Time-Domain Decomposition for Mixed-Integer Optimal Control Problems

Falk M. Hante, Richard Krug, Martin Schmidt

Abstract. We consider mixed-integer optimal control problems, whose optimality conditions involve global combinatorial optimization aspects for the corresponding Hamiltonian pointwise in time. We propose a time-domain decomposition, which makes this problem class accessible for mixed-integer programming using parallel-in-time direct discretizations. The approach is based on a decomposition of the optimality system and the interpretation of the resulting subproblems as suitably chosen mixed-integer optimal control problems on subintervals in time. An iterative procedure then ensures continuity of the states at the boundaries of the subintervals via co-state information encoded in virtual controls. We prove convergence of this iterative scheme for discrete-continuous linear-quadratic problems and present numerical results both for linear-quadratic as well as nonlinear problems.

1. Problem Statement and Optimality Conditions

We consider optimal control problems of the form

$$\begin{aligned} \min \quad & \varphi_0(x(t_0)) + \varphi_{\mathrm{f}}(x(t_{\mathrm{f}})) && \text{(1a)}\\ \text{s.t.} \quad & \dot{x} = f(x,u) \quad \text{a. e. in } (t_0,t_{\mathrm{f}}), && \text{(1b)}\\ & \chi_j(x(t_0)) = 0, \quad j = 1,\dots,p, && \text{(1c)}\\ & \psi_j(x(t_{\mathrm{f}})) = 0, \quad j = 1,\dots,q, && \text{(1d)}\\ & u(t) \in U \quad \text{a. e. in } [t_0,t_{\mathrm{f}}], && \text{(1e)} \end{aligned}$$

where t_0 and t_{f} define a fixed and finite time horizon, $x : [t_0,t_{\mathrm{f}}] \to \mathbb{R}^n$ is a state function, $u : [t_0,t_{\mathrm{f}}] \to \mathbb{R}^m$ is a control function, and U is an arbitrary subset of $\mathbb{R}^m$. The functions $\chi_j, \psi_j : \mathbb{R}^n \to \mathbb{R}$ impose constraints on the initial and terminal state, respectively, and the functions $\varphi_0, \varphi_{\mathrm{f}} : \mathbb{R}^n \to \mathbb{R}$ model initial and terminal costs. The minimum is taken over all absolutely continuous functions $x(\cdot)$ as well as over all measurable and essentially bounded functions $u(\cdot)$.

Such problems are numerically solved predominantly by gradient-based optimization methods. Firstly, because of their close relation to classic problems from the calculus of variations leading to multiplier-based optimality conditions such as Pontryagin's principle, which then in turn can be solved using Newton-type shooting approaches. Secondly, because of the difficulty that their discretization naturally leads to large-scale finite-dimensional problems being typically solved locally based on finite-dimensional optimality conditions. Hence, these problems are mostly studied for compact and convex constraints on the control values.

In this work, it is important to observe that we do not impose any connectivity or convexity assumptions on the control constraint set U. Thus, (1) includes integrality constraints for, e.g., modeling discrete decisions on some components

Date: November 24, 2022.

2010 *Mathematics Subject Classification.* 49-XX, 49M05, 49M27, 90C11.

Key words and phrases. Mixed-integer optimal control problems, Time-domain decomposition, Mixed-integer nonlinear optimization, Convergence.

of u. Such optimal control problems exhibit combinatorial aspects and received a lot of attention in recent years. Indeed, many applications require to include integer restrictions on the control value as to model, e.g., decisions such as opening or closing valves in gas network operation [21]. Similar logical restrictions occur, e.g., in autonomous driving in case of vehicles with gear shifting [14], in contact problems such as robotic multi-arm transport [6], or in the control of chemical reactors in process engineering [53].

Due to the discrete nature of U, these problems cannot be solved directly with gradient-based methods from optimal control. Therefore, several approaches have been proposed in the literature to overcome this limitation such as partial outer convexification and combinatorial integral approximation [26, 46], alternating direction methods [16], extended multi-interval Legendre–Gauß–Radau pseudospectral methods [38], multiple shooting [47], bilevel approaches [7], adaptive wavelet methods [11], model reduction [25], genetic algorithms [54], mixed-integer linear programming techniques [19], instantaneous control [18], time-transformation methods [13, 32, 44], hybrid extensions of the maximum principle [15, 51, 52], the competing Hamiltonian approach [4], or outer-approximation, Benders-type decomposition, etc. for linear-quadratic problems [43]. Available analytical results concern, e.g., turnpike properties [10] and value-function regularity [17].

Similar to solution approaches based on hybrid extensions of the maximum principle, we consider in this work the infinite-dimensional optimality conditions of (1) for discrete U. To this end, our goal will be to find a Pontryagin-minimum of (1), i.e., an admissible state-control point (x^*, u^*), such that for any constant N there exists an $\varepsilon = \varepsilon(N) > 0$ with the properties that for any admissible point (x, u) satisfying

$$\|x^* - x\|_{\mathcal{C}^0([t_0,t_\mathrm{f}];\mathbb{R}^n)} < \varepsilon, \quad \|u^* - u\|_{\mathcal{L}^1(t_0,t_\mathrm{f};\mathbb{R}^m)} < \varepsilon, \quad \|u^* - u\|_{\mathcal{L}^\infty(t_0,t_\mathrm{f};\mathbb{R}^m)} \leq N,$$

it holds

$$\varphi_0(x^*(t_0)) + \varphi_\mathrm{f}(x^*(t_\mathrm{f})) \leq \varphi_0(x(t_0)) + \varphi_\mathrm{f}(x(t_\mathrm{f}));$$

see, e.g., [8, 42]. Note that global minima are Pontryagin-minima.

We suppose that $f : Q \to \mathbb{R}^n$ is continuous with a continuous partial derivative f_x on a set $Q \subseteq \mathbb{R}^{n+m}$ and the functions $\chi_j, \psi_j, \varphi_0, \varphi_\mathrm{f} : P \to \mathbb{R}$ are continuously differentiable on a set $P \subseteq \mathbb{R}^n$. Here, Q and P are open super-sets of the admissible state-control set and its projection onto the states, respectively.

For $\lambda \in \mathbb{R}^n$, $x \in \mathbb{R}^n$, and $u \in \mathbb{R}^m$ we let

$$H(\lambda, x, u) = \langle \lambda, f(x, u) \rangle_{\mathbb{R}^n}$$

denote the Hamiltonian associated with Problem (1). If (x^*, u^*) is a Pontryagin-minimum of Problem (1), Pontryagin's maximum principle yields that there exists a collection $\eta = (\alpha, \beta, \lambda)$ with $\alpha \geq 0$, $\beta = (\beta_1, \ldots, \beta_{p+q}) \in \mathbb{R}^{p+q}$, and λ being a Lipschitz-function $\lambda : [t_0, t_\mathrm{f}] \to \mathbb{R}^n$ such that the following conditions hold:

(i) $(\alpha, \beta) \neq 0$,
(ii) $\dot\lambda(t) = -H_x(\lambda(t), x^*(t), u^*(t)) = -f_x(x^*(t), u^*(t))^\top \lambda(t)$ a. e. in (t_0, t_f),
(iii) $\lambda(t_0) = \alpha\varphi_0'(x^*(t_0)) + \sum_{j=1}^p \beta_j \chi_j'(x^*(t_0))$,
(iv) $\lambda(t_\mathrm{f}) = -\alpha\varphi_\mathrm{f}'(x^*(t_\mathrm{f})) - \sum_{j=1}^q \beta_{p+j} \psi_j'(x^*(t_\mathrm{f}))$,
(v) $\max_{u \in U} H(\lambda(t), x^*(t), u) = H(\lambda(t), x^*(t), u^*(t))$ a. e. in (t_0, t_f);

see, e.g., [8, Theorem 1]. Note that these conditions are stated in a non-qualified form, i.e., $\alpha = 0$ is possible. However, it is well-known that these conditions hold with $\alpha \neq 0$ if the initial- and end-point constraints are regular in the sense that

$$\operatorname{rank}[\chi_1'(x), \ldots, \chi_p'(x)] = p \quad \text{for all } x \text{ such that } \chi_1(x) = \cdots = \chi_p(x) = 0$$

and

$$\operatorname{rank}[\psi_1'(x),\dots,\psi_q'(x)] = q \quad \text{for all } x \text{ such that } \psi_1(x) = \dots = \psi_q(x) = 0;$$

see, e.g., [1]. Therefore, we assume in the following that the constraints $\chi_1,\dots,\chi_p$ and $\psi_1,\dots,\psi_q$ are regular.

In particular, under this assumption, the Pontryagin maximum principle together with the constraints of Problem (1) yield the following necessary optimality conditions for (x,u) being a Pontryagin-minimum:

$$\begin{aligned}
\dot{x} &= f(x,u) \quad \text{a. e. in } (t_0,t_\mathrm{f}),\\
\chi_j(x(t_0)) &= 0, \quad j = 1,\dots,p,\\
\psi_j(x(t_\mathrm{f})) &= 0, \quad j = 1,\dots,q,\\
u(t) &\in U \quad \text{a. e. in } [t_0,t_\mathrm{f}],\\
\dot{\lambda} &= -f_x(x,u)^\top \lambda \quad \text{a. e. in } (t_0,t_\mathrm{f}),\\
\lambda(t_0) &= \varphi_0'(x(t_0)) + \sum_{j=1}^{p} \beta_j \chi_j'(x(t_0)),\\
\lambda(t_\mathrm{f}) &= -\varphi_\mathrm{f}'(x(t_\mathrm{f})) - \sum_{j=1}^{q} \beta_{p+j} \psi_j'(x(t_\mathrm{f})),\\
\max_{u\in U} H(\lambda(t),x(t),u) &= H(\lambda(t),x(t),u(t)) \quad \text{a. e. in } (t_0,t_\mathrm{f}).
\end{aligned} \tag{2}$$

So far, this is well-known and, indeed, does not require any assumption on U in the original proof in [42] using needle-variations, the terminal cone, and the separating hyperplane theorem as, for instance, also being exploited for the so-called hybrid maximum principle [8]. It is also well-known that Problem (1) being in so-called Mayer form is not restrictive. One can, for instance, include running costs

$$\int_{t_0}^{t_\mathrm{f}} L(x,u)\,\mathrm{d}t$$

in Problem (1) by defining a new state variable y with

$$\begin{aligned}
\dot{y} &= L(x,u) \quad \text{a. e. in } (t_0,t_\mathrm{f}),\\
y(t_0) &= 0
\end{aligned} \tag{3}$$

and adding $y(t_\mathrm{f})$ to the objective function. Nevertheless, we note that our convergence result in Section 4 applies to problems in Mayer form with linear dynamics and a quadratic objective function. Hence, nonlinear terms in $L(x,u)$ that depend on x are not covered. For running costs L that are independent of x, however, (3) is equivalent to considering $H(t,x,u) := \langle \lambda, f(x,u)\rangle_{\mathbb{R}^n} - L(u)$. Besides this, we choose the Mayer form to keep the optimality system as simple as possible. Moreover, a free terminal time can be transformed into a fixed terminal time, etc.; see, e.g., [33]. We also note that the regularity assumptions on f and on the constraint functions can be relaxed [2].

As already indicated above, most approaches in the literature yet focused on solving (2) using further necessary optimality conditions for the pointwise maximization of the Hamiltonian. Additional properties on U are then needed. Most prominently, e.g., for convex sets U, the pointwise maximization of the Hamiltonian can be replaced using Karush–Kuhn–Tucker type optimality conditions and the two-point boundary value problem (2) can then be solved using Newton's method in the fashion of single or multiple shooting. If U is a discrete-continuous set, switching-time re-parameterization can be used. This, however, leaves a gap on

efficient treatments concerning the combinatorial aspects. For an overview of shooting techniques we refer to [5]. In any case, all of these approaches solve (2) to some form of local optimality with respect to the maximization of the Hamiltonian. However, the conditions in (2) hold with global optimality even for (locally optimal) Pontryagin-minima.

This motivates us to consider a solution approach for (2) based on (global) mixed-integer nonlinear programming (MINLP) techniques. This, of course, needs to overcome the difficulty that (2) is an MINLP subject to differential equations with no direct advantage over the original problem (1). Instead, (2) has even more variables for the adjoint equation. In particular, a direct discretization of (2) as, e.g., considered in [20] yields a large-scale MINLP that in many applications becomes computationally intractable.

We overcome this limitation by using time-domain decomposition methods originally developed for partial differential equations (PDEs); see, e.g., [23, 24, 27–31, 34, 35, 37]. The main idea of these methods is to split the time domain $[t_0, t_\mathrm{f}]$ into non-overlapping subintervals and then to iteratively decouple the optimality system (2) such that on each subinterval, subproblems are solved together with conditions at the breakpoints that couple the states over two subsequent iterations. For PDEs, this decomposition approach goes back to the so-called parareal-scheme introduced in [36]. In the following, we show that this approach can also be used to decompose (2) for arbitrary sets U. As in [29–31], we exploit that the decomposition yields subproblems that correspond to so-called virtual control problems on the subintervals. For discrete-continuous U, this yields an iterative scheme of mixed-integer optimal control problems on smaller time horizons. For sufficiently fine decompositions, these can eventually be solved using direct transcription methods such as collocation [22] or Runge–Kutta discretizations [49], both resulting in reasonably sized finite-dimensional mixed-integer nonlinear problems. The limit behavior of such approximations are discussed in [20]. The method can be interpreted as an alternating approach [12] or as a structure-exploiting decomposition for MINLPs rather than a generic one [41].

Our main contribution is that we prove convergence of this iterative scheme for the important case of linear-quadratic problems. We support our theoretical result with encouraging numerical results and also include experiments for nonlinear problems. In particular, we demonstrate that the iterative method can provide solutions for problems where the same solver applied to full direct discretization fails to reach an optimal solution within amply time limits.

The remainder of the paper is structured as follows. In Section 2, we introduce the time-domain decomposition of the optimality system and discuss the iterative idea to recover the transmission conditions at the boundaries of the subintervals. Afterward, we re-interpret the decomposed (primal-dual) optimality conditions as so-called (primal) virtual control problems in Section 3 and state the overall iterative procedure. For the case of linear-quadratic problems, we prove convergence of the method in Section 4. We present some case studies for linear as well as nonlinear problems in Section 5 to give a numerical proof of concept. Finally, we conclude in Section 6.

2. Time-Domain Decomposition of the Optimality System

We now consider the time-domain decomposition

$$[t_0, t_\mathrm{f}] = \bigcup_{k=0}^{K} [t_k, t_{k+1}] \quad \text{with} \quad t_k < t_{k+1},\ t_{K+1} = t_\mathrm{f}$$

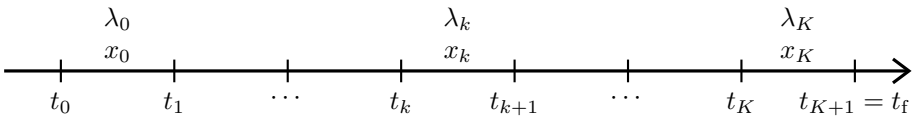

FIGURE 1. Schematic illustration of the time-domain decomposition

of the entire time horizon of Problem (1). Accordingly, we define

$$x_k := x|_{[t_k,t_{k+1}]}, \quad \lambda_k := \lambda|_{[t_k,t_{k+1}]}, \quad \text{and } u_k := u|_{[t_k,t_{k+1}]}$$

for $k = 0, \dots, K$; see Figure 1. The optimality conditions (2) then are equivalent to the following sets of conditions.

