An algorithm for multi-parametric quadratic programming and explicit MPC solutions

Petter Tøndel\textsuperscript{a,*}, Tor Arne Johansen\textsuperscript{a}, Alberto Bemporad\textsuperscript{b}

\textsuperscript{a} Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7491 Trondheim, Norway
\textsuperscript{b} Dipartimento di Ingegneria dell’Informazione, University of Siena, 53100 Siena, Italy

Received 8 June 2001; received in revised form 30 May 2002; accepted 22 October 2002

Abstract

Explicit solutions to constrained linear model predictive control problems can be obtained by solving multi-parametric quadratic programs (mp-QP) where the parameters are the components of the state vector. We study the properties of the polyhedral partition of the state space induced by the multi-parametric piecewise affine solution and propose a new mp-QP solver. Compared to existing algorithms, our approach adopts a different exploration strategy for subdividing the parameter space, avoiding unnecessary partitioning and QP problem solving, with a significant improvement of efficiency.

Keywords: Linear quadratic regulators; Piecewise linear controllers; Constraints; Predictive control

1. Introduction

Our motivation for investigating multi-parametric quadratic programming (mp-QP) comes from linear model predictive control (MPC). This refers to a class of control algorithms that compute a manipulated variable trajectory from a linear process model to minimize a quadratic performance index subject to linear constraints on a prediction horizon. The first control input is then applied to the process. At the next sample, measurements are used to update the optimization problem, and the optimization is repeated. In this way, this becomes a closed-loop approach. There has been some limitation to which processes MPC could be used on, due to the computationally expensive on-line optimization which was required. There has recently been derived explicit solutions to the constrained MPC problem, which could increase the area of use for this kind of controllers. Explicit solutions to MPC problems are not mainly intended to replace traditional implicit MPC, but rather to extend its area of use. MPC functionality can with this be applied to applications with sampling rates in the \(\mu\)-sec range, using low cost embedded hardware. Software complexity and reliability is also improved, allowing the approach to be used on safety-critical applications. Methods for efficient online implementation of PWA function evaluation in explicit MPC has been developed by exploiting convexity (Borrelli, Baotic, Bemporad, & Morari, 2001) or an associated binary search tree data structure (Tøndel & Johansen 2002; Tøndel, Johansen, & Bemporad, 2003). Independent works by Bemporad, Morari, Dua, and Pistikopoulos (2002), Bemporad, Morari, Dua, and Pistikopoulos (2000b), Johansen, Petersen, & Slupphaug (2002) and Seron, De Dona, and Goodwin (2000) have reported how a piecewise affine (PWA) solution can be computed off-line, while the on-line effort is limited to evaluate this PWA function. In particular, in Bemporad et al. (2002, 2000b) such a PWA function is obtained by treating the MPC optimization problem as a parametric program. Parametric programming is a term for solving an optimization problem for a range of parameter values. One can distinguish between parametric programs, in which only one parameter is considered, and multi-parametric programs, in which a vector of parameters is considered. The algorithm reported in Bemporad

\textsuperscript{a} This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Kenko Uchida under the direction of the Editor Tamer Basar.

\textsuperscript{*} Corresponding author. Tel.: +47-7359-4348; fax: +47-7359-4399.

E-mail addresses: petter.tondel@itk.ntnu.no (P. Tøndel),
tor.arne.johansen@itk.ntnu.no (T.A. Johansen), bemporad@dii.unisi.it (A. Bemporad).
et al. (2002) is the only mp-QP algorithm known to the authors for solving general linear MPC problems, while single parameter parametric QP is treated in Berkelaar, Roos, and Terlaky (1997). Multi-parametric LP (mp-LP) is treated in Gal (1995) and Borrelli, Bemporad, and Morari (in press), mp-LP in connection with MPC based on linear programming is investigated in Bemporad, Borrelli, and Morari (2003), and multi-parametric mixed-integer linear programming (Dua & Pistikopoulos, 2000) is used in Bemporad, Borrelli, and Morari (2000a) for obtaining explicit solutions to hybrid MPC. The mp-LP algorithm of Gal (1995) and the mp-QP algorithm presented in this paper are similar, but while Gal (1995) uses simplex steps to solve the mp-LP, our algorithm proceeds similar to an active set QP solver. The problem of reducing the complexity of the PWA solution to linear quadratic MPC problems is addressed in Johansen et al. (2002) and Bemporad and Filippi (2003), and efficient on-line computation schemes of explicit MPC controllers are proposed in Borrelli et al. (2001). This paper extends the theoretical results of Bemporad et al. (2002), by analyzing several properties of the geometry of the polyhedral partition and its relation to the combination of active constraints at the optimum of the quadratic program. Based on these results, we derive a new exploration strategy for subdividing the parameter space, which avoids: (i) unnecessary partitioning, (ii) the solution to LP problems for determining an interior point in each new region of the parameter space, and (iii) the solution to the QP problem for such an interior point. As a consequence, there is a significant improvement of efficiency with respect to the algorithm of Bemporad et al. (2002). Some preliminary results were presented in Tøndel, Johansen, and Bemporad (2001a).

