Quadratic programming with ramp functions and fast online QP-MPC solutions

Giorgio Valmorbida, Morten Hovd

Quadratic programming with ramp functions and fast online QP-MPC solutions ⋆

Giorgio Valmorbida [a] Morten Hovd [b]

[a] *Université Paris-Saclay, CNRS, CentraleSupélec, Inria, Laboratoire des Signaux et Systèmes, 91190, Gif-sur-Yvette, France.*

[b] *Department of Engineering Cybernetics, Norwegian University of Science & Technology, 7491 Trondheim, Norway.*

Abstract

A novel method is proposed for solving quadratic programming problems arising in model predictive control. The method is based on an implicit representation of the Karush-Kuhn-Tucker conditions using ramp functions. The method is shown to be highly efficient on both small and fairly large Quadratic Program problems, can be implemented using simple computer code, and has modest memory requirements.

Key words: PWA systems. Ramp functions. Model Predictive Control.

1 Introduction

Model Predictive Control (MPC) has been a great success in industry [10], and since its initial development in the 1970's (see, *e.g.* [12]) it has found application in a wide range of industrial processes. The main feature of MPC distinguishing it from classical control design methods such as Linear Quadratic Regulator is the ability to take into account constraints in both inputs, state, and outputs. However, the strong abilities of MPC do not come without a cost. In the conventional MPC formulation, an optimization problem has to be solved online, and for large systems and/or systems requiring high sampling rates, the computational load may become prohibitive. Many approaches have been studied in order to reduce the computational requirements of MPC, including input blocking [2] and utilizing structure in the optimization problem (*e.g.* [11]). This paper will make no attempt at covering all such approaches.

This article proposes a new approach to solving strictly convex Quadratic Programs (QP) problems resulting from MPC problem formulations, based on an implicit problem formulation via ramp functions. Numerical results are provided that demonstrate that the QP problems are solved very fast, for a wide range of problem sizes. The MPC problem is represented exactly, without any simplifications or approximations. The computer code required to implement the solution method is very simple, and the computer memory requirement is modest. The paper is organized as follows. Section 2 provides preliminary results, and describes the MPC formulation used in the paper. Section 3 provides the main result, and in Section 4 the application of the main result for efficiently solving QP problems is described. Section 5 demonstrates the efficiency of the method on a range of MPC problems. A discussion is provided in Section 6, focusing on infeasibility detection and infeasibility handling, while Section 7 concludes the paper.

Notation. For a vector $y \in \mathbb{R}^n$, y_i indicates its ith component. For a matrix $M \in \mathbb{R}^{n\times m}$, $M_{(i,j)}$ denotes its (i,j) component, $M_{(i,\cdot)}$ denotes its ith row, and $M_{(\cdot,j)}$ indicates its jth column. Let $\mathbb{D}^N$ denote the set of diagonal matrices of dimension N, the set of positive (semi-)definite matrices as $\mathbb{S}^n_{>0(\geq 0)} = \{M \in \mathbb{R}^{n\times n} \mid M = M^\top, M >(\geq) 0\}$, and $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ defines the matrix $I_{\mathcal{A}} \subset \mathbb{D}^N$, with $I_{\mathcal{A}(i,i)} \in \{0, 1\}$ and $I_{\mathcal{A}(i,i)} = 1$ if $i \in \mathcal{A}$, $I_{\mathcal{A}(i,i)} = 0$ if $i \notin \mathcal{A}$. We also denote $\mathcal{A}^c = \{1, 2, \ldots, N\} \setminus \mathcal{A}$. We denote e_i the ith vector of the canonical basis of $\mathbb{R}^N$.

⋆ This work was supported by the French-Norwegian exchange programme AURORA under contract 294676. This research is also funded in part by ANR via project HANDY, number ANR-18-CE40-0010.

Email addresses:
giorgio.valmorbida@centralesupelec.fr (Giorgio Valmorbida), morten.hovd@itk.ntnu.no (Morten Hovd).

2 MPC formulation and preliminary results

2.1 QP-MPC formulation

Consider the discrete state-space model

$$x_{k+1} = Ax_k + Bu_k \tag{1}$$

with $x_k \in \mathcal{X} \subseteq \mathbb{R}^n$ and $u_k \in \mathcal{U} \subseteq \mathbb{R}^m$, $\forall k \in \mathbb{Z}_{\geq 0}$, where $\mathcal{X}$ and $\mathcal{U}$ are polytopic sets. Given the objective function

$$J(\mathbf{u}, x_k) = x_{k+N}^\top P_T x_{k+N} + \sum_{i=0}^{N-1} x_{k+i}^\top P_i x_{k+i} + u_{k+i}^\top R_i u_{k+i} \tag{2}$$

where $\mathbf{u} = [u_k^\top \ \cdots \ u_{k+N-1}^\top]^\top$, and $P_i \in \mathbb{S}^n_{\geq 0}$, $R_i \in \mathbb{S}^n_{>0}$ $\forall i = 1, \ldots, N-1$, and $P_T \in \mathbb{S}^n_{\geq 0}$. As discussed in [1, Sec 2.1], the MPC problem is formulated as:

$$\begin{aligned} \underset{\mathbf{u}}{\text{minimize}} \ & \frac{1}{2}\mathbf{u}^\top H \mathbf{u} + x_k^\top F^\top \mathbf{u} \\ \text{subject to} \ & G\mathbf{z} \leq S_u x_k + w \end{aligned} \tag{3}$$

