# FINITE ELEMENT APPROXIMATION OF VARIATIONAL INEQUALITIES IN OPTIMAL CONTROL

KARIN KRAFT AND STIG LARSSON

ABSTRACT. The optimal control of a linear system of ordinary differential equations with quadratic goal functional and inequality constraints is considered. Lagrange's method in the calculus of variations leads to necessary conditions for optimality in the form of a system of variational inequalities, which are discretized by a finite element method. An *a posteriori* error estimate based on an augmented Lagrangian is derived by means of the methodology of dual weighted residuals. An adaptive algorithm is implemented and tested on model problems.

## 1. INTRODUCTION

We consider the optimal control of systems whose state equation is a linear system of ordinary differential equations together with boundary conditions at the initial and final times. The goal functional is assumed to be quadratic and we allow inequality constraints on the state and control variables. We call such problems quadratic/linear with inequality constraints. After sufficient regularization we apply Lagrange's method in the calculus of variations, which leads to necessary conditions for optimality in the form of a system of variational inequalities. Our goal is to develop and analyze an adaptive finite element method for the numerical solution of such problems.

The use of finite elements for spatial discretization of optimization and optimal control problems for partial differential equations goes back to [17] and was more recently developed together with *a posteriori* error analysis in [1, 16] and many other works. The use of finite element methods for temporal discretization of optimal control problems is less common [6, 7, 5, 12, 10, 11].

The dual weighted residuals approach [1] for a posteriori error analysis is very suitable for optimal control problems because it is formulated using the Lagrangian framework. In our previous work [10, 11] we exploited this idea for quadratic/linear and nonlinear problems without inequality constraints. However, inequality constraints on state and control variables are important

1991 *Mathematics Subject Classification.* 65L60, 49K15.
*Key words and phrases.* finite element, a posteriori, error estimate, variational inequality, adaptive, state constraint, control constraint, vehicle dynamics.
Research supported by the Swedish Research Council (VR) and by the Swedish Foundation for Strategic Research (SSF) through GMMC, the Gothenburg Mathematical Modelling Centre.

in practice. The present work therefore includes inequality constraints but it is limited to quadratic/linear problems.

*A posteriori* error estimation based on dual weighted residuals for variational inequalities for partial differential equations has been studied in [2] and [16]. The main difficulty is that an inequality is solved instead of an equation, so that the residual of the finite element solution is no longer orthogonal to the finite element space, that is, we have no Galerkin orthogonality. This leads only to a one-sided bound for the error. In order to overcome this, we follow the approach of [16], which is based on using an augmented Lagrange functional with additional Lagrange multipliers for the inequality constraints.

The outline of the article is as follows. Section 2 contains the mathematical setting of the optimal control problem. The Lagrange framework is presented and the optimality conditions are derived. The next section, Section 3, includes the discretization of the problem and the derivation of a projected algorithm for solving the resulting algebraic complementarity equations. In the following Section 4 the *a posteriori* error estimate is derived, the computation of the error estimator is discussed, and the adaptive algorithm is briefly described. In Section 5 presents numerical examples: a lane change maneuver from vehicle dynamics and a simple problem displaying Fuller's phenomenon. The last section contains a discussion of the results and presents some directions for future research.

## 2. An optimal control problem

2.1. **Setting of the problem.** We consider a quadratic/linear optimal control problem with inequality constraints on the controls $u(t) \in \mathbb{R}^m$ and states $x(t) \in \mathbb{R}^d$:

$$
\begin{aligned}
\text{Minimize} \qquad & \mathcal{J}(x,u) = \tfrac{1}{2}\|x(0)-\bar{x}_0\|_{Q_0}^2 + \tfrac{1}{2}\|x(T)-\bar{x}_T\|_{Q_T}^2 \\
& \qquad + \tfrac{1}{2}\int_0^T \left(\|x(t)-\bar{x}(t)\|_Q^2 + \|u(t)-\bar{u}(t)\|_R^2\right) \mathrm{d}t, \\
\text{such that} \qquad & \dot{x}(t) = A(t)x(t) + B(t)u(t), \quad \text{for } t \in [0,T], \\
& I_0 x(0) = x_0, \quad I_T x(T) = x_T, \\
& \|u(t)\| \le r_u, \quad \|x(t)\| \le r_x, \quad \text{for } t \in [0,T],
\end{aligned}
$$

where $Q_0$, $Q_T$, $Q(t) \in \mathbb{R}^{d\times d}$ and $R(t) \in \mathbb{R}^{m\times m}$ are symmetric positive definite matrices, $A(t) \in \mathbb{R}^{d\times d}$ and $B(t) \in \mathbb{R}^{m\times d}$. The matrices $I_0$, $I_T \in \mathbb{R}^{d\times d}$ are binary diagonal matrices and $x_0$, $x_T$, $\bar{x}_0$, $\bar{x}_T$, $\bar{x}(t) \in \mathbb{R}^d$, $\bar{u}(t) \in \mathbb{R}^m$, $r_x$, $r_u \in \mathbb{R}$ are given parameters such that $\|x_0\| \le r_x$ and $\|x_T\| \le r_u$. Here $(\cdot,\cdot)$ and $\|\cdot\|$ denote the scalar product and norm in $\mathbb{R}^d$ or $\mathbb{R}^m$, and $\|x\|_S^2 = (x, Sx)$, for any positive definite symmetric matrix $S$. In cases where the matrices $Q_0$, $Q_T$, $Q(t)$, and $R(t)$ are not positive definite in the original problem, a regularization has to be done by adding positive terms to the goal functional. We assume that the given data depend continuously on $t$.

Inequality constraints of the box form $u_{\min} \le u_i(t) \le u_{\max}$, $i = 1, \dots, m$, $t \in [0,T]$, (similarly for $x$) can also be treated by the methods that we consider here, see [16]. We choose to study constraints of the above circular type in order to allow some specific applications from vehicle dynamics, see Section 5, but also to deviate from the presentation in [16].

2.2. **Lagrange framework.** In order to set the optimal control problem in the Lagrange framework we introduce some function spaces. Let $\mathcal{C}^k$ denote $k$ times continuously differentiable functions and $H^1$ denote functions with square integrable derivative. Further, $\mathcal{C}^1_{\mathrm{PW}}$ denotes piecewise continuously differentiable functions $[0,T] \to \mathbb{R}^d$; more precisely, functions that are $\mathcal{C}^1$ except at a finite number of points in $[0,T]$ and with left and right limits $w(t^-) = \lim_{s \downarrow t} w(s)$, $w(t^+) = \lim_{s \uparrow t} w(s)$ for all points $t \in [0,T]$. We introduce the function spaces

$$\begin{aligned}
\mathcal{W} &= \mathbb{R}^d \times \mathcal{C}^1_{\mathrm{PW}}([0,T], \mathbb{R}^d) \times \mathbb{R}^d, \\
\dot{\mathcal{W}} &= R(I_0^{\mathrm{C}}) \times \mathcal{C}^1_{\mathrm{PW}}([0,T], \mathbb{R}^d) \times R(I_T^{\mathrm{C}}) \\
&= \{w \in \mathcal{W} : I_0 w(0^-) = 0,\ I_T w(T^+) = 0\}, \\
\mathcal{U} &= H^1([0,T], \mathbb{R}^m), \\
\mathcal{V} &= H^1([0,T], \mathbb{R}^d).
\end{aligned}$$

Here $R(I_0^{\mathrm{C}})$ and $R(I_T^{\mathrm{C}})$ denote the ranges of the matrices $I_0^{\mathrm{C}} = I - I_0$ and $I_T^{\mathrm{C}} = I - I_T$. The two factors $\mathbb{R}^d$ in $\mathcal{W}$ are used to accomodate the boundary values $w(0^-)$ and $w(T^+)$. The space $\mathcal{W}$ will contain the state variable $x$, which is expected to be smooth. But here we anticipate that $\mathcal{W}$ will also contain its finite element approximation $x_h$, which is only piecewise smooth. Therefore, $\mathcal{W}$ is defined to be a space of piecewise differentiable functions.

These spaces are linear spaces. For some $\hat{x} \in \mathcal{W}$ such that $I_0 \hat{x}(0^-) = x_0$ and $I_T \hat{x}(T^+) = x_T$, we also define the affine space

$$\hat{\mathcal{W}} = \hat{x} + \dot{\mathcal{W}} = \big\{w \in \mathcal{W} : w - \hat{x} \in \dot{\mathcal{W}}\big\}.$$

Thus, functions in $\hat{\mathcal{W}}$ satisfy the prescribed boundary conditions. In order to incorporate the inequality constraints we define the convex sets

$$\begin{aligned}
\mathcal{K}_x &= \big\{w \in \hat{\mathcal{W}} : \|w(t^{\pm})\| \le r_x,\ t \in [0,T]\big\}, \\
\mathcal{K}_u &= \big\{u \in \mathcal{U} : \|u(t)\| \le r_u,\ t \in [0,T]\big\}.
\end{aligned}$$

For ease of notation we write $\mathcal{X} = \mathcal{W} \times \mathcal{U} \times \mathcal{V}$ and $\mathcal{K} = \mathcal{K}_x \times \mathcal{K}_u \times \mathcal{V}$, and note that $\mathcal{K}$ is a convex subset of $\mathcal{X}$.

The functional

$$\mathcal{J} \colon \mathcal{W} \times \mathcal{U} \to \mathbb{R},$$

is given by

$$\begin{aligned}\mathcal{J}(x,u) &= \tfrac{1}{2}\|x(0^-)-\bar{x}_0\|_{Q_0}^2 + \tfrac{1}{2}\|x(T^+)-\bar{x}_T\|_{Q_T}^2 \\ &\quad + \tfrac{1}{2}\int_0^T \left(\|x(t)-\bar{x}(t)\|_Q^2 + \|u(t)-\bar{u}(t)\|_R^2\right)\mathrm{d}t,\end{aligned}$$

and we define the functional

$$\mathcal{F}\colon \mathcal{W}\times\mathcal{U}\times\mathcal{V}\to\mathbb{R},$$

by

$$\begin{aligned}\mathcal{F}(w,u,v) &= \sum_{n=1}^{N}\int_{t_{n-1}}^{t_n}\Big(\dot{w}(t)-A(t)w(t)-B(t)u(t),v(t)\Big)\,\mathrm{d}t \\ &\quad + \sum_{n=0}^{N}([w]_n,v(t_n)) \\ &= \int_0^T\Big((w(t),-\dot{v}(t)-A(t)^{\mathrm{T}}v(t))-(u(t),B(t)^{\mathrm{T}}v(t))\Big)\,\mathrm{d}t \\ &\quad + (w(T^+),v(T))-(w(0^-),v(0)).\end{aligned} \tag{2.1}$$

Here $[w]_n = w(t_n^+) - w(t_n^-)$ and the sum is taken over all points $\{t_n\}_{n=0}^N$ of discontinuity of $w$, and we set $t_0 = 0$, $t_N = T$. The second form of $\mathcal{F}$ is obtained by integration by parts using the fact that functions in $\mathcal{V}$ are continuous.

