A Distributed Augmented Lagrangian Method for Model Predictive Control

Soomin Lee, Nikolaos Chatzipanagiotis and Michael M. Zavlanos

***Abstract*— In this paper we present a distributed Augmented Lagrangian (AL) algorithm to solve model predictive control (MPC) problems that involve a finite number of subsystems which interact with each other via a general network. We focus on discrete-time control systems with time-varying linear dynamics. Our method relies on the Accelerated Distributed Augmented Lagrangian (ADAL) algorithm, which can handle globally coupled linear constraints in a distributed manner based on a locally estimated AL. We prove that the theoretical complexity of ADAL to reach an ϵ-optimal solution both in terms of primal optimality gap and feasibility residual is $O(\frac{1}{\epsilon})$ iterations. As suggested by our numerical analysis, ADAL achieves very fast convergence rates compared to the popular ADMM for distributed MPC.**

I. INTRODUCTION

Model predictive control (MPC) [1, 2], also called receding horizon control, is an optimization-based control method for dynamical systems, which has received much attention due to its wide range of applications, see e.g., [3]. The basic concept in MPC is that, at any given time instant, an online finite or infinite horizon open-loop optimal control problem is solved, where the current state of the system is used as the initial state. The solution to this problem is an open-loop sequence of inputs for the selected horizon.

MPC is one of the most successful control frameworks implemented on embedded systems due to its ability to handle complex dynamics with hard input and state constraints. However, as the sampling times for embedded systems are very short, any iterative optimization algorithm implemented on such systems must be able to precondition the execution time by providing an explicit number of iterations that is sufficient to obtain a reasonably good solution both in terms of suboptimality and feasibility residual. For this reason, recently, there has been a growing interest in enhancing MPC methods by providing the worst-case computational complexity [4]–[8].

Distributed MPC (DMPC) schemes, see e.g., [9, 10], have gained increased popularity recently, since they can improve on computational efficiency by conjoining resources of multiple inexpensive devices (or subsystems) and alleviate the drawbacks of centralized control. Since computing projections on the primal feasible set is hard for constrained MPC problems, a practical choice is to use Lagrangian relaxation and dual decomposition, see e.g., [11]. In general, classical dual decomposition methods suffer from slow convergence due to the non-differentiability of the dual functions induced by the ordinary Lagrangian [12]. This drawback can be avoided by using the Augmented Lagrangian (AL) framework. However, AL based methods lose the decomposable structure of the ordinary Lagrangian, which makes distributed computation difficult. This calls for the development of specialized AL decomposition techniques.

The contributions of this paper are the following: 1) We revisit the general purpose AL method, the Accelerated Distributed Augmented Lagrangian (ADAL for short) [13], and provide a computational complexity certifications both in terms of primal optimality gap and feasibility residual. Compared to our previous work [13], where only an asymptotic convergence result is provided, we show that the number of iterations to reach an ϵ-optimal and ϵ-feasible solution is $O(\frac{1}{\epsilon})$, under the assumption that the objective function is generally convex and not necessarily differentiable. As mentioned above, this analysis is extremely useful for MPC problems where obtaining a guaranteed solution of predefined accuracy is really important; 2) We compare our method to the ADMM algorithm, which has also been studied in the context of DMPC problems [14, 15]. ADAL and ADMM have the same theoretical rate of convergence, but ADAL converges faster than ADMM in practice. We verify this via simulation results.

II. ACCELERATED DISTRIBUTED AUGMENTED LAGRANGIAN

This section describes ADAL method, a specialized *Augmented Lagrangian* (AL) decomposition technique which has been proposed in [13], for solving a distributed problem of the form:

$$\begin{aligned} \min_{\mathbf{x}_i} \quad & \sum\nolimits_{i=1}^{N} f_i(\mathbf{x}_i) \\ \text{subject to} \quad & \sum\nolimits_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i = \mathbf{b}, \\ & \mathbf{x}_i \in \mathcal{X}_i, \quad i = 1, 2, \ldots, N, \end{aligned} \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^{n_i}$ denotes the decision variables that belong to subsystem i, and $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ is its local objective function. Problem (1) models situations where a set $\mathcal{I} = \{1, 2, \ldots, N\}$ of decision makers, henceforth referred to as agents, need to determine local decisions $\mathbf{x}_i \in \mathcal{X}_i$ that minimize the summation of the local functions $f_i(\mathbf{x}_i)$, while respecting a set of affine coupling constraints $\sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i = \mathbf{b}$. Here,

This work is supported by NSF under grant CNS #1261828, and by ONR under grant #N000141410479.

Soomin Lee is with the Dept. of Industrial and Systems Engineering, Georgia Tech, Atlanta, GA, 30318, USA, soomin.lee@isye.gatech.edu, Nikolaos Chatzipanagiotis and Michael M. Zavlanos are with the Dept. of Mechanical Engineering and Materials Science, Duke University, Durham, NC, 27708, USA, {n.chatzip,michael.zavlanos}@duke.edu.

we assume the functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ are convex (not necessarily differentiable) for all $i \in \mathcal{I}$, the local sets $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$ for $i \in \mathcal{I}$ are convex, closed and bounded, $\mathbf{A}_i \in \mathbb{R}^{m \times n_i}$, $\mathbf{b} \in \mathbb{R}^m$, and $n = \sum_{i=1}^N n_i$.

Furthermore, we let

$$F(\mathbf{x}) := \sum\nolimits_{i=1}^{N} f_i(\mathbf{x}_i),$$

where $\mathbf{x} = [\mathbf{x}_1^\top, \ldots, \mathbf{x}_N^\top]^\top \in \mathbb{R}^n$. Denoting $\mathbf{A} = [\mathbf{A}_1 \ldots \mathbf{A}_N] \in \mathbb{R}^{m \times n}$, the constraint $\sum_{i=1}^N \mathbf{A}_i \mathbf{x}_i = \mathbf{b}$ in problem (1) becomes $\mathbf{A}\mathbf{x} = \mathbf{b}$. Also, we define the maximum degree q as a measure of sparsity of the matrix $\mathbf{A}$, i.e., for each constraint $j = 1, \ldots, m$, we denote by q_j the number of all $i \in \mathcal{I}$ such that $[\mathbf{A}_i]_j \neq \mathbf{0}$, where $[\mathbf{A}_i]_j$ is the j-th row of matrix $\mathbf{A}_i$ and $\mathbf{0}$ stands for a vector of all zeros. Then, q is defined as $q = \max_{j=1,\ldots,m} q_j$. It will be shown below that q plays a critical role in the convergence properties of the proposed method.

