Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Complexity of an Inexact Augmented Lagrangian Method: Application to Constrained MPC

Valentin Nedelcu* Ion Necoara* Ioan Dumitrache*

* *University Politehnica Bucharest, Automatic Control and Systems Engineering Department, 060042 Bucharest, Romania (e-mail: {valentin.nedelcu, ion.necoara,ioan.dumitrache}@acse.upb.ro.)*

Abstract: We propose in this paper an inexact dual gradient algorithm based on augmented Lagrangian theory and inexact information for the values of dual function and its gradient. We study the computational complexity certification of the proposed method and we provide estimates on primal and dual suboptimality and also on primal infeasibility. We also discuss implementation aspects of the proposed algorithm on constrained model predictive control problems for embedded linear systems and provide numerical tests to certify the efficiency of the method.

1. INTRODUCTION

Embedded control systems have been widely used in many applications and their usage in industrial plants has increased concurrently. The concept behind embedded control is to design a control scheme that can be implemented on autonomous electronic hardware [6], e.g a programmable logic controller, a microcontroller circuit board or field-programmable gate arrays. One of the most successful advanced control schemes implemented in industry is model predictive control (MPC) and this is due to its ability to handle complex systems with hard input and state constraints. In the recent decades there has been a growing focus on developing faster MPC schemes, improving the computational efficiency [13] and providing worst case computational complexity certificates for the applied solution methods [7, 14], making these schemes feasible for implementation on hardware with limited computational power. Even if second order methods (e.g. interior point methods) can offer fast rates of convergence in practice, the worst case complexity bounds are high [2]. Further, these methods have complex iterations, involving inversion of matrices. Therefore, first order methods are more suitable in these situations.

When the projection on the primal feasible set is hard to compute, e.g. for constrained MPC problems, an alternative to primal gradient methods is to use the Lagrangian relaxation to handle the complicated constraints and then to apply dual gradient schemes. The computational complexity certification of gradient-based methods for solving the (augmented) Lagrangian dual of a primal convex problem is studied e.g. in [4, 7, 8, 12, 14, 15]. In [4] the authors present a general framework for gradient methods with inexact oracle, i.e. only approximate information is available for the values of the function and of its gradient, and give convergence rate analysis. The authors also apply their approach to gradient augmented Lagrangian methods and provided estimates only for dual suboptimality. In [12] a dual fast gradient method is proposed for solving quadratic programs with linear inequality constraints and estimates on primal suboptimality and infeasibility of the primal solution are provided. In [7] the authors analyze the iteration complexity of an inexact dual gradient augmented Lagrangian method. The authors provides upper bounds on the total number of iterations which have to be performed by the algorithm for obtaining a primal suboptimal solution. In [8] a dual method based on fast gradient schemes and smoothing techniques of the ordinary Lagrangian is presented. Using an averaging scheme the authors are able to recover a suboptimal solution.

Despite widespread use of the dual gradient methods for solving Lagrangian dual problems, there are some aspects of these methods that have not been fully studied. First, the focus is mainly on the convergence analysis of the dual variables. Second, there is no full convergence rate analysis (i.e. no estimates in terms of dual and primal suboptimality and primal feasibility violation) for fast dual gradient methods while using inexact dual information. Therefore, in this paper we focus on solving convex optimization problems (possibly nonsmooth) approximately by using an augmented Lagrangian approach and inexact dual fast gradient methods. We show how approximate primal solutions can be generated based on averaging for general convex problems and we give a full convergence rate analysis for the proposed method.

Contribution: We propose and analyze an inexact dual fast gradient algorithm producing approximate primal feasible and optimal solutions. For solving the outer (dual) problem we propose an inexact dual fast gradient algorithm with complexity $\mathcal{O}(\sqrt{1/\varepsilon_{\text{out}}})$ iterations, provided that the inner problems are solved with accuracy ε_{in} of order $\mathcal{O}(\varepsilon_{\text{out}}\sqrt{\varepsilon_{\text{out}}})$. For the proposed algorithm we provide estimates on primal suboptimality and infeasibility.

Paper outline: The paper is organized as follows. In Section 1, motivated by embedded MPC, we introduce the augmented Lagrangian framework for solving constrained convex problems and discuss the complexity of solving the

Copyright © 2014 IFAC

inner problems. In Section 2 we propose an inexact dual fast gradient algorithm for solving the outer problem and we provide estimates on the dual and primal suboptimality and also on the primal infeasibility. In Section 3 we discuss different implementation aspects of the proposed algorithm in the context of constrained linear MPC. We also provide numerical tests to prove the efficiency of the proposed algorithm.

Notation and terminology We work in the space $\mathbb{R}^n$ composed by column vectors. For $x, y \in \mathbb{R}^n$, $\langle x, y\rangle := x^T y = \sum_{i=1}^n x_i y_i$ and $\|x\| := (\sum_{i=1}^n x_i^2)^{1/2}$ denote the standard Euclidean inner product and norm, respectively. We also denote by $R_\mathrm{p} := \max_{z,y\in Z} \|z - y\|$ the diameter of a convex, compact set Z. For a real number x, $\lfloor x \rfloor$ denotes the largest integer number which is less than or equal to x.

