Stability of SISO Quadratic Dynamic Matrix Control with Hard Output Constraints

Evanghelos Zafiriou and André L. Marchal
Chemical Engineering Dept. and Systems Research Center, University of Maryland, College Park, MD 20742

The inclusion of output constraints in the on-line optimization problem solved by quadratic dynamic matrix control (QDMC) can result in a unstable, closed-loop system, even when the corresponding unconstrained algorithm is stable. The presence of constraints in the optimization problem produces a nonlinear, closed-loop system, although the plant and model dynamics are assumed linear. This article quantifies the effect of output constraints on both nominal and robust stability for single-input/single-output processes. Stability conditions provided can be used to select the QDMC parameters and constraint window. Examples demonstrate that these conditions can capture the nonlinear effect of the constraints and that tuning rules developed for the unconstrained case should not be used with output constraints. An example with dead time error illustrates the use of this framework to guarantee the robustness of constrained QDMC with respect to modeling error.

Introduction

Model predictive control (MPC) encompasses a large class of process control algorithms sharing the common characteristic of using explicitly a model of the process to predict future behavior and taking control action by optimizing some performance objective. A performance measure made popular because of its simplicity and its successful use in industrial applications is a quadratic objective function that includes the predicted deviation from desired setpoint values over a future horizon. In the QDMC formulation (Garcia and Morshedi, 1986), the objective function also includes a penalty term on excessive control moves, and its minimization is carried out on-line at each sampling point, subject to satisfaction of hard constraints on several process variables.

The great attraction of QDMC is that the straightforward formulation of an optimization problem will satisfy the control specifications. Saturation constraints on the manipulated variables, as well as performance and safety constraints on outputs and other state variables, can be taken care of by simply listing them and minimizing the quadratic objective function subject to their satisfaction. When the model used for prediction is linear, the on-line optimization is a quadratic program (QP), for which efficient algorithms exist, especially if the similarity of the optimization problems that are solved at successive sampling points is taken into account (Ricker, 1985). Formulations that use nonlinear models for prediction have also been developed. In this case, the on-line optimization is a Nonlinear Program, which with appropriate mathematical techniques and/or approximations can be transformed into a series of QP's (Li and Biegler, 1989; Peterson et al., 1990). Eaton and Rawlings (1990) also consider the parametric sensitivity of the optimal solution. An industrial application of QDMC that utilizes a nonlinear model is described by Garcia (1984).

There are, however, certain issues that make the use of QDMC more complex than it is apparent. The on-line optimization solves an open-loop control problem, given the information available up to that point. The control action that is calculated at a sampling point is optimal, only if the sequence of control moves found by the optimization is implemented uninterrupted. This, however, will not happen, because a new optimization problem will be solved at the next sampling point, utilizing the newly acquired information from the measurements in predicting future behavior. The fact that QDMC is implemented as a closed-loop control system is not incorporated in the on-line optimization. Closed-loop stability cannot be assumed simply because the on-line optimization finds a solution. This issue of closed-up stability is complicated by

Correspondence concerning this article should be addressed to E. Zafiriou.
A. L. Marchal is currently at the Rhône-Poulenc Industrialisation, 69151 Décines-Charpieu, France.

two facts: first, there is always uncertainty associated with the model used in the prediction; second, the presence of constraints in the optimization problem results in a nonlinear, closed-loop system even if the model and plant dynamics are assumed linear. In the unconstrained case, robust linear control theory can be used to study robustness with respect to modeling error (see, for example, Prett and Garcia, 1988). For the constrained case, Zafiriou (1989) suggested a framework that allows the translation of the robust stability of the constrained, and therefore nonlinear, closed-loop system into robustness conditions for a set of linear systems.

This article focuses on the effect of hard output constraints on the closed-loop stability of QDMC. The ability to include output constraints in the on-line optimization distinguishes QDMC from other efficient methods that deal with constraints on the manipulated variables (for example, Campo and Morari, 1990). It has been pointed out, however, that output constraints can result in very aggressive controllers (Zafiriou, 1989; Ricker et al., 1989). Their effect on stability is quantified, and a methodology for selecting the QDMC parameters is provided that guarantees robustness to model-plant mismatch. It is demonstrated that tuning rules developed for the unconstrained case (Garcia and Morari, 1982) not only may not work well for the output constraint case, but may actually be the cause of stability problems when output constraints are included in the on-line optimization.

Quadratic Programming Formulation of QDMC

This section sets notation for the quadratic program that is solved on-line. An impulse response model is used (see, for example, Garcia and Morari, 1982):

$$\tilde{y}(k) = \sum_{i=1}^{N} H_i u(k-i) + \tilde{d}(k) \tag{1}$$

where $\tilde{y}$ is the model output and u the manipulated variable. $\tilde{d}(k)$ describes the expected disturbance effect on the output. H_i denotes the impulse response coefficients with N, the truncation number, assuming that $H_i = 0$ for $i > N$. The process is assumed to be open-loop stable. Other types of models can also be used, such as step response models (Garcia and Morshedi, 1986) and state space descriptions (Li et al., 1989; Ricker, 1990). The z-transfer function, $\tilde{p}^*(z)$, describing the process model is related to Eq. 1 through

$$\tilde{p}^*(z) = \sum_{i=1}^{N} H_i z^{-i} \tag{2}$$

QDMC minimizes a quadratic objective function of the squares of the predicted error e (setpoint minus predicted output), the manipulated variable u, and its rate of change $\Delta u(l) \triangleq u(l) - u(l-1)$ over a finite horizon P in the future:

$$\min_{\Delta u(k),\ldots,\Delta u(k+M-1)} \sum_{l=k+1}^{k+P} [\Gamma^2 e(l)^2 + B^2 u(l-1)^2 + D^2 \Delta u(l-1)^2] \tag{3}$$

where k denotes the present sampling point. B, D, and Γ are penalty weights. These weights can, in general, be time-varying; here they are assumed constant. M is the number of moves in the future to be optimized. At time k, it is assumed that $u(k+M-1) = u(k+M) = \ldots = u(k+P-1)$. Since only $u(k)$ will be implemented and a new optimization problem is solved at the next sampling point, the input will not necessarily remain constant after M steps.

One can add constraints on any of the future values of the manipulated variable and output. In this article, only output constraints are considered. For the output constraints, a *constraint window* is specified. This means that the *predicted* output values must respect the bounds for every point of the constraint window. Hence, the on-line optimization at k is subject to output constraints at points $k + w_b, \ldots, k + w_e$, where w_b and w_e denote the beginning and ending points of the constraint window. Note that this prediction is based on the model and does not guarantee that the future values will actually be within the prescribed bounds. Unmeasured disturbances and model error may result in violation, although the on-line optimization finds feasible solutions.

This constrained optimization problem can be written as a standard *QP*, as shown in Garcia and Morshedi (1986):

$$\min_X \Phi(X) = \frac{1}{2} X^T G X + g^T X \tag{4}$$

subject to

$$A^T X \geq b \tag{5}$$

where

$$X = [\Delta u(k) \ldots \Delta u(k+M-1)]^T \tag{6}$$

G, g, A, and b are functions of the following:

$$\begin{cases} G = G(M,P,B,D,\Gamma,H_1,\ldots,H_N), \\ g = g[M,P,B,\Gamma,H_1,\ldots,H_N,y(k),u(k-1),\ldots,u(k-N)], \\ A = A(M,P,H_1,\ldots,H_N), \\ b = b[H_1,\ldots,H_N,y(k),u(k-1),\ldots,u(k-N),y_L,y_U] \end{cases} \tag{7}$$

where $y(k)$ is the current output (measurement), and y_L, y_U are the lower and upper output constraint bounds, respectively.

