View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Heidelberger Dokumentenserver

IWR PREPRINT. Submitted to SICON on February, 26th, 2007.

SOLVING MIXED–INTEGER CONTROL PROBLEMS BY SUM UP ROUNDING WITH GUARANTEED INTEGER GAP

SEBASTIAN SAGER[*], HANS GEORG BOCK[†], AND MORITZ DIEHL[‡]

Abstract. Optimal control problems involving time–dependent decisions from a finite set have gained much interest lately, as they occur in practical applications with a high potential for optimization. Typical examples are the choice of gears in transport or processes involving valves instead of pumps. We present tailored rounding strategies for direct methods such that the resulting trajectory fulfills constraints and reaches the objective function value of any (and in particular the optimal) relaxed solution up to a certain tolerance. We prove that this tolerance depends on the control discretization grid, in other words, that the rounded solution will be arbitrarily close to the relaxed one, if only the underlying grid is chosen fine enough. For the first time we show that a finite number of switches will be enough to do so. This will be shown for the linear as well as for the nonlinear case, involving path and control constraints. A numerical example is supplied to illustrate the procedure.

Key words. Mixed integer programming, Nonlinear programming, Optimal control, Bang-bang

AMS subject classifications. 90C11, 90C30, 49J15, 49J30

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 want to minimize a Mayer term

$$\min_{x,w,u,\rho,p} \Phi(x(t_f), \rho, p) \tag{1.1a}$$

over the differential states $x(\cdot)$, the control functions $(w, u)(\cdot)$ and the parameters (ρ, p) subject to the n_x–dimensional ODE system

$$\dot{x}(t) = f(x(t), w(t), u(t), \rho, p), \qquad t \in [0, t_f], \tag{1.1b}$$

control and path constraints

$$0 \le c(x(t), u(t), \rho, p), \qquad t \in [0, t_f], \tag{1.1c}$$

interior point inequalities and equalities

$$0 \le r^{\text{ieq}}(x(0), x(t_1), \dots, x(t_f), \rho, p), \tag{1.1d}$$

$$0 = r^{\text{eq}}(x(0), x(t_1), \dots, x(t_f), \rho, p), \tag{1.1e}$$

binary admissibility of the control function $w(\cdot)$

$$w(\cdot) \in \{0, 1\}^{n_w}, \tag{1.1f}$$

and integer constraints on some of the parameters

$$\rho \in \{0, 1\}^{n_\rho}. \tag{1.1g}$$

[*] Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, Germany (sebastian.sager@iwr.uni-heidelberg.de).

[†] IWR, University of Heidelberg, Germany (bock@iwr.uni-heidelberg.de).

[‡] Optimization in Engineering Center, K.U. Leuven, Belgium (moritz.diehl@esat.kuleuven.be).

The main focus of this paper lies on the control functions $w(\cdot)$. Typical examples are the choice of gears in transport, [11], or processes involving valves instead of pumps, [9, 6].

Although the first MIOCPs, namely the optimization of operation of subway trains that are equipped with discrete acceleration stages, were already solved in the early eighties for the city of New York, [4], the so–called *indirect methods* used there do not seem appropriate for generic large–scale optimal control problems with underlying nonlinear differential (algebraic) equation systems. Instead *direct methods*, in particular *all–at–once approaches*, [5], [1], [2], have become the methods of choice for most practical problems, see [3] for an overview.

In direct methods infinite–dimensional control functions are discretized by basis functions with local support and corresponding finite–dimensional parameters that enter into the optimization problem. The drawback of direct methods with binary control functions is obviously that they lead to high–dimensional vectors of binary variables. For many practical applications a fine control discretization is required, however, therefore techniques from mixed–integer nonlinear programming like Branch and Bound or Outer Approximation will work only on limited and small time horizons because of the exponentially growing complexity of the problem, [12].

In this paper we will therefore follow a different path. We will first consider the case where $w(\cdot)$ enters linearly in the optimization problem. We will show that for any feasible relaxed[1] solution we obtain a binary solution by the presented rounding strategy that is feasible and reaches the objective function value, both up to a given tolerance that depends on the control discretization grid size.

For this we will deduce some theoretical results on the difference between differential states that are obtained by integration with different control functions in section 2. In section 3 we will present the rounding strategy and give an upper bound on the difference between the integral over the relaxed and the rounded control. In section 4 we will bring together the results and connect them to the optimization problem. In the sequel we will extend the procedure for more general problems nonlinear in $w(\cdot)$ as given by (1.1). In section 6 we investigate a benchmark example to illustrate the procedure. We sum up the results in section 7.

Let in the following $\| \cdot \|$ denote the maximum norm $\| \cdot \|_\infty$.

2. Approximating differential states. In this section we want to show how the difference of the integrals of two differential states depends on the difference of the integrals of their corresponding control functions. Before we come to the main theorem of this section, we need the following lemma.

Lemma 2.1 (**A variant of the Gronwall Lemma**).

