Consistent Approximations for the Optimal Control of Constrained Switched Systems-Part 2: An Implementable Algorithm

SIAM J. Control Optim.
Vol. 51, No. 6, pp. 4484–4503

CONSISTENT APPROXIMATIONS FOR THE OPTIMAL CONTROL OF CONSTRAINED SWITCHED SYSTEMS—PART 2: AN IMPLEMENTABLE ALGORITHM*

RAMANARAYAN VASUDEVAN[†], HUMBERTO GONZALEZ[‡], RUZENA BAJCSY[§], AND S. SHANKAR SASTRY[§]

Abstract. In the first part of this two-paper series, we presented a conceptual algorithm for the optimal control of constrained switched systems and proved that this algorithm generates a sequence of points that converge to a necessary condition for optimality. However, since our algorithm requires the exact solution of a differential equation, the numerical implementation of this algorithm is impractical. In this paper, we address this shortcoming by constructing an implementable algorithm that discretizes the differential equation, producing a finite-dimensional nonlinear program. We prove that this implementable algorithm constructs a sequence of points that asymptotically satisfy a necessary condition for optimality for the constrained switched system optimal control problem. Four simulation experiments are included to validate the theoretical developments.

Key words. optimal control, switched systems, chattering lemma, consistent approximations

AMS subject classifications. Primary, 49M25, 90C11, 90C30, 93C30; Secondary, 49J21, 49J30, 49J52, 49N25

DOI. 10.1137/120901507

1. Introduction.

This paper, which is an extension of [14], continues our presentation on algorithms to solve the following optimal control problem:

$$\begin{aligned}(1.1)\qquad &\inf \big\{ h_0\big(x(1)\big) \mid \dot{x}(t) = f\big(t, x(t), u(t), d(t)\big),\ x(0) = x_0,\ h_j\big(x(t)\big) \le 0, \\ &\quad x(t) \in \mathbb{R}^n,\ u(t) \in U,\ d(t) \in \mathcal{Q},\ \text{for a.e. } t \in [0,1]\ \forall j \in \mathcal{J} \big\},\end{aligned}$$

where $\mathcal{Q}$ and $\mathcal{J}$ are finite sets, $U \subset \mathbb{R}^m$ is a compact, convex, nonempty set, $x_0 \in \mathbb{R}^n$, $f\colon [0,1] \times \mathbb{R}^n \times U \times \mathcal{Q} \to \mathbb{R}^n$ is a vector field that defines a switched system, $\{h_j\}_{j\in\mathcal{J}}$ is a collection of functions defining a feasible set, and $h_0\colon \mathbb{R}^n \to \mathbb{R}$ is a cost function.

In this paper, we discretize (1.1) to generate a finite-dimensional nonlinear program. We present an implementable algorithm, Algorithm 3.1, that constructs a sequence of finite-dimensional optimization problems that approximate the problem in (1.1). Our algorithm does not address the complexities that arise due to limited precision computations [8, 11]. We prove in Theorem 4.20 that the sequence of points generated by Algorithm 3.1 asymptotically satisfies the optimality condition defined in [14].

1.1. Related work.

Understanding the relationship between each of the solutions to a sequence of optimization problems and the solution to the optimization problem that they attempt to approximate is a well-studied problem. For example, the *epiconvergence* of a sequence of optimization problems, as in Definition 7.1 in [13],

*Received by the editors December 6, 2012; accepted for publication (in revised form) September 3, 2013; published electronically December 11, 2013. This material is based upon work supported by the National Science Foundation under award ECCS-0931437.

http://www.siam.org/journals/sicon/51-6/90150.html

[†]CSAIL, Massachusetts Institute of Technology, Cambridge, MA 02139 (ramv@csail.mit.edu).

[‡]ESE, Washington University in St. Louis, St. Louis, MO 63130 (hgonzale@ese.wustl.edu).

[§]EECS, University of California at Berkeley, Berkeley, CA 94720 (bajcsy@eecs.berkeley.edu, sastry@eecs.berkeley.edu).

guarantees the convergence of the sequence of global minimizers for each optimization problem to the set of global minimizers of the limiting problem (Theorem 7.33 in [13]). Unfortunately derivative-based optimization algorithms are only capable of constructing points that satisfy a necessary condition for optimality. However, if the sequence of optimization problems *consistently approximates* some limiting optimization problem, as in Definition 3.3.6 in [12], then every sequence of points that satisfy a necessary condition for optimality for each optimization problem converges to the set of points that satisfy a necessary condition for optimality for the limiting optimization problem (Theorem 3.3.11 in [12]).

1.2. Our contribution and organization.

In this paper, we devise an implementable algorithm that generates a sequence of optimization problems and corresponding points that consistently approximates the constrained switched optimal control problem defined in (3.13) in [14]. In sections 2 and 3, we formalize the finite-dimensional optimization problem we solve at each iteration of our implementable algorithm. In section 4, we prove that the sequence of points generated by our implementable algorithm asymptotically satisfies the necessary condition for optimality for the switched system optimal control problem in [14]. In section 5, we illustrate the superior performance of our approach on four examples when compared to a commercial mixed integer programming solver.

2. Notation.

First, we summarize the notation used in this paper, defined in the first part of this two-paper series [14].

$\lVert\cdot\rVert_p$	standard vector space p-norm
$\lVert\cdot\rVert_{i,p}$	matrix induced p-norm
$\lVert\cdot\rVert_{L^p}$	functional L^p-norm
$\mathcal{Q} = \{1, \ldots, q\}$	set of discrete mode indices (section 1)
$\mathcal{J} = \{1, \ldots, N_c\}$	set of constraint function indices (section 1)
Σ_r^q	q-simplex in $\mathbb{R}^q$ (equation (3.5))
$\Sigma_p^q = \{e_i\}_{i\in\mathcal{Q}}$	corners of the q-simplex in $\mathbb{R}^q$ (equation (3.6))
$U \subset \mathbb{R}^m$	compact, convex, nonempty set (section 1)
$\mathcal{D}_r$	relaxed discrete input space (section 3.2)
$\mathcal{D}_p$	pure discrete input space (section 3.2)
$\mathcal{U}$	continuous input space (section 3.2)
$(\mathcal{X}, \lVert\cdot\rVert_{\mathcal{X}})$	general optimization space and its norm (section 3.2)
$\mathcal{X}_r$	relaxed optimization space (section 3.2)
$\mathcal{X}_p$	pure optimization space (section 3.2)
d_i or $[d]_i$	ith coordinate of $d\colon [0,1] \to \mathbb{R}^q$
$f\colon [0,1] \times \mathbb{R}^n \times U \times \mathcal{Q} \to \mathbb{R}^n$	switched system (section 1)
$\mathrm{V}\colon L^2([0,1], \mathbb{R}^n) \to \mathbb{R}$	total variation operator (equation (3.3))
$h_0\colon \mathbb{R}^n \to \mathbb{R}$	terminal cost function (section 3.3)
$\{h_j\colon \mathbb{R}^n \to \mathbb{R}\}_{j\in\mathcal{J}}$	collection of constraint functions (section 3.3)
$x^{(\xi)}\colon [0,1] \to \mathbb{R}^n$	trajectory of the system for $\xi \in \mathcal{X}_r$ (equation (3.8))
$\phi_t\colon \mathcal{X}_r \to \mathbb{R}^n$	flow of the system at $t \in [0,1]$ (equation (3.9))
$J\colon \mathcal{X}_r \to \mathbb{R}$	cost function (equation (3.10))
$\Psi\colon \mathcal{X}_r \to \mathbb{R}$	constraint function (equation (3.11))
$\psi_{j,t}\colon \mathcal{X}_r \to \mathbb{R}$	component constraint functions (equation (3.12))
D	derivative operator (equation (4.2))
$\theta\colon \mathcal{X}_p \to (-\infty, 0]$	optimality function (equation (4.3))
$g\colon \mathcal{X}_p \to \mathcal{X}_r$	descent direction function (equation (4.3))

$\zeta\colon \mathcal{X}_p \times \mathcal{X}_r \to \mathbb{R}$	optimality function objective (equation (4.4))
$\mathbb{1}_A\colon [0,1] \to \{0,1\}$	indicator function of $A \subset [0,1]$, $\mathbb{1} = \mathbb{1}_{[0,1]}$ (section 4.4)
$\mathcal{F}_N\colon \mathcal{D}_r \to \mathcal{D}_r$	Haar wavelet operator (equation (4.7))
$\mathcal{P}_N\colon \mathcal{D}_r \to \mathcal{D}_p$	pulse-width modulation operator (equation (4.8))
$\rho_N\colon \mathcal{X}_r \to \mathcal{X}_p$	projection operator (equation (4.9))

Next, we summarize the notation used and defined in this paper. Note that whenever the subscript τ appears, τ is assumed to belong to $\mathcal{T}_N$ with $N \in \mathbb{N}$.

$\mathcal{T}_N$	Nth switching time space (equation (3.1))
$\mathcal{D}_{\tau,r}$	discretized relaxed discrete input space (equation (3.3))
$\mathcal{D}_{\tau,p}$	discretized pure discrete input space (equation (3.2))
$\mathcal{U}_\tau$	discretized continuous input space (equation (3.4))
$\mathcal{X}_\tau$	discretized general optimization space (equation (3.5))
$\mathcal{X}_{\tau,r}$	discretized relaxed optimization space (section 3.1)
$\mathcal{X}_{\tau,p}$	discretized pure optimization space (section 3.1)
$\mathcal{N}_\tau(\xi,\varepsilon)$	ε-ball around ξ in the τ-induced $\mathcal{X}$-norm (Definition 3.2)
$\lvert\tau\rvert$	last index of τ (section 3.1)
$z_\tau^{(\xi)}\colon [0,1] \to \mathbb{R}^n$	discretized trajectory given $\xi \in \mathcal{X}_{\tau,r}$ (equation (3.7))
$\phi_{\tau,t}\colon \mathcal{X}_{\tau,r} \to \mathbb{R}^n$	discretized flow of the system at $t \in [0,1]$ (equation (3.8))
$J_\tau\colon \mathcal{X}_{\tau,r} \to \mathbb{R}$	discretized cost function (equation (3.9))
$\Psi_\tau\colon \mathcal{X}_{\tau,r} \to \mathbb{R}$	discretized constraint function (equation (3.10))
$\psi_{\tau,j,t}\colon \mathcal{X}_{\tau,r} \to \mathbb{R}$	discretized component constraint function (equation (3.11))
P_τ	discretized relaxed optimal control problem (equation (3.12))
$\theta_\tau\colon \mathcal{X}_{\tau,p} \to (-\infty,0]$	discretized optimality function (equation (3.13))
$g_\tau\colon \mathcal{X}_{\tau,p} \to \mathcal{X}_{\tau,r}$	discretized descent direction function (equation (3.13))
$\zeta_\tau\colon \mathcal{X}_{\tau,p} \times \mathcal{X}_{\tau,r} \to \mathbb{R}$	discretized optimality function objective (equation (3.14))
$\mu_\tau\colon \mathcal{X}_{\tau,p} \to \mathbb{N}$	discretized step size function (equation (3.15))
$\sigma_N\colon \mathcal{X}_r \to \mathcal{T}_N$	induced partition function given $N \in \mathbb{N}$ (equation (3.16))
$\nu_\tau\colon \mathcal{X}_{\tau,p} \times \mathbb{N} \to \mathbb{N}$	discretized frequency modulation function (equation (3.17))
$\Gamma_\tau\colon \mathcal{X}_{\tau,p} \to \mathcal{X}_p$	implementable algorithm recursion function (equation (3.18))

3. Implementable algorithm.

In this section, we begin by describing our discretization strategy in order to define our discretized optimization spaces. Next, we construct the discretized trajectories, cost, constraints, and optimal control problems. This allows us to define a discretized optimality function and a notion of *consistent approximation* between the optimality function and its discretized counterpart. We conclude by describing our numerically implementable optimal control algorithm for constrained switched systems.

3.1. Discretized optimization space.

For each $N \in \mathbb{N}$ we define the *Nth switching time space* as

$$(3.1)\quad \mathcal{T}_N = \left\{ \{t_i\}_{i=0}^k \subset [0,1] \mid k \in \mathbb{N},\ 0 = t_0 \leq \cdots \leq t_k = 1,\ \lvert t_i - t_{i-1}\rvert \leq \frac{1}{2^N}\ \forall i \right\}.$$

$\mathcal{T}_N$ is the collection of finite partitions of $[0,1]$ whose samples have a maximum distance of 2^{-N}. Given $\tau \in \mathcal{T}_N$, we abuse notation and let $\lvert\tau\rvert$ denote the last index of τ, i.e., $\tau = \{t_k\}_{k=0}^{\lvert\tau\rvert}$. We use this convention since it simplifies our indexing in this paper. Notice that for each $N \in \mathbb{N}$, $\mathcal{T}_{N+1} \subset \mathcal{T}_N$.

