Optimal preconditioning and iteration complexity bounds for gradient-based optimization in model predictive control

Pontus Giselsson
Department of Automatic Control LTH
Lund University
Box 118, SE-221 00 Lund, Sweden
pontusg@control.lth.se

Abstract— **In this paper, optimization problems arising in model predictive control (MPC) and in distributed MPC are solved by applying a fast gradient method to the dual of the MPC optimization problem. Although the development of fast gradient methods has improved the convergence rate of gradient-based methods considerably, they are still sensitive to ill-conditioning of the problem data. Since similar optimization problems are solved several times in the MPC controller, the optimization data can be preconditioned offline to improve the convergence rate of the fast gradient method online. A natural approach to precondition the dual problem is to minimize the condition number of the Hessian matrix. However, in MPC the Hessian matrix usually becomes positive semi-definite only, i.e., the condition number is infinite and cannot be minimized. In this paper, we show how to optimally precondition the optimization data by solving a semidefinite program, where optimally refers to the preconditioning that minimizes an explicit iteration complexity bound. Although the iteration bounds can be crude, numerical examples show that the preconditioning can significantly reduce the number of iterations needed to achieve a prespecified accuracy of the solution.**

I. Introduction

Gradient-based optimization methods are known to have iterations of low complexity, but also to have slow convergence rate properties [10]. The development of fast (or accelerated) gradient methods has improved the theoretical convergence rate significantly compared to traditional gradient methods, with approximately unaltered complexity within each iteration, see [9] for the first presented fast gradient method and [10], [1], [16] for extensions and generalizations. When minimizing convex differentiable functions with a Lipschitz continuous gradient using gradient methods, a quadratic upper bound to the minimized function with the same curvature in all directions is minimized in every iteration. For ill-conditioned problems, the quadratic upper bound does not well approximate the function to be minimized. This leads to deteriorated performance of gradient-based methods when applied to such problems.

In this paper, we apply the fast gradient method presented in [5] to solve the Lagrange dual problem to model predictive control (MPC) optimization problems. If the underlying system consists of sparsely interconnected subsystems, the algorithm can be implemented in distributed fashion to enable distributed MPC (DMPC), as in [5]. In MPC and DMPC, similar optimization problems are solved several times. This implies that the problem data can be preconditioned to achieve a more well-conditioned problem with consequent improved convergence rate of the fast gradient method. The objective of the preconditioning differs if applied in an MPC or in an DMPC context. In MPC, the total number of flops should be minimized, while in DMPC, the number of iterations should be minimized while keeping the communication pattern of the distributed controller intact. Since the minimized function is approximated with a quadratic upper bound with the same curvature in all directions in gradient methods, a natural way to precondition the data is to minimize the condition number of the Hessian of the minimized function. However, if there are more constraints than variables, which is often the case for MPC-problems, the dual problem has positive semi-definite Hessian matrix only. This implies that the condition number is not defined, and cannot be minimized.

In this paper, we instead precondition the optimization data by minimizing a bound on the number of iterations needed to guarantee a prespecified accuracy of the solution. To compute such bounds, quantifiable convergence rate results are needed. The convergence rate of fast gradient methods when applied to the dual problem depends on the Lipschitz constant of the dual gradient, which is known [5], and on the norm of the optimal dual variables. Thus, by quantifying the norm of the optimal dual variables and its dependence on the initial condition, iteration complexity bounds that guarantee a prespecified accuracy of the solution can be computed. Computation of the bounds is complicated by the fact that dual variables are not constrained to be in a compact set a priori. For the case where only equality constraints are dualized, iteration complexity bounds are provided in [14] that are based on a result in [3]. These bounds are reported to be conservative [14]. In [4] and [11] similar approaches are taken to compute iteration complexity bounds when only inequality constraints are dualized. In [13], an accelerated gradient method is applied to the primal problem for input constrained MPC. An iteration bound is obtained by bounding the difference between the initial iterate and the optimal control trajectory. In this paper, the iteration complexity bounds from [4] are generalized to include also when the equality constraints are dualized. Using these bounds, we show how to precondition the optimization data optimally

by solving a semi-definite program, where optimally refers to the preconditioners that minimize the iteration complexity bound. The provided numerical example shows, although the bounds can be crude, that the preconditioning can decrease significantly the number of iteration needed to achieve a prespecified accuracy of the solution.

II. Problem setup and preliminaries

The following MPC optimization problem with initial condition $\bar{x} \in \mathbb{R}^n$ is considered

$$\begin{aligned} V_N(\bar{x}) := \min_{\mathbf{x},\mathbf{u}} \quad & \frac{1}{2}\sum_{t=0}^{N-1}(x_t^T Q x_t + u_t^T R u_t) + \frac{1}{2}x_N^T Q_N x_N \\ \text{s.t.} \quad & (x_t, u_t) \in \mathcal{X} \times \mathcal{U}, \qquad t = 0, \dots, N-1 \\ & x_{t+1} = Ax_t + Bu_t, \quad t = 0, \dots, N-1 \\ & x_N \in \mathcal{X}_f,\, x_0 = \bar{x} \end{aligned} \tag{1}$$

where $x_t \in \mathbb{R}^n$, $u_t \in \mathbb{R}^m$, $\mathbf{x} = [x_1^T, \dots, x_N^T]^T$ and $\mathbf{u} = [u_0^T, \dots, u_{N-1}^T]^T$. The cost matrices are assumed to satisfy $Q \succ 0$, $Q_N \succ 0$ and $R \succ 0$ and the constraint sets are assumed to be polytopes defined by

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid C_x x \le d_x\}, \quad \mathcal{X}_f = \{x \in \mathbb{R}^n \mid C_f x \le d_f\},$$
$$\mathcal{U} = \{u \in \mathbb{R}^m \mid C_u u \le d_u\},$$

where $C_x \in \mathbb{R}^{n_x \times n}$, $C_u \in \mathbb{R}^{n_u \times m}$ and $C_f \in \mathbb{R}^{n_f \times n}$. We also assume that $\mathcal{X}$, $\mathcal{X}_f$ and $\mathcal{U}$ are compact and contain zero in their respective interiors which implies that $d_x, d_f, d_u > 0$. By stacking all decision variables into one vector, $\mathbf{y} = [x_0, \dots, x_N, u_0, \dots, u_{N-1}]$ and introducing the cost

$$J_N(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T \mathbf{H} \mathbf{y}$$

where $\mathbf{H} \in R^{(n+m)N+n \times (n+m)N+n}$ is chosen accordingly, the optimization problem (1) can more compactly be written as

