

Linear MPC with Optimal Prioritized Infeasibility Handling: Application, Computational Issues and Stability*

JOSTEIN VADA†, OLAV SLUPPHAUG†, TOR A. JOHANSEN†‡
and BJARNE A. FOSS†

Keywords: Model based control; infeasibility handling; linear programming; linear systems.

All practical MPC implementations should have a means to recover from infeasibility. We present a recently developed infeasibility handler which computes optimal relaxations of the relaxable constraints subject to a user-defined prioritization, by solving only a single linear program on-line in addition to the standard quadratic programming problem on-line. A stability result for this infeasibility handler combined with the Rawlings–Muske MPC controller is provided, and various practical and computational issues are discussed. From a simulated FCCU main fractionator case study, we conclude that the proposed strategy for designing the proposed infeasibility handler is applicable on problems of realistic size.

1. Introduction

During the last years, model predictive control (MPC) has become an attractive control strategy within the process industries. Important stability results within the area of linear MPC are given in Rawlings and Muske (1993) under the assumption of feasibility. In order to fully exploit this stabilizing property, a means to recover from infeasibility of the associated optimization problem whenever possible is required. Infeasibility problems may occur due to e.g. disturbances, operator intervention, modelling errors, or plant failures. Note that in the MPC controller proposed by Rawlings and Muske (1993), an approach for handling infeasibilities caused by the state constraints is included.

Constraints representing physical limitations must be enforced at all times (non-relaxable). Other constraints should be satisfied whenever possible (relaxable), but may be relaxed when necessary. When infeasibility occurs, it is often not obvious which relaxable constraints to relax and the amount that these constraints should be relaxed in order to render a feasible set of constraints. There are some existing techniques which take prioritization levels into account when recovering from infeasibility. Aspen-DMC, IDCOM-M, HIECON and PFC (the last two from Adersa) provide a means of recovering from infeasibilities which involves prioritization of the constraints. When the on-line optimization problem becomes infeasible, the lowest prioritized constraints are dropped (Qin and Badgwell, 1997). In the research literature, the few contributions to this field include Rawlings and Muske (1993), Scokaert and Rawlings (1999), Garcia and Morshedi (1986), Kerrigan *et al.* (2000), Tyler and Morari (1999), Scokaert (1994), Alvarez and de Prada (1997) and Vada *et al.* (2001). To the best of the authors knowledge, the strategy presented in Vada

* Reprinted from Automatica, vol. 37, no. 11, 2001, pp. 1835–1843.

† Department of Engineering Cybernetics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

‡ Corresponding author: Tor.Arne.Johansen@itk.ntnu.no

et al. (2001) is the only optimal infeasibility handler which considers *hard prioritized* constraints without the use of a sequential solution approach. The focus in the present paper is on the application of this infeasibility handler, including guidelines for addressing computational efficiency. The usefulness of the method is illustrated on a simulated distillation column. Moreover, we present a novel stability result for this infeasibility handler combined with the Rawlings–Muske MPC controller.

The following notation is used throughout the paper: Let $n \geq 1$ be an integer and $x, y \in \mathbb{R}^n$. Then $\mathbb{I}_n^+ := \{1, \dots, n\}$, $x \geq (>)y \Leftrightarrow x_i \geq (>)y_i, i \in \mathbb{I}_n^+$, and 0_n is an n -dimensional vector with zeros. (x, y) is used to express $[x^T, y^T]^T$. I_n is the $n \times n$ identity matrix, $I_0 = \emptyset$, $|J|$ denotes the cardinality of the set J , $\text{int}X$ denotes the interior of the set X .

2. Infeasibility handling

Let the model of the plant to be controlled by

$$x_{t+1} = Ax_t + Bu_t \quad (1)$$

for some $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, where x_t and u_t denote the state and control vectors at time t respectively. Note that the model defines the nominal case, while the need for feasibility handling often arises from model/plant mismatch or disturbances. The presentation is based on the well known linear MPC problem (Rawlings and Muske, 1993):

$$\min_{\pi_t} \phi(x_t, \pi_t) = \sum_{j=t}^{\infty} x_{j|t}^T Q x_{j|t} + u_{j|t}^T R u_{j|t}$$

subject to:

$$\begin{aligned} x_{t|t} &= x_t \\ x_{t+N|t}^u &= 0 \\ x_{j+1|t} &= Ax_{j|t} + Bu_{j|t}, t \leq j \\ Hx_{j|t} &\leq h, \quad t < j \leq j_2 + t \\ Du_{j|t} &\leq d, \quad t \leq j \\ u_{j|t} &= 0 \quad t + N \leq j \end{aligned} \quad (2)$$

where $Q \geq 0$, $R > 0$, $\pi_t = (u_{t|t}, \dots, u_{t+N-1|t})$, and $x_{j|t} \in \mathbb{R}^n$, $u_{j|t} \in \mathbb{R}^m$, are the predicted state and control input vector at future time j , respectively, and $x_{j|t}^u \in \mathbb{R}^{n_u}$ denote the unstable mode of the predictor at future time j . Further, $H \in \mathbb{R}^{n_h \times n}$, $h > 0 \in \mathbb{R}^{n_h}$, $D \in \mathbb{R}^{n_d \times m}$, and $d > 0 \in \mathbb{R}^{n_d}$. Assume that (A, B) is stabilizable and $N \geq \max\{n_u, 1\}$. Due to the constraint $u_{j|t} = 0$, $t + N \leq j$, there exists a constraint horizon $j_2 \geq N$ such that satisfaction of $Hx_{j|t} \leq h$, $t < j \leq j_2 + t$ implies $Hx_{j|t} \leq h$, $t < j$ (Rawlings and Muske, 1993).