2. From linear MPC to an mp-QP problem

The main aspects of formulating a linear MPC problem as a mp-QP will be repeated here for convenience (see Bemporad et al., 2002, for further details). Consider the linear system

\[ x(t + 1) = Ax(t) + Bu(t) \]

\[ y(t) = Cx(t), \]

where \( x(t) \in \mathbb{R}^n \) is the state vector, \( u(t) \in \mathbb{R}^m \) is the input vector, \( A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m} \) and \( (A, B) \) is a controllable pair. For the current \( x(t) \), MPC solves the optimization problem

\[
\min_U \left\{ J(U, x(t)) = \sum_{t=0}^{N-1} x_{t+k}^T P x_{t+k} + u_{t+k}^T R u_{t+k} \right\}
\]

subject to \( y_{\min} \leq y_{t+k} \leq y_{\max}, \quad k = 1, \ldots, N, \]

\( u_{\min} \leq u_{t+k} \leq u_{\max}, \quad k = 0, \ldots, M - 1, \]

\( u_{t+k} = K x_{t+k}, \quad M \leq k \leq N - 1, \]

\( x_{t+k} = x(t), \]

\( x_{t+k+1} = Ax_{t+k} + Bu_{t+k}, \quad k \geq 0, \]

\( y_{t+k} = C x_{t+k}, \quad k \geq 0 \)

with respect to \( U = [u_t^T, u_{t+M-1}^T]^T \), where \( y_{\min} < 0 < y_{\max}, \quad u_{\min} < 0 < u_{\max}, \quad R = R^T > 0, \quad Q = Q^T > 0, \quad P = P^T > 0, \quad x_{t+k} \) is the prediction of \( x_{t+k} \) at time \( t \), and \( M \) is the control input horizon. When the final cost matrix \( P \) and gain \( K \) are calculated from the algebraic Riccati equation, under the assumption that the constraints are not active for \( k \geq N \) (2) exactly solve the constrained (infinite-horizon) LQR problem for (1) with weights \( Q, R \) (see also, Sznaier & Damborg, 1987; Chmielewski & Manousiouthakis, 1996; Scokaert & Rawlings, 1998).

For simplicity we consider the regulator problem (2), but the algorithm developed in this paper is directly applicable to tracking and measured disturbance rejection problems as described in Bemporad et al. (2002). These problems can by some algebraic manipulation be reformulated as

\[
V_z(x(t)) = \min_z \frac{1}{2} z^T H z
\]

subject to \( G z \leq W + S x(t) \), where \( z \triangleq U + H^{-1} F^T x(t) \) and \( x(t) \) is the current state, which can be treated as a vector of parameters. Note that \( H > 0 \) since \( R > 0 \). The number of inequalities is denoted by \( q \) and the number of free variables is \( n_z = m \cdot N \). Then \( z \in \mathbb{R}^q, \quad H \in \mathbb{R}^{n_z \times n_z}, \quad G \in \mathbb{R}^{n_z \times n}, \quad W \in \mathbb{R}^{n_z}, \quad S \in \mathbb{R}^{n \times n}, \quad F \in \mathbb{R}^{n \times q} \). The problem we consider here is to find the solution of the optimization problems (3) and (4) in an explicit form \( z^* = z^*(x(t)) \). Bemporad et al. (2002) showed that the solution \( z^*(x(t)) \) (and \( U^*(x(t)) \)) is a continuous PWA function defined over a polyhedral partition of the parameter space, and \( V_z(x(t)) \) is a convex (and therefore continuous) piecewise quadratic function.

3. Background on mp-QP

As shown in Bemporad et al. (2002), the mp-QP problems (3) and (4) can be solved by applying the Karush–Kuhn–Tucker (KKT) conditions

\[ H z + G^T \lambda = 0, \quad \lambda \in \mathbb{R}^q, \]

\[ \lambda_i (G^T z - W^T - S^T x) = 0, \quad i = 1, \ldots, q, \]

\[ \lambda \geq 0, \]

\[ G z - W - S x \leq 0. \]

For ease of notation we write \( x \) instead of \( x(t) \). In the sequel, let the superscript index denote a subset of the rows of a
matrix or vector. Since $H$ has full rank, (5) gives
\[ z = -H^{-1}G^T \lambda. \] (9)

**Definition 1.** Let $z^*(x)$ be the optimal solution to (3) and (4) for a given $x$. We define active constraints the constraints with $G^Tz^*(x) - W^i - S^i x = 0$, and inactive constraints the constraints with $G^Tz^*(x) - W^i - S^i x < 0$. The optimal active set $\mathcal{A}^*(x)$ is the set of indices of active constraints at the optimum. $\mathcal{A}^*(x) = \{i | G^Tz^*(x) = W^i + S^i x\}$. We also define as weakly active constraint an active constraint with an associated zero Lagrange multiplier $\lambda^i$, and as strongly active constraint an active constraint with a positive Lagrange multiplier $\lambda^i$.