The optimal solution X^* satisfies (Fletcher, 1981):

$$\begin{bmatrix} G & -\hat{A} \\ -\hat{A}^T & 0 \end{bmatrix} \begin{bmatrix} X^* \\ \lambda^* \end{bmatrix} = - \begin{bmatrix} g \\ \hat{b} \end{bmatrix} \tag{8}$$

where $\hat{A}^T$ and $\hat{b}$ are the rows of A^T, b for the constraints that are active at the optimum, and λ^* is the vector of Lagrange multipliers that correspond to these constraints.

The control action is found from Eq. 8:

$$u(k) = [1 \ 0 \ \ldots \ 0] X^* + u(k-1) \tag{9}$$

Stability of Constrained QDMC

After solving the QP, one finds that some constraints are predicted to be active at the optimum. Let $J_1, \ldots, J_n$ be sets,

each of which includes a combination of constraints that can become active at the optimum of the on-line optimization while operating the control system. Note that by simply considering all possible combinations of constraints one may expect n to be very large for the general multi-input multi-output (MIMO) case; however, only a few J_i's can actually occur. Zafiriou and Chiou (1989) proposed a procedure for finding all practically relevant J_i's.

Example 1

Let us specify an output constraint at the next two sample times, $w_b=1$ and $w_e=2$:

$$-0.3\le\tilde{y}(k+1)\le 0.3 \text{ and } -0.3\le\tilde{y}(k+2)\le 0.3 \quad (10)$$

The possible J_i's are:
J_0: no constraint active
J_1: lower constraint active at $k+1$
J_2: upper constraint active at $k+1$
J_3: lower constraint active at $k+2$
J_4: upper constraint active at $k+2$
J_5: both lower constraints active
J_6: both upper constraints active
J_7: upper constraint at $k+1$ and lower constraint at $k+2$ active
J_8: lower constraint at $k+1$ and upper constraint at $k+2$ active

For linear model dynamics, Zafiriou (1990) showed that the *constrained* QDMC is *piecewise linear*, meaning that the dynamics of QDMC for a certain constraint set J_i active, are those of a discrete linear controller. This linear controller, denoted $c_{J_i}(z)$, depends explicitly only on J_i; it depends implicitly on the past and current values of the plant inputs and outputs. These values together with external inputs (setpoints, disturbances) determine the J_i that corresponds to a sampling point. However, if at different sampling points the QP solution results in the same J_i, the QDMC dynamics at those points are those of the *same* linear controller.

Let us define:

$$x_j(k)\triangleq u(k-j), \quad 1\le j\le N \quad (11)$$

and let f describe the QP solution given by Eq. 9 for a particular J_i (which determines the rows of A^T, b that are included in $\hat{A}^T$, $\hat{b}$):

$$f[y(k),u(k-1),\ldots,u(k-N);$$
$$r_P(k),d(k)]_{J_i}\triangleq[1\ 0\ \ldots\ 0]X_{J_i}^*+u(k-1) \quad (12)$$

where $r_P(k)$ includes all the values of the setpoint from $k+1$ to $k+P$, and $d(k)$ is the disturbance effect at the output at k. Zafiriou (1990) showed that the c_{J_i} controller is given by:

$$c_{J_i}(z)=\frac{-(\nabla_y f)_{J_i}}{1-(\nabla_{x_1}f)_{J_i}z^{-1}-\ldots-(\nabla_{x_N}f)_{J_i}z^{-N}} \quad (13)$$

Let us define the "state" of the system as $x(k)\triangleq[x_1(k),\ldots,x_N(k)]^T$ where the x_j's are those of Eq. 11. Knowledge of $x(k)$ and of the external inputs, $r_P(k)$ and $d(k)$, allows the computation of $x(k+1)$ by applying the plant and controller equations on it. Let us denote by F the operator that maps $x(k)$ to $x(k+1)$:

$$x(k+1)=F[x(k);r_P(k),d(k)] \quad (14)$$

Let us define the transfer functions:

$$Q_{J_i}(z)\triangleq\frac{-(\nabla_y f)_{J_i}}{1-(\psi_1)_{J_i}z^{-1}-\ldots-(\psi_N)_{J_i}z^{-N}}, \quad 1\le j\le N \quad (15)$$

where

$$(\psi_j)_{J_i}\triangleq(\nabla_{x_j}f)_{J_i}+(\nabla_y f)_{J_i}H_j, \quad 1\le j\le N \quad (16)$$

Then, the following theorem holds.

Theorem 1 (Zafiriou, 1990). F can be α contraction for all plants in a set Π, only if all feedback controllers $c_{J_i}(z)$, for i such that $(\nabla_y f)_{J_i}\neq 0$, stabilize all unconstrained plants in the set Π, and all transfer functions $Q_{J_i}(z)$, for i such that $(\nabla_y f)_{J_i}=0$, are stable.

Theorem 1 allows one to handle any set Π that robust linear control theory can (for a discussion of common types of Π see Morari and Zafiriou, 1989). One should note that F being a contraction implies stability of the nonlinear, closed-loop system. However, a violation of the necessary contraction condition does not always imply instability, but it must be considered as a warning that the control parameters should be modified. A sufficient condition (Zafiriou, 1990) can be derived but it is often conservative. The examples in the next sections demonstrate that, at least for SISO systems, the necessary condition is a very good indicator of stability.

It can be shown that for some J_i's the corresponding c_{J_i}'s are identical; this reduces the number of cases for which one has to check stability.

Theorem 2. Let J_a be a set of active output constraints of the QP, and let J_b be another set of constraints that correspond to the same points in the constraint window as the ones in J_a, but some of them correspond to opposite bounds being reached. Then, $c_{J_a}(z)$ and $c_{J_b}(z)$ are identical.

Proof. See Appendix A for proof. This theorem can easily be generalized to include constraints on the manipulated variables and rate of change of the manipulated variables for MIMO processes.

Example 2

Theorem 2 applied to the system of example 1 yields: $c_{J_1}=c_{J_2}$, $c_{J_3}=c_{J_4}$, $c_{J_5}=c_{J_6}=c_{J_7}=c_{J_8}$.

Special Cases

This section considers the effect of some choices of output constraints and tuning parameters on closed-loop stability. These cases derive their importance from the fact that they may be considered "natural" or "safe" choices. The procedure of using the previously mentioned conditions to predict potential problems is explained in the following.