The inequality constraints in equation (2) may be restated as

$$S\pi_t \leq s_0 + S_1 x_t \quad (3)$$

where S and S_1 are matrices, and s_0 is a vector. Note that the right hand side of equation (3) is parameterized by the state x_t . Due to e.g. disturbances, operator interventions, modelling errors or plant failures, the state may take a value such that

equation (3) has no solution and the MPC optimization problem is infeasible. In the following, for the ease of presentation, we assume that h and d are constants. However, the proposed infeasibility handler can be used for timevarying h and d as well.

The constraints in an MPC optimization problem can be divided into the following classes:

Non-relaxable hard constraints: Hard inequality constraints that are absolute in the sense that they cannot under any circumstances be violated. Constraints related to physical limitations belongs to this class.

Relaxable hard constraints: Hard inequality constraints related to desirables. These constraints are relaxed only in cases when the whole set of hard constraints (i.e. relaxable and non-relaxable) is inconsistent.

Soft constraints: Inequality constraints related to desirables. Violation of these constraints are allowed, but a term is included in the cost function (2) which penalizes constraint violations (see e.g. Zheng and Morari (1995), or Scokaert and Rawlings (1999)).

End point constraints: (I.e. $x_{t+N|t}^u = 0$.) These are equality constraints related to stability. If these constraints are violated, nominal stability is not guaranteed (Rawlings and Muske, 1993).

In the rest of this paper, unless otherwise stated, we make the following assumptions:

- (i) There are no soft constraints.
- (ii) The MPC optimization problem is always feasible when all relaxable hard constraints are removed.
- (iii) Hard prioritization is utilized, i.e. the prioritization among the relaxable hard constraints is absolute in the sense that a higher prioritized constraint is 'infinitely' more important to fulfill than a lower prioritized constraint.
- (iv) All necessary degrees of freedom are used to minimize the violation of the constraints, and all remaining degrees of freedom are used to minimize the cost function in equation (2).

3. Optimal weight design problem (OWDP)

In this section we formulate the problem of computing optimal constraint violations subject to hard prioritization as a single LP problem to be solved on-line at each sample, see also Vada *et al.* (2001).

The constraints in the MPC optimization problem equation (2) can be transformed into the following three constraint sets:

$$G^1 \pi_t = g^1(x_t), g^1(x_t) := g^{10} + g^{11} x_t$$

$$G^2 \pi_t \leq g^2(x_t), g^2(x_t) := g^{20} + g^{21} x_t$$

$$G^3 \pi_t \leq g^3(x_t), g^3(x_t) := g^{30} + g^{31} x_t$$

$$\pi_t \geq 0$$

where $G^1 \in \mathbb{R}^{n_u \times m \cdot N}$, $G^2 \in \mathbb{R}^{m_2 \times m \cdot N}$, $G^3 \in \mathbb{R}^{m_3 \times m \cdot N}$, $g^{10} \in \mathbb{R}^{n_u}$, $g^{11} \in \mathbb{R}^{n_u \times n}$, $g^{20} \in \mathbb{R}^{m_2}$, $g^{21} \in \mathbb{R}^{m_2 \times n}$, $g^{30} \in \mathbb{R}^{m_3}$, $g^{31} \in \mathbb{R}^{m_3 \times n}$, and $\pi_t := \pi_t - \pi^{\min} \in \mathbb{R}^{m \cdot N}$ is a modified vector of control inputs, where π^{\min} is the lower limit on each control input. Such a limit will

always be present in a practical MPC problem, since each element of π_t is related to a physical quantity. (However, if for some reason, π_i^{\min} does not exist, just replace π_i with $u_i - v_i$ in equation (4), with $u_i^{\min} = v_i^{\min} = 0$.) In equation (4), $G^1 \pi_t = g^1(x_t)$ corresponds to the stability constraint $x_{t+N|t}^u = 0$ in equation (2). Further, the inequality constraints in equation (2) are partitioned into the following two sets of constraints: $G^2 \pi_t \leq g^2(x_t)$, which is the set of all non-relaxable hard constraints, and $G^3 \pi_t \leq g^3(x_t)$, which is the set of all relaxable hard constraints. The total number of inequality constraints in equation (2) is $n_d \cdot N + n_h \cdot j_2$, and thus $m_2 + m_3 = n_d \cdot N + n_h \cdot j_2$. The relation between equations (4) and (2) is easily established by, in equation (2), inserting the 1st, 3rd, and 6th constraint into the 2nd, 4th, and 5th constraint and by replacing $\bar{\pi}_t$ with $\pi_t + \pi^{\min}$. Further, assume that there exists a hard prioritization among the inequalities in $G^3 \pi_t \leq g^3(x_t)$, and that G^3 and g^3 are constructed such that the i th row of $G^3 \pi_t \leq g^3(x_t)$ have higher priority than the $(i+1)$ th row. This implies that minimizing the violations of the i th row of $G^3 \pi_t \leq g^3(x_t)$ is 'infinitely' more important than minimizing the violations of the $(i+1)$ th row.

Assume that, at a given sample, there is no feasible solution to equation (4). Since the 3rd constraint in equation (4) is the only relaxable hard constraint, in order to transform equation (4) into a feasible optimization problem, we introduce a vector of constraint violations $z_t \in \mathbb{R}^{m_3}$ as follows

$$\begin{aligned} G^1 \pi_t &= g^1(x_t) \\ G^2 \pi_t &\leq g^2(x_t) \\ G^3 \pi_t &= g^3(x_t) + z_t \\ \pi_t, z_t &\geq 0 \end{aligned} \tag{5}$$

Next we introduce the notion of *lexicographic minimum*: $y^o \in Y \subseteq \mathbb{R}^n$ is the lexicographic minimum of Y if it is not possible to find another $y \in Y$ and an $i \in I_n^+$ such that $y_i < y_i^o$ and $y_j = y_j^o, j \in I_{i-1}^+$. As an example $[0.10, 0.01, 10000]$ is lexicographically less than $[0.10, 0.011, 0]$, since the first element of both vectors are equal, while minimizing the second element is 'infinitely' more important than minimizing the third.

