Preconditioned Krylov iterations and condensing in interior point MPC method

Malyshev, A.; Quirynen, R.; Knyazev, A.

TR2018-121 August 25, 2018

Abstract

We investigate using Krylov subspace iterative methods in model predictive control (MPC), where the prediction model is given by linear or linearized systems with linear inequality constraints on the state and the input, and the performance index is quadratic. The inequality constraints are treated by the primal-dual interior point method. We indicate condition numbers of several linear systems, which determine the search direction in the Newton method, and propose a new preconditioner for one of the systems. Numerical results illustrate convergence of Krylov methods with and without preconditioning and demonstrate that our preconditioning reduces the number of Krylov iterations 2-10 times.

IFAC Conference on Nonlinear Model Predictive Control

Preconditioned Krylov iterations and condensing in interior point MPC method

Alexander Malyshev * **Rien Quirynen** ** **Andrew Knyazev** **

* *University of Bergen, Department of Mathematics, Postbox 8700, 5020 Bergen, Norway. (e-mail: alexander.malyshev@uib.no)*
** *Mitsubishi Electric Research Laboratories, 201 Broadway, 8th floor, Cambridge, MA 02139, USA. (e-mails: quirynen@merl.com, knyazev@merl.com)*

Abstract: We investigate using Krylov subspace iterative methods in model predictive control (MPC), where the prediction model is given by linear or linearized systems with linear inequality constraints on the state and the input, and the performance index is quadratic. The inequality constraints are treated by the primal-dual interior point method. We indicate condition numbers of several linear systems, which determine the search direction in the Newton method, and propose a new preconditioner for one of the systems. Numerical results illustrate convergence of Krylov methods with and without preconditioning and demonstrate that our preconditioning reduces the number of Krylov iterations 2–10 times.

Keywords: Model predictive control, Preconditioning, Numerical algorithms.

1. INTRODUCTION

We assess several numerical algorithms for solving optimal control problems in a model predictive control (MPC) approach. The design of MPC is discussed, e.g., in (Rawlings and Mayne, 2013). The survey (Diehl et al., 2009) gives a comprehensive view of existing numerical techniques based on Newton-type methods for solving the MPC problems and proposes various combinations of these techniques to design robust and efficient numerical procedures. Among hundreds of publications about numerical issues and solutions for MPC, we particularly refer to the prior work in (Rao et al., 1998; Ohtsuka, 2004; Shimizu et al., 2009; Zavala and Biegler, 2009; Wang and Boyd, 2010; Shahzad et al., 2010, 2012; Freund and Jarre, 1996).

We focus on a linear or linearized dynamic system whose state and control input satisfy linear inequality constraints and the performance index is given by a quadratic form, i.e., a standard quadratic programming formulation; cf. (Wang and Boyd, 2010; Shahzad et al., 2012). The inequality constraints are treated by means of a primal-dual interior point method (IPM) (Rao et al., 1998; Zavala and Biegler, 2009; Wright, 1997; Boyd and Vandenberghe, 2004; Gondzio, 2012).

The main computationally intensive part of the interior point method for solving the control prediction problem over a finite horizon, corresponds to solving a structured system of linear equations to compute the Newton step. This system is solved either exactly by a direct method or approximately using iterative techniques. The most efficient direct methods take advantage of the block banded structure of the Hessian under suitable orderings of variables. If we denote by n, m, and N the dimensions of the state variable and control input, and the horizon length, respectively, then the arithmetic complexity of a structured direct method equals $O\left(N(n+m)^3\right)$.

When the dimensions of the state, n, and of the control input, m, are large, say $(n+m) \gg 100$, the cubic dependence on n and m of such direct methods makes the numerical solution too expensive, especially in case of a real-time implementation of MPC under tight timing constraints. To overcome the cubic complexity, one can use Krylov subspace iterative methods (Greenbaum, 1997), which, in theory, reduce the asymptotic complexity to a quadratic dependence on both n and m. Such an approach is developed, e.g., in (Ohtsuka, 2004; Shahzad et al., 2010, 2012; Freund and Jarre, 1996; Gould et al., 2001; Dollar, 2005; Cafieri, 2006).

In our previous papers (Knyazev et al., 2015; Knyazev and Malyshev, 2016), we have further developed Ohtsuka's method (Ohtsuka, 2004), which has been designed for online numerical solution of nonlinear MPC problems. The method uses GMRES iterations (Kelly, 1995) to solve a linear system to compute the Newton-type continuation step, where the system is derived from the Karush-Kuhn-Tucker (KKT) equations by eliminating the state and costate variables. We have proposed a preconditioner for the GMRES iterations in (Knyazev and Malyshev, 2016), which is efficient for problems with inequality constraints on the control input. In the present paper, we consider the general framework of quadratic programming, similar to the work in (Wang and Boyd, 2010; Shahzad et al., 2010, 2012). A similarly structured quadratic program (QP) forms the subproblem within a sequential quadratic programming method for nonlinear MPC (Diehl et al., 2009).

The goal of our work is to design preconditioning techniques for the efficient use of Krylov subspace iterations during calculation of the Newton search direction step in the primal-dual interior point method. In particular, we consider several size reductions of the KKT system and indicate the resulting condition number; cf. (Greif et al., 2014). We complement the investigation started in (Shahzad et al., 2010, 2012) with a condensing procedure that is related to the so-called sequential method reviewed in (Diehl et al., 2009) and used, e.g., in (Ohtsuka, 2004). We propose a sparse preconditioner for the condensed problem, similar to (Knyazev and Malyshev, 2016), and we suggest an alternative preconditioner for the additional size reduction that was proposed in (Shahzad et al., 2010, 2012).