1.1 Motivation: Constrained linear MPC

We consider a discrete time linear system given by:

$$x_{k+1} = A_x x_k + B_u u_k,$$

where $x_k \in \mathbb{R}^{n_x}$ represents the state and $u_k \in \mathbb{R}^{n_u}$ the input of the system. We also assume hard state and input constraints:

$$x_k \in X \subseteq \mathbb{R}^{n_x}, \quad u_k \in U \subseteq \mathbb{R}^{n_u} \ \ \forall k \geq 0.$$

We can define the linear MPC problem over the prediction horizon of length N, for an initial state x, as follows [16]:

$$f^*(x) := \begin{cases} \min\limits_{x_i, u_i} \ \sum\limits_{i=0}^{N-1} \ell(x_i, u_i) + \ell_\mathrm{f}(x_N) \\ \text{s.t.} \ \ x_{i+1} = A_x x_i + B_u u_i, \ x_0 = x, \\ \qquad x_i \in X, \ u_i \in U \ \forall i, \ x_N \in X_\mathrm{f}, \end{cases} \tag{1}$$

where both the stage cost ℓ and the terminal cost ℓ_f are convex functions (possibly nonsmooth). Note that in our formulation we do not require strongly convex costs. Further, the terminal set X_f is chosen so that stability of the closed-loop system is guaranteed. We assume the sets X, U and X_f to be compact, convex and simple (by simple we understand that the projection on these sets can be done easily, e.g. boxes).

Furthermore, we introduce the following notations: $z := \left[x_1^T \cdots x_N^T \ u_0^T \cdots u_{N-1}^T\right]^T$, $Z := \prod_{i=1}^{N-1} X \times X_\mathrm{f} \times \prod_{i=1}^{N} U$ and $f(z) := \sum_{i=0}^{N-1} \ell(x_i, u_i) + \ell_\mathrm{f}(x_N)$. We can also write compactly the linear dynamics $x_{i+1} = A_x x_i + B_u u_i$ for all $i = 0, \cdots, N-1$ and $x_0 = x$ as $Az = b(x)$ (see [16] for details). Note that $b(x) \in \mathbb{R}^{Nn_x}$ depends linearly on x, i.e. $b(x) := \left[(A_x x)^T \ 0^T \cdots 0^T\right]^T$. In these settings, for linear MPC we need to solve, for a given initial state x, the primal convex optimization problem:

$$\min_z \left\{ f(z) \mid Az = b(x), \ z \in Z \right\}, \tag{$\mathbf{P}(x)$}$$

where f is a convex function (possibly nonsmooth) and A is a matrix of appropriate dimension. Moreover, the set Z is simple as long as X, X_f and U are simple sets. In the following sections we discuss how we can efficiently solve optimization problem ($\mathbf{P}(x)$) approximately with a dual fast gradient method based on inexact information.

1.2 Augmented Lagrangian framework

Motivated by MPC problems, we are interested in solving convex optimization problems of the form:

$$f^* := \begin{cases} \min\limits_{z\in\mathbb{R}^n} \ f(z) \\ \text{s.t.} \ \ Az = b, \ z \in Z, \end{cases} \tag{$\mathbf{P}$}$$

where f is convex function (possibly nonsmooth), $A \in \mathbb{R}^{m\times n}$ is a full row-rank matrix and Z is a simple set (i.e. the projection on this set is computationally cheap), compact and convex. We will denote problem ($\mathbf{P}$) as the primal problem and f as the primal objective function.

As we have already mentioned, in contrast with second order methods, for first order methods the number of iterations predicted by the worst case complexity analysis is close to the actual number of iterations performed by the method [10], which is crucial in the context of fast embedded systems. First order methods applied directly to problem ($\mathbf{P}$) imply projection on the feasible set $\{z \mid z \in Z, \ Az = b\}$ which is very hard to compute due to the complicating constraints $Az = b$. An efficient alternative is to move the complicating constraints into the cost via Lagrange multipliers and solve the dual problem approximately by using first order methods and then recover a primal suboptimal solution for ($\mathbf{P}$). This is the approach that we follow in this paper.

First let us define the dual function:

$$d(\lambda) := \min_{z\in Z} \mathcal{L}(z, \lambda), \tag{2}$$

where $\mathcal{L}(z,\lambda) := f(z) + \langle \lambda, Az - b\rangle$ represents the partial Lagrangian with respect to the constraints $Az = b$ and λ the associated Lagrange multipliers. Now, we can write the corresponding dual problem as follows:

$$\max_{\lambda\in\mathbb{R}^m} d(\lambda). \tag{$\mathbf{D}$}$$

We assume that Slater's constraint qualification holds, so that problems ($\mathbf{P}$) and ($\mathbf{D}$) have the same optimal value. We denote by z^* an optimal solution of ($\mathbf{P}$) and by λ^* the corresponding multiplier (i.e. an optimal solution of ($\mathbf{D}$)).

In general, the dual function d is not differentiable [1] and therefore any subgradient method for solving ($\mathbf{D}$) has a slow convergence rate. We will see in the sequel how we can avoid this drawback by means of the augmented Lagrangian framework. We define the augmented Lagrangian function [15]:

$$\mathcal{L}_\rho(z,\lambda) := f(z) + \langle \lambda, Az - b\rangle + \frac{\rho}{2}\|Az - b\|^2, \tag{3}$$

where $\rho > 0$ represents a penalty parameter. The augmented dual problem, called also the *outer* problem, is:

$$\max_{\lambda\in\mathbb{R}^m} d_\rho(\lambda), \tag{$\mathbf{D}_\rho$}$$

where $d_\rho(\lambda) := \min\limits_{z\in Z} \mathcal{L}_\rho(z,\lambda)$ and we denote by $z^*(\lambda)$ an optimal solution of the *inner* problem $\min_{z\in Z} \mathcal{L}_\rho(z,\lambda)$ for a given λ. It is well-known [1, 7] that the optimal value and the set of optimal solutions of the dual problems ($\mathbf{D}$) and ($\mathbf{D}_\rho$) coincide. Furthermore, the function d_ρ is concave and differentiable and its gradient is given by [11]:

$$\nabla d_\rho(\lambda) := Az^*(\lambda) - b.$$

Moreover, the gradient mapping $\nabla d_\rho(\cdot)$ is Lipschitz continuous with a Lipschitz constant [1] $L_\mathrm{d} := \rho^{-1}$.

