2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Linearized ADMM for nonsmooth nonconvex optimization with nonlinear equality constraints

Lahcen El Bourkhissi[1], Ion Necoara[1,2] and Panagiotis Patrinos[3]

Abstract— **This paper proposes a new approach for solving a structured nonsmooth nonconvex optimization problem with nonlinear equality constraints, where both the objective function and constraints are 2-blocks separable. Our method is based on a 2-block linearized ADMM, where we linearize the smooth part of the cost function and the nonlinear term of the functional constraints in the augmented Lagrangian at each outer iteration. This results in simple subproblems, whose solutions are used to update the iterates of the 2 blocks variables. We prove global convergence for the sequence generated by our method to a stationary point of the original problem. To demonstrate its effectiveness, we apply our proposed algorithm as a solver for the nonlinear model predictive control problem of an inverted pendulum on a cart.**

I. INTRODUCTION

Many applications in the recently most concerned fields such as nonlinear model predictive control, machine learning, and signal processing can be formulated as the following structured nonconvex nonsmooth optimization problem with nonlinear equality constraints [10], [16], [18]:

$$\begin{aligned} \min_{x\in\mathbb{R}^n, y\in\mathbb{R}^p} \quad & f(x)+g(x)+h(y) \\ \text{s.t.} \quad & F(x)+Gy=0, \end{aligned} \tag{1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ and $F(x)\triangleq(f_1(x),...,f_m(x))^T$, with $f_i:\mathbb{R}^n\to\mathbb{R}$, for all $i\in\{1,...,m\}$ nonlinear functions. We assume $f,h,f_i\in\mathcal{C}^2$, for all $i=1,...,m$, f,h possibly nonconvex, g is a nonsmooth but proximal-friendly function, i.e., the proximal operator is easy to compute. Moreover, we require matrix $G\in\mathbb{R}^{m\times p}$ to have full row rank i.e., there exists $\sigma>0$ such that $\sigma_{\min}(G^T)\geq\sigma$, where $\sigma_{\min}(G^T)$ denotes the smallest singular value of the matrix G^T.

In this paper, we propose an augmented Lagrangian approach to address problem (1). The augmented Lagrangian method, or method of multipliers, was introduced in [13], [15] to minimize an objective function under equality constraints. It provides many theoretical advantages, even for non-convex problems (e.g., no duality gap and exact penalty representation), highlighted by Rockafellar in [22]. Researchers started to investigate the augmented Lagrangian framework in detail when the Alternating Direction Method of Multiplier (ADMM) showed strong advantages (recall that the augmented Lagrangian framework is at the heart of the ADMM), see [4]–[6], [8], [11], [12], [25].

In the literature, the augmented Lagrangian approach has been widely studied for convex problems, see e.g., [1]–[3], [6] and references therein, and recently its extension to non-convex (smooth/non-smooth) problems with linear constraints have been proposed in [14], [16], [17], [19]–[21], [25], [26]. However, there are very few studies on the use of the augmented Lagrangian framework for nonconvex optimization with nonlinear constraints, such as [8], [9], [24]. In particular, in [24], a proximal augmented Lagrangian (Proximal AL) algorithm is proposed to solve the problem (1); in this method, a static proximal term is added to the original augmented Lagrangian function. It is proved that when an approximate first- (second-) order solution of the subproblem is found, then an ϵ first- (second-) order solution of the original problem (1) is obtained within $\mathcal{O}(1/\epsilon^{2-\eta})$ outer iterations, for some parameter $\eta\in[0,2]$. Note that when η is close to 2, the efficiency is reduced to $\mathcal{O}(1)$ outer iterations, but the subproblem, which is already non-convex, becomes very ill-conditioned as the penalty parameter of the augmented Lagrangian is inversely proportional to ϵ^η. Furthermore, in [8], the authors proposed an augmented Lagrangian-based method to deal with the same problem considered in this paper and under the same assumptions. In that algorithm, the authors linearize the smooth part of the augmented Lagrangian function and added a dynamic quadratic regularization, and prove global convergence of their method using Lyapunov analysis. Finally, in [9] a linearized augmented Lagrangian method is proposed, where the functional constraints are linearized within the augmented Lagrangian, while keeping the objective function unchanged, yielding a simple subproblem at each iteration. Under mild assumptions, it is proved that the iterates converge globally to a first-order optimality point within a rate of order $\mathcal{O}(1/\epsilon^2)$.

In addition to the high cost of solving the subproblem in [24], neither [24] nor [9] employ specific schemes that can leverage the unique structure of the problem, such as its separability. Moreover, in contrast to the large approximation error in the model considered in [8], which arises from linearizing the entire smooth part of the augmented Lagrangian function, in this paper we propose a linearized two-block alternating direction method of multipliers (linearized ADMM) to solve problem (1). Specifically, in our algorithm, we linearize the smooth part of the objective function and the nonlinear term of the functional constraints in the augmented

[1]Department of Automatic Control and Systems Engineering, Universitatea Nationala de Stiinta si Tehnologie Politehnica Bucuresti, 060042 Bucharest, Romania. lel@stud.acs.upb.ro

[2]Gheorghe Mihoc-Caius Iacob Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, 050711 Bucharest, Romania. ion.necoara@upb.ro

[3]Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, 3001 Leuven, Belgium. panos.patrinos@esat.kuleuven.be

Lagrangian function, and we add a dynamic regularization. Thus, our method borrows the advantages of methods from [8] since it takes advantage of the separability structure of the problem at hand and from [9] since it considers more accurate model approximations by linearizing inside the augmented Lagrangian. We prove global convergence of the iterates to a stationary point of the original problem (1). Furthermore, we compare the efficiency of the proposed method with IPOPT [23] and the method in [8] to solve the predictive control problem of a nonlinear model of an inverted pendulum on a cart.