where the matrices H, F, G,and S_u depend on the constraints defining $\mathcal{X}, \mathcal{U}$, and on the matrices defining system (1) and the objective function (2). Since $R_i \in \mathbb{S}^n_{>0}$, we have $H \in \mathbb{S}^n_{>0}$ thus an invertible matrix. Also following [1], the variable change $\mathbf{z} = \mathbf{u} + H^{-1}Fx_k$ results in

$$\begin{aligned} \underset{\mathbf{z}}{\text{minimize}} \ & \frac{1}{2}\mathbf{z}^\top H \mathbf{z} \\ \text{subject to} \ & G\mathbf{z} \leq Sx_k + w \end{aligned} \tag{4}$$

where $S = S_u + GH^{-1}F$. In this paper we search for solutions to the above problem based on the solution of an implicit equation in terms of ramp functions. Some properties of the ramp functions as detailed below

Definition 1 *The ramp function $r(y)$ is given by*

$$r(y) = \begin{cases} 0 & \text{if } y < 0 \\ y & \text{if } y \geq 0 \end{cases} \tag{5}$$

Lemma 1 *The ramp function is the only function satisfying, $\forall y \in \mathbb{R}$*

$$(r(y) - y)r(y) = 0 \tag{6a}$$
$$r(y) \geq 0 \tag{6b}$$
$$(r(y) - y) \geq 0. \tag{6c}$$

Proof. First note that the ramp function can be expressed as the unique solution to the convex optimization problem parameterized in y (thus depending on y) with strictly convex objective function as follows

$$\underset{r}{\text{minimize}} \ \frac{1}{2}(r-y)^2 \quad \text{subject to} \ \ r \geq 0. \tag{7}$$

With the Lagrangian associated to the optimization problem, $\mathcal{L}(r, \lambda) = \frac{1}{2}(r-y)^2 - \lambda r$, we obtain the Karush-Kuhn-Tucker (KKT) conditions

$$(r - y) - \lambda = 0; \ \lambda r = 0; \ r \geq 0; \ \lambda \geq 0$$

which are necessary for optimality and also sufficient since the problem is strictly convex. To obtain a description in the variables (y, r) one can use $\lambda = (r - y)$ above to obtain $r \geq 0, \ \ (r - y) \geq 0, \ \ r(r - y) = 0$. ■

For $y \in \mathbb{R}^m$ let us define the function $\mathbf{r} : \mathbb{R}^m \to \mathbb{R}^m$, the vector-valued ramp function, as

$$\mathbf{r}(y) = \begin{bmatrix} r(y_1) & r(y_2) & \cdots & r(y_m) \end{bmatrix}^\top$$

Remark 1 Note that, due to the piecewise definition of the ramp function in (5), we have that vector $\mathbf{r}$ can also be expressed as the product $\mathbf{r}(y) = I_{\mathcal{A}} y$ where $I_{\mathcal{A}} \in \mathbb{D}^m$, with its diagonal elements verifying $I_{\mathcal{A}(i,i)} \in \{0, 1\}$. Clearly, the set $\mathcal{A} \subseteq \{1, 2, \ldots, N\}$ depends on y. ⌟

3 Main results

This section shows how a Linear Complementarity Problem (LCP) can be associated to a QP-MPC to obtain a piece-wise affine (PWA) representation implicitly defined in terms of the vector-valued ramp function $\mathbf{r}$. Section 4 details how this implicit PWA representation can be exploited to quickly calculate solutions to the QP-MPC problem.

Theorem 1 *The solution of the QP-MPC is a PWA function on the variable x, given by*

$$\mathbf{u}(x) = -H^{-1}F^\top x - H^{-1}G^\top \mathbf{r}(y) \tag{9a}$$
$$y = -Sx + (I - GH^{-1}G^\top)\mathbf{r}(y) - w. \tag{9b}$$

Proof. The following are the KKT conditions associated to (3) as presented in [1, eq.(15)]

$$H\mathbf{z} + G^\top \lambda = 0 \tag{10a}$$
$$\lambda_i(S_{(i,\cdot)}x + w_i - G_{(i,\cdot)}\mathbf{z}) = 0 \tag{10b}$$
$$\lambda \geq 0 \tag{10c}$$
$$Sx + w - G\mathbf{z} \geq 0. \tag{10d}$$

From (10a) we obtain $\mathbf{z} = -H^{-1}G^\top \lambda$, and using this expression in (10b)-(10d), gives the following LCP constraints

$$\begin{aligned} \lambda_i(S_{(i,\cdot)}x + w_i + G_{(i,\cdot)}H^{-1}G^\top \lambda) &= 0 \\ \lambda &\geq 0 \\ Sx + w + GH^{-1}G^\top \lambda &\geq 0. \end{aligned}$$

The above set of constraints can be rewritten as

$$\begin{aligned}\lambda_i \left(\lambda_i - \left(-S_{(i,\cdot)}x + (I - GH^{-1}G^\top)_{(i,\cdot)}\lambda - w_i\right)\right) &= 0\\ \lambda &\geq 0\\ \left(\lambda - \left(-Sx + (I - GH^{-1}G^\top)\lambda - w\right)\right) &\geq 0.\end{aligned}$$

By defining

$$y = -Sx + (I - GH^{-1}G^\top)\lambda - w, \tag{11}$$

we observe that the above set of inequalities and the complementarity constraint become

$$\lambda_i \left(\lambda_i - y_i\right) = 0,\ \lambda \geq 0,\ (\lambda - y) \geq 0.$$

According to Lemma 1, the set of multipliers λ is given by the ramp function of variable y, namely

$$\lambda = \mathbf{r}(y) \tag{12}$$

then, from (11), we have that y satisfies the implicit algebraic equation (9b), while (9a) follows from the change of variables $\mathbf{u} = \mathbf{z} - H^{-1}Fx$, from (10a) and (12). ■

From the above theorem, we have that, given x, the computation of the control action, u, can be carried out by solving the implicit equation in (9b), yielding y. From this solution, the Lagrange multipliers λ can be computed according to (12). The computation of $\mathbf{u}(x)$ in (9a) thus boils down to the solution of the implicit equation.