Now we are ready to state a problem whose solution can be used to compute optimal constraint violations (according to the given hard prioritization) by solving only one LP problem on-line in addition to the original MPC QP problem:

Optimal weight design problem (OWDP)

Let $X \neq \emptyset$ denote the set of all x_t such that there exists (π_t, z_t) satisfying equation (5). Given an $x_t \in X$, let $Z(x_t)$ denote the set of all $z_t \geq 0$ such that there exists a π_t satisfying the inequalities in equation (5). Design the weight vector \tilde{c} in equation (6) such that $\forall x_t \in X, z_t^*$ defined by

$$(\pi_t^*, z_t^*) := \operatorname{argmin} \tilde{c}^T z \quad \text{subject to (5)} \tag{6}$$

is equal to the lexicographic minimum of $Z(x_t)$.

Note that since we have assumed that $d, h > 0$, (A, B) stabilizable, and $N \geq \max\{n_u, 1\}$, we have that $X \neq \emptyset$, and $0 \in \operatorname{int} X$. In Vada *et al.* (2001), existence of a solution to the OWDP under these assumptions is established. A consequence of this result is, at each sample, if:

- (i) the state x_t has a value making equation (2) infeasible, and
- (ii) $x_t \in X$, that is, with the given x_t there exists a relaxation of the relaxable hard constraints such that equation (2) becomes feasible, then an optimal relaxation z_t^* can be computed by solving the LP problem in equation (6).

4. Stability

In this section we show that by combining the proposed infeasibility handler with the MPC controller defined in Section 2, the region of attraction of the original MPC controller without an infeasibility handler is increased. For a certain prioritization, Theorem 4.1 below establishes nominal asymptotic stability for the receding horizon implementation of equation (2) if the constraints in equation (2) are replaced by equation (5) with $z_t = z_t^*$, where the weights \tilde{c} in equation (6) is a solution to the OWDP. First, we present Lemma 4.1 which is needed in the proof of Theorem 4.1:

Lemma 4.1 Assume that the constraints $x_{t+j|t} \in \tilde{X}, \forall j > 1$, are hard non-relaxable constraints in equation (2), where $\tilde{X} \subset \mathbb{R}^n, 0 \in \text{int}\tilde{X}$, is an arbitrary bounded subset of X . Then, in equation (2), there exists a sufficiently large $j_2 \geq N$ such that $\forall x_t \in \tilde{X}, Hx_{t+j_2|t} \leq h \Rightarrow Hx_{t+j_2+i|t} \leq h, i = 1, 2, \dots$

Proof. Follows from Rawlings and Muske (1993) and boundedness of \tilde{X} . \square

Next, we define a prioritization among the constraints which is used in Theorem 4.1:

Priority Assumption

Assume that a unique priority level is assigned to each relaxable row of $Hx_{j|t} \leq h$ and $Du_{j|t} \leq d$ in equation (2), such that all constraints on the horizon related to a certain relaxable row of $Hx_{j|t} \leq h$ or $Du_{j|t} \leq d$ with a given priority level have higher priority than any constraint on the horizon related to rows with a lower priority level constraints. Let $H_i(D_i)$ denote the i th row of $H(D)$, and assume that $\forall j \in \mathbb{J}_{j_2-1}^+, i \in \mathbb{I}_{n_h}^+, t \geq 0, H_i x_{t+j+1|t} \leq h_i$ has higher priority than $H_i x_{t+j|t} \leq h_i$, and that $\forall j \in \mathbb{J}_N^+, i \in \mathbb{I}_{n_d}^+, t \geq 0, D_i u_{t+j|t} \leq d_i$ has higher priority than $D_i u_{t+j-1|t} \leq d_i$.

Theorem 4.1 Assume that the constraints $x_{t+j|t} \in \tilde{X}, \forall j > 1$ are hard non-relaxable constraints in equation (2), and let j_2 be given as in Lemma 4.1. Let G^3 and g^3 in equation (5) be constructed according to the Priority Assumption. Assume that $\forall t \geq 0, z_t = z_t^*$, the solution to equation (6) with \tilde{c} a solution to the OWDP, and assume further that $\forall t \geq 0, u_t = u_{t|t}^*$, where $u_{t|t}^*$ is the first m elements of the solution of equation (2) where the constraints are replaced by equation (5). Then, $\forall x_0 \in \tilde{X}, \{z_t\}_{t=0}^\infty$ becomes zero within finite time, and the origin is an asymptotically stable equilibrium point with \tilde{X} contained in the region of attraction.

Proof. First we prove that z_t^* becomes zero in finite time: Given any $x_0 \in \tilde{X}$, let z'_0 denote the constraint violations obtained by shifting the constraint violations in z_0^* one step ahead and filling up with zeros in the locations corresponding to prediction $j_2 + 1|1(N|1)$ for the state (control input) constraints. Thus, since $x_1 = x_{1|0}$ (nominal case), at time $t = 1, z_1 = z'_0$ is feasible, and hence zero violation of the constraint at the end of the horizon of the highest prioritized constraint is feasible. Then, due to the choice of \tilde{c} in equation (6), the corresponding element of z_1^* is equal to zero.