A. The ADAL Algorithm

Associating Lagrange multipliers $\boldsymbol{\lambda} \in \mathbb{R}^m$ with the affine constraint $\mathbf{A}\mathbf{x} = \mathbf{b}$, the Lagrangian for (1) is defined as

$$L(\mathbf{x}, \boldsymbol{\lambda}) = F(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{A}\mathbf{x} - \mathbf{b} \rangle = \sum\nolimits_{i=1}^{N} L_i(\mathbf{x}_i, \boldsymbol{\lambda}) - \langle \mathbf{b}, \boldsymbol{\lambda} \rangle, \tag{2}$$

where $L_i(\mathbf{x}_i, \boldsymbol{\lambda}) = f_i(\mathbf{x}_i) + \langle \boldsymbol{\lambda}, \mathbf{A}_i \mathbf{x}_i \rangle$, and $\langle \cdot, \cdot \rangle$ denotes inner product. The associated AL is then given by

$$\Lambda_\rho(\mathbf{x}, \boldsymbol{\lambda}) = F(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{A}\mathbf{x} - \mathbf{b} \rangle + \frac{\rho}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2, \tag{3}$$

where $\rho > 0$ is a penalty parameter. A major drawback of the Augmented Lagrangian Method stems from the fact that (3) is not separable with respect to each $\mathbf{x}_i$ due to the additional quadratic penalty term.

The lack of decomposability of the AL calls for the development of specialized AL decomposition techniques. ADAL is a primal-dual iterative method utilizing a local AL function Λ_ρ^i which is defined as:

$$\begin{aligned} \Lambda_\rho^i(\mathbf{x}_i, \mathbf{x}_{-i}^k, \boldsymbol{\lambda}) \;&=\; f_i(\mathbf{x}_i) \;+\; \langle \boldsymbol{\lambda}, \mathbf{A}_i \mathbf{x}_i \rangle \\ &+\; \frac{\rho}{2} \|\mathbf{A}_i \mathbf{x}_i + \sum\nolimits_{j \in \mathcal{I}}^{j \neq i} \mathbf{A}_j \mathbf{x}_j^k - \mathbf{b}\|^2, \end{aligned} \tag{4}$$

where $\mathbf{x}_{-i}^k = [\mathbf{x}_1^k, \ldots, \mathbf{x}_{i-1}^k, \mathbf{x}_{i+1}^k, \ldots, \mathbf{x}_N^k]^\top$. The ADAL method is summarized in Alg. 1. ADAL has two parameters: a positive penalty parameter ρ and a stepsize parameter $\tau \in (0, 1/q)$. Each iteration of ADAL consists of three steps: i) every agent solves a local subproblem in a parallel fashion based on the local approximation of the AL in (4); ii) the agents update and communicate their primal variables to neighboring agents; and iii) they update their dual variables based on the values of the communicated primal variables.

We emphasize here again that the quantities $\mathbf{A}_j \mathbf{x}_j^k$, appearing in the penalty term of the local AL (4), correspond to the local primal variables of agent j that are communicated to agent i. With respect to agent i, these are considered fixed parameters. The penalty term of each Λ_ρ^i can be equivalently expressed as

$$\begin{aligned} &\|\mathbf{A}_i \mathbf{x}_i + \sum\nolimits_{j \in \mathcal{I}}^{j \neq i} \mathbf{A}_j \mathbf{x}_j^k - \mathbf{b}\|^2 \;= \\ &\qquad = \;\sum\nolimits_{l=1}^{m} \Big(\big[\mathbf{A}_i \mathbf{x}_i\big]_l + \sum\nolimits_{j \in \mathcal{I}}^{j \neq i} \big[\mathbf{A}_j \mathbf{x}_j^k\big]_l - b_l \Big)^2. \end{aligned}$$

The above penalty term is involved only in the minimization computation (5), in Alg. 1. Hence, for those l such that $[\mathbf{A}_i]_l = \mathbf{0}$, the terms $\sum_{j \in \mathcal{I}}^{j \neq i} \big[\mathbf{A}_j \mathbf{x}_j^k\big]_l - b_l$ are just constant terms in the minimization step, and can be excluded. Here, $[\mathbf{A}_i]_l$ denotes the l-th row of $\mathbf{A}_i$ and $\mathbf{0}$ stands for a zero vector of proper dimension. This implies that agent i needs access only to the decisions $\big[\mathbf{A}_j \mathbf{x}_j^k\big]_l$ from all agents $j \neq i$ that are involved in the same constraints l as i. Moreover, regarding the term $\langle \boldsymbol{\lambda}, \mathbf{A}_i \mathbf{x}_i \rangle$ in (4), we have that $\langle \boldsymbol{\lambda}, \mathbf{A}_i \mathbf{x}_i \rangle = \sum_{j=1}^m \lambda_j [\mathbf{A}_i \mathbf{x}_i]_j$. Hence, we see that, in order to compute (5), each agent i needs access only to those λ_j for which $[\mathbf{A}_i]_j \neq \mathbf{0}$.

Remark 1 *Compared to the original non-distributed AL*

$$\Lambda_\rho(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i=1}^{N} f_i(\mathbf{x}_i) + \langle \boldsymbol{\lambda}, \sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i \rangle + \frac{\rho}{2} \| \sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i - \mathbf{b}\|^2,$$

the sum of the local AL in ADAL has N local estimates of the quadratic penalty term, i.e.,

$$\begin{aligned} \sum_{i=1}^{N} \Lambda_\rho^i(\mathbf{x}_i, \mathbf{x}_j^k, \boldsymbol{\lambda}) = \sum_{i=1}^{N} f_i(\mathbf{x}_i) \;+\; \langle \boldsymbol{\lambda}, \sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i \rangle \\ + \frac{\rho}{2} \sum_{i=1}^{N} \|\mathbf{A}_i \mathbf{x}_i + \sum\nolimits_{j \in \mathcal{I}}^{j \neq i} \mathbf{A}_j \mathbf{x}_j^k - \mathbf{b}\|^2. \end{aligned}$$

As it will be discussed again in Section IV, this allows us to set the penalty parameter ρ much smaller compared to any other AL methods including ADMM; cf. see the AL in [16]. Roughly speaking, the ρ in our local AL can be set to N times smaller than that in other AL methods. As a consequence, ADAL can focus more on minimizing the objective function than the penalty term, which results in faster convergence.