In conclusion, we want to solve within an accuracy ε_out the equivalent smooth outer problem ($\mathbf{D}_\rho$) by using first order methods with inexact gradients (e.g. fast gradient algorithms) and then recover an approximate primal solution. In other words, the goal of this paper is to generate a primal-dual pair $(\hat{z}, \hat{\lambda})$, with $\hat{z} \in Z$, for which we can ensure bounds on dual suboptimality, primal infeasibility and primal suboptimality of order ε_{out}, i.e.:

$$f^* - d_\rho(\hat{\lambda}) \leq \mathcal{O}(\varepsilon_{\text{out}}), \ \|A\hat{z} - b\| \leq \mathcal{O}(\varepsilon_{\text{out}}) \ \text{and} \\ |f(\hat{z}) - f^*| \leq \mathcal{O}(\varepsilon_{\text{out}}). \quad (4)$$

We will discuss in the following sections how we can ensure conditions (4).

1.3 Complexity estimates of solving the inner problems

As we have seen in the previous section, in order to compute the gradient ∇d_ρ we have to find, for a given λ, an optimal solution of the *inner* convex problem:

$$z^*(\lambda) \in \arg\min_{z \in Z} \mathcal{L}_\rho(z, \lambda). \quad (5)$$

Since an exact minimizer of the inner problem (5) is usually hard to compute, we are interested in finding an approximate solution of this problem instead of its optimal one. Therefore, we have to consider an inner accuracy ε_{in} which measures the suboptimality of such an approximate solution for (5):

$$\bar{z}(\lambda) \approx \arg\min_{z \in Z} \left\{ f(z) + \langle \lambda, Az - b \rangle + \frac{\rho}{2}\|Az - b\|^2 \right\}.$$

Since there exist several ways to characterize an ε_{in}-optimal solution [4, 7, 9], we will further discuss different stopping criteria which can be used in order to find such a solution. A well-known stopping criterion, which measures the distance to optimal value of (5), is given by:

$$\bar{z}(\lambda) \in Z, \ \ \mathcal{L}_\rho(\bar{z}(\lambda), \lambda) - \mathcal{L}_\rho(z^*(\lambda), \lambda) \leq \varepsilon_{\text{in}}^2. \quad (6)$$

As a direct consequence of the optimality condition for problem (5), one can use the following stopping criterion:

$$\bar{z}(\lambda) \in Z, \ \ \langle \nabla \mathcal{L}_\rho(\bar{z}(\lambda), \lambda), z - \bar{z}(\lambda) \rangle \geq -\mathcal{O}(\varepsilon_{\text{in}}) \ \forall z \in Z. \quad (7)$$

Note that if $\nabla \mathcal{L}_\rho$ is Lipschitz continous with constant L_{p} then (6) implies (7) with $\mathcal{O}(\varepsilon_{\text{in}}) = \left(1 + \sqrt{L_{\text{p}}} R_{\text{p}}\right) \varepsilon_{\text{in}}$. A detailed discussion regarding the relation between criterions (6) and (7) can be found in [4, 9].

The next theorem provides estimates on the number of iterations that are required by fast gradient schemes to obtain an ε_{in} approximate solution for inner problem (5).

Theorem 1.1. [10] Assume that the function $\mathcal{L}_\rho(\cdot, \lambda)$ has Lipschitz continuous gradient w.r.t. variable z, with a Lipschitz constant L_{p} and a fast gradient scheme [10] is applied for finding an ε_{in} approximate solution $\bar{z}(\lambda)$ of (5) such that stopping criterion (6) holds, i.e. $\mathcal{L}_\rho(\bar{z}(\lambda), \lambda) - \mathcal{L}_\rho(z^*(\lambda), \lambda) \leq \varepsilon_{\text{in}}^2$. Then, the complexity of finding $\bar{z}(\lambda)$ is $\mathcal{O}\left(\sqrt{L_p/\varepsilon_{\text{in}}^2}\right)$ iterations. If, in addition $\mathcal{L}_\rho(\cdot, \cdot)$ is strongly convex with a convexity parameter $\sigma_{\text{p}} > 0$, then $\bar{z}(\lambda)$ can be computed in at most $\mathcal{O}\left(\sqrt{L_{\text{p}}/\sigma_{\text{p}}} \ln\left(\sigma_{\text{p}}/\varepsilon_{\text{in}}^2\right)\right)$ iterations by using a fast gradient scheme.

2. COMPLEXITY ESTIMATES OF SOLVING THE OUTER PROBLEM USING APPROXIMATE DUAL GRADIENTS

In this section we solve the augmented Lagrangian dual problem $(\mathbf{D}_\rho)$ approximately by using a dual fast gradient method and derive computational complexity certificates for this methods. Since we solve the inner problem inexactly, we have to use inexact gradients and approximate values of the augmented dual function d_ρ defined in terms of $\bar{z}(\lambda)$, i.e. we introduce the following pair:

$$\bar{d}_\rho(\lambda) := \mathcal{L}_\rho(\bar{z}(\lambda), \lambda) \quad \text{and} \quad \nabla \bar{d}_\rho(\lambda) := A\bar{z}(\lambda) - b.$$

The next theorem, whose proof can be found in [4, 9], provides bounds on the dual function when the inner problem (5) is solved approximately.