We seek to minimize $\mathcal{J}(x,u)$ over $\mathcal{K}_x\times\mathcal{K}_u$ subject to $\mathcal{F}(x,u,\varphi)=0$ for all $\varphi\in\mathcal{V}$. We define the Lagrange functional

$$\mathcal{L}\colon \mathcal{X}=\mathcal{W}\times\mathcal{U}\times\mathcal{V}\to\mathbb{R},$$

by

$$\mathcal{L}(x,u,z) = \mathcal{J}(x,u)+\mathcal{F}(x,u,z), \tag{2.2}$$

and assume that there is a minimizer $\xi=(x,u,z)\in\mathcal{K}$ of the Lagrangian $\mathcal{L}$. Since $\mathcal{K}$ is convex, we have for $\eta\in\mathcal{K}$ and $s\in[0,1]$,

$$0\le \mathcal{L}(\xi+s(\eta-\xi))-\mathcal{L}(\xi) = s\mathcal{L}'(\xi,\eta-\xi)+o(s),\quad \text{as } s\to 0^+.$$

Here and below the notation $\mathcal{L}'(\xi,\varphi)$ denotes the derivative of $\mathcal{L}$ at $\xi$ acting linearly on a vector $\varphi$. By dividing by $s$ and letting $s\to 0^+$ we obtain the necessary condition for a minimum $\xi\in\mathcal{K}$:

$$\mathcal{L}'(\xi,\eta-\xi)\ge 0\quad\forall\eta\in\mathcal{K}. \tag{2.3}$$

Expressed in terms of the partial derivatives and $\eta=(\varphi_x,\varphi_u,\varphi_z)\in\mathcal{K}$ this means that we need to find $(x,u,z)\in\mathcal{K}$ such that

$$\mathcal{L}_x'(x,u,z,\varphi_x-x)\ge 0\quad\forall\varphi_x\in\mathcal{K}_x, \tag{2.4a}$$

$$\mathcal{L}_u'(x,u,z,\varphi_u-u)\ge 0\quad\forall\varphi_u\in\mathcal{K}_u, \tag{2.4b}$$

$$\mathcal{L}_z'(x,u,z,\varphi_z)=0\quad\forall\varphi_z\in\mathcal{V}. \tag{2.4c}$$

These are variational inequalities; the last one reduces to an equation because $\varphi_z$ ranges over a full vector space. Computing the derivatives, we obtain first

$$\mathcal{L}'_x(x,u,z,\varphi_x) = \mathcal{J}'_x(x,u,\varphi_x) + \mathcal{F}'_x(x,u,z,\varphi_x),$$

where

$$\begin{aligned}\mathcal{J}'_x(x,u,\varphi_x) = \int_0^T &(Q(x-\bar{x}),\varphi_x)\,\mathrm{d}t \\ &+ (Q_0(x(0^-)-\bar{x}_0),\varphi_x(0^-)) \\ &+ (Q_T(x(T^+)-\bar{x}_T),\varphi_x(T^+)),\end{aligned}$$

and where $\mathcal{F}'_x(x,u,z,\varphi_x) = \mathcal{F}(\varphi_x,0,z)$ is given by the second form of (2.1). Computing $\mathcal{L}'_u(x,u,z,\varphi_u) = \mathcal{J}'_u(x,u,z,\varphi_u) + \mathcal{F}(0,\varphi_u,z)$ using the second form of $\mathcal{F}$ and $\mathcal{L}'_z(x,u,z,\varphi_z) = \mathcal{F}(x,u,\varphi_z)$ with the first form of $\mathcal{F}$ and assuming that $x$ is smooth, we obtain the following equations and inequalities

$$\begin{aligned}
&\text{(2.5a)} \quad \int_0^T (-\dot{z} - A^{\mathrm{T}}z + Q(x-\bar{x}),\varphi_x - x)\,\mathrm{d}t \\
&\qquad + (Q_0(x(0^-)-\bar{x}_0) - z(0),\varphi_x(0^-) - x(0^-)) \\
&\qquad + (Q_T(x(T^+)-\bar{x}_T) + z(T),\varphi_x(T^+) - x(T^+)) \geq 0 \qquad \forall \varphi_x \in \mathcal{K}_x, \\
&\text{(2.5b)} \quad \int_0^T (R(u-\bar{u}) - B^{\mathrm{T}}z,\varphi_u - u)\,\mathrm{d}t \geq 0 \qquad \forall \varphi_u \in \mathcal{K}_u, \\
&\text{(2.5c)} \quad \int_0^T (\dot{x} - Ax - Bu,\varphi_z)\,\mathrm{d}t = 0 \qquad \forall \varphi_z \in \mathcal{V}, \\
&\text{(2.5d)} \quad I_0 x(0^-) = x_0, \quad I_T x(T^-) = x_T.
\end{aligned}$$

We note that (2.5c) and (2.5d) are the equations of state and that (2.5a) is based on the corresponding adjoint.

## 3. The Discrete optimal control problem

3.1. **The finite element method.** In order to discretize the necessary conditions for optimality (2.4) we define a mesh $0 = t_0 < t_1 < t_2 < \ldots < t_N = T$, with steps $h_n = t_n - t_{n-1}$ and intervals $J_n = (t_{n-1}, t_n)$. With $P^k$ denoting polynomials of degree $k$, we introduce the function spaces

$$\begin{aligned}
\mathcal{W}_h &= \mathbb{R}^d \times \Big\{ w \in \mathcal{W} : w|_{J_n} \in P^0(J_n,\mathbb{R}^d),\ n = 1,\ldots,N \Big\} \times \mathbb{R}^d, \\
\dot{\mathcal{W}}_h &= R(I_0^{\mathrm{C}}) \times \Big\{ w \in \mathcal{W} : w|_{J_n} \in P^0(J_n,\mathbb{R}^d),\ n = 1,\ldots,N \Big\} \times R(I_T^{\mathrm{C}}) \\
&= \Big\{ w \in \mathcal{W}_h : I_0 w(0^-) = 0,\ I_T w(T^+) = 0 \Big\}, \\
\mathcal{U}_h &= \Big\{ u \in \mathcal{C}^0([0,T],\mathbb{R}^m) : u|_{J_n} \in P^1(J_n,\mathbb{R}^m),\ n = 1,\ldots,N \Big\}, \\
\mathcal{V}_h &= \Big\{ v \in \mathcal{C}^0([0,T],\mathbb{R}^d) : v|_{J_n} \in P^1(J_n,\mathbb{R}^d),\ n = 1,\ldots,N \Big\}.
\end{aligned}$$

In the definition $\hat{\mathcal{W}} = \hat{x} + \dot{\mathcal{W}}$ we may choose $\hat{x} \in \mathcal{W}_h$ and define

$$\hat{\mathcal{W}}_h = \hat{x} + \dot{\mathcal{W}}_h.$$

Then we have $\mathcal{W}_h \subset \mathcal{W}$, $\dot{\mathcal{W}}_h \subset \dot{\mathcal{W}}$, $\hat{\mathcal{W}}_h \subset \hat{\mathcal{W}}$, $\mathcal{U}_h \subset \mathcal{U}$, and $\mathcal{V}_h \subset \mathcal{V}$. Moreover, we define the convex sets

$$\mathcal{K}_{x,h} = \big\{x_h \in \hat{\mathcal{W}}_h : \|x_h(t_n^{\pm})\| \le r_x,\ n = 0, \dots, N\big\},$$
$$\mathcal{K}_{u,h} = \big\{u_h \in \mathcal{U}_h : \|u_h(t_n)\| \le r_u,\ n = 0, \dots, N\big\},$$

and

$$\mathcal{K}_h = \mathcal{K}_{x,h} \times \mathcal{K}_{u,h} \times \mathcal{V}_h.$$

It follows that $\mathcal{K}_{x,h} \subset \mathcal{K}_x$, $\mathcal{K}_{u,h} \subset \mathcal{K}_u$, and $\mathcal{K}_h \subset \mathcal{K}$. We formulate the finite element approximation of (2.4): Find $(x_h, u_h, z_h) \in \mathcal{K}_h$ such that

$$\mathcal{L}'_x(x_h, u_h, z_h, \varphi_{x,h} - x_h) \ge 0 \quad \forall \varphi_{x,h} \in \mathcal{K}_{x,h}, \tag{3.1a}$$
$$\mathcal{L}'_u(x_h, u_h, z_h, \varphi_{u,h} - u_h) \ge 0 \quad \forall \varphi_{u,h} \in \mathcal{K}_{u,h}, \tag{3.1b}$$
$$\mathcal{L}'_z(x_h, u_h, \varphi_{z,h}) = 0 \quad \forall \varphi_{z,h} \in \mathcal{V}_h. \tag{3.1c}$$

Note that $x_h \in \mathcal{K}_{x,h} \subset \hat{\mathcal{W}}_h$ means that the boundary conditions (2.5d) are also prescribed for $x_h$.

3.2. **Implementation.** In the implementation of (3.1) we use the piecewise constant basis functions $\{\phi_n\}_{n=0}^{N+1}$ which are defined by $\phi_n(t) = 0$ except for

$$\phi_0(t) = 1, \quad t < 0,$$
$$\phi_n(t) = 1, \quad t_{n-1} < t < t_n,$$
$$\phi_{N+1}(t) = 1, \quad t > t_N,$$

and the piecewise linear basis functions $\{\varphi_n\}_{n=0}^{N}$ defined by $\varphi_n(t_k) = \delta_{nk}$.

We make the Ansatz

$$x_h(t) = \sum_{i=0}^{N+1} X_i \phi_i(t), \quad X_i \in \mathbb{R}^d,$$
$$u_h(t) = \sum_{i=0}^{N} U_i \varphi_i(t), \quad U_i \in \mathbb{R}^m,$$
$$z_h(t) = \sum_{i=0}^{N} Z_i \varphi_i(t), \quad Z_i \in \mathbb{R}^d.$$

Note that

$$\dim(\mathcal{W}_h) = (N+2)d,$$
$$\dim(\dot{\mathcal{W}}_h) = (N+2)d - d_0 - d_T,$$
$$\dim(\mathcal{V}_h) = (N+1)d,$$
$$\dim(\mathcal{U}_h) = (N+1)m,$$

where $d_0 = \operatorname{rank}(I_0)$, $d_T = \operatorname{rank}(I_T)$.

