathrm{t}}{t_\mathrm{f}}\ \eta^{\mathrm{SUR}}$ are below 0.5, as predicted by Theorem 2. It is interesting to observe that these values approach 0.5 as n_t increases, due to the increased probability to find a maximum close to the upper bound 0.5.

The next blocks show results for the solutions of MILPs and MINLPs corresponding to different upper limits $\sigma_{\max}$. If this limit is large enough, then in accordance with Theorem 3 the MILP and SUR solutions coincide (e.g., $n_\mathrm{t} = 25, \sigma_{\max} \geq 4$). If not, the value of $\frac{1}{\Delta t}\ \eta^{\mathrm{milp}}$ necessarily increases above 0.5. The objective function values Φ^{milp} and Φ^{minlp} will *both* converge against the value of Φ^{rel}, as $n_\mathrm{t} \to \infty$ and $\sigma_{\max}$ large enough. If switching constraints are active, the objective function value is bounded by a constant multiple of η. Although the MILP is not necessarily optimal for the MINLP, it has the advantage to be feasible, to have asymptotic properties, and to be a priori bounded.

As can be observed, the CPU times for the Branch and Bound algorithm are below those of *cplex* (τ^{bb} vs. τ^{cplex}), which in turn are considerably below those of the

[1] using Cbc 2.4stable and Ipopt 3.8stable.

Table 2 Results for Lotka Volterra fishing problem with MILP (10) solved by our structure exploiting Branch and Bound algorithm (*bb*) or *cplex*

n_t	10	20	25	50	80	100	200
τ^{rel}	2.59616	2.61616	2.75217	2.18814	2.25214	2.33214	1.89612
Φ^{rel}	1.34915	1.34741	1.34718	1.34683	1.34659	1.34649	1.34626
τ^{SUR}	0	0	0	0	0	0	0
Φ^{SUR}	1.60251	1.40651	1.37175	1.38366	1.35234	1.35561	1.35328
σ^{SUR}	2	4	4	8	10	14	24
$\frac{n_t}{t_f}\eta^{\text{SUR}}$	0.316526	0.47577	0.492702	0.499711	0.483694	0.497768	0.49956
Maximum of $\sigma_{\max} = 3$ switches							
τ^{bb}	0.00	0.00	0.00	0.00	0.00	0.01	0.14
τ^{cplex}	0.008001	0.016001	0.020002	0.100006	0.412026	0.584037	5.54835
Φ^{milp}	1.60251	1.60251	1.52323	1.67052	1.48912	1.55515	1.70474
σ^{milp}	2	2	3	2	2	3	3
$\frac{n_t}{t_f}\eta^{\text{milp}}$	0.316526	0.753893	0.807746	0.970736	1.55474	1.85555	3.49649
τ^{bonmin}	63.212	134.204	164.106	420.134	998.922	1600.61	1800*
Φ^{minlp}	1.60251	1.57489	1.52323	1.38746	1.39481	1.38741	*
σ^{minlp}	2	3	3	2	3	3	*
Maximum of $\sigma_{\max} = 4$ switches							
τ^{bb}	0.00	0.00	0.00	0.00	0.00	0.01	0.09
τ^{cplex}	0.008	0.016001	0.020001	0.080006	0.428027	1.00806	4.8443
Φ^{milp}	1.60251	1.40651	1.37175	1.36718	1.4576	1.39684	1.40632
σ^{milp}	2	4	4	4	4	4	4
$\frac{n_t}{t_f}\eta^{\text{milp}}$	0.316526	0.47577	0.492702	0.671702	0.951219	1.17408	1.98732
τ^{bonmin}	62.8599	106.903	145.381	610.482	1800*	1800*	1800*
Φ^{minlp}	1.60251	1.40651	1.37175	1.35883	1.36079	1.35643	3.36001
σ^{minlp}	2	4	4	4	4	4	4
Maximum of $\sigma_{\max} = 5$ switches							
τ^{bb}	0.00	0.00	0.00	0.00	0.00	0.02	0.56
τ^{cplex}	0.008001	0.016001	0.020001	0.088006	1.36809	3.1562	32.378
Φ^{milp}	1.60251	1.40651	1.37175	1.36718	1.4576	1.41056	1.40632
σ^{milp}	2	4	4	4	4	5	4
$\frac{n_t}{t_f}\eta^{\text{milp}}$	0.316526	0.47577	0.492702	0.671702	0.951219	1.17408	1.98732
τ^{bonmin}	60.0358	114.095	153.706	979.285	1800*	1800*	1800*
Φ^{minlp}	1.60251	1.40651	1.37175	1.35883	1.37073	1.35896	*
σ^{minlp}	2	4	4	4	5	5	*
Maximum of $\sigma_{\max} = 6$ switches							
τ^{bb}	0.00	0.00	0.00	0.00	0.00	0.01	0.71
τ^{cplex}	0.012001	0.016001	0.016002	0.096007	0.884056	2.92418	41.9546
Φ^{milp}	1.60251	1.40651	1.37175	1.3654	1.45852	1.39149	1.39471