Assume for the moment that we know the set $\mathcal{A}$ of constraints that are active at the optimum for a given $x$. We can now form matrices $G^d$, $W^d$, and $S^d$, and the Lagrange multipliers $\lambda^d \geq 0$, corresponding to the optimal active set $\mathcal{A}$.

**Definition 2.** For an active set, we say that the *linear independence constraint qualification* (LICQ) holds if the set of active constraint gradients are linearly independent, i.e., $G^d$ has full row rank.

Assuming that LICQ holds, (6) and (9) lead to
\[ \lambda^d = - (G^d H^{-1}(G^d)^T)^{-1}(W^d + S^d x). \] (10)

Eq. (10) can now be substituted into (9) to obtain
\[ z = H^{-1}(G^d)^T(G^d H^{-1}(G^d)^T)^{-1}(W^d + S^d x). \] (11)

We have now characterized the solution to (3) and (4) for a given optimal active set $\mathcal{A} \subseteq \{1, \ldots, q\}$, and a fixed $x$. However, as long as $\mathcal{A}$ remains the optimal active set in a neighborhood of $x$, the solution (11) remains optimal, when $z$ is viewed as a function of $x$. Such a neighborhood where $\mathcal{A}$ is optimal is determined by imposing that $z$ must remain feasible (8)
\[ GH^{-1}(G^d)^T(G^d H^{-1}(G^d)^T)^{-1}(W^d + S^d x) \leq W + Sx \] (12)

and that the Lagrange multipliers $\lambda$ must remain non-negative (7)
\[ - (G^d H^{-1}(G^d)^T)^{-1}(W^d + S^d x) \geq 0. \] (13)

Eqs. (12) and (13) describe a polyhedron in the state space. This region is denoted as the critical region $CR_0$ corresponding to the given set $\mathcal{A}$ of active constraints, is a convex polyhedral set, and represents the largest set of parameters $x$ such that the combination $\mathcal{A}$ of active constraints at the minimizer is optimal (Bemporad et al., 2002).

The recursive algorithm of (Bemporad et al., 2002) can be briefly summarized as follows: Solve an LP to find a feasible parameter $x_0 \in X$, where $X$ is the range of parameters for which the mp-QP is to be solved. Solve the QP (3) and (4) with $x = x_0$, to find the optimal active set $\mathcal{A}$ for $x_0$, and then use (10)–(13) to characterize the solution and critical region $CR_0$ corresponding to $\mathcal{A}$. Then divide the parameter space as in Fig. 1(b) and (c) by reversing one by one the hyperplanes defining the critical region. Iteratively subdivide each new region $R_i$ in a similar way as was done with $X$. The main drawback of this algorithm is that the regions $R_i$ are not related to optimality, as they can split some of the critical regions like $CR_i$ in Fig. 1(d). A consequence is that $CR_1$ will be detected at least twice.

The following theorem characterizes the primal and dual parametric solutions, and will be useful in the sequel.

**Theorem 1.** Consider Problems (3) and (4) with $H \succ 0$. Let $X \subset \mathbb{R}^n$ be a polyhedron. Then the solution $z^*(x)$ and the Lagrange multipliers $\lambda^*(x)$ of a mp-QP are piecewise affine, functions of the parameters $x$, and $z^*(x)$ is continuous. Moreover, if LICQ holds for all $x \in X$, $\lambda^*(x)$ is also continuous.

**Proof.** Follows easily from uniqueness (due to $H \succ 0$ and LICQ) of $z^*(x)$ and $\lambda^*(x)$, cf. Bemporad et al., (2002) and Fiacco (1983). \qed

4. Characterization of the partition

Below, we denote by $z^*_k(x)$ the linear expression of the PWA function $z^*(x)$ over the critical region $CR_k$.

**Definition 3.** Let a polyhedron $X \subset \mathbb{R}^n$ be represented by the linear inequalities $A_0 x \leq b$. Let the $i$th hyperplane $A_0 x = b^i$ be denoted by $H^i$. If $X \cap H^i$ is $(n-1)$-dimensional then $X \cap H^i$ is called a facet of the polyhedron.

**Definition 4.** Two polyhedra are called neighboring polyhedra if they have a common facet.
Definition 5. Let a polyhedron \( X \) be represented by \( A_0 x \leq b \). We say that \( A_0 x \leq b \) is redundant if \( A_j x \leq b^j \) \( \forall j \neq i \Rightarrow A_i x \leq b^i \) (i.e., it can be removed from the description of the polyhedron). The inequality \( i \) is redundant with degree \( h \) if it is redundant and there exists a \( h \)-dimensional subset \( Y \) of \( X \) such that \( A^i_X = b^i \) for all \( x \in Y \).

Definition 6. A representation of a polyhedron (12) and (13) is \( l \)-minimal if all redundant constraints have degree \( h \geq l \). It is minimal if there are no redundant constraints.