Constraint at the first possible point

The output constraint window is often defined to start at

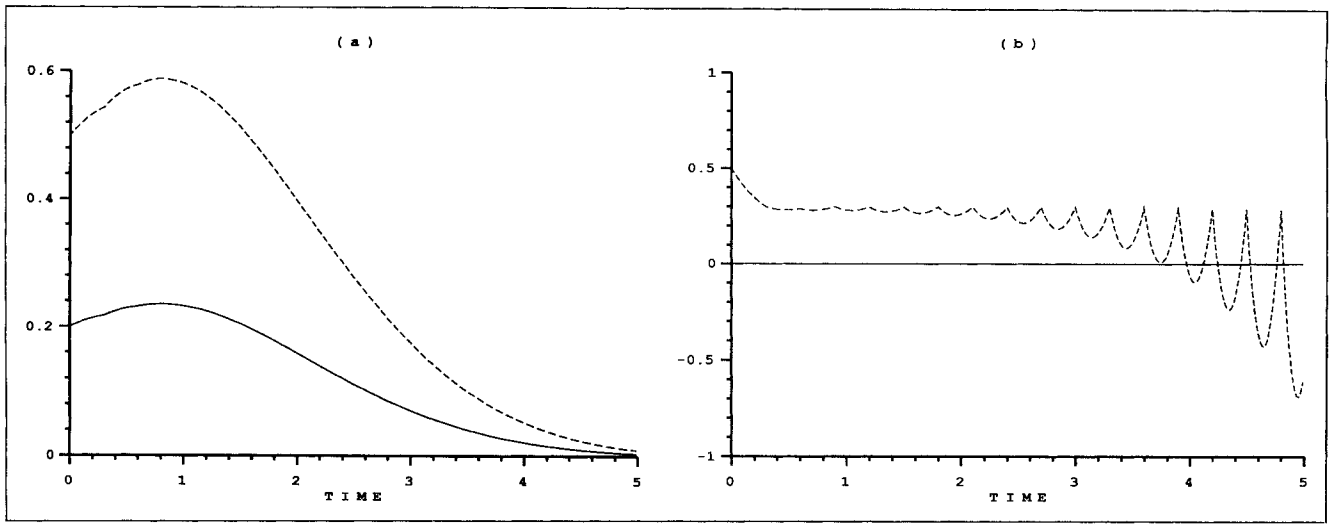

Figure 1. Example 3: (a) unconstrained; (b) constrained.

Solid line in (a), output response for $d(s)=0.2/s$ for both unconstrained and constrained; dashed line, output response for $d(s)=0.5/s$.

the earliest possible future point, by selecting w_b equal to the dead-time plus one (see Prett and Garcia, 1988). This choice, however, can result in a very aggressive controller. The following theorem quantifies this observation for SISO systems.

Theorem 3. Let the dead-time in the discrete model be N_d sampling intervals: $H_1=\ldots=H_{N_d}=0$ and $H_{N_d+1}\neq 0$. For the J_i set that corresponds to either the upper or lower output constraint at $k+N_d+1$, $c_{J_i}(z)$ is a (deadtime plus) one-step-ahead feedback controller $c_{J_i}(z)=(z^{N_d+1}-1)^{-1}(\tilde{p}^*(z))^{-1}$.

Proof. See Appendix B for proof. Computation of the equivalent Internal Model Control (IMC) structure (e.g., Morari and Zafiriou, 1989) controller $q(z)$ yields for the c_{J_i} of theorem 3:

$$q(z)=\frac{c_{J_i}}{1+\tilde{p}^* c_{J_i}}=z^{-(N_d+1)}\tilde{p}^{*-1} \tag{17}$$

This IMC controller inverts the causal part of the model and is the "perfect" controller (Garcia and Morari, 1982), known to suffer from a number of problems. A direct consequence is that stability problems will arise if the discrete model has zeros outside the Unit Circle. Also, this controller is likely to destabilize the system if there is a plant-model mismatch.

Example 3. Consider the following process:

$$\tilde{p}(s)=\frac{-s+1}{(s+1)(2s+1)} \tag{18}$$

with output constraint: $-0.3\leq\tilde{y}\leq 0.3$. A sampling time $T=0.3$ and a truncation number $N=40$ are used to get the discrete model $\tilde{p}^*(z)$. The discrete model has a zero outside the unit circle. Thus, if the constraint window includes point $k+1$ ($w_b=1$), Theorem 3 predicts that in the operating regions of J_1 and J_2 (as defined in example 1) the control system will behave as an unstable controller. The simulations of Figure 1 demonstrate this fact. The responses of both the unconstrained and the constrained QDMC are shown for two step output disturbances $d(s)$ of different sizes; the setpoint is zero, and $w_b=w_e=1$ for the constrained case. For the smaller disturbance, the response is the same for both cases, and the system is stable because it remains in the operating region for J_0 (unconstrained), in which the choice of parameters, $M=5$, $P=50$, $\Gamma=1$, $B=0$, and $D=1$, is good in terms of stability. For the larger disturbance, in the unconstrained case the response is simply appropriately scaled, when compared to the smaller disturbance. In the constrained case, however, the system moves in the J_1/J_2 region and instability occurs as predicted. Note that if one looks only at the sampling points, the instability is not apparent at the output because of exact pole/zero cancellation; the input (not plotted), however, does grow exponentially. Also, the introduction of a slight model-plant mismatch will drive the discrete output unstable as well. The nonlinear behavior is the result of the inclusion of the output constraint in the on-line optimization problem. No "physical" bound is reached.

Small number of input moves (M)

The choice of a small M is considered "safe," especially in conjunction with the selection of a large prediction horizon P. This is valid for unconstrained systems, where these choices detune the controller. The theory behind them can be summarized in the following theorem:

Theorem 4 (Garcia and Morari, 1982; Rephrased). Let $B=0$, $D=0$, and $\Gamma=1$. Then for no model-plant mismatch: (i) for $M=1$ and a sufficiently large $P>N+M-1$, the unconstrained system is stable; (ii) if the model has a discrete monotonic step response, $M=1$ and $P=N$ result in a stable unconstrained control system.

We shall now consider the case $M=1$, when output constraints are included in the on-line optimization. The following theorem gives the expression for the c_{J_i} that corresponds to one of the constraints on the future output values becoming active at the optimum of the QP. (Note that since $M=1$, no more than one constraint can be active, assuming linear independence and feasibility for the QP.)

Theorem 5. Let $M=1$. Then, for the J_i set that corresponds to the upper or lower constraints on the output at point $k+N_a$ in the future, where $N_a\geq N_d+1$, we have:

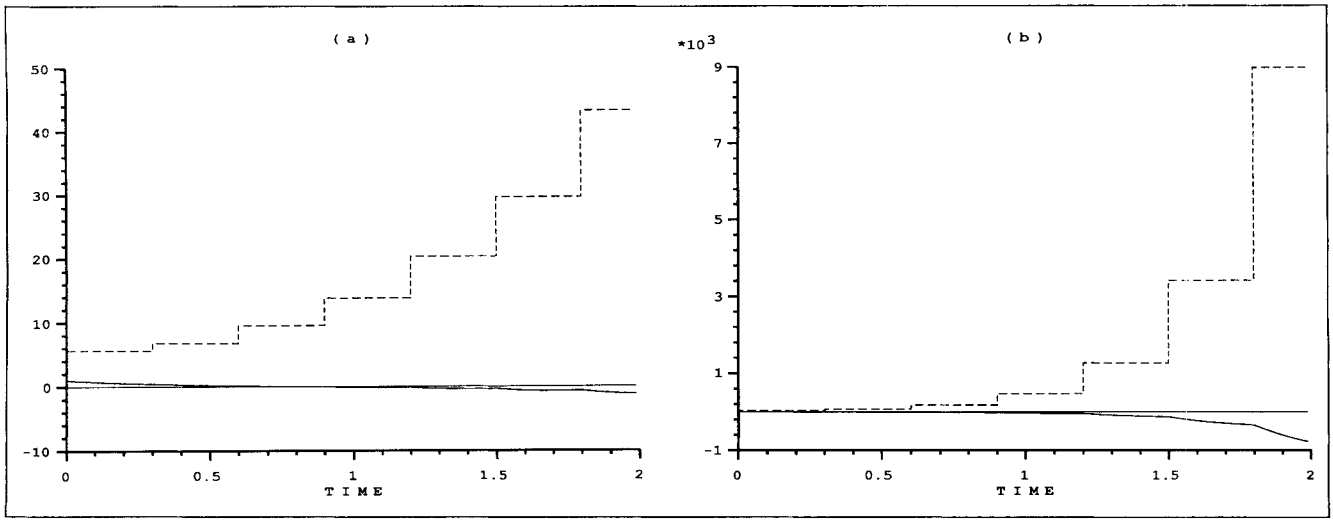

Figure 2. Example 4: (a) $M=1$; (b) $M=2$.