The paper is structured as follows. In the next section, we introduce some mathematical preliminaries, in section III we present the linearized ADMM method followed in section IV by its convergence analysis. Finally, section V presents some preliminary numerical results.

II. Preliminaries

We use $\|.\|$ to denote the $2-$norm of a vector or of a matrix, respectively. For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we denote by $\nabla f(x) \in \mathbb{R}^n$ its gradient at a point x. For a differentiable function $F : \mathbb{R}^n \to \mathbb{R}^m$, we denote its Jacobian at a given point x by $\nabla F(x) \in \mathbb{R}^{m\times n}$. We further introduce the following notations:

$$l_f(x;\bar{x}) := f(\bar{x}) + \langle \nabla f(\bar{x}), x-\bar{x}\rangle \qquad \forall x,\bar{x},$$
$$l_h(y;\bar{y}) := h(\bar{y}) + \langle \nabla h(\bar{y}), y-\bar{y}\rangle \qquad \forall y,\bar{y},$$
$$l_F(x;\bar{x}) := F(\bar{x}) + \nabla F(\bar{x})(x-\bar{x}) \qquad \forall x,\bar{x}.$$

Let $(x^*,y^*) \in \mathbb{R}^n \times \mathbb{R}^m$ be a local minimizer of (1) such that x^* satisfies the following constraint qualification [8]:

(i) The function f is subdifferential regular at x^*.
(ii) $\partial^\infty f(x^*) \cap \operatorname{range} \nabla F(x^*)^T = \{0\}$.

Then, there exists a KKT point for problem (1) at (x^*,y^*). Note that KKT points are critical points of the Lagrangian function. The main purpose of this work is to devise an algorithm that converges to such a point. Let's also consider the following assumption, which is used in our analysis of the optimization problem (1), throughout the paper:

Assumption 2.1: Given a compact set $\mathcal{S} \subseteq \mathbb{R}^n$, there exist positive constants $M_f, M_h, M_F, \sigma, L_f, L_h, L_F$ such that the functions f, h and F satisfy the following conditions for all $x, y \in \mathcal{S}$ (a given compact set):

(i) $\|\nabla f(x)\| \le M_f$,
(ii) $\|\nabla f(x) - \nabla f(y)\| \le L_f\|x-y\|$,
(iii) $\|\nabla h(x)\| \le M_h$,
(iv) $\|\nabla h(x) - \nabla h(y)\| \le L_h\|x-y\|$,
(v) $\|\nabla F(x)\|_2 \le M_F$,
(vi) $\|\nabla F(x) - \nabla F(y)\|_2 \le L_F\|x-y\|$,

Note that these assumptions are standard in nonconvex optimization, see e.g., [8], [9], [24]. In fact, it covers a large class of problems; more precisely, (i), (ii), (iii) and (iv) hold if f and h are smooth and $\nabla f, \nabla h$ are locally Lipschitz continuous on a neighborhood of $\mathcal{S}$. Similarly, (v) and (vi) are valid if F is smooth and ∇F is locally Lipschitz continuous on a neighborhood of $\mathcal{S}$. Note that these assumptions are not very restrictive because they are satisfied locally for any $f, h, F \in \mathcal{C}^2$ and for any matrix G having full rank. We further introduce the following notations:

$$\psi_\rho(x,y,\lambda) = f(x) + \langle \lambda, F(x)+Gy\rangle + \frac{\rho}{2}\|F(x)+Gy\|^2.$$

Then, the augmented Lagrangian associated with (1) is:

$$\begin{aligned}&\mathcal{L}_\rho(x,y,\lambda)\\ &= f(x)+g(x)+h(y)+\langle\lambda,F(x)+Gy\rangle+\tfrac{\rho}{2}\|F(x)+Gy\|^2\\ &= g(x)+h(y)+\psi_\rho(x,y,\lambda).\end{aligned}$$

Given a pair $(\bar{x},\bar{\lambda})$, we denote:

$$\begin{aligned}\bar{\mathcal{L}}_\rho(x,y,\lambda;\bar{x},\bar{y}) &= l_f(x;\bar{x}) + g(x) + l_h(y;\bar{y})\\ &+ \langle\lambda, l_F(x;\bar{x})+Gy\rangle + \frac{\rho}{2}\|l_F(x;\bar{x})+Gy\|^2.\end{aligned}$$

The gradient of ψ_ρ is:

$$\begin{cases}\nabla_x\psi_\rho(x,y,\lambda) = \nabla f(x) + \nabla F(x)^T\left(\lambda + \rho\left(F(x)+Gy\right)\right),\\ \nabla_y\psi_\rho(x,y,\lambda) = G^T\left(\lambda+\rho\left(F(x)+Gy\right)\right),\\ \nabla_\lambda\psi_\rho(x,y,\lambda) = F(x)+Gy.\end{cases}$$

Note that from Assumption 2.1 it follows that ψ_ρ is smooth, (i.e., it has Lipschitz continuous gradient).