Given $N \in \mathbb{N}$ and $\tau = \{t_i\}_{i=0}^{|\tau|} \in \mathcal{T}_N$, let the *discretized relaxed discrete input space* be

$$\mathcal{D}_{\tau,r} = \left\{ d \in \mathcal{D}_r \mid d = \sum_{k=0}^{|\tau|-1} \bar{d}_k \mathbb{1}_{[t_k,t_{k+1})},\ \{\bar{d}_k\}_{k=0}^{|\tau|-1} \subset \Sigma_r^q \right\}. \tag{3.2}$$

Similarly, let the *discretized pure discrete input space* be

$$\mathcal{D}_{\tau,p} = \left\{ d \in \mathcal{D}_p \mid d = \sum_{k=0}^{|\tau|-1} \bar{d}_k \mathbb{1}_{[t_k,t_{k+1})},\ \{\bar{d}_k\}_{k=0}^{|\tau|-1} \subset \Sigma_p^q \right\}. \tag{3.3}$$

Finally, let the *discretized continuous input space* be

$$\mathcal{U}_{\tau} = \left\{ u \in \mathcal{U} \mid u = \sum_{k=0}^{|\tau|-1} \bar{u}_k \mathbb{1}_{[t_k,t_{k+1})},\ \{\bar{u}_k\}_{k=0}^{|\tau|-1} \subset U \right\}. \tag{3.4}$$

Now, we can define the *discretized pure optimization space* as $\mathcal{X}_{\tau,p} = \mathcal{U}_\tau \times \mathcal{D}_{\tau,p}$ and the *discretized relaxed optimization space* as $\mathcal{X}_{\tau,r} = \mathcal{U}_\tau \times \mathcal{D}_{\tau,r}$. Recall that the general optimization space, $\mathcal{X}$, is defined as $\mathcal{X} = L^2([0,1],\mathbb{R}^m) \times L^2([0,1],\mathbb{R}^q)$ and is endowed with the norm $\|\xi\|_{\mathcal{X}} = \|u\|_{L^2} + \|d\|_{L^2}$ for each $\xi = (u,d) \in \mathcal{X}$. We define a subspace of $\mathcal{X}$, the *discretized general optimization space*:

$$\mathcal{X}_{\tau} = \left\{ (u,d) \in \mathcal{X} \mid (u,d) = \sum_{k=0}^{|\tau|-1} \big(\bar{u}_k, \bar{d}_k\big) \mathbb{1}_{[t_k,t_{k+1})},\ \big\{\big(\bar{u}_k, \bar{d}_k\big)\big\}_{k=0}^{|\tau|-1} \subset \mathbb{R}^m \times \mathbb{R}^q \right\}. \tag{3.5}$$

We endow the discretized pure and relaxed optimization spaces with the $\mathcal{X}$-norm as well.

The following lemma, which follows since every measurable function can be approximated in the L^2-norm by sums of indicator functions (Theorem 2.10 in [5]), proves that the spaces $\mathcal{X}_{\tau,r}$ and $\mathcal{X}_{\tau,p}$ can be used to build approximations of points in $\mathcal{X}_p$ and $\mathcal{X}_r$, respectively.

LEMMA 3.1. *Let $\{\tau_k\}_{k\in\mathbb{N}}$ such that $\tau_k \in \mathcal{T}_k$ for each $k \in \mathbb{N}$.*

(1) *For each $\xi \in \mathcal{X}_p$ there exists $\{\xi_k\}_{k\in\mathbb{N}}$, with $\xi_k \in \mathcal{X}_{\tau_k,p}$, such that $\|\xi_k - \xi\|_{\mathcal{X}} \to 0$ as $k \to \infty$.*

(2) *For each $\xi \in \mathcal{X}_r$ there exists $\{\xi_k\}_{k\in\mathbb{N}}$, with $\xi_k \in \mathcal{X}_{\tau_k,r}$, such that $\|\xi_k - \xi\|_{\mathcal{X}} \to 0$ as $k \to \infty$.*

3.2. Discretized trajectories, cost, constraint, and optimal control problem.

Let $N \in \mathbb{N}$, $\tau = \{t_i\}_{i=0}^{|\tau|} \in \mathcal{T}_N$, $\xi = (u,d) \in \mathcal{X}_{\tau,r}$, and $x_0 \in \mathbb{R}^n$. We define the *discretized trajectory*, $z_\tau^{(\xi)}\colon [0,1] \to \mathbb{R}^n$, by performing linear interpolation over points computed via the forward Euler integration formula, i.e., $z_\tau^{(\xi)}(0) = x_0$, for each $k \in \{0,\ldots,|\tau|-1\}$,

$$z_\tau^{(\xi)}(t_{k+1}) = z_\tau^{(\xi)}(t_k) + (t_{k+1} - t_k) f\big(t_k, z_\tau^{(\xi)}(t_k), u(t_k), d(t_k)\big) \tag{3.6}$$

and

$$z_\tau^{(\xi)}(t) = \sum_{k=0}^{|\tau|-1} \Big(z_\tau^{(\xi)}(t_k) + (t - t_k) f\big(t_k, z_\tau^{(\xi)}(t_k), u(t_k), d(t_k)\big) \Big) \mathbb{1}_{[t_k,t_{k+1})}(t). \tag{3.7}$$

For notational convenience, we suppress the dependence on τ in $z_\tau^{(\xi)}$ when it is clear in context. The *discretized flow of the system*, $\phi_{\tau,t}\colon \mathcal{X}_{\tau,r} \to \mathbb{R}^n$, for each $t \in [0,1]$ is

$$\phi_{\tau,t}(\xi) = z_\tau^{(\xi)}(t). \tag{3.8}$$

The *discretized cost function*, $J_\tau \colon \mathcal{X}_{\tau,r} \to \mathbb{R}$, is

$$J_\tau(\xi) = h_0\big(z_\tau^{(\xi)}(1)\big). \tag{3.9}$$

Similarly, the *discretized constraint function*, $\Psi_\tau \colon \mathcal{X}_{\tau,r} \to \mathbb{R}$, is

$$\Psi_\tau(\xi) = \max\big\{h_j\big(z_\tau^{(\xi)}(t_k)\big) \mid j \in \mathcal{J},\ k \in \{0,\dots,|\tau|\}\big\}. \tag{3.10}$$

For any positive integer N and $\tau \in \mathcal{T}_N$, the *discretized component constraint function*, $\psi_{\tau,j,t} \colon \mathcal{X}_{\tau,r} \to \mathbb{R}$, for each $t \in [0,1]$ and each $j \in \mathcal{J}$ is

$$\psi_{\tau,j,t}(\xi) = h_j\big(z_\tau^{(\xi)}(t)\big). \tag{3.11}$$

3.3. Local minimizers and a discretized optimality condition. Let $N \in \mathbb{N}$ and $\tau \in \mathcal{T}_N$. Proceeding as we did in section 4.1 in [14], we define the *discretized relaxed optimal control problem* as

$$\inf\left\{J_\tau(\xi) \mid \Psi_\tau(\xi) \le 0,\ \xi \in \mathcal{X}_{\tau,r}\right\}. \qquad (P_\tau) \tag{3.12}$$

The local minimizers of this problem are then defined as follows.

DEFINITION 3.2. *Let $N \in \mathbb{N}$ and $\tau \in \mathcal{T}_N$. Let an* ε-ball in the τ-induced $\mathcal{X}$-norm *around ξ be defined as $\mathcal{N}_\tau(\xi,\varepsilon) = \big\{\bar\xi \in \mathcal{X}_{\tau,r} \mid \big\|\xi - \bar\xi\big\|_{\mathcal{X}} < \varepsilon\big\}$. A point $\xi \in \mathcal{X}_{\tau,r}$ is a* local minimizer of P_τ *if $\Psi_\tau(\xi) \le 0$ and there exists $\varepsilon > 0$ such that $J_\tau(\hat\xi) \ge J_\tau(\xi)$ for each $\hat\xi \in \mathcal{N}_\tau(\xi,\varepsilon) \cap \big\{\bar\xi \in \mathcal{X}_{\tau,r} \mid \Psi_\tau(\bar\xi) \le 0\big\}$.*

Let the *discretized optimality function*, $\theta_\tau \colon \mathcal{X}_{\tau,p} \to (-\infty, 0]$, and the *discretized descent direction*, $g_\tau \colon \mathcal{X}_{\tau,p} \to \mathcal{X}_{\tau,r}$, be defined by

$$\theta_\tau(\xi) = \min_{\xi' \in \mathcal{X}_{\tau,r}} \zeta_\tau(\xi,\xi') + \mathrm{V}(\xi' - \xi), \qquad g_\tau(\xi) = \operatorname*{arg\,min}_{\xi' \in \mathcal{X}_{\tau,r}} \zeta_\tau(\xi,\xi') + \mathrm{V}(\xi' - \xi), \tag{3.13}$$

where

$$\zeta_\tau(\xi,\xi') = \begin{cases} \max\left\{ \displaystyle\max_{(j,k) \in \mathcal{J} \times \{0,\dots,|\tau|\}} \mathrm{D}\psi_{\tau,j,t_k}(\xi;\xi'-\xi) + \gamma\Psi_\tau(\xi), \right. \\ \qquad \left. \mathrm{D}J_\tau(\xi;\xi'-\xi) \right\} + \|\xi'-\xi\|_{\mathcal{X}} & \text{if } \Psi_\tau(\xi) \le 0, \\ \max\left\{ \displaystyle\max_{(j,k) \in \mathcal{J} \times \{0,\dots,|\tau|\}} \mathrm{D}\psi_{\tau,j,t_k}(\xi;\xi'-\xi), \right. \\ \qquad \left. \mathrm{D}J_\tau(\xi;\xi'-\xi) - \Psi_\tau(\xi) \right\} + \|\xi'-\xi\|_{\mathcal{X}} & \text{if } \Psi_\tau(\xi) > 0, \end{cases} \tag{3.14}$$

where $\gamma > 0$ is the same design parameter used for θ in [14]. Recall that any element of the discretized relaxed optimization space by definition can be naturally identified with an element of the relaxed optimization space. The total variation is applied to each component of the discretized relaxed optimization space once this natural identification has taken place. Note that $\theta_\tau(\xi) \le 0$ for each $\xi \in \mathcal{X}_{\tau,p}$ as shown in Theorem 4.10. Also note that the directional derivatives in (3.14) consider directions $\xi' - \xi$ in order to ensure that all constructed descent directions belong to $\mathcal{X}_{\tau,r}$. As we did in the infinite-dimensional case, we prove in Theorem 4.10 that if $\theta_\tau(\xi) < 0$ for some $\xi \in \mathcal{X}_{\tau,p}$, then ξ is not a local minimizer of P_τ, or in other words, ξ *satisfies the discretized optimality condition* if $\theta_\tau(\xi) = 0$. Note that the total variation of terms in $\mathcal{X}_\tau$ can be computed using a finite sum by (3.3) in [14].

Theorem 5.10 in [14], which is an extension of the chattering lemma, describes a constructive method, with a known error bound, to find pure discrete-valued approximations of points in $\mathcal{X}_r$ in the weak topology. The result of that theorem is fundamental in this paper, since we use it not only to construct suitable approximations but also to determine when to increase the discretization precision by choosing a finer switching time τ. The following definition and theorem lay the foundation for the technique we use to prove that our implementable algorithm converges to the same points as Algorithm 4.1 in [14] does by recursive application.

DEFINITION 3.3 (Definition 3.3.6 in [12]). *Let $\{\tau_i\}_{i\in\mathbb{N}}$ be such that $\tau_i \in \mathcal{T}_i$ for each $i \in \mathbb{N}$. We say that $\{P_{\tau_i}\}_{i\in\mathbb{N}}$ is a* consistent approximation *of P_p if given $\{\xi_i\}_{i\in\mathbb{N}}$ such that $\xi_i \in \mathcal{X}_{\tau_i,p}$ for each $i \in \mathbb{N}$, then* $\liminf_{i\to\infty} \theta_{\tau_i}(\xi_i) \le \liminf_{i\to\infty} \theta(\xi_i)$.

THEOREM 3.4. *Let $\{\tau_i\}_{i\in\mathbb{N}}$ and $\{\xi_i\}_{i\in\mathbb{N}}$ be such that $\tau_i \in \mathcal{T}_i$ and $\xi_i \in \mathcal{X}_{\tau_i,p}$ for each $i \in \mathbb{N}$, and suppose that $\{P_{\tau_i}\}_{i\in\mathbb{N}}$ is a consistent approximation of P_p. If* $\lim_{i\to\infty} \theta_{\tau_i}(\xi_i) = 0$, *then* $\lim_{i\to\infty} \theta(\xi_i) = 0$.

We omit the proof of Theorem 3.4 since it follows trivially from Definition 3.3. Note that Theorem 3.4 implies that our discretized optimality condition is asymptotically equivalent to the optimality condition in [14]. In Theorem 4.12 we show that $\{P_\tau\}_{i\in\mathbb{N}}$ is indeed a consistent approximation of P_p.