Our main contribution consists in a recommendation to use the size reduction by condensing, without actually forming the condensed QP formulation, and in combination with the new sparse preconditioner that treats inequality constraints for the control input as well as for the state variables. Further reduction to the δ-active inequality constraints, described in detail in (Shahzad et al., 2010, 2012), can be desirable; see similar approach in (Jung et al., 2012). Suitable Krylov subspace methods for the condensing approach include GMRES, QMR and BiCG. The main motivation for our recommendation is that the reduction from (Shahzad et al., 2010, 2012) is too expensive for large problems since the computational cost of their block elimination equals to that of a control problem without inequality constraints, i.e., $O\left(N(n+m)^3\right)$.

In the last section, numerical experiments in MATLAB illustrate convergence of the IPM. We show the number of the inner Krylov iterations with and without preconditioning and demonstrate that our preconditioning reduces the number of Krylov iterations 2–10 times. Note that our tests use small matrices, where the fastest direct methods perform better in terms of CPU time.

2. MPC PROBLEM FORMULATION

Model predictive control (MPC) determines the current control input $u(t)$ by solving an optimal control problem on a finite horizon $t = \tau_0 < \tau_1 < \cdots < \tau_N = t + T$. Over the finite horizon of length N, we consider a prediction model given by the discrete-time linear system

$$x_{i+1} = A_i x_i + B_i u_i, \quad i = 0, 1, \ldots, N-1, \tag{1}$$

where $A_i \in \mathbb{R}^{n\times n}$, $B_i \in \mathbb{R}^{n\times m}$, $x_i \in \mathbb{R}^n$ is the state vector, $u_i \in \mathbb{R}^m$ is the control input vector. Let $\bar{x} = x_0 \in \mathbb{R}^n$ be the measurement or estimate of the state at the current time instant. The objective is to find a sequence of optimal control inputs $u_0, \ldots, u_{N-1}$ subject to the equality constraints (1) and the inequality constraints

$$G_{x,i} x_i + G_{u,i} u_i \le g_i, \quad i = 0, \ldots, N-1, \tag{2}$$

$$G_{x,N} x_N \le g_N, \tag{3}$$

while minimizing the quadratic performance index

$$x_N^\top Q_N x_N + \sum_{i=0}^{N-1} \left(x_i^\top Q_i x_i + u_i^\top R_i u_i + 2u_i^\top M_i x_i \right), \tag{4}$$

where we define the matrices $G_{x,i} \in \mathbb{R}^{l_i\times n}$, $G_{u,i} \in \mathbb{R}^{l_i\times m}$, $Q_i \in \mathbb{R}^{n\times n}$, $R_i \in \mathbb{R}^{m\times m}$ and $M_i \in \mathbb{R}^{m\times n}$.

The optimal control problem (OCP) is the sparse convex quadratic program (QP)

$$\min_d \frac{1}{2} d^\top H d + d^\top h \quad \text{subject to } Fd = f(\bar{x}),\ Gd \le g, \tag{5}$$

with respect to the vector of decision variables

$$d = \left[x_0^\top\ u_0^\top\ x_1^\top\ u_1^\top\ \cdots\ u_{N-1}^\top\ x_N^\top\right]^\top, \tag{6}$$

where $d, h \in \mathbb{R}^{n_d}$, $H \in \mathbb{R}^{n_d\times n_d}$, $F \in \mathbb{R}^{n_e\times n_d}$, $G \in \mathbb{R}^{n_s\times n_d}$ with $n_d = (n+m)N+n$, $n_e = n(N+1)$, and $n_s = \sum_{i=0}^N l_i$. The sparse matrices H, F, and G and the vectors h, $f(\bar{x})$ and g are defined as follows:

$$H = \begin{bmatrix} Q_0 & M_0^\top & \cdots & 0 & 0 & 0 \\ M_0 & R_0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & Q_{N-1} & M_{N-1}^\top & 0 \\ 0 & 0 & \cdots & M_{N-1} & R_{N-1} & 0 \\ 0 & 0 & \cdots & 0 & 0 & Q_N \end{bmatrix},$$

$$F = \begin{bmatrix} -I & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ A_0 & B_0 & -I & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & A_{N-1} & B_{N-1} & -I \end{bmatrix},$$

$$G = \begin{bmatrix} G_{x,0} & G_{u,0} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & G_{x,N-1} & G_{u,N-1} & 0 \\ 0 & 0 & \cdots & 0 & 0 & G_{x,N} \end{bmatrix},$$

$$f(\bar{x}) = \left[-\bar{x}^\top\ 0\ \ldots\ 0\right]^\top, \quad g = \left[g_0^\top\ g_1^\top\ \ldots\ g_N^\top\right]^\top.$$

Note that the gradient vector is assumed to be equal to zero, $h = 0$, for simplicity of notation in (4).