Before we present our algorithm, let us discuss an alternative approach to address problem (1) that involves eliminating the constraints and the second block of the primal variables y. Indeed, this can be achieved using the relation $y = -G^\dagger F(x)$, where $G^\dagger$ represents the pseudoinverse of matrix G. By applying this approach, we obtain the following equivalent problem for (1):

$$\min_{x\in\mathbb{R}^n} f(x) + g(x) + h(-G^\dagger F(x))$$

Reformulating the original problem as above may bring some potential advantages. For example, it reduces the dimension of the problem and can be more suitable for some optimization methods, such as proximal gradient, as they do not require handling constraints explicitly. However, there are also disadvantages for the above formulation. For example, eliminating the variable y, the (sparse) structure of the original problem will be lost. Furthermore, this reformulation might induce ill-conditioning since the pseudoinverse of a matrix can be ill-conditioned and consequently affecting the numerical stability and accuracy of the solution. As in [8], in the following we consider the original structured formulation (1).

III. Linearized ADMM

In this section we propose an augmented Lagrangian based method (Algorithm 1), which is similar to the one proposed in [8] but with a difference in the update of the first block of primal variables (see Step 5 below). In [8], the authors linearized the smooth part of the augmented Lagrangian function ψ_ρ and added a dynamic quadratic regularization, while in our paper we linearize the nonlinear functional constraints inside the augmenting function, which makes our model to approximate better the original augmented Lagrangian function than the model considered in [8]. We will also see in simulations that this more accurate approximation yields a better behavior in practice.

Algorithm 1 Linearized ADMM

1: **Initialization:** $x_0, y_0, \lambda_0,$ and $\rho, \theta_0 > 0$
2: $k \leftarrow 0$
3: **while** stopping criterion is not satisfied **do**
4: generate a proximal parameter $\beta_{k+1} > 0$
5: $x_{k+1} \leftarrow \arg\min_x \bar{\mathcal{L}}_\rho(x, y_k, \lambda_k; x_k, y_k) + \frac{\beta_{k+1}}{2}\|x - x_k\|^2$
6: generate a proximal parameter $\theta_{k+1} \geq \theta_0$
7: $y_{k+1} \leftarrow \arg\min_y l_h(y : y_k) + \psi_\rho(x_{k+1}, y, \lambda_k) + \frac{\theta_{k+1}}{2}\|y - y_k\|^2$
8: $\lambda_{k+1} \leftarrow \lambda_k + \rho\left(F(x_{k+1}) + Gy_{k+1}\right)$
9: $k \leftarrow k + 1$
10: **end while**

Note that the dominant step in Algorithm 1 is Step 5, as it involves the nonsmooth function g in addition to a quadratic term. When g is convex or weakly convex, the objective function of the subproblem in Step 5 is strongly convex. The objective function of the subproblem in Step 7 of Algorithm 1 is always a strongly convex quadratic function, even when h is nonconvex. Therefore, solving this subproblem is equivalent to solving a linear system. The dual variables are updated in Step 8 using the conventional update of the dual in traditional augmented Lagrangian-based methods. Let us denote the difference of the steps in x, y and λ, for all $k \geq 1$ as:

$$\Delta x_k = x_k - x_{k-1}, \ \ \Delta y_k = y_k - y_{k-1} \text{ and } \Delta\lambda_k = \lambda_k - \lambda_{k-1}.$$

Let $\nu > 0$ be a user-defined parameter. By using the fact that ψ_ρ is smooth, we can always find β_{k+1} satisfying:

$$\psi_\rho(x_{k+1}, y_k, \lambda_k) - \psi_\rho(x_k, y_k, \lambda_k) - \langle \nabla_x \psi_\rho(x_k, y_k, \lambda_k), x_{k+1} - x_k \rangle \leq \frac{\beta_{k+1} - \nu}{2}\|x_{k+1} - x_k\|^2 \quad \forall k \geq 0. \tag{2}$$

Similarly, let $\theta_0 > 0$ and $\alpha \in (0, 1)$ be a user-defined parameters. Since h is smooth, one can always find, for any $k \geq 0$, θ_{k+1} satisfying:

$$h(y_{k+1}) - l_h(y_{k+1}; y_k) \leq \tfrac{(1-\alpha)\theta_{k+1}}{2}\|y_{k+1} - y_k\|^2. \tag{3}$$

Note that for any $k \geq 1, \beta_k$ and θ_k are well defined since ψ_ρ is smooth, as shown in Lemma 4.1 [8] and h is smooth from Assumption 2.1. To determine these regularization parameters, one approach is to use a backtracking scheme, as described in [8]. Assuming the iterates $\{y_k\}_{k\geq 1}$ are bounded, we have, for any $k \geq 1$, $\theta_k \leq \frac{L_h}{1-\alpha}$. Going forward, we will assume that β_k satisfies the following condition:

Assumption 3.1: Assume that the sequence $\{\beta_k\}_{k\geq 1}$ is bounded, i.e.,

$$\beta := \sup_{k\geq 1} \beta_k < \infty.$$

IV. Convergence analysis

In this section, we derive the asymptotic convergence of the proposed scheme (Algorithm 1). Let us start by proving the decrease with respect to the first argument for the augmented Lagrangian function.

