Inexact Fast Alternating Minimization Algorithm for Distributed Model Predictive Control

Ye Pu[1], Melanie N. Zeilinger[2] and Colin N. Jones[1]

***Abstract*— This paper presents a new distributed optimization technique, the inexact fast alternating minimization algorithm (inexact FAMA), that allows for inexact local computation as well as for errors resulting from limited communication. We show that inexact FAMA is equivalent to the inexact accelerated proximal-gradient method applied to the dual problem and derive an upper-bound on the number of iterations for convergence for inexact FAMA. The second contribution of this work is that a weakened assumption for FAMA, as well as for its inexact version, is presented. The new assumption allows the strongly convex objective in the optimization problem to be subject to convex constraints, while still guaranteeing convergence of the algorithm, which facilitates its application to control problems. We apply inexact FAMA to distributed MPC problems and derive the convergence properties of the algorithm for this special case. By employing the upper-bound on the number of iterations, sufficient conditions on the errors are provided, which ensure converge of the algorithm. Finally, we demonstrate the performance of the algorithm and the theoretical findings using a randomly generated distributed MPC example.**

I. Introduction

Various distributed MPC schemes have been proposed for large-scale applications, which consist of many coupled sub-systems, and must act based on local communication. Examples include power grids, transportation and building networks. Available distributed MPC methods can be roughly classified into two categories: non-iterative, e.g. [8] and [9], and iterative methods, e.g. [6]. While iterative methods are generally less conservative, they require the solution of a global optimization problem in a distributed way subject to communication constraints.

A key challenge is therefore to solve this global optimization problem efficiently by distributed optimization techniques. Powerful tools are offered by decomposition based optimization methods, which allow one to decompose the original large optimization problem into small problems and solve each of them separately. Various decomposition based techniques have been proposed for distributed MPC, for example [11] and [12] based on primal decomposition and [18] and [5] based on dual decomposition.

[1]Laboratoire d'Automatique, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland {y.pu,colin.jones}@epfl.ch.

[2]Department of Electrical Engineering and Computer Sciences, UC Berkeley, CA 94702, USA and Department of Empirical Inference, Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany mzeilinger@berkeley.edu.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ ERC Grant Agreement n. 307608. The research of M. N. Zeilinger has received funding from the EU FP7 under grant agreement no. PIOF-GA-2011-301436-"COGENT".

Alternating direction methods are also dual decomposition based methods, which have attracted a lot of attention in recent years due to their good performance for solving large or distributed optimization problems, e.g. [10] and [4]. In [16], the alternating direction method of multipliers (ADMM), a subclass of alternating direction methods, has been proposed for the solution of distributed MPC problems.

In this work, we propose another alternating direction method called the fast alternating minimization algorithm (FAMA) for solving distributed MPC problems and derive an inexact FAMA algorithm providing convergence despite computation errors in the iteration steps of the algorithm. The difference between ADMM and FAMA is that FAMA requires stronger assumptions on the objective of the optimization problem, which can, however, be satisfied by standard MPC problems. In return, FAMA offers strong theoretical results, i.e. a better convergence rate of $O(\frac{1}{k^2})$ and a complexity upper-bound on the number of iterations to achieve a certain solution accuracy [13].

In practice, distributed optimization methods suffer from inexact solutions of the local problems and from communication errors. Inexact optimization algorithms address this problem, analysing the effect of errors on the overall algorithm and providing conditions under which convergence can still be guaranteed. Seminal works include [7] and [15]. In [7], the authors propose an inexact decomposition algorithm for solving distributed optimization problems by employing smoothing techniques and an excessive gap condition. In [15], an inexact proximal-gradient method, as well as its accelerated version, are introduced. The proximal gradient method, also known as the iterative shrinkage-thresholding algorithm (ISTA) [2], has two main steps: the first one is to compute the gradient of the smooth objective and the second one is to solve the proximal minimization. The conceptual idea of the inexact proximal-gradient method is to allow errors in these two steps. The results in [15] show convergence properties of the inexact proximal-gradient method and provide conditions on the errors, under which convergence of the algorithm can be guaranteed.

In this paper, we present a new inexact algorithm, inexact FAMA, and apply it to distributed MPC problems. Inexact FAMA is an extension of the FAMA algorithm, see [10] and [13]. By extending the results in [15], we show the convergence properties of inexact FAMA and derive a complexity upper-bound. The main contributions of this work are the following:

- We show that applying the proposed inexact FAMA to the primal problem is equivalent to applying the inexact

accelerated proximal-gradient method (inexact APGM) in [15] to the dual problem. Based on the results in [15], an upper bound on the number of iterations to achieve a given accuracy of the dual value function is derived.
- Extending previous results in [10] and [14], we show that all convergence properties of FAMA and inexact FAMA are preserved when imposing convex constraints on the strongly convex objective.
- We apply inexact FAMA to distributed MPC problems, where computation errors of the local optimization problems and consensus errors are considered. By simplifying the upper bound of inexact FAMA for this special case, sufficient conditions on the errors for convergence are provided.
- We demonstrate the performance and the theoretical results of inexact FAMA for a randomly generated example of a distributed MPC problem.