The next section proposes an algorithm to solve this implicit equation by exploiting the fact that the ramp function can be written as a matrix multiplication $\mathbf{r}(y) = I_{\mathcal{A}(x)}y$ where we indicate the dependence of the set of *active constraints* $\mathcal{A}$ on the value of x. Indeed, if the solution, depending on x, gives $y_i < 0$ then $i \notin \mathcal{A}$. If instead, $y_i \geq 0$, then $i \in \mathcal{A}$.

4 Implementation

It is clear from Theorem 1 that the KKT conditions for the MPC optimization problem is an LCP. This is well known both for QPs in general (see, *e.g.* [9]), and within the MPC literature (*e.g.* [3, 6]). However, this knowledge has not been utilized to devise an algorithm for solving the MPC QP problem, using ramp functions to take advantage of the structure and continuity of the problem.

In this section, we present an algorithm to solve the algebraic equation (9b). The algorithm exploits the fact that the ramp function can be expressed as a matrix multiplication by a diagonal matrix as pointed out in Remark 1. That is, that (9b) can be written as

$$y - (I - GH^{-1}G^\top)I_{\mathcal{A}}y = -Sx - w \tag{13}$$

for some matrix $I_{\mathcal{A}}$ to be determined. The set $\mathcal{A}$ corresponds to the set of *active constraints*, namely to the set for which the values of multipliers λ are not zero. Therefore, according to the ramp function defining the multipliers, if $y_i < 0$ then the corresponding value of the multiplier is $\lambda_i = 0$, and in this case, $I_{\mathcal{A}(i,i)} = 0$, that is, the the corresponding diagonal element of $I_{\mathcal{A}}$ is zero. If instead $y_i \geq 0$ then $I_{\mathcal{A}(i,i)} = 1$. The following definition gives the definition of the pairs $(y, I_{\mathcal{A}})$ allowing to establish an equivalence between (9b) and (13).

Definition 2 *For a given x, a pair is said $(y, I_{\mathcal{A}})$* compatible *if $(y, I_{\mathcal{A}})$ satisfy* (13) *and $I_{\mathcal{A}(i,i)} = 1$ if $y_i \geq 0$ and $I_{\mathcal{A}(i,i)} = 0$ if $y_i < 0$.*

Based on the above observations, the solution to the implicit equation (9b) is obtained whenever we satisfy (13) with a *compatible* pair $(y, I_{\mathcal{A}})$. We shall denote $\mathcal{A}(x)$ as the set of active constraints for a given x defining the right hand side of (9b). To search for a solution to (9b), we propose below an algorithm that searches for the set of active constraints by adding elements to or removing elements from the set $\mathcal{A}$, thus modifying matrix $I_{\mathcal{A}}$ one element at the time, and updating the solutions y using the matrix inversion lemma, which is recalled below.

Lemma 2 (Matrix Inversion Lemma) *Let $Q \in \mathbb{R}^{N\times N}$ be an invertible matrix, we have*

$$(Q+q_uq_cq_v)^{-1} = Q^{-1} - Q^{-1}q_u(q_c^{-1}+q_vQ^{-1}q_u)^{-1}q_vQ^{-1}$$

where $q_c \in \mathbb{R}\setminus\{0\}$ is a scalar, $q_u \in \mathbb{R}^N$, and $q_v^\top \in \mathbb{R}^N$, that is q_u is a column vector and q_v is a row vector.

Remark 2 Noting that $(q_c^{-1} + q_vQ^{-1}q_u)$ is a scalar, it is clear that the update of Q^{-1} when adding the rank one term $Q^{-1}q_uq_vQ^{-1}$ requires only sums and multiplications and a single division by a scalar. ⌟

Let us define

$$Q(\mathcal{A}) = \left(I_{\mathcal{A}^c} + GH^{-1}G^\top I_{\mathcal{A}}\right)$$

and observe that the solution to the linear system (13) is obtained following

$$\begin{aligned}y - (I - GH^{-1}G^\top)I_{\mathcal{A}}y &= -Sx - w\\ Q(\mathcal{A})y &= -Sx - w\\ y &= Q(\mathcal{A})^{-1}\left(-Sx - w\right).\end{aligned}$$

Consider the case where $i \notin \mathcal{A}$ and define $\mathcal{A}_{+i} = \mathcal{A}\cup\{i\}$, namely i *enters* the set $\mathcal{A}$. Since $I_{\mathcal{A}_{+i}} = I_{\mathcal{A}} + e_ie_i^\top$, and $I_{\mathcal{A}^c_{+i}} = I_{\mathcal{A}^c} - I_{\mathcal{A}^c(\cdot,i)}e_i^\top$, we have

$$Q(\mathcal{A}_{+i}) = Q(\mathcal{A}) - (I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)}e_i^\top.$$