$$\begin{aligned} V_N(\bar{x}) := \min_{\mathbf{y}} \quad & J_N(\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{y} = \mathbf{b}\bar{x} \\ & \mathbf{C}\mathbf{y} \le \mathbf{d} \end{aligned} \tag{2}$$

where matrices $\mathbf{A} \in \mathbb{R}^{nN \times (n+m)N+n}$, $\mathbf{b} \in \mathbb{R}^{nN \times n}$, $\mathbf{C} \in \mathbb{R}^{(n_x+n_u)N+n_f \times (n+m)N+n}$ and $\mathbf{d} \in \mathbb{R}^{(n_x+n_u)N+n_f}$ are built according to the introduced vector $\mathbf{y}$. Dual variables $\boldsymbol{\lambda} \in \mathbb{R}^{nN}$ for the equality constraints and $\boldsymbol{\mu} \in \mathbb{R}_{\ge 0}^{(n_x+n_u)N+n_f}$ for the inequality constraints are introduced. Under the assumption that Slater's condition holds, the following dual problem is obtained (cf. [2])

$$\max_{\boldsymbol{\lambda},\boldsymbol{\mu} \ge 0} \min_{\mathbf{y}} \frac{1}{2}\mathbf{y}^T \mathbf{H} \mathbf{y} + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{y} - \mathbf{b}\bar{x}) + \boldsymbol{\mu}^T(\mathbf{C}\mathbf{y} - \mathbf{d}).$$

The dual problem can be rewritten as (cf. [5])

$$\max_{\boldsymbol{\lambda},\boldsymbol{\mu} \ge 0} -\frac{1}{2}(\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{C}^T\boldsymbol{\mu})^T \mathbf{H}^{-1}(\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{C}^T\boldsymbol{\mu}) - \boldsymbol{\lambda}^T\mathbf{b}\bar{x} - \boldsymbol{\mu}^T\mathbf{d}. \tag{3}$$

The dual function is defined as the maximand in (3), i.e.,

$$D_N(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = -\frac{1}{2}(\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{C}^T\boldsymbol{\mu})^T \mathbf{H}^{-1}(\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{C}^T\boldsymbol{\mu}) - \boldsymbol{\lambda}^T\mathbf{b}\bar{x} - \boldsymbol{\mu}^T\mathbf{d}$$

and satisfies the properties stated in the following proposition (cf. [5]).

Proposition 1: The gradient of the dual function ∇D_N is Lipschitz continuous with constant $L = \|[\mathbf{A}^T\mathbf{C}^T]^T\mathbf{H}^{-1}[\mathbf{A}^T\mathbf{C}^T]\|$. The gradient w.r.t. $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are given by

$$\nabla_{\boldsymbol{\lambda}} D_N(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = -\mathbf{A}\mathbf{H}^{-1}(\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{C}^T\boldsymbol{\mu}) - \mathbf{b}\bar{x},$$
$$\nabla_{\boldsymbol{\lambda}} D_N(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = -\mathbf{C}\mathbf{H}^{-1}(\mathbf{A}^T\boldsymbol{\lambda} + \mathbf{C}^T\boldsymbol{\mu}) - \mathbf{d}$$

respectively.

These properties are such that an accelerated gradient method [10], [1], [16], [5] can be used to solve the dual problem. Below, a cold-starting variant, i.e., with $\boldsymbol{\lambda}^0 = 0$ and $\boldsymbol{\mu}^0 = 0$, of the algorithm in [5] is presented.

Algorithm 1: **Accelerated gradient algorithm**

Initialize $\boldsymbol{\lambda}^0 = \boldsymbol{\lambda}^{-1} = 0$, $\boldsymbol{\mu}^0 = \boldsymbol{\mu}^{-1} = 0$ and $\mathbf{y}^{-1} = 0$.
For $k \ge 0$

$$\mathbf{y}^k = -\mathbf{H}^{-1}(\mathbf{A}^T\boldsymbol{\lambda}^k + \mathbf{C}^T\boldsymbol{\mu}^k)$$
$$\widetilde{\mathbf{y}}^k = \mathbf{y}^k + \frac{k-1}{k+2}(\mathbf{y}^k - \mathbf{y}^{k-1})$$
$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \frac{k-1}{k+2}(\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1}) + \frac{1}{L}\left(\mathbf{A}\widetilde{\mathbf{y}}^k - \mathbf{b}\bar{x}\right)$$
$$\boldsymbol{\mu}^{k+1} = \max\left\{0, \boldsymbol{\mu}^k + \frac{k-1}{k+2}(\boldsymbol{\mu}^k - \boldsymbol{\mu}^{k-1}) + \right.$$
$$\left. + \frac{1}{L}\left(\mathbf{C}\widetilde{\mathbf{y}}^k - \mathbf{d}\right)\right\}$$

The set of optimal dual variables is denoted by

$$M^*(\bar{x}) = \{\boldsymbol{\lambda}, \boldsymbol{\mu} \mid D_N(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \ge V_N(\bar{x})\}.$$

The set of initial conditions for which (2) is feasible is denoted by $\mathbb{X}_N$. The optimal solution to (2) with initial condition $\bar{x} \in \mathbb{X}_N$ is denoted by $\mathbf{y}^*(\bar{x})$. Next, we state the convergence rate properties of Algorithm 1.

Proposition 2: Suppose that $\bar{x} \in \mathbb{X}_N$. For every $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in M^*(\bar{x})$, Algorithm 1 has the following convergence rate properties:

1) For all $k \ge 1$ the dual function converges as

$$D_N(\bar{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) - D_N(\bar{x}, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) \le \frac{2L}{(k+1)^2}\left\|\begin{bmatrix}\boldsymbol{\lambda}^* \\ \boldsymbol{\mu}^*\end{bmatrix}\right\|^2 \tag{4}$$

2) The primal variable rate of convergence is

$$\|\mathbf{y}^k - \mathbf{y}^*(\bar{x})\|^2 \le \frac{4L}{\underline{\sigma}(\mathbf{H})(k+1)^2}\left\|\begin{bmatrix}\boldsymbol{\lambda}^* \\ \boldsymbol{\mu}^*\end{bmatrix}\right\|^2, \forall k \ge 1 \tag{5}$$

where $\underline{\sigma}(\mathbf{H})$ denotes the smallest eigenvalue to $\mathbf{H}$.