Theorem 2.1. [4, 9] If $\bar{z}(\lambda)$ is computed such that the stopping criterion (7) is satisfied, i.e. $\bar{z}(\lambda) \in Z$ and $\min_{z \in Z} \langle \nabla \mathcal{L}_\rho(\bar{z}(\lambda), \lambda), z - \bar{z}(\lambda) \rangle \geq -\left(1 + \sqrt{L_{\text{p}}} R_{\text{p}}\right) \varepsilon_{\text{in}}$, then the following inequalities hold:

$$\bar{d}_\rho(\lambda) + \langle \nabla \bar{d}_\rho(\lambda), \mu - \lambda \rangle - \frac{L_{\text{d}}}{2}\|\mu - \lambda\|^2 - \left(1 + \sqrt{L_{\text{p}}} R_{\text{p}}\right) \varepsilon_{\text{in}} \\ \leq d_\rho(\mu) \leq \bar{d}_\rho(\lambda) + \langle \nabla \bar{d}_\rho(\lambda), \mu - \lambda \rangle \quad \forall \lambda, \mu \in \mathbb{R}^m. \quad (8)$$

Note that the previous theorem helps us to construct a model which bounds the function d_ρ, when the exact values of the dual function and its gradients are unknown.

2.1 Inexact dual fast gradient method

In this subsection we discuss a fast gradient scheme for solving the augmented Lagrangian dual problem $(\mathbf{D}_\rho)$. Fast gradient schemes were first proposed by Nesterov [11] and have also been discussed in the context of dual decomposition in [8]. A modification of these schemes for the case of inexact information can be also found in [4]. We shortly present such a scheme as follows. Given a positive sequence $\{\theta_k\}_{k \geq 0} \subset (0, +\infty)$ with $\theta_0 = 1$, we define $S_k := \sum_{j=0}^{k} \theta_j$. Let us assume that the sequence $\{\theta_k\}_{k \geq 0}$ satisfies $\theta_{k+1}^2 = S_{k+1}$ for all $k \geq 0$. This condition leads to:

$$\theta_{k+1} := \frac{1}{2}(1 + \sqrt{4\theta_k^2 + 1}) \ \ \forall k \geq 0 \ \text{ and } \theta_0 := 1. \quad (9)$$

Note that the sequence $\{\theta_k\}_{k \geq 0}$ generated by (9) satisfies:

$$0.5(k+1) \leq \theta_k \leq k+1 \ \ \forall k \geq 0. \quad (10)$$

We can also obtain $0.25(k+1)(k+2) < S_k < 0.5(k+1)(k+2)$ and $\sum_{j=0}^{k} S_j < (k+1)(k+2)(k+3)/6$. Given an initial point $\lambda_0 \in \mathbb{R}^m$, we define two sequences of the dual variables $\{\lambda_k\}_{k \geq 0}$ and $\{\mu_k\}_{k \geq 0}$ as:

$$(\mathbf{IDFGM}): \begin{cases} \mu_k & := \lambda_k + L_{\text{d}}^{-1} \nabla \bar{d}_\rho(\lambda_k) \\ \lambda_{k+1} & := (1 - a_{k+1})\,\mu_k \\ & \quad + a_{k+1}\left[\lambda_0 + L_{\text{d}}^{-1} \sum_{i=0}^{k} \theta_i \nabla \bar{d}_\rho(\lambda_i)\right], \end{cases}$$

where the sequence $a_{k+1} := S_{k+1}^{-1} \theta_{k+1}$.

The following lemma, which represents an extension of the results in [8, 11] to the inexact case (see also [4]), will be useful for the analysis of our proposed method.

Lemma 2.2. [4, 8] Under the assumptions of Theorem 2.1, the two sequences $\{(\lambda_k, \mu_k)\}_{k \geq 0}$ generated by the dual fast gradient scheme (**IDFGM**) satisfy:

$$S_k d_\rho(\mu_k) \geq \max_{\lambda\in\mathbb{R}^m} \sum_{j=0}^{k} \theta_j \left[\bar{d}_\rho(\lambda_j) + \left\langle \nabla \bar{d}_\rho(\lambda_j), \lambda - \lambda_j \right\rangle\right] \quad (11)$$
$$- \frac{L_{\rm d}}{2}\|\lambda-\lambda_0\|^2 - \left(1+\sqrt{L_{\rm p}}R_{\rm p}\right)\varepsilon_{\rm in}\sum_{j=0}^{k} S_j \;\; \forall k \geq 0.$$

The next theorem gives an estimate on dual suboptimality.

Theorem 2.3. Under the assumptions of Theorem 2.1, let $\{(\lambda_k, \mu_k)\}_{k\geq 0}$ be the two sequences generated by the scheme (**IDFGM**) and $R_{\rm d} := \|\lambda_0 - \lambda^*\|$. Then, an estimate on dual suboptimality is given by the following expression:

$$f^* - d_\rho(\mu_k) \leq \frac{2L_{\rm d}R_{\rm d}^2}{(k+1)(k+2)} + \frac{2(k+3)}{3}\left(1+\sqrt{L_{\rm p}}R_{\rm p}\right)\varepsilon_{\rm in}.$$

Proof For simplicity, we introduce the notation $C_Z := 1+\sqrt{L_{\rm p}}R_{\rm p}$. Using inequality (8) in (11) we obtain:

$$S_k d_\rho(\mu_k) \geq S_k d_\rho(\lambda^*) - \frac{L_{\rm d}}{2}\|\lambda^* - \lambda_0\|^2 - C_Z\varepsilon_{\rm in}\sum_{j=0}^{k} S_j.$$

Now, using the fact that $S_k > 0.25(k+1)(k+2)$ and $\sum_{j=0}^{k} S_j < (k+1)(k+2)(k+3)/6$ and the definition of C_Z, we obtain our result. □

We further define the following primal average sequence:

$$\hat{z}_k := S_k^{-1}\sum_{j=0}^{k}\theta_j \bar{z}_i, \quad (12)$$

where $\bar{z}_i := \bar{z}(\lambda_i)$. The next theorem gives an estimate on infeasibility of $\hat{z}_k$ for the original problem (**P**).