B. DMPC with linear coupling constraints

Consider a discrete-time linear dynamical system, which can be expressed in terms of the dynamics of a set $\mathcal{I} = \{1, \ldots, N\}$ of subsystems:

$$\begin{aligned} \mathbf{x}_i^{t+1} &= \sum\nolimits_{j \in \mathcal{C}_i^t} \left(\mathbf{A}_{ij}^t \mathbf{x}_j^t + \mathbf{B}_{ij}^t \mathbf{u}_j^t \right) \\ \mathbf{x}_i^t &\in \mathcal{X}_i^t, \quad \mathbf{u}_i^t \in \mathcal{U}_i^t, \quad \forall\, i \in \mathcal{I}, \end{aligned} \tag{8}$$

where $\mathbf{x}_i^t \in \mathcal{X}_i^t \subseteq \mathbb{R}^{n_i}$ and $\mathbf{u}_i^t \in \mathcal{U}_i^t \subseteq \mathbb{R}^{p_i}$ represent a local state and input at time t. We assume that the local constraint sets $\mathcal{X}_i^t$, $\mathcal{U}_i^t$ satisfy $\mathcal{X}^t = \mathcal{X}_1^t \times \cdots \times \mathcal{X}_N^t$, $\mathcal{U}^t = \mathcal{U}_1^t \times \cdots \times \mathcal{U}_N^t$, and $n = \sum_{i \in \mathcal{I}} n_i$, $p = \sum_{i \in \mathcal{I}} p_i$. The dynamic interconnections at time t among the subsystems are modeled by a directed graph $\mathcal{G}^t = (\mathcal{I}, \mathcal{E}^t)$. The set of edges $\mathcal{E}^t \subseteq \mathcal{I} \times \mathcal{I}$ contains a directed edge (v_i, v_j) if the state or input of subsystem i at time t affects the dynamics of subsystem

Algorithm 1 Accelerated Distributed Augmented Lagrangians (ADAL)

Set $k=0$, $\tau \in (0, \frac{1}{q})$ and define initial Lagrange multipliers $\boldsymbol{\lambda}^0$ and initial primal variables $\mathbf{x}^0$.

1. For fixed Lagrange multipliers $\boldsymbol{\lambda}^k$, determine $\hat{\mathbf{x}}_i^k$ for every $i \in \mathcal{I}$ as the solution of the following problem:
$$\min_{\mathbf{x}_i \in \mathcal{X}_i} \Lambda_\rho^i(\mathbf{x}_i, \mathbf{x}_{-i}^k, \boldsymbol{\lambda}^k). \tag{5}$$
2. Set for every $i \in \mathcal{I}$
$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \tau(\hat{\mathbf{x}}_i^k - \mathbf{x}_i^k). \tag{6}$$
3. If the constraints $\sum_{i=1}^N \mathbf{A}_i \mathbf{x}_i^{k+1} = \mathbf{b}$ are satisfied and $\mathbf{A}_i \hat{\mathbf{x}}_i^k = \mathbf{A}_i \mathbf{x}_i^k$ for all $i \in \mathcal{I}$, then stop (optimal solution found). Otherwise, set:
$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \rho\tau\Big(\sum\nolimits_{i=1}^N \mathbf{A}_i \mathbf{x}_i^{k+1} - \mathbf{b}\Big), \tag{7}$$
increase k by one and return to Step 1.

j at time $t+1$. More formally, $(v_j, v_i) \in \mathcal{E}^t$ if and only if $\mathbf{A}_{ij}^t \neq 0 \ \vee \ \mathbf{B}_{ij}^t \neq 0$, where the matrices $\mathbf{A}_{ij}^t \in \mathbb{R}^{n_i \times n_j}$ and $\mathbf{B}_{ij}^t \in \mathbb{R}^{n_i \times p_j}$, define the dynamic coupling between subsystems i and j at time t. We define the coupling in-neighborhood $\mathcal{C}_i^t$ (resp. out-neighborhood $\tilde{\mathcal{C}}_i^t$) of subsystem i at time t as the set of sybsystems j whose dynamics at t affect (resp. is affected by) the evolution of subsystem i, i.e., $\mathcal{C}_i^t = \{j \in \mathcal{I} : (v_j, v_i) \in \mathcal{E}^t\}$ (resp. $\tilde{\mathcal{C}}_i^t = \{j \in \mathcal{I} : (v_i, v_j) \in \mathcal{E}^t\}$).

Determining optimal control sequences for (8) using MPC consists of solving online a finite horizon open-loop optimal control problem, subject to the aforementioned system dynamics and constraints involving states and control inputs. Specifically, the MPC problem for the dynamical system (8) is parametric to the initial state $\mathbf{x}^1$ and can be formulated as

$$\begin{aligned} \min_{\mathbf{x},\mathbf{u}} \ & \sum_{i=1}^N \Big[\sum_{t=1}^{H-1} \ell_i^t(\mathbf{x}_i^t, \mathbf{u}_i^t) + \mathcal{F}_i(\mathbf{x}_i^H) \Big] \\ \text{s.t.} \ & \mathbf{x}_i^{t+1} = \sum\nolimits_{j \in \mathcal{C}_i^t} \left(\mathbf{A}_{ij}^t \mathbf{x}_j^t + \mathbf{B}_{ij}^t \mathbf{u}_j^t \right), \\ & \mathbf{x}_i^{t+1} \in \mathcal{X}_i^{t+1}, \ \mathbf{u}_i^t \in \mathcal{U}_i^t, \ \forall \ i \in \mathcal{I} \text{ and } 1 \leq t \leq H-1. \end{aligned} \tag{9}$$

where the functions $\ell_i^t(\mathbf{x}_i^t, \mathbf{u}_i^t) : \mathbb{R}^{n_i} \times \mathbb{R}^{p_i} \to \mathbb{R}$ denote the running cost and the function $\mathcal{F}_i(\mathbf{x}_i^H) : \mathbb{R}^{n_i} \to \mathbb{R}$ denotes the terminal cost of subsystem i.