3. INTERIOR POINT METHOD (IPM)

Necessary optimality conditions for the OCP are derived by means of the Lagrangian function

$$\mathcal{L} = \frac{1}{2} x_N^\top Q_N x_N + \frac{1}{2}\sum_{i=0}^{N-1} \left(x_i^\top Q_i x_i + u_i^\top R_i u_i + 2u_i^\top M_i x_i \right)$$

$$+ y_0^\top(\bar{x} - x_0) + \sum_{i=0}^{N-1} y_{i+1}^\top (A_i x_i + B_i u_i - x_{i+1})$$

$$- \mu \sum_{i=0}^{N} \mathbf{1}^\top \log s_i + \sum_{i=0}^{N} z_i^\top (G_{x,i} x_i + G_{u,i} u_i - g_i + s_i),$$

where $\mathbf{1} = [1 \ldots 1]^\top$, and the slack variables $s_i > 0$ and small parameter $\mu > 0$ are required by the interior point method. The optimality conditions, also known as the KKT conditions, then read as follows:

$$\begin{aligned} &Hd + F^\top y + G^\top z = 0, \quad Fd - f = 0, \\ &Gd - g + s = 0, \qquad Z\mathbf{1} - \mu S^{-1}\mathbf{1} = 0, \end{aligned} \tag{7}$$

where the matrices $Z = \operatorname{diag}(z)$ and $S = \operatorname{diag}(s)$ are diagonal. The last KKT condition is nonlinear and usually substituted with the equation $ZS\mathbf{1} - \mu\mathbf{1} = 0$, which is better scaled for solving the system in (7).

A search direction at iteration k of Newton's method is determined by the following system of linear equations:

$$\underbrace{\begin{bmatrix} H & F^\top & G^\top & 0 \\ F & 0 & 0 & 0 \\ G & 0 & 0 & I \\ 0 & 0 & S^k & Z^k \end{bmatrix}}_{\mathcal{A}_0^k} \begin{bmatrix} \Delta d^k \\ \Delta y^k \\ \Delta z^k \\ \Delta s^k \end{bmatrix} = - \underbrace{\begin{bmatrix} r_H^k \\ r_F^k \\ r_G^k \\ r_S^k \end{bmatrix}}_{b_0^k} \tag{8}$$

with the residuals

$$r_H^k = Hd^k + F^\top y^k + G^\top z^k, \tag{9}$$
$$r_F^k = Fd^k - f, \tag{10}$$
$$r_G^k = Gd^k - g + s^k, \tag{11}$$
$$r_S^k = Z^k S^k \mathbf{1} - \sigma\mu^k \mathbf{1}, \tag{12}$$

where $\sigma \in (0,1)$ is called a centering parameter, and the value for $\mu^k = (z^k)^\top s^k / n_s$ is directly related to the current duality gap; see (Shahzad et al., 2012; Wright, 1997) for more detail. We solve the KKT system (8) at each iteration of the interior point method (IPM) implemented in Algorithm 1 from (Shahzad et al., 2012).

4. SIZE REDUCTION BY GAUSSIAN ELIMINATIONS

In this section, several KKT systems are introduced based on Schur complements of the matrix in (8), and their condition numbers are indicated near convergence of the IPM with the tolerance $\epsilon = 10^{-3}$ for the test example in Section 9. The initial KKT matrix $\mathcal{A}_0^k$ has a condition number at iteration $k = 11$ of $\mathrm{cond}(\mathcal{A}_0^{11}) = 2.8 \cdot 10^6$.

The system (8) can be reduced to a system of smaller size by block Gaussian elimination, resulting in

$$\underbrace{\begin{bmatrix} H & F^\top & G^\top \\ F & 0 & 0 \\ G & 0 & -W^k \end{bmatrix}}_{\mathcal{A}_1^k} \begin{bmatrix} \Delta d^k \\ \Delta y^k \\ \Delta z^k \end{bmatrix} = -\underbrace{\begin{bmatrix} r_H^k \\ r_F^k \\ r_W^k \end{bmatrix}}_{b_1^k}, \tag{13}$$

$$\Delta s^k = -(Z^k)^{-1}\left(r_S^k + S^k \Delta z^k\right), \tag{14}$$

with the diagonal matrix $W^k = (Z^k)^{-1} S^k$ and the residual $r_W^k = r_G^k - (Z^k)^{-1} r_S^k$. The condition number reads $\mathrm{cond}(\mathcal{A}_1^{11}) = 4.1 \cdot 10^{10}$ for our test example.

Another block Gaussian elimination further reduces the number of unknowns in the system (13):

$$\underbrace{\begin{bmatrix} H + G^\top (W^k)^{-1} G & F^\top \\ F & 0 \end{bmatrix}}_{\mathcal{A}_2^k} \begin{bmatrix} \Delta d^k \\ \Delta y^k \end{bmatrix} = -\underbrace{\begin{bmatrix} r_E^k \\ r_F^k \end{bmatrix}}_{b_2^k}, \tag{15}$$

$$\Delta z^k = (W^k)^{-1}\left(G \Delta d^k + r_W^k\right), \tag{16}$$

where $r_E^k = r_H^k + G^\top (W^k)^{-1} r_W^k$, $\mathrm{cond}(\mathcal{A}_2^{11}) = 3.3 \cdot 10^{10}$.

The three linear systems in (8), (13), and (15) can be transformed into systems with banded matrices by a localized reordering of unknowns. The resulting banded systems are usually solved by Gaussian elimination in order to avoid numerical issues, e.g., caused by large diagonal elements in W^k. The arithmetic complexity of such direct solution methods typically amounts to $O(N(n+m)^3)$ operations.

If the block diagonal matrix $D^k = H + G^\top (W^k)^{-1} G$ is nonsingular, e.g., when H is positive definite, then the linear system in (15) reduces to the block tridiagonal system with respect to Δy^k (cf. (Wang and Boyd, 2010)):

$$\left[F (D^k)^{-1} F^\top\right] \Delta y^k = r_F^k - F (D^k)^{-1} r_E^k, \tag{17}$$
$$\Delta d^k = -(D^k)^{-1}\left(F^\top \Delta y^k + r_E^k\right).$$

For our test example in Section 9, $\mathrm{cond}(D^k) = \infty$ such that the system in (17) can not be formed.