(i) For $k = 0$:

$$\dot{x}_0 = f(x_0, u_0) \quad \text{a. e. in } (t_0, t_1), \tag{4a}$$
$$\chi_j(x_0(t_0)) = 0, \quad j = 1, \dots, p, \tag{4b}$$
$$u_0(t) \in U \quad \text{a. e. in } [t_0, t_1], \tag{4c}$$
$$\dot{\lambda}_0 = -f_x(x_0, u_0)^\top \lambda_0 \quad \text{a. e. in } (t_0, t_1), \tag{4d}$$
$$\lambda_0(t_0) = \varphi_0'(x_0(t_0)) + \sum_{j=1}^{p} \beta_j \chi_j'(x_0(t_0)), \tag{4e}$$
$$\max_{u \in U} H(\lambda_0(t), x_0(t), u) = H(\lambda_0(t), x_0(t), u_0(t)) \quad \text{a. e. in } (t_0, t_1). \tag{4f}$$

(ii) For $k = 1, \dots, K-1$:

$$\dot{x}_k = f(x_k, u_k) \quad \text{a. e. in } (t_k, t_{k+1}), \tag{5a}$$
$$u_k(t) \in U \quad \text{a. e. in } [t_k, t_{k+1}], \tag{5b}$$
$$\dot{\lambda}_k = -f_x(x_k, u_k)^\top \lambda_k \quad \text{a. e. in } (t_k, t_{k+1}), \tag{5c}$$
$$\max_{u \in U} H(\lambda_k(t), x_k(t), u) = H(\lambda_k(t), x_k(t), u_k(t)) \quad \text{a. e. in } (t_k, t_{k+1}). \tag{5d}$$

(iii) For $k = K$:

$$\dot{x}_K = f(x_K, u_K) \quad \text{a. e. in } (t_K, t_\mathrm{f}), \tag{6a}$$
$$\psi_j(x_K(t_\mathrm{f})) = 0, \quad j = 1, \dots, q, \tag{6b}$$
$$u_K(t) \in U \quad \text{a. e. in } [t_K, t_\mathrm{f}], \tag{6c}$$
$$\dot{\lambda}_K = -f_x(x_K, u_K)^\top \lambda_K \quad \text{a. e. in } (t_K, t_\mathrm{f}), \tag{6d}$$
$$\lambda_K(t_\mathrm{f}) = -\varphi_\mathrm{f}'(x_K(t_\mathrm{f})) - \sum_{j=1}^{q} \beta_{p+j} \psi_j'(x_K(t_\mathrm{f})), \tag{6e}$$
$$\max_{u \in U} H(\lambda_K(t), x_K(t), u) = H(\lambda_K(t), x_K(t), u_K(t)) \quad \text{a. e. in } (t_K, t_\mathrm{f}). \tag{6f}$$

These conditions are completed with the conditions

$$x_k(t_{k+1}) = x_{k+1}(t_{k+1}), \qquad k = 0, \dots, K-1, \tag{7a}$$
$$\lambda_k(t_k) = \lambda_{k-1}(t_k), \qquad k = 1, \dots, K, \tag{7b}$$

which ensure the continuity of the state x and of the adjoint λ at the boundaries of the sub-intervals.

The main idea of the time-domain decomposition method is to decouple the transmission conditions (7) and to construct an iterative procedure that converges to points that satisfy the decomposed optimality system above. To this end, we

consider an iteration index ℓ, iterates x_k^ℓ, λ_k^ℓ, and u_k^ℓ for x_k, λ_k, and u_k, respectively, and a scalar parameter $\gamma > 0$, to decouple the transmission conditions (7) as

$$x_k^\ell(t_{k+1}) + \gamma\lambda_k^\ell(t_{k+1}) = \phi_{k,k+1}^{\ell-1}, \quad k = 0, \dots, K-1, \tag{8a}$$

$$x_k^\ell(t_k) - \gamma\lambda_k^\ell(t_k) = \phi_{k,k-1}^{\ell-1}, \quad k = 1, \dots, K, \tag{8b}$$

with the update rules

$$\begin{aligned}\phi_{k,k+1}^{\ell-1} &= (1-\varepsilon)\left(x_{k+1}^{\ell-1}(t_{k+1}) + \gamma\lambda_{k+1}^{\ell-1}(t_{k+1})\right)\\ &\quad + \varepsilon\left(x_k^{\ell-1}(t_{k+1}) + \gamma\lambda_k^{\ell-1}(t_{k+1})\right), \quad k = 0, \dots, K-1,\end{aligned} \tag{9a}$$

$$\begin{aligned}\phi_{k,k-1}^{\ell-1} &= (1-\varepsilon)\left(x_{k-1}^{\ell-1}(t_k) - \gamma\lambda_{k-1}^{\ell-1}(t_k)\right)\\ &\quad + \varepsilon\left(x_k^{\ell-1}(t_k) - \gamma\lambda_k^{\ell-1}(t_k)\right), \quad k = 1, \dots, K\end{aligned} \tag{9b}$$

for $\varepsilon \in (0,1)$; see, e.g., [29–31, 34, 35, 37].

Let us make a first observation. To this end, assume that the iterates $x_k^{\ell-1}$, $\lambda_k^{\ell-1}$ for $k = 0, \dots, K$ converge for $\ell \to \infty$ to x_k and λ_k, respectively. Then, substituting $\phi_{k,k+1}$ and $\phi_{k,k-1}$ and dividing by $(1-\varepsilon)$ yields

$$x_k(t_{k+1}) + \gamma\lambda_k(t_{k+1}) = x_{k+1}(t_{k+1}) + \gamma\lambda_{k+1}(t_{k+1}), \tag{10a}$$

$$x_k(t_k) - \gamma\lambda_k(t_k) = x_{k-1}(t_k) - \gamma\lambda_{k-1}(t_k). \tag{10b}$$

Next, we shift the k-index in Equation (10b) by 1 and obtain

$$x_{k+1}(t_{k+1}) - \gamma\lambda_{k+1}(t_{k+1}) = x_k(t_{k+1}) - \gamma\lambda_k(t_{k+1}).$$

Adding this to (10a) yields

$$\gamma\lambda_k(t_{k+1}) - \gamma\lambda_{k+1}(t_{k+1}) = \gamma\lambda_{k+1}(t_{k+1}) - \gamma\lambda_k(t_{k+1}),$$

which is, for $\gamma > 0$, equivalent to

$$\lambda_{k+1}(t_{k+1}) = \lambda_k(t_{k+1}).$$

Finally, (10a) implies

$$x_k(t_{k+1}) = x_{k+1}(t_{k+1}).$$

Hence, we have shown that if the state x_k^ℓ and the adjoint λ_k^ℓ converge, then the iteratively decoupled transmission conditions (8) tend to the continuity conditions (7).

If we combine the decomposed optimality conditions (4)–(6) for all $k = 0, \dots, K$ with the iterative transmission conditions (8), we get

$$\dot{x}_k^\ell = f(x_k^\ell, u_k^\ell) \quad \text{a. e. in } (t_k, t_{k+1}), \tag{11a}$$

$$\chi_j(x_0^\ell(t_0)) = 0 \quad \text{if } k = 0, \tag{11b}$$

$$\psi_j(x_K^\ell(t_\mathrm{f})) = 0 \quad \text{if } k = K, \tag{11c}$$

$$u_k^\ell(t) \in U \quad \text{a. e. in } [t_k, t_{k+1}], \tag{11d}$$

$$\dot{\lambda}_k^\ell = -f_x(x_k^\ell, u_k^\ell)^\top \lambda_k^\ell \quad \text{a. e. in } (t_k, t_{k+1}), \tag{11e}$$

$$\lambda_0^\ell(t_0) = \varphi_0'(x_0^\ell(t_0)) + \sum_{j=1}^{p} \beta_j \chi_j'(x_0^\ell(t_0)) \quad \text{if } k = 0, \tag{11f}$$

$$\lambda_K^\ell(t_\mathrm{f}) = -\varphi_\mathrm{f}'(x_K^\ell(t_\mathrm{f})) - \sum_{j=1}^{q} \beta_{p+j} \psi_j'(x_K^\ell(t_\mathrm{f})) \quad \text{if } k = K, \tag{11g}$$

$$\max_{u \in U} H(\lambda_k^\ell(t), x_k^\ell(t), u) = H(\lambda_k^\ell(t), x_k^\ell(t), u_k^\ell(t)) \quad \text{a. e. in } (t_k, t_{k+1}), \tag{11h}$$

$$x_k^\ell(t_{k+1}) + \gamma\lambda_k^\ell(t_{k+1}) = \phi_{k,k+1}^{\ell-1} \quad \text{if } k \neq K, \tag{11i}$$

$$x_k^\ell(t_k) - \gamma\lambda_k^\ell(t_k) = \phi_{k,k-1}^{\ell-1} \quad \text{if } k \neq 0 \tag{11j}$$

together with the update rules (9).

3. Virtual Control Problems and the Iterative Procedure

We will now observe that the iteratively decomposed optimality systems (11) have a primal interpretation as other optimal control problems with additional and so-called virtual controls h; see, e.g., [29–31, 34, 35, 37]. For the inner time sub-intervals k, the respective primal problem reads

$$\min_{x_k,u_k,h_k} \quad \frac{1}{2\gamma}\|h_k\|^2 + \frac{1}{2\gamma}\|x_k(t_{k+1}) - \phi_{k,k+1}^{\ell-1}\|^2 \tag{12a}$$

$$\text{s.t.} \quad \dot{x}_k = f(x_k,u_k) \quad \text{a. e. in } (t_k,t_{k+1}), \tag{12b}$$

$$x_k(t_k) = \phi_{k,k-1}^{\ell-1} + h_k, \tag{12c}$$

$$h_k \in \mathbb{R}^n, \tag{12d}$$

$$u_k(t) \in U \quad \text{a. e. in } [t_k,t_{k+1}]. \tag{12e}$$

The ODE constraint (12b) and the control constraint (12e) carry over from Problem (1) but are now restricted to the time interval $[t_k,t_{k+1}]$. The virtual controls enter in the initial conditions (12c). These controls allow the state to be bounded away from the current transmission conditions at t_k. The violation of the current transmission conditions at t_k and t_{k+1} is penalized in the objective function. Since this virtual control problem is defined for the inner time sub-intervals, there are no initial or final conditions as in Problem (1) in the constraints or in the objective function. In order to bring (12) into the form of Problem (1), one can remove (12d) and model the virtual control h_k as a trivial state variable, i.e.,

$$h_k(t_k) = h_k, \quad \dot{h}_k = 0 \quad \text{a. e. in } (t_k,t_{k+1}).$$

Technically speaking, h_k is not a control but a constant state. We will, however, keep referring to it as a virtual control to be in line with the pertinent literature; see, e.g., [29, 35]. We use the Hamiltonian

$$H(\lambda_k,x_k,u_k) = \lambda_k^\top f(x_k,u_k),$$

in which we omit the adjoint state corresponding to h_k because it vanishes. The corresponding optimality conditions are given by

$$\dot{x}_k = f(x_k,u_k) \quad \text{a. e. in } (t_k,t_{k+1}), \tag{13a}$$

$$\dot{h}_k = 0 \quad \text{a. e. in } (t_k,t_{k+1}), \tag{13b}$$

$$x_k(t_k) = \phi_{k,k-1}^{\ell-1} + h_k(t_k), \tag{13c}$$

$$u_k(t) \in U \quad \text{a. e. in } [t_k,t_{k+1}], \tag{13d}$$

$$\dot{\lambda}_k = -f_x(x_k,u_k)^\top \lambda_k \quad \text{a. e. in } (t_k,t_{k+1}), \tag{13e}$$

$$\lambda_k(t_k) = \frac{1}{\gamma}h_k(t_k), \tag{13f}$$

$$\lambda_k(t_{k+1}) = -\frac{1}{\gamma}\left(x_k(t_{k+1}) - \phi_{k,k+1}^{\ell-1}\right), \tag{13g}$$

$$\max_{u\in U} H(\lambda_k(t),x_k(t),u) = H(\lambda_k(t),x_k(t),u_k(t)) \quad \text{a. e. in } (t_k,t_{k+1}). \tag{13h}$$

If we substitute $h_k(t_k)$ in (13c) with (13f) and rewrite (13g), we obtain

$$\begin{aligned}
\dot{x}_k &= f(x_k, u_k) \quad \text{a. e. in } (t_k, t_{k+1}),\\
u_k(t) &\in U \quad \text{a. e. in } [t_k, t_{k+1}],\\
\dot{\lambda}_k &= -f_x(x_k, u_k)^\top \lambda_k \quad \text{a. e. in } (t_k, t_{k+1}),\\
x_k(t_k) - \gamma\lambda_k(t_k) &= \phi_{k,k-1}^{\ell-1},\\
x_k(t_{k+1}) + \gamma\lambda_k(t_{k+1}) &= \phi_{k,k+1}^{\ell-1},\\
\max_{u\in U} H(\lambda_k(t), x_k(t), u) &= H(\lambda_k(t), x_k(t), u_k(t)) \quad \text{a. e. in } (t_k, t_{k+1}).
\end{aligned}$$

This exactly corresponds to the iteratively decomposed optimality system (11) for $k = 1, \dots, K-1$. Note that we scaled the objective function (12a) with the factor $1/(2\gamma)$ to be in line with with the notation of System (11). However, any other positive factor would be valid as well.

For the first time sub-interval, i.e., $k = 0$, the respective primal problem reads

$$\begin{aligned}
\min_{x_0,u_0} \quad & \varphi_0(x_0(t_0)) + \frac{1}{2\gamma}\|x_0(t_1) - \phi_{0,1}^{\ell-1}\|^2 && \text{(14a)}\\
\text{s.t.} \quad & \dot{x}_0 = f(x_0, u_0) \quad \text{a. e. in } (t_0, t_1), && \text{(14b)}\\
& \chi_j(x_0(t_0)) = 0, && \text{(14c)}\\
& u_0(t) \in U \quad \text{a. e. in } [t_0, t_1], && \text{(14d)}
\end{aligned}$$

in which we have no transmission conditions at t_0 but the original initial conditions instead. For the last time sub-interval, i.e., $k = K$, we get

$$\begin{aligned}
\min_{x_K,u_K,h_K} \quad & \frac{1}{2\gamma}\|h_K\|^2 + \varphi_{\mathrm{f}}(x_K(t_{\mathrm{f}})) && \text{(15a)}\\
\text{s.t.} \quad & \dot{x}_K = f(x_K, u_K) \quad \text{a. e. in } (t_K, t_{K+1}), && \text{(15b)}\\
& x_K(t_K) = \phi_{K,K-1}^{\ell-1} + h_K, && \text{(15c)}\\
& \psi_j(x_K(t_{\mathrm{f}})) = 0, && \text{(15d)}\\
& h_K \in \mathbb{R}^n, && \text{(15e)}\\
& u_K(t) \in U \quad \text{a. e. in } [t_K, t_{K+1}], && \text{(15f)}
\end{aligned}$$

in which we have no transmission conditions at t_{K+1} but the original final conditions instead.