To use the ADAL framework in Alg. 1 for solving (9), we introduce a local AL for each subsystem i as

$$\begin{aligned} \Lambda_\rho^i(\mathbf{x}_i, \mathbf{u}_i, \boldsymbol{\lambda}) \ = \ & \sum_{t=1}^{H-1} \ell_i^t(\mathbf{x}_i^t, \mathbf{u}_i^t) \ + \mathcal{F}_i(\mathbf{x}_i^H) \\ + \sum_{t=1}^{H-1} \Bigg[& (\boldsymbol{\lambda}_i^{t+1})^T \mathbf{x}_i^{t+1} - \sum_{j \in \tilde{\mathcal{C}}_i^t} (\boldsymbol{\lambda}_j^{t+1})^T \left(\mathbf{A}_{ji}^t \mathbf{x}_i^t + \mathbf{B}_{ji}^t \mathbf{u}_i^t \right) \\ + \frac{\rho}{2} \| & \mathbf{x}_i^{t+1} - \mathbf{A}_{ii}^t \mathbf{x}_i^t - \mathbf{B}_{ii}^t \mathbf{u}_i^t - \sum_{j \in \mathcal{C}_i^t \setminus \{i\}} \left(\mathbf{A}_{ij}^t \tilde{\mathbf{x}}_j^t + \mathbf{B}_{ij}^t \tilde{\mathbf{u}}_j^t \right) \|^2 \\ + \sum_{j \in \tilde{\mathcal{C}}_i^t} \frac{\rho}{2} \| & \tilde{\mathbf{x}}_j^{t+1} - \mathbf{A}_{ji}^t \mathbf{x}_i^t - \mathbf{B}_{ji}^t \mathbf{u}_i^t \\ & - \sum_{m \in \mathcal{C}_j^t \setminus \{i\}} \left(\mathbf{A}_{jm}^t \tilde{\mathbf{x}}_m^t + \mathbf{B}_{jm}^t \tilde{\mathbf{u}}_m^t \right) \|^2 \Bigg], \end{aligned} \tag{10}$$

where we introduce $\tilde{\mathbf{x}}_j, \tilde{\mathbf{u}}_j$, denoting the primal variables that are controlled by subsystem j but communicated to subsystem i for optimization of its local Lagrangian Λ_ρ^i. With respect to subsystem i, these are just considered as fixed parameters. That is, the local AL is created by taking all the terms involving $\mathbf{x}_i$ in the original AL and setting the remaining variables as fixed parameters, i.e., $\mathbf{x}_j$ as $\tilde{\mathbf{x}}_j$ for all $i \neq j$.

Observe that the local AL (10) of each subsystem i includes only locally available information. Regarding the dual variables, the necessary information includes $\boldsymbol{\lambda}_i^{t+1}$ and all $\boldsymbol{\lambda}_j^{t+1}$ for every $t \in \{1, \ldots, H-1\}$ and $j \in \tilde{\mathcal{C}}_i^t$, i.e., the dual variables corresponding to the dynamical constraints of i and also those of the out-neighbors of subsystem i in all coupling graphs $\mathcal{E}^t$. Regarding the primal variables, the necessary information for the local AL of subsystem i includes all $\tilde{\mathbf{x}}_j^t$, $\tilde{\mathbf{u}}_j^t$ for every $t \in \{1, \ldots, H\}$ from the in-neighbors $j \in \mathcal{C}_i^t$, the out-neighbors $j \in \tilde{\mathcal{C}}_i^t$, and the in-neighbors of the out-neighbors of i, namely $\{m \in \mathcal{I} : m \in \mathcal{C}_j^t, \ \forall j \in \tilde{\mathcal{C}}_i^t\}$ for all the coupling graphs $\mathcal{E}^t$. In other words, each subsystem i needs to be able to exchange messages with all subsystems j that belong to its 2-hop communication neighborhood $\mathcal{I}_i = \bigcup_{t=1}^H \Big(\mathcal{C}_i^t \cup \tilde{\mathcal{C}}_i^t \cup \{m \in \mathcal{I} : m \in \mathcal{C}_j^t, \ \forall j \in \tilde{\mathcal{C}}_i^t\} \Big)$.

In practice (9) is solved repeatedly, and after each solve, the first few inputs are applied to (8) and the horizon is shifted according, providing a new initial condition for a subsequent solution of (9). In this framework, solving (9) until convergence is time consuming, therefore, early termination is highly desired, while ensuring a good quality solution.

III. Rate of Convergence

In this section we characterize the rate of convergence of the ADAL method. In what follows, we denote the subgradient of a convex function f at a point $\mathbf{x} \in \mathcal{X}$ by $\mathbf{s}_\mathbf{x}$, i.e., a vector $\mathbf{s}_\mathbf{x} \in \mathbb{R}^n$ is a subgradient of f at $\mathbf{x} \in \mathcal{X}$ if

$$f(\tilde{\mathbf{x}}) \geq f(\mathbf{x}) + \langle \mathbf{s}_\mathbf{x}, \tilde{\mathbf{x}} - \mathbf{x} \rangle, \quad \forall \tilde{\mathbf{x}} \in \mathcal{X}.$$

We also denote the convex subdifferential of f at $\mathbf{x} \in \mathcal{X}$ by $\partial f(\mathbf{x})$, which is the set of all subgradients $\mathbf{s}_\mathbf{x}$.

The convergence of ADAL relies on the following three assumptions, which are typically required in the analysis of convex optimization methods:

(A1) The functions f_i are convex, and the sets $\mathcal{X}_i$ are convex, closed, and bounded for all $i \in \mathcal{I}$.

(A2) The Lagrangian function L has a saddle point $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathbb{R}^n \times \mathbb{R}^m$ so that
$$L(\mathbf{x}^*, \boldsymbol{\lambda}) \leq L(\mathbf{x}^*, \boldsymbol{\lambda}^*) \leq L(\mathbf{x}, \boldsymbol{\lambda}^*), \ \forall \, \mathbf{x} \in \mathcal{X}, \ \boldsymbol{\lambda} \in \mathbb{R}^m.$$

(A3) All subproblems (5) are exactly solvable at every iteration.