The authors in (Shahzad et al., 2010, 2012) instead favor the following block elimination from Eq. (13):

$$\underbrace{\left([G\ 0] \begin{bmatrix} H & F^\top \\ F & 0 \end{bmatrix}^{-1} \begin{bmatrix} G^\top \\ 0 \end{bmatrix} + W^k \right)}_{\mathcal{A}_3^k} \Delta z^k$$
$$= \underbrace{r_W^k - [G\ 0] \begin{bmatrix} H & F^\top \\ F & 0 \end{bmatrix}^{-1} \begin{bmatrix} r_H^k \\ r_F^k \end{bmatrix}}_{b_3^k}, \tag{18}$$

$$\begin{bmatrix} \Delta d^k \\ \Delta y^k \end{bmatrix} = -\begin{bmatrix} H & F^\top \\ F & 0 \end{bmatrix}^{-1} \left(\begin{bmatrix} G^\top \\ 0 \end{bmatrix} \Delta z^k + \begin{bmatrix} r_H^k \\ r_F^k \end{bmatrix} \right).$$

However, the pivot matrix $\Pi = \begin{bmatrix} H & F^\top \\ F & 0 \end{bmatrix}$ is nonsingular if and only if the linear system (1) with the performance index (4) is consistent, i.e., controllable and observable without inequality constraints. Otherwise, the block elimination method in (18) does not work. We have $\mathrm{cond}(\Pi) = 8 \cdot 10^5$ and $\mathrm{cond}(\mathcal{A}_3^{11}) = 9.2 \cdot 10^{10}$ in our test example.

4.1 Numerical Condensing Procedure

Let us propose an alternative block elimination approach based on the numerical condensing of the state variables (Frison, 2015). We first introduce the square block bidiagonal matrix

$$\mathfrak{A} = \begin{bmatrix} -I & & & & \\ A_0 & -I & & & \\ & A_1 & -I & & \\ & & \ddots & \ddots & \\ & & & A_{N-1} & -I \end{bmatrix}, \tag{19}$$

and the block diagonal matrices

$$\mathfrak{B} = \mathrm{blkdiag}(0, B_0, B_1, \ldots, B_{N-1}),$$
$$\mathfrak{Q} = \mathrm{blkdiag}(Q_0, Q_1, \ldots, Q_N),$$
$$\mathfrak{R} = \mathrm{blkdiag}(R_0, R_1, \ldots, R_{N-1}),$$
$$\mathfrak{M} = \mathrm{blkdiag}(M_0, M_1, \ldots, M_{N-1}),$$
$$\mathfrak{G}_x = \mathrm{blkdiag}(G_{x,0}, G_{x,1}, \ldots, G_{x,N}),$$
$$\mathfrak{G}_u = \mathrm{blkdiag}(G_{u,0}, G_{u,1}, \ldots, G_{u,N-1}).$$

We split the decision vector d into two vectors $x = \begin{bmatrix} x_0^\top & x_1^\top & \cdots & x_N^\top \end{bmatrix}^\top$ and $u = \begin{bmatrix} u_0^\top & u_1^\top & \cdots & u_{N-1}^\top \end{bmatrix}^\top$ and rewrite (8), permuting its rows and columns, as

$$\begin{bmatrix} \mathfrak{Q} & \mathfrak{A}^\top & \mathfrak{M}^\top & \mathfrak{G}_x^\top & 0 \\ \mathfrak{A} & 0 & \mathfrak{B} & 0 & 0 \\ \mathfrak{M} & \mathfrak{B}^\top & \mathfrak{R} & \mathfrak{G}_u^\top & 0 \\ \mathfrak{G}_x & 0 & \mathfrak{G}_u & 0 & I \\ 0 & 0 & 0 & S^k & Z^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta u^k \\ \Delta z^k \\ \Delta s^k \end{bmatrix} = -\begin{bmatrix} r_x^k \\ r_F^k \\ r_u^k \\ r_G^k \\ r_S^k \end{bmatrix}. \tag{20}$$

As in (13), elimination of Δs^k from (20) gives

$$\begin{bmatrix} \mathfrak{Q} & \mathfrak{A}^\top & \mathfrak{M}^\top & \mathfrak{G}_x^\top \\ \mathfrak{A} & 0 & \mathfrak{B} & 0 \\ \mathfrak{M} & \mathfrak{B}^\top & \mathfrak{R} & \mathfrak{G}_u^\top \\ \mathfrak{G}_x & 0 & \mathfrak{G}_u & -W^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \\ \Delta u^k \\ \Delta z^k \end{bmatrix} = -\begin{bmatrix} r_x^k \\ r_F^k \\ r_u^k \\ r_W^k \end{bmatrix}, \tag{21}$$

where $r_W^k = r_G^k - (Z^k)^{-1} r_S^k$. The matrix $\mathfrak{A}$ is always nonsingular, and the inverse of the pivot equals

$$\mathfrak{P} = \begin{bmatrix} \mathfrak{Q} & \mathfrak{A}^\top \\ \mathfrak{A} & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & \mathfrak{A}^{-1} \\ \mathfrak{A}^{-T} & -\mathfrak{A}^{-T}\mathfrak{Q}\mathfrak{A}^{-1} \end{bmatrix}.$$