Continuing this argument, due to the prioritization along the horizon, we obtain that in z_t^* , all violations of $Hx_{j|t} \leq h$ (or $Du_{j|t} \leq d$) corresponding to the highest priority level, becomes zero after at least j_2 (or N) samples. Continuing this for the row of $Hx_{j|t} \leq h$ or $Du_{j|t} \leq d$ corresponding with the next priority level, and so on, we obtain that $z_t^* = 0_{m_3}$, $t = m_3, m_3 + 1, \dots$

Finally we prove asymptotic stability with \tilde{X} contained in the region of attraction: Let X' be the set of all x_t such that there exists a π_t satisfying equation (5) with $z_t = 0$. It follows from Rawlings and Muske (1993) that $\forall x \in X'$, by using the control law defined by receding-horizon implementation of the solution of equation (2), the origin is an asymptotically stable solution. Combining this with the fact that $z_t^* = 0_{m_3}$, $t = m_3, m_3 + 1, \dots$, the result follows. \square

Note that we need to assume that $\forall t \leq 0$, x_t is contained in a bounded region \tilde{X} . This is done to obtain a fixed j_2 which is sufficiently large to be valid for all $x_t \in \tilde{X}$. Also note that a result similar to Theorem 4.1 is stated in Rawlings and Muske (1993) and Scokaert and Rawlings (1999) for the case when only the state constraints can be relaxed, and when all rows of $Hx_{j|t}$ have equal priority. Recall that the strategy in (Scokaert and Rawlings 1999) is based on solving a sequence of optimization problems. An important consequence of Theorem 4.1 is that by using the proposed controller, the region of attraction of the MPC controller equation (2) without infeasibility handling is at least enlarged from X' to \tilde{X} (cf. the proof of Theorem 4.1 for X'). Also note that, in the case when all state constraints and none of the input constraints are relaxable, the region of attraction of the approaches proposed in Rawlings and Muske (1993) and Scokaert and Rawlings (1999) (the optimal minimal time approach) are equal to the region of attraction obtained by using the infeasibility handler we propose.

5. Solving the OWDP

In Vada *et al.* (2001), an algorithm which solves the OWDP is presented. In order to give an intuitive understanding of this algorithm, we give in the following an outline of the main ideas behind the algorithm.

The constraints in the OWDP, i.e. equation (5), can be restated as a set of equality constraints by introducing nonnegative auxiliary variables v_t and w_t

$$\begin{aligned} G^1 \pi_t &= g^1(x_t) \\ G^2 \pi_t + v_t &= g^2(x_t) \\ G^3 \pi_t + w_t - z_t &= g^3(x_t) \\ \pi_t, v_t, w_t, z_t &\geq 0 \end{aligned} \quad (7)$$

and by defining $x^{LP} := (\pi_t, v_t, w_t, z_t)$, equation (6) can be transformed into an LP problem on standard form

$$\begin{aligned} \min_{x^{LP}} c^T x^{LP} \\ \text{subject to: } \begin{cases} A^{LP} x^{LP} = b(x_t) \\ x^{LP} \geq 0 \end{cases} \end{aligned} \quad (8)$$

where $b(x_t) := (g^1(x_t), g^2(x_t), g^3(x_t))$, $c := (0_{Nm+m_2+m_3}, \tilde{c})$ and

$$A^{LP} := \begin{bmatrix} G^1 & 0_{m_1 \times m_2} & 0_{m_1 \times m_3} & 0_{m_1 \times m_3} \\ G^2 & I_{m_2} & 0_{m_2 \times m_3} & 0_{m_2 \times m_3} \\ G^3 & 0_{m_3 \times m_2} & I_{m_3} & -I_{m_3} \end{bmatrix} \in \mathbb{R}^{(m_1+m_2+m_3) \times (Nm+m_2+2m_3)} \quad (9)$$

The problem in equation (8) is called a *parametric LP*, (Gal, 1995), since the right hand side of the equality-constraints in equation (8) is parameterized by x_t . Recall that the problem stated in the OWDP is to design c in equation (8) (or, more precisely \tilde{c} , since $c_i = 0$, $i \in I_{Nm+m_2+m_3}^+$) such that for each $x_t \in X$, any optimal solution to equation (8) has the property that the z_t -part of this solution is equal to the lexicographically least feasible $z_t \geq 0$. By using theory from parametric programming, it can be shown that X can be covered by a set of polytopes, where each of the polytopes is uniquely defined as

$$X_{B^{LP}} := \{x_t \in X \mid (B^{LP})^{-1}(g^1(x_t), g^2(x_t), g^3(x_t)) \geq 0\} \quad (10)$$

where $B^{LP} \in \mathbb{R}^{(m_1+m_2+m_3) \times (m_1+m_2+m_3)}$ is a basis for $\mathbb{R}^{m_1+m_2+m_3}$ that consists of $m_1+m_2+m_3$ linearly independent columns of A^{LP} . Each of the polytopes $X_{B^{LP}}$ is associated with a separate basis B^{LP} . Further, each of the bases considered has the property that if $x_t \in X_{B^{LP}}$, the non-zero elements of $z^0(x_t)$, the lexicographically minimum of $Z(x_t)$, are equal to corresponding elements of the vector $(B^{LP})^{-1}(g^1(x_t), g^2(x_t), g^3(x_t))$. Let \mathcal{B} denote the set of bases such that X is covered by the corresponding set of $X_{B^{LP}}$ s. In Vada *et al.* (2001) it is shown that each basis in \mathcal{B} defines a set of linear constraints on \tilde{c} in equation (6) in order for \tilde{c} to solve the OWDP. The main idea is to compute a \tilde{c} which satisfies the set of constraints defined by all bases in \mathcal{B} . \mathcal{B} is computed by a sequential algorithm that finds a new basis B^{LP} in \mathcal{B} by moving into a new region of X defined by neighbors of already computed regions. This algorithm is continued until X is covered and terminates in finite time since \mathcal{B} is finite.