Proof. Argument 1 is proven in [1], [16], [5] and argument 2 is proven in [5]. □

To compute lower iteration bounds for the cold starting case, i.e., with $\boldsymbol{\lambda}^0 = 0$ and $\boldsymbol{\mu}^0 = 0$, to guarantee a prespecified dual function value or primal variable accuracy is the objective of this paper. We will show how to compute

bounds that hold for every $\bar{x} \in \beta\mathbb{X}_N$ with $\beta \in (0,1)$ and where $\beta\mathbb{X}_N$ is defined as

$$\beta\mathbb{X}_N := \{\bar{x} \in \mathbb{R}^n \mid \frac{1}{\beta}\bar{x} \in \mathbb{X}_N\}.$$

Since the set $\mathbb{X}_N$ is convex and $0 \in \mathbb{X}_N$ (cf. [12]) we have that $\beta\mathbb{X}_N \subseteq \mathbb{X}_N$ and that $0 \in \beta\mathbb{X}_N$. Before we proceed with the presentation we introduce

$$P := \mathbf{b}^T(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T)^{-1}\mathbf{b}$$

which characterizes the optimal solution with equality constraints only and satisfies

$$\frac{1}{2}\bar{x}^T P\bar{x} = \max_{\boldsymbol{\lambda}} D_N(\bar{x}, \boldsymbol{\lambda}, 0) \leq V_N(\bar{x}). \tag{6}$$

We also make the following definition.

Definition 1: The scalar $\kappa \geq 1$ is defined as the smallest scalar such that for every $\bar{x} \in \mathbb{X}_N$ we have

$$V_N(\bar{x}) \leq \frac{\kappa}{2}\bar{x}^T P\bar{x}. \tag{7}$$

Remark 1: The exact value of κ can be found by solving a mixed integer linear program (MILP), since minimizing κ subject to (7) can be cast as a bilevel optimization problem with convex inner problem and indefinite outer cost, see [6, Theorem 2]. The iterations in the MILP solver can be stopped before convergence to obtain an upper bound to the κ value. Finally, we make the following assumption.

Assumption 1: We assume that $\mathbf{A}$ has full row rank.

A. Notation

The Euclidean norm and the induced Euclidean norm are denoted by $\|\cdot\|$ and $\langle x, y\rangle = x^T y$. Further, $\bar{\sigma}(H)$ and $\underline{\sigma}(H)$ are the largest and smallest singular value of H respectively. Finally, the i:th element in a vector is denoted by $[\cdot]_i$.

III. Lagrange multiplier norm bounds

The only unknown quantity in the bounds in Proposition 2 is the norm of the optimal dual variables. The topic of this section is to show how such norms can be computed for any initial condition $\bar{x} \in \beta\mathbb{X}_N$ with $\beta \in (0,1)$. The following result is a straightforward generalization of the result in [8, Lemma 1].

Lemma 1: Assume that there exists a vector $\bar{\mathbf{y}}(\bar{x})$ such that $\mathbf{C}\bar{\mathbf{y}}(\bar{x}) < \mathbf{d}$ and $\mathbf{A}\bar{\mathbf{y}}(\bar{x}) = \mathbf{b}\bar{x}$. Then for every $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in M^*(\bar{x})$ we have that $\boldsymbol{\mu}^*$ satisfies

$$\|\boldsymbol{\mu}^*\| \leq \frac{1}{\gamma(\bar{\mathbf{y}}(\bar{x}))}(J_N(\bar{\mathbf{y}}(\bar{x})) - V_N(\bar{x}))$$

where $\gamma(\bar{\mathbf{y}}(\bar{x})) := \min_{1\leq j\leq (n_x+n_u)N+n_f} -[\mathbf{C}\bar{\mathbf{y}}(\bar{x}) - \mathbf{d}]_j$.

Proof. For every $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in M^*(\bar{x})$ we have

$$\begin{aligned}
V_N(\bar{x}) &= \inf_{\mathbf{y}} J_N(\mathbf{y}) + (\boldsymbol{\lambda}^*)^T(\mathbf{A}\mathbf{y} - \mathbf{b}\bar{x}) + (\boldsymbol{\mu}^*)^T(\mathbf{C}\mathbf{y} - \mathbf{d}) \\
&\leq J_N(\bar{\mathbf{y}}(\bar{x})) + (\boldsymbol{\lambda}^*)^T(\mathbf{A}\bar{\mathbf{y}}(\bar{x}) - \mathbf{b}\bar{x}) + \\
&\qquad + (\boldsymbol{\mu}^*)^T(\mathbf{C}\bar{\mathbf{y}}(\bar{x}) - \mathbf{d}) \\
&\leq J_N(\bar{\mathbf{y}}(\bar{x})) - \gamma(\bar{\mathbf{y}}(\bar{x}))(\boldsymbol{\mu}^*)^T\mathbf{1} \\
&= J_N(\bar{\mathbf{y}}(\bar{x})) - \gamma(\bar{\mathbf{y}}(\bar{x}))\|\boldsymbol{\mu}^*\|_1 \\
&\leq J_N(\bar{\mathbf{y}}(\bar{x})) - \gamma(\bar{\mathbf{y}}(\bar{x}))\|\boldsymbol{\mu}^*\|.
\end{aligned}$$

Rearranging the terms gives the result. □

By constructing a strictly feasible vector $\bar{\mathbf{y}}(\bar{x})$ in Lemma 1, referred to as a *Slater vector*, a bound on the norm of the optimal Lagrange multipliers associated with the inequality constraints can be computed. Next, a straightforward generalization to [4, Lemma 2] is presented where it was shown how a Slater vector to (2) for every initial state $\bar{x} \in \beta\mathbb{X}_N$ can be constructed. Before the lemma is presented we introduce $d_{\min} := \min_j [\mathbf{d}]_j > 0$.

Lemma 2: For every $\bar{x} \in \beta\mathbb{X}_N$ with $\beta \in (0,1)$, a Slater vector to the optimization problem (2) is given by $\bar{\mathbf{y}}(\bar{x}) = \beta\mathbf{y}^*(\bar{x}/\beta)$. Further, $\gamma(\bar{\mathbf{y}}(\bar{x})) \geq (1-\beta)d_{\min}$.

Proof. We first note that

$$\mathbf{A}\bar{\mathbf{y}}(\bar{x}) = \beta\mathbf{A}\mathbf{y}^*(\frac{\bar{x}}{\beta}) = \beta\mathbf{b}\frac{\bar{x}}{\beta} = \mathbf{b}\bar{x}$$

which implies that the equality constraints are satisfied. Further,

$$\mathbf{C}\bar{\mathbf{y}}(\bar{x}) = \beta\mathbf{C}\mathbf{y}^*(\frac{\bar{x}}{\beta}) \leq \beta\mathbf{d} = \mathbf{d} - (1-\beta)\mathbf{d}.$$

Hence, $-(\mathbf{C}\bar{\mathbf{y}}(\bar{x}) - \mathbf{d}) \geq (1-\beta)\mathbf{d}$ which by definition of the function γ and $d_{\min}$ gives the result. □

Next, we present a theorem that, using Lemma 1 and Lemma 2, shows how a bound on the norm of the optimal dual variables can be computed. Before we present the theorem, we introduce the matrices

$$\Phi := \mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T, \quad \Psi := (\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{H}^{-1}\mathbf{C}^T. \tag{8}$$

From Assumption 1 we know that $\mathbf{A}$ has full row rank. Further, $\mathbf{H}$ is positive definite. Hence, $\Phi = \mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T$ is invertible and Ψ exists.