3.4. Choosing a discretized step size and projecting the discretized relaxed discrete input.

We employ a line search equivalent to the one formulated in (4.5) in [14], which is inspired by the traditional Armijo algorithm [1], in order to choose a step size. Letting $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, and $\alpha, \beta \in (0,1)$, a step size for a point ξ is chosen as $\beta^{\mu_\tau(\xi)}$, where the *discretized step size function*, $\mu_\tau \colon \mathcal{X}_{\tau,p} \to \mathbb{N}$, is defined as

(3.15)
$$\mu_\tau(\xi) = \begin{cases} \min\big\{k \in \mathbb{N} \mid J_\tau\big(\xi + \beta^k(g_\tau(\xi) - \xi)\big) - J_\tau(\xi) \le \alpha\beta^k\theta_\tau(\xi), \\ \qquad\qquad \Psi_\tau\big(\xi + \beta^k(g_\tau(\xi) - \xi)\big) \le \alpha\beta^k\theta_\tau(\xi)\big\} & \text{if } \Psi_\tau(\xi) \le 0, \\ \min\big\{k \in \mathbb{N} \mid \Psi_\tau\big(\xi + \beta^k(g_\tau(\xi) - \xi)\big) - \Psi_\tau(\xi) \le \alpha\beta^k\theta_\tau(\xi)\big\} & \text{if } \Psi_\tau(\xi) > 0. \end{cases}$$

Lemma 4.13 proves that if $\theta_\tau(\xi) < 0$ for some $\xi \in \mathcal{X}_{\tau,p}$, then $\mu_\tau(\xi) < \infty$, which means that we can construct a point $\big(\xi + \beta^{\mu_\tau(\xi)}(g_\tau(\xi) - \xi)\big) \in \mathcal{X}_{\tau,r}$ that produces a reduction in the cost (if ξ is feasible) or a reduction in the infeasibility (if ξ is infeasible).

Continuing as we did in section 4.4 in [14], notice that the Haar wavelet operator $\mathcal{F}_N$ induces a partition in $\mathcal{T}_N$ according to the times at which the constructed relaxed discrete input switched. That is, let the *induced partition function*, $\sigma_N \colon \mathcal{X}_r \to \mathcal{T}_N$, be defined as

$$\sigma_N(u, d) = \{0\} \cup \left\{ \frac{k}{2^N} + \frac{1}{2^N} \sum_{j=1}^{i} \left[\mathcal{F}_N(d)\right]_j \left(\frac{k}{2^N}\right) \right\}_{\substack{k\in\{0,\ldots,2^N-1\} \\ i\in\{1,\ldots,q\}}}. \tag{3.16}$$

Thus, $\rho_N(\xi) \in \mathcal{X}_{\sigma_N(\xi),p}$ for each $\xi \in \mathcal{X}_r$.

Let $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, $\bar\alpha \in (0,\infty)$, $\bar\beta \in (\frac{1}{\sqrt{2}}, 1)$, $\omega \in (0,1)$, and $k_{\mathrm{M}} \in \mathbb{N}$. Then, similar to our definition of ν in [14], we modulate a point ξ with frequency $2^{-\nu_\tau(\xi,k_{\mathrm{M}})}$, where the *discretized frequency modulation function*, $\nu_\tau \colon \mathcal{X}_{\tau,p} \times \mathbb{N} \to \mathbb{N}$, is defined as

$$\begin{aligned}
&\nu_\tau(\xi, k_{\mathrm{M}}) \\
&= \begin{cases}
\min\big\{k \in \mathbb{N} \mid k \le k_{\mathrm{M}},\ \xi' = \xi + \beta^{\mu_\tau(\xi)}\big(g_\tau(\xi) - \xi\big), \\
\qquad J_{\sigma_k(\xi')}\big(\rho_k(\xi')\big) - J_\tau(\xi) \le \big(\alpha\beta^{\mu_\tau(\xi)} - \bar{\alpha}\bar{\beta}^k\big)\theta_\tau(\xi), \\
\qquad \Psi_{\sigma_k(\xi')}\big(\rho_k(\xi')\big) \le \big(\alpha\beta^{\mu_\tau(\xi)} - \bar{\alpha}\bar{\beta}^k\big)\theta_\tau(\xi), \\
\qquad \bar{\alpha}\bar{\beta}^k \le (1-\omega)\alpha\beta^{\mu_\tau(\xi)}\big\} & \text{if } \Psi_\tau(\xi) \le 0, \\
\min\big\{k \in \mathbb{N} \mid k \le k_{\mathrm{M}},\ \xi' = \xi + \beta^{\mu_\tau(\xi)}\big(g_\tau(\xi) - \xi\big), \\
\qquad \Psi_{\sigma_k(\xi')}\big(\rho_k(\xi')\big) - \Psi_\tau(\xi) \le \big(\alpha\beta^{\mu_\tau(\xi)} - \bar{\alpha}\bar{\beta}^k\big)\theta_\tau(\xi), \\
\qquad \bar{\alpha}\bar{\beta}^k \le (1-\omega)\alpha\beta^{\mu_\tau(\xi)}\big\} & \text{if } \Psi_\tau(\xi) > 0.
\end{cases}
\end{aligned} \tag{3.17}$$

There is no guarantee that the optimization problem in ν_τ is feasible. Without loss of generality, we let $\nu_\tau(\xi, k_{\mathrm{M}}) = \infty$ for each $\xi \in \mathcal{X}_{\tau,p}$ when there is no feasible solution. Importantly, in Lemma 4.16 we prove that given $N_0 \in \mathbb{N}$, $\tau \in \mathcal{T}_{N_0}$, and $\xi \in \mathcal{X}_{\tau,p}$, if $\theta(\xi) < 0$, then for each $\eta \in \mathbb{N}$ there exists $N \ge N_0$ such that $\nu_{\sigma_N(\xi)}(\xi, N+\eta)$ is finite, i.e., that using (3.17) we can compute a new point that moves closer to the optimality condition with a known switching time sequence.

3.5. An implementable switched system optimal control algorithm. Algorithm 3.1 describes our numerical method to solve P_p, as defined in [14]. Note that at each step of this algorithm, $\xi_j \in \mathcal{X}_{\tau_j,p}$. Also note that the condition in line 4 checks if θ_τ gets too close to zero or if ν_τ equals infinity, in which case the discretization precision is increased. To understand why the first condition in line 4 is required, recall that we look for points ξ such that $\theta(\xi) = 0$ and that $\theta_{\tau_j}(\xi_j) = 0$ does not imply that $\theta(\xi_j) = 0$, and thus if $\theta_{\tau_j}(\xi_j)$ is too close to zero the algorithm may terminate without reaching our objective. Finally, notice that due to the definitions of $\mathrm{D}J_\tau$ and $\mathrm{D}\psi_{\tau,j,t_k}$ for each $j \in \mathcal{J}$ and $k \in \{0, \dots, |\tau|\}$ which appear in Corollary 4.5, θ_τ can be solved via a quadratic program.

Algorithm 3.1 Numerically tractable optimization algorithm for P_p.

```
Require: $N_0 \in \mathbb{N}$, $\tau_0 \in \mathcal{T}_{N_0}$, $\xi_0 \in \mathcal{X}_{\tau_0,p}$, $\alpha, \beta, \omega \in (0,1)$, $\bar{\alpha}, \gamma, \Lambda \in (0,\infty)$, $\bar{\beta} \in \big(\frac{1}{\sqrt{2}}, 1\big)$,
    $\eta \in \mathbb{N}$, $\chi \in \big(0, \frac{1}{2}\big)$.
 1: Set $j = 0$.
 2: loop
 3:   Compute $\theta_{\tau_j}(\xi_j)$, $g_{\tau_j}(\xi_j)$, $\mu_{\tau_j}(\xi_j)$, and $\nu_{\tau_j}(\xi_j, N_j + \eta)$.
 4:   if $\theta_{\tau_j}(\xi_j) > -\Lambda 2^{-\chi N_j}$ or $\nu_{\tau_j}(\xi_j, N_j + \eta) = \infty$ then
 5:     Set $\xi_{j+1} = \xi_j$.
 6:     Set $N_{j+1} = N_j + 1$.
 7:   else
 8:     Set $\xi_{j+1} = \rho_{\nu_{\tau_j}(\xi_j, N_j+\eta)}\big(\xi_j + \beta^{\mu_{\tau_j}(\xi_j)}(g_{\tau_j}(\xi_j) - \xi_j)\big)$.
 9:     Set $N_{j+1} = \max\big\{N_j, \nu_{\tau_j}(\xi_j, N_j + \eta)\big\}$.
10:   end if
11:   Set $\tau_{j+1} = \sigma_{N_{j+1}}(\xi_{j+1})$.
12:   Replace $j$ by $j + 1$.
13: end loop
```

For analysis purposes, we define the *implementable algorithm recursion function*, $\Gamma_\tau \colon \{\xi \in \mathcal{X}_{\tau,p} \mid \nu_\tau(\xi, k_{\mathrm{M}}) < \infty\} \to \mathcal{X}_p$, as

$$\Gamma_\tau(\xi) = \rho_{\nu_\tau(\xi, k_{\mathrm{M}})}\big(\xi + \beta^{\mu_\tau(\xi)}(g_\tau(\xi) - \xi)\big). \tag{3.18}$$

We prove several important properties about the sequence generated by Algorithm 3.1. First, we prove in Lemma 4.17 that as the iteration index i increases $\lim_{i\to\infty} N_i = \infty$ which implies that the discretization mesh size tends to zero. Second, letting $\{N_i\}_{i\in\mathbb{N}}$, $\{\tau_i\}_{i\in\mathbb{N}}$, and $\{\xi_i\}_{i\in\mathbb{N}}$ be the sequences produced by Algorithm 3.1, then, as we prove in Lemma 4.18, there exists $i_0 \in \mathbb{N}$ such that if $\Psi_{\tau_{i_0}}(\xi_{i_0}) \leq 0$, then $\Psi(\xi_i) \leq 0$ and $\Psi_{\tau_i}(\xi_i) \leq 0$ for each $i \geq i_0$. Finally, as we prove in Theorem 4.20, $\lim_{j\to\infty} \theta(\xi_j) = 0$, i.e., Algorithm 3.1 produces sequences that asymptotically satisfy the optimality condition.

4. Implementable algorithm analysis.

In this section, we derive the various components of Algorithm 3.1 and prove that it produces sequences that asymptotically satisfy our optimality condition. Our argument proceeds as follows: first, we prove the continuity and convergence of the discretized state, cost, and constraints to their infinite-dimensional analogues; second, we construct the components of the optimality function and prove the convergence of these discretized components to their infinite-dimensional analogues; finally, we prove the convergence of our algorithm.

4.1. Continuity and convergence of the discretized components.

In this subsection, we note the continuity and convergence of the discretized state, cost, and constraint. The following two lemmas follow from the discrete Bellman–Gronwall inequality and Jensen's inequality.

LEMMA 4.1. *There exists $C > 0$ such that for each $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, $\xi \in \mathcal{X}_{\tau,r}$, and $t \in [0,1]$, $\left\| z_\tau^{(\xi)}(t) \right\|_2 \leq C$.*

LEMMA 4.2. *There exists $L > 0$ such that for each $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, $\xi_1, \xi_2 \in \mathcal{X}_{\tau,r}$, and $t \in [0,1]$, $\left\| \phi_{\tau,t}(\xi_1) - \phi_{\tau,t}(\xi_2) \right\|_2 \leq L \left\| \xi_1 - \xi_2 \right\|_{\mathcal{X}}$.*

We now find the rate of convergence of our discretization scheme.

LEMMA 4.3. *There exists $C > 0$ such that for each $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, $\xi \in \mathcal{X}_{\tau,r}$, and $t \in [0,1]$,*

(1) $\left\| z_\tau^{(\xi)}(t) - x^{(\xi)}(t) \right\|_2 \leq \frac{C}{2^N}$,

(2) $\left| J_\tau(\xi) - J(\xi) \right| \leq \frac{C}{2^N}$, *and*

(3) $\left| \Psi_\tau(\xi) - \Psi(\xi) \right| \leq \frac{C}{2^N}$.

Proof. The first result follows directly from Picard's lemma (Lemma 5.6.3 in [12]). The second result follows from the Lipschitz continuity of h_0 and Lemma 4.3.

To prove the third result, first note that for each $\tau \in \mathcal{T}_N$, $k \in \{0, \ldots, |\tau| - 1\}$, and $t \in [t_k, t_{k+1}]$, $\left| h_j\big(x^{(\xi)}(t)\big) - h_j\big(x^{(\xi)}(t_k)\big) \right| \leq \frac{C'}{2^N}$, which follows from Theorem 3.3 and Assumption 3.4 in [14]. Thus, by the triangle inequality there exists $C > 0$ such that for each $t \in [t_k, t_{k+1}]$, $\left| h_j\big(x^{(\xi)}(t)\big) - h_j\big(z_\tau^{(\xi)}(t_k)\big) \right| \leq \frac{C}{2^N}$.