6. Practical modifications and computational issues

In some MPC implementations it might not be desirable or natural to distinguish between each (scalar) constraint by assigning different priority levels to each of them. In such cases, two or more constraints can be collected into the same priority level and thus share the same element of z_t . Note that by this, the size of the OWDP is also reduced (fewer elements in \tilde{c}) at the cost of a possibly increased number of constraint violations.

Another modification of the proposed infeasibility handler which also allows for more than one constraint having the same priority level and which also reduces the size of the OWDP, is to divide the relaxable constraints into two parts: hard prioritized constraints and soft prioritized constraints. With soft prioritization the original cost function is extended with appropriated weighted penalties on the constraint violations (Zheng and Morari, 1995). This can be done by classifying the m_3^{red} most important relaxable constraints as hard prioritized constraints and the $m_3 - m_3^{red}$ other relaxable hard constraints as soft prioritized constraints. The OWDP is then solved without the set of the soft prioritized constraints, and by this the size of the corresponding LP which computes \tilde{c} is reduced equivalently as with the modification proposed above.

Whether or not each constraint should have a separate priority level, or whether or not the lower prioritized constraints can be treated as soft constraints, is of course

dependent on the application. However, it is important to note that it is not straightforward (if possible at all) to choose the weights in a soft constrained approach so as to obtain the desired hard prioritization for all possible initial states. Thus, in order to ensure that the violation of the most important relaxable constraints, such as shut-down and alarm limits (if they are relaxable), are minimized according to a given prioritization, the violation of these constraints should be computed by a hard prioritized infeasibility handler.

The computational load and memory capacity required for computing and storing the index set to each basis is proportional to $|\mathcal{B}|$. The computational load can only be reduced by reducing the size of the original OWDP, or by designing a more efficient algorithm to compute \mathcal{B} . The latter issue is discussed in Vada *et al.* (2001), while we here focus on the first. Recall that the above suggested modifications reduces the size of the OWDP. Another modification which also reduces the size of the OWDP is to reduce the horizon of the state constraints (i.e. j_2). Note that if j_2 is less than the minimal j_2 satisfying the condition given in Lemma 4.1, nominal stability of the controller is no longer guaranteed. Simulations indicate that the method we have used for computing j_2 give a very conservative estimate (see Section 7), thus much can be gained by improving the procedure for computing j_2 .

7. Simulation example

In this section, we illustrate the use of the proposed infeasibility handler for a linear model of the top section of a fluid catalytic cracker unit (FCCU) main fractionator, see Figure 1, which is a critical unit for separating gasoline and LCO (diesel) from the feedstock from an upstream riser reactor. A rigorous model of the

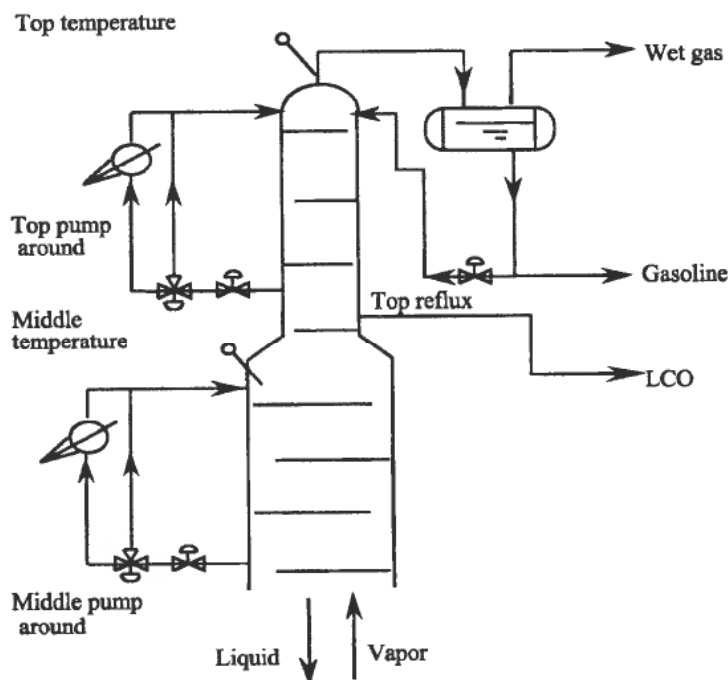


Figure 1. Top part of a FCCU fractionator.

fractionator has been developed and fitted to real plant data (Cong *et al.*, 1998), and a linear model has been derived by discretization and linearization of this model around a nominal operating point:

$$A = 0.01 \begin{bmatrix} 90.9028 & -6.4632 & 5.7545 & 1.5616 \\ -1.2803 & 92.1012 & 1.5499 & 3.3564 \\ -0.7713 & -4.2529 & 85.3340 & 3.4314 \\ 0.2413 & -0.7122 & -0.1415 & 85.6537 \end{bmatrix}$$

$$B = 0.01 \begin{bmatrix} -3.6674 & -12.4077 & -0.0648 & -5.7626 & 0.1049 \\ -2.0582 & -0.5071 & 0.1088 & -7.2823 & 0.2324 \\ -1.3759 & -1.6691 & 1.5177 & -1.8955 & -0.1224 \\ 0.1182 & -0.0777 & -0.0002 & -0.3235 & 1.4456 \end{bmatrix}$$

and $x_t := x_t^{abs} - x^{nom}$, and $u_t := u_t^{abs} - u^{nom}$, where $x^{nom} := (x_1^{nom}, \dots, x_4^{nom})$ and $u^{nom} := (u_1^{nom}, \dots, u_5^{nom})$. (x^{nom}, u^{nom}) is the nominal operating point. The sampling time is 30 s. The legend for the states and control inputs is given in Table 1, and the nominal operating point in addition to the absolute upper and lower bounds are given in Table 2. Non-relaxable hard constraints are defined as $x^{abs, lb} \leq x_t + x^{nom} \leq x^{abs, ub}$, and $u^{abs, lb} \leq u_t + u^{nom} \leq u^{abs, ub}$. Further, the relaxable hard constraints and their corresponding priority levels are given in Table 3. The prioritizations are based on assumptions such as: gasoline is assumed to be more valuable than LCO (this assumption determines the prioritization between priority level 1 and 2, which are related to product quality, and between priority level 3 and 4, which are related to minimizing the content of a valuable product in a less valuable product), and high

Table 1. States and control inputs.

