An Algorithm for Multi-Parametric Quadratic Programming and Explicit MPC Solutions

Petter Tøndel [a], Tor Arne Johansen [a], Alberto Bemporad [b,c]

[a] *Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7491 Trondheim, Norway*

[b] *Dipartimento di Ingegneria dell'Informazione, University of Siena, 53100 Siena, Italy*

[c] *ETH Zentrum, 8092 Zurich, Switzerland*

Abstract

Explicit solutions to constrained linear MPC 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.

Key words: 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 μ-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 on-line implementation of PWA function evaluation in explicit MPC has been developed by exploiting convexity (Borrelli *et al.*, 2001) or an associated binary search tree data structure (Tøndel and Johansen, 2002; Tøndel *et al.*, submitted). Independent works by (Bemporad *et al.*, 2002), (Bemporad *et al.*, 2000*b*), (Johansen *et al.*, 2000) and (Seron *et al.*, 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) and (Bemporad *et al.*, 2000*b*) 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 *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 *et al.*, 1997). Multi-parametric LP (mp-LP) is treated in (Gal, 1995) and (Borrelli *et al.*, 2000), mp-LP in connection with MPC based on linear programming is investigated in (Bemporad *et al.*, accepted for publication), and multi-parametric mixed-integer linear programming (Dua and Pistikopoulos, to appear) is used in (Bemporad *et al.*, 2000*a*) 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.*, 1999; Bemporad and Filippi, Conditionally accepted for publication with minor revisions), and efficient on-line com-

⋆ This paper was not presented at any IFAC meeting. Corresponding author P. Tøndel. Tel. +47 7359 4348. Fax +47 7359 4399 .

Email addresses: Petter.Tondel@itk.ntnu.no (Petter Tøndel), Tor.Arne.Johansen@itk.ntnu.no (Tor Arne Johansen), bemporad@dii.unisi.it (Alberto Bemporad).

putation 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 *et al.*, 2001*a*).

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

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t), \end{aligned} \tag{1}$$

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

$$\begin{aligned} \min_U \quad & \{J(U,x(t)) = x^T_{t+N|t}Px_{t+N|t} \\ & + \textstyle\sum_{k=0}^{N-1} x^T_{t+k|t}Qx_{t+k|t} + u^T_{t+k}Ru_{t+k}\} \\ \text{s.t.} \quad & y_{\min} \le y_{t+k|t} \le y_{\max}, k=1,...,N \\ & u_{\min} \le u_{t+k} \le u_{\max}, k=0,...,M-1 \\ & u_{t+k} = Kx_{t+k|t},\ M \le k \le N-1 \\ & x_{t|t} = x(t) \\ & x_{t+k+1|t} = Ax_{t+k|t} + Bu_{t+k},\ k \ge 0 \\ & y_{t+k|t} = Cx_{t+k|t},\ k \ge 0 \end{aligned} \tag{2}$$

with respect to $U \triangleq [u_t^T,...,u^T_{t+M-1}]^T$, where $y_{\min} < 0 < y_{\max}$, $u_{\min} < 0 < u_{\max}$, $R = R' > 0$, $Q = Q' \ge 0$, $P = P' > 0$ and $x_{k+t|t}$ is the prediction of x_{t+k} at time t. 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 $k \ge N$, (2) exactly solve the constrained (infinite-horizon) LQR problem for (1) with weights Q, R (see also (Sznaier and Damborg, 1987), (Chmielewski and Manousiouthakis, 1996) and (Scokaert and 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