Clearly, a representation of a polyhedron \( X \subset \mathbb{R}^n \) is minimal if it contains all inequalities defining facets, and does not contain two or more coincident hyperplanes. Let us consider a hyperplane defining the common facet between two polyhedra \( CR_0 \), \( CR \) in the optimal partition of the state space. There are two different kinds of hyperplanes. The first (Type I) are those described by (12), which represents a non-active constraint of (4) that becomes active at the optimum as \( x \) moves from \( CR_0 \) to \( CR \). As proved in the following theorem, this means that if a polyhedron is bounded by a hyperplane which originates from (12), the corresponding constraint will be activated on the other side of the facet defined by this hyperplane. In addition, the corresponding Lagrange multiplier may become positive. The other kind (Type II) of hyperplanes which bound the polyhedra are those described by (13). In this case, the corresponding constraint will be non-active on the other side of the facet defined by this hyperplane.

Theorem 2. Consider an optimal active set \( \{i_1, i_2, \ldots, i_k\} \) and its corresponding minimal representation of the critical region \( CR_0 \) obtained by (12) and (13) after removing all redundant inequalities. Let \( CR \) be a full-dimensional neighboring critical region to \( CR_0 \) and assume LICQ holds on their common facet \( \mathcal{F} = CR_0 \cap \mathcal{H} \) where \( \mathcal{H} \) is the separating hyperplane between \( CR_0 \) and \( CR \). Moreover, assume that there are no constraints which are weakly active at the optimizer \( z^*(x) \) for all \( x \in CR_0 \). Then:

Type I: If \( \mathcal{H} \) is given by \( G^{i_1} z^*_0(x) = W^{i_1} + S^{i_1} x \), then the optimal active set in \( CR \) is \( \{i_1, \ldots, i_k, i_{k+1}\} \).

Type II: If \( \mathcal{H} \) is given by \( \lambda_0^j(x) = 0 \), then the optimal active set in \( CR \) is \( \{i_1, \ldots, i_k-1\} \).

Proof. Let us first prove Type I. In order for some constraint \( i_j \in \{i_1, \ldots, i_k\} \) not to be in the optimal active set in \( CR \), by continuity of \( \lambda^j(x) \) (due to Theorem 1 and LICQ), it follows that \( \lambda^j(x) = \lambda_0^j(x) = 0 \) for all \( x \in \mathcal{F} \). Since there are no constraints which are weakly active for all \( x \in CR_0 \), this would mean that constraint \( i_j \) becomes non-active at \( \mathcal{F} \). But this contradicts the assumption of minimality since \( \lambda_0^j(x) \geq 0 \) and \( G^{i_1} z^*_0(x) \leq W^{i_1} + S^{i_1} x \) would be coincident. On the other hand \( \{i_1, \ldots, i_k\} \) cannot be the optimal active set on \( CR \) because \( CR_0 \) is the largest set of \( x \)'s such that \( \{i_1, \ldots, i_k\} \) is the optimal active set. Then, the optimal active set in \( CR \) is a superset of \( \{i_1, \ldots, i_k\} \). Now assume that another constraint \( i_{k+2} \) is active in \( CR_i \). That means \( G^{i_{k+2}} z^*_0(x) = W^{i_{k+2}} + S^{i_{k+2}} x \) in \( CR_i \), and by continuity of \( z^*(x) \), the equality also holds for \( x \in \mathcal{F} \). However, \( G^{i_{k+2}} z^*_0(x) = W^{i_{k+2}} + S^{i_{k+2}} x \) would then coincide with \( G^{i_1} z^*_0(x) = W^{i_1} + S^{i_1} x \), which contradicts the assumption of minimality. Therefore, only \( \{i_1, \ldots, i_k, i_{k+1}\} \) can be the optimal active set in \( CR_i \). The proof for Type II is similar. □

Corollary 1. Consider the same assumptions as in Theorem 2, except that the assumption of minimality is relaxed into \((n-1)\)-minimality, i.e., two or more hyperplanes can coincide. Let \( \mathcal{I} \subset \{i_1, \ldots, i_k\} \) be the set of indices corresponding to coincident hyperplanes in the \((n-1)\)-minimal representation of (12) and (13) of \( CR_0 \).

- every constraint \( i_j \) where \( i_j \in \{i_1, i_2, \ldots, i_k\} \setminus \mathcal{I} \) is active in \( CR \),
- every constraint \( i_j \) where \( i_j \notin \{i_1, i_2, \ldots, i_k\} \cup \mathcal{I} \) is inactive in \( CR \).

We remark that coincident hyperplanes are rare, as from (12) and (13) one can see that special structures of \( H, F, G, W \), and \( S \) are required for two or more hyperplanes to be coincident. Anyway, when for instance two hyperplanes are coincident, by Corollary 1 there are three possible active sets which have to be checked to find the optimal active set in \( CR \).

One should always a priori remove redundant constraints from \( Gz-Sx \leq W \). This reduces the complexity of the mp-QP, and may also avoid some degeneracies (see Section 5).