Now we can state the iterative time-domain decomposition method; see Algorithm 1. In Step 3 the virtual control problems (12), (14), and (15) are solved. This is equivalent to solving the iteratively decomposed optimality systems (11) for all $k = 0, \dots, K$. One can compute the values of the adjoint variables λ_k^ℓ at the transmission points t_k and t_{k+1} by using (11i) and (11j) in Step 4. In Step 6, the update rules in (9) are used to obtain the transmission conditions for the next iteration.

To conclude this section, also note that the solution of the $K+1$ problems in Step 3 of Algorithm 1 can be done in parallel.

4. Convergence Analysis for Linear-Quadratic Problems

In this section, we will prove the convergence of Algorithm 1 in the sense of continuity at the boundaries of the sub-intervals, i.e., we show that

$$\begin{aligned}
x_k^\ell(t_{k+1}) - x_{k+1}^\ell(t_{k+1}) &\to 0,\\
\lambda_k^\ell(t_{k+1}) - \lambda_{k+1}^\ell(t_{k+1}) &\to 0
\end{aligned}$$

Algorithm 1 Iterative Time-Domain Decomposition Method

Require: Problem (1), a time-domain decomposition $t_0 < t_1 < \cdots < t_{K+1} = t_\mathrm{f}$, weights $\gamma > 0$, $\varepsilon \in (0,1)$, and the transmission conditions $\phi^0_{k,k-1} \in \mathbb{R}^n$ for $k = 1, \dots, K$ as well as $\phi^0_{k,k+1} \in \mathbb{R}^n$ for $k = 0, \dots, K-1$.

1: **for** $\ell = 1, 2, 3, \dots$ **do**
2: **for** $k = 0, \dots, K$ **do**
3: Solve Problem (12), (14) if $k = 0$, or (15) if $k = K$.
4: Set

$$\lambda^\ell_k(t_{k+1}) = -\frac{1}{\gamma}\left(x^\ell_k(t_{k+1}) - \phi^{\ell-1}_{k,k+1}\right),$$
$$\lambda^\ell_k(t_k) = \frac{1}{\gamma}\left(x^\ell_k(t_k) - \phi^{\ell-1}_{k,k-1}\right).$$

5: **end for**
6: Set

$$\begin{aligned}\phi^\ell_{k,k+1} &= (1-\varepsilon)\left(x^\ell_{k+1}(t_{k+1}) + \gamma\lambda^\ell_{k+1}(t_{k+1})\right)\\ &\quad + \varepsilon\left(x^\ell_k(t_{k+1}) + \gamma\lambda^\ell_k(t_{k+1})\right), \quad k = 0, \dots, K-1,\\ \phi^\ell_{k,k-1} &= (1-\varepsilon)\left(x^\ell_{k-1}(t_k) - \gamma\lambda^\ell_{k-1}(t_k)\right)\\ &\quad + \varepsilon\left(x^\ell_k(t_k) - \gamma\lambda^\ell_k(t_k)\right), \quad k = 1, \dots, K.\end{aligned}$$

7: **end for**

holds for all $k = 0, \dots, K-1$ as $\ell \to \infty$. To this end, we restrict ourselves to the case of linear dynamics with objective functions being quadratic with respect to the state and include control costs, i.e., we consider

$$\min \quad \frac{1}{2}x(t_0)^\top Q_0 x(t_0) + q_0^\top x(t_0) + \frac{1}{2}x(t_\mathrm{f})^\top Q_\mathrm{f} x(t_\mathrm{f}) + q_\mathrm{f}^\top x(t_\mathrm{f}) + \int_{t_0}^{t_\mathrm{f}} L(u)\,\mathrm{d}t \tag{16a}$$
$$\text{s.t.} \quad \dot{x} = Ax + Bu + c \quad \text{a. e. in } (t_0, t_\mathrm{f}), \tag{16b}$$
$$R_0 x(t_0) = c_0, \tag{16c}$$
$$R_\mathrm{f} x(t_\mathrm{f}) = c_\mathrm{f}, \tag{16d}$$
$$u(t) \in U \quad \text{a. e. in } [t_0, t_\mathrm{f}] \tag{16e}$$

with $U \subseteq \mathbb{R}^m$ arbitrary, $x : [t_0, t_\mathrm{f}] \to \mathbb{R}^n$, $u : [t_0, t_\mathrm{f}] \to \mathbb{R}^m$, $A \in \mathbb{R}^{n\times n}$, $Q_0, Q_\mathrm{f} \in \mathbb{R}^{n\times n}$ being positive semi-definite, $B \in \mathbb{R}^{n\times m}$, $c, q_0, q_\mathrm{f} \in \mathbb{R}^n$, $R_0 \in \mathbb{R}^{p\times n}$, $R_\mathrm{f} \in \mathbb{R}^{q\times n}$, $c_0 \in \mathbb{R}^p$, $c_\mathrm{f} \in \mathbb{R}^q$, and $L(u)$ being continuous on an open super-set of the admissible controls in $\mathbb{R}^m$. Note, that we are still dealing with an arbitrary control set U. The Hamiltonian is given by

$$H(\lambda, x, u) = \langle \lambda, f(x,u)\rangle - L(u);$$

cf. (3) and the subsequent remark.

The decomposed optimality conditions (4)–(6) of Problem (16) for $k = 0, \dots, K$ read

$$\dot{x}_k = Ax_k + Bu_k + c \quad \text{a. e. in } (t_k, t_{k+1}), \tag{17a}$$
$$R_0 x_0(t_0) = c_0 \quad \text{if } k = 0, \tag{17b}$$
$$R_\mathrm{f} x_K(t_\mathrm{f}) = c_\mathrm{f} \quad \text{if } k = K, \tag{17c}$$
$$u_k(t) \in U \quad \text{a. e. in } [t_k, t_{k+1}], \tag{17d}$$
$$\dot{\lambda}_k = -A^\top \lambda_k \quad \text{a. e. in } (t_k, t_{k+1}), \tag{17e}$$
$$\lambda_0(t_0) = Q_0 x_0(t_0) + q_0 + R_0^\top (\beta_j)_{j=1}^p \quad \text{if } k = 0, \tag{17f}$$
$$\lambda_K(t_\mathrm{f}) = -Q_\mathrm{f} x_K(t_\mathrm{f}) - q_\mathrm{f} - R_\mathrm{f}^\top (\beta_j)_{j=p+1}^{p+q} \quad \text{if } k = K, \tag{17g}$$
$$\max_{u \in U} H(\lambda_k(t), x_k(t), u) = H(\lambda_k(t), x_k(t), u_k(t)) \quad \text{a. e. in } (t_k, t_{k+1}), \tag{17h}$$
$$x_k(t_k) = x_{k-1}(t_k) \quad \text{if } k \neq 0, \tag{17i}$$
$$\lambda_k(t_{k+1}) = \lambda_{k+1}(t_{k+1}) \quad \text{if } k \neq K. \tag{17j}$$

The iterative version (11) is now given by

$$\dot{x}_k^\ell = Ax_k^\ell + Bu_k^\ell + c \quad \text{a. e. in } (t_k, t_{k+1}), \tag{18a}$$
$$R_0 x_0^\ell(t_0) = c_0 \quad \text{if } k = 0, \tag{18b}$$
$$R_\mathrm{f} x_K^\ell(t_\mathrm{f}) = c_\mathrm{f} \quad \text{if } k = K, \tag{18c}$$
$$u_k^\ell(t) \in U \quad \text{a. e. in } [t_k, t_{k+1}], \tag{18d}$$
$$\dot{\lambda}_k^\ell = -A^\top \lambda_k^\ell \quad \text{a. e. in } (t_k, t_{k+1}), \tag{18e}$$
$$\lambda_0^\ell(t_0) = Q_0 x_0^\ell(t_0) + q_0 + R_0^\top (\beta_j)_{j=1}^p \quad \text{if } k = 0, \tag{18f}$$
$$\lambda_K^\ell(t_\mathrm{f}) = -Q_\mathrm{f} x_K^\ell(t_\mathrm{f}) - q_\mathrm{f} - R_\mathrm{f}^\top (\beta_j)_{j=p+1}^{p+q} \quad \text{if } k = K, \tag{18g}$$
$$\max_{u \in U} H(\lambda_k^\ell(t), x_k^\ell(t), u) = H(\lambda_k^\ell(t), x_k^\ell(t), u_k^\ell(t)) \quad \text{a. e. in } (t_k, t_{k+1}), \tag{18h}$$
$$x_k^\ell(t_{k+1}) + \gamma \lambda_k^\ell(t_{k+1}) = \phi_{k,k+1}^{\ell-1} \quad \text{if } k \neq K, \tag{18i}$$
$$x_k^\ell(t_k) - \gamma \lambda_k^\ell(t_k) = \phi_{k,k-1}^{\ell-1} \quad \text{if } k \neq 0 \tag{18j}$$

for $\ell \in \mathbb{N}$ with the update rules (9).

We assume that (17) and (18) have a solution. In general, it is not to be expected that this solution is unique, which is one of the main differences to the assumptions made for similar time-domain decompositions of PDEs in, e.g., [29–31]. Let (x_k, u_k, λ_k) be a solution of (17) and let $(x_k^\ell, u_k^\ell, \lambda_k^\ell)$ be a solution of the iteratively decomposed optimality conditions (18) for $k = 0, \dots, K$ and $\ell \in \mathbb{N}$. We introduce the errors

$$\tilde{x}_k^\ell := x_k^\ell - x_k, \quad \tilde{u}_k^\ell := u_k^\ell - u_k, \quad \tilde{\lambda}_k^\ell := \lambda_k^\ell - \lambda_k.$$

These errors satisfy the conditions

$$\dot{\tilde{x}}_k^\ell = A\tilde{x}_k^\ell + B\tilde{u}_k^\ell \quad \text{a. e. in } (t_k, t_{k+1}), \tag{19a}$$
$$R_0\tilde{x}_0^\ell(t_0) = 0 \quad \text{if } k = 0, \tag{19b}$$
$$R_\mathrm{f}\tilde{x}_K^\ell(t_\mathrm{f}) = 0 \quad \text{if } k = K, \tag{19c}$$
$$\tilde{u}_k^\ell(t) \in \tilde{U} \quad \text{a. e. in } [t_k, t_{k+1}], \tag{19d}$$
$$\dot{\tilde{\lambda}}_k^\ell = -A^\top\tilde{\lambda}_k^\ell \quad \text{a. e. in } (t_k, t_{k+1}), \tag{19e}$$
$$\tilde{\lambda}_0^\ell(t_0) = Q_0\tilde{x}_0^\ell(t_0) \quad \text{if } k = 0, \tag{19f}$$
$$\tilde{\lambda}_K^\ell(t_\mathrm{f}) = -Q_\mathrm{f}\tilde{x}_K^\ell(t_\mathrm{f}) \quad \text{if } k = K, \tag{19g}$$
$$\tilde{x}_k^\ell(t_{k+1}) + \gamma\tilde{\lambda}_k^\ell(t_{k+1}) = \tilde{\phi}_{k,k+1}^{\ell-1} \quad \text{if } k \neq K, \tag{19h}$$
$$\tilde{x}_k^\ell(t_k) - \gamma\tilde{\lambda}_k^\ell(t_k) = \tilde{\phi}_{k,k-1}^{\ell-1} \quad \text{if } k \neq 0 \tag{19i}$$

together with the update rules

$$\begin{aligned}\tilde{\phi}_{k,k+1}^{\ell-1} = {}& (1-\varepsilon)\left(\tilde{x}_{k+1}^{\ell-1}(t_{k+1}) + \gamma\tilde{\lambda}_{k+1}^{\ell-1}(t_{k+1})\right)\\ &+ \varepsilon\left(\tilde{x}_k^{\ell-1}(t_{k+1}) + \gamma\tilde{\lambda}_k^{\ell-1}(t_{k+1})\right),\end{aligned} \tag{20a}$$
$$\begin{aligned}\tilde{\phi}_{k,k-1}^{\ell-1} = {}& (1-\varepsilon)\left(\tilde{x}_{k-1}^{\ell-1}(t_k) - \gamma\tilde{\lambda}_{k-1}^{\ell-1}(t_k)\right)\\ &+ \varepsilon\left(\tilde{x}_k^{\ell-1}(t_k) - \gamma\tilde{\lambda}_k^{\ell-1}(t_k)\right).\end{aligned} \tag{20b}$$

Moreover, we set

$$\tilde{U} := U - U = \{u - v \colon u, v \in U\}.$$

To derive System (19) we use that both (17) and (18) are linear. This is the case, because Problem (16) has linear dynamics and in the objective function, the only quadratic terms in x are initial and terminal costs. There are approaches to extend this method to more general nonlinear right-hand sides $f(x, u)$ by applying Lipschitz-type conditions on the nonlinearities and their derivatives; see, e.g., [27] where this was done for hyperbolic semilinear PDEs. This, however, exceeds the scope of this work.

First, we prove the following result.

Lemma 1. *For all $k = 0, \dots, K$ and $\ell \in \mathbb{N}$, it holds*

$$\left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \geq 0, \quad \text{a. e. in } (t_k, t_{k+1}). \tag{21}$$

Proof. From (17h) we have

$$\begin{aligned}
& H(\lambda_k, x_k, u_k) \geq H(\lambda_k, x_k, u)\\
\iff \quad & \langle \lambda_k, Ax_k + Bu_k + c\rangle - L(u_k) \geq \langle \lambda_k, Ax_k + Bu + c\rangle - L(u)\\
\iff \quad & \langle \lambda_k, Bu_k\rangle - L(u_k) \geq \langle \lambda_k, Bu\rangle - L(u)\\
\iff \quad & \langle \lambda_k, B(u_k - u)\rangle \geq L(u_k) - L(u)
\end{aligned}$$

for all $u \in U$ a. e. in (t_k, t_{k+1}). Analogously, using (18h) we get

$$\left\langle \lambda_k^\ell, B(u_k^\ell - u)\right\rangle \geq L(u_k^\ell) - L(u) \tag{22}$$

for all $u \in U$ a. e. in (t_k, t_{k+1}). If we set $u = u_k$ in the last inequality, we can write

$$\begin{aligned}
& \langle \lambda_k^\ell, B(u_k^\ell - u_k) \rangle \geq L(u_k^\ell) - L(u_k) \\
\iff \quad & \langle \lambda_k^\ell - \lambda_k + \lambda_k, B(u_k^\ell - u_k) \rangle \geq L(u_k^\ell) - L(u_k) \\
\iff \quad & \left\langle \tilde{\lambda}_k^\ell, B(u_k^\ell - u_k) \right\rangle - \langle \lambda_k, B(u_k - u_k^\ell) \rangle \geq L(u_k^\ell) - L(u_k) \\
\iff \quad & \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle - \big(L(u_k) - L(u_k^\ell)\big) \geq L(u_k^\ell) - L(u_k) \\
\iff \quad & \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \geq 0
\end{aligned}$$

a. e. in (t_k, t_{k+1}). □

Definition 1. *We define the state of all updates in iteration ℓ as*

$$\begin{aligned}
X^\ell &:= \left(\left(\tilde{\phi}_{k,k+1}^{\ell-1} \right)_{k=0}^{K-1}, \left(\tilde{\phi}_{k,k-1}^{\ell-1} \right)_{k=1}^{K} \right) \\
&= \left(\left(\tilde{x}_k^\ell(t_{k+1}) + \gamma \tilde{\lambda}_k^\ell(t_{k+1}) \right)_{k=0}^{K-1}, \left(\tilde{x}_k^\ell(t_k) - \gamma \tilde{\lambda}_k^\ell(t_k) \right)_{k=1}^{K} \right).
\end{aligned}$$