Theorem 1: For every $\bar{x} \in \beta\mathbb{X}_N$ we have for every $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in M^*(\bar{x})$ that

$$\left\|\begin{bmatrix}\boldsymbol{\lambda}^* \\ \boldsymbol{\mu}^*\end{bmatrix}\right\| \leq \left\|\begin{bmatrix}\Psi \\ I\end{bmatrix}\right\| \frac{\kappa - 1}{2(1-\beta)d_{\min}}\bar{x}^T P\bar{x} + \|\Phi^{-1}\mathbf{b}\bar{x}\|. \tag{9}$$

Proof. Using the Slater vector $\bar{\mathbf{y}}(\bar{x}) = \beta\mathbf{y}^*(\bar{x}/\beta)$ we get

$$\begin{aligned}
J_N(\bar{\mathbf{y}}(\bar{\bar{x}})) &= J_N(\beta\mathbf{y}^*(\frac{\bar{x}}{\beta})) = \beta^2\frac{1}{2}\mathbf{y}^*(\frac{\bar{x}}{\beta})^T\mathbf{H}\mathbf{y}^*(\frac{\bar{x}}{\beta}) \\
&= \beta^2 V_N(\frac{\bar{x}}{\beta}) \leq \beta^2\frac{\kappa}{2}\begin{bmatrix}\frac{\bar{x}}{\beta}\end{bmatrix}^T P\begin{bmatrix}\frac{\bar{x}}{\beta}\end{bmatrix} = \frac{\kappa}{2}\bar{x}^T P\bar{x}
\end{aligned}$$

where the inequality comes from Definition 1. Further, KKT conditions to (2) and Proposition 1 gives that for every $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in M^*(\bar{x})$ we have

$$-\mathbf{A}\mathbf{H}^{-1}(\mathbf{A}^T\boldsymbol{\lambda}^* + \mathbf{C}^T\boldsymbol{\mu}^*) = \mathbf{b}\bar{x}.$$

This implies that

$$\begin{aligned}
\boldsymbol{\lambda}^* &= -(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{H}^{-1}\mathbf{C}^T\boldsymbol{\mu}^* + \mathbf{b}\bar{x}) \\
&= -\Psi\boldsymbol{\mu}^* - \Phi^{-1}\mathbf{b}\bar{x}
\end{aligned}$$

where the last equality comes from the definitions of Φ and Ψ in (8). This gives

$$\begin{aligned}\left\|\begin{bmatrix}\boldsymbol{\lambda}^*\\ \boldsymbol{\mu}^*\end{bmatrix}\right\| &= \left\|\begin{bmatrix}-\Psi\\ I\end{bmatrix}\boldsymbol{\mu}^* + \begin{bmatrix}-\Phi^{-1}\mathbf{b}\bar{x}\\ 0\end{bmatrix}\right\| \\ &\leq \left\|\begin{bmatrix}\Psi\\ I\end{bmatrix}\right\| \|\boldsymbol{\mu}^*\| + \|\Phi^{-1}\mathbf{b}\bar{x}\| \\ &\leq \left\|\begin{bmatrix}\Psi\\ I\end{bmatrix}\right\| \frac{1}{\gamma(\bar{\mathbf{y}}(\bar{x}))}(J_N(\bar{\mathbf{y}}(\bar{x}) - V_N(\bar{x})) + \|\Phi^{-1}\mathbf{b}\bar{x}\| \\ &\leq \left\|\begin{bmatrix}\Psi\\ I\end{bmatrix}\right\| \frac{\kappa-1}{2(1-\beta)d_{\min}}\bar{x}^T P\bar{x} + \|\Phi^{-1}\mathbf{b}\bar{x}\|\end{aligned}$$

where the second inequality comes from Lemma 1 and the final inequality from Lemma 2, Definition 1 and (6). This completes the proof. □

In the following section, the bound on the optimal dual variables is used, together with the convergence rate results in Proposition 2, to compute lower iteration bounds to achieve a prespecified dual function value and primal variable accuracy.

IV. Algorithm iteration bounds

Lower iteration bounds to achieve prespecified dual function value and primal variable tolerances are presented in this section. We consider bounds for the cold starting case, i.e., when $\boldsymbol{\lambda}^0 = 0$ and $\boldsymbol{\mu}^0 = 0$.

A. Iteration bound to guarantee dual ϵ-solution

First, a lower iteration bound to achieve a prespecified dual function value accuracy is presented. As in [4], a relative tolerance is used to avoid that a scaling of the Q and R matrices affects the bound.

Theorem 2: Suppose that Algorithm 1 is initialized with $\boldsymbol{\lambda}^0 = 0$ and $\boldsymbol{\mu}^0 = 0$. Then for every $\bar{x} \in \beta\mathbb{X}_N$ with $\beta \in (0,1)$ we have

$$V_N(\bar{x}) - D(\bar{x}, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) \leq \epsilon_d V_N(\bar{x}) \tag{10}$$

for every $k \geq k_d(\bar{x})$ where

$$k_d(\bar{x}) = 2\sqrt{\frac{L}{\epsilon_d}}\left(\frac{(\kappa-1)\sqrt{\bar{x}^T P\bar{x}}}{2(1-\beta)}\nu + \rho\right) - 1 \tag{11}$$

and $\rho = \|\Phi^{-1}\mathbf{b}P^{-1/2}\|$ and $\nu = \|[\Psi^T\ I^T]^T\|/d_{\min}$.