Table 2 continued

n_t	10	20	25	50	80	100	200
σ^{milp}	2	4	4	6	6	6	6
$\frac{n_t}{t_f}\eta^{milp}$	0.316526	0.47577	0.492702	0.505287	0.793561	0.86204	1.50351
τ^{bonmin}	59.7637	114.447	147.777	374.347	1800*	1800*	1800*
Φ^{minlp}	1.60251	1.40651	1.37175	1.35233	1.35122	1.35211	1.90096
σ^{minlp}	2	4	4	6	6	6	6
Maximum of $\sigma_{max} = 7$ switches							
τ^{bb}	0.00	0.00	0.00	0.00	0.00	0.02	2.39
τ^{cplex}	0.008	0.016001	0.020001	0.096006	1.73611	6.59241	250.428
Φ^{milp}	1.60251	1.40651	1.37175	1.36533	1.45852	1.35481	1.39471
σ^{milp}	2	4	4	7	6	7	6
$\frac{n_t}{t_f}\eta^{milp}$	0.316526	0.47577	0.492702	0.50359	0.793561	0.858368	1.50351
τ^{bonmin}	57.8996	111.763	147.473	364.447	1800*	1800*	1800*
Φ^{minlp}	1.60251	1.40651	1.37175	1.35233	1.35439	1.3539	*
σ^{minlp}	2	4	4	6	6	6	*
Maximum of $\sigma_{max} = 8$ switches							
τ^{bb}	0.00	0.00	0.00	0.00	0.00	0.01	1.15
τ^{cplex}	0.008	0.008	0.016002	0.084005	0.780049	5.34833	388.74
Φ^{milp}	1.60251	1.40651	1.37175	1.38366	1.34997	1.38437	1.35297
σ^{milp}	2	4	4	8	8	8	8
$\frac{n_t}{t_f}\eta^{milp}$	0.316526	0.47577	0.492702	0.499711	0.602692	0.725728	1.23938
τ^{bonmin}	57.8636	112.363	139.653	356.37	1800*	1800*	1800*
Φ^{minlp}	1.60251	1.40651	1.37175	1.35233	1.34964	1.34956	1.43779
σ^{minlp}	2	4	4	6	8	8	8

For reference the original MINLP is solved relaxed (*rel*), with Sum Up Rounding (*SUR*), and with *Bonmin*. τ CPU time, Φ MINLP objective, η MILP objective, σ number switches

MINLP solver (τ^{cplex} vs. τ^{bonmin}). For all larger problems *Bonmin* violated the upper time limit of 1800 seconds.

Remark 3 It is interesting to observe that, as σ_{max} increases for given n_t, the computational effort increases, due to the fact that more Branch and Bound subtrees need to be evaluated. However, once the value σ_{max} reaches σ^{SUR}, the solution of MILP (10) can be determined in linear time with the Sum Up Rounding strategy, compare Theorem 3.

5.2 Using the MILP solution for cutoff in the MINLP tree

The MILP solution can itself be used as a solution that will get arbitrarily close to the lower bound, if n_t and σ_{max} are large enough. If the global solution on a given grid is an issue, and MINLP solvers have to be used, the solution can still be used to obtain a reduction in the MINLP Branch and Bound tree. The MILP solution is a

Table 3 Results for Lotka Volterra fishing problem as MINLP resulting from (3)