After elimination of the variables Δx^k and Δy^k from (21), we denote the blocks of the Schur complement as

$$\mathfrak{S}_{11} = \mathfrak{R} + \mathfrak{B}^\top \mathfrak{A}^{-T} \mathfrak{Q} \mathfrak{A}^{-1} \mathfrak{B} - \mathfrak{M} \mathfrak{A}^{-1} \mathfrak{B} - \mathfrak{B}^\top \mathfrak{A}^{-T} \mathfrak{M}^\top,$$

$$\mathfrak{G}_{21} = \mathfrak{G}_u - \mathfrak{G}_x \mathfrak{A}^{-1}\mathfrak{B},$$

and arrive at the symmetric system

$$\underbrace{\begin{bmatrix} \mathfrak{G}_{11} & \mathfrak{G}_{21}^\top \\ \mathfrak{G}_{21} & -W^k \end{bmatrix}}_{\mathcal{A}_4^k} \begin{bmatrix} \Delta u^k \\ \Delta z^k \end{bmatrix} = - \underbrace{\begin{bmatrix} r_{\mathfrak{G}}^k \\ r_{\mathfrak{W}}^k \end{bmatrix}}_{b_4^k}, \tag{22}$$

where the residuals read $r_{\mathfrak{G}}^k = r_u^k - \left[\mathfrak{M} \ \mathfrak{B}^\top\right] \mathfrak{P} \begin{bmatrix} r_x^k \\ r_F^k \end{bmatrix}$ and $r_{\mathfrak{W}}^k = r_W^k - [\mathfrak{G}_x \ 0] \mathfrak{P} \begin{bmatrix} r_x^k \\ r_F^k \end{bmatrix}$. The corresponding condition numbers are $\text{cond}(\mathfrak{P}) = 2.2 \cdot 10^3$ and $\text{cond}(\mathcal{A}_4^{11}) = 2 \cdot 10^{10}$.

Eliminating Δx^k and Δy^k from (20), but keeping Δs^k, results in the nonsymmetric system

$$\underbrace{\begin{bmatrix} \mathfrak{G}_{11} & \mathfrak{G}_{21}^\top & 0 \\ \mathfrak{G}_{21} & 0 & I \\ 0 & S^k & Z^k \end{bmatrix}}_{\mathcal{A}_5^k} \begin{bmatrix} \Delta u^k \\ \Delta z^k \\ \Delta s^k \end{bmatrix} = - \underbrace{\begin{bmatrix} r_{\mathfrak{G}}^k \\ r_{\mathfrak{G}}^k \\ r_S^k \end{bmatrix}}_{b_5^k}, \tag{23}$$

where $r_{\mathfrak{G}}^k = r_G^k - [\mathfrak{G}_x \ 0] \mathfrak{P} \begin{bmatrix} r_x^k \\ r_F^k \end{bmatrix}$. For our test example, the condition number is equal to $\text{cond}(\mathcal{A}_5^{11}) = 1.6 \cdot 10^7$, which is lower than the condition number for the system matrices $\mathcal{A}_4$, $\mathcal{A}_3$, $\mathcal{A}_2$ and $\mathcal{A}_1$, and relatively comparable to the initial KKT matrix $\text{cond}(\mathcal{A}_0^{11}) = 2.8 \cdot 10^6$.

5. SIZE REDUCTION BY CHOOSING δ-ACTIVE INEQUALITY CONSTRAINTS

The authors of (Shahzad et al., 2010, 2012) present a simple technique that reduces the size of the problem and it can improve the condition number, by introducing the concept of a δ-active set of inequality constraints for a given scalar $\delta > 0$. The δ-active set at iteration k is the set of indices $\mathcal{N}_A^k(\delta) = \{i \in \mathcal{N} | 0 < w_i^k \leq \delta\}$, where $w_i^k = \frac{s_i^k}{z_i^k}$ and $\mathcal{N} := \{1, 2, \ldots, n_s\}$. The δ-inactive set at iteration k is the complement set $\mathcal{N}_I^k = \mathcal{N} \backslash \mathcal{N}_A^k(\delta)$. The value of δ is usually selected sufficiently large such that all inequality constraints in the first IPM iteration are δ-active, i.e., $\delta > \frac{s_i^0}{z_i^0}$ for all indices i.

Let us illustrate the idea on the formulation of the linear KKT system in (13). Permuting and splitting the variables Δz^k according to the δ-active and δ-inactive constraints yields the following block partitioning of (13),

$$\begin{bmatrix} M_1^k & (M_2^k)^\top \\ M_2^k & -W_2^k \end{bmatrix} \begin{bmatrix} p_1^k \\ p_2^k \end{bmatrix} = \begin{bmatrix} r_1^k \\ r_2^k \end{bmatrix}, \tag{24}$$

where

$$M_1^k = \begin{bmatrix} H & F^\top & (G_1^k)^\top \\ F & 0 & 0 \\ G_1^k & 0 & -W_1^k \end{bmatrix}, \quad M_2^k = \begin{bmatrix} G_2^k & 0 & 0 \end{bmatrix},$$

$$p_1^k = \begin{bmatrix} \Delta d^k \\ \Delta y^k \\ \Delta z_1^k \end{bmatrix}, \quad r_1^k = \begin{bmatrix} r_H^k \\ r_F^k \\ r_{W_1}^k \end{bmatrix}, \quad p_2^k = \Delta z_2^k, \quad r_2^k = r_{W_2}^k.$$

The diagonal matrix W_1^k corresponds to the δ-active set, and the diagonal matrix W_2^k to the δ-inactive part.

