An Augmented Lagrangian for Optimal Control of DAE Systems: Algorithm and Properties

Marco Aurelio Aguiar, Eduardo Camponogara, and Bjarne Foss

***Abstract*—This work proposes a relax-and-discretize approach for optimal control of continuous-time differential algebraic systems (DAE). It works by relaxing the algebraic equations and penalizing the violation into the objective function using the augmented Lagrangian, which converts the original problem into a sequence of optimal control problems (OCPs) of ordinary differential equations (ODEs). The relax-and-discretize approach brings about flexibility, by allowing the OCPs of ODEs to be solved by the method of choice, such as direct or indirect methods. Conditions are developed for global, local, and sub-optimal convergence in terms of the solution of the underlying OCPs. The method is applied to an illustrative example.**

***Index Terms*—Optimal control, Differential-Algebraic Systems, Nonlinear systems, Optimization algorithms.**

I. Introduction

The augmented Lagrangian method is well established in constrained optimization, arguably because of its efficient algorithms and strong theory [1]. It has enjoyed diverse applications that include model predictive control [2], distributed and parallel optimization [3], to name a few.

For optimal control, an equivalent method has been proposed and applied to some academic systems [4], [5]. The algorithms thereof relax the algebraic equations of an optimal control problem (OCP) with a system of a differential algebraic equations (DAE). However, little effort has been put into developing a theoretical endorsement for such algorithms. An exception is [6], which provides conditions that ensure converge for problems with a convex objective and a linear system.

The contributions are as follows. This paper proposes an algorithm for solving optimal control problems of continuous-time DAE systems, providing conditions for global and local convergence, and convergence with sub-optimal iterations. This algorithm facilitates the solution of OCPs of DAEs in embedded hardware with limited computational power by eliminating the need of a DAE solver. For a network system with subsystems coupled by algebraic input-output equations, the proposed algorithm enables the decoupling by relaxing these equations. In a distributed solution scheme, such a decoupling allows subsystem problems to be solved with distinct approaches, unlike discretize-and-relax approaches. The paper illustrates the computational and implementation aspects of the algorithm in a simple, but representative example.

Funded by an INTPart grant from the Research Council of Norway and a doctoral scholarship from CAPES/Brasil.

M. de Aguiar and E. Camponogara are with the Department of Automation and Systems Engineering, Federal University of Santa Catarina, Florianópolis, Brazil, e-mail: marcoaaguiar@gmail

B. Foss is with the Engineering Cybernetics Department, Norwegian University of Science and Technology, Trondheim, Norway

II. Optimal Control Problem

This work is concerned with the optimal control problem for a system of differential-algebraic equations of the form:

$$
\begin{aligned}
\mathcal{P}: \quad \min \quad & J(x,y,u) = \int_{t_0}^{t_f} L(x,y,u,t)\,dt && \text{(1a)}\\
\text{s.t.:} \quad & \dot{x} = f(x,y,u,t) && \text{(1b)}\\
& g(x,y,u,t) = 0 && \text{(1c)}\\
& x(t_0) = x_0 && \text{(1d)}\\
& u(t) \in U_B,\ t \in [t_0, t_f] && \text{(1e)}
\end{aligned}
$$

with $U_B = \{\mathrm{u} \in U \,|\, u_L \leq \mathrm{u} \leq u_U\}$ and where $x(t) \in X = \mathbb{R}^{N_x}$ is the state variable, $y(t) \in Y = \mathbb{R}^{N_y}$ is the algebraic variable, $u(t) \in U_B \subset U = \mathbb{R}^{N_u}$ is the control variable, and t is the time variable. The function of dynamics f, the function of algebraic relations g, and the function of dynamic cost L are assumed to be continuously differentiable with respect to their arguments.

The DAE system formed by (1b) and (1c) is assumed to be in the semi-explicit index-1 form, which means that it is solvable for y and the Jacobian $\frac{\partial g}{\partial y}$ is invertible. The algebraic equation can also be used to model equality constraints, e.g. $u_1 + u_2 = 0$, where either u_1 or u_2 can be represented as an algebraic variable y. Problems of the form $\mathcal{P}$ with a final cost function can be framed to this approach by transforming the objective [7].

The Hamiltonian function of the OCP (1) is

$$
\begin{aligned}
H(x,\lambda,y,\nu,u,t) = L(x,y,u,t) + \lambda^T f(x,y,u,t) \\
+ \nu^T g(x,y,u,t) \quad (2)
\end{aligned}
$$

where $\lambda : [t_0, t_f] \to \mathbb{R}^{N_x}$ is the multiplier associated with the state equations, and $\nu(t) \in \mathbb{R}^{N_y}$ is the multiplier associated with the algebraic equations.