n_t	10	20	25	50	80	100	200
Maximum of σ_{max} = 3 switches							
$\tau^{scratch}$	63.212	134.204	164.106	420.134	998.922	1600.61	1800*
$\Phi^{scratch}$	1.60251	1.57489	1.52323	1.38746	1.39481	1.38741	*
$\sigma^{scratch}$	2	3	3	2	3	3	*
τ^{milp_init}	32.75	118.675	128.58	425.671	912.821	1391.44	1800*
Φ^{milp_init}	1.60251	1.57489	1.52323	1.38746	1.39481	1.38741	1.70474
σ^{milp_init}	2	3	3	2	3	3	3*
Maximum of σ_{max} = 4 switches							
$\tau^{scratch}$	62.8599	106.903	145.381	610.482	1800*	1800*	1800*
$\Phi^{scratch}$	1.60251	1.40651	1.37175	1.35883	1.36079	1.35643	3.36001
$\sigma^{scratch}$	2	4	4	4	4	4	4
τ^{milp_init}	31.522	73.2166	89.9296	522.173	1800*	1800*	1800*
Φ^{milp_init}	1.60251	1.40651	1.37175	1.35883	1.36079	1.35643	1.40632
σ^{milp_init}	2	4	4	4	4	4	4*
Maximum of σ_{max} = 5 switches							
$\tau^{scratch}$	60.0358	114.095	153.706	979.285	1800*	1800*	1800*
$\Phi^{scratch}$	1.60251	1.40651	1.37175	1.35883	1.37073	1.35896	*
$\sigma^{scratch}$	2	4	4	4	5	5	*
τ^{milp_init}	30.0979	79.965	119.463	824.568	1800*	1800*	1800*
Φ^{milp_init}	1.60251	1.40651	1.37175	1.35883	1.36917	1.35896	1.40632
σ^{milp_init}	2	4	4	4	5	5	4*
Maximum of σ_{max} = 6 switches							
$\tau^{scratch}$	59.7637	114.447	147.777	374.347	1800*	1800*	1800*
$\Phi^{scratch}$	1.60251	1.40651	1.37175	1.35233	1.35122	1.35211	1.90096
$\sigma^{scratch}$	2	4	4	6	6	6	6
τ^{milp_init}	30.3499	80.8731	129.78	377.04	1800*	1800*	1800*
Φ^{milp_init}	1.60251	1.40651	1.37175	1.35233	1.35122	1.35098	1.38382
σ^{milp_init}	2	4	4	6	6	6	6
Maximum of σ_{max} = 7 switches							
$\tau^{scratch}$	57.8996	111.763	147.473	364.447	1800*	1800*	1800*
$\Phi^{scratch}$	1.60251	1.40651	1.37175	1.35233	1.35439	1.3539	*
$\sigma^{scratch}$	2	4	4	6	6	6	*
τ^{milp_init}	29.9899	78.4889	114.931	350.93	1800*	1800*	1800*
Φ^{milp_init}	1.60251	1.40651	1.37175	1.35233	1.35354	1.35471	1.39471
σ^{milp_init}	2	4	4	6	6	6	6*
Maximum of σ_{max} = 8 switches							
$\tau^{scratch}$	57.8636	112.363	139.653	356.37	1800*	1800*	1800*
$\Phi^{scratch}$	1.60251	1.40651	1.37175	1.35233	1.34964	1.34956	1.43779

Table 3 continued

n_t	10	20	25	50	80	100	200
σ^{scratch}	2	4	4	6	8	8	8
$\tau^{\text{milp\_init}}$	30.1859	79.313	112.163	359.162	1800*	1800*	1800*
$\Phi^{\text{milp\_init}}$	1.60251	1.40651	1.37175	1.35233	1.34952	1.34977	1.35297
$\sigma^{\text{milp\_init}}$	2	4	4	6	8	8	8*

Solutions and computation times for *Bonmin* runs without initialization (*scratch*) as in Table 2 and using the solution Φ^{milp} of (10) for initial cutoff in the Branch & Bound tree (*milp_init*)

feasible solution that respects the switching constraint (3e). *Bonmin* provides a `bonmin.cutoff` option that can be used to eliminate branches with a lower bound exceeding this value. In Table 3 numerical results are presented that show the effect of this additional information.

It results either in a reduction of the overall computation time (up to approximately 50%) when comparing $\tau^{\text{milp\_init}}$ to τ^{scratch}, or in better solutions $\Phi^{\text{milp\_init}}$ compared to Φ^{scratch}, if the computation time is bounded. For the rightmost column with $n_t = 200$ all results obtained by making use of the information from the MILP solution resulted in a better solution.

6 Conclusions

We presented a novel method to solve optimal control problems including control functions with a discrete feasible set and switching constraints. The approach is based on a *first discretize, then optimize* approach which results in MINLPs that need to be solved. To avoid the high computational burden of solving the MINLP with standard methods, we propose to decompose the problem into a NLP and a MILP.

Although the MILP solution is not necessarily optimal for the MINLP, it has the advantage to be feasible, to have asymptotic properties as n_t increases, and to be a priori bounded. We proved that it converges against the solution of the nonlinear mixed-integer optimal control problem, if the switching constraint does not become active and the time discretization is refined. If the switching constraint is active, knowledge of system properties, such as the Lipschitz constant of the right-hand side function of the differential equation, allows to formulate an upper bound on the deviation of the MILP based solution from the solution of the relaxed optimal control problem. This upper bound depends linearly on the objective function value of the MILP.