Assumption (A1) implies that there exists a constant $D_{\mathcal{X}}$ such that

$$D_{\mathcal{X}} := \max_{\mathbf{x},\tilde{\mathbf{x}}\in\mathcal{X}} \|\mathbf{x}-\tilde{\mathbf{x}}\| \tag{11}$$

and also there exist Lipschitz subgradients, i.e., there exists a constant G such that for all $i \in \mathcal{I}$

$$\|\mathbf{s}_{\mathbf{x}}\| \le G, \quad \forall \mathbf{s}_{\mathbf{x}} \in \partial f_i(\mathbf{x}),\ \mathbf{x} \in \mathcal{X}. \tag{12}$$

Assumption (A2) implies that the point $\mathbf{x}^*$ is a solution of problem (1) and the point $\boldsymbol{\lambda}^*$ is a solution of (**??**). Since (1) is a convex program with linear constraints, strong duality holds, i.e., the optimal values of the primal and dual problems are equal, as long as (1) is feasible without the need of any constraint qualification. Assumption (A3) is satisfied for most practical MPC problems, see e.g., [4, Section V], or for general problems if the constraint set $\mathcal{X}$ is simple, e.g., boxes or balls.

A. Lemmas

In this subsection, we provide a few lemmas that will help us prove the convergence of ADAL. Our analysis relies on the following ergodic average of the primal variables up to iteration k, i.e., $\tilde{\mathbf{x}}^k := \frac{1}{k}\sum_{p=0}^{k-1}\hat{\mathbf{x}}^p$. We define the *residual* $\mathbf{r}(\mathbf{x}) \in \mathbb{R}^m$ as the vector containing the amount of all constraint violations with respect to primal variable $\mathbf{x}$, i.e., $\mathbf{r}(\mathbf{x}) = \sum_i \mathbf{A}_i\mathbf{x}_i - \mathbf{b}$. We also define the auxiliary dual variable $\bar{\boldsymbol{\lambda}}^k$ as

$$\bar{\boldsymbol{\lambda}}^k \ := \ \boldsymbol{\lambda}^k + \rho(1-\tau)\mathbf{r}(\mathbf{x}^k). \tag{13}$$

In the next lemma, we obtain an iterative relation for this auxiliary variable. The proof can be found in [13, Theorem 1].

Lemma 1 *The dual update step* (7) *of ADAL is equivalent to the update rule*

$$\bar{\boldsymbol{\lambda}}^{k+1} \ = \ \bar{\boldsymbol{\lambda}}^k + \tau\rho\mathbf{r}(\hat{\mathbf{x}}^k).$$

In the following lemma, we utilize Lemma 1 and the first order optimality conditions for each local subproblem (5) to bound the function value at each iteration, which later will allow us to obtain a telescoping sum. For this, we make use of the following Lyapunov/Merit function $\phi^k(\boldsymbol{\lambda})$ defined by

$$\phi^k(\boldsymbol{\lambda}) \ = \ \rho\sum\nolimits_i \|\mathbf{A}_i(\mathbf{x}_i^k - \mathbf{x}_i^*)\|^2 + \frac{1}{\rho}\|\bar{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}\|^2, \tag{14}$$

for all $k \ge 0$ and any arbitrary $\boldsymbol{\lambda} \in \mathbb{R}^m$. We omit the proof due to space limitation. A similar proof can be found in [17, Lemma 1-3], where the merit function is defined with respect to a fixed dual variable $\boldsymbol{\lambda}^*$.

Lemma 2 *Assume (A1)–(A3). Then, for any $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $k \ge 0$, the following holds:*

$$F(\hat{\mathbf{x}}^k) - F(\mathbf{x}^*) + \langle \boldsymbol{\lambda}, \mathbf{r}(\hat{\mathbf{x}}^k)\rangle \le \ \frac{1}{2\tau}\big(\phi^k(\boldsymbol{\lambda}) - \phi^{k+1}(\boldsymbol{\lambda})\big).$$

B. Primal Optimality and Feasibility

Using Lemma 2 and the properties of convex functions, we now provide two theorems regarding the convergence rate of ADAL. More specifically, in Theorem 1, we consider the objective value difference $F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*)$ and the constraint violation $\|\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{b}\|$ together and show that the sum of these two decreases at a worst-case $O(1/k)$ rate. In Theorem 2, we upper bound the objective value difference and constraint violation separately, and show that each of them decreases at a worst-case $O(1/k)$ rate.

Theorem 1 *Assume (A1)–(A3). Recall that $\tilde{\mathbf{x}}^k = \frac{1}{k}\sum_{p=0}^{k-1}\hat{\mathbf{x}}^p$ denotes the ergodic average of the primal variable sequence generated by ADAL up to iteration k and $\mathbf{r}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$ denotes the residual at $\mathbf{x}$. Then, for all k*

$$F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*) + \|\mathbf{r}(\tilde{\mathbf{x}}^k)\| \ \le \ \frac{1}{2k\tau}\phi, \tag{15}$$

where $\phi = \sum_{i=1}^N \rho\|\mathbf{A}_i(\mathbf{x}_i^0 - \mathbf{x}_i^*)\|^2 + \frac{1}{\rho}(\|\bar{\boldsymbol{\lambda}}^0\| + 1)^2$.