Using this Hamiltonian, the necessary conditions for $(x^*, \lambda^*, y^*, \nu^*, u^*)$ to be optimal are given by [8]:

$$
\begin{aligned}
\frac{\partial H}{\partial x}^T &= -\dot{\lambda}^* = \frac{\partial L}{\partial x}^T + \frac{\partial f}{\partial x}^T \lambda^* + \frac{\partial g}{\partial x}^T \nu^* && \text{(3a)}\\
\frac{\partial H}{\partial y}^T &= \frac{\partial L}{\partial y}^T + \frac{\partial f}{\partial y}^T \lambda^* + \frac{\partial g}{\partial y}^T \nu^* = 0 && \text{(3b)}
\end{aligned}
$$

$$u^*(t) = \underset{\mathrm{u} \in U_B}{\arg\min}\, H(x^*(t), \lambda^*(t), y^*(t), \nu^*(t), \mathrm{u}, t) \qquad (3c)$$

$$\frac{\partial H}{\partial \lambda}^T = \dot{x}^* = f(x^*, y^*, u^*), \quad \lambda^*(t_f) = 0 \qquad (3d)$$

$$\frac{\partial H}{\partial \nu}^T = g(x^*, y^*, u^*) = 0, \qquad x^*(t_0) = x_0 \qquad (3e)$$

for all $t \in [t_0, t_f]$. More general necessary conditions are found in [9] which concerns OCPs with mixed constraints. Less restrictive conditions were recently developed for pure and mixed constraints in [10].

Methods that solve the boundary value problem (BVP) resulting of the necessary optimality conditions (3) are known as indirect methods [11]. The proposed relax-and-discretize algorithm is presented and its properties are demonstrated, both of which make use of the Hamiltonian.

III. Augmented Lagrangian Algorithm

The algorithm proposed in this work solves the OCP in the form $\mathcal{P}$ by first relaxing the algebraic constraint (1c), and then introducing a new objective functional,

$$J_\mu(x, y, u, \nu) = \int_{t_0}^{t_f} \mathcal{L}_\mu(x, y, u, \nu, t)\, dt \qquad (4)$$

where the function $\mathcal{L}_\mu$ is defined by

$$\begin{aligned} \mathcal{L}_\mu(x, y, u, \nu, t) = L(x, y, u, t) + \nu(t)^T g(x, y, u, t) \\ + \frac{\mu}{2} \left\| g(x, y, u, t) \right\|^2, \end{aligned} \qquad (5)$$

where $\mu > 0$ is a scalar, and the function $\nu : [t_0, t_f] \to \mathbb{R}^{N_y}$ is an estimate of the multiplier function ν^*, which will be driven by the algorithm towards satisfying the optimality conditions (3) of problem $\mathcal{P}$.

The functional (4) is the objective of the auxiliary OCP solved by the algorithm at each iteration k, given by

$$\mathcal{P}_{\mathcal{L}}(\mu_k, \nu_k) : \min_{y,u}\; J_{\mu_k} = \int_{t_0}^{t_f} \mathcal{L}_{\mu_k}(x, y, u, \nu_k, t)\, dt \quad (6a)$$

$$\text{s.t.: } \dot{x} = f(x, y, u, t) \qquad (6b)$$

$$x(t_0) = x_0 \qquad (6c)$$

$$u \in U_B,\; t \in [t_0, t_f] \qquad (6d)$$

Notice that without an algebraic equation, the variable y is free to be optimized. In this sense, the algebraic variable plays the same role as the control variable u. Therefore, an extended control variable $\widehat{u} = [u, y]$ can be defined, where $\widehat{u}(t) \in \widehat{U} = U_B \times Y$. Using $\widehat{u}$, problem $\mathcal{P}_{\mathcal{L}}$ meets the standard form of an OCP of ODE, whose optimality conditions are well established [7].

A. Algorithm

The proposed algorithm follows the same structure of the augmented Lagrangian for standard constrained optimization [1]. Let μ_0 be an initial value for the sequence of penalty values $\{\mu_k\}$, ν_0 be an initial estimate for the sequence of multipliers $\{\nu_k\}$, and ε_g be a tolerance on the violation of the algebraic constraint. Starting with these parameters, at each iteration k, the problem (6) is solved, the multiplier estimate and penalty are updated, and the process is repeated until an acceptable tolerance is achieved, as detailed in Algorithm 1.

Algorithm 1 Augmented Lagrangian for Optimal Control

Require: μ_0, ν_0, and ε_g:
for $k = 1, 2, \dots$ **do**
    $(J_k, x_k, y_k, u_k) \leftarrow solve\{\mathcal{P}_{\mathcal{L}}(\mu_k, \nu_k)\}$
    $\nu_{k+1} \leftarrow \nu_k + \mu_k g(x_k, y_k, u_k)$
    $\mu_{k+1} \leftarrow update\_mu\{\mu_k\}$
    if $\|g(x_k, y_k, u_k)\| < \varepsilon_g$ **then**
        return u_k
    end if
end for

The pseudo-function $solve$ yields a solution for the sub-problem $\mathcal{P}_{\mathcal{L}}$ and returns the functional values J_k and the trajectories for the states, algebraic and control variables. The pseudo-function $update\_mu$ represents the use of an update rule for the penalization μ_k. For the convergence analysis it is assumed that $\mu_{k+1} = \beta \mu_k$ with a $\beta > 1$ to ensure that $\mu_k \to \infty$. In practice, however, a $\mu_k \to \infty$ will cause ill-conditioning on the Hessian of the sub-problem $\mathcal{P}_{\mathcal{L}}$, therefore when performing a computational implementation, it is recommended to use an upper bound $\mu_{\max}$ for the penalization.

