CONSTRAINT HANDLING IN REDUCED ORDER MPC APPLICATION TO PAPER MACHINES

Yaman Arkun [*] **Apostolos Rigopoulos** [**]

[*] *Koç University, College of Engineering, Sariyer, Istanbul, 34450, Turkey*
[**] *Weyerhaeuser Corp., Tacoma,WA 98477-2999, USA*

Abstract: This paper presents a method to handle input constraints when a large scale system is to be controlled by a model predictive control algorithm which uses a reduced order model of the process under consideration. A paper machine is used throughout to motivate and illustrate the method.

1. INTRODUCTION

In control of large scale systems controller reduction is one method to handle the computational and implementational difficulties arising in a real-time environment. In their earlier work (Arkun and Kayihan, 1998; Rigopoulos, 1999) the authors have addressed the cross-directional (CD) control of paper machines which are equipped with large number of CD actuators. In order to reject disturbances certain transformations are computed to map a given large scale input-output system to a lower dimensional subspace which captures most of the process dynamics. Next reduced order controllers are designed in this lower dimensional space and the resulting controller inputs are transformed back to the original dimension and implemented on the real plant. Arkun and Kayihan (1998) has used reduced order unconstrained IMC as controller, whereas (Rigopoulos, 1999) has used a reduced order constrained model predictive controller. The goal of this paper is to show how the original constraints are preserved during the three steps of model reduction, reduced order MPC design and final implementation.

2. RATIONALE FOR A REDUCED ORDER CONTROLLER DESIGN. A MOTIVATING EXAMPLE: PAPER MACHINE

Paper machines are equipped with large number of CD actuators (slice lips on the headbox) and scanners provide measurements of the property of interest (e.g. thickness, basis weight) at many measurement points across the paper sheet. In addition control inputs can be tightly constrained due to the physical limitations of the actuators. In this work the full order system model used for CD control is given by (see (Rigopoulos, 1999)):

$$\mathbf{y}^N(k) = g(q^{-1})\mathbf{G}_s\mathbf{u}^N(k) + \mathbf{d}(k) \qquad (1)$$

where $\mathbf{y}^N(k) \in \Re^N$ is the measured output (sheet) to be controlled at sampling time k ; $g(q^{-1})$ accounts for CD dynamics; $\mathbf{G}_s$ is the steady state CD actuator gain matrix; $\mathbf{u}^N(k) \in \Re^N$ is the vector of CD control elements; and $\mathbf{d}(k)$ is the disturbance affecting the property of interest. Here dimension N can be very large (several hundreds).

The idea of building a reduced order representation of the full system (1) originated from the reduced order modeling of the disturbances using the method of Karhunen Loeve Expansion (KLE). KLE generates a model with only a few degrees of freedom ($\mathbf{d}^L(k)$) that capture the most significant disturbance patterns (Rigopoulos *et al.*, 1997)

$$\mathbf{d}(k) = \mathbf{\Phi}^L \mathbf{t}^L(k) + \mathbf{\Phi}^{N-L} \mathbf{t}^{N-L}(k) \quad (2)$$

where $\mathbf{\Phi} = \{\phi_1, \ldots, \phi_L, \ldots, \phi_N\}$ consists of the orthonormal basis vectors. They are in fact the eigenvectors of the covariance matrix of the random process $\mathbf{d}(k)$. They can be computed since $\mathbf{d}(k)$ is available through (1) (we assume that control inputs and outputs are measured). The vector $\mathbf{t}$ is computed by projecting $\mathbf{d}$ onto the set of basis functions, i.e.

$$\mathbf{t}^L(k) = (\mathbf{\Phi}^L)^T \mathbf{d}(k) \quad (3)$$

The subspace order L directly identifies the amount of sheet variance that is captured by using only the L most significant modes. Here the interest is in the design of a constrained feedback controller that is capable of rejecting these L significant modes. The reduced order subspace in which controller design takes place has been derived in the following way (Rigopoulos, 1999) Start with the original system (1) and substitute for disturbance its KLE (1):

$$\begin{aligned} \mathbf{y}^N(k) = g(q^{-1})\mathbf{G}_s \mathbf{u}^N(k) \\ + \mathbf{\Phi}^L \mathbf{t}^L(k) + \mathbf{\Phi}^{N-L} \mathbf{t}^{N-L}(k) \end{aligned} \quad (4)$$

Perform an orthogonal projection from $\Re^N$ to $\Re^L$ by multiplying both sides by $(\mathbf{\Phi}^L)^T$:

$$\mathbf{y}^L(k) = g(q^{-1})(\mathbf{\Phi}^L)^T \mathbf{G}_s \mathbf{u}^N(k) + \mathbf{t}^L(k) \quad (5)$$

Defining

$$\mathbf{u}^L(k) = (\mathbf{\Phi}^L)^T \mathbf{G}_s \mathbf{u}^N(k) \quad (6)$$

the reduced order model for controller design becomes

$$\mathbf{y}^L(k) = g(q^{-1})\mathbf{u}^L(k) + \mathbf{t}^L(k) \quad (7)$$

Once the optimal solution $\mathbf{u}^{L^*}(k)$ is computed for (7) it needs to be projected to the full order system (1) so that it can be implemented on the real plant i.e

$$\mathbf{u}^N(k) = \tilde{\mathbf{A}} \mathbf{u}^L(k) \quad (8)$$

For an uncontrained minimum variance type controller Rigopoulos (1999) has shown that the following transformation is optimal:

$$\tilde{\mathbf{A}} = \mathbf{G}_s^+ \mathbf{\Phi}^L \quad (9)$$

where $\mathbf{G}_s^+$ is a generalized inverse of $\mathbf{G}_s$. Final feedback configuration is schematically shown in Figure 1.

In many applications disturbances may not be classified as stationary. In this case KLE and above transformations can still be applied using the most recent disturbance data; thus, they become time-dependent and we use subscript k to denote the time dependence of retained basis functions $\mathbf{\Phi}_k^L$.

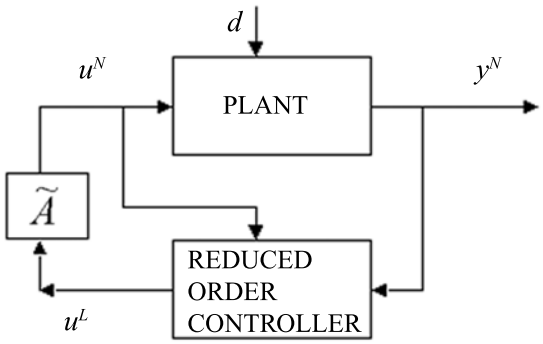

Fig. 1. Feedback configuration.

The transformation of the full order system to a system that has, potentially, much fewer variables (transformed actuators) is done without explicitly considering the original actuator constraints. This is because, there is no way of a-*priori* knowing which of the original constraints would be active at the optimum, at every iteration. Finding the active set would amount to solving the full-order system. When the actuator constraints are mapped onto the reduced order space, the (transformed) feasible region may be empty. The proposed reduced order design is able to recover from this by splitting the problem in two steps:

- **Step 1:** The original N dimesional input/output system is transformed into an M_k-dimensional system with $L \leq M_k$ disturbance modes. The subscript 'k', which denotes sampling time, is included to explicitly show that the input/output dimensions of the transformed system are time-varying. The same transformation matrix that was used in the unconstrained case is also used here to map the actuator values of the reduced order system to the original full order system. Construct the mapping of constraints from the full to the reduced order system. Check for feasibility of that set, possibly by doing a phase-I simplex. If the set is feasible, go to step 2. Otherwise, increase M_k by one, and redo this step. Since the original full order system is considered to be always feasible, there will be an $M_k \leq N$ value for which the reduced order system will also be feasible.
- **Step 2:** Construct all other quantities necessary to form the objective function of the QP for the MPC, and solve the QP, *using the feasible solution of step 1 as the initial value*.

Since this method reduced the dimensions of both input and output spaces, the resulting system is (potentially much) smaller in size than the original one, hence the reduced memory requirements. It is also faster, because although the number of constraints stays the same in the reduced order system, the number of decision variables (transformed actuators) has

(considerably) decreased, thus, it takes less time to compute the active set, and thus the optimum solution.

2.1 Actuator Constraints and their Impact on the Reduced Order Controller Design

There are three types of actuator constraints that are usually encountered in the production of paper and other sheet forming processes:

- *Lower and upper bound constraints*

$$\mathbf{u}_{\min} \leq \mathbf{u}^N(k) \leq \mathbf{u}_{\max} \tag{10}$$

  where usually, because of the problem geometry and that the actuators are expressed in deviation form, $\mathbf{u}_{\max} = -\mathbf{u}_{\min} \geq \mathbf{0}$, and $\mathbf{u}_{\max} = \mathbf{I}_N u_{\max}$.
- *Adjacent actuator constraints*

$$\mathbf{m}_{\min} \leq D\mathbf{u}^N(k) \leq \mathbf{m}_{\max} \tag{11}$$

  where $D \in \Re^{N\times N}$, and for the same reasons as above, $\mathbf{m}_{\max} = -\mathbf{m}_{\min} \geq \mathbf{0}$, with $\mathbf{m}_{\max} = \mathbf{I}_N m_{\max}$. In paper machines this constraint effectively penalizes the bending stress of the slice lip.
- *Rate constraints*

$$|\Delta\mathbf{u}^N(k)| \leq \Delta\mathbf{u}_{\max} \tag{12}$$

  where, $\Delta\mathbf{u}^N(k) = \mathbf{u}^N(k) - \mathbf{u}^N(k-1)$. Again, it is common to have $\Delta\mathbf{u}_{\max} = \mathbf{I}_N \Delta u_{\max}$. This constraint is imposed in order to avoid drastic changes in the magnitude of each actuator *within two consecutive time periods*, which can lead to excessive wear and tear of the actuator hardware.

It is assumed that the above set of inequalities is *always consistent*, i.e. for the operating conditions for which the system was designed, there always exists a feasible solution vector $\mathbf{u}^N(k)$. The situation is different in the case of the reduced order constrained controller design. In particular, the above set of inequalities in the transformed domain becomes

$$\mathbf{u}_{\min} \leq \tilde{\mathbf{A}}_k \mathbf{u}^L(k) \leq \mathbf{u}_{\max} \tag{13}$$

$$\mathbf{m}_{\min} \leq D\tilde{\mathbf{A}}_k \mathbf{u}^L(k) \leq \mathbf{m}_{\max} \tag{14}$$

$$-\Delta\mathbf{u}_{\max} + \tilde{\mathbf{A}}_{k-1}\mathbf{u}^L(k-1) \leq \tilde{\mathbf{A}}_k \mathbf{u}^L(k) \leq \Delta\mathbf{u}_{\max} + \tilde{\mathbf{A}}_{k-1}\mathbf{u}^L(k-1) \tag{15}$$

Thus the *number of constraints has stayed the same*, but the *number of decision variables was reduced from N to L*. This point is very important because, assuming that the objective function is quadratic, the computation time for solving a QP problem will depend heavily on the number of variables, because the latter sets the upper bound on the number of constraints that may be active at the optimum. Finding the active set is one of the most time consuming operations, especially in the presence of tight constraints. This is why the reduced order controller design with a low ratio $\alpha = \frac{L}{N}$ becomes so appealing.

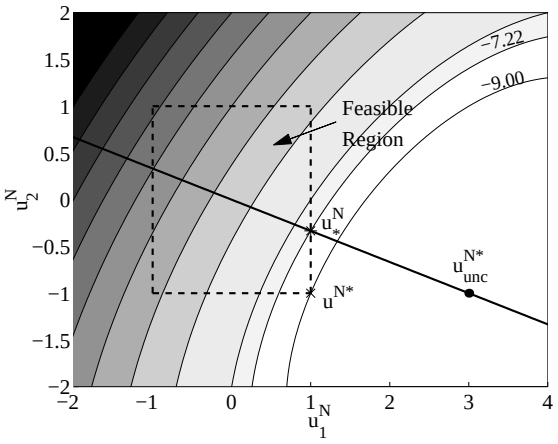

Fig. 2. Contour plot of the objective function arising from a controller with no dynamics ($g(q^{-1}) = 1$), $N = 2$, and $L = 1$.

On the other hand, because transformation matrix $\tilde{\mathbf{A}}$ was constructed based on the unconstrained case, *some of the above inequalities may become infeasible.*

Even in the case where no rate constraints are imposed, *the closed loop performance* obtained from application of the reduced order controller may be unacceptable. This arises again from the fact that the range of $\tilde{\mathbf{A}}_k$ forces $\mathbf{u}^L(k)$ to lie in that region of $\Re^N$ where significant disturbances also lie. But in the presence of constraints this locus may be far from the full order constrained minimum. Fig. 2 shows the problem for a simple example with no dynamics ($g = 1$), $N = 2$, $L = 1$ and a quadratic objective function. The feasible region is the rectangle $S = \{(x,y),\ -1 \leq x \leq 1,\ -1 \leq y \leq 1\}$. The unconstrained minimum denoted by $\mathbf{u}^{N\star}_{unc}$ is located at $[3\ -1]^T$, and it is the same for the full and for the reduced controller designs. In the presence of constraints, however, the optimal full order minimum $\mathbf{u}^{N\star} = [1\ -1]^T$ with an objective function minimum of -9. On the other hand, the locus of $\tilde{\mathbf{A}}\mathbf{u}^L$ is the straight line that passes through points $[0\ \ 0]^T$ and $\mathbf{u}^{N\star}_{unc}$. The constrained optimum arising from the reduced order controller is at $\mathbf{u}^N_\star = [1\ \ -\frac{1}{3}]^T$ and is equal to -7.22, which corresponds to a 20% drop in performance over the full order case that may be deemed unacceptable.

A modification to the reduced order design will be described now that addresses any feasibility issues and allows for improvement of the closed loop performance. The key idea is to *allow the actuator subspace order to be different from the disturbance subspace order.* Thus the reduced order system will have L disturbances, and $M_k \geq L$ inputs and outputs. Following an approach very similar to the one used for the derivation of Eq. (7) one gets:

$$\mathbf{y}^{M_k}(k) = g(q^{-1})\mathbf{u}^{M_k}(k) + \mathbf{R}_k^T \mathbf{t}^L(k) \tag{16}$$

where $\mathbf{R}_k = [\mathbf{I}_L\ \ \mathbf{0}_{L\times(M_k-L)}]$.

With regard to the above system the following observations are in order

- Obtaining $\mathbf{u}_\star^N(k)$ from $\mathbf{u}^{M_k\star}(k)$ presents the same problems as before. Again, a linear relation is postulated

$$\mathbf{u}^N(k) = \tilde{\mathbf{A}}_k \mathbf{u}^{M_k}(k) \tag{17}$$

where $\tilde{\mathbf{A}}_k \in \Re^{N\times M_k}$. Eq. 16 can be written more explicitly as

$$\begin{bmatrix} \mathbf{y}^L(k) \\ \mathbf{y}^{M_k-L}(k) \end{bmatrix} = g(q^{-1}) \begin{bmatrix} \mathbf{u}^L(k) \\ \mathbf{u}^{M_k-L}(k) \end{bmatrix} + \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0}_{(M_k-L)\times L} \end{bmatrix} \mathbf{t}^L(k) \tag{18}$$

where the top block is completely separate from the bottom and identical to the L-dimensional system. Thus, one can apply the results of that design directly to get $\tilde{\mathbf{A}}_k^L = \mathbf{G}_s^+ \mathbf{\Phi}_k^L$, where $\tilde{\mathbf{A}}_k = [\tilde{\mathbf{A}}_k^L \;\; \tilde{\mathbf{A}}_k^{M_k-L}]$.
- $\tilde{\mathbf{A}}_k^{M_k-L}$ is obtained by considering the constrained minimization problem. In particular, the inclusion of additional inputs $\mathbf{u}^{M_k-L}(k)$ is done so as to increase the number of elements of set $S_u \subset \Re^N$ where $\mathbf{u}_\star^N(k)$ belongs. This increase is maximized by appropriately selecting $\tilde{\mathbf{A}}_k^{M_k-L}$ to maximize the number of elements that belong to the range of $\tilde{\mathbf{A}}_k$. This is accomplished by designing $\tilde{\mathbf{A}}_k^{M_k-L}$ to be orthogonal to $\tilde{\mathbf{A}}_k^L = \mathbf{G}_s^+ \mathbf{\Phi}_k^L$, which is guaranteed by setting $\tilde{\mathbf{A}}_k^{M_k-L} = \mathbf{G}_s^+ \mathbf{\Phi}_k^{M_k-L}$, since $\mathbf{\Phi}_k^{M_k-L}$ is orthogonal to $\mathbf{\Phi}_k^L$.

  Finally,

$$\tilde{\mathbf{A}}_k = \mathbf{G}_s^+ \mathbf{\Phi}_k^{M_k} \tag{19}$$

2.2 Implementation of Constrained Control through State-Space Model Predictive Control

In the present work state space MPC as detailed in (Ricker, 1992) is used. Here we will present only the important features which are unique to our problem setting. The control algorithm is based on the reduced order system given by Eq. 16:

$$\begin{aligned} \mathbf{y}^{M_k}(k) &= g(q^{-1})\mathbf{u}^{M_k}(k) + \mathbf{R}_k^T \mathbf{t}^L(k) \\ &= \mathbf{y}_u^{M_k}(k) + \mathbf{y}_d^{M_k}(k) \end{aligned} \tag{20}$$

State-space transformation of $\mathbf{y}_d^{M_k}(k)$: Assuming that the disturbance subspace L has been selected, $\mathbf{y}_d^{M_k}(k)$ is converted to state-space (Rigopoulos, 1999)

$$\mathbf{x}_d(k+1) = \mathbf{A}_d(k)\mathbf{x}_d(k) + \mathbf{\Gamma}_d(k+1)\mathbf{e}(k+1) \tag{21}$$

$$\mathbf{y}_d^{M_k}(k) = \mathbf{R}_k^T \tilde{\mathbf{C}}_d(k)\mathbf{x}_d(k) \tag{22}$$

where $\mathbf{A}_d(k)$, and the size of $\mathbf{\Gamma}_d(k)$ depend on the AutoRegressive (AR) modeling of $\mathbf{t}^L(k)$; and $\tilde{\mathbf{C}}_d(k) = (\mathbf{\Phi}_k^L)^T \mathbf{C}_d(k)$. The order of the AR model selected remains constant throughout the simulation. Thus, although the size of $\mathbf{y}_d^{M_k}(k)$ is determined by the size of $\mathbf{R}_k$ which depends on M_k, the sizes of the vectors and matrices of state equation (21) remain unaffected. As a consequence, varying M_k presents no problem to the state-space modeling of $\mathbf{y}_d^{M_k}(k)$.

State-space transformation of $\mathbf{y}_u^{M_k}(k)$: Because $\mathbf{y}_u^{M_k}(k) = g(q^{-1})\mathbf{u}^{M_k}(k)$ is a *decoupled* system, its state-space equivalent description will be in terms of *block-diagonal* matrices:

$$\mathbf{x}_u^{(nM_{k+1})}(k+1) = \mathbf{A}_u^{(\bar{n}M_k)}\mathbf{x}_u^{(nM_k)}(k) + \mathbf{B}_u^{M_k}\mathbf{u}^{M_k}(k) \tag{23}$$

$$\mathbf{y}_u^{M_k}(k) = \mathbf{C}_u^{M_k}\mathbf{x}_u^{(nM_k)}(k) \tag{24}$$

where, for example $\mathbf{A}_u^{(\bar{n}M_k)} = \mathrm{diag}[\mathbf{A}^{\bar{n}} \cdots \mathbf{A}^{\bar{n}}]$; each $\mathbf{A}^{\bar{n}} \in \Re^{n\times n}$, with $\mathrm{rank}[\mathbf{A}^{\bar{n}}] = \bar{n}$ contains the necessary states for the modeling of each $y_{u,i}^{M_k}(k) = g(q^{-1})u_i^{M_k}(k)$.