II. Preliminaries

A. Notation

Let $f : \Theta \to \Omega$ be a function. The conjugate function of f is defined as $f^\star(y) := \sup_{x\in\Theta}(y^T x - f(x))$. For a conjugate function it holds that $p \in \partial f(q) \Leftrightarrow q \in \partial f^\star(p)$, where $\partial f(\cdot)$ denotes the set of sub-gradients of the function f at a given point. Let f be a strongly convex function. σ_f denotes the convexity modulus $\langle p-q, x-y\rangle \geq \sigma_f \|x-y\|^2$, where $p \in \partial f(x)$, $q \in \partial f(y)$. $L(f)$ denotes a Lipschitz constant of the function f, i.e. $\|f(x_1) - f(x_2)\| \leq L(f)\|x_1 - x_2\|$, $\forall x_1, x_2 \in \Theta$. Let C be a matrix. $\rho(C)$ denotes the matrix norm of $C^T C$. The proximity operator is defined as

$$\text{prox}_f(y) = \text{argmin}_x \quad f(x) + \frac{1}{2}\|x-y\|^2 \ . \tag{1}$$

We note the following equivalence

$$x^\star = \text{prox}_f(y) \Longleftrightarrow y - x^\star \in \partial f(x^\star) \ . \tag{2}$$

We refer to [3] and [1] for details on the definitions and properties above. In this paper, $\tilde{\cdot}$ is used to denote an inexact solution of an optimization problem. The proximity operator with an extra subscript ϵ, i.e. $\tilde{x} = \text{prox}_{f,\epsilon}(y)$, means that a maximum computation error ϵ is allowed in the proximal objective function

$$f(\tilde{x}) + \frac{1}{2}\|\tilde{x}-y\|^2 \ \leq \ \epsilon + \min_x \left\{ f(x) + \frac{1}{2}\|x-y\|^2 \right\} \ . \tag{3}$$

B. Inexact Accelerated Proximal-Gradient Method

In this section, the inexact accelerated proximal-gradient method (inexact APGM) proposed in [15] will be introduced. Both the algorithm and the convergence theorem will be presented. Inexact APGM addresses optimization problems of the form in Problem II.1. The inexact APGM algorithm is given in Algorithm 1.

Problem II.1.

$$\min_{w\in\mathbb{R}^{n_w}} \quad \Phi(w) = \phi(w) + \psi(w) \ .$$

Assumption II.2.
- *ϕ is a convex function with Lipschitz continuous gradient with constant $L(\nabla\phi)$;*
- *ψ is a lower semi-continuous convex function, not necessarily smooth.*

Algorithm 1 Inexact Accelerated Proximal-Gradient Method

Require: Initialize $v^1 = w^0 \in \mathbb{R}^{n_w}$, $\alpha^1 = 1$ and $\tau < \frac{1}{L(\nabla\phi)}$
for $k = 1, 2, \cdots$ **do**
1: $\tilde{w}^k = \text{prox}_{\tau\psi,\epsilon^k}(v^k - \tau(\nabla\phi(v^k) + e^k))$
2: $\alpha^{k+1} = (1 + \sqrt{4(\alpha^k)^2 + 1})/2$
3: $v^{k+1} = \tilde{w}^k + (\alpha^k - 1)(\tilde{w}^k - \tilde{w}^{k-1})/\alpha^{k+1}$
end for

Inexact APGM in Algorithm 1 allows two kinds of errors: $\{e^k\}$ represents the error in the gradient calculations of ϕ, and $\{\epsilon^k\}$ represents the error in the computation of the proximal minimization in (3) at every iteration k. The following proposition states the convergence property of inexact APGM.

Proposition II.3 (Proposition 2 in [15]). *Let $\{\tilde{w}_k\}$ be generated by inexact APGM defined in Algorithm 1. If Assumption II.2 holds, then for any $k \geq 1$ we have*

$$\begin{aligned} &\Phi(\tilde{w}^k) - \Phi(w^\star) \\ &\leq \frac{2L(\nabla\phi)}{(k+1)^2}\left(\|w^0 - w^\star\| + 2\Gamma^k + \sqrt{2\Lambda^k}\right)^2 \ , \end{aligned}$$

where $\Phi(w) = \phi(w) + \psi(w)$,

$$\Gamma^k = \sum_{p=1}^{k} p\left(\frac{\|e^p\|}{L(\nabla\phi)} + \sqrt{\frac{2\epsilon^p}{L(\nabla\phi)}}\right), \ \Lambda^k = \sum_{p=1}^{k}\frac{p^2\epsilon^p}{L(\nabla\phi)} \ ,$$

and w^0 and $w^\star$ denote the initial solution for Algorithm 1 and the optimal solution of Problem II.1, respectively.

As proposed in [15], the upper-bound in Proposition II.3 allows the derivation of sufficient conditions on the error sequences $\{e^k\}$ and $\{\epsilon^k\}$ for the convergence of the algorithm to the optimal solution w^*
- The series $\{k\|e^k\|\}$ and $\{k\sqrt{\epsilon^k}\}$ are finitely summable and decrease at the rate $O(\frac{1}{k^2})$.
- The sequences $\{\|e^k\|\}$ and $\{\sqrt{\epsilon^k}\}$ decrease at the rate $O(\frac{1}{k^{2+\kappa}})$ for $\kappa \geq 0$.

If one of the above conditions holds, then the algorithm converges to the optimal solution. For $\kappa = 0$, i.e. the errors decrease at the rate $O(\frac{1}{k^2})$, the algorithm converges, but the convergence rate is poor $O(\frac{\log^2 k}{k^2})$.

III. Inexact Fast Alternating Minimization Algorithm

The accelerated proximal gradient method, as well as its inexact version, is limited to the case, where both objectives are a function of the same variable. Many problems and most importantly MPC problems are not of this type. In order to generalize the problem formulation, we employ the fast alternating minimization algorithm, which covers

optimization problems of the form in Problem III.1. In this section, we extend FAMA to the inexact case and present the theoretical convergence properties of the new inexact FAMA algorithm.