B. Mathematical Properties

Conditions are now established for the solution sequence produced by the algorithm to arrive at a global solution of the OCP of DAE. Less restrictive conditions are then presented for convergence to local solutions and convergence under a suboptimal solution sequence, which reflect situations typically found in practice.

Before presenting the convergence theorems, some definitions are in order.

Assumption 1 (Regularity)**.** *For problem $\mathcal{P}$ (1) and $\mathcal{P}_{\mathcal{L}}(\mu_k, \nu_k)$ (6) to be well-conditioned, we assume that*

1) *$x : [t_0, t_f] \to \mathbb{R}^{N_x}$ is continuously differentiable; $y : [t_0, t_f] \to \mathbb{R}^{N_y}$, $u : [t_0, t_f] \to U_B$, and $\nu_k : [t_0, t_f] \to \mathbb{R}^{N_y}$ are continuous,*
2) *L, g, and f are continuously differentiable with respect to all the arguments,*
3) *The space of feasible functions for problems $\mathcal{P}$ and $\mathcal{P}_{\mathcal{L}}$ are compact,*
4) *the Jacobian $\frac{\partial g}{\partial y}(x(t), y(t), u(t), t)$ has full rank for all $x(t) \in X$, $y(t) \in Y$, $u(t) \in U_B$, and $t \in [t_0, t_f]$,*
5) *the sequence $\{\mu_k\}$ has the property that $0 < \mu_k < \mu_{k+1}$ for all k, and $\mu_k \to \infty$ as $k \to \infty$,*
6) *problem $\mathcal{P}$ and $\mathcal{P}_{\mathcal{L}}(\mu_k, \nu_k)$ are solvable.*

From condition 4) of the Assumption 1, the algorithm is not applicable to OCP with DAE of index greater than one.

The following theorems will make use of uniform convergence and uniform norm for functions, their definitions are given in Appendix A.

Theorem 1. *Let the functions $\langle x_k, y_k, u_k\rangle$ be global minima of the problem $\mathcal{P}_{\mathcal{L}}(\mu_k, \nu_k)$ (Eq. 6) at each iteration k. In addition, assume that $\{\langle x_k, y_k, u_k\rangle\}_K$ and $\{\nu_k\}_K$ are uniformly convergent subsequences. Then, under Assumption 1, the limiting functions of every subsequences $\{\langle x_k, y_k, u_k\rangle\}_K$ are a global minimizer of problem $\mathcal{P}$ and the subsequence $\{J_{\mu_k}(x_k, y_k, u_k, \nu_k)\}_K$ converges to the optimum objective of $\mathcal{P}$.*

Proof: Let $\langle x^*, y^*, u^*\rangle$ be limiting functions of the subsequence $\{\langle x_k, y_k, u_k\rangle\}_K$. By definition of x_k, y_k, and u_k, for a given k

$$J_{\mu_k}(x_k, y_k, u_k, \nu_k) \leq J_{\mu_k}(x, y, u, \nu_k) \tag{7}$$

for all feasible x, y, and u.

Let J^* denote the optimal value of $\mathcal{P}$. We have that

$$J^* = \min_{\substack{u \\ \text{s.t. (1b)-(1e)}}} J = \min_{\substack{y,u \\ \text{s.t. (6b)-(6d)} \\ g(x,y,u,t)=0}} J_{\mu_k}(\mu_k, \nu_k) \tag{8}$$

the last term implies the minimization of the problem $\mathcal{P}_{\mathcal{L}}$ over y and u with the additional equation $g(x, y, u, t) = 0$. The first equality holds by definition. The second equality holds because $\mathcal{P}$ and $\mathcal{P}_{\mathcal{L}}$ are equivalent when the equation $g(x, y, u, t) = 0$ is included in $\mathcal{P}_{\mathcal{L}}$.

The inequality (7) holds for any x, y, and u, including a minimizer of (8). Therefore, we can substitute the optimum value J^* on the right-hand side of (7), and on the left-hand side we substitute $J_{\mu_k}(x_k, y_k, u_k, \nu_k)$ with its definition to obtain

$$\begin{aligned} \int_{t_0}^{t_f} & L(x_k, y_k, u_k, t) + \nu_k^T g(x_k, y_k, u_k, t) \\ & + \frac{\mu_k}{2} \|g(x_k, y_k, u_k, t)\|^2 \, dt \leq J^* \end{aligned} \tag{9}$$

Given that the subsequence $\{\nu_k\}_K$ is uniformly convergent, it has a limiting function ν^*. By taking the limit with $k \to \infty$ in the inequality (9) we obtain

$$\begin{aligned} \int_{t_0}^{t_f} & \left[L(x^*, y^*, u^*, t) + \nu^{*T} g(x^*, y^*, u^*, t)\right] dt \\ & + \lim_{k\to\infty} \frac{\mu_k}{2} \int_{t_0}^{t_f} \|g(x_k, y_k, u_k, t)\|^2 \, dt \leq J^* \end{aligned} \tag{10}$$