Moreover, we define the mapping

$$\begin{aligned}
TX^\ell &:= \left(\left(\tilde{\phi}_{k,k+1}^{\ell} \right)_{k=0}^{K-1}, \left(\tilde{\phi}_{k,k-1}^{\ell} \right)_{k=1}^{K} \right) \\
&= \left(\left(\tilde{x}_{k+1}^\ell(t_{k+1}) + \gamma \tilde{\lambda}_{k+1}^\ell(t_{k+1}) \right)_{k=0}^{K-1}, \left(\tilde{x}_{k-1}^\ell(t_k) - \gamma \tilde{\lambda}_{k-1}^\ell(t_k) \right)_{k=1}^{K} \right).
\end{aligned}$$

With this notation at hand and I denoting the identity, the update in Step 6 of Algorithm 1 is given by the relaxed mapping $T_\varepsilon := (1-\varepsilon)T + \varepsilon I$, i.e.,

$$T_\varepsilon X^\ell = (1-\varepsilon) T X^\ell + \varepsilon X^\ell.$$

Consequently, we consider the fixed point iteration

$$X^{\ell+1} = T_\varepsilon X^\ell. \tag{23}$$

Clearly, X is a fixed point of T and T_ε if and only if the transmission conditions for the errors

$$\begin{aligned}
\tilde{x}_k(t_{k+1}) &= \tilde{x}_{k+1}(t_{k+1}), && k = 0, \dots, K-1, \\
\tilde{\lambda}_k(t_k) &= \tilde{\lambda}_{k-1}(t_k), && k = 1, \dots, K,
\end{aligned}$$

are satisfied. The errors fulfill these transmission conditions if and only if x_k^ℓ and λ_k^ℓ from the iteratively decomposed optimality conditions (18) do, too. This is only the case in a solution of (17). Since we assume that (17) has at least one solution, T and T_ε have a fixed point.

Remark 1. *For any*

$$\bar{X} = \left((\bar{\phi}_{k,k+1})_{k=0}^{K-1}, (\bar{\phi}_{k,k-1})_{k=1}^{K} \right) \in \mathbb{R}^{2K},$$

the corresponding right-hand sides of the transmission conditions (18i) *and* (18j) *are given by*

$$\begin{aligned}
\phi_{k,k+1} &= \bar{\phi}_{k,k+1} + x_k(t_{k+1}) + \gamma \lambda_k(t_{k+1}), \\
\phi_{k,k-1} &= \bar{\phi}_{k,k-1} + x_k(t_k) - \gamma \lambda_k(t_k),
\end{aligned}$$

where (x_k, u_k, λ_k) is the chosen solution of (17). *If System* (18) *has a solution $(\bar{x}_k, \bar{u}_k, \bar{\lambda}_k)$, then the errors $(\tilde{x}_k, \tilde{u}_k, \tilde{\lambda}_k) = (\bar{x}_k, \bar{u}_k, \bar{\lambda}_k) - (x_k, u_k, \lambda_k)$ are well-defined and, thus, the mapping $T : \mathbb{R}^{2K} \to \mathbb{R}^{2K}$ is well-defined, too.*

Definition 2. *We define the energies*

$$\mathcal{E}_k^\ell(t) := \left\|\tilde{x}_k^\ell(t)\right\|^2 + \gamma^2 \left\|\tilde{\lambda}_k^\ell(t)\right\|^2,$$
$$\mathcal{E}^\ell := \sum_{k=0}^{K-1} \mathcal{E}_k^\ell(t_{k+1}) + \mathcal{E}_{k+1}^\ell(t_{k+1}),$$
$$\mathcal{F}^\ell := 2\gamma \sum_{k=0}^{K} \left(\int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t \right) + 2\gamma \left(\tilde{x}_0^\ell(t_0)^\top Q_0 \tilde{x}_0^\ell(t_0) + \tilde{x}_K^\ell(t_\mathrm{f})^\top Q_\mathrm{f} \tilde{x}_K^\ell(t_\mathrm{f}) \right).$$

Lemma 2. *For all $\ell \in \mathbb{N}$, the energies $\mathcal{E}^\ell$ and $\mathcal{F}^\ell$ are non-negative, i.e.,*

$$\mathcal{E}^\ell \geq 0, \quad \mathcal{F}^\ell \geq 0$$

holds.

Proof. It is clear that $\mathcal{E}^\ell$ is non-negative because it is a sum of norms. The non-negativity of $\mathcal{F}^\ell$ follows from Lemma 1 and of Q_0 and Q_f being positive semi-definite. □

Next, we show that we can use the energies to describe the state of all updates.

Lemma 3. *It holds*

$$\left\|X^\ell\right\|^2 = \mathcal{E}^\ell + \mathcal{F}^\ell.$$

Proof. We multiply the state equation (19a) by $\tilde{\lambda}_k^\ell$ and integrate to obtain

$$\begin{aligned}
0 &= \int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, \dot{\tilde{x}}_k^\ell - A\tilde{x}_k^\ell - B\tilde{u}_k^\ell \right\rangle \mathrm{d}t \\
&= \left[\left\langle \tilde{\lambda}_k^\ell, \tilde{x}_k^\ell \right\rangle \right]_{t_k}^{t_{k+1}} - \int_{t_k}^{t_{k+1}} \left\langle \dot{\tilde{\lambda}}_k^\ell, \tilde{x}_k^\ell \right\rangle \mathrm{d}t - \int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, A\tilde{x}_k^\ell \right\rangle + \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t \\
&= \left[\left\langle \tilde{\lambda}_k^\ell, \tilde{x}_k^\ell \right\rangle \right]_{t_k}^{t_{k+1}} - \int_{t_k}^{t_{k+1}} \left\langle -A^\top \tilde{\lambda}_k^\ell, \tilde{x}_k^\ell \right\rangle \mathrm{d}t - \int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, A\tilde{x}_k^\ell \right\rangle + \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t \\
&= \left\langle \tilde{\lambda}_k^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \right\rangle - \left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle - \int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t.
\end{aligned}$$

It follows

$$\left\langle \tilde{\lambda}_k^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \right\rangle = \left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle + \int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t$$

and

$$\left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle = \left\langle \tilde{\lambda}_k^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \right\rangle - \int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t.$$

Now, we can write

$$\begin{aligned}
\left\|X^\ell\right\|^2 &= \sum_{k=0}^{K-1} \left\|\tilde{x}_k^\ell(t_{k+1}) + \gamma\tilde{\lambda}_k^\ell(t_{k+1})\right\|^2 + \sum_{k=1}^{K} \left\|\tilde{x}_k^\ell(t_k) - \gamma\tilde{\lambda}_k^\ell(t_k)\right\|^2 \\
&= \sum_{k=0}^{K-1} \left(\left\|\tilde{x}_k^\ell(t_{k+1})\right\|^2 + \gamma^2 \left\|\tilde{\lambda}_k^\ell(t_{k+1})\right\|^2 + 2\gamma \left\langle \tilde{\lambda}_k^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \right\rangle \right) \\
&\quad + \sum_{k=1}^{K} \left(\left\|\tilde{x}_k^\ell(t_k)\right\|^2 + \gamma^2 \left\|\tilde{\lambda}_k^\ell(t_k)\right\|^2 - 2\gamma \left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle \right) \\
&= \sum_{k=0}^{K-1} \left(\mathcal{E}_k^\ell(t_{k+1}) + 2\gamma \left\langle \tilde{\lambda}_k^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \right\rangle \right)
\end{aligned}$$

$$
\begin{aligned}
&\quad + \sum_{k=1}^{K} \left(\mathcal{E}_k^\ell(t_k) - 2\gamma \left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle \right) \\
&= \sum_{k=0}^{K-1} \left(\mathcal{E}_k^\ell(t_{k+1}) + \mathcal{E}_{k+1}^\ell(t_{k+1}) \right) \\
&\quad + \sum_{k=0}^{K-1} 2\gamma \left\langle \tilde{\lambda}_k^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \right\rangle - \sum_{k=1}^{K} 2\gamma \left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle \\
&= \mathcal{E}^\ell + 2\gamma \sum_{k=0}^{K-1} \left\langle \tilde{\lambda}_k^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \right\rangle - 2\gamma \sum_{k=1}^{K} \left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle \\
&= \mathcal{E}^\ell + 2\gamma \sum_{k=0}^{K-1} \left(\left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle + \int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t \right) \\
&\quad - 2\gamma \sum_{k=1}^{K} \left\langle \tilde{\lambda}_k^\ell(t_k), \tilde{x}_k^\ell(t_k) \right\rangle \\
&= \mathcal{E}^\ell + 2\gamma \left\langle \tilde{\lambda}_0^\ell(t_0), \tilde{x}_0^\ell(t_0) \right\rangle + 2\gamma \sum_{k=0}^{K-1} \left(\int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t \right) \\
&\quad - 2\gamma \left\langle \tilde{\lambda}_K^\ell(t_K), \tilde{x}_K^\ell(t_K) \right\rangle \\
&= \mathcal{E}^\ell + 2\gamma \left\langle \tilde{\lambda}_0^\ell(t_0), \tilde{x}_0^\ell(t_0) \right\rangle + 2\gamma \sum_{k=0}^{K-1} \left(\int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t \right) \\
&\quad - 2\gamma \left(\left\langle \tilde{\lambda}_K^\ell(t_{K+1}), \tilde{x}_K^\ell(t_{K+1}) \right\rangle - \int_{t_K}^{t_{K+1}} \left\langle \tilde{\lambda}_K^\ell, B\tilde{u}_K^\ell \right\rangle \mathrm{d}t \right) \\
&= \mathcal{E}^\ell + 2\gamma \sum_{k=0}^{K} \left(\int_{t_k}^{t_{k+1}} \left\langle \tilde{\lambda}_k^\ell, B\tilde{u}_k^\ell \right\rangle \mathrm{d}t \right) \\
&\quad + 2\gamma \left(\left\langle Q_0 \tilde{x}_0^\ell(t_0), \tilde{x}_0^\ell(t_0) \right\rangle + \left\langle Q_\mathrm{f} \tilde{x}_K^\ell(t_\mathrm{f}), \tilde{x}_K^\ell(t_\mathrm{f}) \right\rangle \right) \\
&= \mathcal{E}^\ell + \mathcal{F}^\ell. \qquad \square
\end{aligned}
$$

We get a similar result if we apply the mapping T first.

Lemma 4. *It holds*

$$\left\| TX^\ell \right\|^2 = \mathcal{E}^\ell - \mathcal{F}^\ell.$$

Proof. Similarly to the proof of Lemma 3 we can write

$$\begin{aligned}
\left\|TX^\ell\right\|^2 &= \sum_{k=0}^{K-1}\left\|\tilde{x}^\ell_{k+1}(t_{k+1})+\gamma\tilde{\lambda}^\ell_{k+1}(t_{k+1})\right\|^2+\sum_{k=1}^{K}\left\|\tilde{x}^\ell_{k-1}(t_k)-\gamma\tilde{\lambda}^\ell_{k-1}(t_k)\right\|^2\\
&= \sum_{k=0}^{K-1}\left(\left\|\tilde{x}^\ell_{k+1}(t_{k+1})\right\|^2+\gamma^2\left\|\tilde{\lambda}^\ell_{k+1}(t_{k+1})\right\|^2+2\gamma\left\langle\tilde{\lambda}^\ell_{k+1}(t_{k+1}),\tilde{x}^\ell_{k+1}(t_{k+1})\right\rangle\right)\\
&\quad+\sum_{k=1}^{K}\left(\left\|\tilde{x}^\ell_{k-1}(t_k)\right\|^2+\gamma^2\left\|\tilde{\lambda}^\ell_{k-1}(t_k)\right\|^2-2\gamma\left\langle\tilde{\lambda}^\ell_{k-1}(t_k),\tilde{x}^\ell_{k-1}(t_k)\right\rangle\right)\\
&= \sum_{k=0}^{K-1}\left(\mathcal{E}^\ell_{k+1}(t_{k+1})+2\gamma\left\langle\tilde{\lambda}^\ell_{k+1}(t_{k+1}),\tilde{x}^\ell_{k+1}(t_{k+1})\right\rangle\right)\\
&\quad+\sum_{k=1}^{K}\left(\mathcal{E}^\ell_{k-1}(t_k)-2\gamma\left\langle\tilde{\lambda}^\ell_{k-1}(t_k),\tilde{x}^\ell_{k-1}(t_k)\right\rangle\right)\\
&= \sum_{k=0}^{K-1}\mathcal{E}^\ell_{k+1}(t_{k+1})+\sum_{k=1}^{K}2\gamma\left\langle\tilde{\lambda}^\ell_{k}(t_{k}),\tilde{x}^\ell_{k}(t_{k})\right\rangle\\
&\quad+\sum_{k=0}^{K-1}\left(\mathcal{E}^\ell_{k}(t_{k+1})-2\gamma\left\langle\tilde{\lambda}^\ell_{k}(t_{k+1}),\tilde{x}^\ell_{k}(t_{k+1})\right\rangle\right)\\
&= \mathcal{E}^\ell+2\gamma\sum_{k=1}^{K}\left\langle\tilde{\lambda}^\ell_{k}(t_{k}),\tilde{x}^\ell_{k}(t_{k})\right\rangle-2\gamma\sum_{k=0}^{K-1}\left\langle\tilde{\lambda}^\ell_{k}(t_{k+1}),\tilde{x}^\ell_{k}(t_{k+1})\right\rangle\\
&= \mathcal{E}^\ell-\mathcal{F}^\ell.
\end{aligned}$$

□

Remark 2. *A direct consequence of Lemma 3 and 4 is*

$$\left\|TX^\ell\right\|^2=\left\|X^\ell\right\|^2-2\mathcal{F}^\ell.$$

Since $\mathcal{F}^\ell$ is non-negative, it follows

$$\left\|TX^\ell\right\|\le\left\|X^\ell\right\|. \tag{24}$$

Hence, it is sufficient for Algorithm 1 to be well-defined that (18) has a solution for $X\in\{x\in\mathbb{R}^{2K}\colon\|x\|\le\|X^1\|\}$ instead of for all vectors in $\mathbb{R}^{2K}$.

Next, we observe that T is a non-expansive mapping.

Lemma 5. *The mapping $T:\mathbb{R}^{2K}\to\mathbb{R}^{2K}$ satisfies*

$$\|TX^1-TX^2\|\le\|X^1-X^2\|$$

for all $X^1,X^2\in\mathbb{R}^{2K}$.