Problem III.1.

$$\begin{aligned} \min \quad & f(x) + g(z) \\ s.t. \quad & Ax + Bz = c \end{aligned}$$

with variables $x \in \mathbb{R}^{n_x}$ and $z \in \mathbb{R}^{n_z}$, where $A \in \mathbb{R}^{n_c \times n_x}$, $B \in \mathbb{R}^{n_c \times n_z}$ and $c \in \mathbb{R}^{n_c}$. $f : \mathbb{R}^{n_x} \to \mathbb{R}$ and $g : \mathbb{R}^{n_z} \to \mathbb{R}$ are convex functions. The Lagrangian of Problem III.1 is

$$L(x, z, \lambda) = f(x) + g(z) + \lambda^T (Ax + Bz - c) \ , \quad (4)$$

and the dual function is

$$\begin{aligned} D(\lambda) &= \inf_{x,z} \ L(x, z, \lambda) && (5a) \\ &= - \sup_x \left\{ \lambda^T Ax - f(x) \right\} && (5b) \\ &\quad - \sup_z \left\{ \lambda^T Bz - g(z) \right\} + \lambda^T c \\ &= - f^\star(A^T \lambda) - g^\star(B^T \lambda) + \lambda^T c \ , && (5c) \end{aligned}$$

where $f^\star$ and $g^\star$ are the conjugate functions of f and g. The dual problem of Problem III.1 is

Problem III.2.

$$\min \quad -D(\lambda) = \underbrace{f^\star(A^T \lambda)}_{\phi(\lambda)} + \underbrace{g^\star(B^T \lambda) - c^T \lambda}_{\psi(\lambda)} \ .$$

Assumption III.3.
- *f is a strongly convex function with convexity modulus σ_f,*
- *g is a convex function, not necessarily smooth.*

Algorithm 2 Inexact Fast alternating minimization algorithm (Inexact FAMA)

Require: Initialize $\alpha^1 = 1$, $\lambda^1 = \hat{\lambda}^1 = \lambda^0 \in \mathbb{R}^{N_b}$, and $\tau < \sigma_f / \rho(A)$
for $k = 1, 2, \cdots$ **do**
1: $\tilde{x}^k = \text{argmin}_x \ \{f(x) + \langle \hat{\lambda}^k, -Ax \rangle\} + \delta^k$
2: $\tilde{z}^k = \text{argmin}_z \ \{g(z) + \langle \hat{\lambda}^k, -Bz \rangle + \frac{\tau}{2} \|c - A\tilde{x}^k - Bz\|^2\} + \theta^k$
3: $\lambda^{k+1} = \hat{\lambda}^k + \tau(c - A\tilde{x}^k - B\tilde{z}^k)$
4: $\alpha^{k+1} = (1 + \sqrt{4(\alpha^k)^2 + 1})/2$
5: $\hat{\lambda}^{k+1} = \lambda^{k+1} + (\alpha^k - 1)(\lambda^{k+1} - \lambda^k)/\alpha^{k+1}$
end for

The inexact FAMA algorithm is presented in Algorithm 2, which allows errors in Step 1 and in Step 2 of the algorithm, i.e. both minimization problems are solved inexactly with errors δ^k and θ^k, respectively. Lemma III.4 shows that inexact FAMA in Algorithm 2 is equivalent to applying inexact APGM in Algorithm 1 to the dual problem in Problem III.2 with the following correspondence: the gradient computation error in Algorithm 1 is equal to $e^k = A\delta^k$ and the error of solving the proximal minimization is equal to $\epsilon^k = \tau^2 L(\psi) \|B\theta^k\| + \frac{\tau^2}{2} \|B\theta^k\|^2$. With this equivalence, shown in detail in Lemma III.4, the complexity bound in Proposition II.3 can be extended for the inexact FAMA algorithm. The proof of Lemma III.4 is an extension of the proof of Theorem 2 in [10] and the proof in Section 3 in [17].

Lemma III.4. *If Assumption III.3 is satisfied and inexact FAMA and inexact APGM are initialized with the same primal and dual starting points, then applying inexact FAMA in Algorithm 2 to Problem III.1 is equivalent to applying inexact APGM in Algorithm 1 to the dual problem defined in Problem III.2 with the errors $e^k = A\delta^k$ and $\epsilon^k = \tau^2 L(\psi) \|B\theta^k\| + \frac{\tau^2}{2} \|B\theta^k\|^2$, where $L(\psi)$ denotes the Lipschitz constant of the function ψ.*

Proof: The result is shown by proving that Step 1, 2 and 3 in Algorithm 2 are equivalent to Step 1 in Algorithm 1, i.e. the following equality holds:

$$\lambda^k = \text{prox}_{\tau\psi, \epsilon^k}(\hat{\lambda}^k - \tau(\nabla \phi(\hat{\lambda}^k) + e^k)) \ , \quad (6)$$

with $e^k = A\delta^k$ and $\epsilon^k = \tau^2 L(\psi) \|B\theta^k\| + \frac{\tau^2}{2} \|B\theta^k\|^2$. Step 2 in Algorithm 2 implies:

$$B^T \hat{\lambda}^k + \tau B^T (c - A\tilde{x}^k - Bz^k) \in \partial g(z^k) \ ,$$

where $z^k = \text{argmin}_z \ \{g(z) + \langle \hat{\lambda}^k, -Bz \rangle + \frac{\tau}{2} \|c - A\tilde{x}^{k+1} - Bz\|^2\} = \tilde{z}^k - \theta^k$. From the property of the conjugate function $p \in \partial f(q) \Leftrightarrow q \in \partial f^\star(p)$, it follows that

$$z^k \in \partial g^\star(B^T \hat{\lambda}^k + \tau B^T (c - A\tilde{x}^k - Bz^k)) \ .$$

By multiplying with B and subtracting c on both sides, we obtain

$$Bz^k - c \in B \partial g^\star(B^T \hat{\lambda}^k + \tau B^T (c - A\tilde{x}^k - Bz^k)) - c \ .$$

By multiplying with τ and adding $\hat{\lambda}^k + \tau(c - A\tilde{x}^k - Bz^k)$ on both sides, we get

$$\begin{aligned} \hat{\lambda}^k - \tau A\tilde{x}^k \in \tau B \partial g^\star(B^T \hat{\lambda}^k + \tau B^T (c - A\tilde{x}^k - Bz^k)) \\ - \tau c + \hat{\lambda}^k + \tau(c - A\tilde{x}^k - Bz^k) \ . \end{aligned}$$

Since $\psi(\lambda) = g^\star(B^T \lambda) - c^T \lambda$, we have $\partial \psi(\lambda) = B \partial g^\star(B^T \lambda) - c$, which implies

$$\begin{aligned} \hat{\lambda}^k - \tau A\tilde{x}^k \in \tau \partial \psi(\hat{\lambda}^k + \tau(c - A\tilde{x}^k - Bz^k)) \\ + \hat{\lambda}^k + \tau(c - A\tilde{x}^k - Bz^k) \ . \end{aligned}$$