Let $t' \in \arg\max_{t\in[0,1]} h_j\big(x^{(\xi)}(t)\big)$ and $\kappa(t') \in \{0, \ldots, |\tau| - 1\}$ such that $t' \in [t_{\kappa(t')}, t_{\kappa(t')+1}]$. Then,

$$\max_{t\in[0,1]} h_j\big(x^{(\xi)}(t)\big) - \max_{k\in\{0,\ldots,|\tau|\}} h_j\big(z_\tau^{(\xi)}(t_k)\big) \leq h_j\big(x^{(\xi)}(t')\big) - h_j\big(z_\tau^{(\xi)}(t_{\kappa(t')})\big). \tag{4.1}$$

Similarly if $k' \in \arg\max_{k\in\{0,\ldots,|\tau|\}} h_j\big(z_\tau^{(\xi)}(t_k)\big)$, then

$$\max_{k\in\{0,\ldots,|\tau|\}} h_j\big(z_\tau^{(\xi)}(t_k)\big) - \max_{t\in[0,1]} h_j\big(x^{(\xi)}(t)\big) \leq h_j\big(z_\tau^{(\xi)}(t_{k'})\big) - h_j\big(x^{(\xi)}(t_{k'})\big). \tag{4.2}$$

The result follows directly from these inequalities. □

4.2. Derivation of the implementable algorithm terms.

Next, we formally derive the components of θ_τ, prove its well-posedness, and bound the error between θ_τ and θ.

Lemma 4.4. *Let $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, $\xi = (u, d) \in \mathcal{X}_{\tau,r}$, $\xi' = (u', d') \in \mathcal{X}_\tau$. Then, for each $k \in \{0, \dots, |\tau|\}$, the directional derivative of ϕ_{τ,t_k} is*

$$\begin{aligned}(4.3)\quad \mathrm{D}\phi_{\tau,t_k}(\xi;\xi') = \sum_{j=0}^{k-1}(t_{j+1}-t_j)\Phi_\tau^{(\xi)}(t_k,t_{j+1})\bigg(&\frac{\partial f}{\partial u}\big(t_j,\phi_{\tau,t_j}(\xi),u(t_j),d(t_j)\big)u'(t_j)\\ &+ f\big(t_j,\phi_{\tau,t_j}(\xi),u(t_j),d'(t_j)\big)\bigg),\end{aligned}$$

where $\Phi_\tau^{(\xi)}(t_k,t_j)$ is the solution of the following matrix difference equation:

(4.4)

$$\begin{aligned}\Phi_\tau^{(\xi)}(t_{k+1},t_j) = \Phi_\tau^{(\xi)}(t_k,t_j) + (t_{k+1}-t_k)\frac{\partial f}{\partial x}\big(t_k,\phi_{\tau,t_k}(\xi),u(t_k),d(t_k)\big)\Phi_\tau^{(\xi)}(t_k,t_j)\\ \forall k \in \{0,\dots,|\tau|-1\}, \quad \Phi_\tau^{(\xi)}(t_j,t_j) = I.\end{aligned}$$

Proof. For notational convenience, let $z^{(\lambda)} = z^{(\xi+\lambda\xi')}$, $u^{(\lambda)} = u + \lambda u'$, $d^{(\lambda)} = d + \lambda d'$, $\Delta z^{(\lambda)} = z^{(\lambda)} - z^{(\xi)}$, and $\Delta t_k = t_{k+1} - t_k$. Thus, for each $k \in \{0, \dots, |\tau|\}$,

$$\begin{aligned}(4.5)\qquad \Delta z^{(\lambda)}(t_k) = \sum_{j=0}^{k-1}\Delta t_j\bigg(&\lambda\sum_{i=1}^{q} d_i'(t_j) f\big(t_j, z^{(\lambda)}(t_j), u^{(\lambda)}(t_j), e_i\big)\\ &+ \frac{\partial f}{\partial x}\big(t_j, z^{(\xi)}(t_j) + \upsilon_{x,j}\Delta z^{(\lambda)}(t_j), u^{(\lambda)}(t_j), d(t_j)\big)\Delta z^{(\lambda)}(t_j)\\ &+ \lambda\frac{\partial f}{\partial u}\big(t_j, z^{(\xi)}(t_j), u(t_j) + \upsilon_{u,j}\lambda u'(t_j), d(t_j)\big)u'(t_j)\bigg),\end{aligned}$$

where $\{\upsilon_{x,j}\}_{j=0}^{|\tau|}, \{\upsilon_{u,j}\}_{j=0}^{|\tau|} \subset [0,1]$. Let $\{y(t_k)\}_{k=0}^{|\tau|}$ be recursively defined as follows: $y(t_0) = 0$, and for each $k \in \{0, \dots, |\tau| - 1\}$,

$$\begin{aligned}(4.6)\quad y(t_{k+1}) = y(t_k) + \Delta t_k\bigg(&\frac{\partial f}{\partial x}\big(t_k, z^{(\xi)}(t_k), u(t_k), d(t_k)\big)y(t_k)\\ &+ \frac{\partial f}{\partial u}\big(t_k, z^{(\xi)}(t_k), u(t_k), d(t_k)\big)u'(t_k) + f\big(t_k, z^{(\xi)}(t_k), u(t_k), d'(t_k)\big)\bigg).\end{aligned}$$

We want to show that $\lim_{\lambda\downarrow 0}\big\|\frac{1}{\lambda}\Delta z^{(\lambda)}(t_k) - y(t_k)\big\| = 0$. Consider

$$\begin{aligned}(4.7)\qquad &\bigg\|\frac{\partial f}{\partial x}\big(t_k, z^{(\xi)}(t_k), u(t_k), d(t_k)\big)y(t_k)\\ &\qquad - \frac{\partial f}{\partial x}\big(t_k, z^{(\xi)}(t_k) + \upsilon_{x,k}\Delta z^{(\lambda)}(t_k), u^{(\lambda)}(t_k), d(t_k)\big)\frac{\Delta z^{(\lambda)}(t_k)}{\lambda}\bigg\|_2\\ &\le L\left\|y(t_k) - \frac{\Delta z^{(\lambda)}(t_k)}{\lambda}\right\|_2 + L\big(\big\|\Delta z^{(\lambda)}(t_k)\big\|_2 + \lambda\left\|u'(t_k)\right\|_2\big)\left\|y(t_k)\right\|_2,\end{aligned}$$

which follows from Assumption 3.2 in [14] and the triangle inequality. Also,

$$\begin{aligned}(4.8)\qquad &\bigg\|\bigg(\frac{\partial f}{\partial u}\big(t_k, z^{(\xi)}(t_k), u(t_k), d(t_k)\big)\\ &\qquad - \frac{\partial f}{\partial u}\big(t_k, z^{(\xi)}(t_k), u(t_k) + \upsilon_{u,k}\lambda u'(t_k), d(t_k)\big)\bigg)u'(t_k)\bigg\|_2 \le L\lambda\left\|u'(t_k)\right\|_2^2\end{aligned}$$

and

$$\left\| \sum_{i=1}^{q} d_i'(t_k)\Big(f\big(t_k, z^{(\xi)}(t_k), u(t_k), e_i\big) - f\big(t_k, z^{(\lambda)}(t_k), u^{(\lambda)}(t_k), e_i\big)\Big) \right\|_2 \qquad (4.9)$$
$$\leq L \big\| \Delta z^{(\lambda)}(t_k) \big\|_2 + L\lambda \left\| u'(t_k) \right\|_2 .$$

Using the discrete Bellman–Gronwall inequality and the inequalities above,

$$\left\| y(t_k) - \frac{\Delta z^{(\lambda)}(t_k)}{\lambda} \right\|_2 \leq L e^L \sum_{j=0}^{k-1} \Delta t_j \Big(\big(\big\| \Delta z^{(\lambda)}(t_j) \big\|_2 + \lambda \left\| u'(t_j) \right\|_2 \big) \left\| y(t_j) \right\|_2 \qquad (4.10)$$
$$+ \lambda \left\| u'(t_j) \right\|_2^2 + \big\| \Delta z^{(\lambda)}(t_j) \big\|_2 + \lambda \left\| u'(t_j) \right\|_2 \Big),$$

where we used the fact that $\big(1 + \frac{L}{2^N}\big)^{\frac{L}{2^N}} \leq e^L$. By Lemma 4.2, the right-hand side of (4.10) goes to zero as $\lambda \downarrow 0$, thus showing that $\frac{\Delta z^{(\lambda)}(t_k)}{\lambda} \to y(t_k)$. The result of the lemma is obtained after noting that $\mathrm{D}\phi_{\tau,t_k}(\xi;\xi')$ is equal to $y(t_k)$ for each $k \in \{0, \ldots, |\tau|\}$. □

Next, we construct a directional derivative of the discretized cost and component constraint functions, which follow by the chain rule and Lemma 4.4.

Corollary 4.5. *Let $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, $\xi \in \mathcal{X}_{\tau,r}$, $\xi' \in \mathcal{X}_\tau$, $j \in \mathcal{J}$, and $k \in \{0, \ldots, |\tau|\}$. Then the directional derivatives of the discretized cost J_τ and component constraint ψ_{τ,j,t_k} in the ξ' direction are*

(4.11)
$$\mathrm{D}J_\tau(\xi;\xi') = \frac{\partial h_0}{\partial x}\big(\phi_{\tau,1}(\xi)\big)\mathrm{D}\phi_{\tau,1}(\xi;\xi'), \quad \mathrm{D}\psi_{\tau,j,t_k}(\xi;\xi') = \frac{\partial h_j}{\partial x}\big(\phi_{\tau,t_k}(\xi)\big)\mathrm{D}\phi_{\tau,t_k}(\xi;\xi').$$

We omit the proof of the following lemma since it follows from arguments similar to Lemma 5.4 in [14].

Lemma 4.6. *There exists $L > 0$ such that for each $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, $\xi_1, \xi_2 \in \mathcal{X}_{\tau,r}$, $\xi' \in \mathcal{X}_\tau$, and $k \in \{0, \ldots, |\tau|\}$,*
(1) $\left\| \mathrm{D}\phi_{\tau,t_k}(\xi_1;\xi') - \mathrm{D}\phi_{\tau,t_k}(\xi_2;\xi') \right\|_2 \leq L \left\| \xi_1 - \xi_2 \right\|_{\mathcal{X}} \left\| \xi' \right\|_{\mathcal{X}}$,
(2) $\left| \mathrm{D}J_\tau(\xi_1;\xi') - \mathrm{D}J_\tau(\xi_2;\xi') \right| \leq L \left\| \xi_1 - \xi_2 \right\|_{\mathcal{X}} \left\| \xi' \right\|_{\mathcal{X}}$,
(3) $\left| \mathrm{D}\psi_{\tau,j,t_k}(\xi_1;\xi') - \mathrm{D}\psi_{\tau,j,t_k}(\xi_2;\xi') \right| \leq L \left\| \xi_1 - \xi_2 \right\|_{\mathcal{X}} \left\| \xi' \right\|_{\mathcal{X}}$, *and*
(4) $\left| \zeta_\tau(\xi_1,\xi') - \zeta_\tau(\xi_2,\xi') \right| \leq L \left\| \xi_1 - \xi_2 \right\|_{\mathcal{X}}$.

Lemma 4.7. *There exists $C > 0$ such that for each $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, $\xi \in \mathcal{X}_{\tau,r}$, $\xi' \in \mathcal{X}_\tau$ and $k \in \{0, \ldots, |\tau|\}$,*
(1) $\left\| \mathrm{D}\phi_{t_k}(\xi;\xi') - \mathrm{D}\phi_{\tau,t_k}(\xi;\xi') \right\|_2 \leq \frac{C}{2^N}$,
(2) $\left| \mathrm{D}J_\tau(\xi;\xi') - \mathrm{D}J(\xi;\xi') \right| \leq \frac{C}{2^N}$,
(3) $\left| \mathrm{D}\psi_{\tau,j,t_k}(\xi;\xi') - \mathrm{D}\psi_{j,t_k}(\xi;\xi') \right| \leq \frac{C}{2^N}$, *and*
(4) $\left| \zeta_\tau(\xi,\xi') - \zeta(\xi,\xi') \right| \leq \frac{C}{2^N}$.

Proof. By applying Picard's lemma (Lemma 5.6.3 in [12]) together with the triangle inequality, one can show that the error between $\Phi_\tau^{(\xi)}$ and $\Phi^{(\xi)}$ is bounded by $\frac{C}{2^N}$, for some $C > 0$. Conditions (1), (2), and (3) then follow from Lemma 4.6, the boundedness and Lipschitz continuity of the vector fields and their derivatives, and the boundedness of the state transition matrix.

To prove condition (4), let $\kappa(t) \in \{0, \ldots, |\tau| - 1\}$ such that $t \in [t_{\kappa(t)}, t_{\kappa(t)+1}]$ for each $t \in [0, 1]$. Then, applying condition (3) and the triangle inequality we get that there exists $C > 0$ such that $\left|\mathrm{D}\psi_{j,t}(\xi;\xi'-\xi) - \mathrm{D}\psi_{\tau,j,t_{\kappa(t)}}(\xi;\xi'-\xi)\right| \leq \frac{C}{2^N}$ for each $j \in \mathcal{J}$ and $t \in [0,1]$. The result follows using the same argument as in the proof of Lemma 4.3. □

The following result shows that g_τ is a well-defined function.