Proof. Let $(x^1_k,u^1_k,\lambda^1_k)$ and $(x^2_k,u^2_k,\lambda^2_k)$ be the the solutions of (18) corresponding to X^1 and X^2. The errors that solve (19) are given by

$$\begin{aligned}
(\tilde{x}^1_k,\tilde{u}^1_k,\tilde{\lambda}^1_k)&=(x^1_k,u^1_k,\lambda^1_k)-(x_k,u_k,\lambda_k),\\
(\tilde{x}^2_k,\tilde{u}^2_k,\tilde{\lambda}^2_k)&=(x^2_k,u^2_k,\lambda^2_k)-(x_k,u_k,\lambda_k).
\end{aligned}$$

We define the differences

$$(y_k,v_k,\mu_k):=(\tilde{x}^1_k,\tilde{u}^1_k,\tilde{\lambda}^1_k)-(\tilde{x}^2_k,\tilde{u}^2_k,\tilde{\lambda}^2_k)=(x^1_k,u^1_k,\lambda^1_k)-(x^2_k,u^2_k,\lambda^2_k)$$

for all $k = 0, \dots, K$. Because of the linear nature of (19), these differences fulfill the system

$$\dot{y}_k = Ay_k + Bv_k \quad \text{a. e. in } (t_k, t_{k+1}), \tag{25a}$$
$$R_0 y_0(t_0) = 0 \quad \text{if } k = 0, \tag{25b}$$
$$R_\mathrm{f} y_K(t_\mathrm{f}) = 0 \quad \text{if } k = K, \tag{25c}$$
$$\dot{\mu}_k = -A^\top \mu_k \quad \text{a. e. in } (t_k, t_{k+1}), \tag{25d}$$
$$\mu_0(t_0) = Q_0 y_0(t_0) \quad \text{if } k = 0, \tag{25e}$$
$$\mu_K(t_\mathrm{f}) = -Q_\mathrm{f} y_K(t_\mathrm{f}) \quad \text{if } k = K, \tag{25f}$$
$$y_k(t_{k+1}) + \gamma\mu_k(t_{k+1}) = \tilde{\phi}^1_{k,k+1} - \tilde{\phi}^2_{k,k+1} \quad \text{if } k \neq K, \tag{25g}$$
$$y_k(t_k) - \gamma\mu_k(t_k) = \tilde{\phi}^1_{k,k-1} - \tilde{\phi}^2_{k,k-1} \quad \text{if } k \neq 0. \tag{25h}$$

Now, we have

$$\begin{aligned}
&\left\|TX^1 - TX^2\right\|^2 \\
&= \left\| \left(\left(\tilde{x}^1_{k+1}(t_{k+1}) + \gamma\tilde{\lambda}^1_{k+1}(t_{k+1})\right)_{k=0}^{K-1}, \left(\tilde{x}^1_{k-1}(t_k) - \gamma\tilde{\lambda}^1_{k-1}(t_k)\right)_{k=1}^{K} \right) \right. \\
&\qquad \left. - \left(\left(\tilde{x}^2_{k+1}(t_{k+1}) + \gamma\tilde{\lambda}^2_{k+1}(t_{k+1})\right)_{k=0}^{K-1}, \left(\tilde{x}^2_{k-1}(t_k) - \gamma\tilde{\lambda}^2_{k-1}(t_k)\right)_{k=1}^{K} \right) \right\|^2 \\
&= \left\| \left((y_{k+1}(t_{k+1}) + \gamma\mu_{k+1}(t_{k+1}))_{k=0}^{K-1}, (y_{k-1}(t_k) - \gamma\mu_{k-1}(t_k))_{k=1}^{K} \right) \right\|^2 \\
&= \sum_{k=0}^{K-1} \left(y_{k+1}(t_{k+1}) + \gamma\mu_{k+1}(t_{k+1})\right)^2 + \sum_{k=1}^{K} \left(y_{k-1}(t_k) - \gamma\mu_{k-1}(t_k)\right)^2 \\
&= \sum_{k=1}^{K} \left(y_k(t_k) + \gamma\mu_k(t_k)\right)^2 + \sum_{k=0}^{K-1} \left(y_k(t_{k+1}) - \gamma\mu_k(t_{k+1})\right)^2 \\
&= \sum_{k=1}^{K} \left(\left(y_k(t_k) - \gamma\mu_k(t_k)\right)^2 + 4\left\langle y_k(t_k), \gamma\mu_k(t_k)\right\rangle \right) \\
&\qquad + \sum_{k=0}^{K-1} \left(\left(y_k(t_{k+1}) + \gamma\mu_k(t_{k+1})\right)^2 - 4\left\langle y_k(t_{k+1}), \gamma\mu_k(t_{k+1})\right\rangle \right) \\
&= \left\|X^1 - X^2\right\|^2 + 4\sum_{k=1}^{K} \left\langle y_k(t_k), \gamma\mu_k(t_k)\right\rangle - 4\sum_{k=0}^{K-1} \left\langle y_k(t_{k+1}), \gamma\mu_k(t_{k+1})\right\rangle .
\end{aligned}$$

Next, we investigate the relation between $\langle y_k(t_k), \mu_k(t_k)\rangle$ and $\langle y_k(t_{k+1}), \mu_k(t_{k+1})\rangle$. To this end, we multiply (25a) with μ_k and integrate to get

$$\begin{aligned}
0 &= \int_{t_k}^{t_{k+1}} \langle \dot{y}_k, \mu_k\rangle - \langle Ay_k, \mu_k\rangle - \langle Bv_k, \mu_k\rangle \,\mathrm{d}t \\
&= \langle y_k(t_{k+1}), \mu_k(t_{k+1})\rangle - \langle y_k(t_k), \mu_k(t_k)\rangle \\
&\qquad - \int_{t_k}^{t_{k+1}} \langle y_k, \dot{\mu}_k\rangle + \langle Ay_k, \mu_k\rangle + \langle Bv_k, \mu_k\rangle \,\mathrm{d}t \\
&= \langle y_k(t_{k+1}), \mu_k(t_{k+1})\rangle - \langle y_k(t_k), \mu_k(t_k)\rangle \\
&\qquad - \int_{t_k}^{t_{k+1}} \langle y_k, -A^\top\mu_k\rangle + \langle Ay_k, \mu_k\rangle + \langle Bv_k, \mu_k\rangle \,\mathrm{d}t \\
&= \langle y_k(t_{k+1}), \mu_k(t_{k+1})\rangle - \langle y_k(t_k), \mu_k(t_k)\rangle - \int_{t_k}^{t_{k+1}} \langle Bv_k, \mu_k\rangle \,\mathrm{d}t.
\end{aligned}$$

From (22), we have

$$\begin{aligned}\langle Bv_k, \mu_k\rangle &= \left\langle B\left(u_k^1 - u_k^2\right), \lambda_k^1 - \lambda_k^2\right\rangle \\ &= \left\langle B\left(u_k^1 - u_k^2\right), \lambda_k^1\right\rangle + \left\langle B\left(u_k^2 - u_k^1\right), \lambda_k^2\right\rangle \\ &\geq L(u_k^1) - L(u_k^2) + L(u_k^2) - L(u_k^1) \\ &= 0\end{aligned}$$

a. e. in (t_k, t_{k+1}). Therefore, it holds

$$\langle y_k(t_{k+1}), \mu_k(t_{k+1})\rangle \geq \langle y_k(t_k), \mu_k(t_k)\rangle .$$

Finally, we can conclude that

$$\begin{aligned}&\left\|TX^1 - TX^2\right\|^2 \\ =\ &\left\|X^1 - X^2\right\|^2 + 4\sum_{k=1}^{K}\langle y_k(t_k), \gamma\mu_k(t_k)\rangle - 4\sum_{k=0}^{K-1}\langle y_k(t_{k+1}), \gamma\mu_k(t_{k+1})\rangle \\ \leq\ &\left\|X^1 - X^2\right\|^2 + 4\langle y_K(t_K), \gamma\mu_K(t_K)\rangle - 4\langle y_0(t_1), \gamma\mu_0(t_1)\rangle \\ \leq\ &\left\|X^1 - X^2\right\|^2 + 4\langle y_K(t_\mathrm{f}), \gamma\mu_K(t_\mathrm{f})\rangle - 4\langle y_0(t_0), \gamma\mu_0(t_0)\rangle \\ =\ &\left\|X^1 - X^2\right\|^2 + 4\langle y_K(t_\mathrm{f}), \gamma(-Q_\mathrm{f} y_K(t_\mathrm{f}))\rangle - 4\langle y_0(t_0), \gamma Q_0 y_0(t_0)\rangle \\ \leq\ &\left\|X^1 - X^2\right\|^2\end{aligned}$$

holds because Q_f and Q_0 are positive semi-definite. □

We can now apply Schaefer's theorem to T.

Theorem 1 (Schaefer [48])**.** *Let $\mathcal{I}$ be a convex, closed, and bounded set in a uniformly convex Banach space $\mathcal{X}$ and let $T : \mathcal{I} \to \mathcal{I}$ be a non-expansive mapping with at least one fixed point. Then, for any $\varepsilon \in (0, 1)$ the mapping $T_\varepsilon = (1-\varepsilon)T + \varepsilon I$ is asymptotically regular, i.e.,*

$$\|T_\varepsilon^{\ell+1}X - T_\varepsilon^{\ell}X\|_{\mathcal{X}} \to 0 \quad \textit{for all } X \in \mathcal{I}.$$

This theorem allows us to prove convergence for Algorithm 1 applied to the linear-quadratic problem (16).

Theorem 2. *If Algorithm 1 is applied to Problem (16), the iterates $(x_k^\ell, \lambda_k^\ell)$ for $k = 0, \ldots, K$ converge in the sense*

$$\begin{aligned}x_k^\ell(t_{k+1}) - x_{k+1}^\ell(t_{k+1}) &\to 0, \\ \lambda_k^\ell(t_{k+1}) - \lambda_{k+1}^\ell(t_{k+1}) &\to 0\end{aligned}$$

for all $k = 0, \ldots, K-1$ as $\ell \to \infty$.

Proof. By setting $\mathcal{I} = \{x \in \mathbb{R}^{2K} \colon \|x\| \leq \|X^1\|\}$, Schaefer's theorem yields

$$\|TX^\ell - X^\ell\| = \frac{1}{(1-\varepsilon)}\|T_\varepsilon^\ell X^1 - T_\varepsilon^{\ell-1}X^1\| \to 0 \tag{26}$$

as $\ell \to \infty$. We rewrite the scalar product of TX^ℓ and X^ℓ as

$$\begin{aligned}
\langle TX^\ell, X^\ell \rangle &= \sum_{k=0}^{K-1} \left\langle \tilde{x}_{k+1}^\ell(t_{k+1}) + \gamma\tilde{\lambda}_{k+1}^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) + \gamma\tilde{\lambda}_k^\ell(t_{k+1}) \right\rangle \\
&\quad + \sum_{k=1}^{K} \left\langle \tilde{x}_{k-1}^\ell(t_k) - \gamma\tilde{\lambda}_{k-1}^\ell(t_k), \tilde{x}_k^\ell(t_k) - \gamma\tilde{\lambda}_k^\ell(t_k) \right\rangle \\
&= \sum_{k=0}^{K-1} \left\langle \tilde{x}_{k+1}^\ell(t_{k+1}) + \gamma\tilde{\lambda}_{k+1}^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) + \gamma\tilde{\lambda}_k^\ell(t_{k+1}) \right\rangle \qquad (27) \\
&\quad + \sum_{k=0}^{K-1} \left\langle \tilde{x}_k^\ell(t_{k+1}) - \gamma\tilde{\lambda}_k^\ell(t_{k+1}), \tilde{x}_{k+1}^\ell(t_{k+1}) - \gamma\tilde{\lambda}_{k+1}^\ell(t_{k+1}) \right\rangle \\
&= 2\sum_{k=0}^{K-1} \langle \tilde{x}_{k+1}^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \rangle + \gamma^2 \left\langle \tilde{\lambda}_{k+1}^\ell(t_{k+1}), \tilde{\lambda}_k^\ell(t_{k+1}) \right\rangle .
\end{aligned}$$

Using Lemma 3, Lemma 4, and (27) we get

$$\begin{aligned}
\left\| TX^\ell - X^\ell \right\|^2 &= \|TX^\ell\|^2 + \|X^\ell\|^2 - 2\langle TX^\ell, X^\ell \rangle \\
&= 2\left(\mathcal{E}^\ell - \langle TX^\ell, X^\ell \rangle\right) \\
&= 2\sum_{k=0}^{K-1} \Bigg(\left\|\tilde{x}_k^\ell(t_{k+1})\right\|^2 + \gamma^2 \left\|\tilde{\lambda}_k^\ell(t_{k+1})\right\|^2 \\
&\qquad + \left\|\tilde{x}_{k+1}^\ell(t_{k+1})\right\|^2 + \gamma^2 \left\|\tilde{\lambda}_{k+1}^\ell(t_{k+1})\right\|^2 \\
&\qquad - 2\langle \tilde{x}_{k+1}^\ell(t_{k+1}), \tilde{x}_k^\ell(t_{k+1}) \rangle - 2\gamma^2 \left\langle \tilde{\lambda}_{k+1}^\ell(t_{k+1}), \tilde{\lambda}_k^\ell(t_{k+1}) \right\rangle \Bigg) \\
&= 2\sum_{k=0}^{K-1} \left(\left\|\tilde{x}_k^\ell(t_{k+1}) - \tilde{x}_{k+1}^\ell(t_{k+1})\right\|^2 + \gamma^2 \left\|\tilde{\lambda}_k^\ell(t_{k+1}) - \tilde{\lambda}_{k+1}^\ell(t_{k+1})\right\|^2 \right).
\end{aligned}$$

Because of (26), it follows

$$\begin{aligned}
\tilde{x}_k^\ell(t_{k+1}) - \tilde{x}_{k+1}^\ell(t_{k+1}) &\to 0, \\
\tilde{\lambda}_k^\ell(t_{k+1}) - \tilde{\lambda}_{k+1}^\ell(t_{k+1}) &\to 0
\end{aligned}$$

for all $k = 0, \dots, K-1$ as $\ell \to \infty$. Since this holds for the errors, it is also true for the iterates, i.e.,

$$\begin{aligned}
x_k^\ell(t_{k+1}) - x_{k+1}^\ell(t_{k+1}) &\to 0, \\
\lambda_k^\ell(t_{k+1}) - \lambda_{k+1}^\ell(t_{k+1}) &\to 0
\end{aligned}$$

for all $k = 0, \dots, K-1$ as $\ell \to \infty$. □

Remark 3. *Note that Theorem 2 states convergence of Algorithm 1 with respect to the error. In order to conclude for convergence of the iterates, further assumptions are needed. To this end, we recall that if we have continuity at the boundaries of the sub-intervals, i.e.,*

$$\begin{aligned}
x_k^\ell(t_{k+1}) &= x_{k+1}^\ell(t_{k+1}), \\
\lambda_k^\ell(t_{k+1}) &= \lambda_{k+1}^\ell(t_{k+1})
\end{aligned}$$

for all $k = 0, \dots, K-1$, then the optimality conditions (17) are fulfilled as well.

Therefore we can conclude that the iterates $(x_k^\ell, \lambda_k^\ell)$ converge to (x_k, λ_k) in $L^2(t_0, t_\mathrm{f})$ for all $k = 0, \dots, K$ if (17) has a unique solution (x_k, λ_k).