If for constant $c_1, c_2 \in \mathbb{R}, c_2 \neq 0$, *it holds that*

$$\| \dot{x}(t) \| \leq c_1 + c_2 \ \| x(t) \| \quad \forall\, t \in [0, t_f],$$

then

$$\| x(t) \| \leq \frac{c_1}{c_2} \left(e^{c_2 t} - 1 \right) + \| x(0) \| \, e^{c_2 t}.$$

[1] here and in the following the term *relaxed* corresponds to a relaxation of constraint (1.1f) to the hypercube $[0,1]^{n_w}$

Proof.

$$
\begin{aligned}
\| \, x(t) - x(0) \, \| &= \left\| \int_0^t \dot{x}(\tau) \, \mathrm{d}\tau \right\| \le \int_0^t \| \, \dot{x}(\tau) \, \| \, \mathrm{d}\tau \\
&\le tc_1 + c_2 \int_0^t \| \, x(\tau) \, \| \, \mathrm{d}\tau \\
&\le tc_1 + c_2 \int_0^t \| \, x(0) \, \| + \| \, x(\tau) - x(0) \, \| \, \mathrm{d}\tau \\
&= t \underbrace{(c_1 + c_2 \| \, x(0) \, \|)}_{c_3 :=} + c_2 \int_0^t \| \, x(\tau) - x(0) \, \| \, \mathrm{d}\tau
\end{aligned}
$$

If we write $\phi(t) = \| \, x(t) - x(0) \, \|$ and $\Phi(t) = c_3 \, t + c_2 \int_0^t \| \, x(\tau) - x(0) \, \| \, \mathrm{d}\tau$ then we have $\dot{\Phi}(t) = c_3 + c_2 \phi(t)$. Because of $\phi(t) \le \Phi(t)$ it follows that $\dot{\Phi}(t) \le c_3 + c_2 \Phi(t)$. As $\Phi(0) = 0$ we can deduce that the solution $\omega(t) = \frac{c_3}{c_2} \left(e^{tc_2} - 1 \right)$ of the initial value problem

$$\dot{\xi}(t) = c_3 + c_2 \xi(t), \quad \xi(0) = 0$$

is an upper bound on $\Phi(t)$ and therefore also on $\phi(t)$, thus

$$\| \, x(t) - x(0) \, \| \le \frac{c_3}{c_2} \left(e^{tc_2} - 1 \right) = \left(\frac{c_1}{c_2} + \| \, x(0) \, \| \right) \left(e^{tc_2} - 1 \right)$$

and hence

$$
\begin{aligned}
\| \, x(t) \, \| &\le \| \, x(0) \, \| + \| \, x(t) - x(0) \, \| \\
&\le \| \, x(0) \, \| + \frac{c_1}{c_2} e^{c_2 t} - \frac{c_1}{c_2} + \| \, x(0) \, \| \, e^{c_2 t} - \| \, x(0) \, \| \\
&= \frac{c_1}{c_2} \left(e^{c_2 t} - 1 \right) + \| \, x(0) \, \| \, e^{c_2 t}
\end{aligned}
$$

□

Assume now we are given an initial value problem that is of the form

$$\dot{x}(t) = A(x(t)) w(t), \quad x(0) = x_0. \tag{2.1}$$

Here $A(x(t))$ is a matrix in $\mathbb{R}^{n_x \times n_w}$ with entries depending arbitrarily on $x(t)$. Note that we leave away a term independent of $w(\cdot)$, as it may be included easily by fixing one additional component of w to 1. Also, from now on we will often leave the argument (t) away for the sake of notational simplicity. The following theorem states how the difference of solutions to this initial value problem depends on the integrated difference between control functions and the difference between the initial values.

Theorem 2.2. *Let $x(\cdot)$ and $y(\cdot)$ be solutions of the initial value problems*

$$
\begin{aligned}
\dot{x} &= A(x) w, \quad x(0) = x_0, \\
\dot{y} &= A(y) v, \quad y(0) = y_0
\end{aligned}
$$

with $t \in [0, t_f]$. If for all $t \in [0, t_f]$ it holds that

$$\begin{aligned} \| w \| &\le 1, \\ \| v \| &\le 1, \\ \| A(x) \| &\le M \quad \forall\, x \in \mathbb{R}^{n_x}, \\ \| A(y) - A(x) \| &\le L \| y - x \| \quad \forall\, x, y \in \mathbb{R}^{n_x}, \end{aligned}$$

and

$$\left\| \int_0^t v(\tau) - w(\tau) \, \mathrm{d}\tau \right\| \le \epsilon$$

then

$$\| y(t) - x(t) \| \le (2M\epsilon + \| y_0 - x_0 \|) \, e^{Lt}$$

for all $t \in [0, t_f]$.
Proof. Write $\Delta x = y - x$ and $\Delta w = v - w$. Then

$$\begin{aligned} \Delta \dot{x} = \dot{y} - \dot{x} &= A(y)v - A(x)w \\ &= A(x + \Delta x)v - A(x)(v - \Delta w) \\ &= (A(x + \Delta x) - A(x))v + A(x)\Delta w \end{aligned}$$