Lemma 4.1: [Descent of $\mathcal{L}_\rho$ w.r.t. the first block of primal variables] If Assumption 2.1 holds, then for all $k \geq 0$ we have the following:

$$\mathcal{L}_\rho(x_{k+1}, y_k, \lambda_k) \leq \mathcal{L}_\rho(x_k, y_k, \lambda_k) - \frac{\nu}{2}\|x_{k+1} - x_k\|^2.$$

Proof: Using the definition of x_{k+1}, we have:

$$\begin{aligned}&\bar{\mathcal{L}}_\rho(x_{k+1}, y_k, \lambda_k; x_k, y_k) + \frac{\beta_{k+1}}{2}\|x_{k+1} - x_k\|^2 \\ &\leq \bar{\mathcal{L}}_\rho(x_k, y_k, \lambda_k; x_k, y_k) = \mathcal{L}_\rho(x_k, y_k, \lambda_k).\end{aligned}$$

Further, from definition of $\bar{\mathcal{L}}_\rho$ and $\mathcal{L}_\rho$, we get:

$$\begin{aligned}&l_f(x_{k+1}; x_k) + g(x_{k+1}) + \langle \lambda_k, l_F(x_{k+1}; x_k) \rangle \\ &\quad + \frac{\rho}{2}\|l_F(x_{k+1}; x_k) + Gy_k\|^2 \\ &\leq f(x_k) + g(x_k) + \langle \lambda_k, F(x_k) \rangle \\ &\quad + \frac{\rho}{2}\|F(x_k) + Gy_k\|^2 - \frac{\beta_{k+1}}{2}\|\Delta x_{k+1}\|^2.\end{aligned}$$

Rearranging the above inequality, it follows:

$$\begin{aligned}&g(x_{k+1}) - g(x_k) \\ &\leq - \langle \nabla f(x_k), \Delta x_{k+1} \rangle - \langle \nabla F(x_k)\Delta x_{k+1}, \lambda_k \rangle \\ &\quad - \frac{\rho}{2}\langle \nabla F(x_k)\Delta x_{k+1}, 2(F(x_k) + Gy_k) \rangle \\ &\quad - \frac{\rho}{2}\langle \nabla F(x_k)\Delta x_{k+1}, \nabla F(x_k)\Delta x_{k+1} \rangle - \frac{\beta_{k+1}}{2}\|\Delta x_{k+1}\|^2 \\ &= - \frac{\rho}{2}\|\nabla F(x_k)\Delta x_{k+1}\|^2 - \langle \nabla f(x_k), \Delta x_{k+1} \rangle \\ &\quad - \langle \nabla F(x_k)^T(\lambda_k + \rho(F(x_k) + Gy_k), \Delta x_{k+1} \rangle \\ &\quad - \frac{\beta_{k+1}}{2}\|\Delta x_{k+1}\|^2 \\ &\leq - \langle \nabla f(x_k) + \nabla F(x_k)^T(\lambda_k + \rho(F(x_k) + Gy_k)), \Delta x_{k+1} \rangle \\ &\quad - \frac{\beta_{k+1}}{2}\|\Delta x_{k+1}\|^2. \\ &= - \langle \nabla_x \psi_\rho(x_k, y_k, \lambda_k), \Delta x_{k+1} \rangle - \frac{\beta_{k+1}}{2}\|\Delta x_{k+1}\|^2.\end{aligned} \tag{4}$$

Using the definitions of $\mathcal{L}_\rho$ and ψ_ρ, we further obtain:

$$\begin{aligned}&\mathcal{L}_\rho(x_{k+1}, y_k, \lambda_k) - \mathcal{L}_\rho(x_k, y_k, \lambda_k) \\ &\quad = g(x_{k+1}) - g(x_k) + \psi_\rho(x_{k+1}, y_k, \lambda_k) - \psi_\rho(x_k, y_k, \lambda_k) \\ &\quad \overset{(2),(4)}{\leq} - \frac{\nu}{2}\|x_{k+1} - x_k\|^2.\end{aligned}$$

This proves our statement. ■

Let us now prove the decrease with respect to the second argument for the augmented Lagrangian function.

Lemma 4.2: [Descent of $\mathcal{L}_\rho$ w.r.t. second block of primal variables] If Assumption 2.1 holds, then we have the descent:

$$\mathcal{L}_\rho(x_{k+1}, y_{k+1}, \lambda_k) \leq \mathcal{L}_\rho(x_{k+1}, y_k, \lambda_k) - \tfrac{\alpha\theta_{k+1}}{2}\|y_{k+1} - y_k\|^2.$$

Proof: Using the definition of y_{k+1} and $\mathcal{L}_\rho$, we have:

$$\begin{aligned}&\mathcal{L}_\rho(x_{k+1}, y_{k+1}, \lambda_k) - \mathcal{L}_\rho(x_{k+1}, y_k, \lambda_k)\\ \leq &h(y_{k+1}) - l_h(y_{k+1}; y_k) - \frac{\theta_{k+1}}{2}\|y_{k+1} - y_k\|^2\\ \overset{(3)}{\leq} &-\frac{\alpha\theta_{k+1}}{2}\|y_{k+1} - y_k\|^2.\end{aligned}$$

This completes our proof. ■

Let us define the following Lyapunov function inspired from [24] (see also [8], [9]):

$$P(x, y, \lambda, z, \gamma) = \mathcal{L}_\rho(x, y, \lambda) + \frac{\gamma}{2}\|y - z\|^2, \qquad (5)$$

with $\gamma > 0$ to be defined later. The evaluation of the Lyapunov function along the iterates of Algorithm 1 is denoted by:

$$P_k = P(x_k, y_k, \lambda_k, y_{k-1}, \gamma_k) \quad \forall k \geq 1. \qquad (6)$$