Since $z^k = \tilde{z}^k - \theta^k$, it follows that

$$\begin{aligned} \hat{\lambda}^k - \tau A\tilde{x}^k \in \tau \partial \psi(\hat{\lambda}^k + \tau(c - A\tilde{x}^k - B\tilde{z}^k + B\theta^k)) \\ + \hat{\lambda}^k + \tau(c - A\tilde{x}^k - B\tilde{z}^k + B\theta^k) \ . \end{aligned}$$

By Step 3 in Algorithm 2, the above equation results in

$$\hat{\lambda}^k - \tau A\tilde{x}^k \in \tau \partial \psi(\lambda^k + \tau B\theta^k) + \lambda^k + \tau B\theta^k \ .$$

From Step 1 in Algorithm 2 and the property of the conjugate function $p \in \partial f(q) \Leftrightarrow q \in \partial f^\star(p)$, we obtain

$$\hat{\lambda}^k - \tau A(\nabla f^\star(A^T \hat{\lambda}^k) + \delta^k) \in \tau \partial \psi(\lambda^k + \tau B\theta^k) + \lambda^k + \tau B\theta^k \ .$$

By definition of the function ϕ, we get

$$\hat{\lambda}^k - \tau(\nabla\phi(\hat{\lambda}^k) + A\delta^k) \in \tau\partial\psi(\lambda^k + \tau B\theta^k) + \lambda^k + \tau B\theta^k \ ,$$

which is equivalent to

$$\lambda^k = \text{prox}_{\tau\psi}(\hat{\lambda}^k - \tau(\nabla\phi(\hat{\lambda}^k) + e^k)) - \tau B\theta^k \ ,$$

with $e^k = A\delta^k$. In order to complete the proof of equation (6), we need to show that λ^k is an inexact solution of the proximal operator as defined in equation (3) with the error $\epsilon^k = \tau^2 L(\psi)\|B\theta^k\| + \frac{\tau^2}{2}\|B\theta^k\|^2$, i.e. to prove

$$\tau\psi(\lambda^k) + \frac{1}{2}\|\lambda^k - v\|^2 \ \leq \ \epsilon^k + \min_\lambda \left\{\tau\psi(\lambda) + \frac{1}{2}\|\lambda - v\|^2\right\},$$

where $\nu = \hat{\lambda}^k - \tau(\nabla\phi(\hat{\lambda}^k) + A\delta^k)$. Finally, using

$$\begin{aligned}
&\tau\psi(\lambda^k + \tau B\theta^k) + \frac{1}{2}\|\lambda^k + \tau B\theta^k - \nu\|^2 \\
&- \tau\psi(\lambda^k) - \frac{1}{2}\|\lambda^k - \nu\|^2 \\
&\leq \tau(\psi(\lambda^k + \tau B\theta^k) - \psi(\lambda^k)) + \frac{1}{2}\|\tau B\theta^k\|^2 \\
&\leq \tau^2 L(\psi)\|B\theta^k\| + \frac{\tau^2}{2}\|B\theta^k\|^2 = \epsilon^k \ ,
\end{aligned}$$

equation (6) is proved. ■

Based on the equivalence shown in Lemma III.4 , we can now derive an upper-bound on the sub-optimality of the dual function value of the sequence $\{\lambda^k\}$ in Theorem III.5.

Theorem III.5. *Let $\{\lambda^k\}$ be generated by the inexact FAMA in Algorithm 2. If Assumption III.3 holds, then for any $k \geq 1$*

$$D(\lambda^\star) - D(\lambda^k) \leq \frac{2L(\nabla\phi)}{(k+1)^2}\left(\|\lambda^0 - \lambda^\star\| + 2\Gamma^k + \sqrt{2\Lambda^k}\right)^2, \tag{7}$$

where

$$\Gamma^k = \sum_{p=1}^{k} p\left(\frac{\|A\delta^p\|}{L(\nabla\phi)} + \tau\sqrt{\frac{2L(\psi)\|B\theta^p\| + \|B\theta^p\|^2}{L(\nabla\phi)}}\right) \ , \tag{8}$$

$$\Lambda^k = \sum_{p=1}^{k} \frac{p^2\tau^2(2L(\psi)\|B\theta^p\| + \|B\theta^p\|^2)}{2L(\nabla\phi)} \ , \tag{9}$$

and $L(\nabla\phi) = \sigma_f^{-1} \cdot \rho(A)$.

Proof: Lemma III.4 shows the equivalence between Algorithm 2 and Algorithm 1 with $e^k = A\delta^k$ and $\epsilon^k = \tau^2 L(\psi)\|B\theta^k\| + \frac{\tau^2}{2}\|B\theta^k\|^2$. It therefore only remains to be shown that the dual defined in Problem III.2 satisfies Assumption II.2. $\phi(\lambda)$ and $\psi(\lambda)$ are both convex, since the conjugate functions and linear functions as well as their weighted sum are always convex (the conjugate function is the point-wise supremum of a set of affine functions). Furthermore, since $f(x)$ is strongly convex with σ_f by Assumption III.3, we know that $f^\star$ has Lipschitz-continuous gradient with Lipschitz constant

$$L(\nabla f^\star) = \sigma_f^{-1} \ .$$

It follows that the function ϕ has Lipschitz-continuous gradient $\nabla\phi$ with Lipschitz constant

$$L(\nabla\phi) = \sigma_f^{-1} \cdot \rho(A) \ .$$

Hence, the functions ϕ and ψ satisfy Assumption II.2. Proposition II.3 applies and completes the proof of the upper-bound in inequality (7). ■

After showing equivalence of inexact FAMA and inexact APGM in Lemma III.4 and the upper-bound in Theorem III.5, the conditions on the errors for the convergence of inexact APGM presented in Section II-B can directly be extended to inexact FAMA with the errors defined in Lemma III.4.