Theorem 2.4. Under the conditions of Theorem 2.3, the point $\hat{z}_k$ defined by (12) satisfies the following estimate on primal feasibility violation:

$$\|A\hat{z}_k - b\| \leq v(k, \varepsilon_{\rm in}), \quad (13)$$

where $v(k,\varepsilon_{\rm in}) := \frac{8L_{\rm d}R_{\rm d}}{(k+1)(k+2)} + 4\sqrt{\frac{L_{\rm d}(k+3)\left(1+\sqrt{L_{\rm p}}R_{\rm p}\right)\varepsilon_{\rm in}}{3(k+1)(k+2)}}$.

Proof By the definition of $\bar{d}_\rho$, $\nabla\bar{d}_\rho$ and $\hat{z}_k$ we have:

$$\sum_{j=0}^{k}\theta_j\left[\bar{d}_\rho(\lambda_j) + \left\langle\nabla\bar{d}_\rho(\lambda_j), \lambda-\lambda_j\right\rangle\right]$$
$$= \sum_{j=0}^{k}\theta_j f(\bar{z}_j) + S_k\left\langle\lambda, A\hat{z}_k - b\right\rangle + \sum_{j=0}^{k}\theta_j\frac{\rho}{2}\|A\bar{z}_j - b\|^2$$
$$\geq S_k f(\hat{z}_k) + S_k\left\langle\lambda, A\hat{z}_k - b)\right\rangle + \frac{S_k}{2L_{\rm d}}\|A\hat{z}_k - b\|^2,$$

where recall that $S_k = \sum_j \theta_j$ and thus the last relation follows from Jensen's inequality applied to f and $\|\cdot\|^2$.

Substituting the previous inequality into (11) we obtain:

$$d_\rho(\mu_k) \geq f(\hat{z}_k) + \max_{\lambda\in\mathbb{R}^m}\left\{\langle\lambda, A\hat{z}_k - b\rangle - \frac{L_{\rm d}}{2S_k}\|\lambda-\lambda_0\|^2\right\}$$
$$+ \frac{\rho}{2}\|A\hat{z}_k - b\|^2 - C_Z\varepsilon_{\rm in}S_k^{-1}\sum_{j=0}^{k}S_j. \quad (14)$$

On the one hand, we can write:

$$d_\rho(\mu_k) - f(\hat{z}_k) - \frac{\rho}{2}\|A\hat{z}_k - b\|^2$$
$$\leq d_\rho(\lambda^*) - f(\hat{z}_k) - \frac{\rho}{2}\|A\hat{z}_k - b\|^2 \quad (15)$$
$$= \min_{z\in Z}\mathcal{L}_\rho(z,\lambda^*) - f(\hat{z}_k) - \frac{\rho}{2}\|A\hat{z}_k - b\|^2 \leq \langle\lambda^*, A\hat{z}_k - b\rangle.$$

On the other hand, we have:

$$\max_{\lambda\in\mathbb{R}^m}\left\{-\frac{L_{\rm d}}{2S_k}\|\lambda-\lambda_0\|^2 + \langle\lambda, A\hat{z}_k - b\rangle\right\}$$
$$= \frac{S_k}{2L_{\rm d}}\|A\hat{z}_k - b\|^2 + \langle\lambda_0, A\hat{z}_k - b\rangle. \quad (16)$$

Substituting (15) and (16) into (14) we obtain:

$$\frac{S_k}{2L_{\rm d}}\|A\hat{z}_k - b\|^2 + \langle\lambda_0 - \lambda^*, A\hat{z}_k - b\rangle \leq C_Z\varepsilon_{\rm in}S_k^{-1}\sum_{j=0}^{k}S_j.$$

If we define $\xi := \|A\hat{z}_k - b\|$, then the last inequality implies $\frac{(k+1)(k+2)}{8L_{\rm d}}\xi^2 - R_{\rm d}\xi \leq \frac{4(k+3)}{3}C_Z\varepsilon_{\rm in}$. Therefore, using $\sqrt{\zeta_1+\zeta_2} \leq \sqrt{\zeta_1} + \sqrt{\zeta_2}$ we obtain $\xi \leq \nu(k,\varepsilon_{\rm in})$, where $\nu(\cdot,\cdot)$ is defined in (13). □

Finally, we characterize the primal suboptimality for optimization problem (**P**).

Theorem 2.5. Under the conditions of Theorem 2.4, the following estimates hold on primal suboptimality:

$$-\left[\|\lambda^*\| + \frac{\rho}{2}\nu(k,\varepsilon_{\rm in})\right]\nu(k,\varepsilon_{\rm in}) \leq f(\hat{z}_k) - f^*$$
$$\leq \frac{2L_{\rm d}\|\lambda_0\|^2}{(k+1)(k+2)} + \frac{2(k+3)}{3}\left(1+\sqrt{L_{\rm p}}R_{\rm p}\right)\varepsilon_{\rm in}.$$

Proof The left-hand side inequality can be easily proved using the definition of $\mathcal{L}_\rho(\hat{z}_k, \lambda^*)$, Theorem 2.4, the Cauchy-Schwartz inequality and taking into account that $f^* = d_\rho(\lambda^*) \leq \mathcal{L}_\rho(\hat{z}_k,\lambda^*)$.