$$V_z(x(t)) = \min_z \frac{1}{2} z^T H z \tag{3}$$

$$s.t.\ Gz \le W + Sx(t), \tag{4}$$

where $z \triangleq U + H^{-1}F^Tx(t)$ and $x(t)$ is the current state, which can be treated as a vector of parameters. Note that $H \succ 0$ since $R \succ 0$. The number of inequalities is denoted by q and the number of free variables is $n_z = m \cdot N$. Then $z \in \mathbb{R}^{n_z}$, $H \in \mathbb{R}^{n_z \times n_z}$, $G \in \mathbb{R}^{q\times n_z}$, $W \in \mathbb{R}^{q\times 1}$, $S \in \mathbb{R}^{q\times n}$, $F \in \mathbb{R}^{n\times q}$. The problem we consider here is to find the solution of the optimization problem (3)–(4) in an explicit form $z^* = z^*(x(t))$. Bemporad et. al. (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 problem (3)-(4) can be solved by applying the Karush-Kuhn-Tucker (KKT) conditions

$$Hz + G^T\lambda = 0,\ \lambda \in \mathbb{R}^q, \tag{5}$$

$$\lambda_i \left(G^iz - W^i - S^ix\right) = 0,\ i = 1,...,q, \tag{6}$$

$$\lambda \ge 0, \tag{7}$$

$$Gz - W - Sx \le 0. \tag{8}$$

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

$$z = -H^{-1}G^T\lambda. \tag{9}$$

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

Assuming that LICQ holds, (6) and (9) lead to

$$\lambda^\mathcal{A} = -(G^\mathcal{A}H^{-1}(G^\mathcal{A})^T)^{-1}(W^\mathcal{A} + S^\mathcal{A}x). \tag{10}$$

Eq. (10) can now be substituted into (9) to obtain

$$z = H^{-1}(G^\mathcal{A})^T(G^\mathcal{A}H^{-1}(G^\mathcal{A})^T)^{-1}(W^\mathcal{A} + S^\mathcal{A}x). \tag{11}$$

We have now characterized the solution to (3)-(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)

$$GH^{-1}(G^{\mathcal{A}})^T(G^{\mathcal{A}}H^{-1}(G^{\mathcal{A}})^T)^{-1}(W^{\mathcal{A}}+S^{\mathcal{A}}x) \leq W+Sx, \tag{12}$$

and that the Lagrange multipliers λ must remain non-negative (7)

$$-(G^{\mathcal{A}}H^{-1}(G^{\mathcal{A}})^T)^{-1}(W^{\mathcal{A}} + S^{\mathcal{A}}x) \geq 0. \tag{13}$$

Equations (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)-(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 Figure 1b-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_1 in Figure 1d. A consequence is that CR_1 will be detected at least twice.

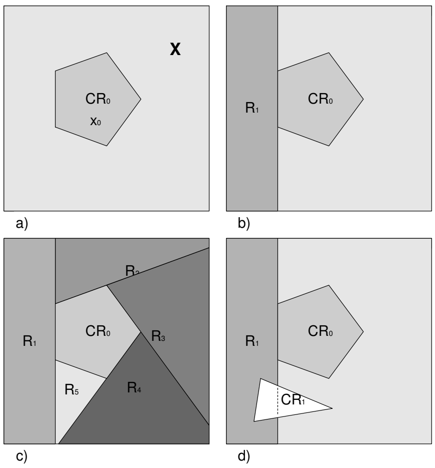

Fig. 1. State space exploration strategy of (Bemporad *et al.*, 2002).

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

Theorem 1 *Consider Problem (3)-(4) with $H \succ 0$. Let $X \in \mathbb{R}^n$ 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 $H > 0$ and LICQ) of $z^*(x)$ and $\lambda^*(x)$, cf. (Bemporad *et al.*, 2002), (Fiacco, 1983). □

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 $A_0^i x = b^i$ be denoted by $\mathcal{H}$. If $X \cap \mathcal{H}$ is $(n-1)$-dimensional then $X \cap \mathcal{H}$ 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 $A_0^i x \leq b^i$ is **redundant** if $A_0^j x \leq b^j \ \forall j \neq i \Rightarrow A_0^i x \leq b^i$ (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 $A_0^i x = b^i$ for all $x \in Y$.*