Moreover, we can conclude that there is a subsequence of $(x_k^\ell, \lambda_k^\ell)$ *that converges to a solution* (x_k, λ_k) *of* (17) *if the boundary points* $x_k^\ell(t_{k+1})$, $x_{k+1}^\ell(t_{k+1})$, $\lambda_k^\ell(t_{k+1})$, $\lambda_{k+1}^\ell(t_{k+1})$ *are contained in a compact set. This is the case if* (17) *has more than one but still finitely many solutions.*

If System (17) *has infinitely many solutions, a similar conclusion can, in general, not be made. However, Algorithm 1 can still be applied in practice, where we have to impose a reasonable stopping criterion anyway such as, e.g.,*

$$\left\|x_k^\ell(t_{k+1}) - x_{k+1}^\ell(t_{k+1})\right\|_\infty \leq \delta_x, \quad \left\|\lambda_k^\ell(t_{k+1}) - \lambda_{k+1}^\ell(t_{k+1})\right\|_\infty \leq \delta_\lambda \tag{28}$$

for all $k = 0, \dots, K-1$ *with given tolerances* $\delta_x, \delta_\lambda > 0$. *This way, we can still compute an approximate solution that is arbitrarily close to a solution of* (17) *by choosing* δ_x *and* δ_λ *sufficiently small.*

5. Case Studies

In this section, we apply Algorithm 1 to test its practical performance on some exemplary problems. We implemented the algorithm in Julia 1.5.3 [3].[1] All computations were done on a machine with an Intel(R) Core(TM) i7-8550U CPU with 4 physical cores (and 8 logical cores), 1.8 GHz to 4.0 GHz, and 16 GB RAM. To solve the virtual control problems in Step 3, we apply a first-discretize-then-optimize approach using a Runge–Kutta method [20, 49]. To this end, we equidistantly partition each time domain $[t_k, t_{k+1}]$, $k = 0, \dots, K$, using $N+1 \in \mathbb{N}$ time points

$$t_{ki} := t_k + i\frac{(t_{k+1} - t_k)}{N}, \quad i = 0, \dots, N.$$

Runge–Kutta methods approximate the solution of the ODE

$$\dot{x}_k = f(x_k, u_k)$$

of the virtual control problems (12), (14), and (15) as

$$x_k(t_{k(i+1)}) = x_k(t_{ki}) + \Delta t \sum_{j=1}^{s} b_j l_j, \quad i = 0, \dots, N,$$

with the step size $\Delta t = (t_{k+1} - t_k)/N$, the intermediate steps

$$l_j = f\left(x_k(t_{ki}) + \Delta t \sum_{m=1}^{s} a_{jm} l_m,\, u_k(t_{ki})\right),$$

given coefficients $a_{jm} \in \mathbb{R}$ and $b_j \geq 0$ for $j, m \in \{1, \dots, s\}$, and the number of stages $s \in \mathbb{N}$. In what follows, we are using the classic fourth-order Runge–Kutta method with coefficients

$$a = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad b = \left(\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\right)^\top. \tag{29}$$

Moreover, we use the initial condition (12c) to solve for the virtual control

$$h_k = x_k(t_k) - \phi_{k,k-1}^{\ell-1}.$$

[1] The implementation of the algorithm is publicly available under https://github.com/m-schmidt-math-opt/miocp-time-domain-decomposition.

We can then substitute h_k in the objective function and remove the constraint (12c). The discretized virtual control problem (12) is thus given by

$$\min_{x_k,u_k} \quad \frac{1}{2\gamma}\|x_k(t_k) - \phi_{k,k-1}^{\ell-1}\|^2 + \frac{1}{2\gamma}\|x_k(t_{k+1}) - \phi_{k,k+1}^{\ell-1}\|^2 \tag{30a}$$

$$\text{s.t.} \quad x_k(t_{k(i+1)}) = x_k(t_{ki}) + \Delta t \sum_{j=1}^{s} b_j l_j, \quad i = 0, \dots, N, \tag{30b}$$

$$u_k(t_{ki}) \in U, \quad i = 0, \dots, N. \tag{30c}$$

The problems (14), (15) can be discretized analogously. Problem (30) is a finite-dimensional optimization problem, which we model using GAMS 25.1.2 [39]. The resulting instances are solved using ANTIGONE 1.1 [40]. All virtual control problems are solved in parallel using Julia's build-in multi-threading capabilities.

In our implementation, the iteration is started with the zero transmission conditions $\phi_{k,k-1}^0 = 0 \in \mathbb{R}^n$ for $k = 1, \dots, K$ and $\phi_{k,k+1}^0 = 0 \in \mathbb{R}^n$ for $k = 0, \dots, K-1$. We initialize the algorithm with all variables set to 0 and after the first iteration, we warmstart all virtual control problems with their solution from the previous iteration. As a stopping criterion for Algorithm 1, we use (28) for all $k = 0, \dots, K-1$ with tolerances $\delta_x = \delta_\lambda = 10^{-2}$. In addition, we use the time limit of 1000 s for all subproblems unless stated otherwise.

After termination of the algorithm, we fix the resulting control u and compute the corresponding state x on the entire time horizon (t_0, t_f) to obtain an accurate objective value.

5.1. **A Mixed-Integer Linear-Quadratic Problem.** We consider the mixed-integer linear-quadratic problem

$$\min \quad x_1(1)^2 + x_2(1)^2 + \int_0^1 0.005u^2 \, \mathrm{d}t \tag{31a}$$

$$\text{s.t.} \quad \dot{x}_1 = 2x_2 \quad \text{a. e. in } (0,1), \tag{31b}$$

$$\dot{x}_2 = -x_1 + x_2 - u \quad \text{a. e. in } (0,1), \tag{31c}$$

$$x_1(0) = -2, \quad x_2(0) = 1, \tag{31d}$$

$$u \in \{0, 1, 2, 3, 4\} \quad \text{a. e. in } [0,1]. \tag{31e}$$

Note that this problem is of the form (16).

We use the time-domain decomposition with $t_i = i/4$ for $i = 0, \dots, 4$ as well as parameters $\gamma = 1$, $\varepsilon = 0.5$, and $\Delta t = 1/100$.

Figure 2 shows the solution of Problem (31) using the direct discretization and solution approach from above on a single interval $[0,1]$ (left) and the one computed using the proposed algorithm with 4 time domains (right). Although the controls differ, the resulting states are almost the same. Additionally, the first part of Table 1 lists the number of iterations, the running time, and the objective value for different decompositions of the time domain. Even though the algorithm does not outperform the solution of the original problem (first row of the table), one can see that they are rather comparable w.r.t. the objective value for 2 and 4 domains. This can be explained by the fact that Problem (31) is quite simple to solve. Thus, solving the fully discretized original problem directly using ANTIGONE (on the entire time horizon) is already rather efficient. The worse objective values in the case of 8 and 16 domains can be attributed to the errors we make at the interfaces of the domains because of the tolerances δ_x and δ_λ. To show this effect we re-ran the cases of 2, 4, 8, and 16 domains with the smaller tolerances $\delta_x = \delta_\lambda = 5 \times 10^{-3}$ and $\delta_x = \delta_\lambda = 10^{-3}$. The corresponding results are shown in the second and third part of Table 1. While these smaller tolerances obviously increase the running time of the algorithm, they

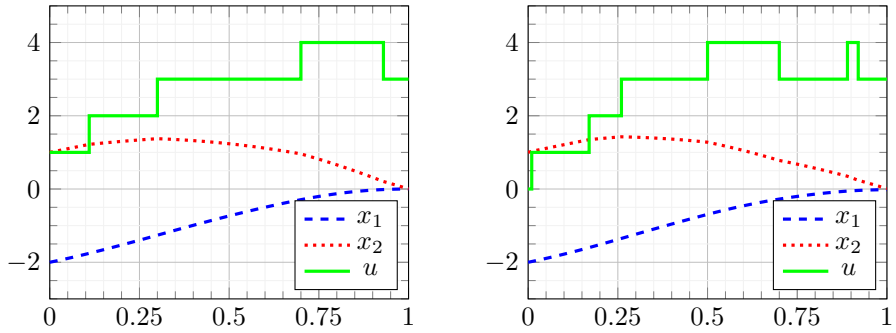

FIGURE 2. Solutions of Problem (31) solved with (right) and without (left) Algorithm 1.

have a positive impact on the objective value. This impact increases if more domains are used. Since (31) is a mixed-integer linear-quadratic problem, it is also possible to solve it with the solver Gurobi 9.1.2. However, our preliminary numerical results showed that for almost all cases, ANTIGONE performs better w.r.t. the running time. For comparison, we include these results in the fourth part of Table 1.

To make sure that our discretization step size Δt is small enough, we used the resulting controls from the first part of Table 1 to compute the corresponding state $\hat{x}$ with the halved step size $\Delta t/2$. The subsequent errors $\max_{i=0,\dots,N+1} \|x(t_i) - \hat{x}(t_i)\|_\infty$ are all smaller than 1.4×10^{-9} which is significantly smaller than the tolerances δ_x and δ_λ at the interfaces.

Next, we study the impact of the parameter γ. It functions as a weighing factor of the state x compared to the adjoint λ in Step 4 of the algorithm. It is thus an important parameter regarding the errors that are made w.r.t. the continuity of the states and adjoints at the interfaces of the time domains. Figure 3 displays these errors for $\gamma \in \{0.2, 1, 5\}$ and the case of 4 domains. One can clearly see that the smallest value $\gamma = 0.2$ leads to larger errors for the adjoint λ compared to the errors for the state x. For the larger value $\gamma = 5$ it is the other way around. For $\gamma = 1$, both errors have roughly the same size. Since in our case, the tolerances δ_x and δ_λ are the same, it takes less iterations for the algorithm to reach the stopping criterion (28) for $\gamma = 1$ than for the other values. More precisely, we need 35 iterations instead of 45 for $\gamma = 0.2$ or 44 for $\gamma = 5$. Figure 3 also shows the typical convergence behavior of Algorithm 1, i.e., the errors are reduced quickly during the first iterations but only decrease slowly when they are already closer to zero. This suggest that the method has an asymptotically sublinear convergence rate. However, estimates on these rates are beyond the scope of this paper.

Lastly for this example, we investigate the role of the parameter ε, which weighs the boundary data of the domain k against the boundary data of its neighboring domain $k-1$ or $k+1$. Table 2 shows the effect of different parameters ε on the number of iterations, the running time, and the objective value of Problem (31) when solved using 4 time domains. One can see that ε does not have a significant impact on the objective value but on the number of iterations and, thus, the running time. We see that we obtain a roughly monotonically increasing relation between ε and the running times of the algorithm. For this example, the fastest algorithm is obtained with $\varepsilon = 0.1$. We also additional sampling points for ε between 0 and 0.2, which confirm that $\varepsilon = 0.1$ is the best choice for this problem.

5.2. **Mixed-Integer Nonlinear Problems.** The convergence theory from Section 4 does not apply to problems with nonlinearities in the differential equation.

TABLE 1. Results for Problem (31) for different numbers of time domains (first column). The overall time ("Time") includes the time spent in the algorithm ("Alg. time") and the time to a posteriori compute the objective value for fixed controls.

Domains	Iterations	Alg. time (s)	Time (s)	Objective value
ANTIGONE, $\delta_x = \delta_\lambda = 10^{-2}$				
1	1	0.337	0.337	0.043909
2	22	3.065	3.107	0.045849
4	35	3.063	3.112	0.044415
8	70	9.203	9.250	0.050030
16	129	26.389	26.437	0.049699
ANTIGONE, $\delta_x = \delta_\lambda = 5 \times 10^{-3}$				
2	37	7.003	7.046	0.044221
4	38	3.165	3.282	0.044415
8	85	11.501	11.548	0.045379
16	166	35.047	35.093	0.045608
ANTIGONE, $\delta_x = \delta_\lambda = 10^{-3}$				
2	57	12.119	12.160	0.044289
4	82	6.898	7.000	0.044259
8	157	21.255	21.307	0.044999
16	258	54.724	54.774	0.044042
Gurobi, $\delta_x = \delta_\lambda = 10^{-2}$				
1	1	0.263	0.263	0.043909
2	11	1.866	2.018	0.044153
4	37	9.838	10.010	0.044478
8	68	15.590	15.793	0.045625
16	126	50.760	50.968	0.047377

TABLE 2. Results for Problem (31) for different parameters ε.

ε	Iterations	Alg. time (s)	Time (s)	Objective value
0	20	3.157	3.238	0.045483
0.025	56	5.753	5.801	0.045185
0.05	50	4.956	5.039	0.044774
0.075	47	4.454	4.507	0.045496
0.1	23	1.926	2.000	0.044690
0.125	25	2.488	2.535	0.044758
0.15	23	2.481	2.526	0.045772
0.175	35	3.283	3.332	0.045013
0.2	26	2.190	2.238	0.045415
0.3	27	2.982	3.027	0.046206
0.4	34	3.088	3.138	0.044874
0.5	35	2.964	3.009	0.044415
0.6	41	3.663	3.746	0.046568
0.7	65	6.622	6.670	0.045552
0.8	85	7.673	7.725	0.044602
0.9	170	16.075	16.126	0.046249

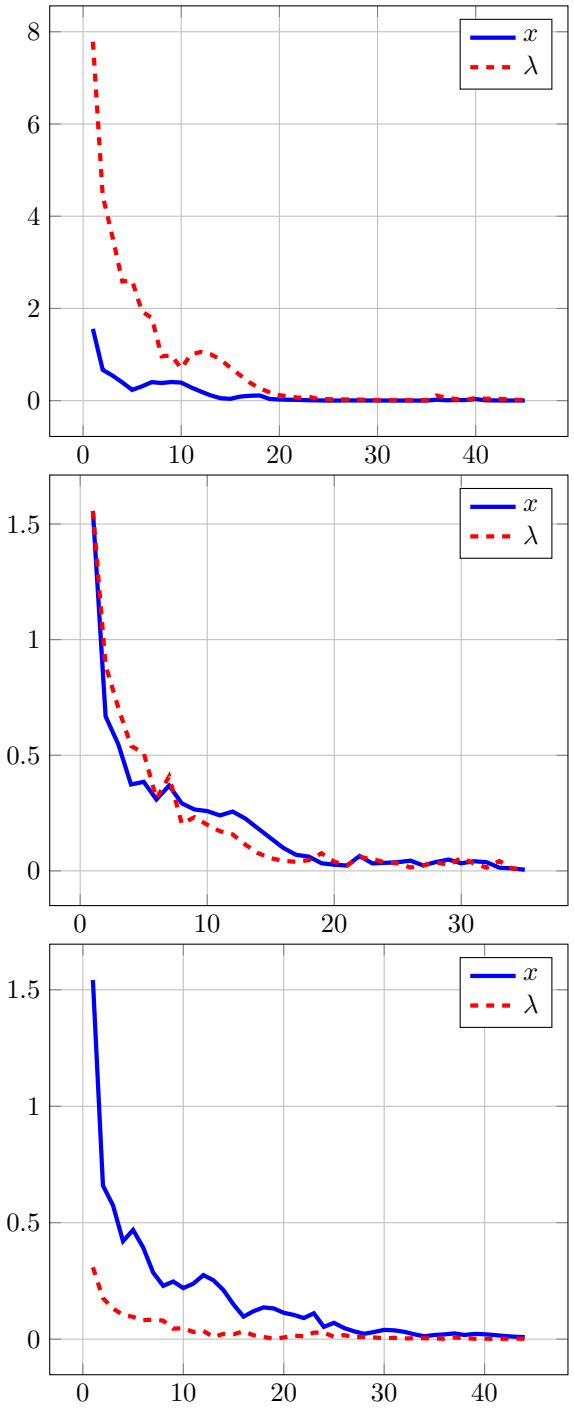

FIGURE 3. The maximum norm of all errors w.r.t. the state x ($\max_{k=0,\dots,K-1}\|x_k^\ell(t_{k+1}) - x_{k+1}^\ell(t_{k+1})\|_\infty$) and the adjoint λ ($\max_{k=0,\dots,K-1}\|\lambda_k^\ell(t_{k+1}) - \lambda_{k+1}^\ell(t_{k+1})\|_\infty$) for each iteration of Problem (31) for different values of γ. Top: $\gamma = 0.2$. Middle: $\gamma = 1$. Bottom: $\gamma = 5$.