For the right-hand side we use first (14) and (16):

$$d_\rho(\mu_k) \geq f(\hat{z}_k) + \frac{S_k}{2L_{\rm d}}\|A\hat{z}_k - b\|^2 + \langle\lambda_0, A\hat{z}_k - b\rangle$$
$$+ \frac{\rho}{2}\|A\hat{z}_k - b\|^2 - C_Z\varepsilon_{\rm in}S_k^{-1}\sum_{j=0}^{k}S_j$$
$$\geq f(\hat{z}_k) - \frac{2L_{\rm d}}{(k+1)(k+2)}\|\lambda_0\|^2 - \frac{2(k+3)}{3}C_Z\varepsilon_{\rm in}.$$

Therefore, we get:

$$f(\hat{z}_k) - d_\rho(\mu_k) \leq \frac{2L_{\rm d}}{(k+1)(k+2)}\|\lambda_0\|^2 + \frac{2(k+3)}{3}C_Z\varepsilon_{\rm in}.$$

Since $d_\rho(\mu_k) \leq f^*$, we obtain the result. □

We assume now that we fix the outer accuracy $\varepsilon_{\rm out}$ to a desired value and the goal is to find $k_{\rm out}$ and a relation between $\varepsilon_{\rm out}$ and $\varepsilon_{\rm in}$ such that after $k_{\rm out}$ outer iterations of the scheme (**IDFGM**) relations (4) holds. We can take e.g.:

$$k_{\rm out} = \left\lfloor 2R_{\rm d}\sqrt{\frac{L_{\rm d}}{\varepsilon_{\rm out}}}\right\rfloor \text{ and } \varepsilon_{\rm in} = \frac{3}{4\left(1+\sqrt{L_{\rm p}}R_{\rm p}\right)(k_{\rm out}+3)}\varepsilon_{\rm out}.$$

Using now Theorems 2.3, 2.4 and 2.5 we can conclude that the following bounds for dual suboptimality, primal infeasibility and primal suboptimality hold:

$f^* - d_\rho(\hat{\lambda}_{k_{\rm out}}) \leq \varepsilon_{\rm out},\ \hat{z}_{k_{\rm out}} \in Z,\ \|A\hat{z}_{k_{\rm out}} - b\| \leq \frac{3}{R_{\rm d}}\varepsilon_{\rm out}$ and

$$-\left(\frac{3\|\lambda^*\|}{R_{\rm d}} + \frac{9\rho}{2R_{\rm d}^2}\varepsilon_{\rm out}\right)\varepsilon_{\rm out} \leq f(\hat{z}_{k_{\rm out}}) - f^* \leq \left(\frac{\|\lambda_0\|^2 + R_{\rm d}^2}{2R_{\rm d}^2}\right)\varepsilon_{\rm out}.$$

3. COMPLEXITY CERTIFICATION FOR LINEAR MPC PROBLEMS

In this section we discuss different implementation aspects and total complexity estimates for the application of the newly developed algorithm in the context of state-input constrained MPC for fast linear embedded systems. We denote by X_N a subset of the region of attraction [16] for the MPC scheme discussed in Section 1.1. A detailed discussion on the stability of suboptimal MPC schemes can be found e.g. in [16]. For a given $x \in X_N$, we denote with $z^*(x)$ an optimal solution for $(\mathbf{P}(x))$ and with $\lambda^*(x)$ an associated optimal multiplier. Usually, in MPC problems the stage and final costs are quadratic functions:

$$\ell(x_i, u_i) := x_i^T Q x_i + u_i^T R u_i \quad \text{and} \quad \ell_{\rm f}(x_N) := x_N^T P x_N,$$

where matrices Q and P are positive semidefinite and R is positive definite. Thus, f becomes quadratic with Hessian H. Further, we characterize the convexity properties of the augmented Lagrangian function.

Lemma 3.1. [9] If the optimization problem $(\mathbf{P}(x))$ comes from a linear MPC problem with quadratic stage and final costs, then the augmented Lagrangian $\mathcal{L}_\rho(z, \lambda, x)$ is a strongly convex quadratic function w.r.t. variable z.

The previous lemma shows that in the linear MPC case with quadratic costs the objective function of the inner subproblems $\mathcal{L}_\rho$ are quadratic strongly convex in the first variable z. Moreover, $\mathcal{L}_\rho$ has also Lipschitz continuous gradient. Note that the convexity parameter $\sigma_{\rm p}$ of this function can be computed easily: $\sigma_{\rm p} := \lambda_{\min}(H + \rho A^T A)$. Also, the Lipschitz constant $L_{\rm p}$ of the gradient of $\mathcal{L}_\rho$ is given by: $L_{\rm p} := \lambda_{\max}(H + \rho A^T A)$.

Note that since $\mathcal{L}_\rho(z, \lambda, x)$ is strongly convex and with Lipschitz continuous gradient in the variable z, by solving the inner problem (5) with a fast gradient scheme we can ensure stopping criterion (6) in a linear number of inner iterations (see Theorem 1.1). Note that the estimate for the number of inner iterations is also dependent on the diameter $R_{\rm p}$ of the set Z. We can see immediately that this diameter can be computed easily for cases when the set Z has a specific structure. Note that the set Z is a Cartesian product and thus we have: $R_{\rm p} := \sqrt{(N-1)D_x^2 + D_{x_{\rm f}}^2 + N D_u^2}$, where D_x, $D_{x_{\rm f}}$ and D_u denotes the diameters of X, $X_{\rm f}$ and U, respectively. These diameters can be computed explicitly for constraints sets defined e.g. by boxes or Euclidean balls, which typically appear in the context of MPC problems.