Solid line, output; dashed line, manipulated variable.

$$c_{J_i}(z)=\frac{1}{z^{N_a}\tilde{p}^*_{N_a}-\tilde{p}^*} \tag{19}$$

where we define

$$\tilde{p}^*_{N_a}(z)=\sum_{j=1}^{N} H_{N_a,j}z^{-j}$$

with:

$$H_{N_a,j}=\begin{cases} 0 & j<N_a \\ \sum_{l=1}^{N_a} H_l & j=N_a \\ H_j & j>N_a \end{cases} \tag{20}$$

Proof. See Appendix C for proof. For $N_a=1$ this reduces to theorem 3. The examples of this section consider the case $N_a>1$ to demonstrate that the potential destabilization by output constraints is not limited to the extreme case $N_a=1$. One can see that the feedback controller given by Eq. 19 does not depend on P, nor on the objective function penalty weights Γ, D, B. As the following examples demonstrate, this results in an unstable constrained controller. The examples also examine the case $M=2$ to show that the problem is not limited to $M=1$.

Example 4. Consider the process of example 3. This time we will study the effect of the constraint at $k+2$ becoming active at the optimum of the optimization solved at k. The roots of $[1+\tilde{p}^*(z)c_{J_i}(z)]$ are computed for the c_{J_i} of Eq. 19. For stability they have to be inside the unit circle. The computation is repeated for $M=2$. In the second case, c_{J_i} depends on P, so we used symbolic computation software to compute the roots at the limit $P\to\infty$. The values of the other parameters are $B=0$, $\Gamma=1$, and $D=0$. The largest absolute value of a root is:

- For $M=1$: 1.45
- For $M=2$: 2.63.

For the simulations, we assume that $P=50>N+M-1$ is a good approximation of $P\to\infty$. Indeed, the largest root is 2.67 for $M=2$, and 1.45 for $M=1$. Thus, we expect potential instability when the output constraint becomes active. This behavior is observed in Figure 2 for an output disturbance $d(s)=1/s$ and $w_b=w_e=2$.

The next example considers a system with discrete monotonic step response.

Example 5. Let:

$$\tilde{p}(s)=\frac{-s+1+e^{-4s}(2s+1)}{(s+1)(2s+1)} \tag{21}$$

The continuous response exhibits inverse response. Part (ii) of theorem 4 is valid only for discrete monotonic step response systems. For a sampling time $T=2$ this is true. The truncation number is $N=15$.

The absolute value of the largest closed-loop pole when using the unconstrained QDMC controller with $M=1$ is 0.29. Hence, the unconstrained system is stable. Let us specify an output constraint at $k+2$ and compute the largest absolute values of the roots of $[1+\tilde{p}^*(z)c_{J_i}(z)]$ for the J_i that corresponds to this constraint being active:

- For $M=1$: 1.54
- For $M=2$: 2.18.

This prediction of instability is verified in the simulations shown in Figure 3. An output constraint $-0.3\le\tilde{y}(k+2)\le 0.3$ is used in the on-line optimization. The parameter values are $M=1$ or 2, $P=15$, $B=0$, $D=0$, and $\Gamma=1$. At time 0, a step disturbance $d(s)=2/s$ is applied to the plant output.

The following theorem provides a sufficient condition for the c_{J_i} that corresponds to an output constraint at $k+N_a$ to be closed-loop stable. We use the notation S_j to denote the value of the unit step response of $\tilde{p}^*(z)$ at time jT. The step response coefficients are related to the impulse response coefficients through

$$S_{j+1}-S_j=H_{j+1} \tag{22}$$

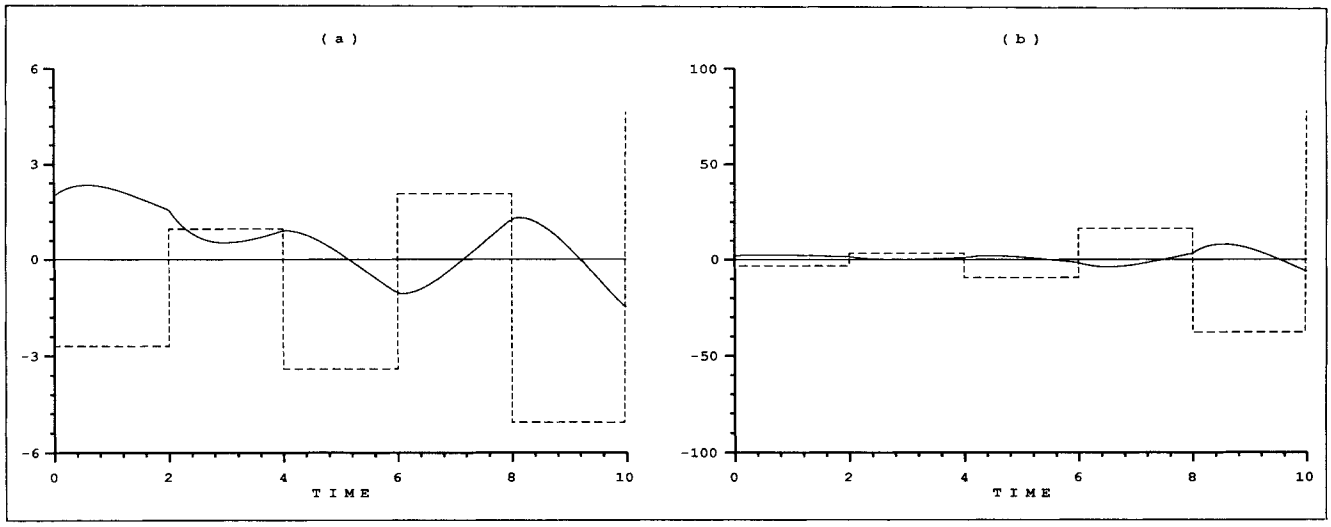

Figure 3. Example 5: (a) $M=1$; (b) $M=2$.

Solid line, output; dashed line, manipulated variable.

Theorem 6. Let $M=1$ and J_i correspond to the upper or lower constraint on the output at point $k+N_a$ in the future, where $N_a \geq N_d+1$. Then, c_{J_i} yields a stable, closed-loop system if both:

(i) $S_j \leq S_{j+1}$ if $\tilde{p}^*(1)>0$ or $S_j \geq S_{j+1}$ if $\tilde{p}^*(1)<0$ for $j=N_a, \ldots, N-1$.

(ii) $S_{N_a}>\tilde{p}^*(1)/2$ if $\tilde{p}^*(1)>0$ or $S_{N_a}<\tilde{p}^*(1)/2$ if $\tilde{p}^*(1)<0$.

Proof. See Appendix D for proof. The theorem states that if the open-loop unit step response of the model is examined at time N_aT and it is found that (i) the response is monotonic from that point on and that (ii) it has reached a value equal to at least half the steady-state value, then placing an output constraint at $k+N_a$ in QDMC will not cause nominal stability problems. The reader is encouraged to compare this result to proposition 1 in Scattolini and Bittanti (1990). Although the predictive controller studied in that article is unconstrained, it so happens that the stability question for that linear control system is mathematically the same as the one under consideration in theorem 6.

Large input penalty weight (B)

Another "safe" choice for the unconstrained case is that of a large weight B in the u penalty term in the objective function. This is based on the following theorem.

Theorem 7 (Garcia and Morari, 1982). There exists a finite $\beta>0$ such that for $B>\beta$ and for no model-plant mismatch, the unconstrained control system is stable for any $M \geq 1$, $P \geq N_d+1$, $D=0$, and $\Gamma \geq 0$.