We furthermore analyzed the structure of the convex hull of feasible points to the MILP and discussed why tailored cutting planes are not likely to be computationally beneficial. We presented a tailored Branch and Bound algorithm to cope with this specific structure. We presented numerical results for a benchmark problem in nonlinear mixed-integer optimal control that illustrate the efficiency of our approach.

Future work will concentrate on related optimization problems, such as the minimization of the number of switches subject to a maximal deviation from the optimal solution without switching constraints, or a weighted sum between penalization of switching and performance with respect to the objective. Open questions include also an efficient determination of model-dependent constants that are needed for the error estimations, and the question of reusage of information in adaptive or moving horizon schemes.

Theorem 3 and the way we efficiently solve the MILP need to be generalized to the case where the integer control functions enter in a nonlinear way. Our main motivation to follow the MILP based way is to be able to incorporate any kind of linear constraints on the binary control functions, hence we will try to identify classes of combinatorial constraints and to incorporate them into our methodology.

Acknowledgments Financial support of the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences and of the EU project EMBOCON under grant FP7-ICT-2009-4 248940 is gratefully acknowledged.

References

Abhishek K, Leyffer S, Linderoth J (2006) Filmint: an outer-approximation-based solver for nonlinear mixed integer programs. Preprint ANL/MCS-P1374-0906, Argonne National Laboratory, Mathematics and Computer Science Division

Applegate D, Bixby R, Chvátal V, Cook W, Espinoza D, Goycoolea M, Helsgaun K (2009) Certification of an optimal TSP tour through 85, 900 cities. Oper Res Lett 37(1):11–15

Belotti P, Lee J, Liberti L, Margot F, Waechter A (2009) Branching and bounds tightening techniques for non-convex MINLP. Optim Methods Softw 24:597–634

Bixby RE, Fenelon M, Gu Z, Rothberg E, Wunderling R (2004) Mixed-integer programming: a progress report. In: The sharpest cut: the impact of manfred padberg and his work. SIAM

Bonami P, Biegler L, Conn A, Cornuéjols G, Grossmann I, Laird C, Lee J, Lodi A, Margot F, Sawaya N, Wächter A (2009) An algorithmic framework for convex mixed integer nonlinear programs. Discr Optim 5(2):186–204

Bonami P, Cornuejols G, Lodi A, Margot F (2009) Feasibility pump for mixed integer nonlinear programs. Math Program 199:331–352

Bonami P, Kilinc M, Linderoth J (2009) Algorithms and software for convex mixed integer nonlinear programs. Technical Report 1664, University of Wisconsin

Burgschweiger J, Gnädig B, Steinbach M (2008) Optimization models for operative planning in drinking water networks. Optim Eng 10(1):43–73, Online first

Christof T, Löbel A PORTA—polyhedron representation transformation algorithm http://www.zib.de/Optimization/Software/Porta/. PORTA Homepage

Christof T, Reinelt G (1996) Combinatorial optimization and small polytopes. TOP 4(1):1–53

Floudas C, Akrotirianakis I, Caratzoulas S, Meyer C, Kallrath J (2005) Global optimization in the 21st century: advances and challenges. Comput Chem Eng 29(6):1185–1202

Grossmann I (2002) Review of nonlinear mixed-integer and disjunctive programming techniques. Optim Eng 3:227–252

Kameswaran S, Biegler L (2006) Simultaneous dynamic optimization strategies: recent advances and challenges. Comput Chem Eng 30:1560–1575

Lee J, Leung J, Margot F (2004) Min-up/ min-down polytopes. Discr Optim 1:77–85

Leineweber D, Bauer I, Bock H, Schlöder J (2003) An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization, Part I: theoretical aspects. Comput Chem Eng 27:157–166

Margaliot M (2007) A counterexample to a conjecture of Gurvits on switched systems. IEEE Trans Automat Control 52(6):1123–1126

Sager S (2005) Numerical methods for mixed–integer optimal control problems. Der andere Verlag, Tönning, Lübeck, Marburg ISBN 3-89959-416-9. Available at http://sager1.de/sebastian/downloads/Sager2005.pdf

Sager S (2009) Reformulations and algorithms for the optimization of switching decisions in nonlinear optimal control. J Process Control 19(8):1238–1247

Sager S, Reinelt G, Bock H (2009) Direct methods with maximal lower bound for mixed-integer optimal control problems. Math Program 118(1):109–149

Sager S, Bock H, Diehl M (2011) The integer approximation error in mixed-integer optimal control. Accepted by Mathematical Programming doi:10.1007/s10107-010-0405-3

Sharon Y, Margaliot M (2007) Third-order nilpotency, finite switchings and asymptotic stability. J Differ Equ 233:135–150

Tawarmalani M, Sahinidis N (2002) Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software, and applications. Kluwer, Dordrecht