Now, suppose that at iteration $k+1$ the feasibility of constraints imposes $M_{k+1} = M_k + 1$ leading to a unit increase of inputs $\mathbf{u}^{M_{k+1}} = \begin{bmatrix} \mathbf{u}^{M_k} \\ u^{new} \end{bmatrix}$ at time $k+1$. However, because of the block diagonal nature of all the state-space matrices involved, the states appropriated to the modeling of the first M_k elements of vector $\mathbf{u}^{M_{k+1}}(k+1)$ will not be influenced by the new states that must be introduced for the additional input. This observation is important because it shows that when the state order increases from nM_k to $nM_k + n$ due to the introduction of additional input u^{new}, only the new states nM_k+1 to nM_k need to be initialized (to zero), before the state equation can be used to compute $\mathbf{x}_u^{(nM_{k+2})}(k+2)$. The first nM_k states evolve normally.

2.3 Computational Issues and Efficiency

Consider a paper machine with $N = 100$ CD actuators where the full order constrained controller is to be applied with p (prediction horizon) $= m_h$ (move horizon) $= 4$ and all three types of constraints are present. Then, an optimization problem with 400 variables and 2,400 constraints would have to be solved *at every sampling time*. Even when a reduced order controller were to be implemented with $M_k = 30$, the task would not be trivial. Therefore, special attention needs to be put on the selection of the most suitable optimization algorithm.

In selecting the most appropriate solver one must, consider the problem that arises from a potential infeasibility due to the dimensionality reduction. The solver should be able to identify this problem as quickly as possible and compensate for it by gradually increasing the actuator subspace order M_k in order to obtain a feasible region. In this "internal" loop one only needs to update matrix $\tilde{A}$ and increase the size of $\Delta U(k)$ before re-checking for feasibility. Updating all other quantities including the objective function should be done only once, *after* feasibility has been ensured.

Of course, should the user decide to reduce M_k in an attempt to speed up the computations, the same type of feasibility test must be made prior to accepting a lower M_k value. Evidently, only a primal active set method works along those lines, because of its inherent need for the *a-priori* calculation of a feasible point. According to this strategy, a phase I simplex is performed repetitively increasing M_k by one until an initial feasible point (thus a feasible region) is obtained. Then, after all appropriate matrices are updated, a search for the optimum is initiated. This procedure may become even more efficient by using an interior point (IP) method to obtain the solution to the phase I simplex problem, instead of using a standard Dantzig-type LP solver.

Another point that can have significant impact on the overall performance of the QP solver is "hot" starts. In the presence of relatively tight input constraints, the optimal solution vector does not vary significantly from one sampling time to the next. Thus, in addition to ensuring that an initial feasible point is available, one can further benefit from providing an initial point that is relatively close to the optimum.

3. EXAMPLES

A paper machine with $N = 200$ CD actuators and measurements positions is considered, with $g(q^{-1}) = \frac{q^{-1}}{1-0.2q^{-1}}$, and

$$\mathbf{G}_s = \text{Toeplitz}[2.0\ 0.8\ -1.0\ -0.8\ -0.6\ -0.4\ 0\ \cdots\ 0]$$

The full disturbance $\mathbf{d}(k)$ consists of 200 CD and 300 MD positions. The last 100 full MD s are illustrated in Fig. 3. For the modeling equations used to create this the reader is referred to (Rigopoulos *et al.*, 1997). Using the first 200 full data, KLE indicated that only $L = 3$ modes were necessary to capture the significant disturbance patterns. Also, an AR(2) model was sufficient for the modeling of temporal vectors $\mathbf{t}^L(k)$.

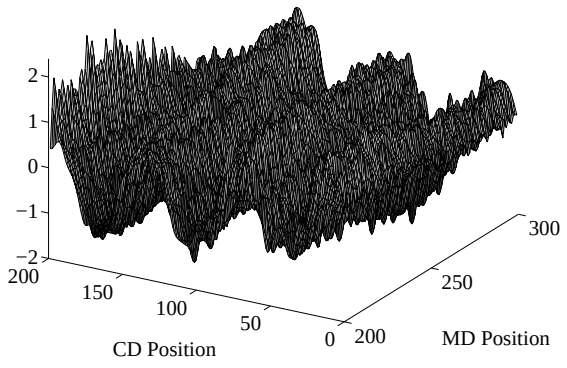

Fig. 3. Last section of the disturbance .

Table 1. Description of the sets of constraints considered in the example.

Constraints Set	$u_{\max} = -u_{\min}$	$m_{\max} = -m_{\min}$	$\Delta u_{\max}$
set 1 — tight	0.15	0.1	0.005
set 2 — moderate	0.15	0.1	∞
set 3 — light	0.25	∞	∞

The parameters of the AR model and basis functions $\mathbf{\Phi}_k^L$ were also updated at every sampling time. Table 1 gives the values of the various types of constraints used.

For constraints sets 2 and 3 M_k remained constant throughout the simulation, as no feasibility issues were encountered. Feasibility problems were encountered, however, when rate constraints were imposed (set 1), and M_k had to be increased in order to be able to compute a solution vector $\mathbf{u}^{M_k}(k)$. Also, for each simulation the time allocated to solving the QP was recorded, as well as the closed loop .

Fig. 4(a) illustrates the standard deviation of the CD profile as a function of the actuator subspace order $M_k = 10$, 30 and 150 for constraint set 3. The open loop standard deviation is also included for comparison. Clearly, the controller was able to reduce the CD variability significantly even when only 10 transformed manipulated variables $\mathbf{u}^{M_k}(k)$ were used. What is more important, however, is the very little improvement achieved by the five-fold increase to the subspace order (from 30 to 150) that shows the efficiency of the reduced order controller design.

Fig. 5(a) and (b) show the number of active constraints at the optimum when the simulation was started with initial subspace order $M_1 = 10$ and $M_1 = 30$, respectively. An increase in the number of active constraints indicates an increase in M_k because an infeasible constraint region was encountered. Notice that in almost all sampling times the number of active constraints is equal to the number of degrees of freedom, illustrating how tight the constraints are.

Finally the trade-off between closed loop performance and computation time is illustrated in Fig. 6 by comparing the drop in the sum-of-square (SSE) errors *vs.* the increase in computation time. For example, in the case of tight constraints, increasing M_k from 50 to 100 would result to a less than 1% performance improvement, while requiring 340% more computation time! This clearly shows the significant computational advantages associated with the implementation of the reduced order controller design.

4. CONCLUSIONS

We have presented a method to handle input constraints when an MPC algorithm is derived based on a reduced order approximation of the process but implemented on the real process. A paper machine example is used to demonstrate the method for CD control.

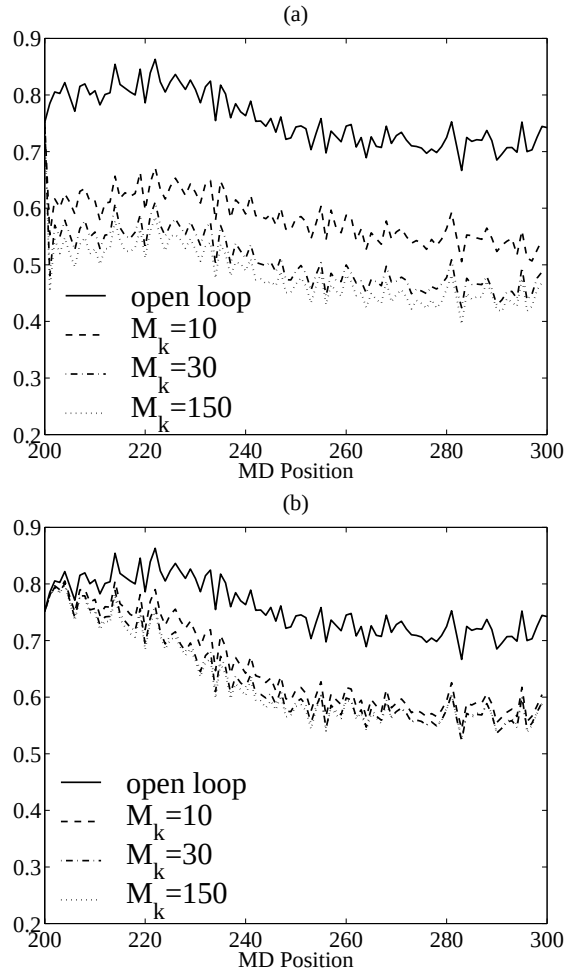

Fig. 4. Standard deviation as a function of initial subspace order for (a) constraints set 3 and (b) constraints set 1.

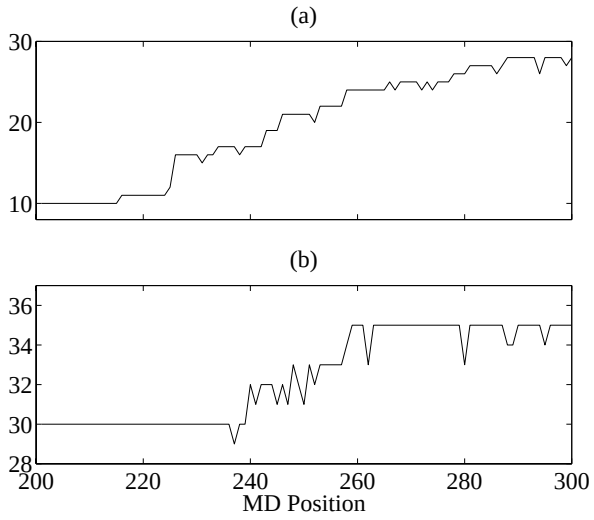

Fig. 5. Number of active constraints for simulations with initial actuator subspace order (a) $M_1 = 10$ and (b) $M_1 = 30$, using constraints set 1.

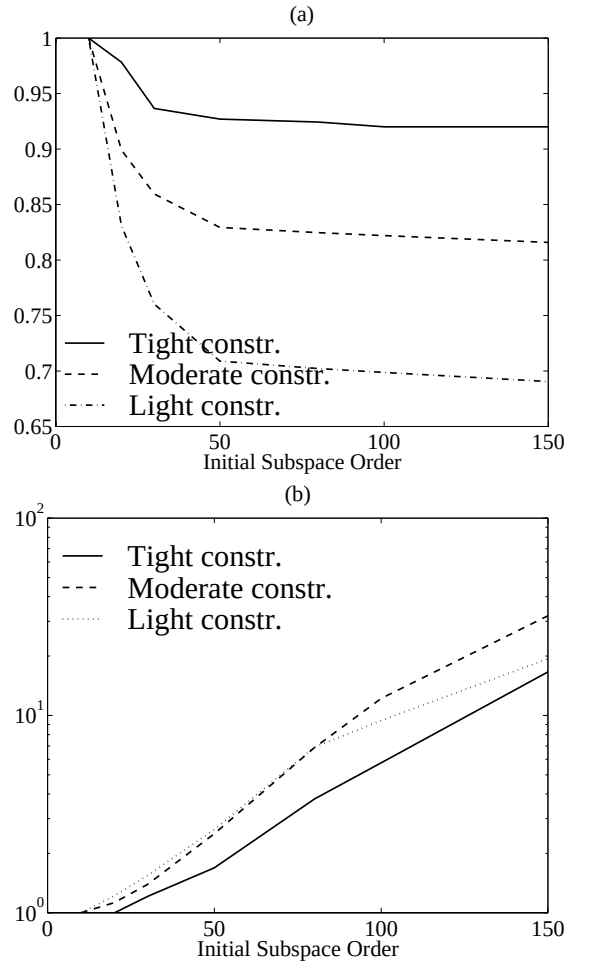

Fig. 6. Trade-off between performance as expressed by the sum-of-square improvement with increasing initial M_k shown in (a), and computation time increase as a function of initial M_k shown in (b).

REFERENCES

Arkun, Yaman and Ferhan Kayihan (1998). A novel approach to full CD profile control of sheet-forming processes using adaptive PCA and reduced-order IMC design. *Computers Chem. Engng* **22**(7–8), 945–962.

Ricker, N. Lawrence (1992). Model-predictive control: State of the art. In: *Proceedings of the CPC-IV Conference*. Padre Island, Texas. pp. 271–296.

Rigopoulos, A. (1999). Application of principal component analysis in the identification and control of sheet-forming processes. *PhD Thesis, Georgia Institute of Technology, Atlanta, GA.*

Rigopoulos, Apostolos, Yaman Arkun and Ferhan Kayihan (1997). Identification of full profile disturbance models for sheet forming processes. *AIChE J.* **43**(3), 727–739.

SIMULATION-BASED DUAL MODE CONTROLLER FOR NONLINEAR PROCESSES

Jong Min Lee * **Jay H. Lee** *,1

* *311 Ferst Dr., School of Chemical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0100, USA*

Abstract: This paper presents a simulation-based strategy for designing a nonlinear override control scheme to improve the performance of a local linear controller. The higher-level nonlinear controller monitors the dynamic state of the system under the local controller and sends an override control action whenever the system is predicted to move outside an acceptable operating regime under the local controller. For this purpose, a cost-to-go function is defined, an approximation of which is constructed by using simulation or historic operation data. The cost-to-go function delineates the "admissible" region of state space within which the local controller is effective, thereby yielding a switching rule. The same cost-to-go function can also be used to calculate override control actions designed to bring the system state back into the admissible region as quickly as possible. One potential problem of this approach is the lack of robustness when the simulation data sparsely cover the state space and the data-based approximation of the cost-to-go function is extrapolated to a region previously unseen. Hence, successful application of the proposed method requires safeguarding against undue extrapolations. For this reason, a kernel-based local approximation, instead of a global approximator like a neural network, is used to interpolate the cost-to-go values. It is shown that the kernel-based local regression provides convenient means to implement a risk-sensitive control scheme which avoids excessive extrapolation. The proposed scheme is demonstrated and discussed with nonlinear examples.

Keywords: Simulation-Based Approach, Nonlinear Predictive Control, Switching Controller, Kernel-Based Approximator, Cost-to-Go Function

1. INTRODUCTION

Model predictive control (MPC) is being widely used in the process industry because of its ability to control multivariable processes with hard constraints. Most of the current commercial MPC solutions are based on linear dynamic models, which are easier in terms of identification and online computation (Qin and Badgewell, 1997). On the other hand, many chemical processes exhibit strong nonlinearities. This disparity has prompted several studies on MPC formulations with nonlinear system models (Lee, 1997). Since most Nonlinear MPC (NMPC) formulations require online solution of a nonlinear program (NLP), issues related to computational efficiency and stability of a control algorithm have received much attention.

The initial focus was on formulating a computationally tractable NMPC method with guaranteed stability. Mayne and Michalska (1990) showed that stability can be guaranteed by introducing a terminal state equality constraint at the end of

[1] to whom correspondence should be addressed: jay.lee@che.gatech.edu

prediction horizon. In this case, the value function for the NMPC can be shown to be a Lyapunov function under some mild assumptions. Because the equality constraint is difficult to handle numerically, Michalska and Mayne (1993) extended their work to suggest a dual-mode MPC scheme with a local linear state feedback controller inside an elliptical invariant region. This effectively relaxed the terminal equality constraint to an inequality constraint for the NMPC calculation. The dual-mode control scheme was designed to switch between the NMPC and the linear feedback controller depending on the location of the state. Chen and Allgöwer (1998) proposed a quasi-infinite horizon NMPC, which solves a finite horizon problem with a terminal cost and a terminal state inequality constraint. The main difference from the Michalska and Mayne's method is that a fictitious local linear state feedback controller is used only to determine the terminal penalty matrix and the terminal region off-line and switching between controllers is not required.

These NMPC schemes have theoretical rigor but have some practical drawbacks. First, these methods still require solving a multi-stage nonlinear program at each sample time. Assurance of a globally optimal solution or even a feasible solution is difficult to guarantee. Second, the optimization problem for determining the invariant region for a local linear controller and the corresponding terminal weight are both conservative and computationally demanding.

Motivated by the drawbacks and the industry's reluctance to adopt full-blown NMPC, we propose an override (or supervisory) control strategy for monitoring and improving the performance of a local controller. Our method is similar to the dual-mode MPC suggested by Michalska and Mayne in that both switch between two different control policies depending on current location of the state. However, we employ a cost-to-go function based approach instead of NMPC. First a *cost-to-go* function under the local controller is defined, which serves to delineate the admissible region within which the local controller can effectively keep the system inside acceptable operating limits. The same cost-to-go function is also shown to facilitate the calculation of override control actions that will bring the system outside the admissible region back into the region as quickly as possible. We propose to use simulation or historic data to construct an approximation to the cost-to-go function. With the cost-to-go function, an override control action can be calculated by solving a single stage nonlinear optimization problem, which is considerably simpler than the multi-stage nonlinear program solved in the NMPC.

One potential problem of using the cost-to-go values approximated using simulation data is that it is only accurate within regions where data existed. Hence, in the on-line calculation, one has to safeguard against unreasonable extrapolation of the cost-to-go function approximator. This leads to a *risk-sensitive* control scheme, where the quality of approximation gets reflected in the cost-to-go value. In this paper, we propose to use a local regression based on Gaussian kernel in order to implement the risk-sensitive control, which avoids unreasonable extrapolations.

2. SIMULATION-BASED CONSTRUCTION OF AN OVERRIDE CONTROLLER

The proposed scheme uses either simulation or actual plant data to identify the region of the state space, in which the local controller can effectively keep the system inside an acceptable operating regime (defined by some inequalities in the state space). We do this by assigning to each state a 'cost-to-go' value, which is defined as

$$J^{\mu}(x_0) = \sum_{i=0}^{\infty} \alpha^i \phi(x_i) \tag{1}$$

where $J^{\mu}(x_0)$ is the cost-to-go for state x_0 under the local control policy $u = \mu(x)$, $0 < \alpha < 1$ is a discount factor, and $\phi(x_i)$ is a stage-wise cost that takes the value of 0 if the state at time i is inside the acceptable operating limit and 1 if outside when x_0 is the state at time 0. This way, if a particular state x_0 under the control policy evolves into a state outside the limit *in some near future* under the policy μ, the cost-to-go value will reflect it. On the other hand, those states that are not a precursor of future violation of the operating limit will have a negligible cost-to-go value. The latter states comprise the "admissible" region.

The cost-to-go function is approximated by first simulating the closed-loop behavior of the nonlinear model under the local linear controller for various possible operating conditions and disturbances. This generates x vs. $J^{\mu}(x)$ data for all the visited states during the simulation. Then the generated data can be interpolated to give an estimate of $J^{\mu}(x)$, $\tilde{J}^{\mu}(x)$, for any given x in the state space.

In the real-time application, whenever the process reaches a state with a significant cost-to-go value, it is considered to be a warning sign that the local controller's action will not be adequate. When this happens, an override control action is calculated and implemented to bring the process back to the "admissible" region where the cost-to-go is insignificant. One can calculate such an action by implementing the override policy of

$$\text{if} \quad \tilde{J}^{\mu}(x_{t+1}(x_t, \mu(x_t))) \geq \eta \quad ,$$
$$u_t = \arg\left(\min_{u'_t} \tilde{J}^{\mu}(x_{t+1}(x_t, u'_t))\right) \qquad (2)$$

where η is a user-given threshold value for triggering the override control scheme. If no u'_t can be found such that $\tilde{J}^{\mu}(x_{t+1}(x_t, u'_t)) < \tilde{J}^{\mu}(x_{t+1}(x_t, \mu(x_t)))$, then $u_t = \mu(x_t)$ is used for the current sample time.

3. A KERNEL-BASED APPROXIMATOR OF COST-TO-GO FUNCTION

In this paper, we propose to use a local regression instead of the usual choice of a feedforward neural network to approximate the cost-to-go values. Empirical studies show that general approximators (e.g. neural network) are not good choices for the approximation of cost-to-go function due to the high nonlinearity and discontinuity of the cost-to-go function in general (Boyan and Moore, 1995). In addition, Gordon (1995) showed that the local averager with non-expansive property (*e.g.* kernel-based approximation) is compatible with dynamic programming operator and effective for representing local characteristics of state spaces.