The following counter-example shows that this is not always true with output constraints, even for a system that has no inverse response and without putting a constraint at the first possible point.

Example 6

Let:

$$\tilde{p}(s)=\frac{1}{(5s+1)(10s+1)(15s+1)(20s+1)} \tag{23}$$

A sampling time $T=3$ is used to get the discrete model $\tilde{p}^*(z)$. A truncation number $N=50$ is used. Note that $\tilde{p}^*(z)$ has a zero outside the unit circle at -7.8, which means that the one-step-ahead controller is unstable. Hence, as discussed earlier, a constraint at point $k+1$ will result in an unstable control system. We will consider the case with the constraint placed at point $k+2$.

Next, we compute the largest absolute value of the roots of $[1+\tilde{p}^*(z)c_{J_i}(z)]$ for the J_i that corresponds to the upper or lower constraint at $k+2$ for several values of B. The other parameters are set to $M=2$, $P=50$, $D=0$, and $\Gamma=1$. These values are listed in Table 1, which also includes the corresponding values for the unconstrained case for comparison. For the unconstrained case, the control system is stable for all values of B. When a constraint is placed at $k+2$, however, the stability condition predicts that the closed-loop system is slightly unstable. Increasing B even further makes no difference. Figure 4 shows the response to a unit step output disturbance for $B=1{,}000$, when the constraint $-0.3 \leq \tilde{y}(k+2) \leq 0.3$ is used. The predicted instability is verified. The simulations are identical for the other values of B in Table 1. Note that if a $B=1{,}000$ was used without the output constraint, essentially no control action would be taken due to the huge penalty on the manipulated variable. The presence of a constraint, even at the second future point, results in very large values for the manipulated variable and in instability.

The above example also serves to demonstrate that even in cases of processes without inverse response characteristics, one cannot assume that nominal stability is guaranteed. The reason is that after sampling, the discrete model $\tilde{p}^*(z)$ often has zeros outside the unit circle. For rational, continuous transfer functions with at least three more poles than zeros, this will always

Table 1. Example 6: Largest Absolute Value of the Roots of $[1+\tilde{p}^*(z)c_{J_i}(z)]$ for Different Values of B

B	2	10	100	1,000
Unconstrained	0.845	0.776	0.685	0.639
Constraint at $k+2$	1.013	1.012	1.011	1.011

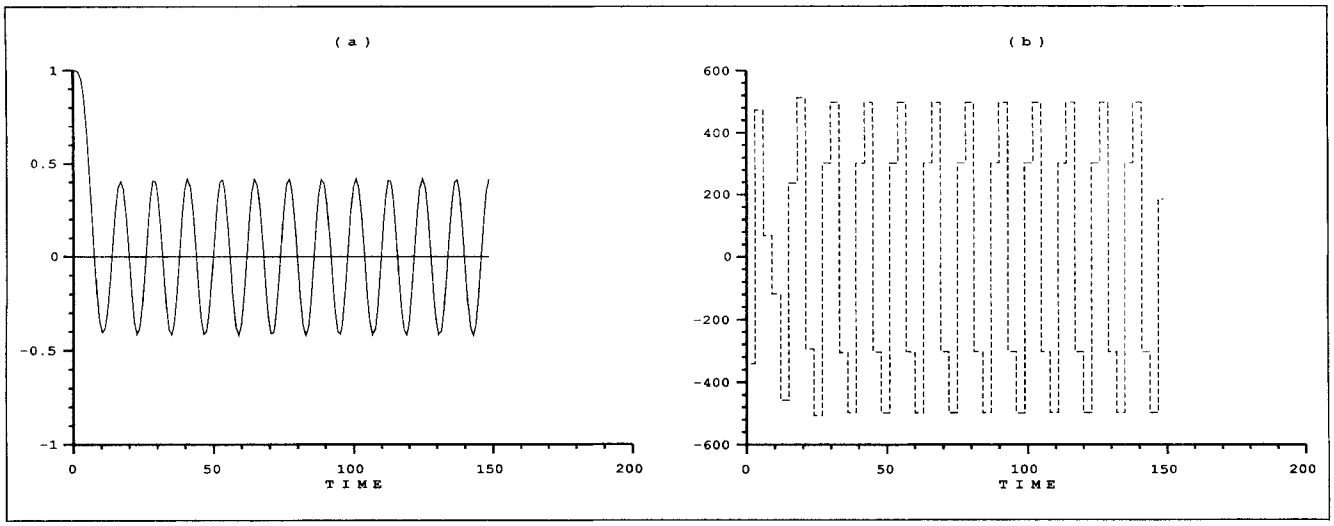

Figure 4. Example 6: (a) output; (b) manipulated variable.

$B = 1{,}000$; constraint at $w_b = w_e = 2$.

happen for small sampling times. In such cases, the one-step-ahead controller will be unstable, and stability problems that can be caused by output constraints are likely to appear. One should note, however, that such problems may also occur at larger sampling times. The effect of sampling time on the location of the zeros of discrete models is discussed rigorously by Astrom et al. (1984). Zafiriou and Morari (1985) present in detail the effect of sampling time selection on the performance of some standard discrete linear control algorithms.

Robust Design

This section demonstrates the use of the contraction condition in designing a robust QDMC controller with output constraints.

We select a simple first-order system with a time delay:

$$p(s) = \frac{e^{-\tau_d s}}{s+1} \tag{24}$$

For the dead time, we use the value $\tau_d = 1$ in the model $\tilde{p}(s)$ and allow up to 15% error in the true deadtime. This error can be described as additive error $\bar{l}_a^*$ in the discrete model. This computation was made by Zafiriou and Morari (1986). The set Π of possible plants is described as:

$$\Pi = \{p(s): |p^*(e^{j\omega T}) - \tilde{p}^*(e^{j\omega T})| \le \bar{l}_a^*(\omega),\ 0 \le \omega \le \pi/T\} \tag{25}$$

where $p^*(z)$ is the discrete plant, $\tilde{p}^*(z)$ the discrete model, and ω is the frequency.

Robust linear control theory provides a robust stability condition for a linear controller applied to a plant in this type of set Π (Zafiriou and Morari, 1986). Hence, theorem 1 reduces, in this case, to the requirement that all feedback controllers $c_{J_i}(z)$ stabilize the linear (with all constraints removed), closed-loop system for the model $\tilde{p}^*(z)$ and satisfy:

$$\mu_{J_i} \triangleq \left| \frac{c_{J_i}(e^{j\omega T})}{1 + \tilde{p}^*(e^{j\omega T}) c_{J_i}(e^{j\omega T})} \right| \bar{l}_a^*(\omega) < 1,\ 0 \le \omega \le \frac{\pi}{T} \tag{26}$$

In the simulations we introduce model-plant mismatch by using:

$$p(s) = \frac{e^{-0.85s}}{s+1} \tag{27}$$

as the actual plant. A sampling time $T = 0.1$ is used.

The set point is 0 and the constraint specification on the output is: $-0.3 \le \tilde{y} \le 0.3$. The theoretical tools that were developed in the previous sections apply to constraint windows of any length; however, we choose the simple case of a window length of 1 ($w_b = w_e$) to highlight the effect of the constraint. Since our model has a dead-time of ten sampling intervals, the beginning point w_b cannot be less than 11.