Example 1. Consider the double integrator (Johansen et al., 2002)

\[
A = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} T_s^2 \\ T_s \end{bmatrix},
\]

where the sampling interval \( T_s = 0.05 \), and consider the MPC problem over the prediction horizon \( N=2 \) with cost matrices \( Q = \text{diag}([1 \, 0]) \), \( R = 1 \). The constraints in the system are \( -0.5 \leq x_2 \leq 0.5, -1 \leq u \leq 1 \). The mp-QP associated with this problem has the form (3) and (4) with

\[
H = \begin{bmatrix} 1.079 & 0.076 \\ 0.076 & 1.073 \end{bmatrix}, \quad F = \begin{bmatrix} 1.109 & 1.036 \\ 1.573 & 1.517 \end{bmatrix},
\]

\[
G^T = \begin{bmatrix} 1 & 0 & -1 & 0 & 0.05 & 0.05 & -0.05 & -0.05 \\ 0 & 1 & 0 & -1 & 0 & 0.05 & 0 & -0.05 \end{bmatrix},
\]

\[
W^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 0.5 & 0.5 & 0.5 \end{bmatrix},
\]
\[ S^T = \begin{bmatrix} 1.0 & 0.9 & -1.0 & -0.9 & 0.1 & 0.1 & -0.1 & -0.1 \\ 1.4 & 1.3 & -1.4 & -1.3 & -0.9 & -0.9 & 0.9 & 0.9 \end{bmatrix}. \]

We start the partitioning with the region where no constraints are active, which is full-dimensional because the mp-QP is created from an MPC problem where upper (lower) bounds on inputs and outputs are strictly positive (negative), see Bemporad and Filippi (2003, Lemma 6). This means that \( \mathcal{A}_0 = \emptyset \), and \( G^d, W^d \) and \( S^d \) are empty matrices, \( z^*(x) = 0 \) and the first component of \( U^*(x) \) is the unconstrained LQR law. This critical region is then described by \( 0 \leq W + Sx \), which contains eight inequalities. Two of these inequalities are redundant with degree 0 (#2 and #4), the remaining six hyperplanes are facet inequalities of the polyhedron (see Fig. 2a). By crossing the facet given by \( \mathcal{H}_1 \), defined by inequality 1 and of Type I, as predicted by Theorem 2 the optimal active set across this facet is \( \mathcal{A}_1 = \{1\} \), which leads to the critical region \( CR_1 \) (see Fig. 2b). After removing redundant inequalities we are left with a minimal representation of \( CR_1 \) containing four facets. The first of these is of Type II, \( \lambda^1(x) = 0 \). The other three are of Type I. These are inequalities #2, #6 and #7. Consider first the other side of the facet which comes from \( \lambda^1(x) = 0 \), see Fig. 2c. The region should not have constraint 1 active, so the optimal active set is \( \mathcal{A}_2 = \emptyset \). This is the same combination of active constraints as \( \mathcal{A}_0 \), as expected, so \( \mathcal{A}_2 \) is not pursued. Next, consider crossing the respective facets of inequalities 2, 6 and 7, see Figs. 2d–f. This results in three different active sets: \( \mathcal{A}_3 = \{1, 2\} \), \( \mathcal{A}_4 = \{1, 6\} \) and \( \mathcal{A}_5 = \{1, 7\} \). \( \mathcal{A}_3 \) and \( \mathcal{A}_4 \) lead to new polyhedra as shown in the figures. The combination \( \mathcal{A}_5 \) leads to an interesting case of “degeneracy”. The associated matrix \( G^d \) has linearly dependent rows, which violates the LICQ assumption. In this case, \( \mathcal{A}_5 \) leads to an infeasible part of the state space. A general treatment of degeneracy is given in the next section.

Theorem 2 and Corollary 1 show how to find the optimal active set across a facet only by using the knowledge of which kind of hyperplane the facet corresponds to, except in degenerate cases, which is the topic of the next section.

5. Degeneracy in mp-QP

We have so far assumed that LICQ holds on the common facet between two polyhedra, and that there are no constraints which are weakly active for all \( x \) within a critical region. Such cases are referred to as degenerate. We will first consider how to handle cases where LICQ is violated, and then consider weakly active constraints.

Theorem 3. Consider a generic combination \( \mathcal{A} \subseteq \{1, \ldots, q\} \) of active constraints and assume that the corresponding rows \( \{G^{\mathcal{A}}|S^{\mathcal{A}}|W^{\mathcal{A}}\} \) are linearly independent. If LICQ is violated, then the corresponding critical region is not full-dimensional.