Proof. Inequality (10) is equivalent to

$$D_N(\bar{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) - D_N(\bar{x}, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k) \leq \epsilon_d D_N(\bar{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

for any $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in M^*(\bar{x})$. From Proposition 2 and (6) we conclude that (10) holds if

$$\frac{2L}{(k+1)^2}\left\|\begin{bmatrix}\boldsymbol{\lambda}^*\\ \boldsymbol{\mu}^*\end{bmatrix}\right\|^2 \leq \epsilon_d \frac{1}{2}\bar{x}^T P\bar{x}. \tag{12}$$

Insertion of the bound in Theorem 1 into (12) and rearranging the terms gives

$$k_d(\bar{x}) = 2\sqrt{\frac{L}{\epsilon_d}}\left(\frac{(\kappa-1)\sqrt{\bar{x}^T P\bar{x}}}{2(1-\beta)}\nu + \frac{\|\Phi^{-1}\mathbf{b}\bar{x}\|}{\sqrt{\bar{x}^T P\bar{x}}}\right) - 1.$$

We have

$$\begin{aligned}\frac{\|\Phi^{-1}\mathbf{b}\bar{x}\|}{\sqrt{\bar{x}^T P\bar{x}}} \leq \rho &\iff \frac{\|\Phi^{-1}\mathbf{b}\bar{x}\|^2}{\|P^{1/2}\bar{x}\|^2} \leq \rho^2 \\ &\iff \bar{x}^T\mathbf{b}^T\Phi^{-2}\mathbf{b}\bar{x} \leq \rho^2\bar{x}^T P\bar{x}.\end{aligned}$$

Since $0 \in \text{int}(\beta\mathbb{X}_N)$ this holds for every $\bar{x} \in \beta\mathbb{X}_N$ if and only if ρ is such that

$$\begin{aligned}\mathbf{b}^T\Phi^{-2}\mathbf{b} \preceq \rho^2 P &\iff P^{-1/2}\mathbf{b}^T\Phi^{-2}\mathbf{b}P^{-1/2} \preceq \rho^2 I \\ &\iff \|\Phi^{-1}\mathbf{b}P^{-1/2}\| \preceq \rho.\end{aligned} \tag{13}$$

Choosing ρ such that the last step in (13) holds with equality completes the proof. □

Remark 2: As in [4], the lower iteration bound is not affected by scaling the cost matrices by a factor $a > 0$. This is true since for cost matrices $Q_a = aQ$ and $R_a = aR$ we get $L_a = \frac{1}{a}L$, $P_a = aP$, $\rho_a = \sqrt{a}\rho$, and $\nu_a = \nu$. By insertion into (11) the factor a is cancelled.

Remark 3: It is desirable to compute a lower iteration bound for all $\bar{x} \in \beta\mathbb{X}_N$. By maximizing $k_d(\bar{x})$ subject to $\bar{x} \in \beta\mathbb{X}_N$ this can be obtained. Since it is often difficult to describe the set $\bar{x} \in \beta\mathbb{X}_N$ an over estimator to the lower iteration bound is found by maximizing $k_d(\bar{x})$ subject to $\bar{x} \in \beta\mathcal{X}$. The only $\bar{x}$-dependency in the iteration bound is $\sqrt{\bar{x}^T P\bar{x}}$ which affects the bound affinely. By maximizing $\bar{x}^T P\bar{x}$ over $\beta\mathcal{X}$, which is a quadratic convex maximization problem, the maximizing $\bar{x}$ can be found. Such problems are NP-complete but can be rewritten as MILPs as shown in [6, Lemma 2]. There are efficient MILP solvers that, in every iteration, produce upper and lower bounds to the optimal value. An upper bound to the optimal value is enough to compute an iteration bound, hence the MILP solver can be stopped when sufficient accuracy has been reached.

B. Primal variable iteration bound

In this section, a bound on the number of iterations needed to guarantee a prespecified distance between the algorithm primal variables and the optimal primal variables is presented.

Theorem 3: Suppose that Algorithm 1 is initialized with $\boldsymbol{\lambda}^0 = 0$ and $\boldsymbol{\mu}^0 = 0$. Then for every $\bar{x} \in \beta\mathbb{X}_N$ we have

$$\|\mathbf{y}^k - \mathbf{y}^*(\bar{x})\| \leq \epsilon_p \tag{14}$$

for every $k \geq k_p(\bar{x})$ where

$$k_p(\bar{x}) = \frac{2}{\epsilon_p}\sqrt{\frac{L}{\underline{\sigma}(\mathbf{H})}}\left(\left\|\begin{bmatrix}\Psi\\ I\end{bmatrix}\right\| \frac{(\kappa-1)\bar{x}^T P\bar{x}}{2(1-\beta)d_{\min}} + \|\Phi^{-1}\mathbf{b}\bar{x}\|\right) - 1.$$

Proof. From Proposition 2 we have that (14) holds if

$$\frac{2}{k+1}\sqrt{\frac{L}{\underline{\sigma}(\mathbf{H})}}\left\|\begin{bmatrix}\boldsymbol{\lambda}^*\\ \boldsymbol{\mu}^*\end{bmatrix}\right\| \leq \epsilon_p$$

Insertion of the bound in Theorem 1 and rearranging gives the result. □

V. Optimal preconditioning

In this section, we show how to precondition the problem data such that the iteration complexity bound in Theorem 2 is minimized. We precondition the equality constraints with an invertible matrix E such that $E\mathbf{A}\mathbf{y} = E\mathbf{b}\bar{x}$ and the inequality constraints with a diagonal matrix F with positive diagonal elements such that $F\mathbf{C}\mathbf{y} \leq F\mathbf{d}$. To keep the sparse structure of the equality constraints, E should satisfy $E \in \mathcal{E}$ where $\mathcal{E}$ defines the sparsity structure. The set $\mathcal{E}$ should be chosen such that if $E \in \mathcal{E}$ then also $E^TE \in \mathcal{E}$. This holds, e.g., for diagonal matrices, for block-diagonal matrices, and for matrices that become block-diagonal after left multiplication of a specific permutation matrix and right multiplication of its transpose. In the following lemma we show that the optimal preconditioner for the inequality constraint has the form $F = tD^{-1}$ where $t > 0$ and $D := \operatorname{diag}(\mathbf{d})$.

Lemma 3: The optimal preconditioner for the inequality constraints satisfies $F = tD^{-1}$ for some $t > 0$ and for any fixed Lipschitz constant $L > 0$ to ∇D_N, where optimal refers to the preconditioners that minimize the iteration bound in Theorem 2.