Lemma 4.8. *Let $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, and $\xi \in \mathcal{X}_{\tau,p}$. Then the map $\xi' \mapsto \zeta_\tau(\xi,\xi') + \mathrm{V}(\xi'-\xi)$, with domain $\mathcal{X}_{\tau,r}$, is strictly convex and has a unique minimizer.*

Proof. The map is strictly convex since it is the sum of the maximum of linear functions of $\xi'-\xi$ (which is convex), the total variation of $\xi'-\xi$ (which is convex), and the $\mathcal{X}$-norm of $\xi'-\xi$ (which is strictly convex). The uniqueness of the minimizer follows after noting that $\mathcal{X}_{\tau,r}$ is a compact set. □

Employing these results we can prove the continuity of θ_τ. This result is not strictly required in order to prove the convergence of Algorithm 3.1. However, it is useful from an implementation point of view, since continuity gives a guarantee that our numerical computations are good approximations.

Lemma 4.9. *Let $N \in \mathbb{N}$ and $\tau \in \mathcal{T}_N$. Then θ_τ is continuous.*

Proof. First note that V, when restricted to elements in $\mathcal{X}_{\tau,r}$, is a continuous operator, since the supremum in its definition is always attained at the partition induced by τ.

Consider a sequence $\{\xi_i\}_{i\in\mathbb{N}} \subset \mathcal{X}_{\tau,r}$ converging to $\xi \in \mathcal{X}_{\tau,r}$. Since $\theta_\tau(\xi_i) \leq \zeta_\tau\big(\xi_i, g_\tau(\xi)\big) + \mathrm{V}(g_\tau(\xi)-\xi_i)$ for each $i \in \mathbb{N}$, then by condition (4) in Lemma 4.6 and the continuity of V we have $\limsup_{i\to\infty}\theta_\tau(\xi_i) \leq \theta_\tau(\xi)$, which proves the upper semicontinuity of θ_τ. Using a similar argument and the reverse triangle inequality, we get that

$$\begin{aligned}\theta_\tau(\xi) &\leq \zeta_\tau\big(\xi, g_\tau(\xi_i)\big) + \mathrm{V}(g_\tau(\xi_i)-\xi) - \theta_\tau(\xi_i) + \theta_\tau(\xi_i)\\ &\leq L\left\|\xi-\xi_i\right\|_{\mathcal{X}} + \mathrm{V}(\xi-\xi_i) + \theta_\tau(\xi_i).\end{aligned} \tag{4.12}$$

Taking a lim inf proves the lower semicontinuity of θ_τ and our desired result. □

Next, we prove that the local minimizers of P_τ are in fact zeros of θ_τ.

Theorem 4.10. *Let $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$. Then, θ_τ is nonpositive valued. Moreover, if $\xi \in \mathcal{X}_{\tau,p}$ is a local minimizer of P_τ, then $\theta_\tau(\xi) = 0$.*

Proof. Note that $\zeta_\tau(\xi,\xi) = 0$ and $\theta_\tau(\xi) \leq \zeta_\tau(\xi,\xi)$, which proves the first part.

To prove the second part, we begin by making several observations. Given $\xi' \in \mathcal{X}_{\tau,r}$ and $\lambda \in [0,1]$, using the mean value theorem and condition (2) in Lemma 4.6, there exists $L > 0$ such that

$$J_\tau\big(\xi+\lambda(\xi'-\xi)\big) - J_\tau(\xi) \leq \lambda\,\mathrm{D}J_\tau\big(\xi;\xi'-\xi\big) + L\lambda^2\left\|\xi'-\xi\right\|_{\mathcal{X}}^2. \tag{4.13}$$

Let $\mathcal{A}_\tau(\xi) = \arg\min\big\{h_j\big(\phi_{\tau,t_k}(\xi)\big) \mid (j,k) \in \mathcal{J}\times\{0,\dots,|\tau|\}\big\}$; then there exists a pair $(j,k) \in \mathcal{A}\big(\xi+\lambda(\xi'-\xi)\big)$ such that using condition (3) in Lemma 4.6

$$\Psi_\tau\big(\xi+\lambda(\xi'-\xi)\big) - \Psi_\tau(\xi) \leq \lambda\,\mathrm{D}\psi_{\tau,j,t_k}\big(\xi;\xi'-\xi\big) + L\lambda^2\left\|\xi'-\xi\right\|_{\mathcal{X}}^2. \tag{4.14}$$

We proceed by contradiction, i.e., we assume that $\theta_\tau(\xi) < 0$ and show that for each $\varepsilon > 0$ there exists $\hat{\xi} \in \mathcal{N}_\tau(\xi,\varepsilon) \cap \big\{\bar{\xi} \in \mathcal{X}_{\tau,r} \mid \Psi_\tau(\bar{\xi}) \leq 0\big\}$ such that $J_\tau(\hat{\xi}) < J_\tau(\xi)$. Note that since $\theta_\tau(\xi) < 0$, $g_\tau(\xi) \neq \xi$. For each $\lambda > 0$ by using (4.13), $J_\tau\big(\xi+\lambda(g_\tau(\xi)-\xi)\big) - J_\tau(\xi) \leq \theta_\tau(\xi)\lambda + 4A^2L\lambda^2$, where $A = \max\big\{\|u\|_2 + 1 \mid u \in U\big\}$ and we used the fact that $\left\|\xi'-\xi\right\|_{\mathcal{X}}^2 \leq 4A^2$ and $\mathrm{D}J_\tau(\xi; g(\xi)-\xi) \leq \theta_\tau(\xi)$. Hence for each $\lambda \in \big(0, \frac{-\theta_\tau(\xi)}{4A^2L}\big)$, $J_\tau\big(\xi+\lambda(g_\tau(\xi)-\xi)\big) - J_\tau(\xi) < 0$. Similarly, for each $\lambda > 0$, by using (4.14), we have

$$\Psi_\tau\big(\xi+\lambda(g_\tau(\xi)-\xi)\big) \leq \Psi_\tau(\xi) + \big(\theta_\tau(\xi) - \gamma\Psi_\tau(\xi)\big)\lambda + 4A^2L\lambda^2, \tag{4.15}$$

where we used the fact that $\mathrm{D}\psi_{\tau,j,t_k}(\xi; g(\xi)-\xi) \leq \theta_\tau(\xi) - \gamma\Psi_\tau(\xi)$ for each $(j,k) \in \mathcal{J} \times \{0,\ldots,|\tau|\}$. Hence for each $\lambda \in \big(0, \min\big\{\frac{-\theta_\tau(\xi)}{4A^2L}, \frac{1}{\gamma}\big\}\big)$, $\Psi_\tau(\xi + \lambda(g_\tau(\xi)-\xi)) \leq (1-\gamma\lambda)\Psi_\tau(\xi) \leq 0$.

Summarizing, suppose ξ is a local minimizer of P_τ and $\theta_\tau(\xi) < 0$. For each $\varepsilon > 0$, by choosing any

$$\lambda \in \left(0, \min\left\{\frac{-\theta_\tau(\xi)}{4A^2L}, \frac{1}{\gamma}, \frac{\varepsilon}{\|g_\tau(\xi)-\xi\|_{\mathcal{X}}}\right\}\right), \tag{4.16}$$

we can construct a new point $\hat{\xi} = \xi + \lambda\big(g_\tau(\xi) - \xi\big) \in \mathcal{X}_{\tau,r}$ such that $\hat{\xi} \in \mathcal{N}_\tau(\xi,\varepsilon)$, $J_\tau(\hat{\xi}) < J_\tau(\xi)$, and $\Psi_\tau(\hat{\xi}) \leq 0$. This is a contradiction and proves our result. □

LEMMA 4.11. *Let $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, and $S_\tau\colon L^2([0,1],\mathbb{R}^n) \cap BV([0,1],\mathbb{R}^n) \to L^2([0,1],\mathbb{R}^n)$ defined by $S_\tau(f) = \sum_{k=0}^{|\tau|-1} f(t_k)\mathbb{1}_{[t_k,t_{k+1})}$. Then, $\|S_\tau(f) - f\|_{L^2} \leq 2^{-\frac{N}{2}}\mathrm{V}(f)$.*

Proof. Let $\varepsilon > 0$, $\bar{\varepsilon} = \frac{\varepsilon}{|\tau|}$, and $\alpha_k \in [t_k, t_{k+1}]$ such that $\|f(t_k) - f(\alpha_k)\|_2^2 + \bar{\varepsilon}^2 \geq \sup_{s\in[t_k,t_{k+1}]}\|f(t_k) - f(s)\|_2^2$ for each $k \in \{0,\ldots,|\tau|-1\}$. Consider

$$\begin{aligned}\|S_\tau(f) - f\|_{L^2} &= \sqrt{\sum_{k=0}^{|\tau|-1}\int_{t_k}^{t_{k+1}} \|f(t_k) - f(s)\|_2^2\, ds} \leq \sqrt{\frac{1}{2^N}\sum_{k=0}^{|\tau|-1}\|f(t_k) - f(\alpha_k)\|_2^2 + \bar{\varepsilon}^2} \\ &\leq 2^{-\frac{N}{2}}\sum_{k=0}^{|\tau|-1}\big(\|f(t_k) - f(\alpha_k)\|_2 + \bar{\varepsilon}\big) \leq 2^{-\frac{N}{2}}\big(\mathrm{V}(f) + \varepsilon\big),\end{aligned} \tag{4.17}$$

where we used the fact that $\|x\|_2 \leq \|x\|_1$ for each $x \in \mathbb{R}^n$, and the last inequality follows by the definition of V ((3.3) in [14]). The result is obtained after noting that (4.17) is valid for each $\varepsilon > 0$. □

Finally, we prove that P_τ consistently approximates P_p when the cardinality of the sequence of switching times satisfies an upper bound, which is satisfied by the sequence of switching times generated by Algorithm 3.1.

THEOREM 4.12. *Let $\{\tau_i\}_{i\in\mathbb{N}}$ and $\{\xi_i\}_{i\in\mathbb{N}}$ such that $\tau_i \in \mathcal{T}_i$ and $\xi_i \in \mathcal{X}_{\tau_i,p}$ for each $i \in \mathbb{N}$. Then $\liminf_{i\to\infty}\theta_{\tau_i}(\xi_i) \leq \liminf_{i\to\infty}\theta(\xi_i)$.*

Proof. By the nonpositivity of θ, shown in Theorem 5.6 in [14], as well as Theorem 3.3 and Lemma 5.4 in [14], $\mathrm{V}(g(\xi) - \xi)$ is uniformly bounded for each $\xi \in \mathcal{X}_p$. Let $\xi_i' = S_\tau\big(g(\xi_i) - \xi_i\big)$ for each $i \in \mathbb{N}$, where S_τ is as defined in Lemma 4.11 and where we abuse notation by writing $S_\tau(\xi) = \big(S_\tau(u), S_\tau(d)\big)$ for each $\xi = (u,d)$. Hence, also by Lemma 4.11, there exists $C' > 0$ such that $\|g(\xi_i) - \xi_i - \xi_i'\|_{\mathcal{X}} \leq C'\,2^{-\frac{N}{2}}$.

Now, using condition (4) in Lemma 4.7, there exists $C > 0$ such that

$$\begin{aligned}\theta_{\tau_i}(\xi_i) - \theta(\xi_i) &\leq \zeta_{\tau_i}(\xi_i, \xi_i' + \xi_i) + \mathrm{V}(\xi_i') - \zeta(\xi_i, \xi_i' + \xi_i) - \mathrm{V}(\xi_i') \\ &\quad + \zeta(\xi_i, \xi_i' + \xi_i) + \mathrm{V}(\xi_i') - \zeta\big(\xi_i, g(\xi_i)\big) - \mathrm{V}(g(\xi_i) - \xi_i) \\ &\leq \frac{C}{2^i} + \zeta(\xi_i, \xi_i' + \xi_i) - \zeta\big(\xi_i, g(\xi_i)\big) + \mathrm{V}(\xi_i') - \mathrm{V}(g(\xi_i) - \xi_i).\end{aligned} \tag{4.18}$$

Using the same argument as in the proof of Lemma 5.4 in [14], there exists $L > 0$ such that $\big|\zeta(\xi_i, \xi_i' + \xi_i) - \zeta\big(\xi_i, g(\xi_i)\big)\big| \leq L\,\|g(\xi_i) - \xi_i - \xi_i'\|_{\mathcal{X}}$. Also notice that by the definitions of the set $\{\xi_i'\}_{i\in\mathbb{N}}$ and V, $V(\xi_i') \leq V(g(\xi_i) - \xi_i)$ for each $i \in \mathbb{N}$. Finally, using all the inequalities above we have $\liminf_{i\to\infty}\theta_{\tau_i}(\xi_i) - \theta(\xi_i) \leq 0$, as desired. □

4.3. Convergence of the implementable algorithm. In this subsection, we prove that the sequence of points generated by Algorithm 3.1 asymptotically satisfies our optimality condition. We begin by proving that μ_τ is bounded whenever θ_τ is negative.