Another reason for adopting the local regression approach is our concern for grossly incorrect cost-to-go estimates that can arise from extrapolating to a region not accounted for in the simulation step. In implementing a risk-averse 'cost-to-go' based controller, Kaisare *et al.* (2002) used a feedforward neural network but gridded the state space in order to separate regions visited by simulation from those not. For those cells with little or no data, a high cost-to-go value was assigned to prevent the controller from driving the state trajectory into these uncertain regions. However, this is difficult to implement for cases with high-dimension state spaces.

For a convenient implementation of the risk-averse or rist-sensitive scheme, we propose to use a variation of Gaussian-kernel-based approximators. This structure decides whether a reliable estimate can be given to a query point based on the available data. For a "reliable" query point it gives local weights calculated from a Gaussian kernel to give more influence over the regression to those training points closer to the query point than those farther away. The suggested structure of kernel-base prediction is

$$\hat{f}(x_0) = \frac{\sum_{i=1}^{N} K_{\lambda}(x_0, x_i) y_i}{\sum_{i=1}^{N} K_{\lambda}(x_0, x_i)} \qquad (3)$$

where

$$K_{\lambda}(x_0, x_i) = \exp\left(-\frac{\|x_0 - x_i\|_2^2}{\lambda^2}\right) \qquad (4)$$

The number of neighbor points N is the number of data points inside a hypersphere, the radius of which is a user-given value r. In addition to r, there are other parameters that user should provide. These are the Gaussian kernel width λ, minimum number of data points inside the hypersphere k_{min}, and the high cost-to-go value J_h to be assigned to an "unreliable" query point. Table 1 describes how the estimate of cost-to-go value for a query point is calculated.

Table 1. "Risk-averse" prediction using Gaussian-kernel-based approximator

Prediction Algorithm
1. Is the query point x_0 in the memory?
a. Yes: Use the value in the memory.
b. No: Go to step 2.
2. Enumerate the data points inside r around the x_0.
Is the number of data points greater than k_{min}?
a. Yes: Average with the kernel.
a. No: $J(x_0)$ cannot be estimated. Assign J_h to x_0.

4. ILLUSTRATIVE EXAMPLES

4.1 Simple Nonlinear Example

4.1.1. Problem Description We consider a system with two states, one output, and one manipulated input described by

$$\begin{aligned} x_1(k+1) &= x_1^2(k) - x_2(k) + u(k) \\ x_2(k+1) &= 0.8\exp\{x_1(k)\} - x_2(k)u(k) \\ y(k) &= x_1(k) \end{aligned} \qquad (5)$$

with an equilibrium point of $x_{eq} = (-0.3898, 0.5418)$, $u_{eq} = 0$.

We also define the *acceptable* operating regime by

$$W(x) = \left\{(x_1 - x_{1eq}) + \sqrt{3}(x_2 - x_{2eq})\right\}^2 + \left\{\frac{(x_2 - x_{2eq}) - \sqrt{3}(x_1 - x_{1eq})}{0.3}\right\}^2 - 4 \leq 0 \qquad (6)$$

A linear MPC controller was designed based on a linearized model around the equilibrium point. The control objective is to regulate y to y_{eq}. The linear MPC is used as the local controller with the following design.

$$\min_{\Delta u} \sum_{i=1}^{p} 5\bar{y}^2(k+i) + \sum_{l=0}^{m-1} \Delta\bar{u}^2(k+l) \qquad (7)$$

with $p = 2$ and $m = 1$.

$$\begin{aligned} -3 \leq \bar{u} \leq 3 \\ \Delta\bar{u} \leq 0.2 \end{aligned} \qquad (8)$$

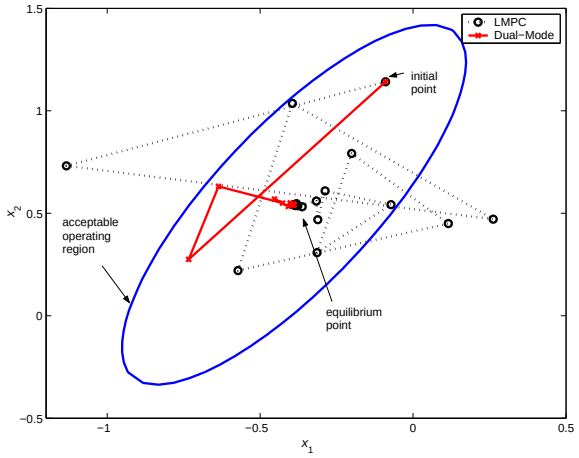

Fig. 1. State trajectories under local MPC and dual-mode controller, $x_0 = [-0.0898\ 1.1418]$

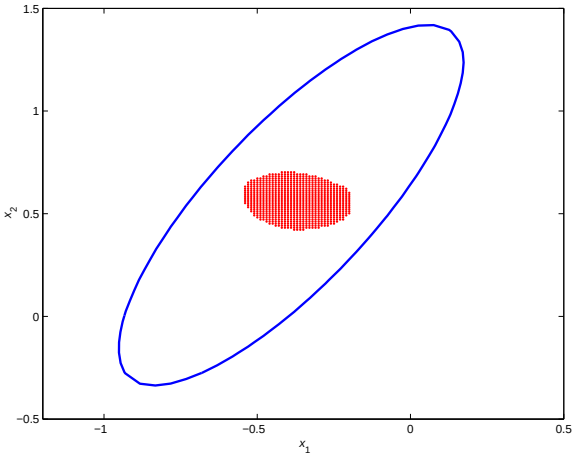

Fig. 2. Regions under local controller with $\tilde{J}(x) < 0.02$

The closed-loop behavior under the local controller starting at $x_0 = x_{eq} + [0.3\ \ 0.6] = [-0.0898\ 1.1418]$ is shown as dotted lines in Fig. 1. Though the initial point is inside the operating limit, the system under the local linear controller violates the limit several times until the system is regulated to the equilibrium point.

4.1.2. Simulation-Based Design To design the proposed override controller, closed-loop simulations under the local controller were performed using 347 initial points inside the operating limit. The simulations generated 17006 data points and cost-to-go values for each state in the trajectory were calculated using Equation (1) with a value of $\alpha = 1$ and

$$\phi(x_t) = \begin{cases} 1 \text{ if } W(x_{1t}, x_{2t}) \leq 0 \\ 0 \text{ if } W(x_{1t}, x_{2t}) > 0 \end{cases} \qquad (9)$$

Next step is to design a Gaussian-kernel approximator. Considering the coverage of state space, following parameters were chosen: $r = 0.05$, $k_{min} = 3$, $\lambda = 0.03$, $J_h = 30$.

The actual value of cost-to-go is zero for the states inside the admissible region of a linear controller and outside the region the cost-to-go will be over unity. This makes the structure of cost-to-go function very stiff. However, the approximator will smoothen out the stiff structure a bit by averaging. Therefore small tolerance value ($\eta = 0.02$) was chosen to illustrate a possible shape of the admissible region under the local controller, which is illustrated in Fig. 2.

4.1.3. Real-Time Application To compare online performances of the local controller alone and the dual mode controller (i.e., the local controller combined with the proposed override controller), eight initial points different from the training set were sampled. We also compare the proposed dual-mode controller with the successive linearization based MPC (SLMPC) scheme suggested by Lee and Ricker (1994). Finally, we also simulated the LMPC and the SLMPC with the state constraints of $-0.95 \leq x_1 \leq 0.2$ and $-0.35 \leq x_2 \leq 0.45$ (denoted by scLMPC and scSLMPC). The prediction and control horizons of SLMPC are the same as those of the LMPC.

The solid lines in Fig. 1 is the state trajectory with the same initial point under the dual-mode controller. For the first three points, the override control actions were used instead of those of LMPC's. The proposed scheme successfully steers the state back to the region with lower cost-to-go values. Table. 2 shows the sum of stage-wise cost (the total number of violation of operating limit) and the suggested control design outperforms for all the test points. We can also see that imposing state constraints did not work here as many infeasible solutions were returned, eventually causing divergence.

Table 2. Comparison of performances (total # of limit violations)

Test pt	LMPC	SLMPC	scLMPC	scSLMPC	Override
1	div.	5	div.	div.	0
2	3	3	div.	div.	0
3	2	0	0	div.	0
4	2	0	0	div.	0
5	0	0	div.	div.	0
6	0	0	0	div.	0
7	7	15	1	div.	0
8	div.	div.	div.	div.	0

4.2 Bioreactor Example

In this section, we consider a bioreactor example with two states: biomass and substrate (Bequette, 1998). With a substrate inhibition for growth rate expression of biomass, the system shows multiple steady states. To operate at the unstable

equilibrium, closed-loop control must be used. The system equation is:

$$\begin{aligned}\frac{dx_1}{dt} &= (\mu - D)x_1 \\ \frac{dx_2}{dt} &= D(x_{2f} - x_2) - \frac{\mu x_1}{Y} \\ \mu &= \frac{\mu_{max} x_2}{k_m + x_2 + k_1 x_2^2}\end{aligned} \tag{10}$$

where x_1 is biomass concentration and x_2 is substrate concentration. Table 3 shows the parameters for the model at the unstable steady state.

Table 3. Model parameters: bioreactor example

μ_{max}	0.53 hr^{-1}	k_m	0.12 g/l
k_1	0.4545 l/g	Y(yield)	0.4
D_s, x_{2fs}	0.3hr^{-1}, 4.0 g/l	x_s	[0.9951 1.5123]

4.2.1. Local Linear Controller A linear MPC was designed based on a linearized model around the unstable equilibrium point with sample time of 0.1h. The control objective is to regulate x to x_s at the equilibrium values and the manipulated variables are the substrate concentration in the feed x_{2f} and the dilution rate D. The LMPC controller parameters we used are $Q = 100I, R = 10I, p = 10$, and $m = 5$, where I is a 2 by 2 identity matrix, Q is a state weighting matrix, and R is an input weighting matrix.

We also define an acceptable operating region as

$$\begin{aligned}W(x) = &\left\{\frac{0.52(x_1 - x_{1eq}) + 0.85(x_2 - x_{2eq})}{7}\right\}^2 \\ + &\left\{\frac{-0.85(x_1 - x_{1eq}) + 0.52(x_2 - x_{2eq})}{0.5}\right\}^2 \\ -1 \le 0 &\end{aligned} \tag{11}$$

which is shown in Fig.3. The input constraints for MPC is

$$\begin{aligned}0 \le D \le 0.5 \quad |\Delta D| \le 0.2 \\ 0 \le x_{2f} \le 8 \quad |\Delta x_{2f}| \le 2\end{aligned} \tag{12}$$

The closed-loop behavior under the LMPC for different initial points are shown in Fig. 3. As in the previous example, the LMPC cannot drive the state back into the equilibrium point without violating the operating limit.

4.2.2. Simulation-Based Dual Mode Controller With the same definition of one-stage cost as in Equation (9), a cost-to-go-based override controller was designed. For the simulation, 109 initial points were sampled inside the operating limit and closed-loop simulations under the LMPC

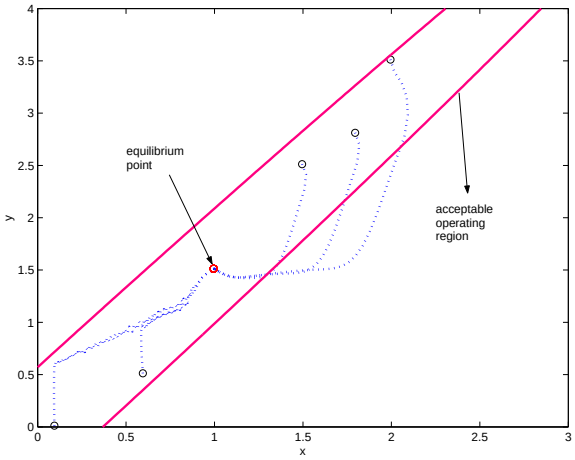

Fig. 3. State trajectories under local MPC

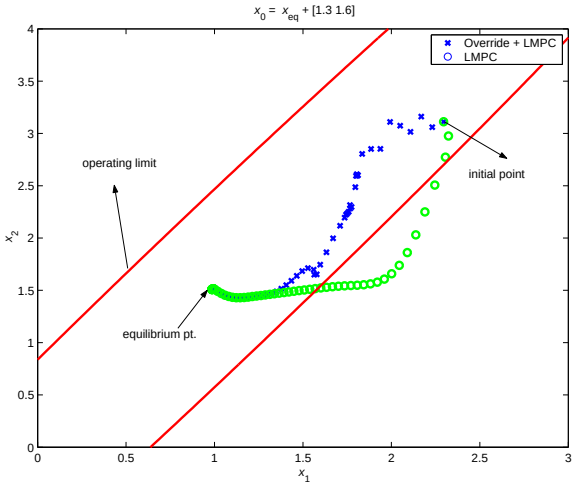

Fig. 4. State trajectory under dual-mode controller

yielded 21909 points. Parameters for a kernel-based approximator were chosen as: $r = 0.1$, $k_{min} = 5$, $\lambda = 0.05$, $J_h = 50$, $\eta = 0.02$.

As in the previous example, the dual mode controller successfully navigated the state to the equilibrium point without violating the operating limit by searching for the path with lowest cost-to-go values. One of the sample trajectories tested is shown in Fig. 4.

5. EVOLUTIONARY IMPROVEMENT OF COST-TO-GO

Because the approximator employed in the calculation of override control action is based on the cost-to-go value of the local linear controller, it is not the optimal cost-to-go. The resulting override controller from the suboptimal cost-to-go approximation is also suboptimal. Hence, further improvement of the override control policy to steer the system back into the admissible region of the linear controller is possible by iteratively solving

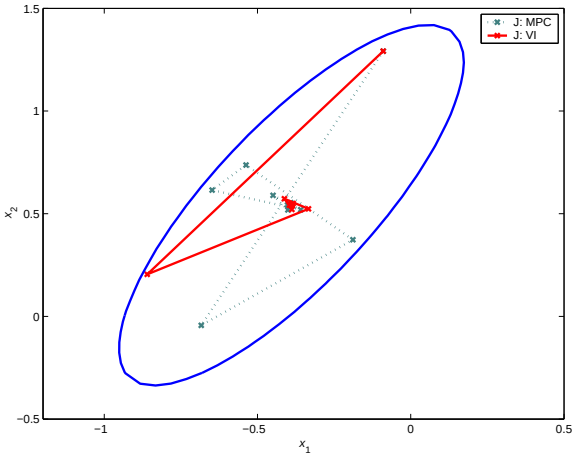

Fig. 5. State trajectory with the dual-mode controller using improved cost-to-go

the following optimality equation (as in *value-iteration*) until $\tilde{J}$ converges.

$$\tilde{J}^{i+1}(x) = \min_u \left[\phi^i(x) + \tilde{J}^{i+1}(f(x,u)) \right] \quad (13)$$

where f is a state transition equation and i denotes iteration index.

For this purpose, the one-stage cost is re-defined differently as

$$\phi^i(x) = \begin{cases} 1 & \tilde{J}^i(x) \geq \eta \\ 0 & \tilde{J}^i(x) < \eta \end{cases} \quad (14)$$

With this change, the aim of the optimal control is to bring the system state back into the "admissible" region as quickly as possible.

The value iteration was performed for the first illustrative example and the iteration converged after 5 steps with the following convergence criterion.

$$\|\tilde{J}^{i+1}(x) - \tilde{J}^i(x)\|_\infty < 0.1 \quad (15)$$

Fig. 5 shows one of the state trajectory with the initial point of $x_0 = x_{eq} + [\,0.3 \quad 0.75\,]$ when the improved cost-to-go function is used in the override control calculation. As shown in the figure, the improved override controller bring the state back into the admissible region more efficiently than that based on the cost-to-go approximation under the LMPC.

6. CONCLUSION

A simulation-based override control scheme was shown to improve the performance and stability of a given local controller. The ease of design and implementation makes it a potentially appealing addition to an existing controller in industrial applications. The suggested framework can give operators indications on the future performance of the local controller and also suggest override control actions, if needed. More realistic situations such as the case with plant/model mismatch will be studied next.

Acknowledgement: The authors gratefully acknowledge the financial support from the National Science Foundation (CTS-#0096326).

REFERENCES

Bequette, B. W. (1998). *Process Dynamics: Modeling, Analysis, and Simulation*. Prentice Hall. Upper Saddle River, New Jersey.

Boyan, J. A. and A. W. Moore (1995). Generalization in reinforcement learning: Safely approximating the value function. In: *Advances in Nueral Information Processing Systems: Proceedings of the 1994 Conference*. MIT Press. Cambridge, MA. pp. 369–376.

Chen, H. and F. Allgöwer (1998). A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability. *Automatica* **34**(10), 1205–1217.

Gordon, G. J. (1995). Stable function approximation in dynamic programming. In: *Proceedings of the Twelfth International Conference on Machine Learning*. Morgan Kaufmann. San Francisco, CA.

Kaisare, Niket S., Jong Min Lee and Jay H. Lee (2002). Simulation based method for nonlinear optimal control: Application to a microbial cell reactor. *International Journal of Robust and Nonlinear Control*.

Lee, J. H. (1997). Recent advances in model predictive control and other related areas. In: *Chemical Process Control – Assessment and New Directions for Research* (Y. C. Kantor, C. E. Garcia and B. Carnahan, Eds.). Vol. 93. AIChE Symposium series. pp. 201–216.

Lee, J. H. and N. L. Ricker (1994). Extended Kalman filter based nonlinear model predictive control. *Ind. Eng. Chem. Res.* **33**(6), 1530–1541.

Mayne, D. Q. and H. Michalska (1990). Receding horizon control of nonlinear systems. *IEEE Transactions on Automatic Control* **35**(7), 814–824.

Michalska, H. and D. Q. Mayne (1993). Robust receding horizon control of constrained nonlinear systems. *IEEE Transactions on Automatic Control* **38**(11), 1623–1633.

Qin, S. J. and T. A. Badgewell (1997). An overview of industrial model predictive control technology. In: *Chemical Process Control - V*. American Insitute of Chemical Engineers. New York. pp. 232–256.

NONLINEAR MODEL PREDICTIVE CONTROL OF MULTICOMPONENT DISTILLATION COLUMNS USING WAVE MODELS

S. Grüner [*] **S. Schwarzkopf** [*] **I. Uslu** [*] **A. Kienle** [**,***]
E.D. Gilles [**,*]

[*] *Institut für Systemdynamik und Regelungstechnik, Universität Stuttgart*
[**] *Max-Planck-Institut für Dynamik komplexer technischer Systeme Magdeburg*
[***] *Institut für Automatisierungstechnik, Otto von Guericke Universität Magdeburg*

Abstract: A novel control concept for multicomponent distillation columns is presented. The concept is based on nonlinear wave-propagation phenomena that occur in counter-current separation processes. On this basis a reduced order model has been developed in previous work that not only considers profile positions but also the profile shape itself. The reduced model gives direct access to key parameters of the plant, such as the separation front positions. Furthermore, it allows real-time computations for multicomponent distillation columns. Such a model is used for both, the nonlinear model predictive control (NMPC) and the observer design. The observer uses temperature measurements and gives estimated temperature and concentration profile positions as well as compositions in the product streams. The robustness of the observer is shown intuitively and in simulation studies. The control of multicomponent distillation is formulated within the NMPC framework by penalising the deviation of the front positions from their reference points and ensuring the product specifications by means of constraints. By directly taking account of product specifications the presented control concept differs from inferential control schemes known from literature. Due to the fact that the concept is based on simple temperature measurements an industrial application seems easily possible.

Keywords: multicomponent distillation, wave phenomena, nonlinear model predictive control, nonlinear observer

1. INTRODUCTION

In the past a vast number of studies has been done in the area of distillation column control. A review of the work produced in this field until early 90s is given by Skogestad (1997). Most of the approaches consider linear control methods. Although there exist a number of studies on nonlinear control of distillation columns, e.g. (Groebel *et al.*, 1995) they mainly concentrate on high purity binary distillation; a study together with a review on this field is presented in (Balasubramhanya and Doyle III, 1997).