The above matrix $Q(\mathcal{A}_{+i})$ presents a sum of $Q(\mathcal{A})$ with a rank one matrix as in Lemma 2, with, respectively $q_c =$

-1, $q_v = e_i^\top$ and $q_u = (I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)}$. Following the matrix inversion lemma, we have the update of the inverse by adding i to $\mathcal{A}$ as

$$\begin{aligned} Q(\mathcal{A}_{+i})^{-1} =& Q(\mathcal{A})^{-1} \\ & - (-1 + e_i^\top Q(\mathcal{A})^{-1}(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)})^{-1} \times \\ & \left(Q(\mathcal{A})^{-1}(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)} e_i^\top Q(\mathcal{A})^{-1}\right). \end{aligned}$$

By defining $v(\mathcal{A}, i) = Q^{-1}(\mathcal{A})(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)}$, we obtain

$$\begin{aligned} Q(\mathcal{A}_{+i})^{-1} =& Q(\mathcal{A})^{-1} \\ & - (-1 + v(\mathcal{A}, i)_i)^{-1} v(\mathcal{A}, i) \left(Q(\mathcal{A})^{-1}\right)_{(i,\cdot)}. \end{aligned} \tag{14}$$

We also have the update of y, denoted $y_+[i]$ as

$$\begin{aligned} y_+[i] =& Q(\mathcal{A}_{+i})^{-1}(-Sx - w) \\ =& Q(\mathcal{A})^{-1}(-Sx - w) \\ & - (-1 + e_i^\top Q(\mathcal{A})^{-1}(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)})^{-1} \times \\ & \left(Q(\mathcal{A})^{-1}(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)} e_i^\top Q(\mathcal{A})^{-1}\right)(-Sx - w) \\ =& y - (-1 + e_i^\top Q(\mathcal{A})^{-1}(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)})^{-1} \times \\ & \left(Q(\mathcal{A})^{-1}(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)}\right) y_i, \end{aligned}$$

that is, $y_+[i] = y - y_i(-1 + v(\mathcal{A}, i)_i)^{-1} v(\mathcal{A}, i)$.

Similarly, consider the case where $i \in \mathcal{A}$ and define $\mathcal{A}_{-i} = \mathcal{A} \setminus \{i\}$, namely i *is removed from* the set $\mathcal{A}$,

$$Q(\mathcal{A}_{-i}) = Q(\mathcal{A}) + e_i^\top (I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)}$$

The above matrix $Q(\mathcal{A}_{-i})$ presents a sum of $Q(\mathcal{A})$ with a rank one matrix as in Lemma 2, with, respectively $q_c = 1$, and, as above, $q_v = e_i^\top$ and $q_u = (I_{\mathcal{A}} - GH^{-1}G^\top)_{(\cdot,i)}$. thus, similarly to above, the following expressions for the inverse and solution updates

$$\begin{aligned} Q(\mathcal{A}_{-i})^{-1} =& Q(\mathcal{A})^{-1} \\ & - (1 + v(\mathcal{A}, i)_i)^{-1} v(\mathcal{A}, i) \left(Q(\mathcal{A})^{-1}\right)_{(i,\cdot)}, \end{aligned} \tag{15}$$

$$y_-[i] = y - y_i(1 + v(\mathcal{A}, i)_i)^{-1} v(\mathcal{A}, i). \tag{16}$$

Given the above steps to update the inverse and y, we use the following criteria to choose which elements to add to or to remove from the current set of active constraints $\mathcal{A}$. These criteria are based on elements of the pair $(y, \mathcal{A})$

- If $y_i < 0$ for $i \in \mathcal{A}$, then constraint i is removed from $\mathcal{A}$ and the inverse of $Q(\mathcal{A}_{-i})$ and an update of y are computed.
- If $y_i \geq 0$ for $i \notin \mathcal{A}$, then constraint i is added to $\mathcal{A}$ and the inverse of $Q(\mathcal{A}_{+i})$ and an update of y are computed.

The order in which we carry out the inclusion and removal of elements in $\mathcal{A}$ is as follows

- Remove first from $\mathcal{A}$ constraints corresponding to negative y_i. If there are more then one such constraint, remove first the one corresponding to the most negative y_i.
- Then add to $\mathcal{A}$ a constraint that is not a member of $\mathcal{A}$ and corresponds to a positive value of y_i. If there is more than one such constraint, add first the one corresponding to the largest y_i.

The above steps are described in Algorithm 1 below. The solution of the algebraic loop in (9b) is obtained with Algorithm 1 and its use in the QP-MPC to obtain the control input u_k from the vector $\mathbf{u}$ as detailed in Algorithm 2.