Proof: Summing the relation in Lemma 2 for all $p = 0, \dots, k-1$, we get

$$\begin{aligned}&\sum\nolimits_{p=0}^{k-1} F(\hat{\mathbf{x}}^p) \ - \ \sum\nolimits_{p=0}^{k-1} F(\mathbf{x}^*) \ + \ \sum\nolimits_{p=0}^{k-1}\Big\langle \boldsymbol{\lambda}, \mathbf{r}(\hat{\mathbf{x}}^p)\Big\rangle \\ &\le \ \frac{1}{2\tau}\Big(\phi^0(\boldsymbol{\lambda}) - \phi^k(\boldsymbol{\lambda})\Big).\end{aligned} \tag{16}$$

By the convexity of F, we have that

$$\sum\nolimits_{p=0}^{k-1}\frac{1}{k}F(\hat{\mathbf{x}}^p) \ \ge \ F\Big(\sum\nolimits_{p=0}^{k-1}\frac{1}{k}\hat{\mathbf{x}}^p\Big),$$

which implies that $\sum_{p=0}^{k-1} F(\hat{\mathbf{x}}^p) \ \ge \ kF(\tilde{\mathbf{x}}^k)$. The analogous relation holds for $\sum_{p=0}^{k-1}\mathbf{r}(\hat{\mathbf{x}}^p) \ \ge \ k\mathbf{r}(\tilde{\mathbf{x}}^k)$, since it is a linear (convex) mapping. We also have that $\sum_{p=0}^{k-1} F(\mathbf{x}^*) = kF(\mathbf{x}^*)$. Hence, (16) can be expressed as

$$kF(\tilde{\mathbf{x}}^k) - kF(\mathbf{x}^*) + k\langle \boldsymbol{\lambda}, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle \ \le \ \frac{1}{2\tau}\Big(\phi^0(\boldsymbol{\lambda}) - \phi^k(\boldsymbol{\lambda})\Big),$$

or,

$$F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*) + \langle \boldsymbol{\lambda}, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle \ \le \ \frac{1}{2k\tau}\phi^0(\boldsymbol{\lambda}), \tag{17}$$

because for any $\boldsymbol{\lambda} \in \mathbb{R}^m$, we have $\phi^k(\boldsymbol{\lambda}) \ge 0$.

The above inequality is true for all $\boldsymbol{\lambda} \in \mathbb{R}^m$, hence it must also hold for any point in the ball $\mathcal{B} = \{\boldsymbol{\lambda} \mid \|\boldsymbol{\lambda}\| \le 1\}$. We now let $\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}^k \triangleq \arg\max_{\boldsymbol{\lambda}\in\mathcal{B}}\langle \boldsymbol{\lambda}, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle$ and rewrite the above relation as

$$F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*) + \|\mathbf{r}(\tilde{\mathbf{x}}^k)\| \ \le \ \frac{1}{2k\tau}\phi^0(\tilde{\boldsymbol{\lambda}}^k),$$

where we used $\langle \tilde{\boldsymbol{\lambda}}^k, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle = \|\mathbf{r}(\tilde{\mathbf{x}}^k)\|$. Finally, the term on the right-hand side can be bounded as

$$\begin{aligned}\phi^0(\tilde{\boldsymbol{\lambda}}^k) &= \sum\nolimits_{i=1}^N \rho\|\mathbf{A}_i(\mathbf{x}_i^0 - \mathbf{x}_i^*)\|^2 + \frac{1}{\rho}\|\bar{\boldsymbol{\lambda}}^0 - \tilde{\boldsymbol{\lambda}}^k\|^2 \\ &\le \sum\nolimits_{i=1}^N \rho\|\mathbf{A}_i(\mathbf{x}_i^0 - \mathbf{x}_i^*)\|^2 + \frac{1}{\rho}(\|\bar{\boldsymbol{\lambda}}^0\| + 1)^2,\end{aligned}$$

which gives the desired result. ∎

The advantage of this bound is that the computation complexity can be specified in advance as long as we can figure out the diameters of the primal constraint sets $\mathcal{X}_i$. When the primal solution $\tilde{\mathbf{x}}^k$ is not feasible, however, it is possible that $F(\tilde{\mathbf{x}}^k) - F^* < 0$. Nonetheless, the bound in (15) can be still useful if the primal residual can be tightly bounded, i.e., as is pointed out in [18], if $\|\mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{b}\| < \delta$ for a relatively small $\delta > 0$, then a lower bound of $F(\tilde{\mathbf{x}}^k) - F^*$ is given by

$$F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*) \geq \langle \boldsymbol{\lambda}^*, \mathbf{A}\tilde{\mathbf{x}}^k - \mathbf{b} \rangle \geq -\delta \|\boldsymbol{\lambda}^*\|,$$

where $\boldsymbol{\lambda}^*$ is a component of $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, which is a saddle point of (2).

Theorem 2 *Assume (A1)–(A3). Recall that* $\tilde{\mathbf{x}}^k = \frac{1}{k}\sum_{p=0}^{k-1} \hat{\mathbf{x}}^p$ *denotes the ergodic average of the primal variable sequence generated by ADAL up to iteration* k *and* $\mathbf{r}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$ *denotes the residual at* $\mathbf{x}$*. Let* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ *be a saddle point of* (2)*. Then, for all* k

(a)

$$|F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*)| \leq \frac{1}{2k\tau} \max\{\phi^0(\mathbf{0}), \phi^0(2\boldsymbol{\lambda}^*)\},$$

where $\phi^0(\boldsymbol{\lambda}) = \sum_{i=1}^N \rho \|\mathbf{A}_i(\mathbf{x}_i^0 - \mathbf{x}_i^*)\|^2 + \frac{1}{\rho}\|\bar{\boldsymbol{\lambda}}^0 - \boldsymbol{\lambda}\|^2$.

(b)

$$\|\mathbf{r}(\tilde{\mathbf{x}}^k)\| \leq \frac{1}{2k\tau}\Bigg[\sum_{i=1}^N \rho \|\mathbf{A}_i(\mathbf{x}_i^0 - \mathbf{x}_i^*)\|^2 + \frac{2}{\rho}\left(\|\bar{\boldsymbol{\lambda}}^0 - \boldsymbol{\lambda}^*\|^2 + 1\right)\Bigg].$$

Proof: (a) The inequality (17) is true for all $\boldsymbol{\lambda} \in \mathbb{R}^m$, hence letting $\boldsymbol{\lambda} = \mathbf{0}$ yields

$$F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*) \;\leq\; \frac{1}{2k\tau}\phi^0(\mathbf{0}). \tag{18}$$

Let $\boldsymbol{\lambda}^*$ be a dual optimal solution. Then, from the saddle point inequality, we have

$$F(\mathbf{x}^*) \;\leq\; F(\tilde{\mathbf{x}}^k) + \langle \boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle. \tag{19}$$

Next, we find an upper bound of the term $\langle \boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle$. We add $\langle \boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle$ to both sides of (19) to obtain

$$\langle \boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle \;\leq\; F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*) + \langle 2\boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle.$$

Using relation (17) again with $\boldsymbol{\lambda} = 2\boldsymbol{\lambda}^*$ to bound the right-hand side of the above equation, we obtain

$$\langle \boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle \;\leq\; \frac{1}{2k\tau}\phi^0(2\boldsymbol{\lambda}^*).$$

Combining this with relation (19), we further obtain

$$F(\mathbf{x}^*) - F(\tilde{\mathbf{x}}^k) \;\leq\; \frac{1}{2k\tau}\phi^0(2\boldsymbol{\lambda}^*).$$

From this and relation (18), the desired result follows.