Let us prove that $\{P_k\}_{k\geq 1}$ is decreasing. Indeed:

$$\begin{aligned}&P_{k+1} - P_k\\ =&\mathcal{L}_\rho(x_{k+1}, y_{k+1}, \lambda_{k+1}) - \mathcal{L}_\rho(x_{k+1}, y_{k+1}, \lambda_k)\\ &+ \mathcal{L}_\rho(x_{k+1}, y_{k+1}, \lambda_k) - \mathcal{L}_\rho(x_{k+1}, y_k, \lambda_k)\\ &+ \mathcal{L}_\rho(x_{k+1}, y_k, \lambda_k) - \mathcal{L}_\rho(x_k, y_k, \lambda_k)\\ &+ \frac{\gamma_{k+1}}{2}\|y_{k+1} - y_k\|^2 - \frac{\gamma_k}{2}\|y_k - y_{k-1}\|^2\\ \leq&\frac{1}{\rho}\|\Delta\lambda_{k+1}\|^2 - \frac{\nu}{2}\|\Delta x_{k+1}\|^2 - \frac{\alpha\theta_{k+1} - \gamma_{k+1}}{2}\|\Delta y_{k+1}\|^2\\ &- \frac{\gamma_k}{2}\|\Delta y_k\|^2, \end{aligned} \qquad (7)$$

where the inequality follows from Lemmas 4.1, 4.2 and from the update of the dual multipliers in Step 8 of Algorithm 1. Now we are going to bound the dual variables by the primal variables.

Lemma 4.3: [Bound for $\|\Delta\lambda_{k+1}\|$] If Assumption 2.1 holds on some compact set $\mathcal{S}$ and the sequence generated by Algorithm 1 is in $\mathcal{S}$, then we have:

$$\|\Delta\lambda_{k+1}\|^2 \leq 2\tfrac{\theta_{k+1}^2}{\sigma^2}\|\Delta y_{k+1}\|^2 + 2\tfrac{(\theta_k + L_h)^2}{\sigma^2}\|\Delta y_k\|^2. \qquad (8)$$

Proof: See Lemma 4.5 in [8]. ■

Now, using the inequality (8) in (7), we obtain:

$$\begin{aligned}P_{k+1} - P_k \leq& -\frac{\nu}{2}\|\Delta x_{k+1}\|^2 - \frac{\gamma_{k+1}}{4}\|\Delta y_{k+1}\|^2 - \frac{\gamma_k}{4}\|\Delta y_k\|^2\\ &- \left(\frac{2\alpha\theta_{k+1} - 3\gamma_{k+1}}{4} - \frac{2\theta_{k+1}^2}{\rho\sigma^2}\right)\|\Delta y_{k+1}\|^2\\ &- \left(\frac{\gamma_k}{4} - \frac{2(\theta_k + L_h)^2}{\rho\sigma^2}\right)\|\Delta y_k\|^2\\ \leq& -\frac{\nu}{2}\|\Delta x_{k+1}\|^2 - \frac{\alpha\theta_{k+1}}{8}\|\Delta y_{k+1}\|^2 - \frac{\alpha\theta_k}{8}\|\Delta y_k\|^2\\ &- \left(\frac{\alpha\theta_{k+1}}{8} - \frac{2(\theta_{k+1} + L_h)^2}{\rho\sigma^2}\right)\|\Delta y_{k+1}\|^2\\ &- \left(\frac{\alpha\theta_k}{8} - \frac{2(\theta_k + L_h)^2}{\rho\sigma^2}\right)\|\Delta y_k\|^2.\end{aligned}$$

The last inequality follows by selecting $\gamma_k = \frac{\alpha\theta_k}{2}$ for all $k \geq 1$, and noting that L_h is positive.

Since $\theta_0 \leq \theta_k \leq \frac{L_h}{1-\alpha}$, $\forall k \geq 1$, then by choosing

$$\rho \geq \frac{16(2-\alpha)^2}{\alpha(1-\alpha)^2}\frac{L_h^2}{\theta_0},$$

it follows that:

$$P_{k+1} - P_k \leq -\tfrac{\nu}{2}\|\Delta x_{k+1}\|^2 - \tfrac{\alpha\theta_{k+1}}{8}\|\Delta y_{k+1}\|^2 - \tfrac{\alpha\theta_k}{8}\|\Delta y_k\|^2. \qquad (9)$$

Before proving the global convergence for the iterates generated by Algorithm 1, let us first bound $\partial\mathcal{L}_\rho$. Here, $\partial\mathcal{L}_\rho$ denotes the limiting subdifferential of the augmented Lagrangian function (see [22] for more details on the limiting subdifferential).

Lemma 4.4: [Subgradient bound for $\partial\mathcal{L}_\rho$] Let $\{z_k := (x_k, y_k, \lambda_k)\}_{k\geq 1}$ be the sequence generated by the Algorithm 1. If Assumption 2.1 holds, then there exists $c > 0$ such that for every $k \geq 1$, there exists $v_{k+1} \in \partial\mathcal{L}_\rho(z_{k+1})$ satisfying:

$$\|v_{k+1}\| \leq c\|z_{k+1} - z_k\|,$$

where $c = L_\psi + \beta + L_h + \|G\| + \rho^{-1}$.

Proof: See Lemma 4.8 in [8] for details. ■

Let us now present the global asymptotic convergence for the iterates of Algorithm 1.