Proof. Let the active constraints be \( G^d z = S^d x + W^d \). Since LICQ is violated, \( G^d \) has not full rank and a reduced set of equations can be defined without changing the solution \( z^*(x) = G^d x + W^d \). Assume without loss of generality that

\[ G^d = \begin{bmatrix} G' \\ G^k \end{bmatrix}, \quad S^d = \begin{bmatrix} S' \\ S^k \end{bmatrix}, \quad W^d = \begin{bmatrix} W' \\ W^k \end{bmatrix}, \]

where \( G' \) and \( S' \) are row-vectors and \( W^k \) is a scalar. Let \( CR' \) and \( CR^k \) be the critical regions where the active sets corresponding to \( G' \) and \( G^k \), respectively, are optimal. The solution is \( z_{CR'}^{*}(x) = z_{CR^k}(x) = Lx + v \) within both \( CR' \) and \( CR^k \), where \( L = H^{-1}GT(G'HT^{-1}G'T)^{-1}S' \), \( v = H^{-1}GT(G'HT^{-1}G'T)^{-1}W' \). It is clear that \( CR^k \subseteq \{x \in \mathbb{R}^n | [G^d - S^d][z_{CR'}^{*}(x)] = W^d\} = M \) and

\[
M = \left\{ x \in \mathbb{R}^n \left| \begin{bmatrix} G' & -S' \\ G^k & -S^k \end{bmatrix} \begin{bmatrix} Lx + v \\ x \end{bmatrix} = \begin{bmatrix} W' \\ W^k \end{bmatrix} \right. \right\} = \left\{ x \in \mathbb{R}^n \left| (G^d L - S^d)x + G^k v = W^k \right. \right\}.
\]

If \( G^k L \neq S^k \) or \( G^k v \neq W^k \) it follows that \( M \) is not a full-dimensional subspace of \( \mathbb{R}^n \), and since \( CR^k \subseteq M \), neither is \( CR^k \). Suppose this does not hold, i.e., \( G^k L = S^k \) and \( G^k v = W^k \). Since \( G^d \) has not full rank, \( G^k = zG' \), where \( z \) is
a row-vector, and \( S^T = xG' H^{-1} G'^T (G'H^{-1} G'^T)^{-1} S' = x S' \), \( W^k = x G' H^{-1} G'^T (G'H^{-1} G'^T)^{-1} W' = x W' \). Then, there is linear dependence between rows of \([G'S' - S'W']\), a contradiction.

In an MPC problem one might avoid full-dimensional critical regions with violation of LICQ by simply slightly perturbing the weight matrices and the constraints, without producing significant changes of the closed-loop behavior. On the other hand, in some situations this may not be possible, for instance equality constraints such as terminal state constraints \( x_{n+1} = 0 \), would lead to violation of LICQ (cf. Berkelaar et al., 1997, Example 6.3). In such cases, full-dimensional critical regions can be handled by solving a QP, as in Tøndel, Johansen, and Bemporad (2001b).

Next Theorem 4 provides a method to find the optimal active set in a neighboring region also when LICQ is violated on the common facet. Before proceeding further, we need a technical Lemma.

**Lemma 1.** Let the optimal active set in a critical region \( CR_0 \) be \( \{i_1, \ldots, i_k\} \), and consider an minimal representation of \( CR_0 \). Assume that there are no constraints which are weakly active for all \( x \in CR_0 \) and that \( G^{\{i_1, \ldots, i_k\}} \) does not have linearly dependent rows. Let \( CR_i \) be a full-dimensional neighboring critical region to \( CR_0 \), and let \( \mathcal{F} \) be their common facet with \( \mathcal{F} = CR_0 \cap \mathcal{H} \) and \( \mathcal{H} \) is the Type I hyperplane \( G^{\{i_1, \ldots, i_k\}} \mathbf{z}_i(x) = W^{i+1} + S^{i+1} \). Suppose \( G^{\{i_1, \ldots, i_k, i_{k+1}\}} \) has linearly dependent rows, such that LICQ is violated at \( \mathcal{F} \). Then, if there is a feasible solution in \( CR_i \), the optimal active set in \( CR_i \) consists of constraint \( i_{k+1} \) and some subset of \( \{i_1, \ldots, i_k\} \).

**Proof.** Clearly, \( G^{\{i_1, \ldots, i_k, i_{k+1}\}} \) collects the active constraints at the optimal solution for \( x \in \mathcal{F} \). Consider now vectors \( x \) in the interior of \( CR_i \). The active sets \( \{i_1, \ldots, i_{k+1}\} \) or any active set including constraint \( i_{k+2} \) can be excluded using similar arguments as in the proof of Theorem 2. Next, assume that \( \{i_1, \ldots, i_{k-1}\} \) is the optimal active set in \( CR_i \). The KKT conditions (5)–(8) together with the full row rank of \( G^{\{i_1, \ldots, i_k\}} \) gives that for each \( x \in CR_i \) (also on the facets) there is a unique solution to

\[
Hz + (G^{\{i_1, \ldots, i_k\}}) \lambda_{\{i_1, \ldots, i_k\}} = 0,
\]

\[
\lambda_{\{i_1, \ldots, i_{k-1}\}} > 0, \quad \lambda_i = 0 \quad \text{for all } j \notin \{i_1, \ldots, i_{k-1}\}.
\]  

(16)

Note that this solution still is unique at \( \mathcal{F} \), since \( \lambda_i = 0 \) for all \( i_j \notin \{i_1, \ldots, i_{k-1}\} \). But in \( CR_0 \) we have a unique solution to

\[
Hz + (G^{\{i_1, \ldots, i_k\}}) \lambda_{\{i_1, \ldots, i_k\}} = 0,
\]

\[
\lambda_{\{i_1, \ldots, i_{k-1}\}} > 0, \lambda_i = 0 \quad \text{for all } i_j \notin \{i_1, \ldots, i_k\}.
\]  

(17)

Due to continuity both of these solutions are valid on \( \mathcal{F} \). This is a contradiction because the solutions are unique, while we require \( \lambda_i = 0 > 0 \).