Since $\|g(x_k, y_k, u_k, t)\|^2 \geq 0$ and $\mu_k \to \infty$, it follows that we must have $g(x_k, y_k, u_k, t) \to 0$ and

$$g(x^*, y^*, u^*, t) = 0 \qquad \forall t \in [t_0, t_f] \tag{11}$$

otherwise the limit on the left-hand side of (10) would go to $+\infty$ which does not hold since J^* is finite. Therefore,

$$J(x^*, y^*, u^*) = \int_{t_0}^{t_f} L(x^*, y^*, u^*, t) \, dt \leq J^* \tag{12}$$

Any solution to problem $\mathcal{P}_{\mathcal{L}}$ satisfies all of the constraints of $\mathcal{P}$ except the relaxed algebraic equations. However (11) ensures that the limiting functions x^*, y^*, and u^* do satisfy the algebraic equation. By definition, J^* is less or equal to the objective of any feasible functions for problem $\mathcal{P}$, therefore we have

$$J^* \leq J(x^*, y^*, u^*) \tag{13}$$

Using (12) and (13), we conclude that

$$J^* \leq J(x^*, y^*, u^*) \leq J^* \implies J^* = J(x^*, y^*, u^*) \tag{14}$$

which proves that the limiting functions x^*, y^*, and u^* are global minimizers for problem $\mathcal{P}$ and that $\{J_{\mu_k}(x_k, y_k, u_k, \nu_k)\}_K \to J^*$.

■

Definition 1. *Let $\mathcal{V}$ be a function space, then a nonempty set $\mathcal{V}^* \subset \mathcal{V}$ is said to be an isolated set of local minima of problem $\mathcal{P}$ if each function $v^* \in \mathcal{V}^*$ is a local minimum of problem $\mathcal{P}$ and, for some $\varepsilon > 0$, the set*

$$\mathcal{V}_\varepsilon^* = \{v \in \mathcal{V} : \|v - v^*\| \leq \varepsilon \text{ for some } v^* \in \mathcal{V}^*\} \tag{15}$$

contains no local minima of problem $\mathcal{P}$ other than the functions of $\mathcal{V}^*$.

An isolated set of local minima consisting of a single function is a strict local minimum.

Theorem 2. *Suppose that the regularity Assumption 1 holds, and that $\mathcal{V}^*$ is a compact and isolated set of local minima of problem $\mathcal{P}$. If $\langle x_k, y_k, u_k\rangle$ is a local minimizer for problem $\mathcal{P}_{\mathcal{L}}$ for each k, then there exists a subsequence $\{\langle x_k, y_k, u_k\rangle\}_K$ converging to a limiting function $\langle x^*, y^*, u^*\rangle \in \mathcal{V}^*$. Furthermore, if $\mathcal{V}^*$ consists of a single function $\langle x^*, y^*, u^*\rangle$, then there exists a sequence $\{\langle x_k, y_k, u_k\rangle\}$ such that $\{\langle x_k, y_k, u_k\rangle\} \to \langle x^*, y^*, u^*\rangle$.*

Proof: Consider the set

$$\mathcal{V}_{\tilde{\varepsilon}}^* = \{v \in \mathcal{V} : \|v - v^*\| \leq \tilde{\varepsilon} \text{ for some } v^* \in \mathcal{V}^*\} \tag{16}$$

where $\mathcal{V}$ is the set of feasible functions of $\mathcal{P}_{\mathcal{L}}$, with some $0 < \tilde{\varepsilon} < \varepsilon$, and ε is as in (15). From (16) and because $\mathcal{V}$ is compact by Assumption 1, it follows that $\mathcal{V}_{\tilde{\varepsilon}}^*$ is also compact, and hence the problem

$$\min_{x,y,u} J_{\mu_k} = \int_{t_0}^{t_f} \mathcal{L}_{\mu_k}(x, y, u, \nu_k, t) \, dt \tag{17a}$$

$$\text{s.t.:} \quad \dot{x} = f(x, y, u, t) \qquad \forall t \in [t_0, t_f] \tag{17b}$$

$$u(t) \in U_B \qquad \forall t \in [t_0, t_f] \tag{17c}$$

$$\langle x, y, u\rangle \in \mathcal{V}_{\tilde{\varepsilon}}^*, \quad x(t_0) = x_0 \tag{17d}$$

has a global minimum $\langle x_k, y_k, u_k\rangle \in \mathcal{V}^*_{\bar{\varepsilon}}$. By Theorem 1, every limiting function $\langle x^*, y^*, u^*\rangle$ of $\{\langle x_k, y_k, u_k\rangle\}_K$ is a global minimum of the problem

$$\min_{x,y,u} J = \int_{t_0}^{t_f} L(x,y,u,t)\,dt \tag{18a}$$
$$\text{s.t.: } \dot{x} = f(x,y,u,t) \qquad \forall t \in [t_0, t_f] \tag{18b}$$
$$g(x,y,u,t) = 0 \qquad \forall t \in [t_0, t_f] \tag{18c}$$
$$u(t) \in U_B \qquad \forall t \in [t_0, t_f] \tag{18d}$$
$$\langle x, y, u\rangle \in \mathcal{V}^*_{\bar{\varepsilon}}, \quad x(t_0) = x_0 \tag{18e}$$