We now compute the matrix form of the variational inequalities (3.1). In order to simplify the presentation we assume temporarily that the coefficients $A(t) = A$, $B(t) = B$, $Q(t) = Q$, $R(t) = R$ are constant and that the mesh is uniform, $h_n = h$. We eliminate the prescribed boundary values,

$$X_0 = I_0 X_0 + I_0^{\mathrm{C}} X_0 = x_0 + I_0^{\mathrm{C}} X_0,$$
$$X_{N+1} = I_T X_{N+1} + I_T^{\mathrm{C}} X_{N+1} = x_T + I_T^{\mathrm{C}} X_{N+1}.$$

The remaining unknowns are kept in the column vector

$$X = [I_0^{\mathrm{C}} X_0, X_1, \dots, X_N, I_T^{\mathrm{C}} X_{N+1}]^{\mathrm{T}} \in \mathbb{R}^{(N+2)d - d_0 - d_T}.$$

Similarly,

$$U = [U_0, \dots, U_N]^{\mathrm{T}} \in \mathbb{R}^{(N+1)m},$$
$$Z = [Z_0, \dots, Z_N]^{\mathrm{T}} \in \mathbb{R}^{(N+1)d}.$$

We first consider the case when there are no inequality constraints, that is, $\mathcal{K}_h = \hat{\mathcal{W}}_h \times \mathcal{U}_h \times \mathcal{V}_h$. Then $\varphi_{x,h} - x_h$ can be chosen freely in $\dot{\mathcal{W}}_h$ and $\varphi_{u,h} - u_h$ in $\mathcal{U}_h$ and the variational inequalities reduce to equations. The first equation is, c.f. (2.5a),

$$\begin{aligned}\mathcal{L}_x'(x_h, u_h, z_h, \varphi_{x,h}) = \int_0^T & (-\dot{z}_h + A^{\mathrm{T}} z_h + Q(x_h - \bar{x}), \varphi_{x,h})\,\mathrm{d}t \\ & + (Q_0(x_h(0^-) - \bar{x}_0) - z_h(0), \varphi_{x,h}(0^-)) \\ & + (Q_T(x_h(T^+) - \bar{x}_T) + z_h(T), \varphi_{x,h}(T^+))\ \forall \varphi_{x,h} \in \dot{\mathcal{W}}_h.\end{aligned}$$

By taking $\varphi_{x,h} = \alpha_j \phi_j$ with arbitrary $\alpha_0 \in R(I_0^{\mathrm{C}})$, $\alpha_j \in \mathbb{R}^d$, $j = 1, \dots, N-1$, $\alpha_N \in R(I_T^{\mathrm{C}})$, this leads to

$$-I_0^{\mathrm{C}} Z_0 + I_0^{\mathrm{C}} Q_0 I_0^{\mathrm{C}} X_0 = I_0^{\mathrm{C}} Q_0 \bar{x}_0 - I_0^{\mathrm{C}} Q_0 x_0,$$
$$(I - \tfrac{h}{2} A^{\mathrm{T}}) Z_0 + (-I - \tfrac{h}{2} A^{\mathrm{T}}) Z_1 + hQ X_1 = Q \int_{J_1} \bar{x}\,\mathrm{d}t,$$
$$\vdots$$
$$(I - \tfrac{h}{2} A^{\mathrm{T}}) Z_{N-1} + (-I - \tfrac{h}{2} A^{\mathrm{T}}) Z_N + hQ X_N = Q \int_{J_N} \bar{x}\,\mathrm{d}t,$$
$$I_T^{\mathrm{C}} Z_N + I_T^{\mathrm{C}} Q_T I_T^{\mathrm{C}} X_{N+1} = I_T^{\mathrm{C}} Q_T \bar{x}_T - I_T^{\mathrm{C}} Q_T x_T.$$

In matrix form this is:

$$\mathcal{Q} X + \mathcal{A}^{\mathrm{T}} Z = F \tag{3.2}$$

with the nodewise block matrices

$$\mathcal{Q} = \begin{bmatrix} I_0^{\mathrm{C}} Q_0 & 0 & \dots & \dots & 0 \\ 0 & hQ & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & hQ & 0 \\ 0 & \dots & \dots & 0 & I_T^{\mathrm{C}} Q_T \end{bmatrix},$$

$$\mathcal{A}^{\mathrm{T}} = \begin{bmatrix} -I_0^{\mathrm{C}} & 0 & \dots & \dots & 0 \\ I - \frac{h}{2}A^{\mathrm{T}} & -I - \frac{h}{2}A^{\mathrm{T}} & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & I - \frac{h}{2}A^{\mathrm{T}} & -I - \frac{h}{2}A^{\mathrm{T}} \\ 0 & \dots & \dots & 0 & I_T^{\mathrm{C}} \end{bmatrix},$$

and where $F$ contains the terms on the right hand side. The second equation is

$$\mathcal{L}_u'(x_h, u_h, z_h, \varphi_{u,h}) = \int_0^T (R(u - \bar{u}) - B^{\mathrm{T}} z, \varphi_{u,h}) \,\mathrm{d}t = 0 \quad \forall \varphi_{u,h} \in \mathcal{U}_h.$$

By taking $\varphi_{u,h} = \beta_j \varphi_j$, $j = 0, \dots, N$, with arbitrary $\beta_j \in \mathbb{R}^m$, we obtain

$$\mathcal{R}U + \mathcal{B}^{\mathrm{T}} Z = G \tag{3.3}$$

with

$$\mathcal{R} = \frac{h}{6} \begin{bmatrix} 2R & R & 0 & \cdots & 0 \\ R & 4R & R & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & R & 4R & R \\ 0 & \cdots & 0 & R & 2R \end{bmatrix},$$

$$\mathcal{B}^{\mathrm{T}} = -\frac{h}{6} \begin{bmatrix} 2B^{\mathrm{T}} & B^{\mathrm{T}} & 0 & \cdots & 0 \\ B^{\mathrm{T}} & 4B^{\mathrm{T}} & B^{\mathrm{T}} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & B^{\mathrm{T}} & 4B^{\mathrm{T}} & B^{\mathrm{T}} \\ 0 & \cdots & 0 & B^{\mathrm{T}} & 2B^{\mathrm{T}} \end{bmatrix},$$

and

$$G_j = R \int_0^T \bar{u} \varphi_j \,\mathrm{d}t.$$

The third equation is

$$\begin{aligned} \mathcal{L}_z'(x_h, u_h, z_h, \varphi_{z,h}) = &\sum_{n=1}^{N} \int_{J_n} (\dot{x}_h - Ax_h - Bu_h, \varphi_{z,h}) \,\mathrm{d}t \\ &+ \sum_{n=0}^{N} ([x_h]_n, \varphi_{z,h}(t_n)) = 0 \quad \forall \varphi_{z,h} \in \mathcal{V}_h. \end{aligned}$$

With $\varphi_{z,h} = \gamma_j \varphi_j$, $j = 0, \dots, N$, arbitrary $\gamma_j \in \mathbb{R}^d$, and recalling that $X_0 = x_0 + I_0^{\mathrm{C}} X_0$ and $X_{N+1} = x_T + I_T^{\mathrm{C}} X_{N+1}$, we get

$$\begin{aligned} -\tfrac{h}{2}AX_1 + X_1 - I_0^{\mathrm{C}}X_0 - \tfrac{h}{3}BU_0 - \tfrac{h}{6}BU_1 &= x_0, \\ -\tfrac{h}{2}AX_1 - \tfrac{h}{2}AX_2 + X_2 - X_1 - \tfrac{h}{6}BU_0 - \tfrac{2h}{3}BU_1 - \tfrac{h}{6}BU_2 &= 0, \\ &\vdots \\ -\tfrac{h}{2}AX_N + I_T^{\mathrm{C}}X_{N+1} - X_N - \tfrac{h}{6}BU_{N-1} - \tfrac{h}{3}BU_N &= -x_T. \end{aligned}$$

This has the matrix form

$$\mathcal{A}X + \mathcal{B}U = H.$$

Therefore, the three equations can be written

$$\begin{bmatrix} \mathcal{Q} & 0 & \mathcal{A}^{\mathrm{T}} \\ 0 & \mathcal{R} & \mathcal{B}^{\mathrm{T}} \\ \mathcal{A} & \mathcal{B} & 0 \end{bmatrix} \begin{bmatrix} X \\ U \\ Z \end{bmatrix} = \begin{bmatrix} F \\ G \\ H \end{bmatrix}. \tag{3.4}$$

Since $\mathcal{Q}$ and $\mathcal{R}$ are symmetric positive definite, this system of equations has a saddle point structure.

We now consider the inequality constraints. In this case we choose

$$\begin{aligned} \varphi_{x,h} &= x_h + \alpha_j \phi_j, \quad j = 0, \dots, N+1, \\ \varphi_{u,h} &= u_h + \beta_j \varphi_j, \quad j = 0, \dots, N, \end{aligned}$$

in (3.1), where $\alpha_j \in \mathbb{R}^d$, $\beta_j \in \mathbb{R}^m$ must be chosen so that $\varphi_{x,h} \in \mathcal{K}_{x,h}$, $\varphi_{u,h} \in \mathcal{K}_{u,h}$. We get

$$\begin{aligned} ((\mathcal{Q}X + \mathcal{A}^{\mathrm{T}}Z - F)_j, \alpha_j) &\geq 0, \quad j = 0, \dots, N+1, \\ ((\mathcal{R}U + \mathcal{B}^{\mathrm{T}}Z - G)_j, \beta_j) &\geq 0, \quad j = 0, \dots, N, \end{aligned}$$

for all admissible $\alpha_j$, $\beta_j$. Here $(\mathcal{Q}X + \mathcal{A}^{\mathrm{T}}Z)_j \in \mathbb{R}^d$, $(\mathcal{R}U + \mathcal{B}^{\mathrm{T}}Z - G)_j \in \mathbb{R}^m$ are nodewise residuals from equations (3.2) and (3.3), respectively. If $U_j = U(t_j)$ is in contact, that is, $\|U_j\| = r_u$, then $\beta_j$ must point into the ball $\|u\| \leq r_u$ from the contact point $U_j$ on the boundary. Hence, $(\mathcal{R}U + \mathcal{B}^{\mathrm{T}}Z - G)_j$ must be anti-parallell to $U_j$, that is,

$$(\mathcal{R}U + \mathcal{B}^{\mathrm{T}}Z - G)_j = -s_{u,j}U_j,$$

for some number $s_{u,j} \geq 0$. If $U_j$ is not in contact, $\|U_j\| < r_u$, then $\beta_j$ may point in any direction and

$$(\mathcal{R}U + \mathcal{B}^{\mathrm{T}}Z - G)_j = 0.$$

We may argue similarly for the first inequality in (3.1). Thus, we obtain the complementarity relations

$$(\mathcal{Q}X + \mathcal{A}^{\mathrm{T}}Z - F)(r_x - \|X\|) = 0, \tag{3.5a}$$
$$(\mathcal{R}U + \mathcal{B}^{\mathrm{T}}Z - G)(r_u - \|U\|) = 0, \tag{3.5b}$$
$$\mathcal{A}X + \mathcal{B}U - H = 0, \tag{3.5c}$$
$$\|X\| \le r_x, \tag{3.5d}$$
$$\|U\| \le r_u, \tag{3.5e}$$
$$\mathcal{Q}X + \mathcal{A}^{\mathrm{T}}Z - F = -s_x X, \tag{3.5f}$$
$$\mathcal{R}U + \mathcal{B}^{\mathrm{T}}Z - G = -s_u U, \tag{3.5g}$$
$$s_x \ge 0, \tag{3.5h}$$
$$s_u \ge 0. \tag{3.5i}$$

These equations should be interpreted nodewise, for example, the first equation is

$$\Big(\sum_{i=0}^{N+1} \mathcal{Q}_{ji}X_i + \sum_{i=0}^{N} \mathcal{A}_{ij}Z_i - F_j\Big)(r_x - \|X_j\|) = 0, \quad j = 0, \dots, N+1,$$

where $\mathcal{Q}_{ij}$, $\mathcal{A}_{ij}$, and $F_j$ are the nodewise blocks of $\mathcal{Q}$, $\mathcal{A}$, and $F$. The matrices $s_x$ and $s_u$ are block diagonal matrices, that is,

$$s_x = \operatorname{diag}(s_{x,0}I, \dots, s_{x,N+1}I),$$
$$s_u = \operatorname{diag}(s_{u,0}I, \dots, s_{u,N}I),$$

where $I$ denotes the unit matrix in $\mathbb{R}^{d\times d}$ respectively $\mathbb{R}^{m\times m}$.