Algorithm 1 Solution to the algebraic equation (9b)

Require: $GH^{-1}G$, $\mathcal{A}$, $invQ$, y satisfying $y = invQ(-Sx - w)$
  while $(\mathcal{A}, y)$ not compatible **do**
    if $ind(sign(y, -1)) \cap \mathcal{A} \neq \emptyset$ **then**
      $L \leftarrow ind(sign(y, -1)) \cap \mathcal{A}$
      $i \leftarrow ind(min(y, L))$
      $v \leftarrow invQ(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)}$
      $\mathcal{A} \leftarrow \mathcal{A} \setminus \{i\}$
      $q_0 \leftarrow 1$
    else if $ind(sign(y, 1)) \neq \mathcal{A}$ **then**
      $L \leftarrow ind(sign(y, 1)) \setminus \mathcal{A}$
      $i \leftarrow ind(max(y, L))$
      $v \leftarrow invQ(I_{\mathcal{A}^c} - GH^{-1}G^\top)_{(\cdot,i)}$
      $\mathcal{A} \leftarrow \mathcal{A} \cup \{i\}$
      $q_0 \leftarrow -1$
    end if
    $invQ \leftarrow invQ - (q_0 + v_i)^{-1} v\, (invQ)_{(i,\cdot)}$
    $y \leftarrow y - y_i (q_0 + v_i)^{-1} v$
  end while
    return $\mathcal{A}$, $invQ$, y

Algorithm 2 Solution to the QP-MPC (3)

Require: $GH^{-1}G$, S, w, F as in (9b) and (3)
  while MPC running **do**
    Obtain x_k
    $invQ \leftarrow I$
    $\mathcal{A} \leftarrow \emptyset$
    $y \leftarrow (-Sx_k - w)$
    $y \leftarrow Algorithm1(GH^{-1}G, \mathcal{A}, invQ, y)$
    $\mathbf{u}(x_k) \leftarrow -H^{-1}F^\top x_k - H^{-1}G^\top \mathbf{r}(y)$
    Apply $u(x_k)$ to the plant
  end while

5 Numerical Results

To illustrate the performance of the proposed approach, numerical results are provided for MPC applied to three

systems of different size. The simulation times are reported for simulating 100 timesteps using the implicit QP solution approach described in the previous section, and compared to the simulation times when using `qpOASES`[4, 5] (starting from the same initial state). The simulations are performed in Matlab on a Windows PC. The calculation times can vary a little from run to run, since Windows is not a real time operating system. The simulations are therefore repeated multiple times, and the average time is reported.

The solver `qpOASES` is a highly regarded QP solver for MPC problems. It takes advantage of the observation that the state x on the right hand side of the constraints in (3) is unlikely to change very much from one timestep to the next. At time k, a parametric active set approach is used to follow a homotopy path from the optimal solution for x_{k-1} to the optimal solution for x_k. In the examples, the upper and lower constraints for $\mathbf{u}$ are separated from the other constraints in (3), and given to `qpOASES` separately, as recommended in [5]. At each timestep except the initial one, the hot start functionality in `qpOASES` is used.

The two optimization solvers give essentially the same results, and they would be indistinguishable in plots showing the closed-loop behavior of the systems. Therefore we do not illustrate the time responses of the closed loop. Instead, the focus here are on the simulation times, which document the efficiency of the proposed method. In each example, constant weights $P_i = P$ and $R_i = R$ are used in (2), and the terminal weight P_T is set equal to the solution of the algebraic Riccati equation for the weights P and R.

Example 1 Consider the double integrator example from [8]. The system dynamics are given by

$$x_{k+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 0.3 \end{bmatrix} u_k$$

with the input constrained between -1 and 1, and each state constrained between -5 and 5, $P = I$, and $R = 1$.

The prediction horizon $N = 10$ is used, and the system with control is simulated for 100 time steps. The average time for 5 runs starting from $x_0 = [5\ -2]^\top$ when solving the QP using ramp functions is $0.0020s$, whereas the average time when using `qpOASES` is $0.0023s$. ⌟

Example 2 We now study the four-state system from [8, 7]. The discrete-time dynamics are given by

$$x_{k+1} = \begin{bmatrix} 0.928 & 0.002 & -0.003 & -0.004 \\ 0.041 & 0.954 & 0.012 & 0.006 \\ -0.052 & -0.046 & 0.893 & -0.003 \\ -0.069 & 0.051 & 0.032 & 0.935 \end{bmatrix} x_k + \begin{bmatrix} 0 & 0.336 \\ 0.183 & 0.007 \\ 0.090 & -0.009 \\ 0.042 & 0.012 \end{bmatrix} u_k$$

$$y_k = Cx_k = \begin{bmatrix} 0 & 0 & -0.098 & 0.269 \\ 0 & 0 & 0.080 & 0.327 \end{bmatrix} x_k$$

with each input and each measurement constrained to lie between -1 and 1, $P = C^\top C$ and $R = I$.

The prediction horizon is $N = 30$, and the initial state $x_0 = [25.5724\ 25.3546\ 9.7892\ 0.2448]^\top$. Simulating 100 timesteps when solving the QP using ramp functions took on average $0.0041s$, while the average time for `qpOASES` was $0.0090s$. ⌟

Example 3 The final example is based on the 82-state binary distillation column model by Skogestad [13]. The control inputs are four in total: the top product flowrate, the top reflux flowrate, the bottom product flowrate, and bottom boil-up (energy supply). The outputs are the top and bottom compositions, and the liquid levels in the top accumulator and the column bottoms. The levels are open loop integrators.

First the model is balanced and reduced to 25 states. Next, the reduced model is converted to discrete time. Both the original system and the reduced model equations and initial state are not listed here for space reasons [1].

Using a prediction horizon $N = 30$, the resulting QP problem has 1562 constraints and 120 degrees of freedom. Simulating 100 timesteps when solving the QP using ramp functions took on average $0.0055s$, while the average time for `qpOASES` was $0.0401s$. ⌟

Table 2 summarizes indicators for the performance. Note, in particular, that few of the time steps required the maximum number of iterations in Algorithm (1). Note also the sparsity of matrix $Q(\mathcal{A})^{-1}$, which is smaller for the example with more states.