Theorem 1: [Limit points are stationary points] Suppose Assumptions 2.1 and 3.1 hold in a compact set $\mathcal{S}$ with radius D_S, and let ρ be chosen as specified above. Further, assume that the sequence $\{z_k := (x_k, y_k, \lambda_k)\}_{k\geq 1}$ generated by Algorithm 1 is bounded. Then, any limit point $z^* := (x^*, y^*, \lambda^*)$ of $\{z_k\}_{k\geq 1}$ is a stationary point, i.e., $0 \in \partial\mathcal{L}_\rho(x^*, y^*, \lambda^*)$. Equivalently:

$$-\nabla f(x^*) - \nabla F(x^*)^T\lambda^* \in \partial g(x^*),$$

$$\nabla h(y^*) + G^T\lambda^* = 0 \quad \text{and} \quad F(x^*) + Gy^* = 0.$$

Proof: Since $\theta_k \geq \theta_0$, for any $k \geq 1$, it then follows from (9) that, for any $k \geq 1$ we have:

$$\frac{\nu}{2}\|\Delta x_{k+1}\|^2 + \frac{\alpha\theta_0}{8}\|\Delta y_{k+1}\|^2 + \frac{\alpha\theta_0}{8}\|\Delta y_k\|^2 \leq P_k - P_{k+1}.$$

Let $k \geq 1$, by summing up the above inequality from $i = 1$ to $i = k$, we obtain:

$$\begin{aligned}&\sum_{i=1}^{k}\left(\frac{\nu}{2}\|\Delta x_{k+1}\|^2 + \frac{\alpha\theta_0}{8}\|\Delta y_{k+1}\|^2 + \frac{\alpha\theta_0}{8}\|\Delta y_k\|^2\right)\\ &\leq P_1 - P_{k+1} \leq P_1 - \bar{P},\end{aligned} \qquad (10)$$

where $\bar{P}$ is the lower bound of the sequence $\{P_k\}_{k\geq 1}$ (it is well-defined see Lemma 4.9 [8]). Since (10) holds for any $k \geq 1$, we have:

$$\sum_{i=1}^{\infty}\left(\frac{\nu}{2}\|\Delta x_{k+1}\|^2 + \frac{\alpha\theta_0}{8}\|\Delta y_{k+1}\|^2 + \frac{\alpha\theta_0}{8}\|\Delta y_k\|^2\right) < \infty.$$

This, together with the fact that $\nu, \alpha, \theta_0 > 0$, yields that:

$$\lim_{k\to\infty}\|\Delta x_k\| = 0 \quad \text{and} \quad \lim_{k\to\infty}\|\Delta y_k\| = 0. \qquad (11)$$

From Lemma 4.3, it follows that

$$\lim_{k\to\infty}\|z_{k+1} - z_k\| = 0.$$

Since the sequence $\{(x_k, y_k, \lambda_k)\}_{k\geq 1}$ is bounded, then there exists a convergent subsequence, let us say $\{(x_k, y_k, \lambda_k)\}_{k\in\mathcal{K}}$, with the limit (x^*, y^*, λ^*). From Lemma 4.4, we have $v_{k+1} \in \partial\mathcal{L}_\rho(z_{k+1})$ such that:

$$\|v^*\| := \lim_{k\in\mathcal{K}} \|v_{k+1}\| \leq c \lim_{k\in\mathcal{K}} \|z_{k+1} - z_k\| = 0.$$

Thus, $0 \in \partial\mathcal{L}_\rho(x^*, y^*, \lambda^*)$, which completes the proof. ∎

Thus, in addition to its straightforward implementation and relative simplicity of steps, our algorithm boasts theoretical convergence results, ensuring that it can reliably find optimal solutions to a wide range of nonconvex problems.

V. Numerical results

In this section, we compare Algorithm 1 with dynamic linearized alternating direction method of multipliers (DAM) from [8] and IPOPT [23] for solving the nonlinear model predictive control (MPC) problem for controlling an inverted pendulum on a cart. The input to the system is the force applied on the cart, denote by $u \in \mathbb{R}$. The horizontal position of the cart is denoted $z_1 \in \mathbb{R}$, and the angular position of the pendulum (relative to the vertical) is denoted $z_3 \in \mathbb{R}$. The continuous-time dynamics of the system can be found in [7]. Hence, the state variables are: $\mathbf{z} \triangleq [z_1\ \dot{z}_1\ z_3\ \dot{z}_3]^T$. Using Euler discretization, discrete-time model is:

$$\mathbf{z}(k+1) := \psi(\mathbf{z}(k), u(k)) \tag{12}$$
$$= \begin{bmatrix} z_1(k) + Tz_2(k) \\ z_2(k) + T\frac{M_p g\cos(z_3(k))\sin(z_3(k)) - M_p L z_4(k)^2 \sin(z_3(k)) + u(k)}{M_c + M_p \sin^2(z_3(k))} \\ z_3(k) + Tz_4(k) \\ z_3(k) + T\frac{(M_p+M_c)g\sin(z_3(k)) - M_p L z_4(k)^2\cos(z_3(k))\sin(z_3(k)) + u(k)\cos(z_3(k))}{L(M_c + M_p\sin^2(z_3(k)))} \end{bmatrix},$$

where T is the sampling time. The pendulum operates under the following state and input constraints:

$$\mathbf{z}_{\text{min}} \leq \mathbf{z}(k) \leq \mathbf{z}_{\text{max}}, \quad u_{\text{min}} \leq u(k) \leq u_{\text{max}}.$$