LEMMA 4.13. *Let $\alpha, \beta \in (0,1)$. For every $\delta > 0$, there exists $M^*_\delta < \infty$ such that if $\theta_\tau(\xi) \leq -\delta$ for $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, and $\xi \in \mathcal{X}_{\tau,p}$, then $\mu_\tau(\xi) \leq M^*_\delta$.*

Proof. Assume that $\Psi_\tau(\xi) \leq 0$. Using (4.13),

$$J_\tau\big(\xi + \beta^k(g_\tau(\xi) - \xi)\big) - J_\tau(\xi) - \alpha\beta^k\theta_\tau(\xi) \leq -(1-\alpha)\delta\beta^k + 4A^2L\beta^{2k}, \tag{4.19}$$

where $A = \max\big\{\|u\|_2 + 1 \mid u \in U\big\}$. Hence, for each $k \in \mathbb{N}$ such that $\beta^k \leq \frac{(1-\alpha)\delta}{4A^2L}$ we have that $J_\tau\big(\xi + \beta^k(g(\xi) - \xi)\big) - J_\tau(\xi) \leq \alpha\beta^k\theta_\tau(\xi)$. Similarly, using (4.14),

$$\Psi_\tau\big(\xi+\beta^k(g(\xi)-\xi)\big) - \Psi_\tau(\xi) + \beta^k\big(\gamma\Psi_\tau(\xi) - \alpha\theta_\tau(\xi)\big) \leq -(1-\alpha)\delta\beta^k + 4A^2L\beta^{2k}; \tag{4.20}$$

thus for each $k \in \mathbb{N}$ such that $\beta^k \leq \min\big\{\frac{(1-\alpha)\delta}{4A^2L}, \frac{1}{\gamma}\big\}$, $\Psi_\tau\big(\xi+\beta^k(g(\xi)-\xi)\big) - \alpha\beta^k\theta_\tau(\xi) \leq \big(1-\beta^k\gamma\big)\Psi_\tau(\xi) \leq 0$.

Now, assume $\Psi_\tau(\xi) > 0$. Using (4.14),

$$\Psi_\tau\big(\xi + \beta^k(g_\tau(\xi) - \xi)\big) - \Psi_\tau(\xi) - \alpha\beta^k\theta_\tau(\xi) \leq -(1-\alpha)\delta\beta^k + 4A^2L\beta^{2k}. \tag{4.21}$$

Hence, for each $k \in \mathbb{N}$ such that $\beta^k \leq \frac{(1-\alpha)\delta}{4A^2L}$, $\Psi_\tau\big(\xi + \beta^k(g_\tau(\xi) - \xi)\big) - \Psi_\tau(\xi) \leq \alpha\beta^k\theta_\tau(\xi)$.

Finally, let $M^*_\delta = 1 + \max\big\{\log_\beta\big(\frac{(1-\alpha)\delta}{4A^2L}\big), \log_\beta\big(\frac{1}{\gamma}\big)\big\}$, then we get that $\mu_\tau(\xi) \leq M^*_\delta$ as desired. □

The following corollary follows directly from the proof of Lemma 4.13.

COROLLARY 4.14. *Let $\alpha, \beta \in (0,1)$. There exists $\delta_0, C > 0$ such that if $\delta \in (0, \delta_0)$ and $\theta_\tau(\xi) \leq -\delta$ for $N \in \mathbb{N}$, $\tau \in \mathcal{T}_N$, and $\xi \in \mathcal{X}_{\tau,p}$, then $\mu_\tau(\xi) \leq 1 + \log_\beta(C\delta)$.*

Next, we show a bound of the error introduced in the trajectory after discretizing and applying the projection operator ρ_N. The proof follows by the triangle inequality, Theorem 5.10 in [14], and condition (1) in Lemma 4.3.

LEMMA 4.15. *There exists $K > 0$ such that for each $N_0, N \in \mathbb{N}$, $\tau \in \mathcal{T}_{N_0}$, $\xi = (u,d) \in \mathcal{X}_{r,\tau}$, and $t \in [0,1]$,*

$$\big\|\phi_{\sigma_N(\xi),t}\big(\rho_N(\xi)\big) - \phi_{\tau,t}(\xi)\big\|_2 \leq K\left(\left(\frac{1}{\sqrt{2}}\right)^N\big(\mathrm{V}(\xi) + 1\big) + \left(\frac{1}{2}\right)^{N_0}\right). \tag{4.22}$$

Now we show that ν_τ is eventually finite for points that do not satisfy the optimality condition infinitely often as the precision increases.

LEMMA 4.16. *Let $N_0 \in \mathbb{N}$, $\tau \in \mathcal{T}_{N_0}$, and $\xi \in \mathcal{X}_{\tau,r}$. Let $\mathcal{K} \subset \mathbb{N}$ be an infinite subsequence and $\delta > 0$ such that $\theta_{\sigma_N(\xi)} \leq -\delta$ for each $N \in \mathcal{K}$. Then for each $\eta \in \mathbb{N}$ there exists a finite $N^* \in \mathcal{K}$, $N^* \geq N_0$, such that $\nu_{\sigma_{N^*}(\xi)}(\xi, N + \eta)$ is finite.*

Proof. For each $N \in \mathbb{N}$, let $\xi'_N = \xi + \beta^{\mu_{\sigma_N(\xi)}(\xi)}\big(g_{\sigma_N(\xi)}(\xi) - \xi\big)$. Note that this lemma is equivalent to showing that the feasible set in the optimization problem solved by $\nu_{\sigma_N(\xi)}$ is nonempty for some $N \in \mathcal{K}$. Also note that by the nonpositivity of $\theta_{\sigma_N(\xi)}$, as shown in Theorem 4.10, together with Lemmas 4.1 and 4.6, there exists $C > 0$ such that for each $N \in \mathbb{N}$, $\mathrm{V}\big(g_{\sigma_N(\xi)}(\xi) - \xi\big) \leq C$, and thus $\mathrm{V}(\xi'_N) \leq \mathrm{V}(\xi) + C$. Using Lemma 4.15, there exists $N^* \in \mathcal{K}$ such that for each $N \geq N^*$, $N \in \mathcal{K}$,

$$J_{\sigma_k(\xi')}\big(\rho_k(\xi')\big) - J_{\sigma_N(\xi)}(\xi'_N) \le LK\left(\frac{1}{\sqrt{2}}\right)^N \big(\mathrm{V}(\xi')+2\big) \le \bar{\alpha}\bar{\beta}^N\frac{1}{2}\delta \le -\bar{\alpha}\bar{\beta}^N\theta_{\sigma_N(\xi)}(\xi), \tag{4.23}$$

and at the same time, $\bar{\alpha}\bar{\beta}^N \le (1-\omega)\alpha\beta^{M^*_\delta} \le (1-\omega)\alpha\beta^{\mu_{\sigma_N}(\xi)}$, where M^*_δ is as in Lemma 4.13. Let $\mathcal{A}_\tau(\xi) = \arg\max\{\psi_{\tau,j,t}(\xi) \mid (j,t) \in \mathcal{J}\times[0,1]\}$ and $(j,t) \in \mathcal{A}_{\sigma_N(\xi')}(\xi')$; then using the same argument as in (4.23),

$$\Psi_{\sigma_k(\xi')}\big(\rho_k(\xi')\big) - \Psi_{\sigma_N(\xi)}(\xi') \le \psi_{\sigma_k(\xi'),j,t}\big(\rho_k(\xi')\big) - \psi_{\sigma_N(\xi),j,t}(\xi') \le -\bar{\alpha}\bar{\beta}^N\theta_{\sigma_N(\xi)}(\xi), \tag{4.24}$$

as desired. □

Lemma 4.17. *Let $\{N_i\}_{i\in\mathbb{N}}$, $\{\tau_i\}_{i\in\mathbb{N}}$, and $\{\xi_i\}_{i\in\mathbb{N}}$ be the sequences generated by Algorithm* 3.1. *Then* $\lim_{i\to\infty} N_i = \infty$.

Proof. Suppose that $N_i \le N^*$ for all $i \in \mathbb{N}$. Then, by definition of Algorithm 3.1, there exists $i_0 \in \mathbb{N}$ such that $\theta_{\tau_i}(\xi_i) \le -\Lambda 2^{-\chi N_i} \le -\Lambda 2^{-\chi N^*}$, $\nu_{\tau_i}(\xi_i, N_i+\eta) < \infty$, and $\xi_{i+1} = \Gamma_{\tau_i}(\xi_i)$ for each $i \ge i_0$. Moreover, by the definition of ν_τ, if there exists $i_1 \ge i_0$ such that $\Psi_{\tau_{i_1}}(\xi_{i_1}) \le 0$, then $\Psi_{\tau_i}(\xi_i) \le 0$ for each $i \ge i_1$. Assume that there exists $i_1 \ge i_0$ such that $\Psi_{\tau_{i_1}}(\xi_{i_1}) \le 0$; then, using Corollary 4.14,

$$J_{\tau_{i+1}}(\xi_{i+1}) - J_{\tau_i}(\xi_i) \le \big(\alpha\beta^{\mu_{\tau_i}(\xi_i)} - \bar{\alpha}\bar{\beta}^{\nu_{\tau_i}(\xi_i,N_i+\eta)}\big)\theta(\xi_i) \le -\omega\alpha\beta^{M^*_{\delta'}}\delta' \tag{4.25}$$

for each $i \ge i_1$, where $\delta' = \Lambda 2^{-\chi N^*}$. But this implies that $J_{\tau_i}(\xi_i) \to -\infty$ as $i \to \infty$, which is a contradiction since $J_{\tau_i}(\xi_i)$ is lower bounded for all $i \in \mathbb{N}$ by Lemma 4.1.

The argument is identical when the sequence is perpetually infeasible. □

We now prove that if Algorithm 3.1 finds a feasible point, then every point generated afterward remains feasible.

Lemma 4.18. *Let $\{N_i\}_{i\in\mathbb{N}}$, $\{\tau_i\}_{i\in\mathbb{N}}$, and $\{\xi_i\}_{i\in\mathbb{N}}$ be the sequences generated by Algorithm* 3.1. *Then there exists $i_0 \in \mathbb{N}$ such that if $\Psi_{\tau_{i_0}}(\xi_{i_0}) \le 0$, then $\Psi(\xi_i) \le 0$ and $\Psi_{\tau_i}(\xi_i) \le 0$ for each $i \ge i_0$.*

Proof. Let $\mathcal{I} \subset \mathbb{N}$ be a subsequence defined by

$$\mathcal{I} = \left\{ i \in \mathbb{N} \mid \theta_{\tau_i}(\xi_i) \le -\frac{\Lambda}{2^{\chi N_i}} \text{ and } \nu_{\tau_i}(\xi_i, N_i+\eta) < \infty \right\}. \tag{4.26}$$

Note that by definition of Algorithm 3.1, $\Psi(\xi_{i+1}) = \Psi(\xi_i)$ for each $i \notin \mathcal{I}$. For each $i \in \mathcal{I}$ such that $\Psi_{\tau_i}(\xi_i) \le 0$, by definition of ν_τ together with Corollary 4.14, there exists $C > 0$ such that

$$\Psi_{\tau_{i+1}}(\xi_{i+1}) \le \big(\alpha\beta^{\mu_{\tau_i}(\xi_i)} - \bar{\alpha}\bar{\beta}^{\nu_{\tau_i}(\xi_i,N_i+\eta)}\big)\theta_{\tau_i}(\xi_i) \le -\omega\alpha\beta C\left(\frac{\Lambda}{2^{\chi N_i}}\right)^2. \tag{4.27}$$

By Lemma 4.3 and the fact that $N_{i+1} \ge N_i$, there exists $C' > 0$ such that

$$\Psi(\xi_{i+1}) \le \frac{C'}{2^{N_i}} - \omega\alpha\beta C\left(\frac{\Lambda}{2^{\chi N_i}}\right)^2 \le \frac{1}{2^{2\chi N_i}}\left(\frac{C'}{2^{(1-2\chi)N_i}} - \omega\alpha\beta C\Lambda^2\right). \tag{4.28}$$

Hence, if $\Psi_{\tau_{i_1}}(\xi_{i_1}) \le 0$ for $i_1 \in \mathbb{N}$ such that N_{i_1} is large enough, then $\Psi(\xi_i) \le 0$ for each $i \ge i_1$. Moreover, by (4.28), for each $N \ge N_i$ and each $\tau \in \mathcal{T}_N$,

$$\Psi_\tau(\xi_{i+1}) \le \frac{1}{2^{2\chi N_i}}\left(\frac{C'}{2^{(1-2\chi)N_i}} - \omega\alpha\beta C\Lambda^2\right) + \frac{C'}{2^N} \le \frac{1}{2^{2\chi N_i}}\left(\frac{2C'}{2^{(1-2\chi)N_i}} - \omega\alpha\beta C\Lambda^2\right). \tag{4.29}$$

Thus, if $\Psi_{\tau_{i_2}}(\xi_{i_2}) \leq 0$ for $i_2 \in \mathbb{N}$ such that N_{i_2} is large enough, then $\Psi_\tau(\xi_{i_2}) \leq 0$ for each $\tau \in \mathcal{T}_N$ such that $N \geq N_{i_2}$. But note that this is exactly the case when there is an uninterrupted finite subsequence of the form $i_2 + k \notin \mathcal{I}$ with $k \in \{1, \ldots, n'\}$; thus we can conclude that $\Psi_{\tau_{i_2+k}}(\xi_{i_2+k}) \leq 0$. Also note that the case of $i \in \mathcal{I}$ is trivially satisfied by the definition of ν_τ. Setting $i_0 = \max\{i_1, i_2\}$, we get the desired result. $\square$

Next, we show that the points produced by our algorithm asymptotically satisfy our discretized optimality condition.