[1] They can be obtained for morten.hovd@itk.ntnu.no on request, together with the weights used in the MPC objective function.

			Algorithm 2						
Ex.	n	N	max. # iterations	k with max. # iterations	average # iterations	max. # elements in $\mathcal{A}$	Worst sparsity of $Q^{-1}(\mathcal{A})$	avg. time per time step (μs)	worst time ($k=1$) (μs)
1	2	10	6	1-2	1.2	7.6%	8.7%	20	465
2	4	30	4	1-14	1.47	0.95%	1.3%	41	388
3	25	30	4	1	1.02	0.26%	0.26%	55	962

Table 1
Order of the system (n), prediction horizon (N). All examples ran 100 time-steps from the reported initial conditions. For Algorithm 2 the table reports the maximum number of iterations of Algorithm 1 within a time step, time steps in which the maximum number iterations of Algorithm 1 were executed, average number of iterations of Algorithm 1 for the 100 time-steps, maximum number of active constratins (as percentage of total), percentage non-zero elements in $Q^{-1}(\mathcal{A})$, average time step for the 100 time steps and maximum time taken within one time-step (corresponding to the first time $k=1$).

	Algorithm 2	`qpOASES`
Ex.	Total time (ms)	Total time (ms)
1	2.0	2.3
2	4.1	9.0
3	5.5	40.1

Table 2
Total execution time for 100 time steps of Algorithm 2 and `qpOASES` considering the same initial conditions (both from the average of several runs).

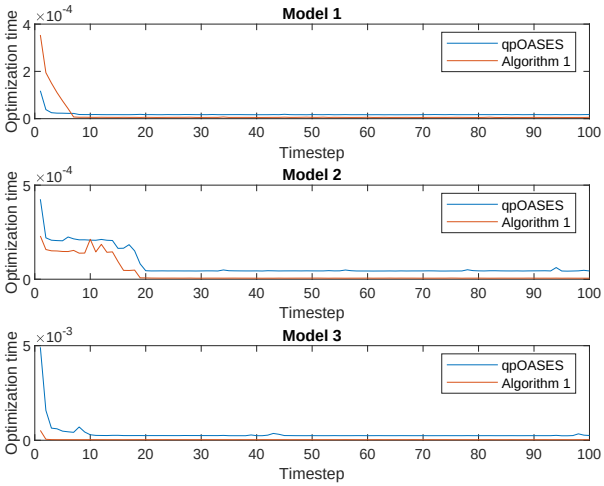

Fig. 1. Solution times in seconds per timestep for a single run of Algorithm 1 and `qpOASES`, for each example.

The solution time per timestep is shown, for each example, in Fig. 1. `qpOASES` is the faster solver initially for Example 1, but the rest of the time Algorithm 1 is faster.

Figure 2 shows the simulation times for 100 simulation runs, each of length 100 timesteps, for Algorithm 1 and `qpOASES`. For run $i, i = \{1, \ldots, 100\}$ the same initial conditions are used for Algorithm 1 and `qpOASES`. The initial conditions are randomly generated, although only feasible initial conditions are retained for comparison of simulation times. Algorithm 1 more often results in shorter simulation times.

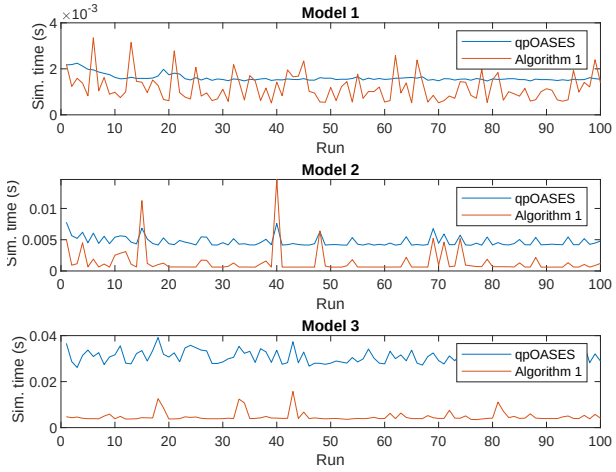

Fig. 2. Solution times in seconds for simulation runs of length 100 timesteps for Algorithm 1 and `qpOASES`, for each example.

6 Discussion

Relationship to active set methods.
Algorithm 1 updates the set of active constraints at each iteration, and may thus be categorized as an active set method. However, conventional active set methods solve the KKT conditions (in the form of a set of linear equations) for each candidate active set tried, in order to determine a step direction and thereby find which constraint to add or drop. In contrast, Algorithm 1 updates $invQ$ through a rank 1 update, and the constraint to be added or dropped is found simply from looking at the values of the elements of the y vector, taking into account whether the corresponding constraint is in the active set.

Memory requirements for Algorithm 1
The computer memory required for implementing Algorithm 1 is clearly very modest, consisting mainly of the storage of $GH^{-1}G^{\top})$, S, and w, which is the minimum required to define the LCP, as well as the leading m rows of $H^{-1}F^{\top}$ and $H^{-1}G^{\top}$ that are required for calculating u from x and λ (see Theorem 1). In addition, Q^{-1} and the set of active constraints $\mathcal{A}$ needs to be stored (and updated), but storing these as sparse matrices can save significant memory, since, as noted above, the num-

ber of active constraints cannot exceed the number of degrees of freedom. Indeed, for the above numerical examples, a full matrix Q^{-1} would contain n_c^2 elements, where $n_c = 66$, $n_c = 316$ and $n_c = 1562$ respectively for examples 1, 2 and 3. However, since only a few of these elements are not zero, as detailed in Table 2, column *Worst sparsity of* $Q^{-1}(\mathcal{A})$, the sparse storage allows to reduce the storage space to a small fraction of the dimensions of $Q^{-1}(\mathcal{A})$.