Var.	Description	Unit
$(x_t)_1$	Top vapor temperature	°C
$(x_t)_2$	Middle vapor temperature	°C
$(x_t)_3$	Top heat exchanger outlet temperature	°C
$(x_t)_4$	Middle heat exchanger outlet temperature	°C
$(u_t)_1$	Top pump-around tee valve position	%
$(u_t)_2$	Top reflux valve position	%
$(u_t)_3$	Top pump-around valve position	%
$(u_t)_4$	Middle pump-around tee valve position	%
$(u_t)_5$	Middle pump-around valve position	%

Table 2. Nominal operating point, lower, and upper bounds.

x_1^{nom}	107.0°C	$x_1^{abs, lb}$	106	$x_1^{abs, ub}$	108
x_2^{nom}	219.0°C	$x_2^{abs, lb}$	218	$x_2^{abs, ub}$	221
x_3^{nom}	87.0°C	$x_3^{abs, lb}$	83	$x_3^{abs, ub}$	91
x_4^{nom}	199.0°C	$x_4^{abs, lb}$	195	$x_4^{abs, ub}$	203
u_1^{nom}	62.3%	$u_1^{abs, lb}$	0	$u_1^{abs, ub}$	100
u_2^{nom}	0.0%	$u_2^{abs, lb}$	0	$u_2^{abs, ub}$	100
u_3^{nom}	50.0%	$u_3^{abs, lb}$	40	$u_3^{abs, ub}$	80
u_4^{nom}	67.9%	$u_4^{abs, lb}$	0	$u_4^{abs, ub}$	100
u_5^{nom}	50.0%	$u_5^{abs, lb}$	40	$u_5^{abs, ub}$	80

Table 3. Relaxable hard constraints.

Pri. level	Constraint	Pri. level	Constraint
1	$(x_t^{abs})_1 \leq 107.5$	6	$(u_t^{abs})_3 \leq 55$
2	$(x_t^{abs})_2 \leq 219.5$	7	$(u_t^{abs})_5 \leq 55$
3	$(x_t^{abs})_1 \leq 106.5$	8	$(u_t^{abs})_1 \leq 67$
4	$(x_t^{abs})_2 \geq 218.5$	9	$(u_t^{abs})_4 \leq 75$
5	$(u_t^{abs})_2 \leq 5.0$		

production rate has higher priority than minimizing the energy use (this assumption determines the prioritization between priority level 5 and the other relaxable input constraints). Note that since the constraint horizon in equation (2) is j_2 , there are j_2 constraints corresponding to each of the above defined state constraints (both relaxable and non-relaxable), and that due to the move horizon, there are N constraints corresponding to each of the above defined control input constraints (both relaxable and non-relaxable). Hence, there are several constraints related to a given priority level. The prioritization implies that minimizing the violation of any of the constraints related to priority level i has higher priority than minimizing any of the constraints related to priority level $i + 1$. Assume that within a given priority level, minimizing the constraint violation at prediction $k + 1$ has higher priority than at prediction k . That is, we assume the same prioritization as in the Priority Assumption (defined in Section 4). We have chosen $N = 5$, and by using a slight modification of (Gilbert and Tan 1991, Algorithm 3.2) to calculate j_2 , assuming that the nonrelaxable hard state constraints are always satisfied, we get $j_2 = 40$. Thus, for the given example, there are $m_3 = 185$ distinct priority levels, and in the OWDP, the dimension of \tilde{c} is thus 185.

In order to solve the OWDP we used Algorithm 4.4 in (Vada *et al.*, 2001), which is briefly described in Section 5. In the algorithm, the parameter determining the lower bound on the weights is set equal to 1.0. The number of bases in the resulting set \mathcal{B} is 167, and the elements of the resulting \tilde{c} which are greater than 1.0 are shown in Table 4.

The algorithm is implemented in MATLAB with NAG Foundation Toolbox, and the computation time was about 76 minutes on a Pentium 450MHz PC with 256MB RAM. Note, however, that the computation of \tilde{c} is done off-line. The on-line computational effort associated with the infeasibility handler (the LP problem in equation (6)) is typically smaller than the QP in equation (2).

Note that only six of the weights are greater than their minimum value. At a first glance, since the weights related to the 78 highest prioritized constraints are all equal to the predetermined minimum value, one might think that it is remarkable that these weights solve the OWDP. However, note that all weights in Table 4 are related to the first or second samples on the horizon for a given priority level. Thus, for the given

Table 4. Weights solving the OWDP with $j_2 = 40$. All other \tilde{c}_i are equal to 1.0.

i	\tilde{c}_i	Corresponds to:	i	\tilde{c}_i	Corresponds to:
79	25.70	$(x_{t+2 t})_2 \leq 219.5$	120	62.910	$(x_{t+1 t})_2 \geq 106.5$
80	113.90	$(x_{t+1 t})_2 \leq 219.5$	165	1.070	$(u_t^{abs})_2 \leq 5.0$
119	14.86	$(x_{t+2 t})_1 \geq 106.5$	180	1.283	$(u_t^{abs})_1 \leq 67.0$

process, minimizing the constraint violations at the beginning of the horizon implies that the constraint violations at the end of the horizon are minimized. (Recall that within a given priority level, the constraints corresponding to the first samples of the horizon have lower priority than the samples at the end of the horizon.) Note that this might not be the case for a different process. Further note that it is by far not intuitive to determine how large the weights should be in order to guarantee the fulfillment of the hard prioritization. The largest weight produced by this algorithm is only two orders of magnitude larger than the smallest weights. This is in strong contrast to a heuristic approach that might rely on a sufficiently large weight ratio between each priority level. The latter approach could lead to a numerically ill-conditioned LP.