Our goal is to stabilize the pendulum in the vertical position and the carriage at the origin. More precisely, we aim to achieve $\mathbf{z}_{\text{ref}} = [0, 0, 0, 0]^T$ and $u_{\text{ref}} = 0$ using nonlinear MPC. To accomplish this we solve a nonlinear MPC problem at each sampling time. This problem is obtained by adopting a single shooting approach, where the state variables are eliminated under the assumption of a piecewise constant input trajectory. Slack variables are added to state constraints to convert them to equality constraints, and the constraints on the slack variables are incorporated into the objective function via soft constraints on the state variables. The decision variables for NMPC are given by $x = (u_0, \ldots, u_{N-1}) \in \mathbb{R}^N$. Before presenting the resulting NMPC, we introduce a sequence of functions $F_k : \mathbb{R}^n \to \mathbb{R}^4$ for $k \in \{0, \ldots, N-1\}$:

$$F_0(x) = z(0), \quad F_{k+1}(x) = \psi(F_k(x), u_k).$$

The resulting NMPC problem that needs to be solved at each sampling time is then given by:

$$\begin{aligned} \min_{(x, \{(s_{k+1}, s'_{k+1})\}_{k=0}^{N-1})} \ & l(x) + \sum_{k=0}^{N-1} V(s_{k+1}, s'_{k+1}) \\ \text{s.t.:}\ & F_{k+1}(x) - \mathbf{z}_{\text{max}} + s_{k+1} = 0 \\ & \mathbf{z}_{\text{min}} - F_{k+1}(x) + s'_{k+1} = 0 \\ & u_{\text{min}} \leq u(k) \leq u_{\text{max}}\ k = 0 : N-1,\ \mathbf{z}(0) \text{ given}, \end{aligned} \tag{13}$$

where $V(s, s') = \frac{\zeta}{2}\left(\max\{0, -s\}^2 + \max\{0, -s'\}^2\right)$, and where $l(x) = \frac{1}{2}\sum_{k=0}^{N-1}\left(F_{k+1}(x)^T Q F_{k+1}(x) + u_k^T R u_k\right)$. The problem described in (13) can be reformulated as problem (1), where G is the identity matrix of dimension $8N$ and N represents the prediction horizon. The nonsmooth function g is the indicator function of the set describing the input constraints. At this point it is worth mentioning that since g is an indicator function, then Step 5 of Algorithm 1 and its counterpart in DAM [8] basically reduces to finding a solution of a QP problem with box constraints. All the codes were implemented in MATLAB and we used the following setup for our system and for nonlinear MPC:

$$T = 0.2, M_p = 0.1, M_c = 1, L = 0.8, g = 9.81,$$
$$N = 10, Q = \text{diag}(3, 3, 1, 1), R = 1, \text{z}_{\text{max}} = [10, 10, 1, 10]^T,$$
$$\text{z}_{\text{min}} = -\text{z}_{\text{max}},\ u_{\text{max}} = 10,\ u_{\text{min}} = -u_{\text{max}},$$

and the simulation horizon is 40. Moreover, the parameters for the DAM algorithm and the Linearized ADMM algorithm are provided in Table I. Note that the above parameters are

Method \ Parameters	ρ	β_k	θ_k	ζ
DAM from [8]	30	50	1	5
Linearized ADMM	3	1	1	5

TABLE I
Parameters for Linearized ADMM and DAM.

the best ones found for each algorithm. For DAM, a large value of β_k is required to cover the big approximation error generated by the linearization of the full smooth part of the augmented Lagrangian function. Additionally, a value of ρ less than the one reported in Table I either does not allow convergence or is extremely slow. On the other hand, we chose $\theta_k = 1$ for both our method and the DAM. The value $\zeta = 5$ is of course the one considered in the case of IPOPT. In Table II, we report the number of iterations required for each algorithm to solve a nonlinear MPC problem (the first NMPC problem in our case), the CPU time required for each algorithm to solve a nonlinear MPC problem, the optimal value found by each algorithm, and the degree of infeasibility at the obtained optimal point. As can be seen from Table II, our algorithm requires fewer iterations to solve the problem, since the model considered in Step 5 approximates better the original augmented Lagrangian than the one considered in [8]. At the same time, the complexity of Step 5 for us does not differ too much from the complexity of the corresponding step in DAM, which results in the fact that our algorithm is much faster than DAM. Also, our algorithm is slightly faster than IPOPT. Both the proposed method and DAM algorithm achieve an optimal value that is very close to that of the IPOPT solver. In Table II, the term ”infeasibility” refers to the infeasibility of the problem (13). This is equivalent to violations of the state constraints when the penalty parameter ζ is sufficiently large.

Method \ Characteristics	# Iter	Cpu (s)	Optimal value	Infeasibility
IPOPT	35	2.88	350.29	1.28×10^{-6}
DAM from [8]	208	13.04	358.32	9.76×10^{-5}
Linearized ADMM	60	0.85	350.30	6.31×10^{-5}

TABLE II
NUMERICAL RESULTS COMPARING LINEARIZED ADMM AND DAM ON A NONLINEAR MPC PROBLEM.

Figure 1 displays the closed-loop trajectories of inputs and states obtained using Algorithm 1 (linearized ADMM), revealing the successful stabilization of the system through the proposed approach.