Initial data for Algorithm 1
One possible input data for the solution of the implicit equation using Algorithm 1 is $\mathcal{A} = \emptyset$ corresponding to no active constraints, giving $Q_{\mathcal{A}} = I$, hence $y = -Sx - w$.

This initial input is clearly very efficient when there are no active constraints. However, with the proposed rules for removing or adding constraints to the active set given in Section 4, the numerical experiments showed that the solution method is still very efficient, and the number of solver iterations within one timestep have not been found to exceed the number of (actually) active constraints by much.

Typically, the value of the state will not change much from one timestep to the next, and it may therefore seem advantageous to start the optimization from the active set at the previous timestep. However, it is important to take advantage of the fact that Q^{-1} typically is a highly sparse matrix. It is observed that when (over multiple time steps) hot starting the optimization from the active set at the previous time step, then Q^{-1} will gradually fill up with elements of magnitude around the order of `eps`, *i.e.*, elements that should have been set to zero when adding and later removing constraints. The small non-zero values are due to finite precision in the calculations. This both slows down the calculations and for large problems significantly increases the requirement for computer memory. The problem is easily handled by cleaning Q^{-1} of such very small elements (*e.g.*, using the Matlab function `clean`), but it is found that this takes more time than starting each iteration with an empty active set, corresponding to $Q^{-1} = I$. However, there are situations where it may be possible to do the cleaning of tiny elements from Q^{-1} without incurring significant time loss, especially if parallel computing is an option:

- There is actually a sequence of calculations that need to be carried out within each timestep: first measurements are obtained, next the model is updated from measurements (typically using state estimation), only thereafter are the MPC calculations performed. It may therefore be possible to clean Q^{-1} while the model is updated.
- In some cases there may be ample time for calculations within each sample interval, but it is desired to minimize the time delay *within each sample interval* associated with the MPC calculations. In such a situation, the cleaning of Q^{-1} can be performed after the MPC calculations and after the new input has been implemented, but before the end of the sample interval.

Infeasibility detection
To detect infeasibily we have relied on the heuristics of checking the value of $(-1 + v(\mathcal{A}, i)_i)$ when adding constraints. Indeed, we observe that these values become very small when an inconsistent set of constraints is selected.

Infeasibility has been investigated for all examples above by scaling the initial state until the problem becomes infeasible. Using a threshold value of 10^{-13} for $(-1 + v(\mathcal{A}, i)_i)$, infeasibility was correctly determined in all three examples, to within an accuracy of three decimal places in the scaling factor.The appropriate threshold value will depend on, and should be significantly larger than, the machine precision. Clearly, setting the threshold too large will result in the algorithm declaring infeasibility unnecessarily. Even though we implemented this procedure for infeasibility detection, it has not been detailed Algorithm 1.

Soft Constraints
From an application point of view, a quite different issue is what *should* happen when an MPC problem is infeasible. The basic options are either to have some backup functionality bringing the system to a safe state (which often means shutting down the system entirely), or to design the MPC to "make the best of it", trying to minimize whatever damage might be caused by operation in an infeasible region of the state space. Such damage minimization can be achieved by using *soft constraints*, by adding slack variables to constraints where violations are physically possible and operationally (temporarily) tolerable, and adding corresponding penalty terms to the objective function.

A desirable property is for the soft constraints to be *exact*. This is achieved by a sufficiently large weight on a penalty term linear in the slack variable. What is sufficiently large can be calculated as described in [8]. Quadratic terms in the penalty function do not affect whether the soft constraint is exact, and quadratic terms are therefore sometimes dropped. However, when solving the MPC QP using ramp functions, the Hessian matrix needs to be invertible (positive definite), and hence weights on quadratic terms in the penalty functions are required.

Precision of the Lagrange Multipliers and complementarity KKT conditions
Although of little importance in practical applications, it is interesting to observe that the complementarity constraint of the KKT conditions is fulfilled with very high accuracy for the proposed method. That is, the λ's for the inactive constraints are identically zero (or of magnitude similar to machine precision), whereas conven-

tional optimization routines will give λ's for inactive constraints in the range of some tolerance specification – often in the range of 10^{-8}.

Need for Rank 2 updates

The expression for the control law (9) in Theorem 1 is obtained from the general convex QP formulation (4). Recall that Algorithm 1 uses rank 1 updates to compute a solution to (9). We point out that Algorithm 1, combined with the infeasibility detection described above, has been effectivelly applied to QP problems arising from MPC formulations discussed in Section 2 and illustrated in examples in in Section 5.