Consider Δw to be the derivative of a function Δa, that is, $\Delta w = \Delta \dot{a} = \frac{\mathrm{d}}{\mathrm{d}t}\Delta a$. Then

$$\Delta \dot{x} = (A(x + \Delta x) - A(x))v + \frac{\mathrm{d}}{\mathrm{d}t}(A(x)\Delta a) - \frac{\mathrm{d}}{\mathrm{d}t}A(x)\Delta a$$

Rearranging yields

$$\frac{\mathrm{d}}{\mathrm{d}t} \underbrace{(\Delta x - A(x)\Delta a)}_{\Delta z :=} = (A(x + \Delta x) - A(x))v - \frac{\mathrm{d}}{\mathrm{d}t}A(x)\Delta a.$$

We insert $\Delta x = \Delta z + A(x)\Delta a$ and obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta z = (A(x + \Delta z + A(x)\Delta a) - A(x))v - \frac{\mathrm{d}}{\mathrm{d}t}A(x)\Delta a.$$

Taking the norms on both sides we get

$$\begin{aligned} \left\| \frac{\mathrm{d}}{\mathrm{d}t}\Delta z \right\| &\le L \| x + \Delta z + A(x)\Delta a - x \| \| v \| + \left\| \frac{\mathrm{d}}{\mathrm{d}t}A(x) \right\| \| \Delta a \| \\ &\le L (\| \Delta z \| + \| A(x) \| \| \Delta a \|) \| v \| + L \| \dot{x} \| \| \Delta a \| \\ &\le LM \| \Delta a \| \| v \| + L \| \dot{x} \| \| \Delta a \| + L \| v \| \| \Delta z \| \\ &\le \underbrace{LM \| \Delta a \| + L \| \dot{x} \| \| \Delta a \|}_{c_1 :=} + L \| \Delta z \| \end{aligned}$$

Applying Lemma 2.1 yields

$$\begin{aligned} \| \Delta z \| &\le \frac{c_1}{L}(e^{Lt} - 1) + \| \Delta z(0) \| \, e^{Lt} \\ &= (M \| \Delta a \| + \| \dot{x} \| \| \Delta a \|)(e^{Lt} - 1) + \| \Delta z(0) \| \, e^{Lt} \end{aligned}$$

Remembering $\Delta z = \Delta x - A(x)\Delta a$ we have

$$\| \Delta x - A(x)\Delta a \| \leq \| \Delta a \| (M + \| \dot{x} \|)(e^{Lt} - 1) + \|\Delta x(0) - A(x(0)) \underbrace{\Delta a(0)}_{=\int_0^0 v - w = 0} \|e^{Lt}$$

and by rearranging

$$\| \Delta x - A(x)\Delta a \| + M \| \Delta a \| \leq \| \Delta a \| \| \dot{x} \| (e^{Lt} - 1) + M \| \Delta a \| e^{Lt} + \| \Delta x(0) \| e^{Lt}$$

It follows

$$\begin{aligned} \| \Delta x \| &= \| \Delta x - A(x)\Delta a + A(x)\Delta a \| \\ &\leq \| \Delta x - A(x)\Delta a \| + \| A(x) \| \| \Delta a \| \\ &\leq \| \Delta x - A(x)\Delta a \| + M \| \Delta a \| \\ &\leq \| \Delta a \| \| \dot{x} \| (e^{Lt} - 1) + M \| \Delta a \| e^{Lt} + \| \Delta x(0) \| e^{Lt} \\ &= (\| \Delta a \| \| \dot{x} \| + M \| \Delta a \| + \| \Delta x(0) \|) \, e^{Lt} - \| \Delta a \| \| \dot{x} \| \\ &\leq (\| \Delta a \| \| \dot{x} \| + M \| \Delta a \| + \| \Delta x(0) \|) \, e^{Lt} \end{aligned}$$

With $\| \dot{x} \| = \| A(x)w \| \leq \| A(x) \| \| w \| \leq M$ we obtain

$$\begin{aligned} \| \Delta x(t) \| &\leq (\| \Delta a \| \, 2M + \| \Delta x(0) \|) \, e^{Lt} \\ &= \left(\left\| \int_0^t v(\tau) - w(\tau) \, \mathrm{d}\tau \right\| 2M + \| y(0) - x(0) \| \right) e^{Lt} \\ &\leq (\epsilon 2M + \| y_0 - x_0 \|) \, e^{Lt} \end{aligned}$$

□

In our context the initial values x_0 and y_0 will be identical. Therefore theorem 2.2 states that we have an upper bound on the difference between differential states that depends linearly on the integrated difference between the two control functions. In the next section we will investigate this term closer.