In previous studies tray temperatures are frequently used as controlled variables instead of product compositions (Luyben, 1973; Yu and Luyben, 1984), since temperatures are easily measured online. For high purity binary distillation, the controlled temperatures are easily selected. In general, sensors are located at points where the temperature profile has a sharp transition and this corresponds to some distance away from the column ends. However such inferential control relies on the correlation between the temperature on the measurement trays and the product composition. This correlation becomes poor for multicomponent systems and consequently controlling temperatures alone may result in a considerable violation of the product specifications (Moore, 1992). These difficulties may be overcome by composition estimators derived on the basis of temperature measurements, as proposed by e.g. (Lang and Gilles, 1990; Mejdell and Skogestad, 1991*b*; Mejdell and Skogestad, 1991*a*; Quintero-Marmol *et al.*, 1991; Baratti *et al.*, 1998; Dodds *et al.*, 2001).

In the last decade a new low order modelling approach based on nonlinear wave propagation theory was developed for counter-current separation processes (Marquardt, 1990) taking into account proper profile shapes (Kienle, 2000). This type of wave models offers a more precise insight to the dynamic mechanisms of distillation processes. Additionally they explicitly take into account the separation fronts that determine the quality of the separation.

In (Shin *et al.*, 2000) a nonlinear profile observer together with a profile position control for bi-

nary distillation columns is presented. That work makes direct use of the fact that in binary distillation the composition follows from the temperature, which is not possible for multicomponent systems. In contrast, this contribution aims at directly controlling the product composition by adjusting the front position. The capabilities of wave position observers was shown recently in (Roffel *et al.*, 2002) for binary distillation.

Recent developments in the area of nonlinear model predictive control (NMPC) (Allgöwer *et al.*, 1999) provide an efficient control technique that is able to deal with the multi-variable nature of distillation processes and the process operating constraints. Furthermore, it is directly possible to utilise the nonlinear process model.

In this contribution, the advantages of the wave model together with the benefits of NMPC are used to control separation front positions such that the product specifications are met. At the same time further operation limits are respected. Together with an observer that is also based on the wave model, this contribution presents a consistently designed control system that is directly applicable to multicomponent distillation columns.

2. NEW CONCEPT FOR DISTILLATION CONTROL

Most common distillation column models are based on modelling each tray separately. In contrast, the wave model is based on integral balances and regards the concentration profiles by the use of suitable wave functions.

In previous studies it was realized that controlling temperatures on individual trays may have problems in the precences of disturbances in the feed composition. I.e. an adjustment of the setpoints may be come necessary even for binary distillations in order to operate in specification.

Due to these problems wave propagation based concepts have been successfully applied to the control of binary distillation columns (Han and Park, 1993; Balasubramhanya and Doyle III, 1997). An extensions of those concepts to multicomponent distillation columns is not trivial and hence such applications are still missing.

Using the wave model introduced in (Kienle, 2000), the concept can be expanded to multicomponent distillation. In this case there are $N_C - 1$ traveling wave fronts, where N_C stands for the number of components. From among $N_C - 1$ fronts the key separation front has to be selected. The position of this front is used as controlled variable afterwards. The selection can be made by analysing the concentration profiles obtained at the desired operating conditions. The key front is the front which performs the main separation with respect to the product specifications, e.g. in Fig. 2 the fronts s^1_{top} and s^2_{bot} near tray 11 and 40 are selected. The key front is typically a balanced front, i.e. it is a front with zero propagation velocity standing in the middle of a column section.

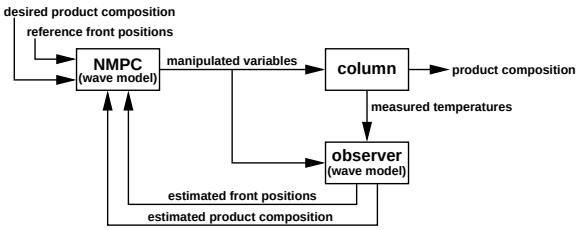

Fig. 1. Control setup

All other fronts are either pushed to the top or bottom of the column section and not able to pass the balanced front. The control aim is to balance the key front in the presence of disturbances and load changes.

2.1 *Control scheme*

Nonlinear model predictive control (NMPC) is chosen as control strategy since it is able to handle constraints on the states. Hence, in contrast to inferential control, the desired product specifications are respected at any time by including them as constraints.

The NMPC technique used in this contribution is based on the following main components: a nonlinear process model, measurements, a state estimator and an optimisation algorithm. As the wave fronts are not measurable, a suitable observer is designed to reconstruct the whole system state by measuring one temperature per column section. In this study, the same nonlinear wave propagation model is employed within both the NMPC and the observer. This makes it less time consuming to set up the complete control environment and only one parameter set has to be identified. The resulting control setup is shown in Fig. 1.

The feasibility of NMPC in real-time by the use of special high performance optimization algorithm is shown in (Diehl *et al.*, 2001*b*; Diehl *et al.*, 2003) for the control of a binary distillation column. Compared to the equilibrium stage model used in (Diehl *et al.*, 2001*b*), which has 42 differential states, a wave model based on similar assumptions only needs 7 differential states. The benefits of such an immense order reduction, are twofold. First it is possible to solve NMPC problems with limited computational power in real-time. Second the NMPC approach can be further exploited by the use of highly sophisticated, more time consuming, optimization strategies.

2.2 *The wave model*

In this section, the used wave model will be sketched, for details the reader is referred to (Kienle, 2000). The column is divided into sections by in- or outflows like feed or side streams. Each of these sections, e.g. the rectifying or the stripping part, is described by a wave model. Wave models are derived from the constant pattern wave phenomena appearing in distillation processes. The main equation is the integral component material balance

$$\frac{dh_i}{dt} = \dot{n}_{i,\mathrm{in}} - \dot{n}_{i,\mathrm{out}} \qquad i = 1 \dots N_C - 1 \qquad (1)$$

over one column section. The integral amount h_i of the component i is calculated with the relation

$$h_i = \sum_{k=1}^{N_S} n_k x_{i,k} \qquad i = 1 \dots N_C - 1 \qquad (2)$$

where n is the molar liquid holdup and x_i the mole fraction at each of the N_S trays. The vapour holdup is neglected. Both n_k and $x_{i,k}$ depend on the wave position and are calculated from the wave function. The slope of this wave mainly depends on the wave asymptotes and a mass transfer coefficient. Furthermore, constant relative volatilities and constant molar holdups are assumed.

The other parts of the column like feed tray, condenser and reboiler are described by standard equilibrium models.

3. OBSERVER DESIGN

As pointed out one of the key components of the proposed operating concept for multicomponent distillation columns is the observer. Therefore the main idea is explained in the following and the robustness of the approach is shown intuitively. All ideas will be shown for the case of a ternary distillation, but they can be transfered to distillations with any number of components.

3.1 *Main Idea*

The observer is built up of a plant model that acts as a simulator part and is augmented by an error injection. The error injection is designed by insight into the process dynamics (Lang and Gilles, 1990).

The main idea for the observer design is that the error injection has to try to match the estimated fronts with those of the plant. The following rules for the error injection can be figured out by analyzing the temperature and composition

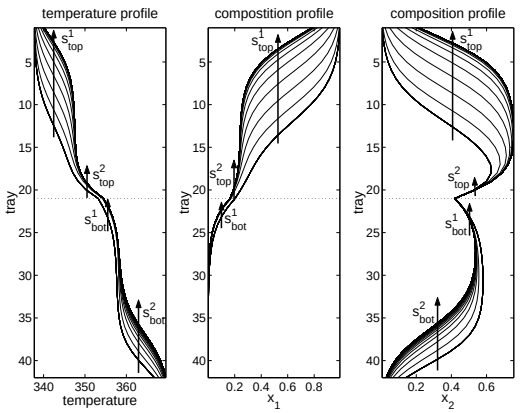

Fig. 2. Columns profiles after a reflux reduction by 25%. The fronts move in the direction of the arrows. The dotted line marks the feed location.

profiles shown in Fig. 2.

The key separating front in the top section, labeled s^1_{top} in Fig. 2, is located near tray 11 at the desired operating point. If the estimated temperature at this tray is too high the front has to be moved down, away from the condenser. This can be achieved by increasing the mole fraction of component 1 and decreasing that of component 2 via error injection.

The same analysis can be applied to the key separating front s^2_{bot} in the bottom section. However, component 1 does not contribute to the movement of the front. Consequently there is no sense in changing the mole fraction of component 1.

This observer design will work in a very robust way since no assumptions regarding the model structure as well as the precise parameter values have been made so far. This idea has been already successfully applied to a reactive distillation column in (Grüner *et al.*, 2001).

In the following the placement of the temperature measurements and subsequently the application of the proposed error injection to the wave model which is used as simulator will be shown.

3.2 *Sensor Placement*

Usually finding the right locations for the sensors is a difficult and important task in the observer design for spatially distributed systems. Wrong sensor placement may even make the process unobservable.

However, for distillation columns nonlinear wave propagation theory provides the necessary information. From the theory it follows, that there can be at most one balanced wave, i.e. a wave with zero propagation velocity in each column section. All other waves are either pushed against the top or the bottom boundary of the column section. Even the smallest changes in the flow rates or the feed composition will make this balanced wave move up or down in the column section. Consequently, this wave is the most sensitive to no matter how small a disturbance to the process is. Thus a temperature measurement in the middle of that front at its nominal location will detect all these movements and is the perfect location for a sensor. In addition to these considerations it should be noted that the control aim is to keep this front at its nominal location. Hence, in stable closed loop operation the wave will never be too far away from the sensor.

3.3 *Error Injection*

As pointed out the idea of the error injection is to move fronts up and down in the column. This can be achieved by injecting the estimation error into the integral component material balances (1) of the wave model. The resulting equation for one section, i.e. either stripping or rectifying section is as follows:

$$\frac{d\hat{\mathbf{h}}}{dt} = \underbrace{\dot{\mathbf{n}}_{\text{in}} - \dot{\mathbf{n}}_{\text{out}}}_{\text{simulator}} + \underbrace{\boldsymbol{\alpha}\left(T_m - \hat{T}_m\right)}_{\text{error injection}}, \qquad (3)$$

where T_m is the measured temperature and $\boldsymbol{\alpha}$ the $N_C - 1$-dimensional vector of correction coefficients. The sign of the components of $\boldsymbol{\alpha}$ in the integral material balances is chosen according to the reasoning in Section 3.1. Elements corresponding to compositions that do not contribute to a front movement may be set to zero. In order to reduce the number of tuning parameters, the absolute value of the elements of $\boldsymbol{\alpha}$ is assumed to be equal, resulting in one tuning parameter per column section.

The final magnitude of $\boldsymbol{\alpha}$ can only be determined in closed loop simulation studies. For the following observer performance analysis the $\boldsymbol{\alpha}$ were chosen to be $\boldsymbol{\alpha}_{\text{rect}} = [20, 20]^T$ and $\boldsymbol{\alpha}_{\text{strip}} = [0, -20]^T$.

It is pointed out, that the proposed observer not only gives estimates for the front positions, but also for the complete temperature and composition profiles. I.e., in contrast to e.g. (Shin *et al.*, 2000), it also estimates the product compositions. In addition, the proposed observer is applicable to multicomponent distillation.

3.4 *Observer Performance*

In order to investigate the performance of the proposed observer in the presence of unmeasured disturbances, simulation studies were carried out by using a much more detailed model, representing the plant. This model is a tray to tray constant molar overflow model using saturation pressures to describe the vapour-liquid equilibrium.

Very difficult disturbances to distillation columns are changes of the feed composition as shown in Fig. 3. But even for such a critical disturbance the

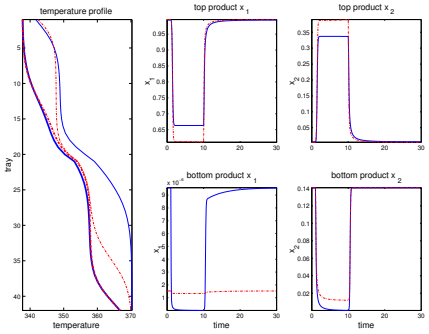

Fig. 3. Response of the observer to a step change of the feed composition at t=1.0 by 10% and back to the nominal value at t=10.0. (Initial profiles in bold, thin profile at t=10.0, solid lines process, dashed-dotted lines observer)

observer shows good performance.

Taking into account that besides other differences the vapour-liquid-equilibrium of the observer model is different from that of the plant model the observer gives good quantitative estimates for the product compositions, especially at the nominal operating point where the estimates are almost indistinguishable from the plant values.

Even more important for a good closed loop performance of the observer is its ability to capture the plant dynamics. The time plots of the product compositions shown in Fig. 3 verify that the observer is well able to render the dynamics of the plant.

Besides the simulation study shown in Fig. 3 numerous other simulation studies were done.

These simulation studies show the robustness and good performance of the proposed observer and give full confidence for good closed loop performance.

4. CONTROLER DESIGN

In this study, the nonlinear wave propagation model is employed within the NMPC framework. In the light of the discussion in Section 2 the control aim can be defined as to maintain the wave front positions at their required set points, while at the same time the product specifications have to be fulfilled. This has to be achieved in the presence of disturbances and constraints on the input and output variables.

The process model required for the NMPC framework is a DAE model of index one in the following form :

$$\begin{aligned} \dot{\mathbf{x}}(t) &= f(\mathbf{x}(t), \mathbf{z}(t), \mathbf{u}(t)), \\ \mathbf{0} &= g(\mathbf{x}(t), \mathbf{z}(t), \mathbf{u}(t)), \end{aligned} \tag{4}$$

together with suitable initial conditions, where $x(t)$ and $z(t)$ are the differential and algebraic state vectors, $u(t)$ is the control vector and t is the time.

The NMPC open-loop optimal control problem to solve for a prediction horizon $[0, T_p]$, with horizon length T_p, is given by

$$\min_{\mathbf{u}(\cdot),\mathbf{x}(\cdot)} \int_0^{T_p} \left\{ \| S_i(t) - S_i^{ref} \|_2^2 \right\} dt. \tag{5}$$

The controlled variables are the key front positions, S_i, and desired front positions are denoted by S_i^{ref}. Subscript i corresponds to the column section. The state and control inequality constraints are formulated by

$$c(\mathbf{x}(t), \mathbf{z}(t), \mathbf{u}(t)) \geq 0 \quad \text{for } t \in [0, T_p]. \tag{6}$$

In this particular case

$$c(\mathbf{u}) = \begin{bmatrix} u - u_{min} \\ u_{max} - u \end{bmatrix} \tag{7}$$

define lower and upper bounds for the controls. The important constraints on the top and bottom product compositions, x_T and x_B, are given by

$$\begin{bmatrix} x_T(t) - x_T^{ref} \\ x_B^{ref} - x_B(t) \end{bmatrix} \geq 0 \quad \text{for } t \in [0, T_p]. \tag{8}$$

The constraints on the states, Eq. (8), are treated as soft constraints using slack variables which are added in the objective function as linear penalty terms. Such an approach is particularly useful when output constraints represent the control objectives rather than hard limits in the process.

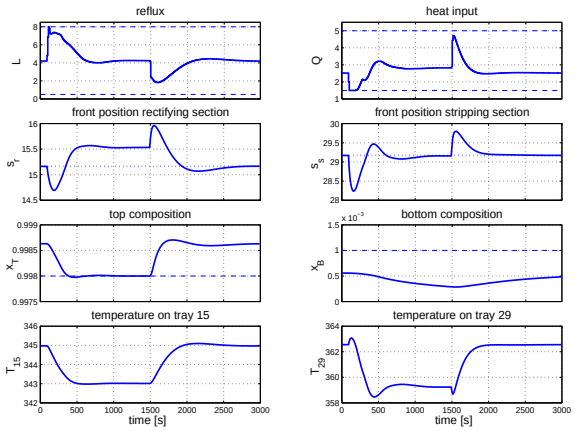

Fig. 4. Closed loop simulation experiment after 50% increase and decreasing to the original value of the light component in the feed. (dashed lines: boundaries, dotted lines: setpoints)

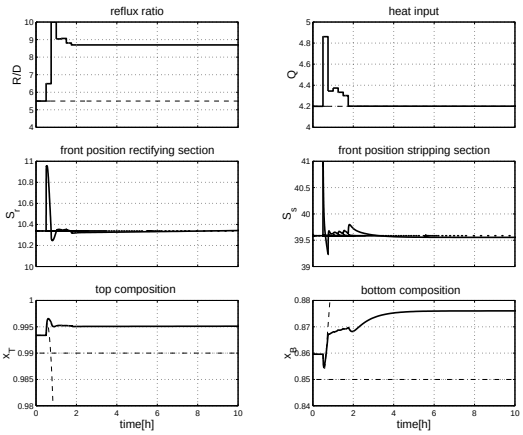

Fig. 5. Open loop simulation experiment after 33% decrease of the light component in the feed. (dash-dotted lines: boundaries, dotted lines: setpoints, dashed lines: uncontrolled)

To ensure nominal stability of the closed system a practical approach based on the result given in (Chen and Allgöwer, 1998) was taken. This is done by dividing the prediction horizon into a control horizon $[0, T_c]$ and a prediction interval $[T_c, T_p]$ along which the controls are kept constant at their final values.

In order to show the advantage of the proposed control concept at first a NMPC closed loop simulation study for the binary separation of methanol and 1-propanol is presented in Fig. 4. Controlling the only existing front in the binary problem is equivalent to controlling the key separation front in the ternary case.

The reflux flow-rate, L, and the heat input, Q, into the reboiler (which corresponds to the vapour flow rate out of the reboiler) are considered as manipulated variables (LV configuration).

The scenario considered in the following is a 50 % step increase of the light component in the feed occurring after 100 seconds and a decrease to its original value after 1500 seconds. T_c, is selected as 1200 seconds with 10 control intervals and T_p is 30000 seconds. The product concentration constraints are set to $x_T(t) \geq 0.998$ and $x_B \leq 0.001$. As shown in Fig. 4 the fronts have to be shifted to fulfill the product specification at the top of the column. After the disturbance disappeared the fronts are shifted back to their reference points.

If one chooses the inferential control scheme for the same binary distillation system, the controlled temperatures would be on trays 15 and 29 similar to (Diehl *et al.*, 2003) and as Fig. 4 clearly shows, set-points of the temperature controllers have to be modified to guarantee the product specifications.

The NMPC computations are carried out on a Unix workstation running under Linux (1 Ghz AMD Athlon processor), using an efficient dynamic optimisation algorithm which is based on a direct multiple shooting approach and available as the dynamic optimisation software, MUSCOD-II (Diehl *et al.*, 2001*a*).

The application of the control concept to multicomponent systems is demonstrated for the ternary system of methanol/ethanol/1-propanol. The two key fronts to be controlled are s^1_{top} and s^2_{bot} located near tray 11 in the rectifying section and near tray 40 in the stripping section at the nominal operating point respectively. Manipulated variables, the reflux ratio, L/D, and the heat input, Q (L/D,V configuration) are computed in 6 control intervals each of 900 seconds length and T_p is 30000 seconds.

Fig. 5 shows the input and state trajectory obtained as a solution to the NMPC open loop optimal control problem in the face of a disturbance in the feed concentration (33 % step decrease of the light component). The key front positions, S_r and S_s are kept nicely around desired reference values. Meanwhile constraints on the product concentrations, $x_T \geq 0.99$ and $x_B \geq 0.85$, are satisfied. The dynamic optimisation problem for *Case II* was solved within the process simulation environment DIVA(Kröner *et al.*, 1990). A standard Sequential Quadratic Programming (SQP) algorithm from NAG library is used. [1]

5. CONCLUSION

A novel concept for the control of multicomponent distillation columns has been proposed. The main idea of the control concept is based on the observation that the product composition is mainly influenced by the key separating front.