The J_i corresponding to a constraint at $k+11$ results in the c_{J_i} of theorem 3. Although this is stable for no model error for this process, it is very sensitive to model error. This is shown by the robustness condition μ_{J_i} plotted as a function of frequency in Figure 5. Since this controller is independent of the tuning parameters, the condition cannot be reduced to below 1 by tuning. Hence, we cannot impose a constraint at the first possible point because of model uncertainty.

We will now study the case, where the constraint is placed at time $k+12$ ($w_b = w_e = 12$). From theorem 5 we know that for $M = 1$, the other tuning parameters have no effect on c_{J_i}. Hence, we will use M greater than or equal to 2.

Let us first consider the following values: $M = 2$, $P = 20$, $B = 0$, $D = 0.13$, and $\Gamma = 1$. Figure 5 shows that the robustness condition is satisfied for the unconstrained QDMC, but the c_{J_i} corresponding to the constraint is not robustly stable. Let us observe how the system rejects a step output disturbance. First, let us simulate the response to $d(s) = 2/s$. Figure 6 shows that the closed-loop system is stable, although close to instability with a response that may not be acceptable in practice. One can verify that the peaks of the values of the manipulated

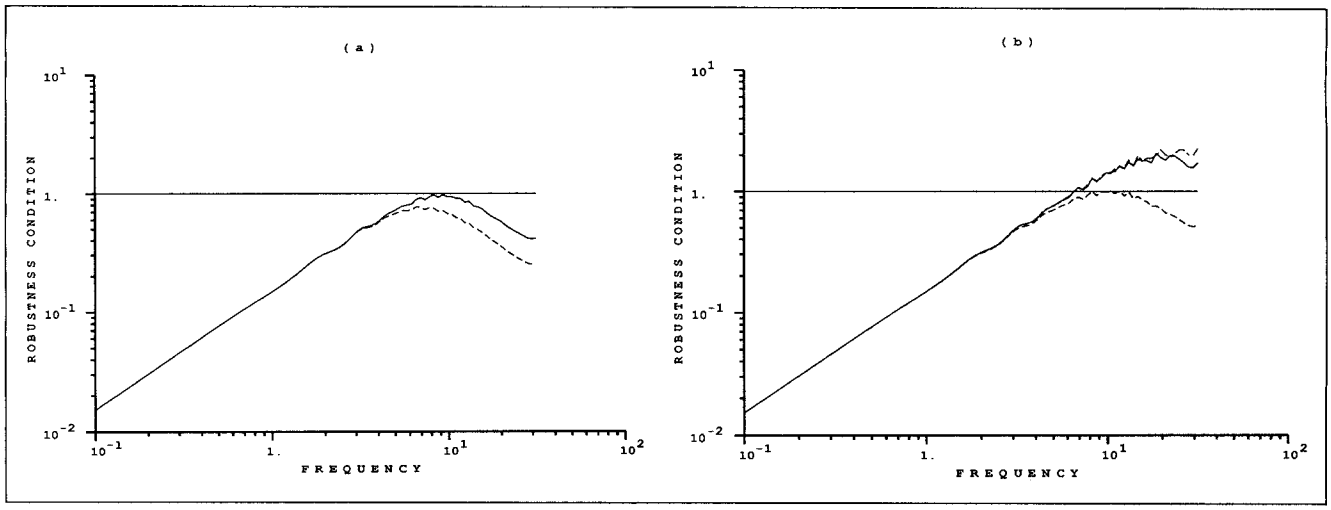

Figure 5. Robustness condition (μ_{J_i}) for up to 15 percent dead time error: (a) no constraint active; (b) constraint at $w_b(= w_e)$ active.

(a) Solid line, $M=2$, $D=0.13$; dashed line, $M=5$, $D=0.20$; (b) dash and dot, $w_b=w_e=11$; solid line, $w_b=w_e=12$, $M=2$, $D=0.13$; dashed line, $w_b=w_e=12$, $M=5$, $D=0.20$.

variable occur when the constraint is predicted active. Next, increase the disturbance to $d(s)=2.5/s$. For the unconstrained case, as expected from linear control theory, the response is simply scaled appropriately and therefore not plotted. However, as Figure 7 shows, the constrained controller is unstable.

Next, we tune the parameters to satisfy the robustness condition. After a trial-and-error procedure, we select $M=5$, $P=20$, $B=0$, $D=0.2$, and $\Gamma=1$. The output constraint is still at $k+12$. To satisfy the contraction condition, first we need to have all the roots of $[1+\tilde{p}^*(z)c_{J_i}(z)]$ inside the unit circle for both J_i's. We can compute that for the new parameter values, the magnitude of the largest root is 0.9048 when the constraint is active and 0.9050 for the unconstrained case. Second, we have to plot μ_{J_i} for both J_i's and check if they are both below 1. One can see in Figure 5 that the robustness condition is satisfied for the unconstrained controller with a relatively large "margin." The tuning parameter values are selected because they satisfy the condition for the J_i that corresponds to the constraint being active as well. These results are confirmed by the simulations shown in Figure 7. Thus, the contraction condition proved efficient in designing a robustly stable QDMC controller with output constraint.

Concluding Remarks

The theory and examples presented in this article show that the inclusion of hard output constraints in the on-line optimization problem solved by QDMC may result in very aggressive control systems that are unstable or sensitive to modeling error. This should not lead one to conclude that output constraints ought not to be used in QDMC. The theory can predict when a stability problem exists and usually offers a way to address it. What is clear though is that one should not rely on tuning rules that work for the unconstrained case. The stability guarantees behind these rules are no longer valid. In the case of the "small M" rule, not only there is no stability guarantee any more, but the rule actually suggests values for M that are inadvisable for the constrained case.

An important aspect of the robust design methodology is the fact that the design is based on the effect of the QDMC

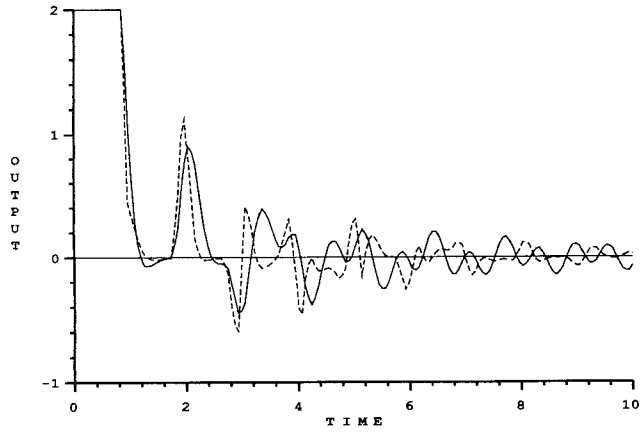

Figure 6. $d(s)=2/s$, $M=2$, $D=0.13$.

Solid line, unconstrained controller; dashed line, $w_b=w_e=12$.

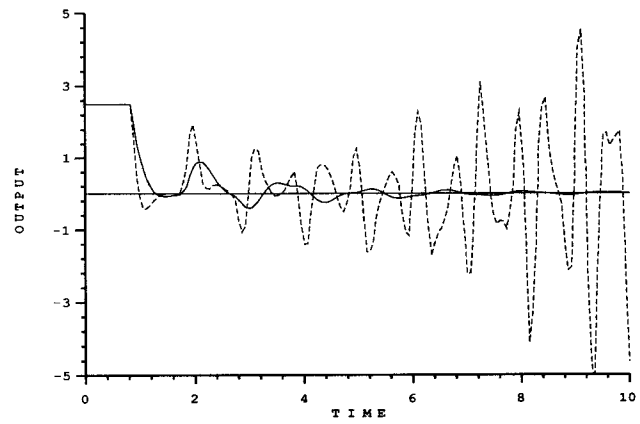

Figure 7. $d(s)=2.5/s$, $w_b=w_e=12$.