The size reduction by means of the δ-active set is done by replacing the block $(M_2^k)^\top$ by the zero matrix, i.e., instead of the system (24), we solve the system of two equations

$$M_1^k \hat{p}_1^k = r_1^k, \tag{25}$$

$$W_2^k \hat{p}_2^k = M_2^k \hat{p}_1^k - r_2^k. \tag{26}$$

The main computational burden corresponds to solving the linear KKT system in (25). Since the matrix W_2^k is diagonal, solving system (26) is cheap. Both numerical experience and theoretical arguments from (Shahzad et al., 2012) show that the use of the solution $(\hat{p}_1^k, \hat{p}_2^k)$ instead of (p_1^k, p_2^k) provides a good performance of the IPM, but maybe at the cost of a few extra iterations of Newton's method. It is straightforward to apply this size reduction technique, by means of the δ-active set of inequality constraints, to all other formulations of the linear KKT system in (8), (15), (17), (18), (22) and (23).

6. REMARKS ON ITERATIVE SOLUTION OF THE IPM SYSTEMS OF LINEAR EQUATIONS

Our main goal is to investigate the considered numerical methods on suitability for problems of large size, e.g., to cope with the case when the matrices A_i, B_i, Q_i, R_i, M_i, and G_i are large and sparse. Specifically, we aim to reduce the problem size and lower condition numbers of the resulting systems of linear equations, to which the Krylov subspace methods are applied after size reduction.

From several size reductions introduced in the previous sections, we propose to consider the following variants:

(a) system (8) with/out δ-active inequality constraints;
(b) system (13) with δ-active inequality constraints;
(c) system (18) with δ-active inequality constraints;
(d) system (22) with δ-active inequality constraints;
(e) system (23) with/out δ-active inequality constraints.

The papers (Shahzad et al., 2010, 2012) propose two preconditioners for the variant c), where the matrix $\mathcal{A}_3^k$ exists only if the pivot block $\Pi = \begin{bmatrix} H & F^\top \\ F & 0 \end{bmatrix}$ is invertible. In this case, $\mathcal{A}_3^k$ is symmetric positive definite and dense. This structure allows applying MINRES or CG iterations for solving the system (18).

Our contribution concerns the variants d) and e), which are always applicable for the quadratic program formulated in Section II. These variants can be particularly useful when the pivot matrix $\Pi = \begin{bmatrix} H & F^\top \\ F & 0 \end{bmatrix}$ is singular, or ill-conditioned, which makes the methods from (Shahzad et al., 2010, 2012) difficult or impossible to use. Suitable Krylov subspace methods for the variants d) and e) include GMRES, QMR and BiCG, with optional preconditioning as discussed in the next section. It is worth to remark that the block Gaussian elimination with the pivot matrix Π, in general, costs $O(N(n+m)^3)$ arithmetic operations, while the numerical condensing in the variants d) and e) costs $O(N(n+m)^2)$ arithmetic operations.

Note that the latter complexity is very different from the typical result of $O\left(N^2(nm^2 + n^2m)\right)$ or $O\left(N^2(n+m)^3\right)$ arithmetic operations to form the condensed Hessian $\mathfrak{G}_{11}$ as in (Frison, 2015). We namely require only $O(N(n+m)^2)$ arithmetic operations to perform a matrix-vector multiplication in an iterative solver for system (22) or (23), without explicitly constructing the matrix directly.

7. PRECONDITIONING TECHNIQUES

In this section, we propose a new preconditioner for the matrix $\mathcal{A}_4^k$ in (22). We recall that the matrix $\mathcal{A}_4^k$ is not well-conditioned, in particular, owing to large diagonals in W^k. The condition number can be sometimes improved by the size reduction with the δ-active inequality constraints. Therefore, we can assume that $\mathcal{A}_4^k$ is given after this reduction.

Let us look at the blocks of the Schur complement

$$\mathfrak{S}_{11} = \mathfrak{R} + \mathfrak{B}^\top \mathfrak{A}^{-T} \mathfrak{Q} \mathfrak{A}^{-1} \mathfrak{B} - \mathfrak{M} \mathfrak{A}^{-1} \mathfrak{B} - \mathfrak{B}^\top \mathfrak{A}^{-T} \mathfrak{M}^\top, \tag{27}$$

$$\mathfrak{S}_{21} = \mathfrak{G}_u - \mathfrak{G}_x \mathfrak{A}^{-1} \mathfrak{B}, \tag{28}$$

as a result of the condensing procedure. Our preconditioner is obtained by setting $A_i = 0$ for all indices $i = 0, 1, \ldots, N-1$, i.e., $\mathfrak{A} = -I$, where the block bidiagonal matrix $\mathfrak{A}$ is given in (19). In other words, we attempt to approximate $\mathfrak{S}_{11}$ by the block diagonal matrix

$$\mathcal{P}_{11} = \mathfrak{R} + \mathfrak{B}^\top \mathfrak{Q} \mathfrak{B} + \mathfrak{M} \mathfrak{B} + \mathfrak{B}^\top \mathfrak{M}^\top$$

and $\mathfrak{S}_{21}$ by the block diagonal matrix

$$\mathcal{P}_{21} = \mathfrak{G}_u + \mathfrak{G}_x \mathfrak{B}.$$

Hence, the preconditioner for the matrices $\mathcal{A}_4^k$ is given by the block partitioned matrix

$$\mathcal{P}^k = \begin{bmatrix} \mathcal{P}_{11} & \mathcal{P}_{21}^\top \\ \mathcal{P}_{21} & -W^k \end{bmatrix}, \tag{29}$$

with the sparse block diagonal submatrices $\mathcal{P}_{11}$ and $\mathcal{P}_{21}$ and the diagonal matrix W^k. An alternative preconditioner can be constructed by setting all A_i to the identity matrix. However, such a preconditioner would be less sparse than (29).