The control achieves the desired product specifications by adjusting the position of the key separation front in each column section, while the product specifications are ensured by constraints in the NMPC. Due to the fact that a direct specification of the product composition is possible,

[1] Current work is on the NMPC closed loop application for the ternary system within MUSCOD-II environment and results are likely to be presented at the conference.

the control concept is a direct control concept in contrast to the many inferential control concepts in the literature. The capability of the concept was shown in two case studies for a binary and a ternary distillation.

The NMPC is provided with the necessary information by an observer. Both, observer and NMPC use the same wave model. The robustness of the observer with respect to parameter errors as well as model-plant mismatches is intuitively shown and validated by simulation studies for a ternary separation.

In the future the control concept will be applied to a ternary separation of e.g. methanol/ethanol/1-propanol in a system of two coupled distillation columns. This will be done first in simulation studies and then at the pilot scale plant at the Institut für Systemdynamik und Regelungstechnik.

6. ACKNOWLEDGEMENTS

The authors thank M. Diehl for getting them into MUSCOD-II and R. Findeisen for past and present discussions on NMPC.
This research is in part supported by DFG within the joint research project SFB 412.

7. REFERENCES

Allgöwer, F., T.A. Badgwell, J.S. Qin and J.B. Rawlings (1999). In: *Advances in Control, Highlights of ECC'99* (P.M. Frank, Ed.). pp. 391–449. Springer.

Balasubramhanya, L.S. and F.J. Doyle III (1997). Nonlinear control of a high-purity distillation column using a traveling-wave model. *AIChE J.* **43**(3), 703–714.

Baratti, R., A. Bertucco, A. Da Rold and M. Morbidelli (1998). A composition estimator for multicomponent distillation columns - development and experimental test on ternary mixtures. *Chem. Eng. Sci.* **53**(20), 3601–3612.

Chen, H. and F. Allgöwer (1998). A quasi-infinite horizon nonlinear predictive control scheme with guarnteed stability. *Automatica* **34**(10), 1205–1218.

Diehl, M., D.B. Leineweber and A.S. Schäfer (2001*a*). *MUSCOD-II Users' Manual*. IWR-preprint 2001-25 ed.. University of Heidelberg.

Diehl, M., I. Uslu, R. Findeisen, S. Schwarzkopf, F. Allgöwer, H.G. Bock, T. Bürner, E.D. Gilles, A. Kienle, J.P. Schlöder and E. Stein (2001*b*). In: *Online Optimization of Large Scale Systems: State of the Art* (M. Grötschel, S.O. Krumke and J. Rambau, Eds.). pp. 363–384. Springer.

Diehl, M., R. Findeisen, S. Schwarzkopf, I. Uslu, F. Allgöwer, H.G. Bock, E.D. Gilles and J.P. Schlöder (2003). An efficient algorithm for nonlinear model predictive control of large-scale sytems, part 2: Experimental evaluation for a distillation column. *Automatisierungstechnik* **51**(1), 7–15.

Dodds, S., G. Adams, W. Heath and G. Goodwin (2001). In: *6th IFAC Symp. on Dynamics and Control of Process Systems*. Jejudo Island, Korea. pp. 497–501.

Groebel, M., F. Allgöwer, M. Storz and E. D. Gilles (1995). In: *Proc. ACC' 95*. pp. 2648–2652.

Grüner, S., K.-D. Mohl, A. Kienle, E.D. Gilles, G. Fernholz and M. Friedrich (2001). In: *6th IFAC Symp. on Dynamics and Control of Process Systems*. Jejudo Island, Korea. pp. 125–130.

Han, M. and S. Park (1993). Control of high-purity distillation column using a nonlinear wave theory. *AIChE J.* **39**(5), 787–796.

Kienle, A. (2000). Low-order dynamic models for ideal multicomponent distillation processes using nonlinear wave propagation theory.. *Chem. Eng. Sci.* **55**, 1817–1828.

Kröner, A., P. Holl, W. Marquardt and E.D. Gilles (1990). Diva–an open architecture for dynamic simulation. *Comput. Chem. Engng.* **14**, 1289–1295.

Lang, L. and E.D. Gilles (1990). Nonlinear observers for distillation columns. *Computers chem. Engng.* **14**(11), 1297–1301.

Luyben, W.L. (1973). Profile position control of distillation columns with sharp profiles. *AIChE J.* **18**(1), 238–240.

Marquardt, W. (1990). Traveling waves in chemical process. *International Chemical Engineering* **30**, 585–606.

Mejdell, T. and S. Skogestad (1991*a*). Composition estimator in a pilot-plant distillation column using multiple temperatures. *Ind. Eng. Chem. Res.* **30**, 2555–2564.

Mejdell, T. and S. Skogestad (1991*b*). Estimation of distillation composition from multiple temperature measurements using partial-least-square regression. *Ind. Eng. Chem. Res.* **30**, 2543–2555.

Moore, C.F. (1992). *Practical Distillation Control*. pp. 140–177. Van Nostrand Reinhold.

Quintero-Marmol, E., W.L. Luyben and C. Georgakis (1991). Application of an extended luenberger observer to the control of multicomponent batch distillation. *Ind. Eng. Chem. Res.* **30**, 1870–1880.

Roffel, B., B.H.L. Betlem and R.M. de Blouw (2002). A comparison of the performance of profile position and composition estimators for quality control in binary distillation. *Computers and Chemical Engineering*.

Shin, J., H. Seo, M. Han and S. Park (2000). A nonlinear profile observer using tray temperatures for high-purity binary distillation column control. *Chem. Eng. Sci.* **55**, 807–816.

Skogestad, S. (1997). Dynamics and control of distallation columns – a critical survey. *Modelling, Identification and Control* **18**, 177–217.

Yu, C.-C. and W.L. Luyben (1984). Use of temperatures for the control of multicomponent distillation columns. *Ind. Eng. Chem. Process Des. Dev.* **23**, 590–597.

NONLINEAR MODEL PREDICTIVE CONTROL OF CEMENT GRINDING CIRCUITS

R. Lepore [*,1] **A. Vande Wouwer** [*] **M. Remy** [*]

[*] *Laboratoire d'Automatique*
Faculté Polytechnique de Mons
Boulevard Dolez 31, B-7000 Mons, Belgium

Abstract: Based on a reduced-order model of a cement grinding circuit, a nonlinear model predictive control strategy is developed. The first step of this NMPC study is the definition of control objectives which consider product fineness, product flow rate and/or grinding efficiency. At this stage, one of the main concerns is to relate these objectives to easily measurable particle weight fractions. Second, NMPC is implemented so as to take the various constraints on the manipulated variables and operating conditions of the mill into account. Third, robustness with respect to model uncertainties is analyzed, and the most critical parameters are highlighted. Finally, an NMPC scheme, combining a stable inner loop for controlling the mill flow rate and a DMC-like compensation of the model mismatch, is proposed.

Keywords: nonlinear systems; modeling; predictive control;
grinding (comminution); cement industry

1. INTRODUCTION

Control of cement grinding circuits is a delicate task. According to (Hulbert, 1989) and (Hodouin and Del Villar, 1994), the difficulties associated with control arise from two major causes:

- process complexity and nonlinearity: grinding depends on the material content of the mill, separation is affected by the material flow rate and the process has recycle;
- lack of measurements: some variables cannot be measured on-line, others are heavily corrupted by noise.

In recent years, some studies have witnessed the relevance of model predictive control for cement grinding processes. In (Martín Sánchez and Rodellar, 1996), a single mill is considered (no downstream classification) and maximization of the product flow rate is achieved while stabilizing the degree of material filling. In (Magni *et al.*, 1999), a nonlinear model of a grinding circuit has been developed and the delicate problem of stability has been treated. These studies essentially consider global variables, e.g., flow rates, total material content of the mill,....

In contrast with these studies, the authors have focused on the transport of the material in the mill and, mostly, on the particle size distribution, which is highly related to the final properties of cement, such as the compressive strength. From this latter philosophy, they have developed:

(1) a distributed-parameter population model, which has been identified on an industrial closed-loop grinding process (C.B.R., Belgium) (see (Boulvin, 2000) and (Boulvin *et al.*, 2002));

[1] Author to whom correspondence should be addressed:
e-mail: Renato.Lepore@fpms.ac.be
phone: +32 (0)65374140 fax: +32 (0)65374136

(2) a simplified distributed-parameter model, based on a reduced number of size intervals (Lepore *et al.*, 2002).

The contribution of the present study is:

- to formulate new control objectives in agreement with the coarser size discretization used in the reduced-order model (2), i.e., a) a well-determined fineness of the product, b) either a maximization of the product flow rate or an optimization of the grinding efficiency;
- to design a multivariable, constrained NMPC scheme achieving these objectives, which considers the input flow rate and the classifier selectivity as manipulated variables. Constraints apply a) on magnitudes and slew rates of the inputs (saturation effects), b) on the mill flow rate variable (preventing temperature increase and/or wear as well as mill overfilling);
- to treat the following aspects of model inaccuracy: to investigate NMPC robustness to model uncertainties, to perform a thorough parameter sensitivity analysis and to study two types of model mismatch compensation: (a) a typical DMC-like scheme, which considers the mismatch as a constant disturbance over the prediction horizon, (b) the DMC-like scheme with prior stabilization of the mill flow rate by a proportional control loop.

In the sequel, the document is divided into five sections. Section 2 contains the description of the process and the equations of the reduced-order model. New control objectives are formulated in Section 3. In Section 4, the NMPC strategy is presented, then robustness analysis is considered in Section 5. Some conclusions and perspectives are finally presented in Section 6.

2. PROCESS DESCRIPTION AND MODELING

2.1 Process description

A typical cement grinding circuit is represented in figure 1, which consists of a single-compartment ball mill in closed-loop with an air classifier. The raw material (usually clinker) flow q_C is fed to the rotating mill, in which balls perform the breakage of the material particles by fracture and/or attrition. At the other end, the output or mill flow q_M is lifted up by a bucket elevator onto the classifier which separates the material into two parts: the product flow q_P and the rejected flow q_R, which is recirculated to the mill inlet. The selectivity of the classifier and, in turn, the product fineness, can be modified by acting on special registers Rp. The sum of q_C and q_R is the total feed flow, denoted by q_F.

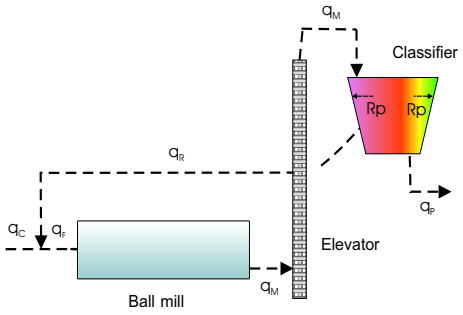

Fig. 1. Closed-loop grinding circuit

2.2 Modeling

Consider the size continuum as divided into three size intervals numbered 1, 2 and 3 for the coarse, intermediate and fine particles, respectively. Mass balances lead to:

$$\frac{\partial X_i}{\partial t} = -u_i \frac{\partial X_i}{\partial x} + D_i \frac{\partial^2 X_i}{\partial x^2} + \sum_{j=1}^{2} k_{ij} \varphi_j \quad ; \text{i=1,2,3}$$

$$(k_{ij}) = \begin{pmatrix} -1 & 0 \\ +k & -1 \\ 1-k & +1 \end{pmatrix} \tag{1}$$

where:

- X_i is the mass per unit of length of the particles in size interval i;
- k is the yield fraction of the particles in size interval 2 appearing from the breakage of the particles in size interval 1; φ_j is the breakage rate of the material in size interval j;
- u_i is the convection velocity and D_i is the diffusion coefficient of the particles in size interval i;

The partial differential equations (1) are supplemented by initial (2) and boundary (3) conditions:

$$X_i(0,x) = H_0(x) w_{0;i}(x) \quad \forall \text{ x; i=1,2,3} \tag{2}$$

$$0 = u_i X_i - D_i \frac{\partial X_i}{\partial x} - q_F w_{F;i} \quad \text{x=0; i=1,2,3}$$

$$0 = \frac{\partial X_i}{\partial x} \quad \text{x=L; i=1,2,3} \tag{3}$$

where:

- $H_0(x)$ is the initial material content per unit of length, $w_{0;i}(x)$ is the corresponding mass fraction in size i;
- q_F is the total feed flow rate, $w_{F;i}$ is the corresponding mass fraction in size i.

The breakage rates are formulated as follows:

$$\varphi_j = \alpha_j X_j e^{-\beta H} \quad ; \text{j=1,2} \tag{4}$$

where:

- α_j is the specific rate of breakage for size interval j;
- H is the hold-up, i.e., $(X_1 + X_2 + X_3)$;
- β is an inhibition coefficient.

The classifier has very fast dynamics compared to the mill and is therefore described by a steady-state model. Selectivity is the fraction of material in each size interval which is recirculated (see the "fish-hook" curves in 2).

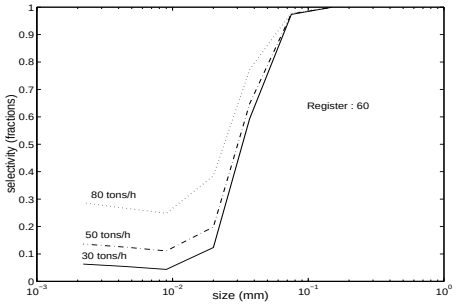

Fig. 2. Classifier selectivity (for a single register position and several mill flow rates)

3. CONTROL OBJECTIVES

The compressive strength specified by the client is the main goal of the cement manufacturer. Among the many characteristics of cement, the chemical composition, which is essentially set during the kiln operation, and the particle size distribution are the variables influencing at most the compressive strength. So, given the chemical composition (as it is considered for the grinding circuit), the compressive strength is dependent a priori on the total particle size distribution, which represents a very complex, hardly interpretable objective.
However, from recent industrial results (experimental data collected at C.B.R., Belgium), it appears that a strong relationship exists between the compressive strength and the weight fraction of the fine particles in the product, denoted by $w_{P;3}$. In figure 3 ($w_{M;2}$ is the weight fraction of the intermediate particles at the mill outlet), the arc $\overline{AB}$ represents all operating points corresponding to some constant $w_{P;3}$ that are compatible with operational restrictions on the mill flow rate, which prevent on the one hand dramatic temperature increase and/or wear of the equipment (e.g., at least 50 t/h), on the other hand mill overfilling (e.g., at most 140 t/h).

The second objective, usually more related to the economical strategy of the company itself, will set a well-determined operating point on the characteristic $\overline{AB}$. One common strategy is the maximization of the product flow rate. Provided that the process characteristic $\overline{AB}$ is available, this objective can be uniquely identified by the corresponding value $w_{M;2}^{Pmax}$ (see point 1). This strategy requires the process to be run at the stability limit. In fact, the arcs $\overline{A1}$ and $\overline{1B}$ correspond to stable and unstable operating points, respectively.
Another strategy could be to optimize the grinding efficiency or, in other words, to avoid overgrinding. From our description based on three size intervals, it is suggested to achieve this goal by avoiding at maximum coarse particles (obviously!) and also fine particles (overgrinding) at the mill outlet. So, increasing $w_{M;2}$ up to a reasonable limit, could be the criterion (e.g., point 2). It is noted that this strategy requires the process to be run completely in the unstable region.

Several advantages arise from using $w_{P;3}$ and $w_{M;2}$ as controlled variables:

- the measurements are simple (two sieve measurements only for each variable) and can be achieved automatically at moderate cost; measurement error is small since one can expect high values for $w_{P;3}$ ($0.7 \sim 0.8$) and $w_{M;2}$ ($0.5 \sim 0.7$); the measurement of $w_{M;2}$ is more reliable than, say, the elevator power which is very affected by mechanical vibrations (low-frequency noise);
- the use of $w_{M;2}$ for the achievement of operating point 1 converts an ill-conditioned optimization problem (well-determined product fineness and maximization of the product flow rate) into a well-conditioned minimization of an output error.

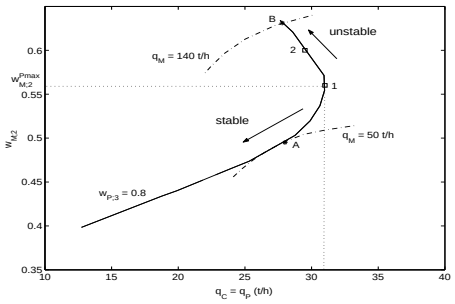

Fig. 3. Steady-state relationship for constant $w_{P;3}$

4. NMPC STRATEGY

NMPC consists in determining a set of manipulated-variable moves over a control horizon of N_u sampling periods that minimizes an objective function J over a prediction horizon of N_h sampling periods. The manipulated variables are the feed flow rate q_C and the register position Rp and the controlled variable y is the vector $[w_{P;3}\ w_{M;2}]^T$. At time instant k, the function J and the reference trajectory $y_{r;i}$ are defined as follows:

$$J(k) = \sum_{i=1}^{N_h} (y_{r;i} - \hat{y}_{k;i})^T Q (y_{r;i} - \hat{y}_{k;i}) \qquad (5)$$

$$y_{r;i} = y^* + (y_k - y^*)e^{-\frac{iT_s}{T_r}} \quad (6)$$

where:

- $\hat{y}_{k;i}$ is the predicted value at time $(k+i)T_s$;
- $y_{r;i}$ is the reference value at time $(k+i)T_s$;
- Q is the weighting matrix
- T_s is the sampling period;
- T_r is the time constant of the reference trajectory;
- y^* is the two-component set point $[w^*_{P;3}\ w^*_{M;2}]^T$
- y_k is the two-component measured value

In addition, the following constraints apply:

- box constraints on the manipulated variables: $0 \leq q_C \leq q_C^{max}$ (saturation of the feeding mechanism), $0 \leq Rp \leq Rp^{max}$ (minimum and maximum displacement of the registers)
- linear constraints on the manipulated variables (limits to the slew rates): $\mid q_C(i+1) - q_C(i) \mid \leq \Delta q_C^{max}$, $\mid Rp(i+1) - Rp(i) \mid \leq \Delta Rp^{max}$
- nonlinear constraints: limits to the mill flow rate value at the end of the prediction horizon preventing high cement temperatures $(q_M((k+H_p)T_s) \geq q_M^{min})$, mill overfilling $(q_M((k+H_p)T_s) \leq q_M^{max})$

Table 1 contains the values of the most important parameters mentioned above.

T_s	10 min	q_C^{max}	50 t/h	q_M^{min}	50 t/h
T_r	10 min	Rp^{max}	100	q_M^{max}	140 t/h
H_u	1	Δq_C^{max}	10 t/h	Q	$I_{(2,2)}$
H_p	5	ΔRp^{max}	50		

Table 1. Parameter values of the NMPC

The minimization of the objective function (5) is performed using the "Optimization toolbox 2.0" from Matlab 6.0. The solution of the partial differential equations is achieved using (a) a "method of lines" Matlab procedure for spatial differentiation (b) standard solvers from Matlab 6.0 for the integration in time of the differential equations.

5. ROBUSTNESS ANALYSIS AND MODEL MISMATCH COMPENSATION

In the sequel, we will (1) study the effect of grinding efficiency on steady-state characteristics , and, in turn, on the performance of the NMPC (2) evaluate systematically the impact of individual changes in the parameters (sensitivity analysis) (3) discuss two correction schemes, i.e., a simple DMC-like scheme and a DMC-like scheme with prior stabilization of the mill flow rate by an internal proportional loop.
For illustration purposes, we consider a step change in the set point y^* from the steady-state value $[0.71\ 0.46]^T$ (stable region) to $[0.80\ 0.56]^T$ (stability limit).

5.1 Effect of grinding efficiency on steady-state characteristics

Occurrences of model mismatch are obtained by modifying the process specific rates of breakage $\alpha_i^{mod} = \alpha_i^{proc} \cdot C$ $(i = 1, 2)$; cases (a) $(C = 0.9)$ and (b) $(C = 1.1)$ correspond to lower and higher efficiency, respectively. The static characteristics for the process and the two model occurrences are represented in figure 4, circles indicate the two corresponding operating points targeted by the optimization algorithm.