Proof. Since F and D are diagonal matrices with positive elements, F can be represented as $F = GD^{-1}$ where G is a diagonal matrix with positive elements. The variables in the iteration bound in Theorem 2 that are affected by the preconditioning are ρ, ν and L. For preconditioners E and $F = GD^{-1}$, ρ satisfies

$$\begin{aligned}\rho &= \|(E\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^TE^T)^{-1}E\mathbf{b}P^{-1/2}\| \\ &= \|E^{-T}(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T)^{-1}E^{-1}E\mathbf{b}P^{-1/2}\| \\ &= \|E^{-T}\Phi^{-1}\mathbf{b}P^{-1/2}\| \end{aligned} \tag{15}$$

and ν satisfies

$$\begin{aligned}\nu &= \left\|\begin{bmatrix}(E\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^TE^T)^{-1}E\mathbf{A}\mathbf{H}^{-1}\mathbf{C}^TD^{-T}G^T \\ I\end{bmatrix}\right\|/d_{\min} \\ &= \frac{\left\|\begin{bmatrix}E^{-T}(\mathbf{A}\mathbf{H}^{-1}\mathbf{A}^T)^{-1}E^{-1}E\mathbf{A}\mathbf{H}^{-1}\mathbf{C}^TD^{-T}G^T \\ I\end{bmatrix}\right\|}{\min_j[GD^{-1}\mathbf{d}]_j} \\ &= \left\|\begin{bmatrix}E^{-T}\Psi D^{-T}G^T \\ I\end{bmatrix}\right\|/\lambda_{\min}(G). \end{aligned} \tag{16}$$

where Φ and Ψ are defined in (8). Further, the Lipschitz constant to ∇D_N is bounded by L for all feasible E and $F = GD^{-1}$, i.e., they must satisfy (see Proposition 1)

$$L \geq \left\|\begin{bmatrix}E\mathbf{A} \\ GD^{-1}\mathbf{C}\end{bmatrix}\mathbf{H}^{-1}\begin{bmatrix}E\mathbf{A} \\ GD^{-1}\mathbf{C}\end{bmatrix}^T\right\|. \tag{17}$$

Next, we show that the optimal F satisfies $F = tD^{-1}$ where $t > 0$, i.e., that the optimal $G = tI$. We represent G as $G = tI + \widetilde{G}$ where $\widetilde{G}$ is diagonal and $\widetilde{G} \succeq 0$. This implies that (16) is equivalent to ν being the smallest scalar such that

$$GD^{-1}\Psi^TE^{-1}E^{-T}\Psi D^{-T}G^T + I \preceq \nu^2t^2I$$

which in turn is equivalent to ν being the smallest scalar such that

$$\begin{aligned}(\nu^2t^2 - 1)I &\succeq E^{-T}\Psi D^{-T}G^TGD^{-1}\Psi^TE^{-1} \\ &= E^{-T}\Psi D^{-T}(t^2I + 2\widetilde{G}t + \widetilde{G}^T\widetilde{G})D^{-1}\Psi^TE^{-1}.\end{aligned}$$

Hence, for given $t > 0$, $\widetilde{G} = 0$ gives the smallest ν independent of E. Further, the Lipschitz constant constraint (17) is equivalent to

$$\begin{aligned}LI \succeq \mathbf{H}^{-1/2}[\mathbf{A}^TE^T, \mathbf{C}^TD^{-T}G^T]\times \\ \times [\mathbf{A}^TE^T, \mathbf{C}^TD^{-T}G^T]^T\mathbf{H}^{-1/2}\end{aligned}$$

which in turn is equivalent to

$$\begin{aligned}L\mathbf{H} &\succeq \mathbf{A}^TE^TE\mathbf{A} + \mathbf{C}^TD^{-T}G^TGD^{-1}\mathbf{C} \\ &= \mathbf{A}^TE^TE\mathbf{A} + \mathbf{C}^TD^{-T}(t^2I + 2\widetilde{G}t + \widetilde{G}^T\widetilde{G})D^{-1}\mathbf{C}.\end{aligned}$$

We introduce $X(\widetilde{G}) = L\mathbf{H} - \mathbf{C}^TD^{-T}(t^2I + 2\widetilde{G}t + \widetilde{G}^T\widetilde{G})D^{-1}\mathbf{C}$ which satisfies $X(\widetilde{G}_1) \prec X(\widetilde{G}_2)$ if $\widetilde{G}_1 \succ \widetilde{G}_2 \succeq 0$. We also define the set of feasible preconditioners for the equality constraints

$$\Sigma(\widetilde{G}) = \{E \in \mathcal{E} \mid \mathbf{A}^TE^TE\mathbf{A} \preceq X(\widetilde{G})\}.$$

For $E \in \Sigma(0)$ we have $\mathbf{A}^TE^TE\mathbf{A} \preceq X(0)$ and for $E \in \Sigma(\widetilde{G})$ with $\widetilde{G} \succ 0$ we have $\mathbf{A}^TE^TE\mathbf{A} \preceq X(\widetilde{G}) \prec X(0)$, which implies $\Sigma(\widetilde{G}) \subset \Sigma(0)$ for every $\widetilde{G} \succ 0$. Hence, by setting $\widetilde{G} = 0$ the variable ν is as small as possible and the set of feasible E is as large as possible. This implies that if $\widetilde{G} = 0$ then E gets maximal freedom in minimizing (15) and (16) while respecting the constraint (17). This concludes the proof. □

Before we state the theorem about how to compute the optimal preconditioner we define $P_x := \max_{x\in\mathbb{X}}\sqrt{x^TPx}$. This implies $\max_{x\in\beta\mathbb{X}}\sqrt{x^TPx} = \beta P_x$.

Theorem 4: The preconditioners E and F that minimize the iteration bound in Theorem 2 for fixed $\beta \in (0, 1)$ are found by solving the following semidefinite program

$$\begin{aligned}\min\ & \frac{(\kappa-1)\beta P_x}{2(1-\beta)}\nu + \rho \\ \text{s.t.}\ & \begin{bmatrix}Z & \theta\Phi^{-1}\mathbf{b} \\ \theta\mathbf{b}^T\Phi^{-T} & P\end{bmatrix} \succeq 0 && (18) \\ & \begin{bmatrix}Z & \phi\Psi D^{-T} \\ \phi D^{-1}\Psi^T & (1-s)I\end{bmatrix} \succeq 0 && (19) \\ & \begin{bmatrix}sI & \phi D^{-1}\mathbf{C} \\ \phi\mathbf{C}^TD^{-T} & L\mathbf{H} - \mathbf{A}^TZ\mathbf{A}\end{bmatrix} \succeq 0 && (20) \\ & \begin{bmatrix}\nu & 1 \\ 1 & \phi\end{bmatrix} \succeq 0 && (21) \\ & \begin{bmatrix}\rho & 1 \\ 1 & \theta\end{bmatrix} \succeq 0 && (22) \\ & Z \succ 0, Z \in \mathcal{E} \\ & \rho > 0, \theta > 0, \phi > 0, \nu > 0, s > 0\end{aligned}$$

where $Z = E^TE$, $s = \frac{\phi^2}{t^2}$, $F = tD^{-1}$, and L is the maximal Lipschitz constant to ∇D_N. Further, the optimal value is the same for any choice of $L > 0$.