Dashed line, $M=2$, $D=0.13$; solid line, $M=5$, $D=0.20$.

parameters on the robustness conditions for the constrained, closed-loop system and requires no exhaustive simulations to check all possible situations. All the simulations given are used to confirm the behavior predicted by the stability conditions. One should note that the simulations are kept as simple as possible to make the nonlinearity of the control system clear.

Future work in this area will look at the effect of "softening" the constraints. As suggested by Ricker et al. (1989), allowing the constraints to be violated should in general result in less aggressive control actions. An open question is how to determine the "degree" of softening, without having to resort to the usually impossible task of simulating all possible cases. The framework used in this article is ideal for this task.

Acknowledgment

Partial support for this project was provided through the National Science Foundation's Engineering Research Centers Program NSFD CDR 8803012 and a grant from Shell Development Co. One of the authors (A. L. Marchal) wishes to thank Rhône-Poulenc for financial support during his graduate studies.

The authors are grateful to H.-W. Chiou for his valuable help with the software package QDMC, developed at the University of Maryland (Dr. Zafiriou's group). Modified software from the control system design package CONSYD, developed at Caltech (Dr. Morari's group) and the University of Wisconsin (Dr. Ray's group), was also used in the simulations.

Notation

A^T = matrix of inequality constraints
b = vector of bounds of inequality constraints
B = penalty weight of manipulated variable
$c_{J_i}(z)$ = feedback controller for J_i
$d(l)$ = disturbance effect on the plant output at lth sampling point
$\bar{d}(l)$ = estimated disturbance on the model output at lth sampling point
$d(s)$ = output disturbance transfer function
D = penalty weight on change of manipulated variable
f = quadratic program solution
F = operator mapping $x(l-1)$ to $x(l)$
g^T = linear term vector in objective function
G = Hessian
H_i = ith impulse response coefficient
J_i = set of active constraints
k = present sampling point
M = number of input moves to be optimized
N = truncation number
N_d = dead time in sampling intervals
$\tilde{p}^*(z)$, $p^*(z)$ = discrete model and plant transfer functions
$\tilde{p}(s)$, $p(s)$ = continuous model and plant transfer functions
P = prediction horizon
$q(z)$ = IMC controller
$Q_{J_i}(z)$ = defined in Eq. 15
$r_P(l)$ = vector of P future setpoint values at lth sampling point
S_i = ith step response coefficient
T = sampling time
$u(l)$ = manipulated variable at lth sampling point
w_b, w_e = beginning and ending points of constraint window
$x(l)$ = closed-loop system state vector at lth sampling point
X = vector of optimization variables
$y(l)$ = output measurement at lth sampling point
$\tilde{y}(l)$ = model output at lth sampling point
y_L, y_U = lower and upper output bounds

Greek letters

Γ = penalty weight on predicted output error
$\Delta u(l)$ = $u(l)-u(l-1)$
μ_{J_i} = defined in Eq. 26
Π = set of possible plants
$(\psi_j)_{J_i}$ = defined in Eq. 16
ω = frequency

Subscripts and superscripts

J_i = corresponding to set J_i
T = transpose
* = corresponding to optimal solution or discrete transfer function
$\hat{}$ = corresponding to active constraints

Literature Cited

Astrom, K. J., P. Hagander, and J. Sternby, "Zeros of Sampled Systems," *Automatica*, **20**, 31 (1984).

Campo, P. J., and M. Morari, "Robust Control of Processes Subject to Saturation Nonlinearities," *Comp. Chem. Eng.*, **14**, 343 (1990).

Eaton, J. W., and J. B. Rawlings, "Feedback Control of Chemical Processes Using On-Line Optimization Techniques," *Comp. Chem. Eng.*, **14**, 469 (1990).

Fletcher, R., *Practical Methods of Optimization: 2. Constrained Optimization*, Wiley (1981).

Garcia, C. E., "Quadratic Dynamic Matrix Control of Nonlinear Processes: An Application to a Batch Reaction Process," AIChE Meeting, San Francisco (1984).

Garcia, C. E., and M. Morari, "Internal Model Control: 1. A Unifying Review and Some New Results," *Ind. Eng. Chem. Proc. Des. Dev.*, **21**, 308 (1982).

Garcia, C. E., and A. M. Morshedi, "Quadratic Programming Solution of Dynamic Matrix Control (QDMC)," *Chem. Eng. Commun.*, **46**, 73 (1986).

Jury, E. I., *Theory and Application of the Z-Transform Method*, Wiley, Chichester (1964).

Li, W. C., and L. T. Biegler, "Multistep, Newton-Type Control Strategies for Constrained, Nonlinear Processes," *Chem. Eng. Res. Des.*, **67**, 562 (1989).

Li, S., K. Y. Lim, and D. G. Fisher, "A State Space Formulation for Model Predictive Control," *AIChE J.*, **35**, 241 (1989).

Morari, M., and E. Zafiriou, *Robust Process Control*, Prentice-Hall, Englewood Cliffs, NJ (1989).

Peterson, T., E. Hernandez, Y. Arkun, and F. J. Schork, "A Nonlinear DMC Algorithm and its Application to a Semibatch Polymerization Reactor," *Chem. Eng. Sci.*, in press (1991).

Prett, D. M., and C. E. Garcia, *Fundamental Process Control*, Butterworth, Stoneham, MA (1988).

Ricker, N. L., "Use of Quadratic Programming for Constrained Internal Model Control," *Ind. Eng. Chem. Proc. Des. Dev.*, **24**, 925 (1985).

Ricker, N. L., "Model Predictive Control with State Estimation," *Ind. Eng. Chem. Res.*, **29**, 374 (1990).

Ricker, N. L., T. Subrahmanian, and T. Sim, "Case Studies of Model-Predictive Control in Pulp and Paper Production," *Model Based Process Control—Proc. IFAC Workshop*, T. J. McAvoy, Y. Arkun, and E. Zafiriou, eds., Pergamon Press, Oxford (1989).

Scattolini, R., and S. Bittanti, "On the Choice of the Horizon in Long-Range Predictive Control—Some Simple Criteria," *Automatica*, **26**, 915 (1990).

Zafiriou, E., "Robustness and Tuning of On-Line Optimizing Control Algorithms," *Model-Based Process Control—Proc. IFAC Workshop*, T. J. McAvoy, Y. Arkun and E. Zafiriou, eds., Pergamon Press, Oxford (1989).

Zafiriou, E., "Robust Model Predictive Control of Processes with Hard Constraints," *Comp. Chem. Eng.*, **14**, 359 (1990).

Zafiriou, E., and H.-W. Chiou, "An Operator Control Theory Approach to the Design and Tuning of Constrained Model Predictive Controllers," AIChE Meeting, San Francisco, Paper 21e (1989).

Zafiriou, E., and M. Morari, "Digital Controllers for SISO Systems: a Review and a New Algorithm," *Int. J. Control*, **42**, 855 (1985).

Zafiriou, E., and M. Morari, "Design of Robust Digital Controllers and Sampling-Time Selection for SISO Systems," *Int. J. Control*, **44**, 711 (1986).

Appendix A: Proof of Theorem 2

If a lower or upper bound on the output is specified for the future time $k+l$, the corresponding expressions of this output constraint in the QP formulation (Eq. 5) are (see, for instance, Garcia and Morshedi, 1986):

$$Z_{2l}X \geq y_L - y(k) - Z_{1l}V \tag{A1}$$

$$-Z_{2l}X \geq -y_U + y(k) + Z_{1l}V \tag{A2}$$

with

$$V = [\Delta u(k-1) \ldots \Delta u(k-N+1)]^T \tag{A3}$$

where Z_{1l}, Z_{2l} are row vectors, independent of the past $u(k-1)$, ..., $u(k-N)$, and the measurement $y(k)$. These inequalities are repeated for each point in the constraint window ($l = w_b$, ..., w_e).