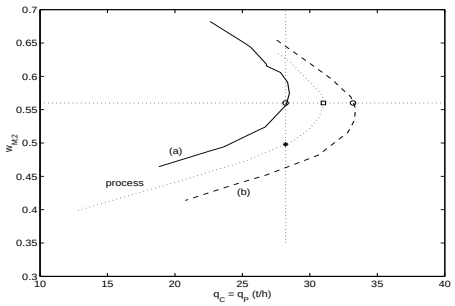

Fig. 4. Steady-state relationship for $w_{P;3} = w^*_{P;3}$; (a): $C = 0.9$; (b): $C = 1.1$

From the temporal evolution of the most relevant variables (see figure 5), it can be deduced that:

- in case (a), the high gain existing between Rp and $w_{P;3}$ allows the desired steady-state value $w_{P;3} = 0.8$ to be reached for the process. On the other hand, the input flow rates computed by the algorithm drive the process to the operating point represented by a star in figure 4 ($w_{M;2} = 0.498$); so, NMPC leads to the desired product fineness but to lower production flow rate than expected, the difference increasing with the model mismatch;
- in case (b), the manipulated-variable values computed by the optimization algorithm (particularly a too high input flow rate) lead inevitably to process overfilling. As a result, the two weight fractions, $w_{P;3}$ and $w_{M;2}$, tend to zero.

5.2 NMPC sensitivity to individual parameter inaccuracies

In the sequel, each parameter is modified by -10% and $+10\%$ from the estimated value and figure 6 represents the resulting effect on the steady-state characteristics. It is mostly noted that:

- variations in α_2 affect substantially the behaviour of the model whereas those in α_1 do

not (in fact, the lower values of α_2 determine the dominant time constants of the model);
- the parameters k and β are as relevant as α_2;
- changes in the velocity u and, particularly, the diffusion D have little influence on the steady-state characteristics.

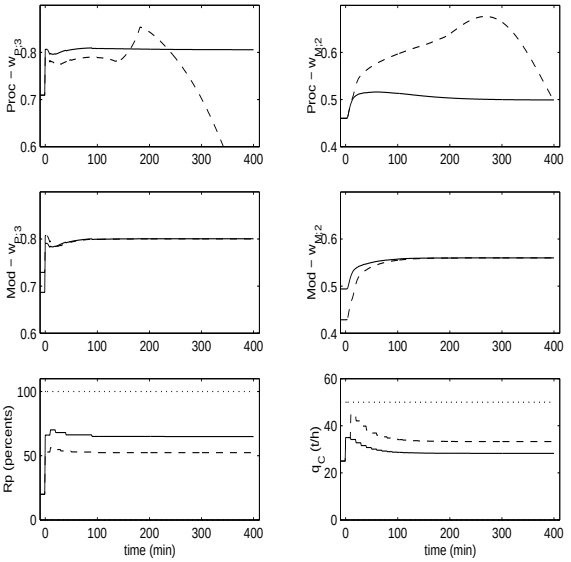

Fig. 5. NMPC performance with model mismatch (solid) : $C = 0.9$; (dashed) : $C = 1.1$; (dotted): maximum value constraints

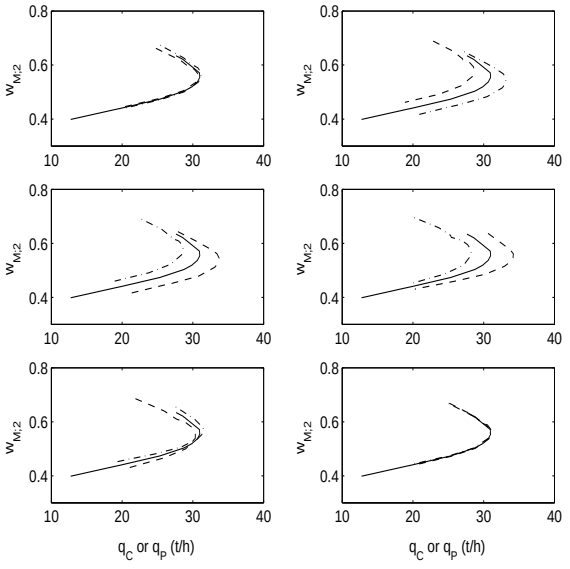

Fig. 6. Parameter sensitivity analysis : α_1 and α_2 (top figures), k and β (intermediate figures), u and D (bottom figures); (dashed) -10%, (dash-dotted) $+10\%$ on the parameter estimate, respectively

5.3 Model mismatch compensation

5.3.1. DMC-like scheme In this scheme, the mismatch at time t_k between the process and the model is viewed as an external, constant disturbance on the state vector all over the prediction horizon. The disturbance d_k is first estimated by $d_k = x_{proc;k} - x_{mod;k}$, then the reference trajectory is adjusted by the corresponding constant value over the prediction horizon. Figures 7 and 8 show the following results when this correction is applied in cases (a) and (b) of model mismatch:

- case (a): the modified reference trajectories (dotted lines) require the model to be run in the unstable region where constraints on the mill flow rate variable ($140\ t/h$) become active (see the evolution of the model prediction in figure 8). This constraint is responsible for the limitation on the input flow rate q_C. The closed-loop process is stable but the steady-state values (particularly $w_{M;2}$) are not satisfactory;
- case (b): the modification of the reference trajectory brings the targeted set point into the stable region, so that the closed-loop process is stable and no steady-state error appears.

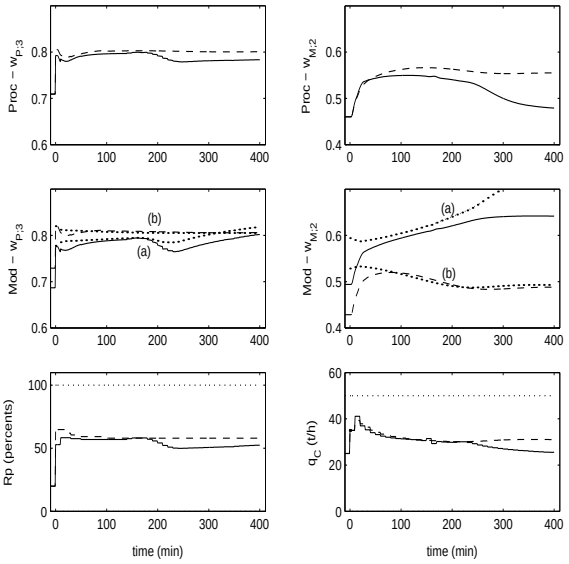

Fig. 7. DMC-like compensation results; (solid): $C = 0.9$, (dashed): $C = 1.1$

In conclusion, the DMC-like correction, which considers the model inaccuracy as a disturbance, guarantees feasibility and stability but does not guarantee satisfactory performance, particularly with respect to the steady-state error of $w_{M;2}$.

5.3.2. DMC-like scheme with prior stabilization Two embedded schemes are used: (a) an inner proportional loop controls the mill flow rate by acting on the input flow rate and ensures stable operation (b) the outer scheme is the NMPC itself which uses the mill flow rate set point q_M^* of the inner loop instead of the input flow rate q_C as the second manipulated-variable component. The

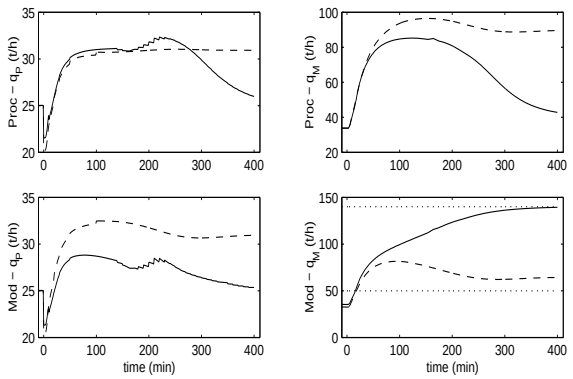

Fig. 8. DMC-like compensation results; (solid): $C = 0.9$, (dashed): $C = 1.1$

capabilities of the NMPC are entirely devoted to the performance achievement. Box constraints apply on q_M^* (0 and 200 t/h) and supplementary nonlinear constraints apply on the absolute value of q_C, which is now an intermediate variable.
Figure 9 shows the results obtained when the correction is applied to cases (a) and (b). Both cases demonstrate stability, satisfactory time responses and negligible steady-state error.

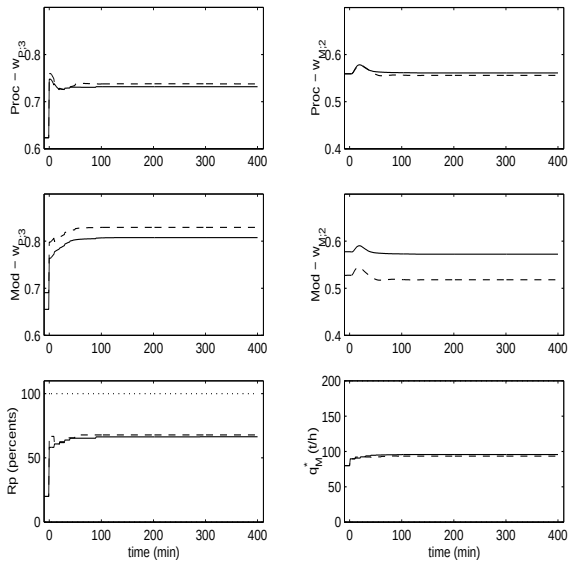

Fig. 9. DMC-like with prior stabilization; (solid): $C = 0.9$, (dashed): $C = 1.1$

6. CONCLUSION

Based on a reduced-order model of a cement grinding plant, a nonlinear model predictive control strategy is developed and analyzed. As a first step, new control objectives are defined, which are based on two weight fractions only: (a) the fraction of fine particles in the product, which is related to the compressive strength (b) the fraction of intermediate-size particles at the mill output, which is related to product maximization or optimum grinding efficiency. One major advantage is that two-sieve measurements of these variables could be achieved at low cost.
NMPC achieves these objectives by using the registers' position and the input flow rate as manipulated variables.
Robustness analysis leads to the following observations:

- model mismatch may lead to closed-loop instability (for example when the model has higher grinding efficiency); otherwise, steady-state errors affect the mill output but not the product fineness;
- not all the individual parameters have the same impact; experiments should be designed to accurately estimate grinding efficiency and, particularly, the appearance and the disappearance mechanisms of the intermediate-size particles and the inhibition effect of the material content;
- a DMC-like scheme cannot guarantee satisfactory performance; however, when a prior stabilization of the mill flow rate is achieved (here with a simple proportional loop using the input flow rate as a manipulated variable), very satisfactory results are obtained with DMC-like scheme in terms of stability and steady-state error.

REFERENCES

Boulvin, M. (2000). Contribution à la modélisation dynamique, à la simulation et à la conduite des circuits de broyage à boulets utilisés en cimenterie. PhD thesis. Faculté Polytechnique de Mons.

Boulvin, M., A. Vande Wouwer, R. Lepore, Chr. Renotte and M. Remy (2002). Modeling and control of cement grinding processes. *IEEE transactions on control systems technology (in press)*. (in press).

Hodouin, D. and R. Del Villar (1994). Conduite des unités de broyage. *Techniques de l'ingénieur* **JP**(J 3 110), 1–26.

Hulbert, D.G. (1989). The state of the art in the control of milling circuits. In: *Automation in mining, mineral and metal processing* (Koppel V., Ed.). IFAC Pergamon Press. Oxford.

Lepore, R., A. Vande Wouwer and M. Remy (2002). Modeling and predictive control of cement grinding circuits. IFAC '02, Barcelona, Spain.

Magni, L., G. Bastin and V. Wertz (1999). Multivariable nonlinear predictive control of cement mills. *IEEE transactions on control systems technology* **7**(4), 502–508.

Martín Sánchez, J. M. and J. Rodellar (1996). *Adaptive Predictive Control: from the concepts to plant optimization*. Prentice Hall.

OPTIMAL OPERATION AND CONTROL OF A REACTIVE SIMULATED MOVING BED PROCESS

Abdelaziz Toumi * **Sebastian Engell** *

* *Process Control Laboratory*
Department of Chemical Engineering, University of Dortmund
D-44221 Dortmund, Germany

Abstract: In this paper, we investigate the continuous production of High Fructose Corn Syrup (HFCS) in a Reactive Simulated Moving Bed process (RSMB). The RSMB process combines a quasi-continuous chromatographic separation with an enzymatic biochemical conversion of glucose to fructose. For the equilibrium limited glucose isomerization such an integration is suitable. The optimal operation of the RSMB process is determined using a sequential approach based on a rigorous mathematical model of the plant. In addition, we propose a new strategy to determine the distribution of the columns over the zones in the RSMB plant circumventing the solution of a Mixed Integer Nonlinear Problem (MINLP). During the operation of the RSMB process, disturbances occur (e. g. continuous decrease of the enzyme activity) which lead to an off-spec product. The control objective is to maintain the product purity while injecting a minimal additional amount of eluent. We propose a nonlinear model predictive controller which can deal with the complex hybrid dynamic of the RSMB plant as well as with hard constraints. The parameters of the non-linear process model are periodically estimated by least-squares fitting to online measurements. The efficiency of the whole control concept is shown in simulation studies for a 6-column RSMB plant.

Keywords: Simulated-Moving-Bed (SMB), Glucose Isomerization, High Fructose Corn Syrup (HFCS), Reactive Chromatography, Asynchronous Chromatography (VARI*COL*), Nonlinear model predictive control (NMPC).

1. INTRODUCTION

Glucose has only about 70% of the sweetness of sucrose and is less soluble in water. At the commercial concentration, glucose syrup must be kept above room temperature to prevent crystallization. Fructose is 30% sweeter than sucrose and twice as soluble as glucose at low temperatures. Using enzyme technology, by at least 50% conversion of glucose to fructose both problems can be overcome giving a stable high fructose corn syrup that is as sweet as a sucrose solution. The present world market for high fructose corn syrup is over 5 million tons and it is still expanding. This is due to the fact that the commercially available 'glucose isomerase' is remarkably resistant to changing temperatures and can be handled at high substrate concentrations. Most of the currently produced fructose syrups is obtained by the hydrolysis of starch into glucose followed by isomerization of glucose to fructose (Asif and Abaseed, 1998). This process produces syrups containing only about 42% fructose which has to be enhanced by selective removal of glucose or by applying multistage chromatographic separation methods.

A 3-zones-SMB process is proposed in this paper for the continuous isomerization of glucose, where reaction and separation are integrated in one apparatus leading to significant improvements in process performance. The optimization and the control of this complex hybrid system represent challenging tasks, and advanced strategies are required.

This work is structured as follows: in section I, the process is described and a detailed mathematical process model is presented. In section II, a mathematical model-based sequential optimization approach (Klatt *et al.*, 2001) is aplied to this process. In addition, we show how to use a new

[1] The authors are indebted to Prof. H. Schmidt-Traub and his research group for their support and valuable input to this work. The financial support of the Deutsche Forschungsgemeinschaft, in the context of the research cluster "Integrated Reaction and Separation Processes" at the University of Dortmund (SCHM 808/5-1) under grant (DFG En 152/26) is very gratefully acknowledged.

multicolumn continuous process called VARI*COL* which was recently introduced by (Ludemann-Hombourger and Nicoud, 2000), in order to determine the distribution of the columns over the zones of the RSMB plant. In section III, the control problem will be discussed. The objective is to maintain the desired product purity using a minimal amount of eluent. A nonlinear model predictive controller is developed, where at every sampling moment the optimal trajectory over a future control horizon is calculated online. The parameters of the rigorous process model are estimated periodically by least-squares fitting to online measurements. The efficiency of the control concept is shown in simulation studies for a 6-column RSMB plant.

2. THE RSMB PROCESS FOR GLUCOSE ISOMERIZATION

2.1 Process description

The process consists of a number of fixed beds, which are interconnected to form a closed-loop arrangement. All columns are homogeneously packed with an ion exchange resin (Amberlite CR-13Na) and the imobilized enzyme Sweetzyme T (supplied by Novo Nordisk Bioindustriale). A counter-current movement between the solid and the liquid phase is achieved by simultaneously advancing the inlet and outlet ports in direction of the liquid flow. In this special Simulated Moving Bed process, no attempt is made to achieve a complete separation of glucose and fructose, since the most common type of fructose syrup, usually called high-fructose syrup, is made in two categories: HFCS42 (42% fructose) and HFCS55 (55% fructose). For some purposes, a syrup with more than 55% fructose , called a higher-fructose syrup is desirable. Thus, the objective is to transform a syrup containing only pure glucose to one, where the glucose is **partially** converted to fructose.

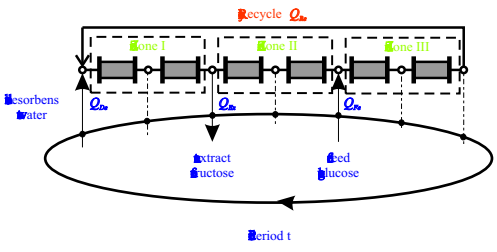

Figure 1. Schematic diagram of the 3-zones RSMB process for glucose isomerization

After start-up, a reactive True Moving Bed unit would achieve a steady-state, in which every process variable remains constant in time at any spatial location. In contrast, due to the discrete switching of the columns, SMB units do not reach a steady-state but rather a *cyclic* steady or periodic one (see figure 2).

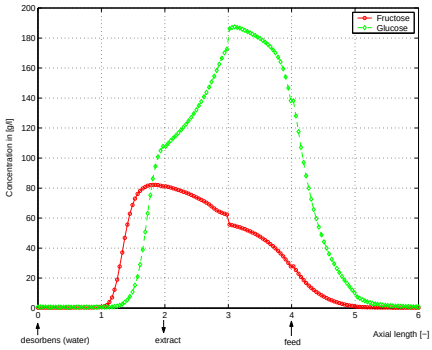

Figure 2. Axial concentration profile at the begin of the period for $Pur_{Ex} = 70\%$

2.2 Mathematical Modeling

A lot of work has been published on modeling of chromatographic processes (Gu, 1995). Accurate dynamic models of multi-column continuous chromatographic processes consist of coupled dynamic models of each column under consideration of the periodic port switching. A single chromatographic column is described by the *General Rate Model* which accounts for all important non-idealities of the column, axial dispersion, pore diffusion and the mass transfer between liquid and solid phase:

$$\begin{aligned} \epsilon_b \frac{\partial c_{b,i}}{\partial t} + \frac{(1-\epsilon_b)3k_{l,i}}{R_p}(c_{b,i} - c_{p,i}|_{r=R_p}) \\ + \epsilon_b r^{\text{liq}}_{\text{kin},i} = \epsilon_b D_{ax} \frac{\partial^2 c_{b,i}}{\partial x^2} + \epsilon_b u \frac{\partial c_{b,i}}{\partial x} \\ (1-\epsilon_p)\frac{\partial q_i}{\partial t} + \epsilon_p \frac{\partial c_{p,i}}{\partial t} = \epsilon_p D_{p,i} \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial c_{p,i}}{\partial r}\right). \end{aligned} \quad (1)$$

The model is completed by the mass balances at the nodes, and boundary and initial conditions which represent the switching process. The adsorption equilibrium and the reaction kinetic have to be determined experimentally. For the adsorption isotherm we assume a parabolic behavior:

$$q_i = H_i c_{b,i} + k_i c_{b,i}^2 + k_{ij} c_{b,i} c_{b,j},\ i,j = A,B. \quad (2)$$

The reaction kinetic can be accurately described by a first order pseudo-kinetic model (Fricke and Schmidt-Traub, 2002):

$$r^{\text{liq}}_{\text{kin},i} = \nu_i \frac{k_m (c_{b,i} - c_{b,j})}{k_{eq}},\ i,j = A,B. \quad (3)$$

The parameters are taken from (Fricke and Schmidt-Traub, 2002) and are listed in Appendix A. The resulting system of coupled partial differential equations can be solved efficiently using the approach introduced by (Gu, 1995), where a finite element discretization of the bulk phase is combined with an orthogonal collocation of the solid phase.