Definition 6 *A representation of a polyhedron (12)-(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_i 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_i. 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 $\{i_1, i_2, ..., i_k\}$ and its corresponding minimal representation of the critical region CR_0 obtained by (12)-(13) after removing all redundant inequalities. Let CR_i 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_i. 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 $G^{i_{k+1}} z_0^*(x) = W^{i_{k+1}} + S^{i_{k+1}} x$, then the optimal active set in CR_i is $\{i_1, ..., i_k, i_{k+1}\}$.*

Type II *If $\mathcal{H}$ is given by $\lambda_0^{i_k}(x) = 0$, then the optimal active set in CR_i is $\{i_1, ..., i_{k-1}\}$.*

Proof: Let us first prove Type I. In order for some constraint $i_j \in \{i_1, ..., i_k\}$ not to be in the optimal active set in CR_i, by continuity of $\lambda^*(x)$ (due to Theorem 1 and LICQ), it follows that $(\lambda^*)^{i_j}(x) = \lambda_0^{i_j}(x) = 0$ 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 $\lambda_0^{i_j}(x) \geq 0$ and $G^{i_{k+1}} z_0^*(x) \leq W^{i_{k+1}} + S^{i_{k+1}} x$ would be coincident. On the other hand $\{i_1, ..., i_k\}$ can not be the optimal active set on CR_i because CR_0 is the largest set of x's such that $\{i_1, ..., i_k\}$ is the optimal active set. Then, the optimal active set in CR_i is a superset of $\{i_1, ..., i_k\}$. Now assume that another constraint i_{k+2} is active in CR_i. That means $G^{i_{k+2}} z_i^*(x) = W^{i_{k+2}} + S^{i_{k+2}} x$ 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 $G^{i_{k+1}} z_0^*(x) = W^{i_{k+1}} + S^{i_{k+1}} x$, which contradicts the assumption of minimality. Therefore, only $\{i_1, ..., i_k, i_{k+1}\}$ 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 $\mathcal{I} \subset \{i_1, ..., i_k\}$ be the set of indices corresponding to coincident hyperplanes in the $(n-1)$-minimal representation of (12)-(13) of CR_0.*

- *Every constraint i_j where $i_j \in \{i_1, i_2, ..., i_k\} \setminus \mathcal{I}$ is active in CR_i.*
- *Every constraint i_j where $i_j \notin \{i_1, i_2, ..., i_k\} \cup \mathcal{I}$ is inactive in CR_i.*

We remark that coincident hyperplanes are rare, as from (12)-(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_i.

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.*, 2000)

$$A = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} T_s^2 \\ T_s \end{bmatrix}$$

where the sampling interval $T_s = 0.05$, and consider the MPC problem over the prediction horizon $N = 2$ with cost matrices $Q = diag([1 \;\; 0])$, $R = 1$. 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)-(4) with

$$H = \begin{bmatrix} 1.079 & 0.076 \\ 0.076 & 1.073 \end{bmatrix}, \; F = \begin{bmatrix} 1.109 & 1.036 \\ 1.573 & 1.517 \end{bmatrix},$$

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

$$W^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 0.5 & 0.5 & 0.5 & 0.5 \end{bmatrix}, \; S^T =$$

$$\begin{bmatrix} 1.0 & 0.9 & -1.0 & -0.9 & 0.1 & 0.1 & -0.1 & -0.1 \\ 1.4 & 1.3 & -1.4 & -1.3 & -0.9 & -0.9 & 0.9 & 0.9 \end{bmatrix}$$