IV. Inexact Fast alternating minimization algorithm with weakened assumption

For many problems, it is desirable to choose a splitting where a convex constraint is additionally imposed on the strongly convex objective f. An example is distributed MPC discussed in the next section, where the global optimization problem is split into local problems according to the dynamic couplings of the subsystems. However, the local problems have local constraints, which can be considered as indicator functions in the cost. Since indicator functions are only weakly convex, Assumption III.3 will be violated. In this section, we present a weakened assumption in Assumption IV.1, which allows the strongly convex objective in Problem III.1 to be subject to convex constraints, and still guarantees convergence of the algorithm.

Assumption IV.1.
- *f is a strongly convex function defined on a convex set $\mathbb{C}$. σ_f denotes the convexity modulus of f.*
- *g is a convex function, not necessarily smooth.*

We define the dual function $\bar{f}$ of f in equality (10). It has the similar meaning to the conjugate function $f^\star$, but the domain of the function is given by the convex constraint. Before showing the convergence proof of FAMA under Assumption IV.1, we first show Lipschitz continuity of $\nabla\bar{f}(\cdot)$, which is required by the convergence proof in Theorem IV.3.

$$\bar{f}(y) := \sup_{x\in\mathbb{C}} \ y^T x - f(x) \ = - \inf_{x\in\mathbb{C}} \ f(x) - y^T x \ . \tag{10}$$

Lemma IV.2. *If the function f is strongly convex with the convexity modulus σ_f and the constraint $\mathbb{C}$ is a convex set, then the function $\bar{f}$ has a Lipschitz continuous gradient*

$$\nabla\bar{f}(y) \ = \ -x^\star(y) \ , \tag{11}$$

with

$$x^\star(y) \ = \ \textit{argmin}_{x\in\mathbb{C}} \ f(x) - y^T x \ . \tag{12}$$

For any y_1 and y_2, we have

$$\|\nabla\bar{f}(y_1) - \nabla\bar{f}(y_2)\| \leq L(\nabla\bar{f})\|y_1 - y_2\| \ , \tag{13}$$

with a Lipschitz constant $L(\nabla\bar{f}) = \frac{1}{\sigma_f}$.

Proof: The proof follows directly from Theorem 1 in [14]. ■

Adding a convex constraint $\mathbb{C}$ to the objective f in Problem III.1 results in the following modification of Step 1 in Algorithm 2

$$\tilde{x}^k = \text{argmin}_{x\in\mathbb{C}} \left\{f(x) + \langle \hat{\lambda}^k, -Ax\rangle\right\} + \delta^k \ . \tag{14}$$

The following theorem shows the convergence property of the inexact FAMA algorithm with modified Step 1 in (14).

Theorem IV.3. *If Assumption IV.1 holds and the inexact solution $\tilde{x}^k$ is feasible for all $k \geq 0$, i.e. $\tilde{x}^k \in \mathbb{C}$, the sequence $\{\lambda_k\}$ generated by the inexact FAMA algorithm with the modified first step in (14) satisfies inequality (7).*

Proof: Since the definition of the function $\bar{f}(\cdot)$ corresponds to the definition of the conjugate function of $f(\cdot)$ with the function domain given by $\mathbb{C}$, then the property $p \in \partial f(q) \Leftrightarrow q \in \partial f^\star(p)$ is preserved. Hence Lemma III.4 still holds, namely that the inexact FAMA with the modified first step in (14) is equivalent to the inexact accelerated proximal-gradient method on the dual problem. It remains to be shown that the dual defined in (15b) satisfies the convergence assumption of the inexact accelerated proximal-gradient method in Assumption 1

$$\begin{aligned} D(\lambda) = &- \sup_{x\in\mathbb{C}} \left\{\lambda^T Ax - f(x)\right\} \\ &- \sup_{z} \left\{\lambda^T Bz - g(z)\right\} + \lambda^T c \qquad (15a) \\ = &\underbrace{-\bar{f}(A^T\lambda)}_{-\bar{\phi}(\lambda)} + \underbrace{-g^\star(B^T\lambda) + c^T\lambda}_{-\psi(\lambda)} \ . \qquad (15b) \end{aligned}$$

The dual problem is to minimize $-D(\lambda) = \bar{\phi}(\lambda) + \psi(\lambda)$. $\bar{\phi}(\lambda)$ and $\psi(\lambda)$ are both convex by definition. From Lemma IV.2, it follows that $\bar{\phi}(\lambda)$ has a Lipschitz continuous gradient $\nabla\bar{\phi}(\lambda) = A\nabla\bar{f}(A^T\lambda)$ with a Lipschitz constant $L(\nabla\phi(\lambda)) = \sigma_f^{-1} \cdot \rho(A)$. Hence, Assumption 1 is satisfied and (7) holds by Theorem III.5. ■

Theorem IV.3 allows for the application of FAMA, as well as its inexact version, to problems with additional convex constraints on the strongly convex part of the objective, which in particular enables the application of inexact FAMA to distributed MPC problems, discussed in the next section.

V. Inexact FAMA for Distributed Model Predictive Control

A. Distributed Model Predictive Control

We consider a network of M agents. The agents interact and communicate according to a fixed undirected graph $G = (V, E)$. The vertex set $V = \{1, 2, \cdots, M\}$ represents the agents and the edge set $E \subseteq V \times V$ specifies pairs of agents that interact and can communicate. If $(i, j) \in E$, we say that agents i and j are neighbors, and we denote by $\mathcal{N}_i = \{j|(i, j) \in E\}$ the set of the neighbors of agent i. Note that $\mathcal{N}_i$ includes i. The dynamics of the ith agent are given by the discrete-time linear dynamics

$$x_i(t+1) = \sum_{j\in\mathcal{N}_i} A_{ij}x_j(t) + B_{ij}u_j(t) \quad i = 1, 2, \cdots, M. \tag{16}$$

where A_{ij} and B_{ij} are the dynamic matrices. The states and inputs of agent i are subject to local convex constraints

$$x_i(t) \in \mathbb{X}_i \quad u_i(t) \in \mathbb{U}_i \quad i = 1, 2, \cdots, M \ . \tag{17}$$

The distributed MPC problem is given in Problem V.1. A similar problem formulation is considered in [6], where it is shown how stability can be guaranteed with the considered problem structure.