3.3. **Projected solver.** In order to solve the complementarity equations (3.5) a projected solver is proposed. This solver is described in Algorithm 1. In short, we start by setting $s_x = 0$, $s_u = 0$ and by solving the system in (3.4). The resulting states and controls may be too large, that is, violate (3.5d) or (3.5e). For each node, we project large states and controls to the boundary of the set of admissible functions, hence violating equations (3.5c), (3.5f), (3.5g). By adding appropriate block diagonal matrices $s_x$ and $s_u$ to the diagonal of the system matrix, the two latter equations are fulfilled. However, since we take no measure to satisfy (3.5c), we must iterate the procedure until convergence.

We claim that all complementarity conditions are fulfilled by the solution obtained by this algorithm. In (3.5a) and (3.5b) either the left or right factor is zero. Equations (3.5c), (3.5f), and (3.5g) are fulfilled at convergence and the remaining equations are trivial consequences of the algorithm.

The implementation of the solver is done in Matlab [14] and the solution of the indefinite system of equations in Algorithm 1 is done by using the built-in Matlab operator \.

**Algorithm 1**: The projected solver

**set** $\tilde{X} = 0,\ \tilde{U} = 0$

**solve**
$$\begin{bmatrix} \mathcal{Q} & 0 & \mathcal{A}^{\mathrm{T}} \\ 0 & \mathcal{R} & \mathcal{B}^{\mathrm{T}} \\ \mathcal{A} & \mathcal{B} & 0 \end{bmatrix} \begin{bmatrix} X \\ U \\ Z \end{bmatrix} = \begin{bmatrix} F \\ G \\ H \end{bmatrix}$$

**while** $\|([X\ U] - [\tilde{X}\ \tilde{U}]\| > \mathrm{tol}$ **do**
  **for** $i = 0, \ldots, N+1$ **do**
    **if** $\|X_i\| > r_x$ **then**
$$\tilde{X}_i = r_x \frac{X_i}{\|X_i\|}$$
    **end**
  **end**
  **for** $i = 0, \ldots, N$ **do**
    **if** $\|U_i\| > r_u$ **then**
$$\tilde{U}_i = r_u \frac{U_i}{\|U_i\|}$$
    **end**
  **end**
  **for** $i = 1, \ldots, N+1$ **do**
$$s_{x,i} = \tfrac{1}{r_x} \|(F - \mathcal{A}^{\mathrm{T}} Z - \mathcal{Q}\tilde{X})_i\|$$
  **end**
  **for** $i = 1, \ldots, N$ **do**
$$s_{u,i} = \tfrac{1}{r_u} \|(G - \mathcal{B}^{\mathrm{T}} Z - \mathcal{R}\tilde{U})_i\|$$
  **end**

**solve**
$$\begin{bmatrix} \mathcal{Q} + s_x & 0 & \mathcal{A}^{\mathrm{T}} \\ 0 & \mathcal{R} + s_u & \mathcal{B}^{\mathrm{T}} \\ \mathcal{A} & \mathcal{B} & 0 \end{bmatrix} \begin{bmatrix} X \\ U \\ Z \end{bmatrix} = \begin{bmatrix} F \\ G \\ H \end{bmatrix}$$

**end**

## 4. A posteriori error estimate

We present an *a posteriori* analysis of the error in the goal functional $\mathcal{J}$ based on the dual weighted residuals methodology. Due to the lack of Galerkin orthogonality, a direct application of the methodology to the discrete variational inequality leads only to a one-sided bound in Subsection 4.1. We therefore use an augmented Lagrangian in Subsection 4.2.

### 4.1. An upper bound for the error in the objective functional.

In a previous article [10] we adapted the dual weighted residuals methodology of [1] to an optimal control problem without inequality constraints. This

results in a representation formula for the error in the goal functional. In the present situation we obtain only an upper bound for the error.

**Theorem 4.1.** *Let $(x,u,z) \in \mathcal{K}$ be a solution to the optimality conditions in* (2.4) *and $(x_h,u_h,z_h) \in \mathcal{K}_h$ be a solution to the finite element approximation in* (3.1). *Then the error in the goal functional satisfies*

$$\mathcal{J}(x,u) - \mathcal{J}(x_h,u_h) \le \tfrac{1}{2}\rho_x + \tfrac{1}{2}\rho_u + \tfrac{1}{2}\rho_x, \tag{4.1}$$

*with the residuals*

$$\begin{aligned}
\rho_x &= \mathcal{J}'_x(x_h,u_h,x-x_h) + \mathcal{F}'_x(x_h,u_h,z_h,x-x_h),\\
\rho_u &= \mathcal{J}'_u(x_h,u_h,u-u_h) + \mathcal{F}'_u(x_h,u_h,z_h,u-u_h),\\
\rho_z &= \mathcal{F}(x_h,u_h,z-\tilde{z}_h),
\end{aligned}$$

*where $\tilde{z}_h \in \mathcal{V}_h$ is arbitrary.*

*Proof.* Introduce the notation $e = (e_x,e_u,e_z) = (x-x_h,u-u_h,z-z_h)$. By the definition of $\mathcal{L}$ in (2.2) and $\mathcal{F}(x,u,z) = 0$, $\mathcal{F}(x_h,u_h,z_h) = 0$, which follow from (2.4c) and (3.1c), we get

$$\begin{aligned}
&\mathcal{J}(x,u) - \mathcal{J}(x_h,u_h)\\
&\quad= \mathcal{L}(x,u,z) - \mathcal{F}(x,u,z) - \mathcal{L}(x_h,u_h,z_h) + \mathcal{F}(x_h,u_h,z_h)\\
&\quad= \mathcal{L}(x,u,z) - \mathcal{L}(x_h,u_h,z_h)\\
&\quad= \int_0^1 \frac{d}{ds}\mathcal{L}(x_h+se_x,u_h+se_u,z_h+se_z)\,\mathrm{d}s\\
&\quad= \int_0^1 \mathcal{L}'(x_h+se_x,u_h+se_u,z_h+se_z,e)\,\mathrm{d}s\\
&\quad= \tfrac{1}{2}\mathcal{L}'(x_h,u_h,z_h,e) + \tfrac{1}{2}\mathcal{L}'(x,u,z,e).
\end{aligned}$$

The last step is the trapezoidal rule for the integral. Since $\mathcal{J}$ is bi-quadratic and $\mathcal{F}$ is tri-linear the integrand is a polynomial in $s$ of degree 1 so that the trapezoidal rule is exact. By (2.3) we have

$$\mathcal{L}'(x,u,z,e) = -\mathcal{L}'(\xi,\xi_h-\xi) \le 0,$$

since $\xi = (x,u,z) \in \mathcal{K}$ and $\xi_h = (x_h,u_h,z_h) \in \mathcal{K}_h \subset \mathcal{K}$. Therefore,

$$\begin{aligned}
\mathcal{J}(x,u) - \mathcal{J}(x_h,u_h) &\le \tfrac{1}{2}\mathcal{L}'(x_h,u_h,z_h,e)\\
&= \tfrac{1}{2}\rho_x(x_h,u_h,z_h,x-x_h) + \tfrac{1}{2}\rho_u(x_h,u_h,z_h,u-u_h)\\
&\quad + \tfrac{1}{2}\rho_z(x_h,u_h,z_h,z-\tilde{z}_h),
\end{aligned}$$

where we also used the Galerkin orthogonality (3.1c) to replace $z_h$ by an arbitrary $\tilde{z}_h \in \mathcal{V}_h$. □

Note that we cannot use (3.1a)–(3.1b) as in [10] to replace also $x_h$ and $u_h$ by arbitrary finite element functions, because this would give

$$\begin{aligned}
\mathcal{L}'(x_h,u_h,z_h,e) = \mathcal{L}'(\xi_h,\xi-\xi_h) &= \mathcal{L}'(\xi_h,\xi-\tilde{\xi}_h) + \mathcal{L}'(\xi_h,\tilde{\xi}_h-\xi_h)\\
&\ge \mathcal{L}'(\xi_h,\xi-\tilde{\xi}_h) \quad \forall \tilde{\xi}_h \in \mathcal{K}_h,
\end{aligned}$$

which cannot be combined with the above inequality.

A similar argument is presented in [2] for an elliptic variational inequality. However, they consider the error in an output quantity, which is a linear functional $\mathcal{J}(u)$ of the solution $u$ of an elliptic variational inequality. The above argument then applies to both $\mathcal{J}$ and $-\mathcal{J}$, yielding upper bounds for the errors in both $\pm\mathcal{J}$, and hence a bound for the absolute value of the error in $\mathcal{J}$. This does not work in our present situation.

4.2. **An augmented Lagrangian.** In order to derive a two-sided error estimate we introduce an augmented Lagrangian. A similar approach was used in [16] for elliptic optimal control problems.