LEMMA 4.19. *Let* $\{N_i\}_{i\in\mathbb{N}}$, $\{\tau_i\}_{i\in\mathbb{N}}$, *and* $\{\xi_i\}_{i\in\mathbb{N}}$ *be the sequences generated by Algorithm* 3.1. *Then* $\lim_{i\to\infty} \theta_{\tau_i}(\xi_i) = 0$.

Proof. Let us suppose that $\lim_{i\to\infty} \theta_{\tau_i}(\xi_i) \neq 0$. Then there exists $\delta > 0$ such that $\liminf_{i\to\infty} \theta_{\tau_i}(\xi_i) < -2\delta$. Hence, using Lemma 4.17, there exists an infinite subsequence $\mathcal{K} \subset \mathbb{N}$ defined by $\mathcal{K} = \{i \in \mathbb{N} \mid \theta_{\tau_i}(\xi_i) < -2\delta\}$. Let $\mathcal{I} \subset \mathbb{N}$ be as in (4.26). Note that by Lemma 4.16 we get that $\mathcal{K} \cap \mathcal{I}$ is an infinite set.

We analyze Algorithm 3.1 by considering the behavior of each step as a function of its membership to each subsequence. First, for each $i \notin \mathcal{I}$, $\xi_{i+1} = \xi_i$, and thus $J(\xi_{i+1}) = J(\xi_i)$ and $\Psi(\xi_{i+1}) = \Psi(\xi_i)$. Second, let $i \in \mathcal{I}$ such that $\Psi_{\tau_i}(\xi_i) \leq 0$; then arguing just as we did in (4.27), (4.28), and (4.29), we have that

$$(4.30)\quad J(\xi_{i+1}) - J(\xi_i) \leq \frac{2C'}{2^{N_i}} - \omega\alpha\beta C \left(\frac{\Lambda}{2^{\chi N_i}}\right)^2 \leq \frac{1}{2^{2\chi N_i}} \left(\frac{2C'}{2^{(1-2\chi)N_i}} - \omega\alpha\beta C \Lambda^2\right).$$

Since $\chi \in \left(0, \frac{1}{2}\right)$, we get that for N_i large enough $J(\xi_{i+1}) \leq J(\xi_i)$. Using a similar argument, for N_i large enough, $\Psi(\xi_{i+1}) \leq \Psi(\xi_i)$. Third, let $i \in \mathcal{K} \cap \mathcal{I}$ such that $\Psi_{\tau_i}(\xi_i) \leq 0$. Then, by Lemma 4.13,

$$(4.31)\qquad J_{\tau_{i+1}}(\xi_{i+1}) - J_{\tau_i}(\xi_i) \leq \left(\alpha\beta^{\mu_{\tau_i}(\xi_i)} - \bar{\alpha}\bar{\beta}^{\nu_{\tau_i}(\xi_i, N_i+\eta)}\right)\theta_{\tau_i}(\xi_i) \leq -2\omega\alpha\beta^{M^*_{2\delta}}\delta;$$

thus, by Lemmas 4.3 and 4.17, for N_i large enough, $J(\xi_{i+1}) - J(\xi_i) \leq -\omega\alpha\beta^{M^*_{2\delta}}\delta$. Similarly, if $\Psi_{\tau_i}(\xi_i) > 0$, using the same argument, for N_i large enough, $\Psi(\xi_{i+1}) - \Psi(\xi_i) \leq -\omega\alpha\beta^{M^*_{2\delta}}\delta$.

Now let us assume that there exists $i_0 \in \mathbb{N}$ such that N_{i_0} is large enough and $\Psi_{\tau_{i_0}}(\xi_{i_0}) \leq 0$. Then by Lemma 4.18 we get that $\Psi_{\tau_i}(\xi_i) \leq 0$ for each $i \geq i_0$. But as shown above, either $i \notin \mathcal{K} \cap \mathcal{I}$ and $J(\xi_{i+1}) \leq J(\xi_i)$ or $i \in \mathcal{K} \cap \mathcal{I}$ and $J(\xi_{i+1}) - J(\xi_i) \leq -\omega\alpha\beta^{M^*_{2\delta}}\delta$, and since $\mathcal{K} \cap \mathcal{I}$ is an infinite set we get that $J(\xi_i) \to -\infty$ as $i \to \infty$, which is a contradiction as J is lower bounded by Lemma 4.1.

The same argument is valid for the case when $\Psi_{\tau_i}(\xi_i) > 0$ for each $i \in \mathbb{N}$, now using the bounds for Ψ. Together, both contradictions imply that $\theta_{\tau_i}(\xi_i) \to 0$ as $i \to \infty$, as desired. $\square$

In conclusion, we prove that the sequence of points generated by Algorithm 3.1 asymptotically satisfies the optimality condition defined in [14], which follows since P_τ is a consistent approximation of P_p, shown in Theorem 4.12, and Theorem 3.4.

THEOREM 4.20. *Let* $\{N_i\}_{i\in\mathbb{N}}$, $\{\tau_i\}_{i\in\mathbb{N}}$, *and* $\{\xi_i\}_{i\in\mathbb{N}}$ *be the sequences generated by Algorithm* 3.1*; then* $\lim_{i\to\infty} \theta(\xi_i) = 0$.

5. Examples.

In this section, we apply Algorithm 3.1 to calculate an optimal control for four examples. Before describing each example, we begin by describing the numerical implementation of Algorithm 3.1. First, observe that the time interval for the trajectory of the system can be generalized to any interval $[t_0, t_f]$, instead of $[0, 1]$ as used in the previous sections. Second, we employ the LSSOL [6] numerical solver, using TOMLAB [9] as an interface, in order to compute θ_τ at each iteration.

Third, for each example we employ a stopping criterion that terminates Algorithm 3.1 whenever θ_τ gets too close to zero. Next, for the sake of comparison, we solved every example using Algorithm 3.1 and the mixed integer program (MIP) described in [4], which combines branch and bound steps with sequential quadratic programming steps. Finally, all our experiments are performed on an Intel Xeon, 12 core, 3.47 GHz, 92 GB RAM machine.

5.1. Constrained switched linear quadratic regulator. Linear quadratic regulator (LQR) examples have been used to illustrate the utility of many optimal control algorithms [3, 15]. We consider an LQR system with three continuous states, three discrete modes, and a single continuous input. The dynamics in each mode are as described in Table 5.1, where

$$A = \begin{bmatrix} 1.0979 & -0.0105 & 0.0167 \\ -0.0105 & 1.0481 & 0.0825 \\ 0.0167 & 0.0825 & 1.1540 \end{bmatrix}. \tag{5.1}$$

The system matrix is purposefully chosen to have three unstable eigenvalues and the control matrix in each mode is only able to control along a single continuous state. Hence, while the system and control matrix in each mode are not a stabilizable pair, the system and all the control matrices taken together simultaneously are stabilizable and are expected to appropriately switch between the modes to reduce the cost. The objective of the optimization is to have the trajectory of the system at time t_f be at $(1, 1, 1)$ while minimizing the energy of the input, as described in Table 5.2.

Algorithm 3.1 and the MIP are initialized at $x_0 = (0, 0, 0)$, with continuous and discrete inputs as described in Table 5.3, using 16 equally spaced samples in time. Algorithm 3.1 took 11 iterations, ended with 48 time samples, and terminated when $\theta_{\tau_i}(\xi_i) > -10^{-2}$. The result of both optimization procedures is illustrated in Figure 5.1. The computation time and final cost of both algorithms can be found in Table 5.3. Notice that Algorithm 3.1 is able to compute a lower cost than MIP and is able to do it more than 75 times faster.

5.2. Double-tank system. To illustrate the performance of Algorithm 3.1 when there is no continuous input present, we consider a double-tank example. The

TABLE 5.1
Vector fields of each of the modes of the examples considered in section 5.

Example	Mode 1	Mode 2	Mode 3
LQR	$\dot{x}(t) = Ax(t) + \begin{bmatrix} 0.9801 \\ -0.1987 \\ 0 \end{bmatrix} u(t)$	$\dot{x}(t) = Ax(t) + \begin{bmatrix} 0.1743 \\ 0.8601 \\ -0.4794 \end{bmatrix} u(t)$	$\dot{x}(t) = Ax(t) + \begin{bmatrix} 0.0952 \\ 0.4699 \\ 0.8776 \end{bmatrix} u(t)$
Tank	$\dot{x}(t) = \begin{bmatrix} 1 - \sqrt{x_1(t)} \\ \sqrt{x_1(t)} - \sqrt{x_2(t)} \end{bmatrix}$	$\dot{x}(t) = \begin{bmatrix} 2 - \sqrt{x_1(t)} \\ \sqrt{x_1(t)} - \sqrt{x_2(t)} \end{bmatrix}$	N/A
Quadrotor	$\ddot{x}(t) = \begin{bmatrix} \frac{\sin x_3(t)}{M}\,(u(t) + Mg) \\ \frac{\cos x_3(t)}{M}\,(u(t) + Mg) - g \\ 0 \end{bmatrix}$	$\ddot{x}(t) = \begin{bmatrix} g \sin x_3(t) \\ g \cos x_3(t) - g \\ \frac{-Lu(t)}{I} \end{bmatrix}$	$\ddot{x}(t) = \begin{bmatrix} g \sin x_3(t) \\ g \cos x_3(t) - g \\ \frac{Lu(t)}{I} \end{bmatrix}$
Needle	$\dot{x}(t) = \begin{bmatrix} \sin(x_5(t))u_1(t) \\ -\cos(x_5(t))\sin(x_4(t))u_1(t) \\ \cos(x_4(t))\cos(x_5(t))u_1(t) \\ \kappa\cos(x_6(t))\sec(x_5(t))u_1(t) \\ \kappa\sin(x_6(t))u_1(t) \\ -\kappa\cos(x_6(t))\tan(x_5(t))u_1(t) \end{bmatrix}$	$\dot{x}(t) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ u_2(t) \end{bmatrix}$	N/A

Table 5.2

The algorithmic parameters and cost function used for each of the examples during the implementation of Algorithm 3.1.

Example	$L\big(x(t), u(t), t\big)$	$\phi\big(x^{(\xi)}(t_f)\big)$	U	γ	α	β	$\bar{\alpha}$	$\bar{\beta}$	Λ	χ	ω	t_0	t_f
LQR	$0.01 \cdot u(t)^2$	$\left\lVert\begin{bmatrix} x_1(t_f) - 1 \\ x_2(t_f) - 1 \\ x_3(t_f) - 1 \end{bmatrix}\right\rVert_2^2$	$u(t) \in [-20, 20]$	1	0.1	0.87	0.005	0.72	10^{-4}	$\frac{1}{4}$	10^{-6}	0	2
Tank	$2 \cdot \big(x_2(t) - 3\big)^2$	0	N/A	100	0.01	0.75	0.005	0.72	10^{-4}	$\frac{1}{4}$	10^{-6}	0	10
Quadrotor	$5 \cdot u(t)^2$	$\left\lVert\begin{bmatrix} \sqrt{5} \cdot \big(x_1(t_f) - 6\big) \\ \sqrt{5} \cdot \big(x_2(t_f) - 1\big) \\ \sin\big(0.5 \cdot x_3(t_f)\big) \end{bmatrix}\right\rVert_2^2$	$u(t) \in \big[0, 10^{-3}\big]$	10	0.01	0.80	5×10^{-4}	0.72	10^{-4}	$\frac{1}{4}$	10^{-6}	0	7.5
Needle	$0.01 \cdot \left\lVert\begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}\right\rVert_2^2$	$\left\lVert\begin{bmatrix} x_1(t_f) + 2 \\ x_2(t_f) - 3.5 \\ x_3(t_f) - 10 \end{bmatrix}\right\rVert_2^2$	$u_1(t) \in [0, 5]$ $u_2(t) \in [\frac{-\pi}{2}, \frac{\pi}{2}]$	100	0.002	0.72	0.001	0.71	10^{-4}	$\frac{1}{4}$	0.05	0	8

Table 5.3

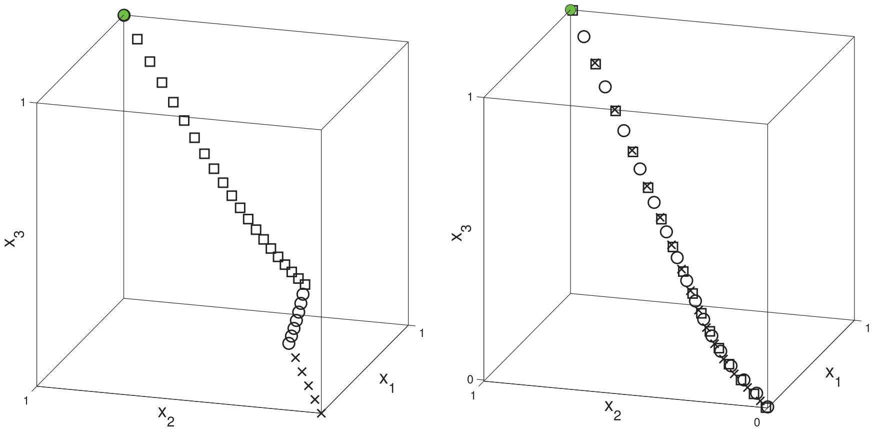
The initialization parameters used for each of the examples, and the computation time, of Algorithm 3.1 *and the MIP described in* [4].