**Theorem 4.** Make the same assumptions as in Lemma 1. Consider the following LP:

\[
\begin{align*}
\max_{\{i_1, \ldots, i_{k+1}\}} & \quad \lambda_{i_{k+1}} \\
\text{s.t.} & \quad Hz(x_0) + (G^{\{i_1, \ldots, i_k, i_{k+1}\}}) \lambda_{\{i_1, \ldots, i_k, i_{k+1}\}} = 0, \\
& \quad \lambda_{\{i_1, \ldots, i_k, i_{k+1}\}} \geq 0
\end{align*}
\]

(18)

for some \( x_0 \) on \( \mathcal{F} \). If this LP has a bounded solution, the optimal active set in \( CR_i \) consists of the elements of \( \{i_1, \ldots, i_k, i_{k+1}\} \) with \( \lambda_i > 0 \) in the solution. If the LP is unbounded, \( CR_i \) is an infeasible area of the parameter space.

**Proof.** The solution \( z^*(x) \) to (5)–(8) on \( \mathcal{F} \) is known from the solution in \( CR_0 \). The optimal Lagrange multipliers \( \lambda^*(x) \) on \( \mathcal{F} \) is then characterized by (19) and (20). The solution to (5)–(8) in \( CR_i \) must be valid also on \( \mathcal{F} \), in particular, \( \lambda^*_i(x) \) must satisfy (19) and (20) on \( \mathcal{F} \). From Lemma 1, the optimal active set in \( CR_i \), consists of constraint \( i_{k+1} \) and a proper subset of \( \{i_1, \ldots, i_k\} \). Therefore, there must be a solution on \( \mathcal{F} \) which satisfies (\( \lambda^*_i(x) \geq 0 \) and (\( \lambda^*_i(x) \) is free for at least one \( i_j \in \{i_1, \ldots, i_k\} \). With a fixed \( \lambda_{i_{k+1}} = 0 \), (19) defines \( n_z \) equations in \( k \) unknowns \( (n_z \geq k) \). But there exists a solution from \( CR_0 \), such that a reduced set of equations can be defined with \( k \) equations in \( k \) unknowns. When \( \lambda_{i_{k+1}} > 0 \) (19) consists of \( k \) equations in \( k + 1 \) unknowns, and \( \lambda_i = f^{i}_{j}(\lambda_{i_{k+1}}) \) for any \( i_j \in \{i_1, \ldots, i_k\} \), where \( f^{i}_{j} \) is an affine function. When \( \lambda_{i_{k+1}} = 0 \), the solution of (19) and (20) has \( \lambda_i > 0 \) for all \( i_j \in \{i_1, \ldots, i_k\} \) (due to minimality and no weakly active constraints for all \( x \) in \( CR_0 \)). To find a solution which satisfies Lemma 1, \( \lambda_{i_{k+1}} \) must be increased from zero until \( \lambda_i = 0 \) for some \( i_j \in \{i_1, \ldots, i_k\} \). This is the only solution of (19) and (20) which satisfies Lemma 1 because if \( \lambda_{i_{k+1}} \) is increased further, \( \lambda_i = f^{i}_{j}(\lambda_{i_{k+1}}) < 0 \) (since \( f^{i}_{j} \) is an affine function).

Constraints that are weakly active for all \( x \) in a critical region, can be handled according to the following result, which can be proven similarly to Theorem 2.

**Theorem 5.** Make the same assumptions as in Theorem 2, except that now constraint \( i_1 \) is weakly active for all \( x \in CR_0 \).

**Type I:** If \( \mathcal{H} \) is given by \( G^{i_{k+1}} \mathbf{z}_0(x) = W^{i+1} + S^{i+1} x \), then the optimal active set in \( CR_i \) is \( \{i_1, \ldots, i_k, i_{k+1}\} \) or \( \{i_2, \ldots, i_k, i_{k+1}\} \).

**Type II:** If \( \mathcal{H} \) is given by \( \lambda^*_0(x) = 0 \), then the optimal active set in \( CR_i \) is \( \{i_1, \ldots, i_k, i_{k+1}\} \) or \( \{i_2, \ldots, i_k, i_{k+1}\} \).
Example 1 (Continued). We want to show how to handle the case when LICQ is violated at a facet. First, notice in Fig. 2 that the polyhedra made from \( \mathcal{A}_3 \) and \( \mathcal{A}_4 \) are neighboring polyhedra, but still there are two elements in \( \mathcal{A}_3 \) which are different from \( \mathcal{A}_4 \). This is caused by a violation of LICQ on the hyperplane separating these regions. Assume we have found \( CR_3 \), and try to detect \( CR_4 \). We cross a hyperplane of Type 1, which defines their common facet \( \mathcal{F} \). This hyperplane says that constraint 6 is becoming active at the optimal solution for \( x \in \mathcal{F} \). Since constraints 1 and 2 was active in \( LP(18)-(20) \), and in this case we use \( Filippi, 2003, Lemma 6 \).