We introduce two additional Lagrange multipliers, $\sigma_x \in \mathcal{Z} = \mathcal{C}([0,T],\mathbb{R})$ and $\sigma_u \in \mathcal{Z} = \mathcal{C}([0,T],\mathbb{R})$. The augmented Lagrangian, $\tilde{\mathcal{L}}\colon \mathcal{X}\times\mathcal{Z}\times\mathcal{Z}\to\mathbb{R}$, is then defined as

$$\begin{aligned}\tilde{\mathcal{L}}(x,u,z,\sigma_x,\sigma_u) &= \mathcal{J}(x,u)+\mathcal{F}(x,u,z)\\ &\quad + \tfrac{1}{2}\int_0^T \sigma_x(t)(\|x(t)\|^2 - r_x^2)\,\mathrm{d}t\\ &\quad + \tfrac{1}{2}\int_0^T \sigma_u(t)(\|u(t)\|^2 - r_u^2)\,\mathrm{d}t.\end{aligned} \tag{4.2}$$

Then we introduce the sets $\omega_x$ and $\omega_u$, where the constraints are active, and the spaces $\mathcal{Z}_x, \mathcal{Z}_u$ of functions supported in these sets,

$$\begin{aligned}\omega_x &= \{t\in[0,T] : \|x(t)\| = r_x\},\\ \omega_u &= \{t\in[0,T] : \|u(t)\| = r_u\},\\ \mathcal{Z}_x &= \{\sigma\in\mathcal{Z} : \sigma(t) = 0 \text{ on } [0,T]\setminus\omega_x\},\\ \mathcal{Z}_u &= \{\sigma\in\mathcal{Z} : \sigma(t) = 0 \text{ on } [0,T]\setminus\omega_u\}.\end{aligned}$$

In order to shorten the notation, we further define

$$\begin{aligned}\mathcal{Y} &= \mathcal{X}\times\mathcal{Z}\times\mathcal{Z},\\ \mathcal{Y}_{\mathrm{ad}} &= \mathcal{K}\times\mathcal{Z}_x\times\mathcal{Z}_u.\end{aligned}$$

We take $\xi = (x,u,z)\in\mathcal{X}$ and note that for $\chi = (\xi,\sigma_x,\sigma_u)\in\mathcal{Y}_{\mathrm{ad}}$ we have

$$\mathcal{L}(\xi) = \tilde{\mathcal{L}}(\chi).$$

The necessary conditions for optimality now become, for $\chi\in\mathcal{Y}_{\mathrm{ad}}$,

$$\tilde{\mathcal{L}}'_x(\chi,\varphi_x) = 0 \quad \forall\varphi_x\in\dot{\mathcal{W}}, \tag{4.3a}$$

$$\tilde{\mathcal{L}}'_u(\chi,\varphi_u) = 0 \quad \forall\varphi_u\in\mathcal{U}, \tag{4.3b}$$

$$\tilde{\mathcal{L}}'_z(\chi,\varphi_z) = 0 \quad \forall\varphi_z\in\mathcal{V}, \tag{4.3c}$$

$$\tilde{\mathcal{L}}'_{\sigma_x}(\chi,\varphi_{\sigma_x}) = 0 \quad \forall\varphi_{\sigma_x}\in\mathcal{Z}_x, \tag{4.3d}$$

$$\tilde{\mathcal{L}}'_{\sigma_u}(\chi,\varphi_{\sigma_u}) = 0 \quad \forall\varphi_{\sigma_u}\in\mathcal{Z}_u, \tag{4.3e}$$

$$\sigma_x,\sigma_u \geq 0 \quad \text{on } [0,T]. \tag{4.3f}$$

Note that these are equations in weak form with test functions taken in linear spaces. The equations are

$$\begin{aligned}
&\int_0^T (-\dot z - A^{\mathrm{T}} z + Q(x-\bar x) + \sigma_x x, \varphi_x)\,\mathrm{d}t \\
&\quad + (Q_0(x(0^-)-\bar x_0) - z(0), \varphi_x(0^-)) \\
&\quad + (Q_T(x(T^+)-\bar x_T) + z(T), \varphi_x(T^+)) = 0 \qquad \forall \varphi_x \in \dot{\mathcal{W}},
\end{aligned} \tag{4.4a}$$

$$\int_0^T (R(u-\bar u) - B^{\mathrm{T}} z + \sigma_u u, \varphi_u)\,\mathrm{d}t = 0 \qquad \forall \varphi_u \in \mathcal{U}, \tag{4.4b}$$

$$\int_0^T (\dot x - Ax + Bu, \varphi_z)\,\mathrm{d}t = 0 \qquad \forall \varphi_z \in \mathcal{V}, \tag{4.4c}$$

$$I_0 x(0^-) = x_0, \quad I_T x(T^-) = x_T, \tag{4.4d}$$

$$\int_0^T \varphi_{\sigma_x}(\|x\|^2 - r_x^2)\,\mathrm{d}t = 0 \qquad \forall \varphi_{\sigma_x} \in \mathcal{Z}_x, \tag{4.4e}$$

$$\int_0^T \varphi_{\sigma_u}(\|u\|^2 - r_u^2)\,\mathrm{d}t = 0 \qquad \forall \varphi_{\sigma_u} \in \mathcal{Z}_u. \tag{4.4f}$$

Note that the former variational inequalities (2.5a), (2.5b) have now become equations (4.4a), (4.4b) by the introduction of the terms $\sigma_x x$ and $\sigma_u u$, which are supported in the contact sets $\omega_x$ and $\omega_u$, respectively.

In order to identify the multipliers $\sigma_x$ and $\sigma_u$ we note that, by (4.3a), (4.3b),

$$\begin{aligned}
\tilde{\mathcal{L}}_x'(\chi, \varphi_x) &= \mathcal{L}_x'(\xi, \varphi_x) + \int_0^T \sigma_x(t)(x(t), \varphi_x(t))\,\mathrm{d}t = 0 \quad \forall \varphi_x \in \dot{\mathcal{W}}, \\
\tilde{\mathcal{L}}_u'(\chi, \varphi_u) &= \mathcal{L}_u'(\xi, \varphi_u) + \int_0^T \sigma_u(t)(u(t), \varphi_u(t))\,\mathrm{d}t = 0 \quad \forall \varphi_u \in \mathcal{U}.
\end{aligned} \tag{4.5}$$

Given a solution $\xi$ of the variational inequality (2.4), these equations determine $\sigma_x$ and $\sigma_u$.

4.3. **Discrete augmented optimality conditions.** In order to derive a discrete version of (4.3) we introduce the discrete versions of the active sets and the discrete multiplier function spaces,

$$\begin{aligned}
\omega_{x,h} &= \{t \in [0,T] : \|x_h(t)\| = r_x\}, \\
\omega_{u,h} &= \{t \in [0,T] : \|u_h(t)\| = r_u\}, \\
\mathcal{Z}_h &= \{\sigma \in \mathcal{C}([0,T], \mathbb{R}) : \sigma|_{J_n} \in P^1(J_n, \mathbb{R}), n = 1, \dots, N\}, \\
\mathcal{Z}_{x,h} &= \{w \in \mathcal{Z}_h : w(t) = 0 \text{ on } [0,T] \setminus \omega_{x,h}\}, \\
\mathcal{Z}_{u,h} &= \{u \in \mathcal{Z}_h : u(t) = 0 \text{ on } [0,T] \setminus \omega_{u,h}\}.
\end{aligned}$$

Additionally, the sets

$$\begin{aligned}
\mathcal{Y}_h &= \mathcal{X}_h \times \mathcal{Z}_h \times \mathcal{Z}_h, \\
\mathcal{Y}_{\mathrm{ad},h} &= \mathcal{K}_h \times \mathcal{Z}_{x,h} \times \mathcal{Z}_{u,h},
\end{aligned}$$

are introduced and the discrete Lagrange multipliers $\sigma_{x,h} \in \mathcal{Z}_{x,h}$, and $\sigma_{u,h} \in \mathcal{Z}_{u,h}$, are defined by, c.f. (4.5),

$$\begin{aligned} \int_0^T \sigma_{x,h}(x_h, \varphi_{x,h})\,\mathrm{d}t &= -\mathcal{L}'_x(x_h, u_h, z_h, \varphi_{x,h}) \quad \forall \varphi_{x,h} \in \dot{\mathcal{W}}_h, \\ \int_0^T \sigma_{u,h}(u_h, \varphi_{u,h})\,\mathrm{d}t &= -\mathcal{L}'_u(x_h, u_h, z_h, \varphi_{u,h}) \quad \forall \varphi_{u,h} \in \mathcal{U}_h. \end{aligned} \tag{4.6}$$

Once $\xi_h = (x_h, u_h, z_h) \in \mathcal{K}_h$ is found by solving the variational inequality (3.1), we can solve for $\sigma_{x,h}$ and $\sigma_{u,h}$. Then $\chi_h = (x_h, u_h, z_h, \sigma_{x,h}, \sigma_{u,h}) \in \mathcal{Y}_{\mathrm{ad},h}$ satisfies

$$\tilde{\mathcal{L}}'_x(\chi_h, \varphi_{x,h}) = 0 \quad \forall \varphi_{x,h} \in \dot{\mathcal{W}}_h, \tag{4.7a}$$

$$\tilde{\mathcal{L}}'_u(\chi_h, \varphi_{u,h}) = 0 \quad \forall \varphi_{u,h} \in \mathcal{U}_h, \tag{4.7b}$$

$$\tilde{\mathcal{L}}'_z(\chi_h, \varphi_{z,h}) = 0 \quad \forall \varphi_{z,h} \in \mathcal{V}_h, \tag{4.7c}$$

$$\tilde{\mathcal{L}}'_{\sigma_x}(\chi_h, \varphi_{\sigma_x,h}) = 0 \quad \forall \varphi_{\sigma_x,h} \in \mathcal{Z}_{x,h}, \tag{4.7d}$$

$$\tilde{\mathcal{L}}'_{\sigma_u}(\chi_h, \varphi_{\sigma_u,h}) = 0 \quad \forall \varphi_{\sigma_u,h} \in \mathcal{Z}_{u,h}. \tag{4.7e}$$

4.4. **An error representation formula.** In order to estimate the absolute value of the error, the one-sided error estimate in Theorem 4.1 is not enough. Therefore we present a representation formula for the error based on the augmented Lagrangian in (4.2). First, we introduce the new residuals by splitting $\tilde{\mathcal{L}}'(\chi_h, \varphi)$ into partial derivatives:

$$\begin{aligned} \tilde{\rho}_x(\chi_h, \varphi_x) &= \mathcal{J}'_x(x_h, u_h, \varphi_x) + \mathcal{F}'_x(x_h, u_h, z_h, \varphi_x) \\ &\quad + \int_0^T \sigma_{x,h}(x_h, \varphi_x)\,\mathrm{d}t && \forall \varphi_x \in \dot{\mathcal{W}}, \\ \tilde{\rho}_u(\chi_h, \varphi_u) &= \mathcal{J}'_u(x_h, u_h, \varphi_u) + \mathcal{F}'_u(x_h, u_h, z_h, \varphi_u) \\ &\quad + \int_0^T \sigma_{u,h}(u_h, \varphi_u)\,\mathrm{d}t && \forall \varphi_u \in \mathcal{U}, \\ \tilde{\rho}_z(\chi_h, \varphi_z) &= \mathcal{F}(x_h, u_h, \varphi_z) && \forall \varphi_z \in \mathcal{V}, \\ \tilde{\rho}_{\sigma_x}(\chi_h, \varphi_{\sigma_x}) &= \tfrac{1}{2}\int_0^T \varphi_{\sigma_x}(\|x_h\|^2 - r_x^2)\,\mathrm{d}t && \forall \varphi_{\sigma_x} \in \mathcal{Z}, \\ \tilde{\rho}_{\sigma_u}(\chi_h, \varphi_{\sigma_u}) &= \tfrac{1}{2}\int_0^T \varphi_{\sigma_u}(\|u_h\|^2 - r_u^2)\,\mathrm{d}t && \forall \varphi_{\sigma_u} \in \mathcal{Z}. \end{aligned}$$

We will evaluate the last two residuals also at $\chi$, that is,

$$\tilde{\rho}_{\sigma_x}(\chi, \varphi_{\sigma_x}) = \tfrac{1}{2}\int_0^T \varphi_{\sigma_x}(\|x\|^2 - r_x^2)\,\mathrm{d}t \quad \forall \varphi_{\sigma_x} \in \mathcal{Z},$$

$$\tilde{\rho}_{\sigma_u}(\chi, \varphi_{\sigma_u}) = \tfrac{1}{2}\int_0^T \varphi_{\sigma_u}(\|u\|^2 - r_u^2)\,\mathrm{d}t \quad \forall \varphi_{\sigma_u} \in \mathcal{Z}.$$

Note that $\tilde{\rho}_z(\chi_h, z-\tilde{z}_h)$ is the same as $\rho_z$ in Theorem 4.1. The other residuals contain additional contributions from the contact sets.