Proof. The variables in the iteration bound in Theorem 2 that are affected by the preconditioning are ρ, ν and L. Variables ρ and ν are given by (15) and (16) respectively when preconditioning is used, and the constraint imposed by L is given in (17) in the preconditioning case. We will show that the posed semidefinite program implies (15), (16), and (17) and that it finds the preconditioners that minimize the iteration bound in Theorem 2.

Schur complement of (22) gives $\rho \geq 1/\theta$. Further, Schur complement gives that (18) and $Z \succ 0$ implies

$$\mathbf{b}^T \Phi^{-T} Z^{-1} \Phi^{-1} \mathbf{b} \preceq \frac{1}{\theta^2} P \preceq \rho^2 P.$$

Hence

$$\|E^{-T} \Phi^{-1} \mathbf{b} P^{-1/2}\| \leq \rho \tag{23}$$

and (15) is implied by choosing the smallest ρ such that (23) holds. Schur complement of (21) gives $\nu \geq 1/\phi$ and Schur complement of (19) gives

$$D^{-1} \Psi^T Z^{-1} \Psi D^{-T} \preceq \frac{1}{\phi^2}(1-s)I \preceq (\nu^2 - \frac{1}{t^2})I.$$

This is equivalent to

$$tD^{-1} \Psi^T Z^{-1} \Psi D^{-T} t + I \preceq \nu^2 t^2 I = \nu^2 d_{\min}^2 I$$

which in turn is equivalent to

$$\left\| \begin{bmatrix} E^{-T} \Psi D^{-T} t \\ I \end{bmatrix} \right\| \frac{1}{d_{\min}} \preceq \nu. \tag{24}$$

By choosing the smallest ν such that (24) holds, (16) is satisfied since by Lemma 3, $G = tI$ is optimal and since $\lambda_{\min}(G) = d_{\min}$. Further, since $s > 0$ Schur complement of (20) gives

$$\begin{aligned} L\mathbf{H} &\succeq \mathbf{A}^T Z \mathbf{A} + \mathbf{C}^T D^{-T} \frac{\phi^2}{s} D^{-1} \mathbf{C} \\ &= \mathbf{A}^T Z \mathbf{A} + \mathbf{C}^T D^{-T} t^2 D^{-1} \mathbf{C} \end{aligned}$$

which is equivalent to

$$L \geq \|\mathbf{H}^{-1/2} [\mathbf{A}^T E^T, \mathbf{C}^T D^{-T} t][\mathbf{A}^T E^T, \mathbf{C}^T D^{-T} t]^T \mathbf{H}^{-1/2}\|$$

which in turn is equivalent to

$$L \geq \left\| \begin{bmatrix} E\mathbf{A} \\ tD^{-1}\mathbf{C} \end{bmatrix} \mathbf{H}^{-1} \begin{bmatrix} E\mathbf{A} \\ tD^{-1}\mathbf{C} \end{bmatrix}^T \right\|. \tag{25}$$

This implies that the Lipschitz constant constraint (17) holds. Next we show that the cost

$$\min \frac{(\kappa - 1)\beta P_x}{2(1-\beta)} \nu + \rho \tag{26}$$

implies that (25) holds with equality which implies that the iteration bound (11) is minimized. Since $t = d_{\min}$ the bound (24) is equivalent to

$$\left\| \begin{bmatrix} E^{-T} \Psi D^{-T} \\ 1/t \end{bmatrix} \right\| \preceq \nu.$$

Further, (23) depends on E^{-1}. This implies that ρ and ν are decreasing when E and t are increasing, while the r.h.s. of (25) is increasing with E and t. Hence, the optimal preconditioners must have equality in (25). Further, since (25) holds with equality, the Lipschitz constant L in (11) is fixed and (26) minimizes the iteration bound (11).

It remains to show that the iteration bound (11) is independent of the choice of $L > 0$. The iteration bound (11) depends on $\sqrt{L}\rho$ and $\sqrt{L}\nu$ and the only hard constraints are (25) and the constraints on positive definiteness. We introduce the set of feasible preconditioners for fixed L as follows

$$\Theta(L) = \{t > 0, E \in \mathcal{E} \mid E \succ 0 \text{ and (25) holds}\}.$$

Since $\mathcal{E}$ is only a sparsity constraint we get for any $L_1 > 0$ and $L_2 > 0$ that $\sqrt{L_2}\Theta(L_1) = \sqrt{L_1}\Theta(L_2)$. Hence every pair $(t_2, E_2) \in \Theta(L_2)$ can be described as $(t_2, E_2) = \sqrt{L_2}(t_1, E_1)$ where $(t_1, E_1) \in \Theta(L_1)$ and $L_1 = 1$. We denote by ρ_1 and ν_1 the bounds (23) and (24) using t_1 and E_1 and by ρ_2 and ν_2 the bound (23) and (24) using t_2 and E_2. We see from (23) that using $E_2 = \sqrt{L_2}E_1$ and $t_2 = \sqrt{L_2}t_1$ gives $\rho_2 = \rho_1/\sqrt{L_2}$ and from (24) we conclude that $\nu_2 = \nu_1/\sqrt{L_2}$. Since the iteration bound depends on $\sqrt{L}\rho$ and $\sqrt{L}\nu$ we get $\sqrt{L_2}\rho_2 = \sqrt{L_2}\rho_1/\sqrt{L_2} = 1\rho_1 = \sqrt{L_1}\rho_1$ and $\sqrt{L_2}\nu_2 = \sqrt{L_2}\nu_1/\sqrt{L_2} = 1\nu_1 = \sqrt{L_1}\nu_1$. Hence, the choice of L does not influence the iteration bound. This completes the proof. □

Remark 4: The matrix $Z \in \mathcal{E}$ is symmetric and positive definite and can hence be decomposed as $Z = U\Sigma U^T$, where U is unitary and Σ is diagonal with positive diagonal elements. This decomposition, together with the requirement on the sparsity, i.e., that $E \in \mathcal{E}$ and $E^T E \in \mathcal{E}$, implies that by choosing preconditioner $E = U\Sigma^{1/2}U^T$ we get $E \in \mathcal{E}$ and $E^T E = U\Sigma^{1/2}U^T U\Sigma^{1/2}U^T = U\Sigma U^T = Z$. The preconditioning matrix F is readily computed by setting $F = \frac{\phi}{\sqrt{s}} D^{-1}$.