3. Approximating the integral over the controls by Sum Up Rounding. We consider a piecewise constant control function

$$v_j(t) = q_{j,i}, \quad t \in [t_i, t_{i+1}] \tag{3.1}$$

with $j = 1 \dots n_w$ and $i = 0 \dots m-1$ on a fixed time grid $0 = t_0 < t_1 < \cdots < t_m = t_f$. Such a function could be the result of an optimization of a control problem with a direct approach that discretizes the control functions by piecewise constant functions. We write $\Delta t_i := t_{i+1} - t_i$ and Δt for the maximum distance between two time points,

$$\Delta t := \max_{i=0\dots m-1} \Delta t_i = \max_{i=0\dots m-1} \{t_{i+1} - t_i\}. \tag{3.2}$$

Let then a function $w(\cdot) : [0, t_f] \mapsto \{0,1\}^{n_w}$ be defined by

$$w_j(t) = p_{j,i}, \quad t \in [t_i, t_{i+1}] \tag{3.3}$$

where the $p_{j,i}$ are binary values given by

$$p_{j,i} = \left\{ \begin{array}{ll} 1 & \text{if } \sum_{k=0}^{i} q_{j,k}\Delta t_k - \sum_{k=0}^{i-1} p_{j,k}\Delta t_k \geq 0.5\Delta t_i \\ 0 & \text{else} \end{array} \right. . \tag{3.4}$$

See figure 6.1 for an example. We have the following estimate on the integral over the difference between the control functions $v(\cdot)$ and $w(\cdot)$.

THEOREM 3.1. *Let the functions $v : [0, t_f] \mapsto [0, 1]^{n_w}$ and $w : [0, t_f] \mapsto \{0, 1\}^{n_w}$ be given by (3.1) and (3.3, 3.4), respectively. Then it holds*

$$\left\| \int_0^t v(\tau) - w(\tau) \; \mathrm{d}\tau \right\| \le 0.5 \; \Delta t$$

Proof. Let $-1 \le n_t \le m - 1$ be the index such that $t_{n_t+1} \le t < t_{n_t+2}$. First observe that the integral

$$\int_0^t v(\tau) - w(\tau) \; \mathrm{d}\tau = \sum_{i=0}^{n_t} (q_{j,i} - p_{j,i})(t_{i+1} - t_i) + (q_{j,n_t+1} - p_{j_{n_t+1}})(t - t_{n_t+1})$$

is a piecewise linear function of t, therefore it suffices to show the claim for all $t = t_{r+1}$. For $r = -1 \dots m - 1$ we have

$$\left\| \int_0^{t_{r+1}} v(\tau) - w(\tau) \; \mathrm{d}\tau \right\| = \max_j \left| \sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i \right| \le 0.5 \; \Delta t \tag{3.5}$$

We use induction to show (3.5). For $r = -1$ the claim follows trivially. So let us assume

$$\max_j \left| \sum_{i=0}^{r-1} (q_{j,i} - p_{j,i}) \Delta t_i \right| \le 0.5 \; \Delta t \tag{3.6}$$

and show that also

$$\max_j \left| \sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i \right| \le 0.5 \; \Delta t.$$

If $p_{j,r} = 1$, then because of (3.4) we have

$$\sum_{k=0}^{r} q_{j,k} \Delta t_k - \sum_{k=0}^{r-1} p_{j,k} \Delta t_k \ge 0.5 \Delta t_r$$

and by adding $-p_{j,k} \Delta t_r = -\Delta t_r$ on both sides

$$\sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i \ge -0.5 \Delta t_r \ge -0.5 \Delta t.$$

By induction hypothesis we also have

$$\underbrace{\sum_{i=0}^{r-1} (q_{j,i} - p_{j,i}) \Delta t_i}_{\le 0.5 \Delta t} + \underbrace{(q_{j,r} - 1) \Delta t_r}_{\le 0} \le 0.5 \Delta t.$$

If $p_{j,r} = 0$, then because of (3.4) we have

$$\sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i = \sum_{k=0}^{r} q_{j,k} \Delta t_k - \sum_{k=0}^{r-1} p_{j,k} \Delta t_k < 0.5 \Delta t_r \le 0.5 \Delta t.$$

By induction hypothesis we also have

$$\underbrace{\sum_{i=0}^{r-1}(q_{j,i}-p_{j,i})\Delta t_i}_{\geq -0.5\Delta t}+\underbrace{(q_{j,r})\Delta t_r}_{\geq 0}\geq -0.5\Delta t,$$

completing the proof.

□

4. Connection to the optimization problem. Let us assume we have found the feasible and (globally) optimal trajectory[2] $(x^*, v^*, u^*, \rho^*, p^*)$ of problem (1.1), where the integer constraint (1.1f) on $v^*(\cdot)$ was relaxed to $[0,1]^{n_w}$. If the control functions $w(\cdot)$ enter linearly in the ODE (1.1b), then we can fix the continuous control functions $u^*(\cdot)$, the binary valued parameters ρ^* and the continuous parameters p^* and write $f(\cdot)$ as a function $A(x; u^*, \rho^*, p^*)w$ of $x(\cdot)$ and $w(\cdot)$ only. Fixing also $x(0)$ to $x^*(0)$, the ODE (1.1b) is in the form of (2.1).