Since the estimate for the number of outer iterations depends on the norm of the dual optimal solution $\|\lambda^*\|$ we make use of the result from [3]:

Lemma 3.2. [3] For the MPC problems $(\mathbf{P}(x))_{x \in X_N}$ we assume that there exists $r > 0$ such that $B(0, r) \subseteq \{Az - b(x) \mid z \in Z, x \in X_N\}$, where $B(0, r)$ denotes the Euclidean ball in $\mathbb{R}^{N(n_x + n_u)}$ with center 0 and radius r. Then, the following upper bound on the norm of the dual optimal solutions of MPC problems $(\mathbf{P}(x))$ holds:

$$\|\lambda^*(x)\| \leq \frac{\max_{z \in Z} \langle Hz^*(x), z - z^*(x)\rangle}{\bar{r}} \quad \forall x \in X_N,$$

where $\bar{r} := \max\left\{r \mid B(0, r) \subseteq \{AZ - b(x) \mid x \in X_N\}\right\}$.

Based on the previous lemma, in [14] upper bounds are derived on $\|\lambda^*(x)\|$ for all $x \in X_N$ for linear MPC problems with X, $X_{\rm f}$, U and X_N polyhedral sets:

$$\mathcal{R}_{\rm d} \geq \max_{x \in X_N} \|\lambda^*(x)\|. \tag{17}$$

We assume now that we fix the outer accuracy $\varepsilon_{\rm out}$ to a desired value and we want to estimate the total number of iterations and also the number of flops per inner and outer iterations which have to be performed by Algorithm (**IDFGM**) in order to solve the MPC problem $(\mathbf{P}(x))$. For simplicity of the exposition we assume that the initialization $\lambda_0 = 0$ and the inner problems are solved using the stopping criterion (6). According to Section 2.1, for the outer iterations we have:

$$k_{\rm out}^{FG} := \left\lfloor 2\mathcal{R}_{\rm d}\sqrt{\frac{L_{\rm d}}{\varepsilon_{\rm out}}}\right\rfloor. \tag{18}$$

Since we have proved that in the MPC case $\mathcal{L}_\rho(\cdot, \lambda, x)$ is strongly convex with convexity parameter $\sigma_{\rm p}$ and has also Lipschitz continuous gradient with constant $L_{\rm p}$, in order to find a point $\bar{z}_{k_{\rm in}^{FG}}(\lambda)$ such that stoping criterion (6) $\mathcal{L}_\rho(\bar{z}_{k_{\rm in}^{FG}}(\lambda), \lambda, x) - \mathcal{L}_\rho(z^*(\lambda), \lambda, x) \leq \varepsilon_{\rm in}^2$ we can apply a fast gradient scheme. Using Theorem 2.2.3 in [10] and taking the inner accuracy as in Section 2.1, the number of inner iterations will be given by:

$$k_{\rm in}^{FG} = \left\lfloor 2\sqrt{\frac{L_{\rm p}}{\sigma_{\rm p}}}\ln\left(\frac{5\sqrt{L_{\rm d}}\mathcal{R}_d\sqrt{L_{\rm p}}R_{\rm p}\left(1+\sqrt{L_{\rm p}}R_{\rm p}\right)}{\varepsilon_{\rm out}\sqrt{\varepsilon_{\rm out}}}\right)\right\rfloor.$$

For solving the inner problem we use a simple fast gradient scheme for smooth strongly convex objective functions, see e.g. [10]. For this scheme, an inner iteration will require $n_{\rm in}^{\rm flops} = N\left(3n_x^2 + 2n_x n_u + 2n_u^2 + 10n_x + 8n_u\right)$ flops. Regarding the number of flops required by an outer iteration, the following value can be established: $n_{\rm out}^{\rm flops,FG} = N\left(2n_x^2 + 2n_x n_u + 10n_x\right) + k_{\rm in}^{FG} n_{\rm in}^{\rm flops}$.

4. NUMERICAL EXPERIMENTS

In this section we apply Algorithm (**IDFGM**) to an MPC problem for the ball on plate system described in [14]. We consider box constraints for states X and $X_{\rm f}$, inputs U and for the region of attraction X_N as in [14], while for the stage costs we take the matrices $Q = q_1 q_1^T$, where $q_1 = [2\ 1]^T$, $R = 1$ and we compute the terminal matrix P as the solution of the LQR problem.

For different prediction horizons ranging from $N = 5$ to $N = 20$, we first analyze the behavior of Algorithm (**IDFGM**) in terms of the number of outer iterations. For each prediction horizon length, we consider two different estimates for the number of outer iterations depending on the way we compute the upper bound on the optimal Lagrange multipliers $\lambda^*(x)$: $k_{\rm out}^{FG}$ is the theoretical number of iterations obtained using relation (18) with $\mathcal{R}_{\rm d}$ computed according to [14] (see (17)), while $k_{\rm out,samp}^{FG}$ is the average number of iterations obtained using relation (18) with $R_{\rm d}$ computed exactly using `Gurobi 5.0.1` solver, iterations which correspond to 50 random initial states $x \in X_N$.

We also compute the average number of outer iterations $k^{FG}_{\text{out,real}}$ observed in practice. In all simulations we take $\rho = 1$. The results are reported in Figure 1 (left). We can notice that $k^{FG}_{\text{out,samp}}$ obtained from our derived bound in Section 2 offer a good approximation for the real number of iterations performed by the algorithm.