Example	Initial continuous input, $\forall t \in [t_0, t_f]$	Initial discrete input, $\forall t \in [t_0, t_f]$	Algorithm 3.1 computation time	Algorithm 3.1 final cost	MIP computation time	MIP final cost
LQR	$u(t) = 0$	$d(t) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$	9.827[s]	1.23×10^{-3}	753.0[s]	1.89×10^{-3}
Tank	N/A	$d(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$	$32.38[s]$	4.829	$119700[s]$	4.828
Quadrotor	$u(t) = 5 \times 10^{-4}$	$d(t) = \begin{bmatrix} 0.33 \\ 0.34 \\ 0.33 \end{bmatrix}$	8.350[s]	0.128	2783[s]	0.165
Needle	$u(t) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$d(t) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$	62.76[s]	0.302	did not converge	did not converge

two states of the system correspond to the fluid levels of an upper and a lower tank. The output of the upper tank flows into the lower tank, the output of the lower tank exits the system, and the flow into the upper tank is restricted to either 1[lt/s] or 2[lt/s]. The dynamics in each mode are then derived using Torricelli's law, as shown in Table 5.1. The objective of the optimization is to have the fluid level in the lower tank equal to 3[m], as reflected in the cost function in Table 5.2.

Algorithm 3.1 and the MIP are initialized at $x_0 = (0, 0)$ with a discrete input described in Table 5.3 and 128 equally spaced samples in time. Algorithm 3.1 took 67 iterations, ended with 256 time samples, and terminated when $\theta_{\tau_i}(\xi_i) > -10^{-2}$. The result of both optimization procedures is illustrated in Figure 5.2. The computation time and final cost of both algorithms can be found in Table 5.3. Notice that Algorithm 3.1 is able to compute a comparable cost to the MIP and is able to do it nearly 3700 times faster.

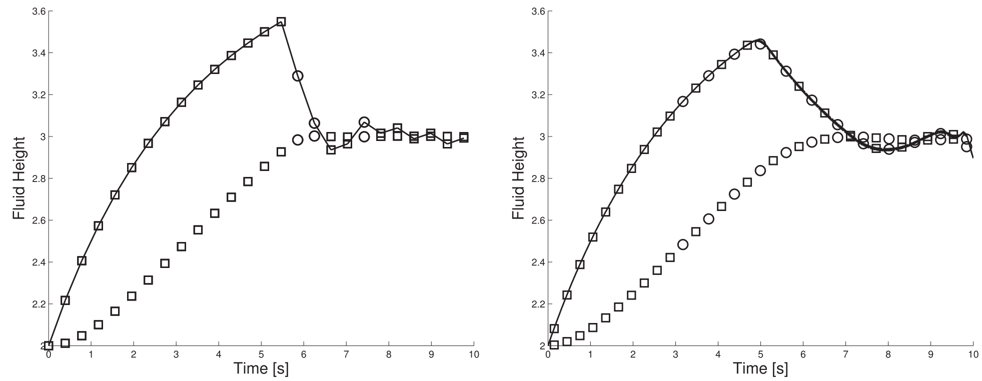
5.3. Quadrotor helicopter control. Next, we consider the optimal control of a quadrotor helicopter in two dimensions using a model described in [7]. The evolution of the quadrotor can be defined with respect to a fixed two dimensional reference frame using six dimensions, where the first three dimensions represent the position along a horizontal axis, the position along the vertical axis, and the roll angle of the helicopter, respectively, and the last three dimensions represent the time derivative of the first three dimensions. We model the dynamics as a three-mode switched system (the first mode describes the dynamics of going up, the second mode describes the dynamics of moving to the left, and the third mode describes the dynamics of moving to the right) with a single input as described in Table 5.1, where $L = 0.3050$ meters, $M = 1.3000$ kilograms, $I = 0.0605$ kilogram meters squared, and $g = 9.8000$ meters



(a) MIP Final Result

(b) Algorithm 3.1 Final Result

FIG. 5.1. *Optimal trajectories for Algorithm* 3.1 *and the MIP, where the point* $(1, 1, 1)$ *is shown as a filled circle, and the trajectory is shown in circles when in mode* 1*, in squares when in mode* 2*, and in crosses when in mode* 3.



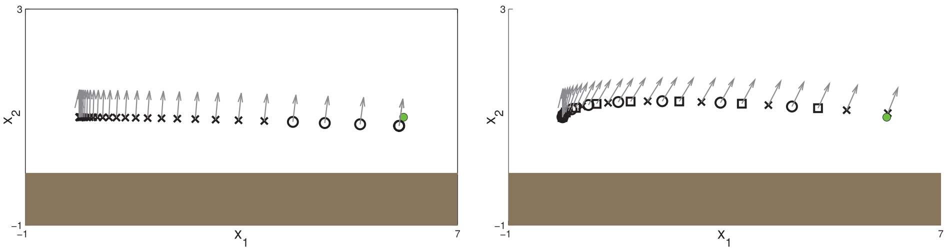
(a) MIP Final Result

(b) Algorithm 3.1 Final Result

FIG. 5.2. *Optimal trajectories for Algorithm* 3.1 *and the MIP, where* $x_1(t)$ *is connected with a line and* $x_2(t)$ *is not connected with a line and where each state trajectory is shown in circles when in mode* 1 *and in squares when in mode* 2.

per second squared. The objective of the optimization is to have the trajectory of the system at time t_f be at position $(6, 1)$ with a zero roll angle while minimizing the input required to achieve this task. This objective is reflected in the chosen cost function which is described in Table 5.2. A state constraint is added to the optimization to ensure that the quadrotor remains above ground.

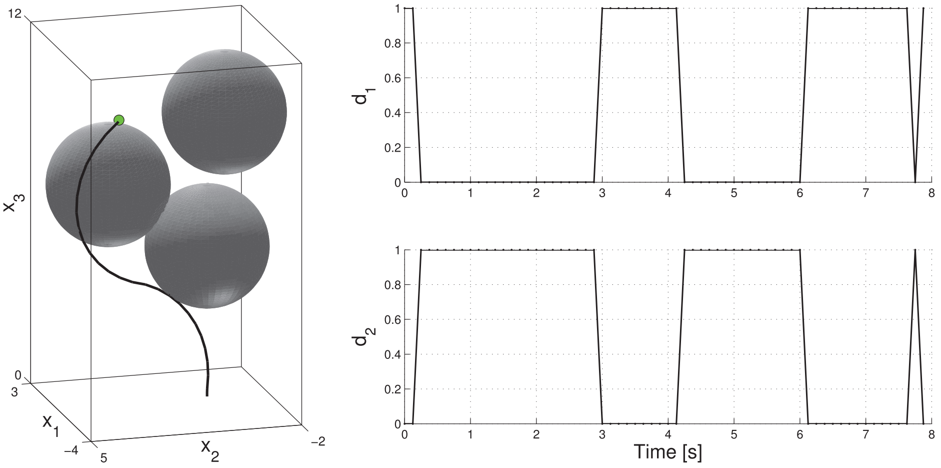
Algorithm 3.1 and the MIP are initialized at position $(0, 1)$ with a zero roll angle, with zero velocity, with continuous and discrete inputs as described in Table 5.3, and with 64 equally spaced samples in time. Algorithm 3.1 took 31 iterations, ended with 192 time samples, and terminated after the optimality condition was bigger than -10^{-4}. The result of both optimization procedures is illustrated in Figure 5.3. The computation time and final cost of both algorithms can be found in Table 5.3. Notice that Algorithm 3.1 is able to compute a lower-cost continuous and discrete input when compared to the MIP and is able to do it more than 333 times faster.



(a) MIP Final Result

(b) Algorithm 3.1 Final Result

Fig. 5.3. *Optimal trajectories for Algorithm* 3.1 *and the MIP, where the point* $(6, 1)$ *is shown as a filled circle, the trajectory is shown in circles when in mode* 1*, in squares when in mode* 2*, and in crosses when in mode* 3*, and where the normal direction to the quadrotor is shown with an arrow.*



(a) Algorithm 3.1 Final Result

(b) Algorithm 3.1 Final Discrete Input

Fig. 5.4. *The optimal trajectory and discrete inputs generated by Algorithm* 3.1*, where the point* $(-2, 3.5, 10)$ *is shown as a filled circle and obstacles are shown as spheres.*

5.4. Bevel-tip flexible needle. Bevel-tip flexible needles are asymmetric needles that move along curved trajectories when a forward pushing force is applied. The three dimensional dynamics of such needles has been described in [10] and heuristics for the path planning in the presence of obstacles have been considered in [2]. The evolution of the needle can be defined using six continuous states, where the first three states represent the position of the needle relative to the point of entry, and the last three states represent the yaw, pitch, and roll of the needle relative to the plane, respectively. As suggested by [2], the dynamics of the needle are naturally modeled as a two-mode switched system: one mode when the needle is being pushed and another mode when the needle is turning. These modes are described in Table 5.1 with two continuous inputs: u_1 representing the insertion speed and u_2 representing the rotation speed of the needle and where κ is the curvature of the needle and is equal to 0.22[1/cm]. The objective of the optimization is to have the trajectory of the system at time t_f be at position $(-2, 3.5, 10)$ while minimizing the energy of the input. This objective is reflected in the cost function described in Table 5.2. Constraints are added to the optimization to ensure that the needle remains outside three spherical obstacles centered at $(0, 0, 5)$, $(1, 3, 7)$, and $(-2, 0, 10)$, all with radius 2.

Algorithm 3.1 and the MIP are initialized at position $(0, 0, 0)$ with inputs as described in Table 5.3 with 64 equally spaced samples in time. Algorithm 3.1 took 103 iterations, ended with 64 time samples, and terminated when $\theta_{\tau_i}(\xi_i) > -10^{-3}$. The computation time and final cost of both algorithms can be found in Table 5.3, and it is important to note that the MIP was unable to find any solution. The result of Algorithm 3.1 is illustrated in Figure 5.4.

Acknowledgments. We are grateful to Maryam Kamgarpour and Claire Tomlin, who helped write our original papers on switched system optimal control. We are also grateful to Magnus Egerstedt and Yorai Wardi, whose comments on our original papers were critical to our realizing the deficiencies of the sufficient descent argument. The chattering lemma argument owes its origin to a series of discussions with Ray DeCarlo. The contents of this article reflect a longstanding collaboration with Elijah Polak.

REFERENCES

[1] L. Armijo, *Minimization of functions having Lipschitz continuous first partial derivatives*, Pacific J. Math., 16 (1966), pp. 1–3.

[2] V. Duindam, R. Alterovitz, S. S. Sastry, and K. Goldberg, *Screw-based motion planning for bevel-tip flexible needles in 3D environments with obstacles*, in Proceedings of the 2008 IEEE International Conference on Robotics and Automation, 2008, pp. 2483–2488.

[3] M. Egerstedt, Y. Y. Wardi, and H. Axelsson, *Transition-time optimization for switched-mode dynamical systems*, IEEE Trans. Automat. Control, 51 (2006), pp. 110–115.

[4] R. Fletcher and S. Leyffer, *Solving mixed integer nonlinear programs by outer approximation*, Math. Program., 66 (1994), pp. 327–349.

[5] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed., John Wiley, New York, 1999.

[6] P. E. Gill, S. J. Hammarling, W. Murray, M. A. Saunders, and M. H. Wright, *User's Guide for LSSOL (Version 1.0): A FORTRAN Package for Constrained Linear Least-Squares and Convex Quadratic Programming*, Tech. report, Systems Optimization Laboratory, Stanford University, Stanford, CA, 1986.

[7] J. H. Gillula, G. M. Hoffmann, H. Huang, M. P. Vitus, and C. J. Tomlin, *Applications of hybrid reachability analysis to robotic aerial vehicles*, Internat. J. Robotics Res., 30 (2011), pp. 335–354.

[8] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.

[9] K. Holmström, *The TOMLAB Optimization Environment in Matlab*, Adv. Model. Optim., 1 (1999), pp. 47–69.

[10] V. Kallem and N. J. Cowan, *Image-guided control of flexible bevel-tip needles*, in Proceedings of the 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 3015–3020.

[11] P. A. Parrilo, *Semidefinite programming relaxations for semialgebraic problems*, Math. Program., 96 (2003), pp. 293–320.

[12] E. Polak, *Optimization: Algorithms and Consistent Approximation*, Springer, New York, 1997.

[13] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, New York, 1998.

[14] R. Vasudevan, H. Gonzalez, R. Bajcsy, and S. S. Sastry, *Consistent approximations for the optimal control of constrained switched systems—Part 1: A conceptual algorithm*, SIAM J. Control Optim., 51 (2013), pp. 4463–4483.

[15] X. Xu and P. J. Antsaklis, *Optimal control of switched autonomous systems*, in Proceedings of the 41st IEEE Conference on Decision and Control, 2002, pp. 4401–4406.