Then we can, given tolerances δ_1, δ_2, δ_3, and δ_4 for objective function, path and control constraints and interior point equality and inequality constraints, respectively, determine a grid size Δt such that all tolerances are kept. This is due to the fact that at all times t every state $x_i(t)$ is closer than $M\Delta t e^{Lt_f}$ to the feasible trajectory $x^*(t)$. As all occurring functions $\Phi(\cdot), c(\cdot), r^{\text{ieq}}(\cdot)$ and $r^{\text{ieq}}(\cdot)$ are continuous functions, also their value will be arbitrarily close to the one of the relaxed solution when Δt gets smaller.

In practice, the values of $u^*(\cdot)$ and p^* do not necessarily need to be fixed, as they allow for further flexibility and may improve the objective function value on a given grid. The binary parameters ρ^* however are typically hard to compute and should be fixed. Our procedure allows thus for a decoupling of the determination of optimal binary parameters and optimal binary control functions.

5. Extension to the nonlinear case. To apply the above results also to the more general nonlinear case, we convexify problem (1.1) with respect to the binary control functions $w(\cdot)$ as first suggested in [7]. This convexification, that is described below, again allows us to generate a tight relaxation of the integer control problem - very similar as before for the affinely entering binary controls, but with one important modification, namely an additional linear constraint to ensure the controls form a special ordered set at each instant in time. We assign one control function $\xi_i(\cdot)$ to every possible control $w^i \in \{0,1\}^{n_w}$. In the worst case, this corresponds to $n_\xi = 2^{n_w}$ vertices of the hypercube. In practice however often there is a finite set of admissible choices resp. most of the vertices can be excluded logically. Here n_ξ would correspond to the number of these feasible choices. Examples are the selection of a gear [11], of a destillation column tray [7] or of an inlet stream port [9]. In all examples n_ξ is linear in the number of choices. Furthermore, in most practical applications the binary control functions enter linearly (such as valves that indicate whether a certain term is present or not). Therefore the drawback of an increased number of control functions is outweighted by the advantages concerning the avoidance of binary variables associated with the discretization in time for most applications we know of.

[2] in the context of direct methods, i.e., for a sufficiently fine approximation of the infinite–dimensional controls

By convexifying with respect to $w(\cdot)$, we obtain the following optimal control problem. We want to minimize a Mayer term

$$\min_{x,\xi,u,\rho,p} \Phi(x(t_f),\rho,p) \tag{5.1a}$$

subject to the n_x–dimensional convexified ODE system

$$\dot{x}(t) = \sum_{i=1}^{n_\xi} f(x(t),w^i,u(t),\rho,p)\ \xi_i(t), \quad t \in [0,t_f], \tag{5.1b}$$

control and path constraints

$$0 \leq c(x(t),u(t),\rho,p), \quad t \in [0,t_f] \tag{5.1c}$$

interior point inequalities and equalities

$$0 \leq r^{\mathrm{ieq}}(x(0),x(t_1),\ldots,x(t_f),\rho,p), \tag{5.1d}$$
$$0 = r^{\mathrm{eq}}(x(0),x(t_1),\ldots,x(t_f),\rho,p), \tag{5.1e}$$

binary admissibility of the control function $\xi(\cdot)$

$$\xi(\cdot) \in \{0,1\}^{n_\xi}, \tag{5.1f}$$

integer constraints on some of the parameters

$$\rho \in \{0,1\}^{n_\rho}, \tag{5.1g}$$

and the special ordered set type one constraint

$$\sum_{i=1}^{n_\xi} \xi_i(t) = 1 \quad \forall\, t \in [0,t_f]. \tag{5.1h}$$

There is obviously a bijection $w(t) = w^i \leftrightarrow \xi_i(t) = 1$ between solutions of problems (1.1) and (5.1), compare [7]. This means, we can find a solution to the convexified linear program by applying the proposed Sum Up Rounding strategy to a solution of its relaxation and then deduce the optimal solution to the nonlinear binary problem from it.

However, the Sum Up Rounding strategy (3.4) does not work for problems with the additional special ordered set property (5.1h), as can be seen by the easy example of two functions that have the constant value $v_1(t) = v_2(t) = 0.5$. Therefore we propose a different technique for functions that have to fulfill this equality. Let us assume we are given a piecewise constant function (3.1) that fulfills (5.1h). Again we define $w(\cdot) = \xi(\cdot)$ via (3.3), but with $p_{j,i}$ given by

$$\hat{p}_{j,i} = \sum_{k=0}^{i} q_{j,k}\Delta t_k - \sum_{k=0}^{i-1} p_{j,k}\Delta t_k \tag{5.2a}$$

$$p_{j,i} = \begin{cases} 1 & \text{if } \hat{p}_{j,i} \geq \hat{p}_{k,i}\ \forall\ k \neq j \text{ and } j < k\ \forall\ k : \hat{p}_{j,i} = \hat{p}_{k,i} \\ 0 & \text{else} \end{cases} \tag{5.2b}$$

and not by (3.4). Note that $w(t)$ also fulfills the special ordered set type one property (5.1h). Again we have an estimation of the integral over $v - w$ that depends on Δt of the underlying grid, compare (3.2).