Since the number of outer iterations is also dependent on the way the inner accuracy ε_{in} is chosen we also solve the optimization problem ($\mathbf{P}(x)$) with a prediction horizon $N = 20$, a fixed outer accuracy $\varepsilon_{\text{out}} = 10^{-3}$ and varying ε_{in}. In Figure 1 (right) we plot the average number of outer iterations performed by the algorithms by taking 10 random samples for the initial state $x \in X_N$. We observe that we can increase the inner accuracy ε_{in} up to a certain value and the algorithm still performs a number of iterations less than the theoretical bounds derived in Section 2.

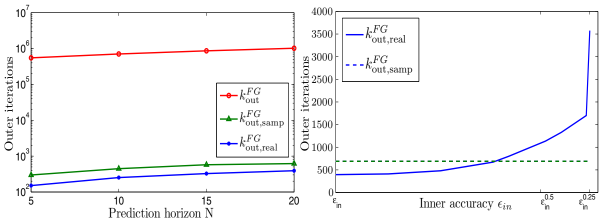

Fig. 1. Variation of k^{FG}_{out}, $k^{FG}_{\text{out,samp}}$ and $k^{FG}_{\text{out,real}}$ w.r.t the prediction horizon N (left). The number of outer iterations performed by Algorithm (**IDFGM**) with fixed $\varepsilon_{\text{out}} = 10^{-3}$ and varying ε_{in} (right).

In Figure 2 we also plot the evolution of the states and inputs for a prediction horizon $N = 5$ and an outer accuracy $\varepsilon_{\text{out}} = 10^{-3}$. We observe that the system is driven to the equilibrium point.

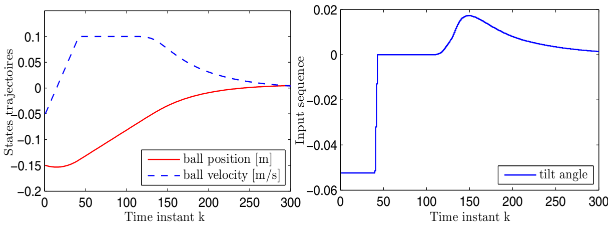

Fig. 2. The trajectories of the states and inputs for a prediction horizon $N = 5$ obtained using Algorithm (**IDFGM**) with accuracy $\varepsilon_{\text{out}} = 10^{-3}$.

5. CONCLUSION

Based on the augmented Lagrangian framework, we have proposed an inexact dual fast gradient method for solving convex optimization problems with complicating linear constraints. We have solved the dual problem using a fast gradient method with inexact information. We have solved the inner subproblems only up to a certain accuracy and derived tight estimates on primal and dual suboptimality and also on primal infeasibility. We have discussed some implementation aspects of the new algorithm for embedded linear MPC problems and tested it on a ball on plate system. Also, an interesting issue which we intend to approach in our future work is a comparison some intelligent control strategies (see e.g. [5]).

REFERENCES

[1] D.P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, 1999.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[3] O. Devolder, F. Glineur and Y. Nesterov, *Double Smoothing Technique for Large Scale Linearly Constrained Convex Optimization*, SIAM J. Optim., 22 (2012), pp. 702–727.

[4] O. Devolder, F. Glineur and Y. Nesterov, *First-order methods of smooth convex optimization with inexact oracle*, CORE Discussion Papers 02/2011, Universite Catholique de Louvain, 2011.

[5] I. Dumitrache, *From model-based strategies to intelligent control systems*, Proc. of the 9th WSEAS International Conference on Automation and Information, pp. 408–415.

[6] J.L. Jerez, K.-V. Ling, G.A. Constantinides and E.C. Kerrigan, *Model predictive control for deeply pipelined field-programmable gate array implementation: Algorithms and circuitry*, IET Control Theory Appl., 6 (2012), pp. 1029–1041.

[7] G. Lan and R.D.C. Monteiro, *Iteration-complexity of first-order augmented Lagrangian methods for convex programming*, Technical Report, School of Industrial and Systems Engineering, Georgia Institute of Technology, 2008.

[8] I. Necoara, J. Suykens, *Application of a Smoothing Technique to Decomposition in Convex Optimization*, IEEE Trans. Automat. Control, 53 (2008), pp. 2674–2679.

[9] V. Nedelcu, I. Necoara and Q.T. Dinh,*Computational Complexity of Inexact Gradient Augmented Lagrangian Methods: Application to Constrained MPC*, submitted.

[10] Y. Nesterov, *Introductory lectures on convex optimization*, Springer, 2004.

[11] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.

[12] P. Patrinos and A. Bemporad, *Simple and Certifiable Quadratic Programming Algorithms for Embedded Linear Model Predictive Control*, Proc. of IFAC NMPC Conference, 2012.

[13] C. V. Rao, S. J. Wright and J. B. Rawlings, *Application of interior-point methods to model predictive control*, J. Optim. Theory Appl., 99 (1998), pp. 723–757.

[14] S. Richter, M. Morari and C.N. Jones, *Towards Computational Complexity Certification for Constrained MPC Based on Lagrange Relaxation and the Fast Gradient Method*, Proc. of IEEE Conference on Decision and Control, 2011.

[15] R.T. Rockafellar, *Augmented Lagrangian and Applications of The Proximal Point Algorithm In Convex Programming*, Math. Oper. Res., 1 (1976), pp. 97–116.

[16] P.O.M Scokaert, D.Q. Mayne and J.B. Rawlings, *Suboptimal model predictive control (feasibility implies stability)*, IEEE Trans. Automat. Control, 44 (1999), pp. 648–654.