The simulation result obtain by combining the proposed infeasibility handler with the closed-loop implementation of equation (2) when a state disturbance of $[-1, 2, -4, 4]^T$ enters the system at $t=0$ is shown in Figure 2. In equation (2), $Q = 100I_n$ and $R = I_m$. Observe that all relaxable constraints are satisfied for all $t \geq 2$: At $t=0$, there are 4 relaxable constraints which are violated. Two of them corresponds to the first sample of the constraints with priority level 2 and 3, and the other two corresponds to the first two samples of the constraint with priority level 9. At $t=1$,

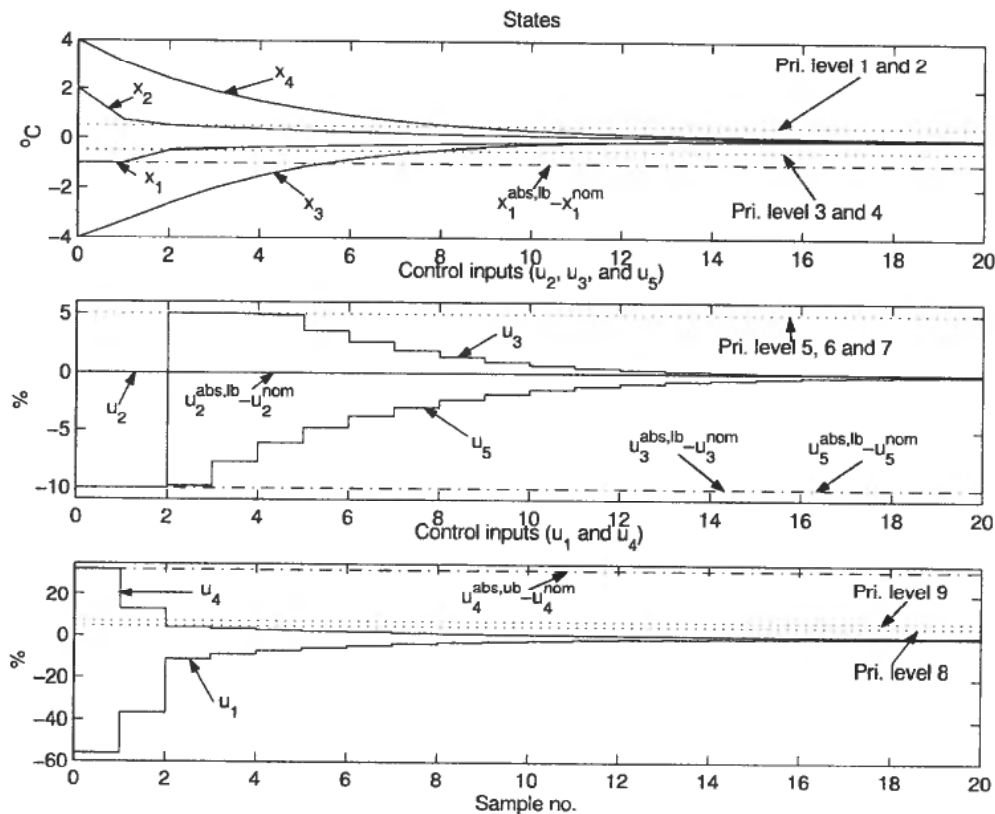


Figure 2. Simulation results using the proposed infeasibility handler combined with linear MPC. All values are deviations from the nominal operating point. Solid lines: states and control inputs, dash-dotted lines: non-relaxable constraints, dotted lines: relaxable constraints. Note that in the upper part, only the hard constraint on x_1 are shown, since the other hard constraints are active only at $t=0$.

Table 5. Weights solving the modified OWDP with 9 priority levels and $j_2 = 40$. All other \tilde{c}_i are equal to 1.0.

i	\tilde{c}_i	Corresponds to:	i	\tilde{c}_i	Corresponds to:
2	113.90	$(x_t^{abs})_2 \leq 219.5$	5	1.070	$(u_t^{abs})_2 \geq 5.0$
3	62.91	$(x_t^{abs})_1 \geq 106.5$	8	1.282	$(u_t^{abs})_1 \leq 67.0$

Table 6. Weights solving the OWDP with $j_2 = 5$. All other \tilde{c}_i are equal to 1.0.

i	\tilde{c}_i	Corresponds to:	i	\tilde{c}_i	Corresponds to:
9	25.70	$(x_{t+2 t}^{abs})_2 \leq 219.5$	15	62.910	$(x_{t+1 t}^{abs})_2 \geq 106.5$
10	113.90	$(x_{t+1 t}^{abs})_2 \leq 219.5$	25	1.070	$(u_{t t}^{abs})_2 \leq 5.0$
14	14.86	$(x_{t+2 t}^{abs})_1 \geq 106.5$	40	1.283	$(u_{t t}^{abs})_1 \leq 67.0$

the only constraint violation corresponds to the first sample of the constraint with priority level 9.