Furthermore, each global minimum of the problem above must belong to $\mathcal{V}^*$ by the definition of $\mathcal{V}^*_{\bar{\varepsilon}}$. Thus there is a subsequence $\{\langle x_k, y_k, u_k\rangle\}_K$ converging to $\langle x^*, y^*, u^*\rangle \in \mathcal{V}^*$. If $\mathcal{V}^*$ contains only one local optimum, then all the subsequences will lead to this local optimum, therefore $\{\langle x_k, y_k, u_k\rangle\} \to \langle x^*, y^*, u^*\rangle \in \mathcal{V}^*$. ■

From a practical point of view, numerical methods are expected to terminate when the optimality conditions of $\mathcal{P}_\mathcal{L}$ are almost satisfied, meaning that for a small scalar $\varepsilon_k > 0$ the necessary optimality conditions [7] are

$$\|f(x_k, y_k, u_k, t) - \dot{x}\| \leq \varepsilon_k, \tag{19a}$$

$$\left\| \frac{\partial \mathcal{L}_{\mu_k}}{\partial x}(x_k, y_k, u_k, \nu_k, t)^T + \right. \left. + \frac{\partial f}{\partial x}(x_k, y_k, u_k, t)^T \lambda_k + \dot{\lambda}_k \right\| \leq \varepsilon_k, \tag{19b}$$

$$\left\| u_k(t) - \arg\inf_{\mathrm{u}\in U_B} H(x_k(t), \lambda_k(t), y_k, \nu_k, \mathrm{u}, t) \right\| \leq \varepsilon_k \tag{19c}$$

$$\left\| \frac{\partial \mathcal{L}_{\mu_k}}{\partial y}(x_k, y_k, u_k, \nu_k, t)^T + \frac{\partial f}{\partial y}(x_k, y_k, u_k, t)^T \lambda_k \right\| \leq \varepsilon_k. \tag{19d}$$

The following theorem shows that if $\varepsilon_k \to 0$, the algorithm still converges.

Theorem 3. *Suppose that Assumption 1 holds and let $\langle x_k, y_k, u_k\rangle$ be a suboptimal solution obtained for $\mathcal{P}_\mathcal{L}(\mu_k, \nu_k)$ such that the violation of the optimality conditions are given by* (19)*, for which inequality* (19d) *is fundamental, where $0 \leq \varepsilon_k$, and $\varepsilon_k \to 0$ as $k \to \infty$, $\{\nu_k\}$ is a uniform convergent sequence, and λ_k is the costate at the k-th algorithm iteration. Assume that a subsequence $\{\langle x_k, y_k, u_k\rangle\}_K$ converges uniformly to $\langle x^*, y^*, u^*\rangle$ such that $\frac{\partial g}{\partial y}(x^*, y^*, u^*, t)$ has full rank and is bounded for all $t \in [t_0, t_f]$.*

Then the subsequence $\{\nu_k + \mu_k g(x_k, y_k, u_k, t)\}_K$ converges uniformly to $\widetilde{\nu}^*$, such that the following relations are obtained, with respect to y

$$\frac{\partial L}{\partial y}(x^*, y^*, u^*, t)^T + \frac{\partial f}{\partial y}(x^*, y^*, u^*, t)^T \lambda^* + \frac{\partial g}{\partial y}(x^*, y^*, u^*, t)^T \widetilde{\nu}^* = 0 \tag{20a}$$

and with respect to λ, u, and x are

$$-\dot{\lambda}^* = \frac{\partial L}{\partial x}(x^*, y^*, u^*, t)^T + \frac{\partial f}{\partial x}(x^*, y^*, u^*, t)^T \lambda^* + \frac{\partial g}{\partial x}(x^*, y^*, u^*, t)^T \widetilde{\nu}^* \tag{20b}$$
$$u^*(t) = \arg\inf_{\mathrm{u}\in U_B} H(x^*(t), \lambda^*(t), y^*, \nu^*, \mathrm{u}, t) \tag{20c}$$
$$\dot{x}^* = f(x^*, y^*, u^*, t). \tag{20d}$$

Proof: The derivative of $\mathcal{L}_{\mu_k}$ w.r.t. y results in

$$\frac{\partial \mathcal{L}_{\mu_k}}{\partial y}(x_k, y_k, u_k, \nu_k, t) = \frac{\partial L}{\partial y}(x_k, y_k, u_k, t) + [\nu_k + \mu_k g(x_k, y_k, u_k, t)]^T \frac{\partial g}{\partial y}(x_k, y_k, u_k, t) \tag{21}$$

Then, by defining for all k

$$\widetilde{\nu}_k = \nu_k + \mu_k g(x_k, y_k, u_k, t) \tag{22}$$

replacing $\widetilde{\nu}_k$ into (21) results in

$$\frac{\partial \mathcal{L}_{\mu_k}}{\partial y}(x_k, y_k, u_k, \nu_k, t) = \frac{\partial L}{\partial y}(x_k, y_k, u_k, t) + \widetilde{\nu}_k^T \frac{\partial g}{\partial y}(x_k, y_k, u_k, t). \tag{23}$$