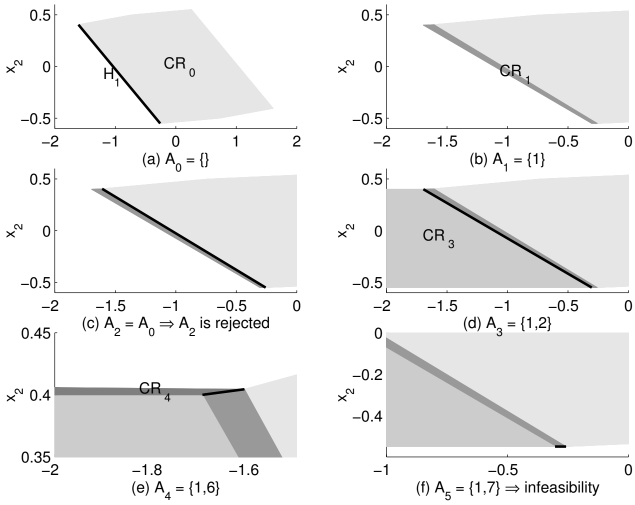

Fig. 2. Critical regions for double integrator

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, Conditionally accepted for publication with minor revisions, Lemma 6). This means that $\mathcal{A}_0 = \emptyset$, and $G^{\mathcal{A}}$, $W^{\mathcal{A}}$ and $S^{\mathcal{A}}$ 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 Figure 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 Figure 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 Figure 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 Figures 2d–2f. 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 $G^{\mathcal{A}}$ 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 $[G^{\mathcal{A}} \mid -S^{\mathcal{A}} \mid -W^{\mathcal{A}}]$ are linearly independent. If LICQ is violated, then the corresponding critical region is not full-dimensional.*

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

$$G^{\mathcal{A}} = \begin{bmatrix} G' \\ G^k \end{bmatrix}, \; S^{\mathcal{A}} = \begin{bmatrix} S' \\ S^k \end{bmatrix}, \; W^{\mathcal{A}} = \begin{bmatrix} W' \\ W^k \end{bmatrix}, \qquad (14)$$

where G^k, and S^k are row-vectors and W^k is a scalar. Let CR' and $CR^{\mathcal{A}}$ be the critical regions where the active sets corresponding to G' and $G^{\mathcal{A}}$, respectively, are optimal. The solution is $z^*_{CR'}(x) = z^*_{CR^{\mathcal{A}}}(x) = Lx + v$ within both CR' and $CR^{\mathcal{A}}$, where $L = H^{-1}G'^T(G'H^{-1}G'^T)^{-1}S'$, $v = H^{-1}G'^T(G'H^{-1}G'^T)^{-1}W'$. It is clear that $CR^{\mathcal{A}} \subset \left\{ x \in \mathbb{R}^n \mid [G^{\mathcal{A}} \;\; -S^{\mathcal{A}}] \begin{bmatrix} z^*_{CR^{\mathcal{A}}}(x) \\ x \end{bmatrix} = W^{\mathcal{A}} \right\} = M$ and

$$\begin{aligned} M &= \left\{ x \in \mathbb{R}^n \mid \begin{bmatrix} G' & -S' \\ G^k & -S^k \end{bmatrix} \begin{bmatrix} Lx + v \\ x \end{bmatrix} = \begin{bmatrix} W' \\ W^k \end{bmatrix} \right\} \\ &= \left\{ \begin{array}{l} x \in \mathbb{R}^n \mid G'Lx - S'x + G'v = W' \\ \quad ,(G^kL - S^k)x + G^kv = W^k \end{array} \right\} \qquad (15) \\ &= \left\{ x \in \mathbb{R}^n \mid (G^kL - S^k)x + G^kv = W^k \right\}. \end{aligned}$$

If $G^kL \neq S^k$ or $G^kv \neq W^k$ it follows that M is not a full-dimensional subspace of $\mathbb{R}^n$, and since $CR^{\mathcal{A}} \subset M$, neither is $CR^{\mathcal{A}}$. Suppose this does not hold, i.e., $G^kL = S^k$ and $G^kv = W^k$. Since $G^{\mathcal{A}}$ has not full rank, $G^k = \alpha G'$, where α is a row-vector, and $S^k = \alpha G'H^{-1}G'^T(G'H^{-1}G'^T)^{-1}S' = \alpha S'$, $W^k = \alpha G'H^{-1}G'^T(G'H^{-1}G'^T)^{-1}W' = \alpha W'$. Then, there is linear dependence between rows of $[G^{\mathcal{A}} \mid -S^{\mathcal{A}} \mid -W^{\mathcal{A}}]$, 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 $x_{t+N|t} = 0$, 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 *et al.*, 2001*b*).