TABLE 3. Results for Problem (32) for different numbers of domains.

Domains	Iterations	Alg. time (s)	Time (s)	Objective value
4	117	56.112	56.197	0.001054
8	483	178.834	178.886	0.000139
16	1665	650.574	650.628	0.000051

Nevertheless, we can conclude from considering Algorithm 1 as the fixed-point iteration (23) that upon convergence, the error in the transmission conditions vanishes and a candidate for a Pontryagin minimum is found. We demonstrate this for two nonlinear examples.

As a first nonlinear example we consider Fuller's initial value problem from the benchmark library mintOC of mixed-integer optimal control problems (see [45] and https://mintoc.de):

$$\begin{aligned}
\min \quad & (x_1(1) - 0.01)^2 + x_2(1)^2 + x_3(1) && \text{(32a)}\\
\text{s.t.} \quad & \dot{x}_1 = x_2 \quad \text{a. e. in } (0,1), && \text{(32b)}\\
& \dot{x}_2 = 1 - 2u \quad \text{a. e. in } (0,1), && \text{(32c)}\\
& \dot{x}_3 = x_1^2 \quad \text{a. e. in } (0,1), && \text{(32d)}\\
& x_1(0) = 0.01, \quad x_2(0) = 0, \quad x_3(0) = 0, && \text{(32e)}\\
& u \in \{0,1\} \quad \text{a. e. in } [0,1]. && \text{(32f)}
\end{aligned}$$

Again, we choose the parameters $\gamma = 1$ and $\varepsilon = 0.5$ for our computations. For this example the step size for the discretization is $\Delta t = 1/50$. We solve Problem (32) with Algorithm 1 using 1, 2, 4, 8, and 16 domains. For a single domain, i.e., without using the presented time-domain decomposition approach, ANTIGONE reaches the time limit and cannot solve the problem. For two domains, our algorithm also reaches the time limit in the first iteration. The remaining results are shown in Table 3. For 4 domains, the algorithm finds a solution in 56 s. Figure 4 shows the solution for 4 domains. For more domains, the number of iterations and the running time increases. The objective values for the last three cases are all close to zero. This shows the potential of Algorithm 1 when applied to nonlinear problems since it can outperform applying a global MINLP solver to the discretized problem directly, i.e., on the entire time horizon. However, let us comment on that this comparison needs to be interpreted carefully since the goal of ANTIGONE as a global MINLP solver is different to the one of our method that aims to compute a Pontryagin minimum of the problem. Thus, we additionally applied ANTIGONE with a time limit of 56 s, which is the solution time of our method, and compared the feasible point found by ANTIGONE within this time limit with the point that our method computed. For the considered instance, ANTIGONE found a feasible point of comparable quality in 26 s. Note, however, that this point comes with no quality guarantee whereas we can guarantee that we computed a Pontryagin minimum.

Again, we used the resulting controls from Table 3 to compute the corresponding state $\hat{x}$ with the halved step size $\Delta t/2$. The errors $\max_{i=0,\dots,N+1} \|x(t_i) - \hat{x}(t_i)\|_\infty$ are all smaller than 1.9×10^{-9} which is, again, significantly smaller than the tolerances δ_x and δ_λ at the interfaces.

As for the example in the linear-quadratic case we again test, which values of ε lead to the best results. Table 4 contains the results for the case of 4 domains. We skip the case of $\varepsilon = 0$ since we reached the time limit for this case. The algorithm also takes rather long for $\varepsilon = 0.1$. For ε between 0.2 and 0.7, one achieves the

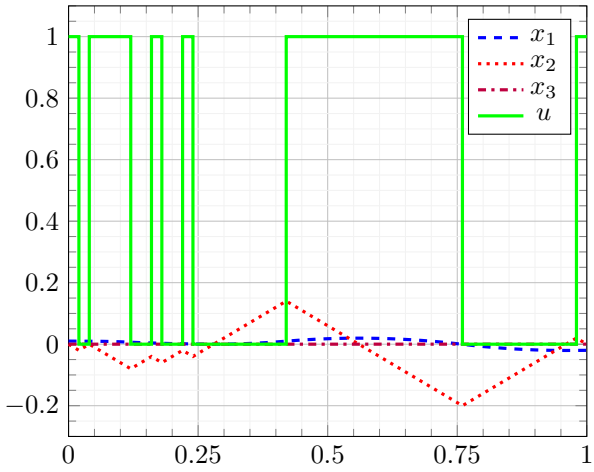

FIGURE 4. Solutions of problem (32) solved with Algorithm 1 using 4 domains.

shortest running times. Again, if the value of ε is too large, the number of iterations and the running time increase considerably.

TABLE 4. Results for problem (32) for different parameters ε.

ε	Iterations	Alg. time (s)	Time (s)	Objective value
0.1	321	588.903	588.984	0.000761
0.2	181	91.456	91.503	0.000383
0.3	139	108.863	108.912	0.000092
0.4	122	67.793	67.873	0.000243
0.5	117	56.112	56.197	0.001054
0.6	123	64.458	64.504	0.000267
0.7	147	74.871	74.925	0.001054
0.8	193	145.817	145.867	0.001824
0.9	341	186.941	186.990	0.001824

As a second example, we consider a variation of Problem (32), namely Fuller's initial value multimode problem from mintOC (see again [45] and https://mintoc.de):

$$\begin{aligned}
\min \quad & (x_1(1) - 0.01)^2 + x_2(1)^2 + x_3(1) && \text{(33a)}\\
\text{s.t.} \quad & \dot{x}_1 = x_2 \quad \text{a. e. in } (0,1), && \text{(33b)}\\
& \dot{x}_2 = 1 - 2u_1 - 0.5u_2 - 3u_3 \quad \text{a. e. in } (0,1), && \text{(33c)}\\
& \dot{x}_3 = x_1^2 \quad \text{a. e. in } (0,1), && \text{(33d)}\\
& x_1(0) = 0.01, \quad x_2(0) = 0, \quad , x_3(0) = 0, && \text{(33e)}\\
& 1 = \sum_{i=1}^{4} u_i \quad \text{a. e. in } [0,1], && \text{(33f)}\\
& u \in \{0,1\}^4 \quad \text{a. e. in } [0,1]. && \text{(33g)}
\end{aligned}$$

Here, we have three more binary controls and an SOS-1 constraint w.r.t. the binary controls. This makes the problem harder to solve, which is why we increased the maximal running time for this example to 2000 s. Note that u_4 does not occur in

TABLE 5. Results for Problem (33) for different numbers of domains.

Domains	Iterations	Alg. time (s)	Time (s)	Objective value
2	33	1580.499	1580.587	0.002083
4	117	1580.026	1580.081	0.000027
8	483	208.538	208.641	0.000018

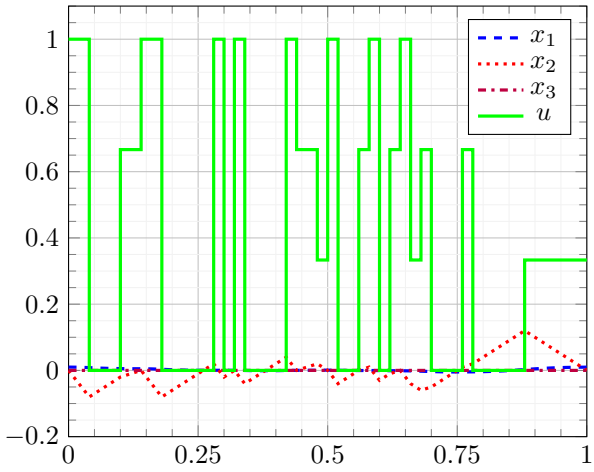

FIGURE 5. Solutions of Problem (33) solved with Algorithm 1 using 8 domains. Because of the SOS-1 condition (33f) exactly one u_i ($i = 1, 2, 3, 4$) is 1 at the same time, thus, we use the aggregated control $u = u_1/3 + 2u_2/3 + u_3 + 0u_4$ to simplify the plot.

the ODE system (33b)–(33d). Therefore, $u_4(t) = 1$ corresponds to not controlling the system at time t.

The parameters $\gamma = 1$, $\varepsilon = 0.5$, and $\Delta t = 1/50$ stay unchanged and we, again, solve the problem using 1, 2, 4, 8, and 16 domains. The results are displayed in Table 5 except for a single domain and the case of 16 domains. For a single domain, we again reach the time limit—which is also the case for our algorithm applied to 16 domains, where the time limit is reached in iteration 737. For this example, we achieved the fastest running time (of 209 s) for 8 domains. One can see that the number of iterations for 2 and 4 domains is lower than for 8 domains but the running times are higher. This is the case because, in a few iterations, the solution time of a virtual control problem is much higher than usual. Thus, we can see the trade-off between (i) a usually higher coordination effort of the method to obtain continuity at the interfaces of the time domains for a larger number of domains and (ii) a usually larger running time for less domains since the virtual control problems then are larger. The solution is shown in Figure 5. We again compared the solution that we obtain with the best feasible point computed by ANTIGONE within the time limit given by the best running time of our method, i.e., within 209 s. In this case, the feasible point found by ANTIGONE is slightly worse than the one we computed. The first feasible point found by ANTIGONE that is better than our solution needs 899 s to be computed by ANTIGONE.

For this example, we also used the resulting controls from Table 5 to compute the corresponding state $\hat{x}$ with the halved step size $\Delta t/2$. This time, the errors $\max_{i=0,\dots,N+1} \|x(t_i) - \hat{x}(t_i)\|_\infty$ are all smaller than 3.1×10^{-9} which is, again, significantly smaller than the tolerances δ_x and δ_λ at the interfaces.

An example for which convergence of our implementation of Algorithm 1 was not observed is a reformulation of the highly nonlinear problem "F-8 aircraft" from the mintOC library (see [45] and https://mintoc.de) to a fixed-time horizon problem, where either the time limit of 2000 s was reached or at some point one of the virtual control problems reported infeasibility. This happened for all combinations of parameters $\varepsilon \in \{0, 0.1, \dots, 0.9\}$, $\Delta t \in \{1/50, 1/100\}$, $\delta_x, \delta_\lambda \in \{10^{-1}, 10^{-2}\}$ and all numbers of domains in $\{1, 2, 4, 8, 16\}$ and could be caused by the presence of multiple local minima. See [15] for a successful application of a global maximum principle to the aircraft problem with numerical results.

Remark 4. *Another instructive exemplary problem is*

$$\begin{aligned}
\min \quad & T && \text{(34a)}\\
\text{s.t.} \quad & \ddot{x} = u \quad \text{a. e. in } (0,T), && \text{(34b)}\\
& x(0) = -7, \quad \dot{x}(0) = 2, && \text{(34c)}\\
& x(T) = 0, \quad \dot{x}(T) = 0, && \text{(34d)}\\
& u \in \{-1, 0, 1\} \quad \text{a. e. in } [0,T]. && \text{(34e)}
\end{aligned}$$

The state $x(t) \in \mathbb{R}$ represents the position of a car with bounded acceleration u. The car starts at -7 and has an initial speed of 2. The goal is to park the car in the origin in the shortest possible time T. The optimal bang-bang solution to this problem is to accelerate ($u(t) = 1$) until $t^* = 1$ and to break ($u(t) = -1$) after this until $T = 4$. To fit this in our setting we use a time transformation to reformulate the problem as

$$\begin{aligned}
\min \quad & x_3^2 && \text{(35a)}\\
\text{s.t.} \quad & \dot{x}_1 = x_2 x_3 \quad \text{a. e. in } (0,1), && \text{(35b)}\\
& \dot{x}_2 = u x_3 \quad \text{a. e. in } (0,1), && \text{(35c)}\\
& x_1(0) = -7, \quad x_2(0) = 2, && \text{(35d)}\\
& x_1(1) = 0, \quad x_2(1) = 0, && \text{(35e)}\\
& u \in \{-1, 0, 1\} \quad \text{a. e. in } [0,1], && \text{(35f)}
\end{aligned}$$

where x_1 is the position, x_2 the speed, and x_3 the time T. This is a nonlinear problem with initial and terminal conditions. The optimal solution stays unchanged but the switching point between acceleration and breaking is now $t^* = 1/4$. If we try to solve this problem with $\Delta t = 1/100$ using ANTIGONE, no feasible point is found. For 2 domains of equal length, the algorithm reaches the time limit of 1000 s in iteration 19. However, for 4 domains of equal length and the 2 domains $[0, 1/4], [1/4, 1]$ the global optimal solution is found in 60.886 s and 182.550 s, respectively. This can be explained by both decompositions containing the switching point $t^* = 1/4$ in an interface between two subdomains, which decreases the difficulty of the virtual control problems. Therefore, it could be an interesting topic for future research to compute the switching points of a bang-bang solution before applying Algorithm 1. This could be combined with considering constant control decisions for each subdomain, which could further decrease the difficulty of solving the virtual control problems.

6. Conclusion

Our results show that linear-quadratic mixed-integer optimal control problems can be solved iteratively by computing solutions to suitably chosen virtual mixed-integer optimal control problems on smaller time-horizons. In order to reach a certain accuracy of the solution, these subproblems then require less discretization variables and they can be solved in parallel. Both aspects can provide significant computational advantages compared to direct discretization of the original problem

on the entire time horizon. We exemplarily demonstrate this with our numerical results, where we also discuss the choice of additional parameters of the proposed algorithm. Our computational experiments show that this advantage also pays off for the case of nonlinear problems if the iterative procedure converges. However, a more detailed numerical study that might also include a comparison with other methods are out of scope of this paper and part of our future work. In particular, the iterative method may even yield solutions where full direct discretization fails. Guarantees for nonlinear problems, however, requires further algorithmic and theoretical investigations.

The presented study is primarily motivated by the need to incorporate integer restrictions on the control values into optimal control problems as to model logical constraints. However, the findings are also interesting for classic optimal control problems with other nonconvex control constraints, because the proposed iterative procedure provides an alternative to shooting-type methods as a solution approach with its well-known limitations such as proper initialization or achieving global optimality in the Hamiltonian maximization condition.

Future work will therefore concern extensions of this approach to more general problem classes including further nonlinearities, state constraints, and also partial differential equations. Moreover, the general type of methods discussed in this paper is known to have slow convergence rates although approximate solutions of acceptable quality might be obtained rather quickly. It is thus another future research topic to consider possible crossover-approaches in which an acceptable solution found with the proposed method is handed over to another method that has favorable local convergence properties. Additionally, investigations on the convergence rates and on accelerating mechanisms would be of great interest. Finally, the method presented in this paper could be combined with techniques that try to compute the time points at which the optimal control switches, see, e.g., [9, 50], and to use this information to set up the time blocks of the domain decomposition.