Problem V.1.

$$\begin{aligned} \min_{x,u} \quad & \sum_{i=1}^{M}\sum_{t=0}^{N-1} l_i(x_i(t), u_i(t)) + \sum_{i=1}^{M} l_i^f(x_i(N)) \\ s.t. \quad & x_i(t+1) = \sum_{j\in\mathcal{N}_i} A_{ij}x_j(t) + B_{ij}u_j(t) \\ & x_i(t) \in \mathbb{X}_i \ , \ u_i(t) \in \mathbb{U}_i \ , \\ & x_i(N) \in \mathbb{X}_i^f \ . \ x_i(0) = \bar{x}_i \ , \ i = 1, 2, \cdots, M \ . \end{aligned}$$

where $l_i(\cdot, \cdot)$ and $l_i^f(\cdot)$ are strictly convex stage cost functions and N is the horizon for the MPC problem. The state and input sequences along the horizon of agent i are denoted by $x_i = [x_i^T(0), x_i^T(1), \cdots, x_i^T(N)]^T$ and $u_i = [u_i^T(0), u_i^T(1), \cdots, u_i^T(N)]^T$. We denote the concatenations of the state sequences and input sequences of agent i and its neighbours by $x_{\mathcal{N}_i}$ and $u_{\mathcal{N}_i}$. The corresponding constraints are $x_{\mathcal{N}_i} \in \mathbb{X}_{\mathcal{N}_i}$ and $u_{\mathcal{N}_i} \in \mathbb{U}_{\mathcal{N}_i}$. We define $v = [x_1^T, x_2^T, \cdots, x_M^T, u_1^T, u_2^T, \cdots, u_M^T]^T$ to be the global variable and $z_i = [x_{\mathcal{N}_i}, u_{\mathcal{N}_i}]$ to be the local variables. $\mathbb{Z}_{\mathcal{N}_i} = \mathbb{X}_{\mathcal{N}_i} \times \mathbb{U}_{\mathcal{N}_i}$ denotes the local constraints on z_i and $H_i z_i = h_i$ denotes the dynamical constraint of sub-system i. Problem V.1 can be rewritten as

Problem V.2.

$$\begin{aligned} \min_{x,u} \quad & \sum_{i=1}^{M} f_i(z_i) \\ s.t. \quad & z_i \in \mathbb{C}_i \ , \ z_i = E_i v \ , \ i = 1, 2, \cdots, M \ . \end{aligned}$$

where f_i is the local cost function for agent i containing all the stage cost functions of the state and input sequences of agent i and its neighbours. The constraint $\mathbb{C}_i$ includes the constraint $\mathbb{Z}_{\mathcal{N}_i}$ and the dynamical constraint $H_i z_i = h_i$. E_i are the matrices, which select the local variables from the global variable.

B. IFAMA for Distributed Model Predictive Control

In this section, we apply inexact FAMA to Problem V.2. The idea is to split the global optimization problem into small and local problems according to the physical couplings of the sub-systems. Algorithm 3 presents the algorithm. We denote the ith component of v by $[v]_i$, i.e. $[v]_i = [x_i, u_i]$. Note that Step 2 in Algorithm 2 is simplified to be Step 3 in Algorithm 3, which requires only local communication. With the considered splitting in Problem V.2, δ_i^k and θ_i^k represent the computation error of the local problems and the consensus error.

Assumption V.3. *Every local cost function f_i in Problem V.2 is a strongly convex function with a convexity*

Algorithm 3 Inexact Fast alternating minimization algorithm for DMPC

Require: Initialize $\alpha^0 = 1$, $\lambda_i^{-1} = \hat{\lambda}_i^0 = \lambda_i^0$, and $\tau < \min_{1\le i\le M}\{\sigma_{f_i}\}$

for $k = 1, 2, \cdots$ **do**

1: $\tilde{z}_i^k = \text{argmin}_{z_i\in\mathbb{C}_i}\{f_i(z_i) + \langle\hat{\lambda}_i^k, -z_i\rangle\} + \delta_i^k$

2: Send $\tilde{z}_i^k$ to all the neighbours of agent i.

3: $[\tilde{v}^k]_i = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}[\tilde{z}_j^k]_i + \theta_i^k$.

4: Send $[\tilde{v}^k]_i$ to all the neighbours of agent i.

5: $\lambda_i^k = \hat{\lambda}_i^k + \tau(E_i\tilde{v}^k - \tilde{z}_i^k)$

6: $\alpha^{k+1} = (1 + \sqrt{4(\alpha^k)^2 + 1})/2$

7: $\hat{\lambda}_i^{k+1} = \lambda_i^k + (\alpha^k - 1)(\lambda_i^k - \lambda_i^{k-1})/\alpha^{k+1}$

8: $i = 1, 2, \cdots, M$.

end for

modulus σ_{f_i} and the set $\mathbb{Z}_{\mathcal{N}_i}$ is a convex set, for all $i = 1, \cdots, M$.

Recall the problem formulation of FAMA defined in Problem III.1. For Problem V.2 and Algorithm 3, the two objectives are $f(z) = \sum_{i=1}^M f_i(z_i)$ subject to $z_i \in \mathbb{C}_i$ for all $i = 1, \cdots, M$ and $g = 0$. The matrices are $A = I$, $B = [E_1^T, E_2^T, \cdots, E_M^T]^T$ and $c = 0$. Hence, Theorem IV.3 can be simplified as follows.