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 $\{i_1, ..., i_k\}$, 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 $G^{i_{k+1}}z^*_0(x) = W^{i_{k+1}} + S^{i_{k+1}}x$. 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 $\{i_1, ..., i_k\}$.*

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 $\{i_1, ..., i_{k+1}\}$ 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 $\{i_1, ..., i_{k-1}\}$ is the optimal active set in CR_i. The KKT conditions (5)-(8) together with the full row rank of $G^{\{i_1,\ldots,i_{k-1}\}}$ gives that for each $x \in CR_i$ (also on the facets) there is a unique solution to

$$Hz + \left(G^{\{i_1,\ldots,i_{k-1}\}}\right)^T \lambda^{\{i_1,\ldots,i_{k-1}\}} = 0, \quad (16)$$
$$\lambda^{\{i_1,\ldots,i_{k-1}\}} > 0, \; \lambda^j = 0 \text{ for all } j \notin \{i_1, ..., i_{k-1}\}$$

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

$$Hz + \left(G^{\{i_1,\ldots,i_k\}}\right)^T \lambda^{\{i_1,\ldots,i_k\}} = 0, \qquad (17)$$
$$\lambda^{\{i_1,\ldots,i_k\}} > 0, \; \lambda^{i_j} = 0 \text{ for all } i_j \notin \{i_1, ..., i_k\}$$

Due to continuity both of these solutions are valid on $\mathcal{F}$. This is a contradiction because the solutions are unique, while we require $\lambda^{i_k} = 0$ and $\lambda^{i_k} > 0$. The only remaining possibility is that the optimal active set in CR_i consists of i_{k+1} and a subset of $\{i_1, ..., i_k\}$. □

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

$$\max_{\lambda^{\{i_1,\cdots,i_k,i_{k+1}\}}} \lambda^{i_{k+1}} \quad (18)$$

$$s.t.\ Hz(x_0) + \left(G^{\{i_1,\cdots,i_k,i_{k+1}\}}\right)^T \lambda^{\{i_1,\cdots,i_k,i_{k+1}\}} = 0 \quad (19)$$

$$\lambda^{\{i_1,\cdots,i_k,i_{k+1}\}} \geq 0, \quad (20)$$

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 $\{i_1, ..., i_k, i_{k+1}\}$ with $\lambda^{i_j} > 0$ 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)-(20). The solution to (5)-(8) in CR_i must be valid also on $\mathcal{F}$, in particular, $\lambda_i^*(x)$ must satisfy (19)-(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 $\{i_1, ..., i_k\}$. Therefore, there must be a solution on $\mathcal{F}$ which satisfies $\left(\lambda_i^{i_{k+1}}\right)^* (x) > 0$ and $\left(\lambda_i^{i_j}\right)^* (x) = 0$ for at least one $i_j \in \{i_1, ..., i_k\}$. 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 $\lambda^{i_{k+1}} \geq 0$, (19) consists of k equations in $k+1$ unknowns, and $\lambda^{i_j} = f^{i_j}(\lambda^{i_{k+1}})$ for any $i_j \in \{i_1, ..., i_k\}$, where f^{i_j} is an affine function. When $\lambda^{i_{k+1}} = 0$, the solution of (19)-(20) has $\lambda^{i_j} > 0$ for all $i_j \in \{i_1, ..., i_k\}$ (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 $\lambda^{i_j} = 0$ for some $i_j \in \{i_1, ..., i_k\}$. This is the only solution of (19)-(20) which satisfies Lemma 1 because if $\lambda^{i_{k+1}}$ is increased further, $\lambda^{i_j} = f^{i_j}(\lambda^{i_{k+1}}) < 0$ (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 $G^{i_{k+1}} z_0^*(x) = W^{i_{k+1}} + S^{i_{k+1}} x$, then the optimal active set in CR_i is $\{i_1, ..., i_k, i_{k+1}\}$ or $\{i_2, ..., i_k, i_{k+1}\}$.*