Acknowledgments

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under projects A03, A05, B08, and C08 of the Sonderforschungsbereich/Transregio 154 “Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks” (project ID: 239904186) and Germany’s Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689). The authors are also grateful to Günter Leugering and Alexander Martin for many fruitful discussions on the topic of this paper.

References

[1] A. Arutyunov and D. Karamzin. “A survey on regularity conditions for state-constrained optimal control problems and the non-degenerate maximum principle.” In: *Journal of Optimization Theory and Applications* 184.3 (2020), pp. 697–723. DOI: 10.1007/s10957-019-01623-7.

[2] A. V. Arutyunov and R. B. Vinter. “A simple ‘finite approximations’ proofs of the Pontryagin maximum principle under reduced differentiability hypotheses.” In: *Set-Valued Analysis* 12.1-2 (2004), pp. 5–24. DOI: 10.1023/B:SVAN.0000023406.16145.a8.

[3] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. “Julia: A Fresh Approach to Numerical Computing.” In: *SIAM Review* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671.

[4] H. Bock and R. Longman. "Optimal control of velocity profiles for minimization of energy consumption in the New York subway system." In: *Proceedings of the Second IFAC Workshop on Control Applications of Nonlinear Programming and Optimization, International Federation of Automatic Control*. 1980, pp. 34–43.

[5] J. F. Bonnans. "The shooting approach to optimal control problems." In: *IFAC Proceedings Volumes* 46.11 (2013), pp. 281–292. DOI: 10.3182/20130703-3-FR-4038.00158.

[6] M. Buss, M. Glocker, M. Hardt, O. Von Stryk, R. Bulirsch, and G. Schmidt. "Nonlinear hybrid dynamical systems: modeling, optimal control, and applications." In: *Modelling, Analysis, and Design of Hybrid Systems*. Springer, 2002, pp. 311–335. DOI: 10.1007/3-540-45426-8_18.

[7] A. De Marchi. "On the mixed-integer linear-quadratic optimal control with switching cost." In: *IEEE Control Systems Letters* 3.4 (2019), pp. 990–995. DOI: 10.1109/LCSYS.2019.2920425.

[8] A. V. Dmitruk and A. M. Kaganovich. "The hybrid maximum principle is a consequence of Pontryagin maximum principle." In: *Systems & Control Letters* 57.11 (2008), pp. 964–970. DOI: 10.1016/j.sysconle.2008.05.006.

[9] M. Egerstedt, Y. Wardi, and H. Axelsson. "Transition-time optimization for switched-mode dynamical systems." In: *IEEE Transactions on Automatic Control* 51.1 (2006), pp. 110–115.

[10] T. Faulwasser and A. Murray. "Turnpike properties in discrete-time mixed-integer optimal control." In: *IEEE Control Systems Letters* 4.3 (2020), pp. 704–709. DOI: 10.1109/LCSYS.2020.2988943.

[11] Y. Ge, S. Li, Y. Shi, and L. Han. "An adaptive wavelet method for solving mixed-integer dynamic optimization problems with discontinuous controls and application to alkali-surfactant-polymer flooding." In: *Engineering Optimization* 51.6 (2019), pp. 1028–1048. DOI: 10.1080/0305215X.2018.1508573.

[12] B. Geißler, A. Morsi, L. Schewe, and M. Schmidt. "Penalty Alternating Direction Methods for Mixed-Integer Optimization: A New View on Feasibility Pumps." In: *SIAM Journal on Optimization* 27.3 (2017), pp. 1611–1636. DOI: 10.1137/16M1069687.

[13] M. Gerdts. "A variable time transformation method for mixed-integer optimal control problems." In: *Optimal Control Applications & Methods* 27.3 (2006), pp. 169–182. DOI: 10.1002/oca.778.

[14] M. Gerdts. "Solving mixed-integer optimal control problems by branch&bound: a case study from automobile test-driving with gear shift." In: *Optimal Control Applications & Methods* 26.1 (2005), pp. 1–18. DOI: 10.1002/oca.751.

[15] M. Gerdts and S. Sager. "Mixed-Integer DAE Optimal Control Problems: Necessary Conditions and Bounds." In: *Control and Optimization with Differential-Algebraic Constraints*. Chap. 9, pp. 189–212. DOI: 10.1137/9781611972252.ch9.

[16] S. Göttlich, F. M. Hante, A. Potschka, and L. Schewe. "Penalty alternating direction methods for mixed-integer optimal control with combinatorial constraints." In: *Math. Program.* 188.2, Ser. B (2021), pp. 599–619. DOI: 10.1007/s10107-021-01656-9.

[17] M. Gugat and F. M. Hante. "Lipschitz continuity of the value function in mixed-integer optimal control problems." In: *Mathematics of Control, Signals, and Systems* 29.1 (2017). DOI: 10.1007/s00498-016-0183-4.

[18] M. Gugat, G. Leugering, A. Martin, M. Schmidt, M. Sirvent, and D. Wintergerst. "MIP-based instantaneous control of mixed-integer PDE-constrained gas

transport problems." In: *Computational Optimization and Applications* 70.1 (2018), pp. 267–294. DOI: 10.1007/s10589-017-9970-1.

[19] M. Gugat, G. Leugering, A. Martin, M. Schmidt, M. Sirvent, and D. Wintergerst. "Towards simulation based mixed-integer optimization with differential equations." In: *Networks* 72.1 (2018), pp. 60–83. DOI: 10.1002/net.21812.

[20] F. M. Hante and M. Schmidt. "Convergence of finite-dimensional approximations for mixed-integer optimization with differential equations." In: *Control and Cybernetics* 48.2 (2019), pp. 209–230.

[21] F. M. Hante, G. Leugering, A. Martin, L. Schewe, and M. Schmidt. "Challenges in Optimal Control Problems for Gas and Fluid Flow in Networks of Pipes and Canals: From Modeling to Industrial Applications." In: *Industrial Mathematics and Complex Systems: Emerging Mathematical Models, Methods and Algorithms*. Ed. by P. Manchanda, R. Lozi, and A. H. Siddiqi. Industrial and Applied Mathematics. 2017, pp. 77–122. DOI: 10.1007/978-981-10-3758-0_5.

[22] C. R. Hargraves and S. W. Paris. "Direct trajectory optimization using nonlinear programming and collocation." In: *Journal of guidance, control, and dynamics* 10.4 (1987), pp. 338–342. DOI: 10.2514/3.20223.

[23] M. Heinkenschloss. "A time-domain decomposition iterative method for the solution of distributed linear quadratic optimal control problems." In: *Journal of Computational and Applied Mathematics* 173.1 (2005), pp. 169–198. DOI: 10.1016/j.cam.2004.03.005.

[24] M. Heinkenschloss. "Time–domain decomposition iterative methods for the solution of distributed linear quadratic optimal control problems." In: *Journal of Computational and Applied Mathematics* 173 (2000), pp. 169–198. DOI: 10.1016/j.cam.2004.03.005.

[25] C. Jäkle and S. Volkwein. "POD-based mixed-integer optimal control of evolution systems." In: *Advances in dynamics, optimization and computation*. Vol. 304. Stud. Syst. Decis. Control. Springer, Cham, 2020, pp. 238–264. DOI: 10.1007/978-3-030-51264-4_10.

[26] C. Kirches, F. Lenders, and P. Manns. "Approximation properties and tight bounds for constrained mixed-integer optimal control." In: *SIAM Journal on Control and Optimization* 58.3 (2020), pp. 1371–1402. DOI: 10.1137/18M1182917.

[27] R. Krug, G. Leugering, A. Martin, M. Schmidt, and D. Weninger. "Time-Domain Decomposition for Optimal Control Problems Governed by Semilinear Hyperbolic Systems." In: *SIAM Journal on Control and Optimization* 59.6 (2021), pp. 4339–4372. DOI: 10.1137/20M138329X.

[28] R. Krug, G. Leugering, A. Martin, M. Schmidt, and D. Weninger. "Time-Domain Decomposition for Optimal Control Problems Governed by Semilinear Hyperbolic Systems with Mixed Two-Point Boundary Conditions." In: *Control and Cybernetics* 50.4 (2021), pp. 427–455. DOI: 10.2478/candc-2021-0026.

[29] J. E. Lagnese and G. Leugering. *Domain decomposition methods in optimal control of partial differential equations*. Vol. 148. International Series of Numerical Mathematics. Birkhäuser Verlag, Basel, 2004, pp. xiv+443. DOI: 10.1007/978-3-0348-7885-2.

[30] J. E. Lagnese and G. Leugering. "Time domain decomposition in final value optimal control of the Maxwell system." In: *ESAIM: Control, Optimisation and Calculus of Variations* 8 (2002), pp. 775–799. DOI: 10.1051/cocv:2002042.

[31] J. E. Lagnese and G. Leugering. "Time-domain decomposition of optimal control problems for the wave equation." In: *Systems & Control Letters* 48.3-4 (2003), pp. 229–242. DOI: 10.1016/S0167-6911(02)00268-2.

[32] H. W. J. Lee, K. L. Teo, V. Rehbock, and L. S. Jennings. "Control parametrization enhancing technique for optimal discrete-valued control problems." In: *Automatica. A Journal of IFAC, the International Federation of Automatic Control* 35.8 (1999), pp. 1401–1407. DOI: 10.1016/S0005-1098(99)00050-3.

[33] D. Liberzon. *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton University Press, 2012. DOI: 10.2307/j.ctvcm4g0s.

[34] J. L. Lions. "Decomposition of energy space and virtual control for parabolic systems." In: *12th International Conference on Domain Decomposition Methods*. 2001, pp. 41–53.

[35] J. L. Lions. "Virtual and effective control for distributed systems and decomposition of everything." In: *Journal d'Analyse Mathématique* 80 (2000), pp. 257–297. DOI: 10.1007/BF02791538.

[36] J.-L. Lions, Y. Maday, and G. Turinici. "Résolution d'EDP par un schéma en temps "pararéel"." In: *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* 332.7 (2001), pp. 661–668. DOI: 10.1016/S0764-4442(00)01793-6.

[37] J.-L. Lions and O. Pironneau. "Domain decomposition methods for CAD." In: *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* 328.1 (1999), pp. 73–80. DOI: 10.1016/S0764-4442(99)80015-9.

[38] Z. Liu, S. Li, and K. Zhao. "Extended multi-interval Legendre-Gauss-Radau pseudospectral method for mixed-integer optimal control problem in engineering." In: *International Journal of Systems Science. Principles and Applications of Systems and Integration* 52.5 (2021), pp. 928–951. DOI: 10.1080/00207721.2020.1849862.

[39] B. A. McCarl. *GAMS User Guide*. Version 23.0. 2009.

[40] R. Misener and C. A. Floudas. "ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations." In: *Journal of Global Optimization* 59.2-3 (2014), pp. 503–526. DOI: 10.1007/s10898-014-0166-2.

[41] I. Nowak. *Relaxation and decomposition methods for mixed integer nonlinear programming*. Vol. 152. International Series of Numerical Mathematics. Birkhäuser Verlag, Basel, 2005, pp. xvi+213. DOI: 10.1007/3-7643-7374-1.

[42] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The mathematical theory of optimal processes*. Translated from the Russian by K. N. Trirogoff; edited by L. W. Neustadt. Interscience Publishers John Wiley & Sons, Inc. New York-London, 1962, pp. viii+360. DOI: 10.1002/zamm.19630431023.

[43] D. Preda and J. Noailles. "Mixed integer programming for a special logic constrained optimal control problem." In: *Mathematical Programming. A Publication of the Mathematical Programming Society* 103.2, Ser. B (2005), pp. 309–333. DOI: 10.1007/s10107-005-0584-5.

[44] M. Ringkamp, S. Ober-Blöbaum, and S. Leyendecker. "On the time transformation of mixed integer optimal control problems using a consistent fixed integer control function." In: *Mathematical Programming* 161.1-2 (2017), pp. 551–581. DOI: 10.1007/s10107-016-1023-5.

[45] S. Sager. "A benchmark library of mixed-integer optimal control problems." In: *Mixed integer nonlinear programming*. Springer, 2012, pp. 631–670. DOI: 10.1007/978-1-4614-1927-3_22.

[46] S. Sager, H. G. Bock, and M. Diehl. "The integer approximation error in mixed-integer optimal control." In: *Mathematical Programming* 133.1-2 (2012), pp. 1–23. DOI: 10.1007/s10107-010-0405-3.

[47] S. Sager, M. Jung, and C. Kirches. "Combinatorial integral approximation." In: *Math. Methods Oper. Res.* 73.3 (2011), pp. 363–380. DOI: 10.1007/s00186-011-0355-4.

[48] H. Schaefer. "Über die Methode sukzessiver Approximationen." In: *Jahresbericht der Deutschen Mathematiker-Vereinigung* 59 (1957), pp. 131–140.

[49] A. Schwartz and E. Polak. "Runge-Kutta Discretization of Optimal Control Problems." In: *IFAC Proceedings Volumes* 29.8 (1996). 10th IFAC Workshop on Control Applications of Optimization 1995, Haifa, Israel, 19-21 December, pp. 123–128. DOI: 10.1016/S1474-6670(17)43687-1.

[50] H. R. Sirisena. "A gradient method for computing optimal bang-bang controls." In: *International Journal of Control* 19.2 (1974), pp. 257–264. DOI: 10.1080/00207177408932627.

[51] H. Sussmann. "A maximum principle for hybrid optimal control problems." In: *Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No.99CH36304)*. Vol. 1. 1999, 425–430 vol.1. DOI: 10.1109/CDC.1999.832814.

[52] N. Tauchnitz. "Das Pontrjaginsche Maximumprinzip für eine Klasse hybrider Steuerungsprobleme mit Zustandsbeschränkung und seine Anwendung." doctoralthesis. BTU Cottbus - Senftenberg, 2010.

[53] F. Trespalacios and I. E. Grossmann. "Review of mixed-integer nonlinear optimization and generalized disjunctive programming applications in process systems engineering." In: *Advances and trends in optimization with engineering applications*. Vol. 24. MOS-SIAM Ser. Optim. SIAM, Philadelphia, PA, 2017, pp. 315–329. DOI: 10.1137/1.9781611974683.ch24.

[54] W. Zhao, T. H. Beach, and Y. Rezgui. "A systematic mixed-integer differential evolution approach for water network operational optimization." In: *Proceedings A* 474.2217 (2018), p. 20170879. DOI: 10.1098/rspa.2017.0879.

(F.M. Hante) Humboldt-Universität zu Berlin, Department of Mathematics, Unter den Linden 6, 10099 Berlin, Germany
Email address: falk.hante@hu-berlin.de

(R. Krug) Friedrich-Alexander-Universität Erlangen-Nürnberg, Department of Data Science, Cauerstr. 11, 91058 Erlangen, Germany
Email address: richard.krug@fau.de

(M. Schmidt) Trier University, Department of Mathematics, Universitätsring 15, 54296 Trier, Germany
Email address: martin.schmidt@uni-trier.de