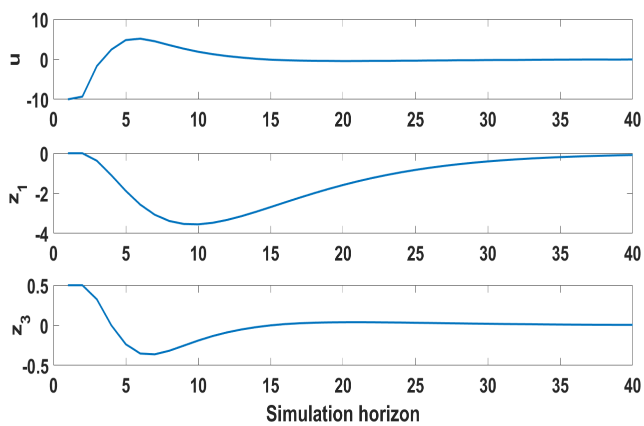

Fig. 1. Nonlinear MPC inputs u (top), cart lateral position z_1 (middle) and pendulum angular position z_3 (bottom) computed using Linearized ADMM.

VI. CONCLUSIONS

In this paper we have introduced a linearized ADMM method for solving structured nonsmooth nonconvex optimization problems. By linearizing the smooth term of the objective function and functional constraints within the augmented Lagrangian, we have been able to derive simple updates. We have established that the iterates of our method globally converge to a critical point of the original problem. Furthermore, the numerical experiments have shown the effectiveness of our proposed algorithm in solving nonlinear MPC problems.

ACKNOWLEDGMENT

The research leading to these results has received funding from: the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 953348; UEFISCDI PN-III-P4-PCE-2021-0720, under project L2O-MOC, nr. 70/2022.

REFERENCES

[1] D. P. Bertsekas *Constrained Optimization and Lagrange Multiplier Methods*, Athena Scientific, 1996.
[2] D. P. Bertsekas *Convex Optimization Algorithms*, Athena Scientific, 2015.
[3] D. P. Bertsekas, and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, 2003.
[4] J. Bolte, S. Sabach and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming, 146(1–2), 459–494, 2014.
[5] R.I. Bot and D.K. Nguyen, *The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates*, Mathematics of Operations Research, 45(2), 682–712, 2020.
[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning 3(1), 1–122, 2011.
[7] A. Brendon and V. Lua, *Inverted Pendulum Stabilization via Learning-Based MPC,* Lecture Notes, 2019.
[8] E. Cohen, N. Hallak and M. Teboulle, *A Dynamic Alternating Direction of Multipliers for Nonconvex Minimization with Nonlinear Functional Equality Constraints*, Journal of Optimization Theory and Applications, 193, 324–353, 2022.
[9] L. El Bourkhissi and I. Necoara, *Complexity of linearized augmented Lagrangian for nonconvex optimization with nonlinear equality constraints*, arXiv:2301.08345, 2022.
[10] J. A. Fessler, *Optimization methods for magnetic resonance image reconstruction*, key models and optimization algorithmsm, IEEE Signal Process. Mag., 37(1), 33–40, 2020.
[11] G. Gabay, and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. Appl., 2(1), 17–40, 1976.
[12] R. Glowinski, and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, SIAM, 9, 1989.
[13] P. C. Haarhoff, J. D. Buys, *A new method for the optimization of a nonlinear function subject to nonlinear constraints*, Computer Journal, 13, 178–184, 1970.
[14] D. Hajinezhad and M. Hong, *Perturbed proximal primal-dual algorithm for nonconvex nonsmooth optimization*, Mathematical Programming, 176, 207–245, 2019.
[15] M. Hestenes, *Multiplier and gradient methods*, Journal of Optimization Theory and Applications, 4, 303–320, 1969.
[16] M. Hong, D. Hajinezhad, and M. M. Zhao, *Prox-PDA: The proximal primal-dual algorithm for fast distributed nonconvex optimization and learning over networks*. In: D. Precup, Y.W. Teh (eds.), International Conference on Machine Learning, 70, 1529–1538, 2017.
[17] B. Jiang, T. Lin, S. Ma and S. Zhang, *Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis*, Computational Optimization and Applications, 72(1), 115–157, 2019.
[18] F. Messerer, K. Baumgärtner and M. Diehl, *Survey of sequential convex programming and generalized Gauss-Newton methods*, ESAIM: Proceedings and Surveys, 2021.
[19] L. Qinghua, S. Xinyue and G. Yuantao, *Linearized ADMM for Nonconvex Nonsmooth Optimization With Convergence Analysis*, IEEE Access, doi:10.1109/ACCESS.2019.2914461.
[20] A. Themelis and P. Patrinos, *Douglas–Rachford splitting and ADMM for nonconvex optimization: Tight convergence results*, SIAM Journal on Optimization, 30(1), 149–181, 2020.
[21] A. Themelis, L. Stella and P. Patrinos, *Douglas–Rachford splitting and ADMM for nonconvex optimization: accelerated and Newton-type linesearch algorithms*, Computational Optimization and Applications, 82(2), 395–440, 2022.
[22] R. Rockafellar and R. Wets, *Variational Analysis*, Springer, 1998.
[23] A. Wächter and L. T. Biegler, *On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming*, Mathematical Programming, 106(1), 25-57, 2006.
[24] Y. Xie, S. J. Wright, *Complexity of Proximal Augmented Lagrangian for Nonconvex Optimization with Nonlinear Equality Constraints*, Journal of Scientific Computing, 86, 2021.
[25] M. Yashtini, *Convergence and rate analysis of a proximal linearized ADMM for nonconvex nonsmooth optimization*, Journal of Global Optimization, 84(4), 913-939, 2022.
[26] J. Zhang and Z. Q. Luo, *A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization*, SIAM Journal Optimization: 30(3), 2272–2302, 2020.