Type II *If $\mathcal{H}$ is given by $\lambda_0^{i_k}(x) = 0$, then the optimal active set in CR_i is $\{i_1, ..., i_{k-1}\}$ or $\{i_2, ..., i_{k-1}\}$.*

Example 1 (cont'd). We want to show how to handle the case when LICQ is violated at a facet. First, notice in Figure 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 constraint 1 and 2 was active in CR_3, constraints $\{1, 2, 6\}$ are active at the optimal solution for $x \in \mathcal{F}$. This obviously leads to linear dependence among the elements in $G^{\mathcal{A}}$, and Theorem 4 is applied to find the optimal active set across $\mathcal{F}$. A point $x_0 \in \mathcal{F}$ is needed to initialize the LP (18)-(20), and in this case we use $x_0 = [-1.8\ 0.4]^T$. We then solve the LP (18)-(20): $\max_{\lambda^{\{1,2,6\}}} \lambda^6,\ s.t.\ Hz(x_0) + \left(G^{\{1,2,6\}}\right)^T \lambda^{\{1,2,6\}} = 0,\ \lambda^{\{1,2,6\}} \geq 0$. The solution of this LP is $\lambda^{\{1,2,6\}} = \begin{bmatrix} 0.11 & 0 & 4.25 \end{bmatrix}^T$. Hence, λ^2 should be removed from the active set, and the optimal active set in CR_4 is $\{1, 6\}$, as expected. Next, consider crossing the facet drawn as a thick segment in Figure 2f. The optimal active set in CR_1 is $\{1\}$, and the inequality corresponding to the facet says that constraint 7 is being activated. G^1 and G^7 are linearly dependent, so LICQ is violated. We therefore solve the LP (18)-(20), with $x_0 = \begin{bmatrix} -0.28 & -0.55 \end{bmatrix}^T$: $\max_{\lambda^{\{1,7\}}} \lambda^7,\ s.t.\ Hz + \left(G^{\{1,7\}}\right)^T \lambda^{\{1,7\}} = 0,\ \lambda^{\{1,7\}} \geq 0$. The solution to this LP is unbounded and according to Theorem 4, we have reached an infeasible part of the state space, which is easily verified. □

6 Off-Line Mp-QP Algorithm

Based on the results of Sections 3, 4 and 5, we finally present an efficient algorithm for the computation of the solution to the mp-QP (3)-(4). Generally, there exist active sets which are not optimal anywhere in the parameter space (typically, most active sets are not optimal anywhere). We need an active set which is optimal in a full-dimensional region to start the algorithm. Generally we can do this by choosing a feasible x, and find the optimal active set for this x by solving a QP. This can be avoided in the special case when we solve a linear MPC problem, where in general the region where no constraint is active at the optimum is full dimensional, and we can choose the active set $\mathcal{A}_0 = \emptyset$ (see (Bemporad and Filippi, Conditionally accepted for publication with minor revisions, Lemma 6)).

Let L_{cand} be a list of active sets which are found, but not yet explored (i.e., are candidates for optimality) and L_{opt} be the set of active sets which have been explored (i.e., are found to be optimal).