THEOREM 5.1. *Let the functions $v : [0, t_f] \mapsto [0, 1]^{n_w}$ and $w : [0, t_f] \mapsto \{0, 1\}^{n_w}$ be given by (3.1) and (3.3, 5.2), respectively, and let both fulfill equation (5.1h) for all $t \in [0, t_f]$. Then it holds*

$$\left\| \int_0^t v(\tau) - w(\tau) \, \mathrm{d}\tau \right\| \leq (n_w - 1) \, \Delta t$$

Proof. As above we can restrict our proof to the case that $t = t_{r+1}$. For $r = -1 \ldots m - 1$ we have

$$\left\| \int_0^{t_{r+1}} v(\tau) - w(\tau) \, \mathrm{d}\tau \right\| = \max_j \left| \sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i \right| .$$

For the sake of notational simplicity we define

$$k := \arg \max_{j=1 \ldots n_w} \left| \sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i \right| .$$

We use a case differentiation to show $|\sum_{i=0}^{r} (q_{k,i} - p_{k,i}) \Delta t_i| \leq (n_w - 1) \Delta t$. Let us first assume that $\sum_{i=0}^{r} (q_{k,i} - p_{k,i}) \Delta t_i > (n_w - 1) \Delta t$. As both $v(\cdot)$ and $w(\cdot)$ fulfill (5.1h), summing up over all control functions yields

$$\sum_{j=1}^{n_w} \sum_{i=0}^{r} q_{j,i} \Delta t_i = \sum_{j=1}^{n_w} \sum_{i=0}^{r} p_{j,i} \Delta t_i = \sum_{i=0}^{r} \Delta t_i,$$

hence

$$\sum_{1=j \neq k}^{n_w} \sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i + \sum_{i=0}^{r} (q_{k,i} - p_{k,i}) \Delta t_i = 0$$

and by assumption

$$\sum_{1=j \neq k}^{n_w} \sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i + (n_w - 1) \Delta t < 0.$$

We can write the left hand side as the sum of $n_w - 1$ terms $\Delta t + \sum_{i=0}^{r} (q_{j,i} - p_{j,i}) \Delta t_i$. Obviously at least one of them has to be negative, thus there exists an index $\hat{j}$ such that

$$\Delta t + \sum_{i=0}^{r} (q_{\hat{j},i} - p_{\hat{j},i}) \Delta t_i < 0.$$

Let $\hat{\imath}$ be the highest index for which the control $\hat{j}$ has been rounded up,

$$\hat{\imath} := \arg \max_{0 \leq i \leq m-1} \{ i : p_{\hat{j},i} = 1 \text{ and } p_{\hat{j},l} = 0 \; \forall \; l > i \}.$$

Note that $\hat{\imath}$ is well defined, as there must be at least two i such that $p_{\hat{j},i} = 1$. Then it holds

$$\sum_{i=0}^{\hat{\imath}} q_{\hat{j},i} \Delta t_i - \sum_{i=0}^{\hat{\imath}-1} p_{\hat{j},i} \Delta t_i \leq \Delta t + \sum_{i=0}^{r} (q_{\hat{j},i} - p_{\hat{j},i}) \Delta t_i < 0$$

and with

$$\hat{j} = \arg \max_{1 \le j \le n_w} \left\{ \sum_{i=0}^{\hat{i}} q_{j,i} \Delta t_i - \sum_{i=0}^{\hat{i}-1} p_{j,i} \Delta t_i \right\}$$

we have

$$\sum_{i=0}^{\hat{i}} (q_{j,i} - p_{j,i}) \Delta t_i \le \sum_{i=0}^{\hat{i}} q_{j,i} \Delta t_i - \sum_{i=0}^{\hat{i}-1} p_{j,i} \Delta t_i < 0 \quad \forall\, j$$

in contradiction to

$$\sum_{j=1}^{n_w} \sum_{i=0}^{\hat{i}} (q_{j,i} - p_{j,i}) \Delta t_i = \sum_{i=0}^{\hat{i}} \Delta t_i \sum_{j=1}^{n_w} (q_{j,i} - p_{j,i}) = 0.$$

Let us now assume the second case, $\sum_{i=0}^{r} (q_{k,i} - p_{k,i}) \Delta t_i < -(n_w - 1) \Delta t$. Let $\hat{i}$ be the highest index for which the control k has been rounded up,

$$\hat{i} := \arg \max_{0 \le i \le m-1} \{ i : p_{k,i} = 1 \text{ and } p_{k,l} = 0 \; \forall\, l > i \}.$$