Remark. As mentioned earlier, the papers (Shahzad et al., 2010, 2012) propose two preconditioners for solving (18) iteratively, where the matrix undergoes the elimination of the δ-inactive inequality constraints. One of the preconditioners is the diagonal matrix W_1^k, the other one is the block diagonal of the matrix $\mathcal{A}_3^k$. We suggest another block diagonal preconditioner for (18),

$$\mathcal{P}_3^k = W_1^k + \mathfrak{G}_u^k \mathfrak{R}^{-1} (\mathfrak{G}_u^k)^\top,$$

which is constructed analogously to the preconditioner (29). Testing $\mathcal{P}_3^k$ is beyond the scope of the present paper.

8. TEST EXAMPLE: OSCILLATING MASSES

We consider a linear system of $n/2$ unit masses connected by springs and to walls at the ends. The stiffness of each spring equals 1, and there is no damping. There are m actuators connected to the first m masses. The vectors of the state, control and output are respectively as

$$x = \left[q_1^\top \; q_2^\top \; \ldots \; q_{n/2}^\top \; \dot{q}_1^\top \; \dot{q}_2^\top \; \ldots \; \dot{q}_{n/2}^\top \right]^\top,$$

$$u = \left[f_1^\top \; f_2^\top \; \ldots \; f_m^\top \right]^\top, \quad y = \left[q_1^\top \; q_2^\top \; \ldots \; q_{n/2}^\top \right]^\top,$$

where q_i denotes the coordinate of the ith mass with respect to its equilibrium position and f_i represents the control force acting on the ith mass. The following inequality constraints are imposed on the inputs and outputs:

$$-0.5 \le u(i) \le 0.5, \qquad i = 0, \ldots, N-1,$$
$$-3.5 \le y(i) \le 3.5, \qquad i = 1, \ldots, N.$$

The corresponding linear discrete-time system is sampled for the time rate $\Delta\tau = 0.5$, the inputs are constant between sample instants. The following matrices are used in the discrete-time model:

$$A_i = \exp(\Delta\tau A_c), \quad B_i = A_c^{-1}(A_i - I_n) \begin{bmatrix} 0_{n/2} \\ I_m \\ 0_{n/2-m} \end{bmatrix},$$

where

$$A_c = \begin{bmatrix} 0_{n/2} & I_{n/2} \\ T_{n/2} & 0_{n/2} \end{bmatrix}, \quad T = \begin{bmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \ddots \\ & & 1 & -2 \end{bmatrix}.$$

The objective of the control is to regulate the displacements with the given constraints on displacements and control inputs. We choose the following regulator tuning matrices $R_i = 10^{-6} I$, $M_i = 0$, and $Q_i = C^\top C = \left[I_{n/2} \; 0 \right]^\top \left[I_{n/2} \; 0 \right]$ and the following initial values

$$\bar{x} = 3.5 \left[1 \; 1 \; 0 \; \cdots \; 0 \right]^\top.$$

9. NUMERICAL EXPERIMENTS

We have carried out numerical experiments in MATLAB with the example from Section 8, where dimension of the state is $n = 12$, the number of control inputs is $m = 3$, the horizon length is $N = 30$. The number of inequality constraints at each time instance is $l = n + 2m = 18$. As the IPM method we choose Algorithm 1 from (Shahzad et al., 2012). Other constants are $\sigma = 0.1$, $\gamma = 10^{-3}$, $\beta = 2$, $\epsilon = 10^{-6}$. The absolute error tolerance for the Krylov iterations is $tol = \epsilon$. The initial values of x, u, y, z, s are chosen to be equal to $\mathbf{1}$.

Our MATLAB code implements the IPM using iterative solvers for the linear system in (22) with the preconditioner from (29) and the size reduction to the δ-active inequality constraints. The available iterative solvers are the MATLAB functions for GMRES, QMR and BiCG. We report only the results of solving MPC at the initial time and from a cold start, where the initial values equal to $\mathbf{1}$.

The example from Section 8 allows a simple MATLAB implementation of a direct solution method exploiting the particular banded structure of the properly permuted system (8) and the sparse Gaussian elimination of MATLAB. However, the direct solution method from (Wang and Boyd, 2010) fails because the matrix D^k in (17) is singular near the convergence of IPM. Our MATLAB implementation of the iterative method from (Shahzad et al., 2010, 2012), which solves (18) by CG, with $\delta = 1000$, converges to a solution after 17 IPM iterations and 19279 unpreconditioned CG iterations inside IPM, or 29689 CG iterations preconditioned with W_1^k.

Figure 1 displays the number of Krylov iterations without preconditioning. The δ-active inequality constraints for GMRES are selected with $\delta = 100$ and for QMR and BiCG with $\delta = 4$. We note that BiCG often requires less inner Krylov iterations, compared to GMRES and QMR, but ends up with many more IPM iterations, resulting in residual values $r_{GMRES} = 2e{-4}$, $r_{QMR} = 4e{-4}$, $r_{BiCG} = 4e{-4}$ at the last IPM step.

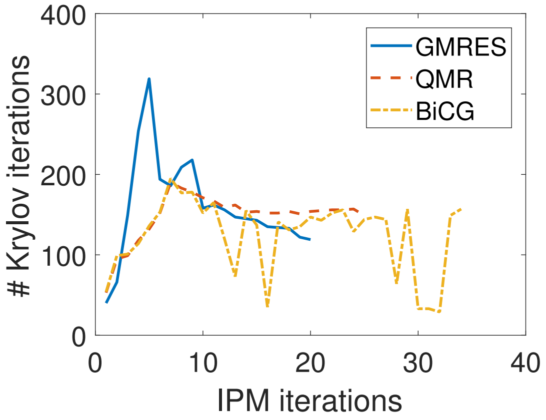

Fig. 1. Krylov iterations without preconditioning.