**Theorem 4.2.** *Let $\chi \in \mathcal{Y}_{\mathrm{ad}}$ be a solution to the necessary conditions of optimality in (4.3) and let $\chi_h \in \mathcal{Y}_{\mathrm{ad},h}$ be a solution to its finite element approximation (4.7). Then the error in the objective functional is given by*

$$\begin{aligned}\mathcal{J}(x,u) - \mathcal{J}(x_h,u_h) = {}& \tfrac{1}{2}\tilde{\rho}_x(\chi_h, z-\tilde{z}_h) + \tfrac{1}{2}\tilde{\rho}_u(\chi_h, u-\tilde{u}_h) + \tfrac{1}{2}\tilde{\rho}_z(\chi_h, x-\tilde{x}_h)\\ &+ \tfrac{1}{2}\tilde{\rho}_{\sigma_x}(\chi_h, \sigma_x - \tilde{\sigma}_{x,h}) + \tfrac{1}{2}\tilde{\rho}_{\sigma_u}(\chi_h, \sigma_u - \tilde{\sigma}_{u,h})\\ &+ \tfrac{1}{2}\tilde{\rho}_{\sigma_x}(\chi, \tilde{\sigma}_x - \sigma_{x,h}) + \tfrac{1}{2}\tilde{\rho}_{\sigma_u}(\chi, \tilde{\sigma}_u - \sigma_{u,h}) + R,\end{aligned}$$

*where $(\tilde{x}_h, \tilde{u}_h, \tilde{z}_h) \in \mathcal{X}_h$, $\tilde{\sigma}_{x,h} \in \mathcal{Z}_{x,h}$, $\tilde{\sigma}_{u,h} \in \mathcal{Z}_{u,h}$, $\tilde{\sigma}_x \in \mathcal{Z}_x$, and $\tilde{\sigma}_u \in \mathcal{Z}_u$ are arbitrary and*

$$R = -\tfrac{1}{4}\int_0^T \Big((\sigma_x - \sigma_{x,h})\|x - x_h\|^2 + (\sigma_u - \sigma_{u,h})\|u - u_h\|^2\Big)\,\mathrm{d}t, \tag{4.8}$$

*is a remainder.*

*Proof.* Let $\chi = (x, u, z, \sigma_x, \sigma_u) \in \mathcal{Y}_{\mathrm{ad}}$, $\chi_h = (x_h, u_h, z_h, \sigma_{x,h}, \sigma_{u,h}) \in \mathcal{Y}_{\mathrm{ad},h}$, and $\xi = (x, u, z) \in \mathcal{K}$. We have

$$\mathcal{J}(x,u) = \mathcal{L}(\xi) = \tilde{\mathcal{L}}(\chi), \quad \mathcal{J}(x_h,u_h) = \mathcal{L}(\xi_h) = \tilde{\mathcal{L}}(\chi_h),$$

so that

$$\begin{aligned}\mathcal{J}(x,u) - \mathcal{J}(x_h,u_h) &= \tilde{\mathcal{L}}(\chi) - \tilde{\mathcal{L}}(\chi_h)\\ &= \int_0^1 \tilde{\mathcal{L}}'(\chi_h + s(\chi - \chi_h), \chi - \chi_h)\,\mathrm{d}s.\end{aligned}$$

This integral is computed by the trapezoidal rule as in the proof of Theorem 4.1 and we obtain

$$\mathcal{J}(x,u) - \mathcal{J}(x_h,u_h) = \tfrac{1}{2}\tilde{\mathcal{L}}'(\chi, \chi - \chi_h) + \tfrac{1}{2}\tilde{\mathcal{L}}'(\chi_h, \chi - \chi_h) + R. \tag{4.9}$$

Since the integrand $f(s) = \tilde{\mathcal{L}}'(\chi_h + s(\chi - \chi_h), \chi - \chi_h)$ contains terms that are quadratic in $s$, there is now a remainder $R = \frac{1}{2}\int_0^1 f''(s)s(s-1)\,\mathrm{d}s$ and a simple calculation leads to (4.8). The first term on the right side of (4.9) is

$$\begin{aligned}\tilde{\mathcal{L}}'(\chi, \chi - \chi_h) = {}& \tilde{\mathcal{L}}'_x(\chi, x - x_h) + \tilde{\mathcal{L}}'_u(\chi, u - u_h) + \tilde{\mathcal{L}}'_z(\chi, z - z_h)\\ &+ \tilde{\mathcal{L}}'_{\sigma_x}(\chi, \sigma_x - \sigma_{x,h}) + \tilde{\mathcal{L}}'_{\sigma_u}(\chi, \sigma_u - \sigma_{u,h}).\end{aligned}$$

From the optimality conditions in (4.3) we see that the first three terms vanish, so that

$$\begin{aligned}\tilde{\mathcal{L}}'(\chi, \chi - \chi_h) &= \tilde{\mathcal{L}}'_{\sigma_x}(\chi, \sigma_x - \sigma_{x,h}) + \tilde{\mathcal{L}}'_{\sigma_u}(\chi, \sigma_u - \sigma_{u,h})\\ &= \tilde{\rho}_{\sigma_x}(\chi, \sigma_x - \sigma_{x,h}) + \tilde{\rho}_{\sigma_u}(\chi, \sigma_u - \sigma_{u,h}).\end{aligned}$$

These terms are not zero, since $\sigma_x - \sigma_{x,h}$ belongs to $\mathcal{Z}$ but not necessarily to $\mathcal{Z}_x$ because $\omega_{x,h} \not\subset \omega_x$. The same argument is valid for $\sigma_u - \sigma_{u,h}$. Using

(4.3d), (4.3e), and that $\tilde{\rho}_{\sigma_x}$ and $\tilde{\rho}_{\sigma_u}$ are linear in the second argument, we can replace $\sigma_x$ and $\sigma_u$ by arbitrary $\tilde{\sigma}_x \in \mathcal{Z}_x$ and $\tilde{\sigma}_u \in \mathcal{Z}_u$, so that

$$\tilde{\mathcal{L}}'(\chi, \chi - \chi_h) = \tilde{\rho}_{\sigma_x}(\chi, \tilde{\sigma}_x - \sigma_{x,h}) + \tilde{\rho}_{\sigma_u}(\chi, \tilde{\sigma}_u - \sigma_{u,h}).$$

The second term in (4.9) is

$$\begin{aligned}\tilde{\mathcal{L}}'(\chi_h, \chi - \chi_h) = \tilde{\rho}_x(\chi_h, z - z_h) + \tilde{\rho}_u(\chi_h, u - u_h) + \tilde{\rho}_z(\chi_h, x - x_h) \\ + \tilde{\rho}_{\sigma_x}(\chi_h, \sigma_x - \sigma_{x,h}) + \tilde{\rho}_{\sigma_u}(\chi_h, \sigma_u - \sigma_{u,h}).\end{aligned}$$

The discrete necessary conditions (4.7) mean that these residuals are orthogonal to the respective finite element function spaces, so that $x_h, u_h, z_h$, $\sigma_{x,h}$, and $\sigma_{u,h}$ can be replaced by arbitrary $\tilde{x}_h \in \dot{\mathcal{W}}_h$, $\tilde{u}_h \in \mathcal{U}_h$, $\tilde{z}_h \in \mathcal{V}_h$, $\tilde{\sigma}_{x,h} \in \mathcal{Z}_{x,h}$, and $\tilde{\sigma}_{u,h} \in \mathcal{Z}_{u,h}$, yielding

$$\begin{aligned}\tilde{\rho}_x(\chi_h, z - z_h) &= \tilde{\rho}_x(\chi_h, z - \tilde{z}_h), \\ \tilde{\rho}_u(\chi_h, u - u_h) &= \tilde{\rho}_u(\chi_h, u - \tilde{u}_h), \\ \tilde{\rho}_z(\chi_h, x - x_h) &= \tilde{\rho}_z(\chi_h, x - \tilde{x}_h), \\ \tilde{\rho}_{\sigma_x}(\chi_h, \sigma_x - \sigma_{x,h}) &= \tilde{\rho}_{\sigma_x}(\chi_h, \sigma_x - \tilde{\sigma}_{x,h}), \\ \tilde{\rho}_{\sigma_u}(\chi_h, \sigma_u - \sigma_{u,h}) &= \tilde{\rho}_{\sigma_u}(\chi_h, \sigma_u - \tilde{\sigma}_{u,h}).\end{aligned}$$

This completes the proof. □

4.5. **An error estimator.** In order to derive a computable *a posteriori* error estimate we expand each term in the error formula separately:

$$\begin{aligned}\tilde{\rho}_x(x_h, u_h, z_h, x - \tilde{x}_h) &= \mathcal{J}'_x(x_h, u_h, x - \tilde{x}_h) + \mathcal{F}'_x(x_h, u_h, z_h, x - \tilde{x}_h) \\ &\quad + \int_0^T \sigma_{x,h}(x_h, x - \tilde{x}_h)\,\mathrm{d}t, \\ \tilde{\rho}_u(x_h, u_h, z_h, u - \tilde{u}_h) &= \mathcal{J}'_u(x_h, u_h, u - \tilde{u}_h) + \mathcal{F}'_u(x_h, u_h, z_h, u - \tilde{u}_h) \\ &\quad + \int_0^T \sigma_{u,h}(u_h, u - \tilde{u}_h)\,\mathrm{d}t, \\ \tilde{\rho}_z(x_h, u_h, z_h, z - \tilde{z}_h) &= \mathcal{F}(x_h, u_h, z - \tilde{z}_h), \\ \tilde{\rho}_{\sigma_x}(x_h, u_h, z_h, \sigma_x - \tilde{\sigma}_{x,h}) &= \tfrac{1}{2}\int_0^T (\sigma_x - \tilde{\sigma}_{x,h})(\|x_h\|^2 - r_x^2)\,\mathrm{d}t, \\ \tilde{\rho}_{\sigma_u}(x_h, u_h, z_h, \sigma_u - \tilde{\sigma}_{u,h}) &= \tfrac{1}{2}\int_0^T (\sigma_u - \tilde{\sigma}_{u,h})(\|u_h\|^2 - r_u^2)\,\mathrm{d}t, \\ \tilde{\rho}_{\sigma_x}(x, u, z, \tilde{\sigma}_x - \sigma_{x,h}) &= \tfrac{1}{2}\int_0^T (\tilde{\sigma}_x - \sigma_{x,h})(\|x\|^2 - r_x^2)\,\mathrm{d}t, \\ \tilde{\rho}_{\sigma_u}(x, u, z, \tilde{\sigma}_u - \sigma_{u,h}) &= \tfrac{1}{2}\int_0^T (\tilde{\sigma}_u - \sigma_{u,h})(\|u\|^2 - r_u^2)\,\mathrm{d}t.\end{aligned}$$