Corollary V.4. *Let $\{\lambda^k = [\lambda_1^{k^T}, \cdots, \lambda_M^{k^T}]^T\}$ be generated by Algorithm 3. If Assumption V.3 is satisfied and the inexact solutions $\tilde{z}_i^k$ are feasible for all $k \ge 1$, i.e. $\tilde{z}_i^k \in \mathbb{C}_i$, then for any $k \ge 1$*

$$D(\lambda^\star) - D(\lambda^k) \le \frac{2L(\nabla\phi)}{(k+1)^2}\left(\|\lambda^0 - \lambda^\star\| + 2\bar{\Gamma}^k + \sqrt{\bar{\Lambda}^k}\right)^2, \tag{18}$$

where $D(\cdot)$ is the dual function of Problem V.2, $\lambda^0 = [\lambda_1^{0^T}, \cdots, \lambda_M^{0^T}]^T$ and $\lambda^\star$ are the initial and the optimal sequences of Lagrangian multipliers, respectively. The Lipschitz constant $L(\nabla\phi)$ is equal to $\sigma_{f_{min}}^{-1}$, where $\sigma_{f_{min}} = \min\{\sigma_{f_1}, \cdots, \sigma_{f_M}\}$. And

$$\bar{\Gamma}^k = M\sum_{p=1}^k p\left(\frac{\delta_{\max}^p}{L(\nabla\phi)} + \tau\frac{\rho(E)\theta_{\max}^p}{\sqrt{L(\nabla\phi)}}\right) \quad , \tag{19}$$

$$\bar{\Lambda}^k = \frac{\tau^2\rho(E)^2M^2}{L(\nabla\phi)}\sum_{p=1}^k p^2\theta_{\max}^{p^2} \quad . \tag{20}$$

where $\delta_{\max}^p = \max\{\|\delta_1^p\|, \cdots, \|\delta_M^p\|\}$, $\theta_{\max}^p = \max\{\|\theta_1^p\|, \cdots, \|\theta_M^p\|\}$ and $E^T = [E_i^T, \cdots, E_M^T]^T$.

Proof: We split Problem V.2 such that the two objectives according to Problem III.1 are $f = \sum_{i=1}^M f_i$ and $g = 0$, and the matrices are $A = I$, $B = E$ and $c = 0$. By Assumption V.3, it follows that all the assumptions for Theorem IV.3 are satisfied. Hence, the sequence $\{\lambda^k\}$ generated by Algorithm 3 satisfies inequality (7) with Γ^k and Λ^k in (8) and (9) with $\delta^k = [\delta_1^{k^T}, \cdots, \delta_M^{k^T}]^T$ and $\theta^k = [\theta_1^{k^T}, \cdots, \theta_M^{k^T}]^T$. Since $g = 0$ and $c = 0$, it follows that $L(\nabla\psi) = 0$. Then we can simplify Γ^k and Λ^k and upper-bound them by $\bar{\Gamma}^k$ and $\bar{\Lambda}^k$

$$\Gamma^k = \sum_{p=1}^k p\left(\frac{\|\delta^p\|}{L(\nabla\phi)} + \tau\sqrt{\frac{\|B\theta^p\|^2}{L(\nabla\phi)}}\right) \le \bar{\Gamma}^k \quad ,$$

$$\Lambda^k = \sum_{p=1}^k \frac{p^2\tau^2\|B\theta^p\|^2}{2L(\nabla\phi)} \le \bar{\Lambda}^k \quad .$$

This proves inequality (18). Since $A = I$, then the Lipschitz constant is $L(\nabla\phi) = \sigma_{f_{min}}^{-1}\cdot\rho(A) = \sigma_{f_{min}}^{-1}$ with $\sigma_{f_{min}} = \min\{\sigma_{f_1}, \cdots, \sigma_{f_M}\}$. Furthermore, the condition on the step-size τ reduces to $\tau < \sigma_f/\rho(A) = \sigma_{f_{min}}$. The proof is completed. ■

Remark V.5. *All the steps in Algorithm 3 can be updated in parallel. Although Step 5 uses the global variable v, each update requires only local information, since the selection matrix E_i only selects sub-system i and its neighbours.*

Remark V.6. *For the case that all local problems are solved exactly and the consensus errors are zero, i.e. $\delta_i^k = 0$ and $\theta_i^k = 0$, Algorithm 3 converges to the optimal point at the rate $O(\frac{1}{k^2})$.*

The upper-bound in Corollary V.4 provides sufficient conditions on the error sequences $\{\delta_{\max}^k\}$ and $\{\theta_{\max}^k\}$ for the convergence of the algorithm to the optimal solution:

- The series $\{k\|\delta_{\max}^k\|\}$ and $\{k\|\theta_{\max}^k\|\}$ are summable and decrease at $O(\frac{1}{k^2})$.
- The sequences $\{\|\delta_{\max}^k\|\}$ and $\{\|\theta_{\max}^k\|\}$ decrease at the rate $O(\frac{1}{k^{2+\kappa}})$ for any $\kappa \ge 0$.

Remark V.7. *The benefit of replacing Γ^k and Λ^k with $\bar{\Gamma}^k$ and $\bar{\Lambda}^k$ is that $\bar{\Gamma}^k$ and $\bar{\Lambda}^k$ are represented by $\delta_{\max}^k$ and $\theta_{\max}^k$, which are upper-bounds for the local errors. Then the conditions on the global errors δ^k and θ^k reduce to conditions on the local computation and consensus errors.*

C. Discussion

After having shown that inexact FAMA allows for solving the local problems inexactly, the following discussion addresses the question of which methods provide the required properties of the local solutions. The local problems have the structure of standard MPC problems with strongly convex cost functions and convex constraints.

The fast gradient method, the interior-point method and active-set method are good candidates for solving the local problems, since these methods have shown good performance for solving MPC problems and ensure primal feasibility of suboptimal iterates. However, the interior-point method and active-set method generally do not provide good bounds on the number of iterations for a given level of suboptimality.