Remark 5: Denoting the dual function for the preconditioned problem $D_{N,\mathrm{pc}}(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ and keeping the notation for the original unconditioned problem $D_N(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$, gives that

$$D_N(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = D_{N,\mathrm{pc}}(\bar{x}, E^{-1}\boldsymbol{\lambda}, F^{-1}\boldsymbol{\mu}).$$

Denoting primal variables computed from unconditioned problem in Algorithm 1 by $\mathbf{y}(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ and primal variables computed from preconditioned problem in Algorithm 1 by $\mathbf{y}_{\mathrm{pc}}(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$, gives that

$$\mathbf{y}(\bar{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{y}_{\mathrm{pc}}(\bar{x}, E^{-1}\boldsymbol{\lambda}, F^{-1}\boldsymbol{\mu}).$$

This implies that for a certain dual accuracy of the preconditioned problem, the same accuracy of the primal solution is obtained as if the original unconditioned problem is used. Hence, the accuracy of the primal solution for a specific dual optimality tolerance is unaffected by the preconditioning. However, the set of optimal dual variables is linearly transformed between the original and preconditioned problems.

VI. Numerical example

The efficiency of the preconditioning and the conservatism of the iteration bound are evaluated by applying the optimization algorithm on a DMPC problem where the dynamics

TABLE I

EXPERIMENTAL RESULTS FOR ALGORITHM 1 WITH MPC-, DMPC-, AND WITHOUT PRECONDITIONING. THE NUMBER OF ALGORITHM ITERATIONS AND ITERATION COMPLEXITY BOUNDS ARE PRESENTED.

ϵ_v	β	precond	# iters		iter bound
			avg.	max.	
0.005	0.25	DMPC	13.92	49	487
0.005	0.25	MPC	56.76	119	1536
0.005	0.25	no	149.51	252	5087
0.005	0.50	DMPC	32.24	71	1333
0.005	0.50	MPC	71.06	120	3399
0.005	0.50	no	155.77	262	12463
0.005	0.75	DMPC	38.08	167	3877
0.005	0.75	MPC	81.28	242	8986
0.005	0.75	no	159.84	330	34585
0.005	0.90	DMPC	41.70	232	11505
0.005	0.90	MPC	87.20	340	25905
0.005	0.90	no	164.84	370	100967

matrix is randomly generated and has sparse structure. The system is unstable since the largest eigenvalue of the dynamics matrix is 1.1. The system has 3 sub-systems with 5 states and 1 input each, i.e., in total 15 states and 3 inputs. The state and input variables are upper and lower bounded by random numbers in the intervals $[0.5\ 1.5]$ and $[-0.15\ -0.05]$ respectively. The cost matrices are diagonal and each diagonal element is randomly chosen from the interval $[1\ 100]$. The control horizon is $N = 6$. We use two different preconditionings, one MPC preconditioning and one DMPC preconditioning. The objective of the MPC preconditioning is to reduce the total number of flops. We use a diagonal matrix for preconditioning of the equality constraints, E, which implies that the number of flops per iteration is unchanged after preconditioning. For the DMPC preconditioning, the objective is to minimize the number of iterations while keeping the communication structure of the distributed controller. Thus, the E matrix is restricted to accompany this request. All simulations are performed in MATLAB and the semidefinite program for the preconditioning is solved through YALMIP [7] using SeDuMi [15].

In Table I we compare the number of iterations needed to achieve a prespecified dual accuracy for the randomly generated DMPC problem for different preconditionings. We also compare the actual number of iterations with the iteration bounds to evaluate the conservatism of the bounds. The first column specifies the relative duality tolerance and the second column specifies the set from which the initial conditions are chosen where β is the scaling factor, i.e., initial conditions are chosen from $\beta\mathbb{X}_N$. The third column specifies which preconditioning that is used, DMPC for DMPC preconditioning and MPC for MPC preconditioning, or if no preconditioning is used. The fourth and fifth columns present average and max number of iterations while the sixth column presents the iteration bound. The data in Table I is obtained by solving the DMPC optimization problem for 10000 randomly generated initial conditions.

In Table I we see that, although the iteration bounds can be crude, the MPC preconditioning reduces the number of iterations needed while keeping the number of flops per iteration constant, and the DMPC preconditioning reduces the number of iterations significantly while keeping the communication structure of the distributed controller intact.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we present a method to precondition the optimization data for MPC and DMPC optimization problems when the dual to the optimization problem is solved using a fast gradient method. The preconditioning relies on minimizing an explicit iteration bound to achieve a prespecified dual accuracy. Although the bounds can be crude, numerical examples suggest that the number of iterations can be reduced significantly by using the proposed method.

REFERENCES

[1] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, October 2009.

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

[3] O. Devolder, F. Glineur, and Y. Nesterov. Double smoothing technique for convex optimization problems in Hilbert spaces with applications to optimal control. Technical report, Center for Operations Research and Econometrics (CORE), Universit catholique de Louvain (UCL), Belgium, January 2011.

[4] P. Giselsson. Execution time certification for gradient-based optimization in model predictive control. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 3165–3170, Maui, HI, December 2012.

[5] P. Giselsson, M. D. Doan, T. Keviczky, B. De Schutter, and A. Rantzer. Accelerated gradient methods and dual decomposition in distributed model predictive control. *Automatica*, 49(3):829–833, 2013.

[6] C.N. Jones and M. Morari. Approximate Explicit MPC using Bilevel Optimization. In *European Control Conference*, Budapest, Hungary, August 2009.

[7] J. Löfberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[8] A. Nedic and A.E. Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Optimization*, 19(4):1757–1780, 2009.

[9] Y. Nesterov. A method of solving a convex programming problem with convergence rate O (1/k^2). *Soviet Mathematics Doklady*, 27(2):372–376, 1983.

[10] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course (Applied Optimization)*. Springer Netherlands, 1st edition, 2003.

[11] P. Patrinos and A. Bemporad. An accelerated dual gradient-projection algorithm for linear model predictive control. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 662–667, Maui, HI, December 2012.

[12] J.B. Rawlings, D. Bonne, J.B. Jorgensen, A.N. Venkat, and S.B. Jorgensen. Unreachable setpoints in model predictive control. *IEEE Transactions on Automatic Control*, 53(9):2209–2215, October 2008.

[13] S. Richter, C.N. Jones, and M. Morari. Real-time input-constrained MPC using fast gradient methods. In *Proceedings of the 48th Conference on Decision and Control*, pages 7387–7393, Shanghai, China, December 2009.

[14] S. Richter, C.N. 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*, 57(6):1391–1403, 2012.

[15] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11–12:625–653, 1999.

[16] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. Technical report, Department of Mathematics, University of Washington, 2008.