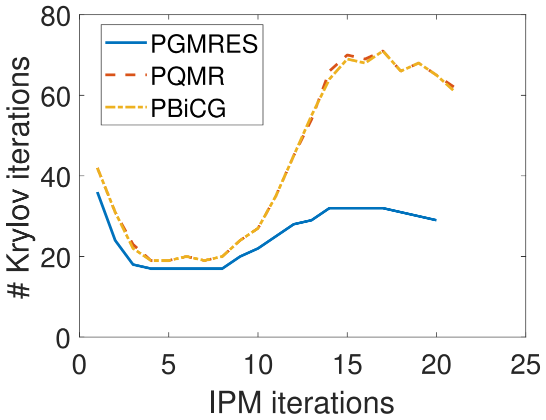

Fig. 2. Krylov iterations with proposed preconditioning.

Figure 2 shows similar results for GMRES, QMR and BiCG, but with preconditioning, where the δ-active inequality constraints are selected with $\delta = 200$. We observe that preconditioning accelerates GMRES by a factor of 5–10, compared to the results in Figure 1.

10. CONCLUSION

Our sparse preconditioners for Krylov methods applied to systems of linear equations that appear at each Newton step in an IPM for quadratic programs show speedups 5–10x in numerical experiments with preconditioned GMRES. The expected computational complexity is $O(N(n+m+l)^2)$, where N is the horizon length in the MPC prediction, n and m are dimensions of the state and the control input, and l is the number of inequality constraints.

REFERENCES

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press, Heidelberg, Germany.

Cafieri, S. (2006). *On the application of iterative solvers to KKT systems in interior point methods for large-scale quadratic programming problems.* Ph.D. thesis, Universitá degli Studi di Napoli Federico II.

Diehl, M., Ferreau, H.J., and Haverbeke, N. (2009). Efficient numerical methods for nonlinear MPC and moving horizon estimation. In L.M. et al. (ed.), *Nonlinear Model Predictive Control*, LNCIS 384, 391–417. Springer, Heidelberg, Germany.

Dollar, H. (2005). *Iterative linear algebra for constrained optimization.* Ph.D. thesis, University of Oxford.

Freund, R.W. and Jarre, F. (1996). A QMR-based interior-point algorithm for solving linear programs. *Math. Program.*, 76, 183–210.

Frison, G. (2015). *Algorithms and Methods for High-Performance Model Predictive Control.* Ph.D. thesis, Technical University of Denmark (DTU).

Gondzio, J. (2012). Interior point methods 25 years later. *European J. Operational Research*, 218, 587–601.

Gould, N.I.M., Hribar, M.E., and Nocedal, J. (2001). On the solution of equality constrained quadratic programming problems arising in optimization. *SIAM J. Sci. Comput.*, 23, 1376–1395.

Greenbaum, A. (1997). *Iterative methods for solving linear systems.* SIAM, Philadelphia, PA.

Greif, C., Moulding, E., and Orban, D. (2014). Bounds on eigenvalues of matrices arising from interior-point methods. *SIAM J. Optim.*, 24, 49–83.

Jung, J.H., O'Leary, D.P., and Tits, A.L. (2012). Adaptive constraint reduction for convex quadratic programming. *Comput. Optim. Appl.*, 51, 125–157.

Kelly, C.T. (1995). *Iterative methods for linear and nonlinear equations.* SIAM, Philadelphia, PA.

Knyazev, A., Fujii, Y., and Malyshev, A. (2015). Preconditioned continuation model predictive control. In *Proc. SIAM Conf. Control Appl.*, 101–108. SIAM, Paris, France.

Knyazev, A. and Malyshev, A. (2016). Sparse preconditioning for model predictive control. In *Proc. American Control Conf. (ACC)*, 4494–4499. Boston, MA, USA.

Ohtsuka, T. (2004). A continuation/GMRES method for fast computation of nonlinear receding horizon control. *Automatica*, 40, 563–574.

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

Rawlings, J.B. and Mayne, D.Q. (2013). *Model Predictive Control. Theory and Design.* Nob Hill Publishing, Madison, WI.

Shahzad, A., Kerrigan, E.C., and Constantinides, G.A. (2010). A fast well-conditionend interior point method for predictive control. In *Proc. IEEE Conf. Decision Control*, 508–513. Atlanta, GA, USA, December 15–17.

Shahzad, A., Kerrigan, E.C., and Constantinides, G.A. (2012). A stable and efficient method for solving a convex quadratic program with application to optimal control. *SIAM J. Optim.*, 22(4), 1369–1393.

Shimizu, Y., Ohtsuka, T., and Diehl, M. (2009). A real-time algorithm for nonlinear receding horizon control using multiple shooting and continuation/Krylov method. *Int. J. Robust Nonlinear Control*, 19, 919–936.

Wang, Y. and Boyd, S. (2010). Fast model predictive control using online optimization. *IEEE Trans. Control Syst. Technology*, 18, 267–278.

Wright, S.J. (1997). *Primal-Dual Interior Point Methods.* SIAM, Philadelphia.

Zavala, V.M. and Biegler, L.T. (2009). Nonlinear programming strategies for state estimation and model predictive control. In L.M. et al. (ed.), *Nonlinear Model Predictive Control*, LNCIS 384, 419–432. Springer, Heidelberg, Germany.