Since $\frac{\partial g}{\partial y}$ is invertible, we can derive the following expression for $\widetilde{\nu}_k$,

$$\widetilde{\nu}_k = \left[\frac{\partial g}{\partial y}(x_k, y_k, u_k, t)^T\right]^{-1} \left[\frac{\partial \mathcal{L}_{\mu_k}}{\partial y}(x_k, y_k, u_k, \nu_k, t)^T - \frac{\partial L}{\partial y}(x_k, y_k, u_k, t)^T\right] \tag{24}$$

From (24) we can say that there exists an F such that

$$\widetilde{\nu}_k = F(x_k, y_k, u_k, \nu_k) \tag{25}$$

which is continuous since all the functions in (24) are continuous. Given that a subsequence $\{\langle x_k, y_k, u_k\rangle\}_K$ converges to $\langle x^*, y^*, u^*\rangle$ and $\{\nu_k\}$ converges to ν^*, Theorem 4 (from Appendix A) is invoked to conclude that

$$\{\widetilde{\nu}_k = F(x_k, y_k, u_k, \nu_k)\}_K \to \widetilde{\nu}^* = F(x^*, y^*, u^*, \nu^*) \tag{26}$$

which shows that $\{\nu_k + \mu_k g(x_k, y_k, u_k, t)\}_K \to \widetilde{\nu}^*$ uniformly, and $\widetilde{\nu}^*$ is given by

$$\widetilde{\nu}^* = \left[\frac{\partial g}{\partial y}(x^*, y^*, u^*, t)^T\right]^{-1} \left[\frac{\partial \mathcal{L}_{\mu^*}}{\partial y}(x^*, y^*, u^*, \nu^*, t)^T - \frac{\partial L}{\partial y}(x^*, y^*, u^*, t)^T\right]. \tag{27}$$

Considering the optimality conditions for y, given in (19d), and taking the limit $k \to \infty$, we obtain

$$\frac{\partial \mathcal{L}_{\mu^*}}{\partial y}(x^*, y^*, u^*, \nu^*, t) = -\lambda^{*T} \frac{\partial f}{\partial y}(x^*, y^*, u^*, t) \tag{28}$$

which can be substituted into (27) to obtain

$$\widetilde{\nu}^* = \left[\frac{\partial g}{\partial y}(x^*, y^*, u^*, t)^T\right]^{-1} \left[-\frac{\partial L}{\partial y}(x^*, y^*, u^*, t)^T - \frac{\partial f}{\partial y}(x^*, y^*, u^*, t)^T \lambda^*\right] \quad (29)$$

which can be rearranged into

$$\frac{\partial L}{\partial y}(x^*, y^*, u^*, t) + \lambda^{*T}\frac{\partial f}{\partial y}(x^*, y^*, u^*, t) + \widetilde{\nu}^{*T}\frac{\partial g}{\partial y}(x^*, y^*, u^*, t) = 0 \quad (30)$$

and related to the necessary conditions (3d) of the original OCP $\mathcal{P}$. Similar approach can be used to obtain the conditions for x, u, and λ.

Since the sequence $\{\nu_k\}$ is bounded and $\{\nu_k + \mu_k g(x_k, y_k, u_k, t)\}_K \to \widetilde{\nu}^*$ from (26), it follows that $\{\mu_k g(x_k, y_k, u_k, t)\}_K$ is bounded. Given that $\mu_k \to \infty$ we must have $g(x_k, y_k, u_k, t) \to 0$ with $g(x^*, y^*, u^*, t) = 0$ for all t. ■

Notice that the sequence $\{\nu_k\}$ was never specified, other than it is a uniformly convergent sequence. From Theorem 3, an update rule can be derived such that $\{\nu_k\} \to \widetilde{\nu}^*$.

Corollary 1. *By defining* $\nu_{k+1} = \nu_k + \mu_k g(x_k, y_k, u_k, t)$ *we have that* $\{\nu_k\} \to \widetilde{\nu}^*$ *and* $\{\mu_k g(x_k, y_k, u_k, t)\} \to 0$.

Proof: For any uniformly convergent sequence $\{\nu_k\}$, Theorem 3 ensures that $\{\nu_k + \mu_k g(x_k, y_k, u_k, t)\} \to \widetilde{\nu}^*$. Therefore, we can define $\nu_{k+1} = \nu_k + \mu_k g(x_k, y_k, u_k, t)$, which makes the sequence become $\{\nu_{k+1}\} \to \widetilde{\nu}^*$. ■

IV. Computational Details and Experiments

To illustrate the algorithm behavior and to remark some implementation details, the algorithm is applied to the optimal control problem of stabilizing a four-tank system.