Here we know only $x_h$, $u_h$, and $z_h$ and the remaining functions must be approximated. In order to do so we compute approximate solutions $x_{\text{fine}}$,

$u_{\text{fine}}$, and $z_{\text{fine}}$ on a finer mesh. These are used to replace $x$, $u$, $z$ and $\tilde{x}$, $\tilde{u}$, $\tilde{z}$ wherever they occur. Likewise, we replace $\tilde{x}_h$, $\tilde{u}_h$, $\tilde{z}_h$ by $x_h$, $u_h$, $z_h$.

The Lagrange multipliers $\sigma_{x,h}$, $\sigma_{u,h}$, $\tilde{\sigma}_{x,h}$, $\tilde{\sigma}_{u,h}$ are computed by solving the equations in (4.6) with $x_h$, $u_h$, and $z_h$ as coefficients. The Lagrange multipliers $\sigma_x$, $\sigma_u$, $\tilde{\sigma}_x$, and $\tilde{\sigma}_u$ are approximated by solving (4.6) with $x_{\text{fine}}$, $u_{\text{fine}}$, and $z_{\text{fine}}$ as coefficients. After these substitutions we write the various residuals as elementwise sums and apply norms to the terms in an appropriate way, see [10] for details. Finally, we ignore the remainder $R$, which is cubic in the errors in $\sigma_{x,h}$, $\sigma_{u,h}$, $x_h$, and $u_h$, while the residuals are formally linear in these errors.

4.6. **An adaptive algorithm.** The error estimator in the previous section is used for implementing an adaptive finite element method. The algorithm is described in detail in [10]. A solution is computed on a coarse mesh, then the contribution to the error estimator from each interval is computed. The mesh is refined according to the principle that the elements with the largest contribution to the total error are refined. This procedure is iterated until the required error tolerance has been achieved.

## 5. Numerical examples

The algorithm described in this work has been tested on several problems. These tests are described in this section.

5.1. **Fuller's problem 1.** The following problem is known as Fuller's problem [3, 8].

$$\text{Minimize} \qquad \int_0^1 x_1(t)^2 \,\mathrm{d}t \tag{5.1}$$

subject to

$$\begin{aligned} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} &= \begin{bmatrix} x_2(t) \\ -u(t) \end{bmatrix}, \quad 0 < t < 1, \\ \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} &= \begin{bmatrix} 0.01 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} 0.01 \\ 0 \end{bmatrix}, \\ |u(t)| &\leq 1, \quad 0 < t < 1. \end{aligned} \tag{5.2}$$

We regularize the problem by adding the following term to (5.1):

$$\omega \int_0^1 (x_2(t)^2 + u(t)^2) \,\mathrm{d}t,$$

with $\omega = 10^{-10}$. It is known, that the non-regularized problem displays Fuller's phenomenon, that is, the control makes infinitely many switchings between $\pm 1$, but that the regularized problem only has a finite, but large, number of switchings [13, 4]. In Figure 5.1 the control is plotted and we can see that it is symmetric around $t = 0.5$. The adaptive solver has started to resolve the switchings. The value of the goal functional for the non-regularized problem computed with an exact formula is $\mathcal{J}(x, u) = 1.5280 \cdot$

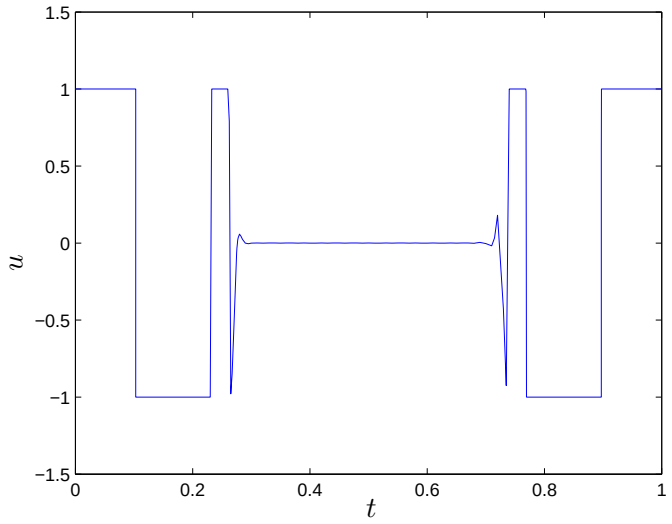


FIGURE 5.1. The control in Fuller's problem 1.

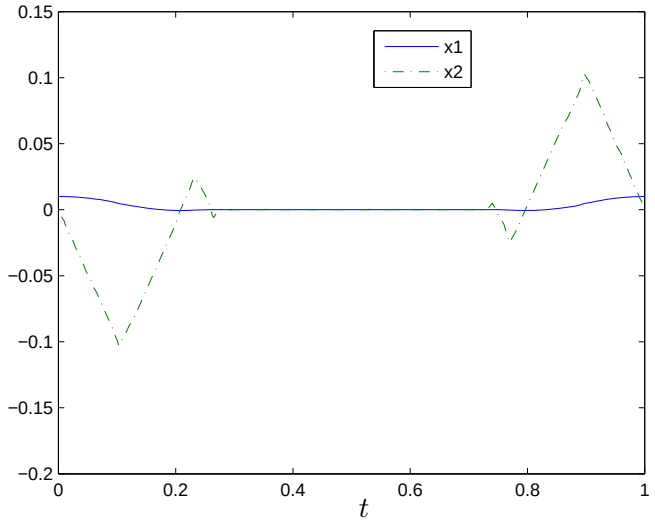


FIGURE 5.2. The states in Fuller's problem 1.

$10^{-5}$ [9]. The value of the goal functional for the numerical solution is is $\mathcal{J}(x_h, u_h) = 1.4760 \cdot 10^{-5}$. This solution is computed with the tolerance $10^{-4}$.

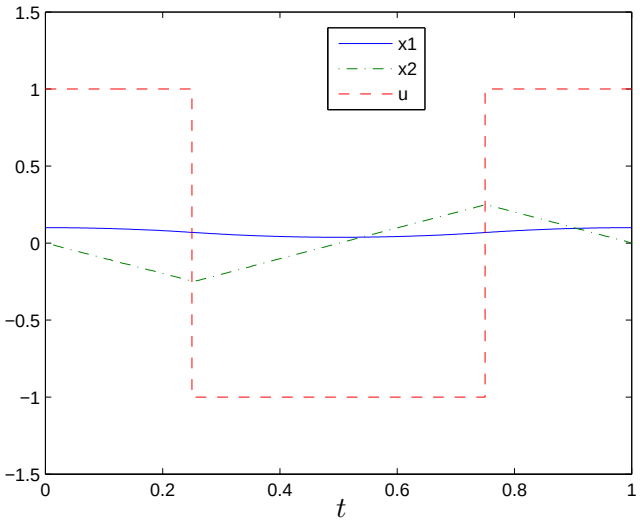


FIGURE 5.3. The computed solution of Fuller's problem 2 with only two switchings.

5.2. **Fuller's problem 2.** Changing the boundary values in (5.2) to

$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix},$$

gives a problem where Fuller's phenomenon does not occur [9]. The result can be found in Figure 5.3.

5.3. **Lane change maneuver.** In this subsection we present a stabilization maneuver taken from vehicle dynamics. A vehicle is described by a particle on which a steering force, $F_Y$, and a braking force, $F_X$, are acting as controls. The vehicle has an initial velocity in both the $X$- and $Y$-directions and it has an initial position $Y = 8$ m. The aim is to find the optimal way to steer and brake in order to reach another lane at $Y = 0$ m, where the vehicle is moving only in the $X$-direction, that is $V_Y = 0$. We minimize $\int_0^T Y^2 \, \mathrm{d}t$ for fixed $T = 4.15$ s, and with additional regularization the problem is:

$$\begin{aligned} \text{Minimize} \quad & \tfrac{1}{2} \int_0^T \Big\{ Y(t)^2 \\ & + \omega \big( X(t)^2 + V_X(t)^2 + V_Y(t)^2 + F_X(t)^2 + F_Y(t)^2 \big) \Big\} \, \mathrm{d}t \end{aligned}$$

subject to the dynamics of the vehicle,

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{V}_X \\ \dot{V}_Y \end{bmatrix} = \begin{bmatrix} V_X \\ V_Y \\ F_X \\ F_Y \end{bmatrix}, \quad 0 < t < T; \qquad \begin{bmatrix} X(0) \\ Y(0) \\ V_X(0) \\ V_Y(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ 24 \\ 14.4 \end{bmatrix}, \quad V_Y(T) = 0,$$

$$F_X^2 + F_Y^2 \leq (\mu g)^2, \quad 0 < t < T.$$

The inequality constraint means that the force (divided by the mass of the vehicle) is limited by the friction circle. We use the friction coefficient $\mu = 1$ and acceleration of gravity $g = 10$ m/s$^2$. We emphasize that the problem that we really want to solve from a vehicle dynamics point of view has $\omega = 0$. However, due to numerical reasons we must regularize and use $\omega > 0$. Here we present a solution with $\omega = 10^{-10}$. The results presented are computed with the tolerance $10^{-4}$ for the discretization error.

The results are shown in Figures 5.4–5.8. In Figure 5.4 the optimal track is shown. The same figure displays the solution obtained with the optimal control module PROPT [15] and we see that the solutions coincide. The velocities of the vehicle in the $X$- and $Y$-directions are shown in Figure 5.5 and we note that $V_X$ is constant because the vehicle is only steering and no braking takes place, which can be seen in Figure 5.6, where the controls are plotted. The force acting in the $X$-direction, $F_X$, is equal to zero during the whole maneuver and only the steering force is active. We also note that the forces are on the friction circle, that is $F_X^2 + F_Y^2 = (\mu g)^2$. This is expected since the optimization problem is convex ($\mathcal{K}$ and $\mathcal{J}$ are convex). Figure 5.7 shows the size of the error when the mesh is refined adaptively or uniformly. The advantage of the adaptive solver is clear when looking at the number of nodes needed to reach a precision of $10^{-4}$. Approximately one third more nodes are used in the case of uniform refinement in order to reach the same accuracy. The figure also implies that this effect will be more clear when decreasing the tolerance further. Finally, the adapted mesh is shown in Figure 5.8.