Algorithm 1

Choose the initial active set $\mathcal{A}_0$ as in (Bemporad *et al.*, accepted for publication, Prop. 2); Let $L_{cand} \leftarrow \{\mathcal{A}_0\}$, $L_{opt} \leftarrow \emptyset$;
while $L_{cand} \neq \emptyset$ **do**
  Pick an element $\mathcal{A}$ from L_{cand}. $L_{cand} \leftarrow L_{cand} \backslash \{\mathcal{A}\}$;
  Build the matrices $G^{\mathcal{A}}$, $W^{\mathcal{A}}$ and $S^{\mathcal{A}}$ from $\mathcal{A}$ and determine the local Lagrange multipliers, $\lambda^{\mathcal{A}}(x)$ and the solution $z(x)$ from (10) and (9);
  Find the CR where $\mathcal{A}$ is optimal from (12) and (13), and remove all hyperplanes from CR which are not coincident to hyperplanes in the minimal representation of CR;
  if CR is full-dimensional **then**
    $L_{opt} \leftarrow L_{opt} \cup \{\mathcal{A}\}$;
    for each facet $\mathcal{F}$ in CR **do**
      Find the optimal active set on $\mathcal{F}$ by examining the type of hyperplane $\mathcal{F}$ is given by;
      Find any possible optimal active sets in CR_i according to Theorem 2, Corollary 1, Theorem 4 and Theorem 5. If none of these are applicable, find the active set in CR_i by solving a QP as in (Tøndel *et al.*, 2001*b*);
      For any new active set $\mathcal{A}_{new}$ found, let $L_{cand} \leftarrow L_{cand} \cup \{\mathcal{A}_{new}\}$
    end for
  end if
end while

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

$$\begin{pmatrix} \text{Final \# regions} \\ \text{found by the algorithm} \end{pmatrix} \times \begin{pmatrix} \text{\# LPs per region} \\ \text{for redundancy check} \end{pmatrix},$$

and the main cost of the algorithm from (Bemporad *et al.*, 2002) is

$$\begin{pmatrix} \text{Final \# regions} \\ \text{found by the algorithm} \end{pmatrix} \times \begin{pmatrix} \text{\# LPs per region} \\ \text{for redundancy check} \end{pmatrix} + \begin{pmatrix} \text{Total \#} \\ \text{regions explored} \end{pmatrix} \times \begin{pmatrix} \text{\# LPs for red. check} \\ \text{+1 LP to find interior point} \\ \text{+1 QP to find active set} \end{pmatrix}.$$

Consequently, the difference between the two algorithms is the last term, which is due to the extra partitioning into regions R_i, as in figure 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

$$3^{n_z} \times \begin{pmatrix} \text{\# LPs per region} \\ \text{for redundancy check} \end{pmatrix}.$$

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.

Table 1
Number of regions explored and computation times for Algorithm 1 and the algorithm of (Bemporad *et al.*, 2002) for Example 2. We have also added the number of solutions that would be explored by the algorithm of (Seron *et al.*, 2000). In this example, the final number of regions in the solution is the same as the number of regions explored by Algorithm 1

Hor.	**Alg. 1**	**Alg. from** [1]	**Alg. from** [2]
2	[9, 0.14s]	[15, 0.77s]	[9, –]
3	[19, 0.33s]	[39, 2.63s]	[27, –]
4	[33, 0.64s]	[79, 5.60s]	[81, –]
5	[51, 1.14s]	[131, 9.01s]	[243, –]
6	[73, 1.79s]	[205, 16.48s]	[729, –]
7	[99, 2.65s]	[261, 22.74s]	[2187, –]
8	[125, 3.62s]	[329, 30.98s]	[6561, –]
9	[143, 4.58s]	[393, 39.71s]	[19683, –]
10	[157, 5.39s]	[415, 44.82s]	[59049, –]

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

$$A = \begin{bmatrix} 1 & 0 & 0.01 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.01 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0.01 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0.01 & 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 0.0001 & -0.0001 \\ 0.0019 & 0.0019 \\ 0.0132 & -0.0132 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The states of the system are

x_1 - elevation

[1] (Bemporad *et al.*, 2002)
[2] (Seron *et al.*, 2000)

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

$$Q = \mathrm{diag}(100, 100, 10, 10, 400, 200)$$
$$R = I_{2\times 2}$$

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, 2, 3 and 4 parameters to describe the control input.