Then it holds by assumption

$$\sum_{i=0}^{\hat{i}} p_{k,i} \Delta t_i = \sum_{i=0}^{r} p_{k,i} \Delta t_i > \sum_{i=0}^{r} q_{k,i} \Delta t_i + (n_w - 1) \Delta t \ge \sum_{i=0}^{\hat{i}} q_{k,i} \Delta t_i + (n_w - 1) \Delta t$$

and as k had the maximum value on interval $\hat{i}$,

$$\sum_{i=0}^{\hat{i}} (p_{j,i} - q_{j,i}) \Delta t_i > (n_w - 1) \Delta t \quad \forall\, j.$$

Summing up over all controls j yields

$$\sum_{j=1}^{n_w} \sum_{i=0}^{\hat{i}} (p_{j,i} - q_{j,i}) \Delta t_i > \sum_{j=1}^{n_w} (n_w - 1) \Delta t \quad \forall\, j$$

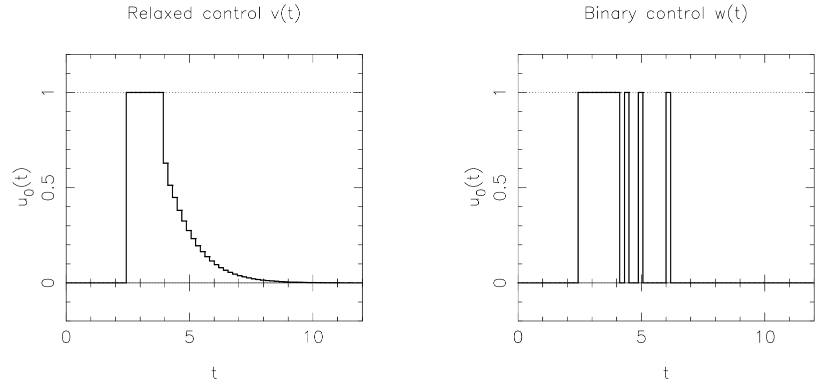
and because of (5.1h) we have the contradiction $0 > n_w^2 - n_w$.

□

Thus, also for control functions that have to obey the property (5.1h), we can get arbitrarily close to the optimal objective function value by making Δt sufficiently small.

In constraints (1.1c) and (5.1c) we assumed that the path and control constraints do not depend explicitly on $w(\cdot)$, or are fulfilled for all binary values. Unfortunately, only few claims can be made for the general case of $c(x(t), w(t), u(t), \rho, p)$, but these constraints have to be investigated problem–specifically, as in [10]. One general idea is to reformulate the constraints to

$$0 \le c(x(t), w^i, u(t), \rho, p)\; \xi_i(t), \qquad t \in [0, t_f], \;\; i = 1, \dots, n_\xi. \tag{5.3}$$


Relaxed control v(t)

Binary control w(t)

FIG. 6.1. *Relaxed and Sum Up Rounding binary controls for $m = 64$ time intervals.*

Note that by constraint (5.3) only positive relaxed solutions are feasible, for which also the corresponding binary vector is feasible. This makes it more unlikely (although not impossible) that the index i corresponding to a value $\xi_i(t) = 0$ is chosen as the maximum in (5.2), whenever $c(x(t), w^i, u(t), \rho, p) < 0$. Furthermore this constraint should be included in the rounding decision to avoid infeasibilities. Investigating this approach more rigorously is beyond the scope of this paper, however.

6. Numerical example. To illustrate the effect of the proposed rounding strategy we investigate a benchmark problem for mixed–integer optimal control including a singular arc. This problem was first proposed in [8] and reads as

$$\begin{array}{lrcl} \min\limits_{x,w} & x_2(t_f) & & \\ \text{s.t.} & \dot{x}_0(t) & = & x_0(t) - x_0(t)x_1(t) - \; c_0 x_0(t)\; w(t), \\ & \dot{x}_1(t) & = & -x_1(t) + x_0(t)x_1(t) - \; c_1 x_1(t)\; w(t), \\ & \dot{x}_2(t) & = & (x_0(t)-1)^2 + (x_1(t)-1)^2, \\ & x(0) & = & (0.5, 0.7, 0)^T, \\ & w(t) & \in & \{0,1\} \end{array}$$

with $t \in [t_0, t_f] = [0, 12]$ and parameters $c_0 = 0.4$ and $c_1 = 0.2$. We solve this problem using the direct multiple shooting [5] based software package `MUSCOD-II` for different equidistant control discretization intervals. Figure 6.1 shows the optimal solution of the relaxed problem and the corresponding control obtained by Sum Up Rounding for a discretization with 64 time intervals. In table 6.1 the values for different grid sizes are given. The equidistant interval length Δt is given by the overall length divided by m. The third column shows the value $\epsilon = \| \int_0^{12} v(\tau) - w(\tau) \; \mathrm{d}\tau \|$ that is bounded by $0.5\Delta t$, compare theorem 3.1. Note that the given upper bound is almost active for $m = 2$, but is way too high for finer discretizations. The fourth and the fifth column show the objective function values of the relaxed (R) and the rounded binary (B) solution. The discretization has been bisected for illustrative purposes. In practice more advanced adaptive schemes are used that neglect bang–bang arcs and take the goal to obtain approximate integral values into account, see [7]. The computational effort is very low. In every step only a relaxed optimization problem has to be solved. The rounding procedure is almost for free and then a simple forward simulation has