(b) We next bound the residual $\|\mathbf{r}(\tilde{\mathbf{x}}^k)\|$. Using relation (17) with $\boldsymbol{\lambda} = \boldsymbol{\lambda}^* + \frac{\mathbf{r}(\tilde{\mathbf{x}}^k)}{\|\mathbf{r}(\tilde{\mathbf{x}}^k)\|}$, we have

$$F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*) + \langle \boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle + \|\mathbf{r}(\tilde{\mathbf{x}}^k)\|$$

$$\leq\; \frac{1}{2k\tau}\phi^0\left(\boldsymbol{\lambda}^* + \frac{\mathbf{r}(\tilde{\mathbf{x}}^k)}{\|\mathbf{r}(\tilde{\mathbf{x}}^k)\|}\right). \tag{20}$$

Using the saddle point inequality together with the fact that $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a primal-dual optimal pair, we obtain

$$F(\tilde{\mathbf{x}}^k) + \langle \boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle \geq F(\mathbf{x}^*) + \langle \boldsymbol{\lambda}^*, \mathbf{r}(\mathbf{x}^*)\rangle,$$

which implies

$$F(\tilde{\mathbf{x}}^k) - F(\mathbf{x}^*) + \langle \boldsymbol{\lambda}^*, \mathbf{r}(\tilde{\mathbf{x}}^k)\rangle \geq 0.$$

Combining this with relation (20), we obtain

$$\|\mathbf{r}(\tilde{\mathbf{x}}^k)\| \;\leq\; \frac{1}{2k\tau}\phi^0\left(\boldsymbol{\lambda}^* + \frac{\mathbf{r}(\tilde{\mathbf{x}}^k)}{\|\mathbf{r}(\tilde{\mathbf{x}}^k)\|}\right).$$

From the definition of the Lyapunov/Merit function $\phi^k(\boldsymbol{\lambda})$ in (14), the right-hand side can be represented as

$$\begin{aligned}
&\phi^0\left(\boldsymbol{\lambda}^* + \frac{\mathbf{r}(\tilde{\mathbf{x}}^k)}{\|\mathbf{r}(\tilde{\mathbf{x}}^k)\|}\right) \\
&= \sum_{i=1}^N \rho\|\mathbf{A}_i(\mathbf{x}_i^0 - \mathbf{x}_i^*)\|^2 + \frac{1}{\rho}\left\|\bar{\boldsymbol{\lambda}}^0 - \boldsymbol{\lambda}^* + \frac{\mathbf{r}(\tilde{\mathbf{x}}^k)}{\|\mathbf{r}(\tilde{\mathbf{x}}^k)\|}\right\|^2 \\
&= \sum_{i=1}^N \rho\|\mathbf{A}_i(\mathbf{x}_i^0 - \mathbf{x}_i^*)\|^2 + \frac{2}{\rho}\left(\|\bar{\boldsymbol{\lambda}}^0 - \boldsymbol{\lambda}^*\|^2 + 1\right),
\end{aligned}$$

from which the desired result follows. ∎

Theorem 2 characterizes that the theoretical complexity for the algorithm to reach an ϵ-optimal solution both in terms of primal optimality gap and feasibility residual is $O(\frac{1}{\epsilon})$ iterations. This result is particularly useful for MPC applications where frequent re-optimization for different time horizon is often required in practice, as discussed in Section II-B.

IV. Numerical Results

In order to compare our approach, we also present results for the popular ADMM method [19]. In all numerical experiments, we randomly generate the coupling graphs $\mathcal{E}^t$. For simplicity, we assume $\mathcal{E}^t = \mathcal{E}$ and the graph is connected.

We pre-determine the initial and final states, by introducing the equality constraints $\mathbf{x}^1 = \mathbf{c}$ and $\mathbf{x}^H = \mathbf{d}$, where the $\mathbf{c}$, $\mathbf{d}$ are randomly generated vectors for each problem. We also define the running cost to be a quadratic control cost, i.e., $V(\mathbf{x}, \mathbf{u}, t) = \sum_{i=1}^N \sum_{t=1}^{H-1} (\mathbf{u}_i^t)^T \mathbf{Q}_i^t \mathbf{u}_i^t$. Here, we define the cost matrices to be positive definite and also time-invariant, that is $\mathbf{Q}_i^t = \mathbf{Q}_i$ for all t. We generate their entries by randomly sampling from a uniform distribution in $[1, 10]$. The state and input dimensions are taken as $n_i = 2$ and $p_i = 2$, respectively, (unless otherwise noted) and we also impose box constraints on the states $\mathbf{x}_i^t \in [-2, 2]$ and inputs $\mathbf{u}_i^t \in [-1, 1]$ for all i and t.

In all the cases presented below, we have set the initial values of the primal and dual variables to 1. For the penalty parameter ρ, we have found that fastest convergence for ADAL is obtained for values $\rho \in [1, 10]$, while at the same time preventing ill-conditioning. After extensive sensitivity analysis in our simulations, we found that ADMM requires relatively larger values of $\rho \in [4, 20]$, compared to ADAL; cf. Section II Remark 1.

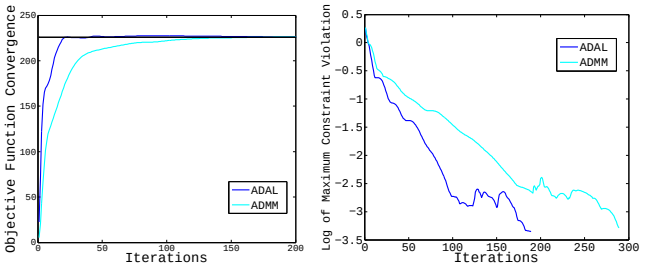

Fig. 1. Comparison of the two different distributed algorithms, ADAL and ADMM, for a problem with $N = 20$ subsystems, a horizon length of $H = 10$ time steps and a maximum degree 6 for the undirected coupling graph.