Although, in the authors' experience, Algorithm 1 works very well for QP problems arising from MPC formulations, strictly convex QP formulations exist for which a modification of the algorithm is required. This modification copes with the need for rank two updates. To illustrate such a case, consider the problem (4) with $S = 0$ and

$$H = \begin{bmatrix} 11 & 9 \\ 9 & 11 \end{bmatrix},\ G = \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix},\ w = \begin{bmatrix} -0.5 \\ -0.8 \\ -\frac{1}{2\sqrt{2}} \\ -\frac{0.15}{\sqrt{10}} \end{bmatrix}.$$

Since there are only two degrees of freedom in the corresponding QP, the maximum number of strongly active constraints is two. Starting Algorithm 1 with an empty active set, $\mathcal{A} = \emptyset$, constraint 2 is the first to add to the active set, and then constraint 4 should be added. The active set $\mathcal{A} = \{2, 4\}$, results in

$$y = \begin{bmatrix} 0.2833 & 5.2444 & -0.0589 & 5.0772 \end{bmatrix}^\top.$$

which, following Algorithm 1, is not a compatible solution. Thus, constraint 1 should be added to the active set - but since the problem has only two degrees of freedom, this would lead to a rank defect Q matrix. Therefore, one of the constraints in the active set has to be removed when introducing constraint 1 into the active set, even though all constraints in the active set have positive y values, namely a rank 2 update.

The following procedure can be used to identify the constraint to be removed. Let the index i denote the constraint to be added to the active set, and note that the problem only arises when the number of constraints already in the active set equals m. The current (non-optimal) solution point can be calculated from (9a) and the current value of y. Clearly, the solution point needs to move in a direction for which the value of constraint i decreases. Start with $v_0 = -G_{(i,:)}^\top$ as the candidate direction for changing the solution point, and set $S_0 = I$. Obviously, the solution point cannot move in a direction in which the values of the constraints in the active set increase.

(1) For $k = 1 : m$, define j as the constraint index (row index in G) for element k of $\mathcal{A}$.
(2) Select a constraint index k such that $G_{(j,:)}v_{k-1} \geq 0$.
(3) Set $v_k = (I - r_j^\top (r_j r_j^\top)^{-1} r_j) v_{k-1}$, where $r_j^T = S_{j-1} G_{(j,:)}^\top$ and $S_j = (I - r_j^\top (r_j r_j^\top)^{-1} r_j) S_{j-1}$. Note that $r_j r_j^\top$ is a scalar, so the inversion is actually a scalar division.
(4) Repeat from (2) until $G_{(j,:)}v_{k-1} \leq 0$ for all remaining unselected constraints.

The constraint to be removed from $\mathcal{A}$ is then the first constraint to become inactive when moving the solution point in the direction v_m. If $v_m = 0$ the problem is infeasible.

Following these steps in the above example, we find that constraint 2 should be removed from the active set. It is easily checked that $\mathcal{A} = \{1, 4\}$ is optimal.

The corresponding rank 2 update to Q^{-1} can be carried out either using the matrix inversion lemma directly, or applying two rank 1 updates, by first removing the constraint to leave the active set.

7 Conclusions and Perspectives

A novel method for solving QPs arising from MPC problems has been proposed. The method is shown to be efficient for a wide range of problem sizes, and can be implemented using short and simple computer code. The method is currently limited to strictly convex QP problems, semi-definite Hessian matrices cannot be accommodated.

Future work will address this limitation, and also attempt to extend the method to LP-MPC. Finally, both for general QPs and for QP-MPC in particular, we did not study the convergence of the proposed algorithm, this is a topic for future research.

References

[1] A. Bemporad, M. Morari, V. Dua, and E. N. Pistikopoulos. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38:3–20, 2002.

[2] R. Cagienard, P. Grieder, E. Kerrigan, and M. Morari. Move blocking strategies in receding horizon control. *Journal of Process Control*, 17(6):563–570, 2007.

[3] E. F. Camacho. Constrained generalized predictive control. *IEEE Transactions on Automatic Control*, 38(2):327–332, 1993.

[4] H.J. Ferreau, H.G. Bock, and M. Diehl. An online active set strategy to overcome the limitations of explicit MPC. *International Journal of Robust and Nonlinear Control*, 18(8):816–830, 2008.

[5] H.J. Ferreau, C. Kirches, A. Potschka, H.G. Bock, and M. Diehl. qpOASES: A parametric active-set algorithm for quadratic programming. *Mathematical Programming Computation*, 6(4):327–363, 2014.

[6] M. Herceg, M. Kvasnica, C.N. Jones, and M. Morari. Multi-Parametric Toolbox 3.0. In *Proc. of the European Control Conference*, pages 502–510, Zürich, Switzerland, July 17–19 2013. `http://control.ee.ethz.ch/~mpt`.

[7] M. Hovd and R. D. Braatz. Handling state and output constraints in MPC using time-dependent weights. In *Proceedings of the American Control Conference*, 2001.

[8] M. Hovd and F. Stoican. On the design of exact penalty functions for MPC using mixed integer programming. *Computers & Chemical Engineering*, 70:104–113, 2014.

[9] K. G. Murty. *Linear Complementarity, Linear and Nonlinear Programming*. Helderman Verlag, Berlin, Germany, 1988. `http://www-personal.umich.edu/~murty/books/linear_complementarity_webbook`.

[10] S. J. Qin and T. A. Badgwell. A survey of industrial model predictive control technology. *Control Engineering Practice*, pages 733–764, 2003.

[11] C. V. Rao, S. J. Wright, and J. B. Rawlings. Application of interior-point methods to model predictive control. *J. of Optimization Theory and Applications*, 99(3):723–757, December 1998.

[12] J. Richalet, A. Raoult, J. L. Testud, and J. Papon. Model predictive heuristic control: Application to industrial processes. *Automatica*, pages 413–428, 1978.

[13] S. Skogestad. Matlab distillation column model, Accessed October 2021. `http://folk.ntnu.no/skoge/distillation/`.