A. Problem Modeling

Instead of modeling the four tank as an ODE system as it is commonly done, this paper represents the process as a DAE system by using an algebraic variable for the outflow of each tank. For every tank i, the outflow is given by

$$q_{t,i} = a_i\sqrt{2gh_i}. \quad (31)$$

where g is the gravity constant, a_i is the cross section area of the orifice and h_i is the fluid level of tank i, given by

$$\dot{h}_1 = \frac{q_3 + \gamma_1 q_{p,1} - q_{t,1}}{A_1}, \dot{h}_2 = \frac{q_4 + \gamma_2 q_{p,2} - q_{t,2}}{A_2} \quad (32a)$$

$$\dot{h}_3 = \frac{(1-\gamma_2) q_{p,2} - q_{t,3}}{A_3}, \dot{h}_4 = \frac{(1-\gamma_1) q_{p,1} - q_{t,4}}{A_4} \quad (32b)$$

where A_i is the cross section area of the tank i, and the flow on each pump j is given by the differential equation

$$\dot{q}_{p,j} = \delta_j. \quad (33)$$

The objective of the controller is to stabilize the tanks 1 and 2, while reducing the variation in the pump flows, which is expressed by the following objective

$$\min_u J = \int_{t_0}^{t_f} \Delta x^T \Delta x + u^T u \, dt \quad (34)$$

with $\Delta x = x - x_{ref}$, $x = [h_1, h_2, h_3, h_4, q_{p,1}, q_{p,2}]$, and $u = [\delta_1, \delta_2]$.

B. Applying the Augmented Lagrangian

By using the algorithm to relax the algebraic equation (31), the following relaxed problem is obtained

$$\min_{u,y} \quad J_{\mu_k}, \qquad \text{s.t.: eq. (32) and (33)} \quad (35)$$

where:

$$J_{\mu_k} = \int_{t_0}^{t_f} \Delta x^T \Delta x + u^T u + \sum_{i=1}^{4}\left[\nu_{i,k}\left(q_{t,i} - a_i\sqrt{2gh_i}\right) + \frac{\mu_k}{2}\left\|q_{t,i} - a_i\sqrt{2gh_i}\right\|^2\right] dt \quad (36)$$

At each algorithm iteration, the relaxed problem (35) is solved and the solution is used to compute the new multiplier estimates $\nu_{i,k+1}$ according to the rule

$$\nu_{i,k+1} = \nu_{i,k} + \mu_k\left[q_{t,i} - a_i\sqrt{2gh_i}\right] \quad (37)$$

As discussed in [4], since $\nu_{i,k}$ is a function that can assume any shape, a piecewise polynomial approximation with a finite number of terms is used instead. For this application, the Lagrangian polynomial was chosen as it facilitates the computation of updates.

C. Computational Experiments

To solve the relaxed subproblem, an indirect collocation method with polynomials of order 3 was used, discretized in 30 finite elements and implemented using YAOCP-Tool and CasADi [12]. The same settings were used to solve the original problem (Eqs. (31)-(34)) for comparison purpose. The resulting nonlinear programming problems were solved with the IPOPT solver. By using indirect methods, the multipliers are easily obtained which allow us to compare with the estimate ν_i yielded by our algorithm. The multiplier estimates ν_i are also approximated with a piecewise polynomial of degree 3 with 30 finite elements.

Since no information on the multipliers is available the algorithm is initialized with the multiplier estimates as zero ($\nu_0 = 0$ for all $t \in [t_0, t_f]$). The penalization term starts with $\mu_0 = 0.1$ and increases at a rate $\beta = 4$.

Although not shown here, the trajectories of the proposed algorithm coincide with the optimal trajectories obtained by the indirect method. To evaluate if the algorithm is converging to the optimal solution of the original problem, in Fig. 1, the relaxed objective (J_{μ_k}) and the evaluation of the solution iteration on the original objective (J_k) are

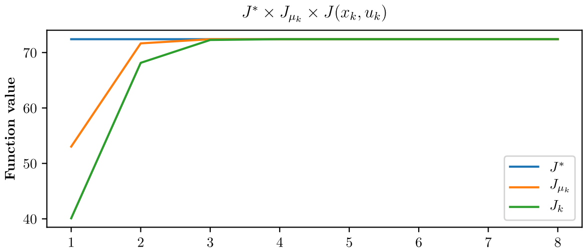

Figure 1. Comparison of the objective functions.

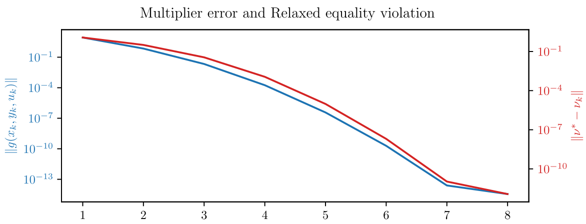

Figure 2. Convergence of the algebraic function to zero (red) and the multiplier estimate converging to the original problem multiplier (blue).

compared to the optimal cost obtained with the indirect method (J^*). It can be seen that the objectives converge to the same objective value J^* as the indirect method. As for the violation of algebraic equations, the line in blue of Fig. 2 shows the violation rapidly converging to zero; the line in red shows the norm of the difference between the multiplier obtained with the indirect method and the multiplier estimate computed by the proposed algorithm, which decreases as the algorithm iterates.

An experiment was performed using the proposed algorithm with direct multiple shooting to solve the subproblems (AL-DMS). To solve the ODE of the suproblems, we apply Sundails CVODES and a 4^{th} order Runge-Kutta method (RK4). The proposed algorithm was compared against DMS applied directly to the original OCP, whereby Sundails IDAS was used to solve the DAE. Table I presents the results that indicate faster convergence of proposed algorithm with RK4, which might be more suitable for embedded applications with restricted computational power.