Table 5 shows the weights when, within each of the priority levels given in Table 3, all constraints along the horizon have the same priority. In this case, there are only 9 priority levels, i.e. $m_3^{red} = 9$. The number of bases in the resulting \mathcal{B}^{red} is 243, i.e. $|\mathcal{B}^{red}|$ is almost 50% larger than obtained by using the same prioritization as in the Priority Assumption. Note that for each priority level, the weights obtained by solving this modified OWDP are equal to the largest weights over the horizon in Table 4. Hence, reducing the number of priority levels does not imply reduced off-line computational load. The simulation results obtained with the same disturbance as above is equal to the one obtained by using the prioritization defined by the Priority Assumption, see Figure 2.

Table 6 shows the weights when $j_2 = 5$. With this choice of j_2 , the corresponding m_3 becomes 45. The number of bases in the resulting \mathcal{B} is 45, and the computation time was less than one minute. Also in this case, the simulation results obtained are equal to the one in Figure 2, where $j_2 = 40$. This shows that for the given disturbance, the original choice of j_2 is rather conservative, and by reducing j_2 to a more realistic value, the computational load is greatly reduced.

8. Discussion/Conclusions

In MPC, it is normally the case that some constraints are more important to fulfill than others. In such cases, this information defines a restriction on how the constraints should be relaxed in order to recover from infeasibility. We assume that the difference in importance can be described by the use of priority levels and we focus on how to relax the constraints when a constraint with a certain priority level is infinitely more important to satisfy than a constraint with a lower priority level. Furthermore, we assume that if a certain constraint must be violated, it is desirable to minimize the violation of this constraint. The main feature of the optimal infeasibility handling algorithm of Vada *et al.* (2001) is that it is faster than alternative algorithms for hard prioritization, since it reduces the problem to a single LP to be solved online. In this paper, this algorithm is applied in simulations of a realistic MPC problem. For this problem, the offline computational load of the algorithm is not prohibitively large, even though the problems includes 185 priority levels. The

elements of the resulting cost function computed by the algorithm are non-intuitive, implying that designing such a cost function by trial and error might be time consuming.

The paper also proves that the proposed strategy guarantees nominal asymptotic stability if avoiding constraint violations at the end of the horizon has the highest priority. This result implies that the region of attraction of the controller without infeasibility handling is at least enlarged by using the proposed infeasibility handler.

Certainly, in some MPC implementations it might not be desirable or natural to distinguish between each (scalar) constraint by giving them different priority levels. Hence, we propose two modifications of the infeasibility handler which assigns the same priority level to several constraints. One approach is based on hard prioritization only, and one approach combines hard prioritization and soft prioritization. This leads to an LP with less variables, which is desirable in large-scale practical applications. Still, the computation complexity of the offline computations may restrict the applicability of the present approach for some problems. Reducing the computational complexity is therefore a problem that calls for further investigation.

Traditionally, when designing constraints which are desirables (not related to physical limitations), one needs to consider whether or not such constraints may cause the controller to run into feasibility problems. By using the proposed approach for infeasibility handling, such considerations become less important. Actually, one might design relaxable hard constraints which one knows can be satisfied in only small regions of the state space.

Acknowledgment

The financial support of the Research Council of Norway, project number 107620/420, is gratefully acknowledged. Dr. Songbo Cong is also gratefully acknowledged for providing us with the linear model of the top section of the fluid catalytic cracker unit he has developed, and for giving us insight into this process.

REFERENCES

- ALVAREZ, T. and C. DE PRADA (1997). Handling Infeasibilities in Predictive Control. *Computers chem. Engng.* 21, pp. 577–582.
- CONG, S. B., P. YUAN and F. SHEN (1998). An integrated non-equilibrium dynamic model of petroleum distillation column. Unpublished manuscript.
- GAL, T. (1995). *Postoptimal Analyses, Parametric Programming, and Related Topics*. 2 ed. Walter de Gruyter.
- GARCIA, C. E. and A. M. MORSHEDI (1986). Quadratic Programming Solution of Dynamic Matrix Control (QDMC). *Chemical Engineering Communications* 46, 73–87.
- GILBERT, E. G. and K. T. TAN (1991). Linear Systems with State and Control Constraints: The Theory and Application of Maximal Output Admissible Sets. *IEEE Transactions on Automatic Control* 36(19), pp. 1008–1020.
- KERRIGAN, E., A. BEMPORAD, D. MIGNONE, M. MORARI and J. M. MACIEJOWSKI (2000). Multi-objective Prioritisation and Reconfiguration for the Control of Constrained Hybrid Systems. In: *Proc. American Control Conference, Chicago*. pp. 1694–1698.
- QIN, S. J. and T. A. BADGWELL (1997). An Overview of Industrial Model Predictive Control Technology. In: *Fifth International Conference on Chemical Process Control* (J. C. KANTOR, C. E. GARCIA and B. CARNAHAN, Eds.). AIChE Symposium Series 316. pp. 232–256.
- RAWLINGS, J. B. and K. R. MUSKE (1993). The Stability of Constrained Receding Horizon Control. *IEEE Transactions on Automatic Control* 38(10), pp. 1512–1516.

- SCOKAERT, P. O. M. (1994). Constrained predictive control. PhD thesis. University of Oxford, UK.
- SCOKAERT, P. O. M. and J. B. RAWLINGS (1999). Feasibility issues in model predictive control. *AIChE Journal* 45(8), pp. 1649–1659.
- TYLER, M. L. and M. MORARI (1999). Propositional logic in control and monitoring problems. *Automatica* 35, pp. 565–582.
- VADA, J., O. SLUPPHAUG and T. A. JOHANSEN (2001). Efficient Optimal Prioritized Infeasibility Handling in Model Predictive Control—a Parametric Preemptive Multi-Objective Linear Programming Approach. *Journal of Optimization Theory and Applications*.
- ZHENG, A. and M. MORARI (1995). Stability of Model Predictive Control with Mixed Constraints. *IEEE Transactions on Automatic Control* 40(10), pp. 1818–1823.