Assume, without loss of generality, that we can write for the $\hat{A}$, $\hat{b}$, that correspond to J_a:

$$\hat{A}_a^T = \begin{bmatrix} \alpha_{a1}^T \\ \alpha_{a2}^T \end{bmatrix} \text{ and } \hat{b}_a = \begin{bmatrix} \beta_{a1} \\ \beta_{a2} \end{bmatrix} \tag{A4}$$

such that subscript 1 corresponds to the constraints that are identical in J_a and J_b, and 2 refers to opposite active constraints. (In general, one has to simply reorder the rows of $\hat{A}$, $\hat{b}$.) Note that the matrix on the far lefthand side in Eq. 8 is independent of the x_i's or $y(k)$. Thus, one can differentiate:

$$\begin{bmatrix} G & -\alpha_1 & -\alpha_2 \\ -\alpha_1^T & 0 & 0 \\ -\alpha_2^T & 0 & 0 \end{bmatrix} \begin{bmatrix} (\nabla_{x_j} X_a^*)_{J_a} \\ (\nabla_{x_j} \lambda_{a_1}^*)_{J_a} \\ (\nabla_{x_j} \lambda_{a_2}^*)_{J_a} \end{bmatrix} = - \begin{bmatrix} (\nabla_{x_j} g)_{J_a} \\ (\nabla_{x_j} \beta_{a_1})_{J_a} \\ (\nabla_{x_j} \beta_{a_2})_{J_a} \end{bmatrix} \quad 1 \leq j \leq N \tag{A5}$$

One can proceed in the same way for differentiation with respect to $y(k)$. Equation A5 represents a square system of equalities, which has a unique solution.

Now, let us do the same for J_b. In this case, from Eqs. A1 and A2 it follows that:

$$\hat{A}_b^T = \begin{bmatrix} \alpha_{a1}^T \\ -\alpha_{a2}^T \end{bmatrix} \tag{A6}$$

and

$$\begin{aligned} (\nabla_{x_j} \beta_{a_1})_{J_a} &= (\nabla_{x_j} \beta_{b_1})_{J_b} \quad 1 \leq j \leq N \\ (\nabla_{x_j} \beta_{a_2})_{J_a} &= -(\nabla_{x_j} \beta_{b_2})_{J_b} \quad 1 \leq j \leq N \\ (\nabla_y \beta_{a_1})_{J_a} &= (\nabla_y \beta_{b_1})_{J_b} \\ (\nabla_y \beta_{a_2})_{J_a} &= -(\nabla_y \beta_{b_2})_{J_b} \end{aligned} \tag{A7}$$

where

$$\hat{b}_b = \begin{bmatrix} \beta_{b_1} \\ \beta_{b_2} \end{bmatrix} \tag{A8}$$

From Eqs. A5, A6 and A7 one obtains:

$$\begin{aligned} (\nabla_{x_j} \lambda_{a_1}^*)_{J_a} &= (\nabla_{x_j} \lambda_{b_1}^*)_{J_b} \quad 1 \leq j \leq N \\ (\nabla_{x_j} \lambda_{a_2}^*)_{J_a} &= -(\nabla_{x_j} \lambda_{b_2}^*)_{J_b} \quad 1 \leq j \leq N \\ (\nabla_y \lambda_{a_1}^*)_{J_a} &= (\nabla_y \lambda_{b_1}^*)_{J_b} \\ (\nabla_y \lambda_{a_2}^*)_{J_a} &= -(\nabla_y \lambda_{b_2}^*)_{J_b} \\ (\nabla_{x_j} X_a^*)_{J_a} &= (\nabla_{x_j} X_b^*)_{J_b} \quad 1 \leq j \leq N \\ (\nabla_y X_a^*)_{J_a} &= (\nabla_y X_b^*)_{J_b} \end{aligned} \tag{A9}$$

The last two equations of the system (Eq. A9) combined with Eq. 12 yield:

$$\begin{aligned} (\nabla_y f)_{J_a} &= (\nabla_y f)_{J_b} \\ (\nabla_{x_j} f)_{J_a} &= (\nabla_{x_j} f)_{J_b} \quad 1 \leq j \leq N \end{aligned} \tag{A10}$$

Then from Eqs. 13 and A10 we conclude that $c_{J_a}(z) = c_{J_b}(z)$.

Appendix B: Proof of Theorem 3

Assume that the lower or upper bound $\bar{y}$ on the output constraint is predicted active at time $k+N_d+1$. The predicted output is:

$$\hat{y}(k+N_d+1) = \sum_{i=1}^{N} H_i u(k+N_d+1-i) + d(k) = \bar{y} \tag{B1}$$

where

$$d(k) \triangleq y(k) - \sum_{i=1}^{N} H_i u(k-i)$$

is the estimate of the disturbance at the current sample point obtained as in Garcia and Morshedi (1986). Since $H_1 = \ldots = H_{N_d} = 0$, we can solve Eq. B1 for $u(k) \triangleq f$ and then differentiate to get:

$$(\nabla_{x_j} f)_{J_i} = \frac{H_j - H_{j+N_d+1}}{H_{N_d+1}} \text{ and } (\nabla_y f)_{J_i} = \frac{-1}{H_{N_d+1}} \tag{B2}$$

Substitution of Eq. B2 into Eq. 13 and use of Eq. 2 yields the desired result.

Appendix C: Proof of Theorem 5

Assume that the lower or upper bound $\bar{y}$ on the output constraint is predicted to be reached at time $k+N_a$.

$$\hat{y}(k+N_a) = \sum_{j=1}^{N} H_j u(k+N_a-j) + y(k) - \sum_{j=1}^{N} H_j u(k-j) = \bar{y} \tag{C1}$$

$$\Leftrightarrow u(k) = \frac{1}{\sum_{j=1}^{N_a} H_j} \left[\bar{y} - y(k) + \sum_{j=1}^{N} (H_j - H_{N_a+j}) u(k-j) \right] \triangleq f \tag{C2}$$

Then

$$(\nabla_{x_j} f)_{J_i} = \frac{H_j - H_{N_a+j}}{\sum_{l=1}^{N_a} H_l} \text{ and } (\nabla_y f)_{J_i} = \frac{-1}{\sum_{l=1}^{N_a} H_l} \tag{C3}$$

Substitution into Eq. 13 produces the desired expressions.

Appendix D: Proof of Theorem 6

We need that all roots of $[1+\tilde{p}^*(z)c_{J_i}(z)]$ lie inside the unit circle for the c_{J_i} given by Eq. 19. This translates into a requirement that the roots of

$$\left(\sum_{l=1}^{N_a} H_l\right) z^{N-N_a} + H_{N_a+1} z^{N-N_a-1} + \ldots + H_N = 0$$

are inside the unit circle. A sufficient condition for this to happen can be obtained from the dominant coefficient theorem (for example, Jury, 1964, p. 116), which in this case results in the condition:

$$\left|\sum_{l=1}^{N_a} H_l\right| > |H_{N_a+1}| + \ldots + |H_N| \tag{D1}$$

By using Eq. 22 one can easily see that satisfaction of conditions (i) and (ii) of the theorem imply satisfaction of Eq. D1 and therefore nominal stability for the c_{J_i} under consideration.

Manuscript received Jan. 7, 1991, and revision received Aug. 15, 1991.