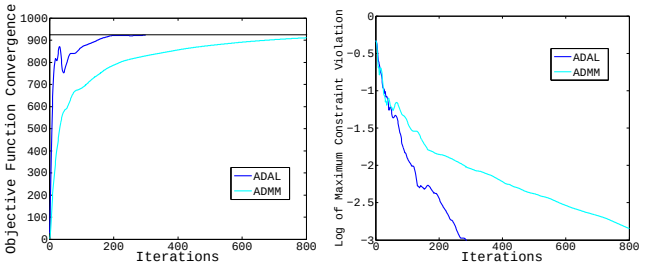

Fig. 2. Comparison of the two different distributed algorithms, ADAL and ADMM, for a problem with $N = 20$ subsystems, a horizon length of $H = 10$ time steps and a maximum degree 12 for the undirected coupling graph.

Fig. 1 depicts typical convergence results obtained after applying ADAL and ADMM on a problem with $N = 20$ subsystems and a horizon length of $H = 10$ time steps. The results correspond to a system with an undirected coupling graph whose maximum degree is 6. Considering the per-iteration computational complexity of ADAL is almost the same as ADMM, we can observe that ADAL converges significantly faster than ADMM. Next, we consider how the density of the coupling graph affects the performance of both methods. Towards this goal, we define a problem with $N = 20$, $H = 10$ as before, but now the maximum degree is 12. The results are shown in Fig. 2. It is obvious that both methods slow down for denser graphs since the increased coupling naturally introduces the need for more coordination between the coupled subsystems. Note that this behavior has also been observed in [14], where the authors test ADMM on DMPC problems. Nevertheless, Fig. 2 shows that the impact of the increased coupling is less severe on ADAL than on ADMM, and ADAL outperforms ADMM by a significant margin.

V. Conclusions

In this paper, we presented an Augmented Lagrangian decomposition method (ADAL) and characterized its computational complexity. We showed that the algorithm generates an ϵ-optimal and ϵ-feasible solution using the ergodic average of the sequence of primal variables under some mild assumptions such as the general convexity of the problems. The results in this paper have the potential to significantly improve the performance of distributed MPC problems, where preconditioning of computational complexity is important. Numerical experiments demonstrate that our algorithm outperforms ADMM by a significant margin.

References

[1] M. Morari and J. H. Lee, "Model predictive control: Past, present and future," *Computers and Chemical Engineering*, vol. 23, pp. 667–682, 1997.

[2] A. Bemporad and M. Morari, "Robust model predictive control: A survey," in *Robustness in identification and control*, ser. Lecture Notes in Control and Information Sciences. Springer London, 1999, vol. 245, pp. 207–226.

[3] S. J. Qin and T. A. Badgwell, "An overview of nonlinear model predictive control applications," in *Nonlinear Predictive Control*. Verlag, 2000, pp. 369–392.

[4] S. Richter, M. Morari, and C. Jones, "Towards computational complexity certification for constrained MPC based on lagrange relaxation and the fast gradient method," in *50th IEEE Conference on Decision and Control and European Control Conference*, Dec 2011, pp. 5223–5229.

[5] S. Richter, C. Jones, and M. Morari, "Computational complexity certification for real-time mpc with input constraints based on the fast gradient method," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1391–1403, June 2012.

[6] P. Patrinos and A. Bemporad, "An accelerated dual gradient-projection algorithm for embedded linear model predictive control," *IEEE Transactions on Automatic Control*, vol. 59, no. 1, pp. 18–33, Jan 2014.

[7] P. Giselsson, "Gradient-based distributed model predictive control," Ph.D. dissertation, Lund University, Nov 2012.

[8] V. Nedelcu, I. Necoara, and Q. Tran-Dinh, "Computational complexity of inexact gradient augmented lagrangian methods: Application to constrained mpc," *SIAM J. Optimization and Control*, vol. 52, no. 5, pp. 3109–3134, 2014.

[9] E. Camponogara, D. Jia, B. Krogh, and S. Talukdar, "Distributed model predictive control," *Control Systems, IEEE*, vol. 22, no. 1, pp. 44–52, 2002.

[10] P. D. Christofides, R. Scattolini, D. M. de la Pena, and J. Liu, "Distributed model predictive control: a tutorial review," *Proc. Chem. Process Control*, vol. 8, p. 22pp, 2012.

[11] Y. Wakasa, M. Arakawa, K. Tanaka, and T. Akashi, "Decentralized model predictive control via dual decomposition," in *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, 2008, pp. 381–386.

[12] D. Bertsekas, A. Nedic, and A. Ozdaglar, "Convex analysis and optimization," in *Series in Computational Mathematics*. Athena Scientific, 2003.

[13] N. Chatzipanagiotis, D. Dentcheva, and M. Zavlanos, "An augmented Lagrangian method for distributed optimization," *Mathematical Programming*, vol. 152, no. 1-2, pp. 405–434, 2015.

[14] C. Conte, T. Summers, M. Zeilinger, M. Morari, and C. Jones, "Computational aspects of distributed optimization in model predictive control," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, 2012, pp. 6819–6824.

[15] J. Mota, J. Xavier, P. Aguiar, and M. Puschel, "Distributed admm for model predictive control and congestion control," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, 2012, pp. 5110–5115.

[16] J. Eckstein, "The alternating step method for monotropic programming on the connection machine CM-2." *ORSA Journal on Computing*, vol. 5, no. 1, p. 84, 1993.

[17] N. Chatzipanagiotis, A. Petropulu, and M. Zavlanos, "A distributed algorithm for cooperative relay beamforming," in *American Control Conference (ACC), 2013*, June 2013, pp. 3796–3801.

[18] G. Lan and R. D. C. Monteiro, "Iteration-complexity of first-order penalty methods for convex programming," *Math. Program.*, vol. 138, pp. 115–139, 2013.

[19] J. Eckstein and D. P. Bertsekas, "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming,*, vol. 55, pp. 293–318, 1992.