3. OPTIMAL OPERATION

3.1 Optimizing the operating parameters

The goal is to minimize specific separation costs for a given plant meeting the required product purities after the process has reached the cyclic steady state (CSS). For the description of the CSS, the operator Φ is introduced which represents the process dynamics $\mathbf{f}(\mathbf{x}, \mathbf{u})$ and the switching operations between two switching intervals:

$$\mathbf{x}_{k+1} = \Phi(x_k) \Leftrightarrow \begin{cases} \mathbf{x}^*_{k+1} &= \int_{t=0}^{\tau} \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))dt, \\ \mathbf{x}_0 &= \mathbf{x}_k, \\ \mathbf{x}_{k+1} &= \mathbf{P}\mathbf{x}^*_{k+1}. \end{cases} \tag{4}$$

The switching operation causes a re-initialization of the initial value of the dynamic simulation and is represented by the permutation matrix $\mathbf{P}$. In the CSS, the axial concentration profile $\mathbf{x}_k$ at the end of period k does not change from period to period, which can be checked numerically as:

$$\|\Phi(\mathbf{x}_k) - \mathbf{x}_k\| \leq eps_{steady}. \tag{5}$$

Then, given a SMB process with a *fixed* column partition, the optimization problem can be stated as follows:

$$\begin{aligned} \min_{Q_{De}, Q_{Ex}, Q_{Fe}, Q_{III}, \tau} \quad & \text{Cost}_{spec}(k) \\ \text{s.t.} \quad \|\Phi(\mathbf{x}_k) - \mathbf{x}_k\| &\leq eps_{steady}, \\ Pur_{Ex,k} &\geq Pur_{Ex,min}, \\ Q_I &\leq Q_{max}. \end{aligned} \tag{6}$$

An inequality constraint is imposed on the product purity. Since the flow rate in zone I is the highest one in the plant, it is constrained in order to avoid violation of the maximal pressure drop delivered by the pumps. The objective function must be specified based on the available data on the operating cost.

The natural degrees of freedom are the flow rates of desorbent Q_{De}, feed Q_{Fe}, recycle Q_{III} and the switching period τ. In the framework of optimization, they are transformed to the β-factors (Hashimoto *et al.*, 1983), where the apparent solid flow rate Q_S is introduced:

$$\begin{aligned} Q_S = \frac{(1-\epsilon_b)A\ L}{\tau}, \beta_1 = (\frac{Q_1}{Q_S} - \frac{1}{F})/H_A \\ \beta_2 = (\frac{Q_2}{Q_S} - \frac{1}{F})/H_B, \frac{1}{\beta_3} = (\frac{Q_3}{Q_S} - \frac{1}{F})/H_A, \end{aligned} \tag{7}$$

and which reflects the fact that in SMB processes absolute flow rates are less important as their relative values. H_A and H_B denote the slope of the isotherm for the different species at the feed concentrations.

We use a direct sequential algorithm for the solution of the problem (6). The process is simulated until the cyclic steady state is reached. The constraints and the objective value are then evaluated and given back to a non-linear optimizer FFSQP (Zhou *et al.*, 1997).

3.2 Optimizing the distribution of the columns

For this purpose, we first introduce the VARI*COL* process developed by (Ludemann-Hombourger and Nicoud, 2000), where in contrast to the SMB process the ports are shifted asynchronously (see figure 3).

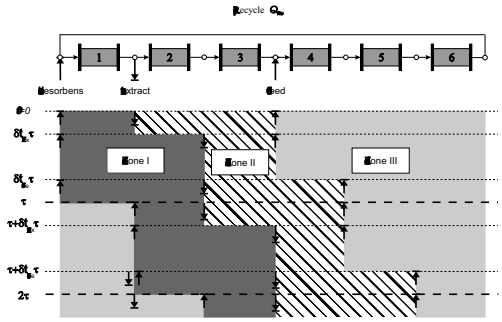

Figure 3. 3-zone VARI*COL* process for glucose isomerization

The numbers of columns per zone change during the period. E. g. the number of columns in zone I is one during a quarter of the period and two for the rest of the period. This corresponds to an average length of

$$\overline{N}_1 = 1 \cdot \frac{1}{4} + 2 \cdot \frac{3}{4} = \frac{7}{4} \tag{8}$$

for zone I. Analogously, $\overline{N}_2 = \frac{6}{4}, \overline{N}_3 = \frac{11}{4}$ result. While the SMB process is described by an integer column distribution, the asynchronous VARI*COL* process corresponds to a *real* number of columns in each zone.

Thus, the VARI*COL* process is more flexible than the SMB process. From the optimization point of view, the zone lengths $\overline{N}_i$ can be simply integrated as further degrees of freedom in the NLP formulated in equation 6 (Toumi *et al.*, 2002*a*; Toumi *et al.*, 2002*b*).

In table 1 (second column), the optimal operating point for the VARI*COL* process is listed. In this study, the objective was to minimize the desorbens consumption for an extract purity of 70%. The optimal distribution obtained in the VARI*COL* case is $[0.99, 1.6, 3.4]$, i. e. the zone III should be chosen much larger than the zones I and II. In comparison to the optimal operating point of the SMB-process with the configuration [2,2,2] (first column), **40%** less desorbens consumption is reached. If the optimal VARI*COL* distribution is rounded to the *next* integer distribution, an SMB process with section lengths $[1, 2, 3]$ results. By this distribution, the desorbens consumption can

Table 1. Optimization of the distribution of the columns among the zones for Pur_{Ex}=70%

	SMB I	VARICOL	SMB II
rel. Q_{De} [%]	100.00	60.00	74.00
Pur_{Ex} [%]	70.00	70.00	70.00
Q_{Fe} [ml/min]	1.30	1.30	1.30
Q_{De} [ml/min]	3.70	2.26	2.31
Q_{Re} [ml/min]	17.60	15.51	15.25
τ [min]	10.40	12.12	12.26
Q_{Ex}	5.00	3.56	3.61
β_i	**[1.04,0.79,1.63]**	**[0.97,0.84,1.50]**	**[0.97,0.82,1.53]**
N_i	**[2,2,2]**	**[0.99,1.61,3.4]**	**[1,2,3]**

be reduced by **26%** in comparison to the original [2,2,2]-configuration.

This approach led us to a considerably better SMB distribution after only two optimization runs. Even if mathematically this approach can lead to a sub-optimal distribution, the search space can be reduced when the optimal distribution of the VARI*COL* process is known. It can be assumed, that the optimal SMB distribution is one of the edges in the neighborhood of the optimal VARI*COL* distribution. In addition, since the VARI*COL* process includes all possible SMB-configurations, we know now a-priori the maximal attainable performance in term of desorbens consumption for a given plant (i. e. a given number of columns).

4. CONTROL STRATEGY

Automatic control of the SMB process was applied to the separation of aromatic hydrocarbons where on-line Raman spectroscopy can be utilized to measure the concentration of the compound at the outlet of the chromatographic columns (Marteau *et al.*, 1994). However this as well as the geometric nonlinear control concept described in (Kloppenburg and Gilles, 1999) are mainly based on a model for the corresponding true moving bed (TMB) process, where the cyclic port switching is neglected. In the case of SMB processes with a few number of columns (e. g. less than 8), the TMB process does not approximate the SMB process accurately, so that the applicability of this control scheme to plants with few columns seems problematic.

(Natarajan and Lee, 2000) investigated the application of the repetitive model predictive control (RMPC) technique on SMB processes. RMPC is a model-based control technique developed by incorporating the basic concept from repetitive control into model predictive control technique. In order to apply this technique, the switching period of the process is assumed to be constant. This is limiting, since the switching time can be used as another manipulated variable to control the process. The rigorous model was then linearised along the optimal trajectory. Afterwards it was reduced to a low dimensional linear model, based on which a linear MPC controller scheme was developed.

(Schramm *et al.*, 2001) presented a model-based control approach for direct control of the product purities of SMB processes. Based on wave theory, they derived relationships between the front movements and the flow rates for the equivalent TMB process. Based on these relationships, they developed a simple concept with two standard PI controllers. This concept is very easy to implement. However, similar relationships are difficult to determine analytically in the case of nonlinear reactive chromatography.

(Klatt *et al.*, 2001) proposed a two-layer control architecture where the optimal operating trajectory is calculated off-line by dynamic optimization based on a rigorous process model. The parameters are adapted based on online measurements. The low-level control task is to keep the process on the optimal trajectory despite disturbances and plant/model mismatch. The controller are based upon identified models gained from simulation data of the rigorous process model along the optimal trajectory. For the linear (linear adsorption isotherm) case, linear ARX models are sufficient (Klatt *et al.*, 2001), whereas in the nonlinear case neural networks (NN) were applied successfully (Wang *et al.*, 2002). A disadvantage of this two-layer concept is that the stabilized front positions may not guarantee the product purities if plant/model mismatch occurs.

4.1 Formulation of the control problem

The essence of model predictive control (MPC) is to optimize, over the future values of the inputs, the future process behavior. The future process behavior is analyzed with a process model over a finite time interval which is called the *prediction horizon*. The first input of the optimal input sequence, which spans the *control horizon*, is applied to the plant and the problem is solved again at the next time interval using updated process measurements and a shifted horizon. In

the framework of MPC control, it is simple to include hard constraints on the state and input variables. Furthermore, the process behavior can be predicted using a linear model or a nonlinear model. However, in the latter case, getting the *exact global* solution of a non-convex optimization problem requires formidable efforts and can not be achieved within a fixed sampling time. Even with state-of-the-art optimization algorithms, this seems to be practically impossible. Therefore we modify the nonlinear model predictive algorithm: we are calculating a *suboptimal* but *feasible* solution which can be applied in real-time.

We propose to solve the following optimal control problem over the finite *control horizon* H_r:

$$\begin{aligned}
\min_{[\beta_k,\ldots,\beta_{k+H_r}]} \quad & J = \sum_{j=k}^{k+H_p} \left(Cost(i) + \mathbf{\Delta}\beta_j^T \mathbf{R}_j \mathbf{\Delta}\beta_j \right) \\
\text{s.t.} \quad & \begin{cases} \dot{\mathbf{x}}_j = \mathbf{f}(\mathbf{x}_j, \beta_j), \\ \mathbf{x_{j+1,0}} = \mathbf{P}\mathbf{x}_j(\tau(j)), \\ j = k, \cdots, k+H_p. \end{cases} \\
& Pur_{Ex,H_r} + \Delta Pur_{Ex} \geq Pur_{Ex,min}, \\
& Pur_{Ex,H_p} + \Delta Pur_{Ex} \geq Pur_{Ex,min}, \\
& Q_{I,j} \leq Q_{max}, \\
& \mathbf{g}(\beta_j) \geq \mathbf{0}, j = k, \cdots, k+H_p.
\end{aligned} \tag{9}$$

We discretize the *prediction horizon* in cycles, where a cycle is a switching time $\tau(k)$ multiplied by the total number of columns. Eq. 9 consists of a dynamic optimization problem including the transient behavior of the process. The objective function J is a sum of stages costs (e. g. desorbens consumption) and a regularizing term added in order to smooth the input sequence avoiding high fluctuations in the input sequence from cycle to cycle. The first equality constraint represents the plant model evaluated over the finite prediction horizon H_p. The switching dynamic is introduced vice the permutation matrix $\mathbf{P}$. Since the maximal attainable pressure drop has not to be exceeded, constraints are imposed on the flow rates in zone I. Further inequality constraints $\mathbf{g}(\beta_j)$ are added in order to avoid negative flow rates during optimization.

The control objective is introduced by the purity constraint over the control Horizon H_r which is additionally corrected with a bias term ΔPur_{Ex} resulting from the difference between the last simulated and the last measured process output. A second purity constraint over the whole prediction horizon acts as a terminal (stability) constraint forcing the process to converge towards the optimal cyclic steady state. It has to be pointed out that the control goal (i. e. to fulfill the extract purity) is introduced as a constraint. We are using a feasible path SQP algorithm for optimization (Zhou *et al.*, 1997), which generates a feasible point before it starts to minimize the objective function.

4.2 Parameter Estimation

We assume that the concentration profiles in the recycling line are measured during a cycle. Since this measurement point is fixed in the closed-loop arrangement, the sampled signal includes information of all three zones. At every cycle and during the start-up phase an on-line estimation of the actual model parameters is started. The quadratic cost functional $J(\mathbf{p})$

$$J(\mathbf{p}) = \sum_{i=1}^{nsp} \left(\int_0^{N_{col,ges}} (c_{i,meas}(t) - c_{i,Re}(t))^2 \, dt \right) \tag{10}$$

is minimized with respect to the parameters $\mathbf{p}$. For this purpose, the Least-Squares solver E04UNF from the NAG-library is used. A by-product of the parameter estimation is the actual value $\mathbf{x}_0(k)$ which is given back to the NMPC controller.

4.3 Simulation study

Figure 4 shows a simulation scenario where the desired extract purity was set to 70% at the beginning of the experiment. At cycle 60, the desired extract purity is then changed to 60%. At cycle 120, the extract purity is increased to 65%. A fast response of the controller in both directions can be seen. Compared to the uncontrolled case, the control concept can keep the desired product purity rejecting a disturbance in the enzyme activity. The evolution of the optimizer-iterations are plotted as dashed-lines and shows that a feasible solution can be found rapidly and that the concept can be realized in real time. In this example the *control horizon* was set to 2 cycles and the *prediction horizon* to 10 cycles. A diagonal matrix $R_j = 0.02 I_{(3,3)}$ was chosen for regularization.

We assumed an exponential drift in the enzyme activity which corresponds to a drift in the reaction rate. Figure 5 shows the result of the parameter estimation. A good fitting was achieved and the estimated parameter follows the drift of the real parameter adequately.

5. CONCLUSIONS AND FUTURE WORK

The optimal operating point, as well as the distribution of the columns over the zones, of a RSMB plant were determined using a model-based mathematical approach. A nonlinear model predictive controller is presented which maintains the product purity while using a minimal additional amount of desorbens.

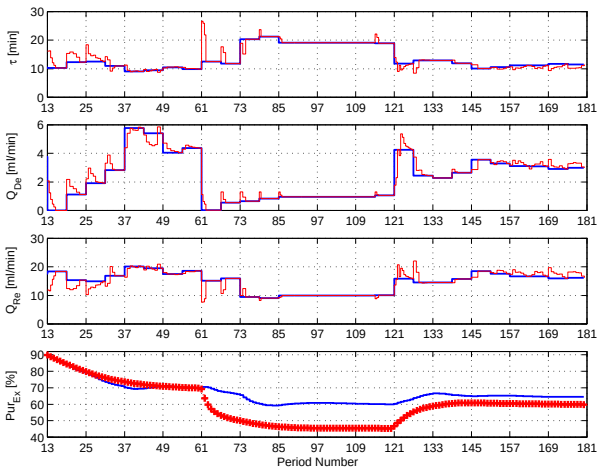

Figure 4. Control scenario: $H_r = 2, H_p = 10$

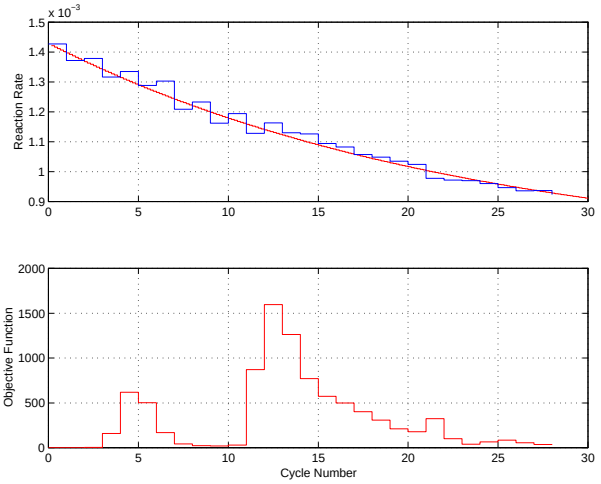

Figure 5. Estimation of the reaction rate

VARI*COL* is still **14%** better in term of desorbens consumption than the best SMB distribution [1,2,3] (see table 1). Therefore, it becomes more attractive to use the VARI*COL* concept, where in fact no additional investments are needed (despite programming the logical control system).

Appendix A. SYSTEM PARAMETERS

L=60.0 [cm]	k_i=[1.46E-07,1.33E-07]
D=2.6 [cm]	k_{ij}=[2.90E-07,9.30E-08]
ϵ_p=0.01 [–]	X=0.1 [-]
ϵ_b=0.4 [–]	k_0=24
D_p=16.25 [mm]	k_m=1.43E-03
k_{eff}=8.88E-05	k_{eq}=1.079 8 [–]
r=1.0 [g/cm^3]	Q_f=1.3 [ml/min]
h=5.8E-3 [g/(cm s)]	c_f=0 0.3 [g/cm^3]
D_p=1.0E-3 [cm^2/s]	Pur_{Ex}=60.0% or 70.0%
ν_i=[+1,-1] [–]	N_i=[2 2 2] or [1 2 3]
H_i=[0.47,0.27]	Δp_{max}=50 [bar]

REFERENCES

Asif, M. and A. E. Abaseed (1998). Modeling of glucose isomerization in a fluidized bed immobilized enzyme bioreactor. *Bioresource Technology* **64**, 229–235.

Fricke, J. and H. Schmidt-Traub (2002). Design of chromatographic SMB-reactors. In: *Inter. Symp. Prep. and Ind. Chromatography and Allied Tech. (SPICA, Heidelberg)*.

Gu, T. (1995). *Mathematical Modelling and Scale Up of Liquid Chromatography*. Springer: New York.

Hashimoto, K., S. Adachi, H. Noujima and H. Maruyama (1983). Models for the separation of glucose/fructose mixture using a simulated moving-bed adsorber. *Journal of Chemical Engineering of Japan* **16**, 400–406.

Klatt, K.-U., G. Dünnebier, F. Hanisch and S. Engell (2001). Optimal operation and control of simulated moving bed chromatography: A model-based approach. In: *Invited Plenary Paper, Preprints Chemical Process Control, CPC*. Tucson, USA. pp. 266–282.

Kloppenburg, E. and E. D. Gilles (1999). Automatic control of the simulated moving bed process for c_8 aromatics seperation using asymptotacally exact input / output - linearization. *J. of Process Control* **9**, 41–50.

Ludemann-Hombourger, O. and R. M. Nicoud (2000). The "VARICOL" process: A new multicolumn continuous chromatographic process. *Sep. Sci. Techn.* **35**, 1829–1862.

Marteau, P., G. Hotier, N. Zanier-Szydlowski, A. Aoufi and F. Cansell (1994). Advanced control of c_8 aromatics separation process with real-time multiport on-line raman spectroscopy. *Process and Quality* **6**, 133–140.

Natarajan, S. and J.H. Lee (2000). Repetitive model predictive control applied to a simulated moving bed chromatography system. *Comp. Chem. Eng.* **24**, 1127–1133.

Schramm, H., S. Grüner, A. Kienle and E.D. Gilles (2001). Control of moving bed chromatographic processes. In: *Proceedings of the European Control Conference*.

Toumi, A., F. Hanisch and S. Engell (2002*a*). Optimal operation of continuous chromatographic processes: Mathematical optimization of the VARI*COL* process. *Ind. Eng. Chem. Res.* **41**, 4328–4337.

Toumi, A., S. Engell, O. Ludemann-Hombourger and R.M. Nicoud (2002*b*). The VARI*COL* process: Principle and optimal operation. *submitted for publication to Journal of Chromatography*.

Wang, C., S. Engell and F. Hanisch (2002). Neural network based identification and adaptive MPC control of SMB chromatography. In: *Paper Tu-M11-6, 15th IFAC World Congress, Barcelona, Spain*.