VI. Numerical Example

This section illustrates the theoretical findings of the paper and demonstrates the performance of inexact FAMA for solving a randomly generated distributed MPC problem. For

this example, we assume that the sub-systems are coupled only in the control input.

$$x_i(t+1) = A_{ii}x_j(t) + \sum_{j\in\mathcal{N}_i} B_{ij}u_j(t) \ , \ i = 1, 2, \cdots, M \ .$$

We randomly generate a connected network with 40 sub-systems. Each sub-system has 3 states and 2 inputs. The dynamical matrices A_{ii} and B_{ij} are randomly generated, i.e. generally dense, and the local systems are controllable and unstable. The input constraint $\mathbb{U}_i$ for sub-system i is set to $\mathbb{U}_i = \{u_i | -0.4 * \mathbf{1} \leq u_i(t) \leq 0.3 * \mathbf{1}\}$, where $\mathbf{1}$ denotes the all-ones vector with the same dimension as u_i. The horizon of the MPC problem is set to $N = 11$. The local cost functions are chosen as quadratic functions, i.e. $l_i(x_i(t), u_i(t)) = x_i^T(t)Qx_i(t) + u_i^T(t)Ru_i(t)$ and $l_i^f(x_i(N)) = x_i^T(N)Px_i(N)$, where Q, R and P are identity matrices. The initial states $\bar{x}_i$ are chosen, such that more than 70% of the elements of the vector $u^\star(0) = [u_1^{\star^T}(0), \cdots, u_M^{\star^T}(0)]^T$ are at the constraints.

In Fig. 1, we demonstrate the convergence properties of inexact FAMA with three different kinds of errors for δ^k and θ^k and compare inexact FAMA with exact FAMA, for which all the errors are equal to zero. Note that these errors are synthetically constructed to specify different error properties. We solve the local problems to optimality and then add the errors to it. The red line shows the performance of exact FAMA. The blue, green and pink lines show the performance of inexact FAMA. For the blue one, the errors δ_{max}^k and θ_{max}^k are set to be decreasing at $O(\frac{1}{k^2})$ rate. For the green line, the rate of decrease is set to $O(\frac{1}{k^3})$. For the pink line, the rate of decrease is set to $O(\frac{1}{k})$. We can observe that the red, blue and green lines basically overlap. Inexact FAMA with errors decreasing at the rates $O(\frac{1}{k^2})$ and $O(\frac{1}{k^3})$, which satisfy the sufficient conditions on the errors discussed in Section V-B, converges to the optimal solution as the iterations increase, and shows almost the same performance as exact FAMA. The pink line, for which the decrease rate of the error is $O(\frac{1}{k})$, which does not satisfy the sufficient conditions in Section V-B, shows a lot of oscillations and the decrease tends to be slower and eventually stalls.

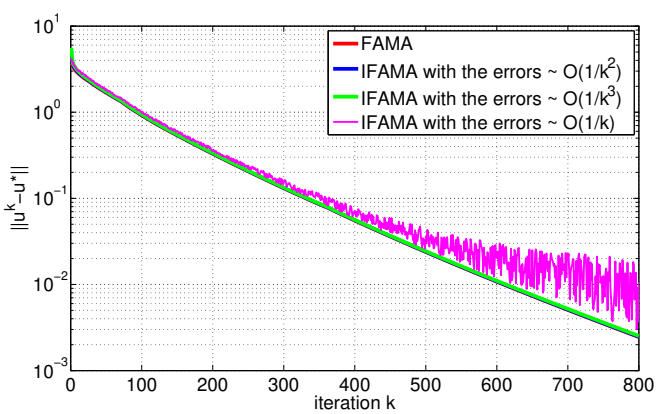

Fig. 1: Comparison of the performance of FAMA and inexact FAMA with errors decreasing at different pre-defined rates.

References

[1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, pages 183–202, 2009.

[3] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex analysis and optimization*. Athena Scientific Belmont, 2003.

[4] 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*, 3:1–122, 2011.

[5] C. Conte, T. Summers, M.N. Zeilinger, M. Morari, and C.N. Jones. Computational aspects of distributed optimization in model predictive control. In *51th IEEE Conference on Decision and Control*, pages 6819–6824, 2012.

[6] C. Conte, N. R. Voellmy, M. N. Zeilinger, M. Morari, and C. N. Jones. Distributed synthesis and control of constrained linear systems. In *American Control Conference*, pages 6017–6022, 2012.

[7] Q. T. Dinh, I. Necoara, and M. Diehl. Fast inexact decomposition algorithms for large-scale separable convex optimization. *arXiv preprint arXiv:1212,4275*, 2012.

[8] W. B. Dunbar. Distributed receding horizon control of dynamically coupled non-linear systems. *IEEE Transactions on Automatic Control*, 52(7):1249–1263, 2007.

[9] M. Farina and R. Scattolini. Distributed non-cooperative MPC with neighbor-to-neighbor communication. In *18th World Congress of the International Federation of Automatic Control*, pages 404–409, 2011.

[10] T. Goldstein, B. O'Donoghue, and S. Setzer. Fast alternating direction optimization methods. *CAM report*, pages 12–35, 2012.

[11] B. Johansson, A. Speranzon, M. Johansson, and K. H. Johansson. On decentralized negotiation of optimal consensus. *Automatica*, 44:1175–1179, 2008.

[12] T. Keviczky and K. H. Johansson. A study on distributed model predictive consensus. In *17th World Congress of the International Federation of Automatic Control*, pages 1516–1521, 2008.

[13] Y. Pu, M. N. Zeilinger, and C. N. Jones. Fast alternating minimization algorithm for model predictive control. In *19th World Congress of the International Federation of Automatic Control*, 2014 to appear.

[14] S. Richter, C. N. Jones, and M. Morari. Certification aspects of the fast gradient method for solving the dual of parametric convex programs. *Mathematical Methods of Operations Research*, 77(3):305–321, 2012.

[15] M. Schmidt, N. L. Roux, and F. Bach. Convergence rates of inexact proximal-gradient methods for convex optimization. In *25th Annual Conference on Neural Information Processing Systems*, pages 6819–6824, 2011.

[16] T. H. Summers and J. Lygeros. Distributed model predictive consensus via the alternating direction method of multipliers. In *50th Annual Allerton Conference on Communication, Control and Computing*, pages 79–84, 2012.

[17] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control and Optimization*, 29:119–138, 1991.

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