6. O7VT-line mp-QP algorithm

We can choose the active set which is optimal in a full-dimensional region to most active sets are not optimal anywhere). We need an efficient algorithm for the computation of the solution to the type of hyperplane \( \mathcal{F} \) is given by.

Algorithm 1.

Choose the initial active set \( \mathcal{A}_0 \) as in (Bemporad et al., 2003, Proposition 2); Let \( L_{\text{cand}} \leftarrow \{ \mathcal{A}_0 \} \), \( L_{\text{opt}} \leftarrow \emptyset \);

while \( L_{\text{cand}} \neq \emptyset \) do

Pick an element \( \mathcal{A} \) from \( L_{\text{cand}} \), \( L_{\text{cand}} \leftarrow L_{\text{cand}} \setminus \{ \mathcal{A} \} \);

Find the CR where \( \mathcal{A} \) is optimal from (12) and (13), and the solution \( z(x) \) from (10) and (9);

end while

An estimate of the cost for solving the mp-QP (3) and (4) by different algorithms is given below. This estimate is given by the number of LPs/QPs which has to be solved, as this is the main cost. For Algorithm 1 this is given by

\[
\frac{\text{Final # regions found by the algorithm}}{\text{# LPs per region for redundancy check}} \times \frac{\text{# LPs per region for redundancy check}}{\text{# LPs for red. check}} \times \frac{\text{Total # regions explored}}{\text{# LPs for red. check}} \times \left( +1 \text{ LP to find interior point} \right) + 1 \text{ QP to find active set}.
\]

Consequently, the difference between the two algorithms is the last term, which is due to the extra partitioning into regions \( R_i \), as in Fig. 1. The removal of redundant constraints
from polyhedra is done by solving one LP for each hyperplane. The cost of the algorithm of (Seron et al., 2000) which only handles input constraints, is

\[ 3^n \times \left( \text{# LPs per region for redundancy check} \right) \].

**Example 2.** We compare the efficiency of Algorithm 1, the algorithm of Bemporad et al. (2002) and the algorithm of Seron et al. (2000) on the double integrator example from Bemporad et al. (2002) in Table 1. All the computation times are achieved on a 650 MHz Pentium III running Matlab 5.3, using the NAG Foundation Toolbox to solve LP/QP subproblems. In this example, both Algorithm 1 and the algorithm of Bemporad et al. (2002) spend more than 60% of the time on removing redundant constraints from the polyhedra, according to the previous complexity analysis. Note that symmetries of this MPC problem could be exploited to further decrease computation times.

**Example 3.** The laboratory model helicopter (Quanser 3-DOF Helicopter) described in Tøndel and Johansen (2002) sampled with \( T = 0.01 \) s, and the following state space representation is obtained

\[
A = \begin{bmatrix}
1 & 0 & 0.01 & 0 & 0 & 0 \\
0 & 1 & 0 & 0.01 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0.01 & 0 & 0 & 0 & 1 & 0 \\
0 & 0.01 & 0 & 0 & 0 & 1 \\
\end{bmatrix},
\]

The states of the system are:

- \( x_1 \) —elevation,
- \( x_2 \) —pitch angle,
- \( x_3 \) —elevation rate,
- \( x_4 \) —pitch angle rate,
- \( x_5 \) —integral of elevation error,
- \( x_6 \) —integral of pitch angle error.

The inputs to the system are:

- \( u_1 \) —front rotor power,
- \( u_2 \) —rear rotor power.

The system is to be regulated to the origin with the following constraints on the inputs and pitch and elevation rates:

\[-1 \leq u_1 \leq 3, \]
\[-1 \leq u_2 \leq 3, \]
\[-0.44 \leq x_3 \leq 0.44, \]
\[-0.6 \leq x_4 \leq 0.6. \]

The LQ cost function is given by

\[ Q = \text{diag}(100, 100, 10, 10, 400, 200), \]

\[ R = I_{2 \times 2} \]

and \( P \) is given by the algebraic Riccati equation.

The system is optimized with a horizon of 50 samples, and as is common in MPC implementations, input parameterization has been used to reduce the dimensions of the optimization problem. Table 2 shows the number of regions in the partition and computation times using 1–4 parameters to describe the control input.

**7. Conclusions**

In this paper, we have proposed a new approach for solving mp-QP problems giving off-line piecewise affine explicit
solutions to MPC control problems. Being based on the exploitation of direct relations between neighboring polyhedral regions and combinations of active constraints, we believe that our contribution significantly advances the field of explicit MPC control, both theoretically and practically, as examples have indicated large improvements of computational efficiency over existing mp-QP algorithms.

References