## 6. Conclusions

This work aimed at investigating the potential use of adaptive finite element methods for solving optimal control problems with inequality constraints on controls and states. The contributions consist of setting the problem in a mathematical framework, deriving an *a posteriori* error estimate, and the implementation of an adaptive algorithm. The results so far

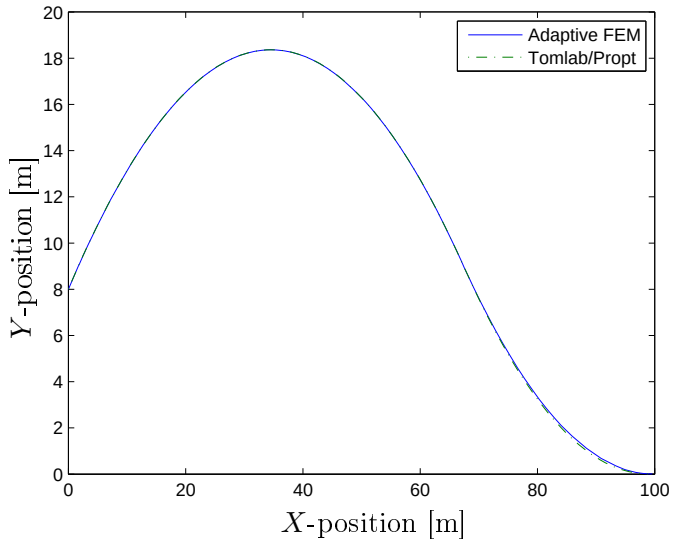


FIGURE 5.4. The track of the vehicle computed with the adaptive finite element solver and the optimal control module PROPT.

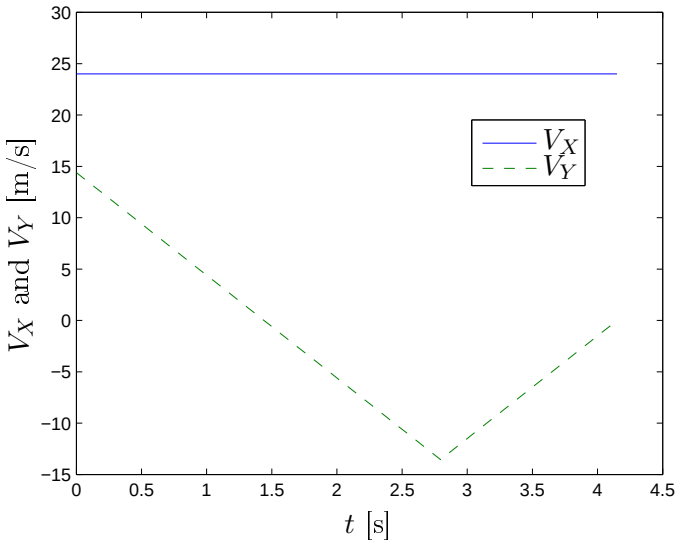


FIGURE 5.5. The velocities in the $X$-direction and the $Y$-direction respectively. Note that $V_X$ is constant.

show that the use of an adaptive finite element method can contribute to automation of the solution of optimal control problems.

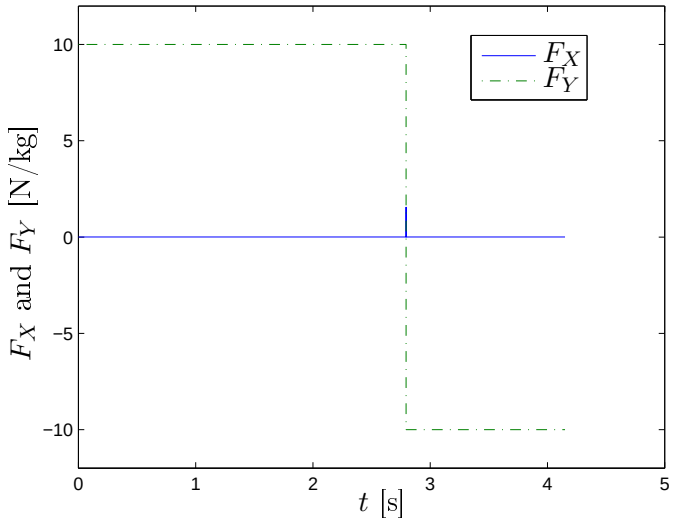


FIGURE 5.6. The forces acting on the vehicle.

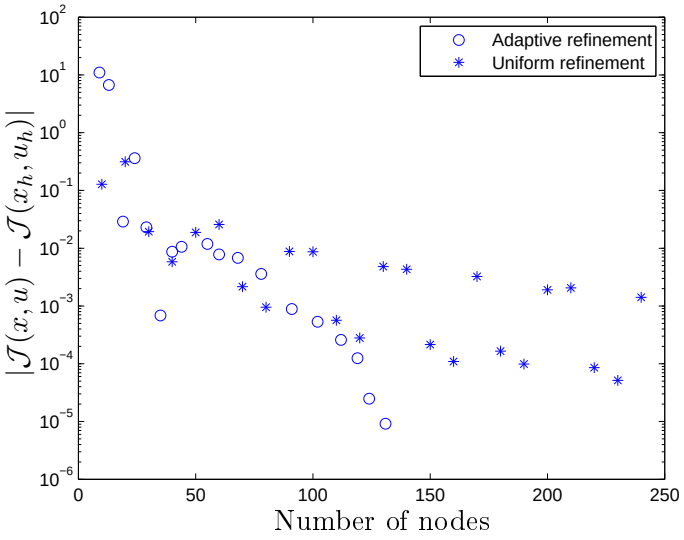


FIGURE 5.7. The error for the solution computed on a uniformly refined mesh and an adaptively refined mesh respectively.

At this stage more focus has to be put on the issue of effectiveness of the implementation of the algorithm. We also have to address the problem of handling inequality constraints for non-linear problems.

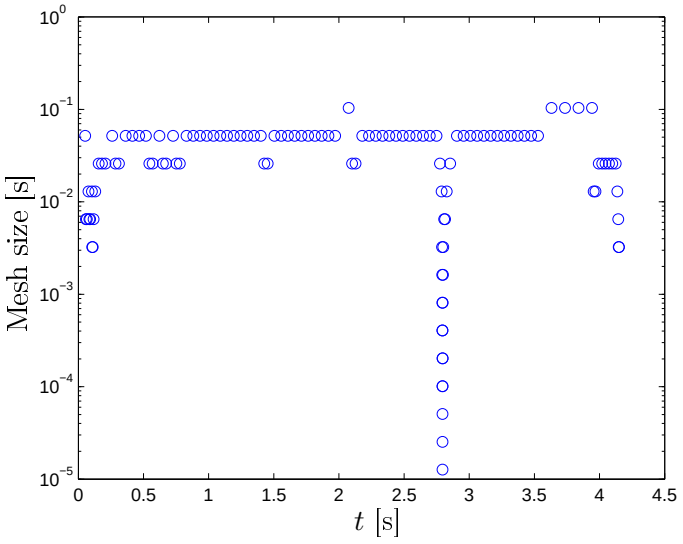


FIGURE 5.8. The mesh size for the adaptively refined mesh.

## 7. Acknowledgement

The authors wish to thank Dr Mathias Lidberg, Department of Applied Mechanics, Chalmers University of Technology, and Dr Kjell Holmåker, Department of Mathematical Sciences, Chalmers University of Technology, for helpful suggestions.

## References

1. R. Becker and R. Rannacher, *An optimal control approach to a posteriori error estimation in finite element methods*, Acta Numer. **10** (2001), 1–102.
2. H. Blum and J.T. Suttmeier, *Weighted error estimates for finite element solutions of variational inequalities*, Computing **65** (2000), 119–134.
3. V. F. Borisov, *Fuller's phenomenon: review*, J. Math. Sci. (New York) **100** (2000), 2311–2354, Dynamical systems, 8.
4. P. Brunovský and J. Mallet-Paret, *Switchings of optimal controls and the equation* $y^{(4)} + |y|^{\alpha}\operatorname{sign} y = 0,\ 0 < \alpha < 1$, Časopis Pěst. Mat. **110** (1985), 302–313.
5. A. Eriksson and A. Nordmark, *Temporal finite element formulation of optimal control in mechanisms*, Comput. Methods Appl. Mech. Engrg. **199** (2010), 1783–1792.
6. D. Estep, D. H. Hodges, and M. Warner, *Computational error estimation and adaptive error control for a finite element solution of launch vehicle trajectory problems*, SIAM J. Sci. Comput. **21** (1999), 1609–1631 (electronic).
7. D. J. Estep, D. H. Hodges, and M. Warner, *The solution of a launch vehicle trajectory problem by an adaptive finite-element method*, Comput. Methods Appl. Mech. Engrg. (2001), 4677–4690.
8. A. T. Fuller, *Relay control systems optimized for various performance criteria.*, Proc. First World Congress IFAC (London), Butterworths, 1961, pp. 510–519.
9. K. Holmåker, personal communication.

10. K. Kraft and S. Larsson, *The dual weighted residuals approach to optimal control of ordinary differential equations*, BIT **50** (2010), 587–607.
11. ———, *An adaptive finite element method for non-linear optimal control problems*, Preprint 2011-01, Mathematical Sciences, Chalmers University of Technology, Gothenburg (2011).
12. K. Kraft, S. Larsson, and M. Lidberg, *Using an adaptive FEM to determine the optimal control of a vehicle during a collision avoidance manoeuvre*, Proceedings of the 48th Scandinavian Conference on Simulation and Modeling (SIMS2007), Linköping University Electronic Press, 2007, `http://www.ep.liu.se/ecp/027/`.
13. I. Kupka, *The ubiquity of Fuller's problem*, Nonlinear Controllability and Optimal Control (New York), Dekker, 1990, pp. 313–350.
14. MATLAB, `http://www.mathworks.com/products/matlab/`.
15. P. Rutquist and M. M. Edvall, *PROPT-MATLAB Optimal Control Software*, `http://tomdyn.com/`.
16. B. Vexler and W. Wollner, *Adaptive finite elements for elliptic optimization problems with control constraints*, SIAM J. Control Optim. **47** (2008), 509–534.
17. R. Winther, *Initial value methods for parabolic control problems*, Math. Comp. **34** (1980), 115–125.

Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE–412 96 Gothenburg, Sweden
*E-mail address*: `karin.kraft@chalmers.se`

Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE–412 96 Gothenburg, Sweden
*E-mail address*: `stig@chalmers.se`