Table 6.1
Numerical results for Lotka example

m	Δt	$\int_0^{12} v - w$	$x_2^R(12)$	$x_2^B(12)$
2	6	2.87924	5.40278	8.51376
4	3	0.631402	2.75402	5.08501
8	1.5	0.585463	1.46812	1.90096
16	0.75	0.0827811	1.35597	1.73284
32	0.375	0.00718493	1.34881	1.61399
64	0.1875	0.00151131	1.34406	1.34680
128	0.09375	0.00135644	1.34405	1.34511
256	0.046875	0.00130135	1.34402	1.34430

to be performed. For the finer discretization the last solution can even be reused to initialize the optimization variables. An additional benefit of this approach is the fact that all previously calculated solutions can be stored and compared a posteriori to compare the trade off between the typically undesired often switching and a loss in the objective function.

7. Conclusions. We presented a novel method to solve mixed–integer optimal control problems that is based on a convexification and a relaxation of the binary control functions plus a tailored rounding strategy. This approach avoids the disadvantage of an exponential increase in complexity when the control discretization grid is chosen finer, from which previous direct methods like, e.g., Branch and Bound, suffer.

For the first time we showed in a constructive way that, under reasonable conditions, the objective function value of the relaxation of a convexified mixed–integer optimal control problem can be approximated arbitrarily close with a bang–bang solution that switches only finitely many times.

Acknowledgments. This work was supported by the *Deutsche Forschungsgemeinschaft (DFG)* under grant BO864/10-1, by Research Council KUL: CoE EF/05/006 Optimization in Engineering Center (OPTEC), and by Belgian Federal Science Policy Office: IUAP P6/04 (Dynamical systems, control and optimization, 2007-2011).

REFERENCES

[1] V. Bär, *Ein Kollokationsverfahren zur numerischen Lösung allgemeiner Mehrpunktrandwertaufgaben mit Schalt– und Sprungbedingungen mit Anwendungen in der optimalen Steuerung und der Parameteridentifizierung*, master's thesis, Universität Bonn, 1984.

[2] L.T. Biegler, *Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation*, Computers and Chemical Engineering, 8 (1984), pp. 243–248.

[3] T. Binder, L. Blank, H.G. Bock, R. Bulirsch, W. Dahmen, M. Diehl, T. Kronseder, W. Marquardt, J.P. Schlöder, and O.v. Stryk, *Introduction to model based optimization of chemical processes on moving horizons*, in Online Optimization of Large Scale Systems: State of the Art, M. Grötschel, S.O. Krumke, and J. Rambau, eds., Springer, 2001, pp. 295–340.

[4] H.G. Bock and R.W. Longman, *Computation of optimal controls on disjoint control sets for minimum energy subway operation*, in Proceedings of the American Astronomical Society. Symposium on Engineering Science and Mechanics, Taiwan, 1982.

[5] H.G. Bock and K.J. Plitt, *A multiple shooting algorithm for direct solution of optimal control problems*, in Proceedings 9th IFAC World Congress Budapest, Pergamon Press, 1984, pp. 243–247.

[6] Y. Kawajiri and L.T. Biegler, *Large-scale optimization strategies for zone configuration of simulated moving beds*, in 16th European Symposium on Computer Aided Process Engi-

neering and 9th International Symposium on Process Systems Engineering, Elsevier, 2006, pp. 131–136.

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

[8] S. Sager, H.G. Bock, M. Diehl, G. Reinelt, and J.P. Schlöder, *Numerical methods for optimal control with binary control functions applied to a Lotka-Volterra type fishing problem*, in Recent Advances in Optimization (Proceedings of the 12th French-German-Spanish Conference on Optimization), A. Seeger, ed., vol. 563 of Lectures Notes in Economics and Mathematical Systems, Heidelberg, 2006, Springer, pp. 269–289.

[9] S. Sager, M. Diehl, G. Singh, A. Küpper, and S. Engell, *Determining SMB superstructures by mixed-integer control*, in Proc. of OR2006, Karlsruhe, 2007.

[10] S. Sager, Y. Kawajiri, and L. Biegler, *On the optimality of superstructures for simulated moving beds: Is one pump sufficient for each stream?*, AIChE Journal, (2007), p. (submitted).

[11] S. Terwen, M. Back, and V. Krebs, *Predictive powertrain control for heavy duty trucks*, in Proceedings of IFAC Symposium in Advances in Automotive Control, Salerno, Italy, 2004, pp. 451–457.

[12] J. Till, S. Engell, S. Panek, and O. Stursberg, *Applied hybrid system optimization: An empirical investigation of complexity*, Control Eng, 12 (2004), pp. 1291–1303.