Table I
ALGORITHM SOLVING TIME (S)

AL-DMS (CVODES)	AL-DMS (RK4)	DMS (IDAS)
24.15	0.11	2.42

V. CONCLUSION

An algorithm based on the augmented Lagrangian for solving OCPs of DAEs by relaxing the algebraic equations was presented. Properties for global and local convergence, and convergence under sub-optimal iterations are shown. The algorithm properties are verified with an illustrative example, showing the convergence of the objective value, multiplier estimate, and optimal trajectory.

The proposed algorithm can be deployed for optimal control in applications where DAE solvers are too costly, which are supported by the convergence conditions established heretofore. As future work, the structure of certain problems could be exploited in the augmented Lagrangian to enable distributed and parallel computations.

APPENDIX A

Definition 2. *Let $f : [t_0, t_f] \to \mathbb{R}^N$ be a continuous function then $\|f\|$ is given by $\|f\| = \max_{t\in[t_0,t_f]} \|f(t)\|_\infty$.*

Definition 3. *Let $f_k : [t_0, t_f] \to \mathbb{R}^N$ be a function for every $k \in \mathbb{N}$. The sequence of functions $\{f_k\}$ converges uniformly to the limiting function $f^* : [t_0, t_f] \to \mathbb{R}^N$ if, for every $\varepsilon > 0$, there exists a number $K \in \mathbb{N}$ such that for all $t \in [t_0, t_f]$ and all $k \geq K$, we have $\|f_k(t) - f^*(t)\| < \varepsilon$.*

Theorem 4. *Let $g : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ be a continuous function, and the sequence of functions $\{f_n\}$ to converge uniformly to f, where $f_n : [0, 1] \to \mathbb{R}^{d_1}$. Let the function norm $\|\cdot\|$ be given by $\|g\| = \max_{x\in[0,1]} \|g(x)\|_\infty$. Then $\{g(f_n)\}$ converges uniformly to $g(f)$.*

Proof: If f_n converges uniformly to f, then for all ε_f exists N, such that $\|f_n - f\| < \varepsilon_f$ for all $n > N$, and exists an upper bound M s.t. $\|f_n\| \leq M$ for all $n \in \mathbb{N}$.

Then, consider $g : [-M, M]^{d_1} \to \mathbb{R}^{d_2}$. As g is continuous in a compact set, for all $\varepsilon_g > 0$, there exists a $\delta_g > 0$ such that $\|g(z_1) - g(z_2)\| < \varepsilon_g$ for all $\|z_1 - z_2\| < \delta_g$. Using $\varepsilon_f = \delta_g$, $\|f_n - f\| < \varepsilon_f = \delta_g$ for all $n > N$. Therefore, $\|g(f_n) - g(f)\| < \varepsilon_g$ for all $n > N$. ■

REFERENCES

[1] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, 1996.
[2] R. R. Negenborn, B. De Schutter, and J. Hellendoorn, “Multi-agent model predictive control for transportation networks: Serial versus parallel schemes,” *Eng. Appl. Artif. Intell.*, vol. 21, no. 3, pp. 353–366, Apr. 2008.
[3] N. Chatzipanagiotis, D. Dentcheva, and M. M. Zavlanos, “An augmented Lagrangian method for distributed optimization,” *Mathematical Programming*, vol. 152, no. 1, pp. 405–434, Aug 2015.
[4] M. A. S. de Aguiar, E. Camponogara, and B. Foss, “An augmented Lagrangian method for optimal control of continuous time DAE systems,” in *IEEE CCA*, 2006, pp. 1185–1190.
[5] S. Hentzelt and K. Graichen, “An augmented Lagrangian method in distributed dynamic optimization based on approximate neighbor dynamics,” in *IEEE SMC*, 2013, pp. 571–576.
[6] H. Attouch and M. Soueycatt, “Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. applications to games, PDE’s and control,” *Pacific Journal of Optimization*, vol. 5, no. 1, pp. 17–37, Nov. 2008.
[7] D. Kirk, *Optimal Control Theory: An Introduction*, 4th ed. Dover Publications, 2004.
[8] L. T. Biegler, *Nonlinear Programming*. SIAM, jan 2010.
[9] M. R. de Pinho, R. B. Vinter, and H. Zheng, “A maximum principle for optimal control problems with mixed constraints,” *IMA Journal of Math. Control and Information*, vol. 18, pp. 189–205, 2001.
[10] A. Boccia, M. D. R. de Pinho, and R. B. Vinter, “Optimal control problems with mixed and pure state constraints,” *SIAM Journal on Control and Optimization*, vol. 54, no. 6, pp. 3061–3083, 2016.
[11] J. T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. Cambridge Univ. Press, 2009.
[12] J. Andersson, J. Gillis, G. Horn, J. Rawlings, and M. Diehl, “CasADi – A software framework for nonlinear optimization and optimal control,” *Math. Progr. Comp.*, vol. 11, no. 1, pp. 1–36, 2019.