Table 2
Computation times, helicopter example

Horizon	Algorithm 1	Number of regions
1	1.1 s	33
2	18.9 s	395
3	163.2 s	2211
4	1830.0 s	12223

□

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

Bemporad, A. and C. Filippi (Conditionally accepted for publication with minor revisions). Suboptimal explicit RHC via approximate multiparametric quadratic programming. *JOTA*.

Bemporad, A., F. Borrelli and M. Morari (2000*a*). Piecewise linear optimal controllers for hybrid systems. In: *Proc. American Contr. Conf.*. Chicago, IL. pp. 1190–1194.

Bemporad, A., F. Borrelli and M. Morari (accepted for publication). Model predictive control based on linear programming - the explicit solution. *IEEE Trans. Automatic Control*.

Bemporad, A., M. Morari, V. Dua and E.N. Pistikopoulos (2000*b*). The explicit linear quadratic regulator for constrained systems. In: *Proc. American Contr. Conf.*. Chicago, IL. pp. 872–876.

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

Berkelaar, A.B., K. Roos and T. Terlaky (1997). Advances in sensitivity analysis and parametric programming. pp. [6–1]–[6–45].

Borrelli, F., A. Bemporad and M. Morari (2000). A geometric algorithm for multi-parametric linear programming. *Tech. Rep. AUT00-04, Automatic Control Laboratory, ETH*.

Borrelli, F., M. Baotic, A. Bemporad and M. Morari (2001). Efficient on-line computation of explicit model predictive control. In: *Proc. 40th IEEE Conf. on Decision and Control*. Orlando, Florida.

Chmielewski, D. and V. Manousiouthakis (1996). On constrained infinite-time linear quadratic optimal control.. *Systems and Control Letters* **29**, 121–129.

Dua, V. and E.N. Pistikopoulos (to appear). An algorithm for the solution of multiparametric mixed integer linear programming problems. *Annals of Operations Research*.

Fiacco, A.V. (1983). *Introduction to sensitivity and stability analysis in nonlinear programming*. Academic Press Inc.. Orlando, FL.

Gal, T. (1995). *Postoptimal Analyses, Parametric Programming, and Related Topics*. second ed.. de Gruyter. Berlin.

Johansen, T.A., I. Petersen and O. Slupphaug (1999). Explicit suboptimal linear quadratic regulation with state and input constraints. *SINTEF report STF72-A00303*.

Johansen, T.A., I. Petersen and O. Slupphaug (2000). Explicit subopitmal linear quadratic regulation with input and state constraints. In: *Proc. 39th IEEE Conf. on Decision and Control*. Sydney.

Scokaert, P.O.M. and J. B. Rawlings (1998). Constrained linear quadratic regulation.. *IEEE Trans. Automatic Control'* **43**, 1163–1169.

Seron, M. M., J.A. De Doná and G.C. Goodwin (2000). Global analytical model predictive control with input constraints. In: *Proc. 39th IEEE Conf. on Decision and Control*. Sydney.

Sznaier, M. and M. J. Damborg (1987). Suboptimal control of linear systems with state and control inequality constraints.. *Proc. 26th Conf. Decision and Control.* pp. 761–762.

Tøndel, P. and T. A. Johansen (2002). Complexity reduction in explicit linear model predictive control. In: *Proc. XV IFAC World Congress*. Barcelona.

Tøndel, P., T. A. Johansen and A. Bemporad (2001*a*). An algorithm for multi-parametric quadratic programming and explicit MPC solutions. In: *Proc. 40th IEEE Conf. on Decision and Control*. Orlando, Fl.

Tøndel, P., T. A. Johansen and A. Bemporad (2001*b*). An algorithm for multi-parametric quadratic programming and explicit MPC solutions. *Tech. report 2001-7-W. Available at http://www.itk.ntnu.no/ansatte/Johansen_Tor.Arne/mpQPAlg.pdf*.

Tøndel, P., T. A. Johansen and A. Bemporad (submitted). Computation of piecewise affine control via binary search tree. *Automatica*.