Zhou, J. L., A. L. Tits and C. T. Lawrence (1997). User's guide for FFSQP version 3.7. Technical report. University of Maryland.

COMBINATIONS OF MEASUREMENTS AS CONTROLLED VARIABLES: APPLICATION TO A PETLYUK DISTILLATION COLUMN.

V.Alstad, **S. Skogestad** [1]

Department of Chemical Engineering,
Norwegian University of Science and Technology, NTNU,
N-7491 Trondheim, Norway

Abstract: A new simple approach for selecting controlled variables, that give near-optimal operation with a constant set-point feedback structure in the presence of uncertainty, is presented. The controlled variables are linear combinations of a subset of the available measurements. A method for selecting the best sub-set and the required number of measurements is derived. The method is illustrated on a Petlyuk (Divided wall) distillation column.

Keywords: Self-optimizing control, Control structure design, Distillation columns

1. INTRODUCTION

Although not widely acknowledged, controlling the right variables is a key element in overcoming uncertainty in operation. Control systems often consist of several layers in a hierarchical structure, each operating on a different time scale. Typically, layers include scheduling (weeks), site-wide optimization (day), local optimization (hours), supervisory/predictive control (minutes) and regulatory control (seconds). The layers are interconnected through the controlled variables **c**. This paper focuses on the interaction between the local optimization layer and the feedback control layer, see Figure 1. The objective is to find good candidate controlled variables **c** with self-optimizing properties. Self-optimizing control follows the idea of Morari *et al.* (1980) where one want to find controlled variables that, when kept at constant set-points, operates near optimally under the influence of disturbances and implementation errors. The disturbances include both exogenous process disturbances and modeling errors. In typical plants the number of disturbances may be very large. In order to reduce the dimension of the problem, only slow varying disturbances that are economically important should be included in the analysis. Morari *et al.* (1980) propose to include disturbances that have a large effect on the objective, ($\frac{\partial J^{opt}}{\partial d_i}$), and let the remaining disturbances be handled by the regulatory layer

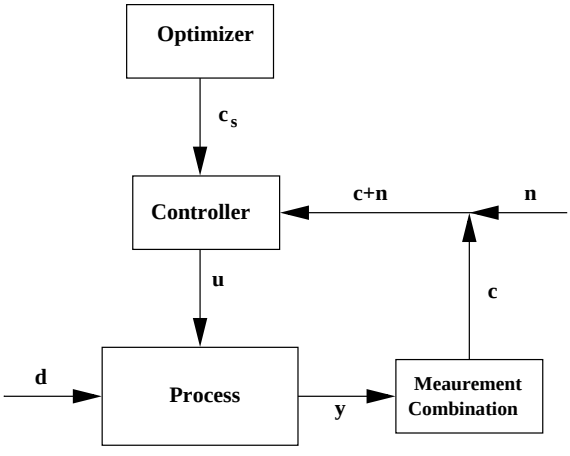

Fig. 1. Self-optimizing feedback control structure

[1] e-mail: skoge@chemeng.ntnu.no; phone: +47-7359-4154; fax: +47-7359-4080

In plant operation the basic goal is to optimize an economic measure of the operation, while satisfying equality and inequality constraints (such as product specifications, safety constraints, environmental regulations etc.). Since plant economics is primarily decided by steady-state behavior, only steady-state information is used in the rest of this paper. For a given disturbance ($\mathbf{d}$), optimal operation is defined as the solution to the following problem:

$$\min_{u_0} J_0(\mathbf{x_0}, \mathbf{u_0}, \mathbf{d}) \quad (1)$$

$$\mathbf{f}(\mathbf{x_0}, \mathbf{u_0}, \mathbf{d}) = 0 \quad (2)$$

$$\mathbf{g}(\mathbf{x_0}, \mathbf{u_0}, \mathbf{d}) \leq 0 \quad (3)$$

$$\mathbf{x_0} \in \mathcal{R}^{n_{x_0}}, u_0 \in \mathcal{R}^{n_{u_0}}, d \in \mathcal{R}^{n_d}$$

where $\mathbf{f}$ are the equality constraints, $\mathbf{g}$ the inequality constraints, $\mathbf{u_0}$ the free independent variables (inputs), $\mathbf{d}$ the disturbances and $\mathbf{x_0}$ the states. At the nominal optimum a subset ($\mathbf{g'}$) of the inequality constraints will be active and for small changes in the disturbance from the nominal point, it is assumed that the active set does not change. Thus, the reduced space optimization problem is:

$$\min_{u} J_0(\mathbf{x}, \mathbf{u}, \mathbf{d}) \quad (4)$$

$$\mathbf{f'}(\mathbf{x}, \mathbf{u}, \mathbf{d}) = 0 \quad (5)$$

where $\mathbf{f'} = [\mathbf{f}\ \mathbf{g'}]^T$, $\mathbf{x} = [\mathbf{x_0}\ \mathbf{u'}]^T$ where $\mathbf{u'} \in \mathbf{u_0}$ is the subset used to fulfill the active constraints ($\mathbf{g'} = 0$) and $\mathbf{u} \in \mathbf{u_0}$ denotes the remaining unconstrained reduced space degrees of freedom. By formally eliminating the states ($\mathbf{x}$) by using the equality constraints ($\mathbf{f'} = 0$), the remaining unconstrained problem, which is the focus in the rest of this paper, becomes:

$$\min_{\mathbf{u}} J(\mathbf{u}, \mathbf{d}) \quad (6)$$

where $\mathbf{u} \in \mathcal{R}^{n_u}$. The solution of the resulting problem in (4) may be categorized into two classes. Let $n_{f'} = dim(\mathbf{f'})$. If $n_u = n_{u_0} + (n_{x_0} - n_{f'}) = 0$, all degrees of freedom must be used to fulfill the constraints and implementation is usually simple by using the ideas of active constraint control (Maarleveld and Rijnsdorp, 1970). In case of $n_u > 0$, implementing the remaining n_u unconstrained degrees of freedom is not straight forward and this will be the focus in the rest of this paper.

.

Online information about the system behavior is available from the measurements in the plant:

$$\mathbf{y_0} = \mathbf{f}_{y_0}(\mathbf{u}, \mathbf{d}) \quad (7)$$

Based on the online information $\mathbf{y_0}$, the most obvious operational policy is to use some sort of optimizing controller with frequent model updates and re-optimization. A much simpler approach for practical applications, is to utilize the ideas of self-optimizing control.

Self-optimizing control (Skogestad, 2000) *is when an acceptable* ***loss*** *can be achieved using constant set-points* $\mathbf{c_s}$ *for the controlled variables* $\mathbf{c}$ *(without the need to re-optimize when disturbances occur).*

The loss is defined as the difference between the objective using the constant feedback structure and the true optimal objective

$$L = J(\mathbf{c_s} + \mathbf{n}, \mathbf{d}) - J_{opt}(\mathbf{d}) \quad (8)$$

where $\mathbf{n}$ is the implementation error (measurement and set-point error) in enforcing $\mathbf{c} = \mathbf{c_s}$. The central issue when searching for the self-optimizing control structure, is to decide how to best *implement* the optimal policy in the presence of uncertainty.

The optimal self-optimizing control structure may be formulated mathematically by:

$$\min_{\mathbf{h}} \int_{d \in \mathcal{D}} \int_{n \in \mathcal{N}} J(\mathbf{u}, \mathbf{d})\ \mathrm{d}n\ \mathrm{d}d \quad (9)$$

$$\mathbf{y} = \mathbf{f}_{\mathbf{y}}(\mathbf{u}, \mathbf{d}) \quad (10)$$

$$\mathbf{h}(\mathbf{y}) = \mathbf{c_s} + \mathbf{n} \quad (11)$$

where $\mathbf{y} \in \mathbf{y_0}$ and $\mathbf{u}$ is an implicit function of $\mathbf{h}$, $\mathbf{d}$, $\mathbf{c_s}$ and $\mathbf{n}$. The goal is to find the optimal function $\mathbf{h}$ interconnecting the measurements and the controlled variables. We assume in this paper that we use nominally optimal set-points, $\mathbf{c_s} = \mathbf{c_{opt}}(\mathbf{d}^*)$, but it is also possible to compute the "robust optimal set-points" by minimizing with respect to $\mathbf{c_s}$ in (9) (Govatsmark and Skogestad, 2002). In practice (9) may be solved by discretizing the disturbance and implementation error space and calculate some weighted average over all points. Clearly, this is a non-convex combinatorial optimization problem, that may be very difficult to solve in practice. A much simpler method for selecting the interconnecting structure $\mathbf{h}$ is needed.

2. PREVIOUS WORK ON SELECTION OF CONTROLLED VARIABLES

Skogestad *et al.* (2003) use a Taylor series expansion of the loss function around the nominal optimal point to develop two methods for selecting controlled variables, the "singular value rule" and the "exact local method". The exact local method, is based on the second order Taylor series expansion of the loss function $L = \frac{1}{2}\|\mathbf{z}\|_2^2$ with

$$\mathbf{z} = J_{uu}^{1/2}[(J_{uu}^{-1^*} J_{du}^* - G^{-1}G_d)(\mathbf{\Delta d}) + G^{-1}\mathbf{n}]$$

where $J_{uu}$and J_{ud} are the second derivatives of J and G and G_d are given by $\Delta\mathbf{c} = G\Delta\mathbf{u} + G_d\Delta\mathbf{d}$. By proper scaling and assuming that $\|[\mathbf{d}\ \mathbf{n}]^T\|_2 \leq 1$, the worst-case loss is:

$$L = \frac{\overline{\sigma}([M_d\ M_n])^2}{2} \quad (12)$$

$$M_d = J_{uu}^{1/2}(J_{uu}^{-1^*}J_{du}^* - G^{-1}G_d)W_d \quad (13)$$

$$M_n = J_{uu}^{1/2}G^{-1}W_n \quad (14)$$

where W_d and W_n are positive diagonal matrices representing the expected magnitudes of the disturbances and implementation errors respectively. This method require that, for each *candidate set* the singular value of the matrix M is calculated. The second method, the singular value rule, is based on scaling the candidate set of controlled variables, and select controlled variables that maximize the minimum singular value of the gain matrix G.

Mahajanam *et al.* (2001) propose a "short-cut" method to eliminate poor choices and to generate rank alternatives without solving the optimization problem. The method is based on scaling all candidate controlled variables so that they have similar effects on the steady-state profit.

3. PROPOSED METHOD FOR SELECTING CONTROLLED VARIABLES AS LINEAR COMBINATIONS OF THE MEASUREMENTS

We here consider the remaining unconstrained optimization problem in (6), and the objective is to find variables $\mathbf{c}$ to be kept at constant setpoints. In general, we have

$$\mathbf{c} = h(\mathbf{y}) \quad (15)$$

where $\mathbf{y} \in \mathbf{y_0}$ is the subset of all available measurements which we choose to make use of. Note that $\mathbf{y_0}$ generally also includes the input variables $\mathbf{u}$. Previous work (Skogestad, 2000) has mainly focused on using single measurements as controlled variables, ie. $\mathbf{c} = \mathbf{y}$. The generally nonlinear function $\mathbf{h}$ is free to choose, except that the controlled variables are assumed independent and that the number of controlled variables (c's) equals the number of remaining unconstrained degrees of freedom (u's). In this paper, we consider only linear combinations of the measurements

$$\Delta\mathbf{c} = H\Delta\mathbf{y} \quad (16)$$

where the matrix H is free to choose. Skogestad *et al.* (2003) use (12) to search for the optimal measurement combination (matrix H), taking into account both disturbances and implementation errors, but this is generally a very difficult problem. However, as shown below, it is actually trivial to find the optimal H for the case with *no implementation error* ($\mathbf{n} = 0$). We use the following insight: *With no implementation error, the constant set-point policy ($\mathbf{c} = \mathbf{c_s}$) is optimal if $\mathbf{c}_{\text{opt}}(\mathbf{d})$ is independent of d.* Of course, the optimal values of the individual measurements $\mathbf{y}$ depend on $\mathbf{d}$, which for a small disturbance change may be written

$$\Delta\mathbf{y_{opt}} = \mathbf{y_{opt}}(\mathbf{d}) - \mathbf{y_{opt}}(\mathbf{d}^*) = F(\mathbf{d} - \mathbf{d}^*) = F\Delta\mathbf{d} \quad (17)$$

where $F = \left(\frac{d\mathbf{y_{opt}}}{d\,\mathbf{d}}\right)^*$. For example, F may be obtained numerically by solving the optimization problem (4) for small changes in the disturbance, and from this obtaining $\mathbf{u_{opt}}(\mathbf{d})$ as well as $\mathbf{y_{opt}}(\mathbf{d})$. F must be understood as an constrained optimal linear mapping. Ganesh and Biegler (1987) give an efficient and rigorous strategy for finding the optimal sensitivity based on a reduced Hessian method. From (16) the corresponding change in the optimal value of $\mathbf{c}$ is

$$\Delta\mathbf{c_{opt}} = H\Delta\mathbf{y_{opt}} \quad (18)$$

Now require that

$$\Delta\mathbf{c_{opt}} = HF\Delta\mathbf{d} = 0 \quad (19)$$

This needs to be satisfied for any $\Delta\mathbf{d}$ so

$$HF = 0 \quad (20)$$

For this to hold, Hshould be in the left null space of F ($H \in \mathcal{N}(F^T)$). This requirement is always possible to fulfill, if there are enough measurements available in the plant. There are n_u unconstrained degrees of freedom (the length of vectors $\mathbf{u}$ and $\mathbf{c}$ are n_u), n_y independent measurements used when forming $\mathbf{c}$, and n_d independent disturbances. Then F is a $n_y \times n_d$ matrix and H a $n_u \times n_y$ matrix. The fundamental theorem of linear algebra (Strang, 1988) gives that $\mathcal{N}(F^T)$, the left null space of F has rank $n_y - r$, where $r = rank(F) = n_d$. Since $H \in \mathcal{N}(F^T)$ it follows that $rank(H) = n_y - n_d$ and by assuming that the number of controlled variables must be equal to the number of inputs, $rank(H) = n_u$.

$$n_y - n_d = n_u \Leftrightarrow n_y = n_u + n_d \quad (21)$$

so that the *minimum number of measurements needed, is equal to the number of inputs plus the number of disturbances.*

3.1 Comparison with the exact local method

The linearized models at the nominal point is

$$\Delta\mathbf{y} = G^y\Delta\mathbf{u} + G_d^y\Delta\mathbf{d} \quad (22)$$

where $G^y = (\partial \mathbf{f_y}/\partial \mathbf{u}^T)^*$ and $G_d^y = (\partial \mathbf{f_y}/\partial \mathbf{d}^T)^*$. For a disturbance change we have (Skogestad *et al.*, 2003).

$$\mathbf{u_{opt}}(\mathbf{d}) - \mathbf{u_{opt}}(\mathbf{d}^*) = -J_{uu}^{*-1} J_{du}^* (\mathbf{d} - \mathbf{d}^* \quad (23)$$

Thus

$$\Delta \mathbf{y_{opt}} = \left[-G^y J_{uu}^{*-1} J_{du}^* + G_d^y \right] (\mathbf{d} - \mathbf{d}^*) \quad (24)$$

By using $G = HG_y$ and $G_d = HG_d^y$ in (13), setting $M_d = 0$, assuming no implementation error and rearranging we get $GJ_{uu}^{-1^*} J_{du}^* - G_d = 0$ and inserting into (24) we re-derive $\Delta \mathbf{c_{opt}} = H \Delta \mathbf{y_{opt}} = 0$. Note that

$$F = -G^y J_{uu}^{-1} J_{du} + G_d^y \quad (25)$$

4. A TWO-STEP METHOD CONSIDERING DISTURBANCES AND IMPLEMENTATION ERRORS.

From the analysis in Section 3, see (12) to (14), it is evident that even if $M_d = 0$ the loss may still be large, since M_n is non-zero due to the implementation error. As stated, the selection matrix H is not unique, since there is freedom in selecting another sub-set of measurements. This may be utilized in order to reduce the effect of the implementation error, while still ensuring $M_d = 0$. The selection of a sub-set $\mathbf{y}$ of the available measurements $\mathbf{y_0}$, should reflect two goals. First, since the feedback structure must correct for disturbances in order to keep the plant optimal, the disturbances must be observable in the process (high gain in G_d). Second, in order to reduce the implementation error, it is evident from (14) that G^{-1} should be small in all directions (e.g $\underline{\sigma}(G)$ should be large). Based on these observations, it is proposed here to select measurements sequentially, that maximize the minimum singular value of the scaled augmented plant $\mathbf{\Delta y'} = \tilde{G}^y \tilde{\mathbf{\Delta u}} = [G^{y\prime} \; G_d^{y\prime}][\mathbf{\Delta u'} \; \mathbf{\Delta d'}]^T$.

The reason for using $\underline{\sigma}(\tilde{G}^y)$ rather than $\underline{\sigma}(G)$ and $\underline{\sigma}(G_d)$, is that H is not known *a priori*. To justify this, the following applies:

$$\underline{\sigma}(H)\underline{\sigma}(\tilde{G}^y) \leq \underline{\sigma}(H\tilde{G}^y) = \underline{\sigma}([G \; G_d]) \leq min(\underline{\sigma}(G), \underline{\sigma}(G_d)) \quad (26)$$

(Skogestad and Postlethwaite, 1996), (Horn and Johnson, 1991), where it is alway possible to select $\underline{\sigma}(H) = 1$. Thus, $\underline{\sigma}(\tilde{G}^y)$ provide a lower bound on $\underline{\sigma}(G)$ and $\underline{\sigma}(G_d)$. In addition, if $\underline{\sigma}(\tilde{G}^y)$ is nonzero this guarantees that F has full rank n_d, see (19), which is required to ensure that $F \neq 0$ for all $\mathbf{d}$. This follows from (25) since G^y and G_d^y has full rank n_u and n_d respectively, and $\tilde{G}^y$ has full rank $n_u + n_d$.

The proposed method of selecting controlled variables as linear combinations of the measurements is summarized in Section 4.1.

4.1 Details of procedure

Assume that $n_{y_0} \geq n_u + n_d$ and the nominal optimal point is $\mathbf{u_{opt}}(\mathbf{d}^*)$

(1) **Linearization**. Linearize the process model around the nominal optimal point. This give G^{y_0} and $G_d^{y_0}$ for all measurements $\mathbf{y_0}$.

(2) **Scaling** Scale each measurement $y_{0,i}$ with its corresponding implementation error ($|n_{y_{0,i}}|$), each input u_j with its corresponding allowable range ($\Delta u_{j,max}$) and each disturbance d_k by its corresponding expected disturbance. This give the scaling matrices $W_{n_{y_0}} = diag(|n_{0,i}|)$, $W_u = diag(\Delta u_{j,max})$ and $W_d = diag(|d_k|)$

(3) **Selection of measurements**.

(a) **Augmented process model**. Calculate the scaled process model $\mathbf{\Delta y_0'} = G^{y_0\prime}\mathbf{\Delta u'} + G_d^{y_0\prime}\mathbf{\Delta d'} = W_{n_{y_0}}^{-1} G^{y_0} W_u \mathbf{\Delta u'} + W_{n_{y_0}}^{-1} G_d^{y_0} W_d \mathbf{\Delta d'}$ and obtain a new process matrix

$$\mathbf{\Delta y_0'} = \tilde{G}^{y_0}\tilde{\mathbf{\Delta u}} = [G^{y_0\prime} \; G_d^{y_0\prime}][\mathbf{\Delta u'} \; \mathbf{\Delta d'}]^T$$

(b) **Selection of the first measurement**. Calculate the row norm $\|\tilde{G}_i^{\,y_0}\|_2$ for all rows i and sort by decreasing row norm. Select the row with highest norm and add the corresponding row of the process matrix to a selection process matrix $\tilde{G}_1^y = max_i \|\tilde{G}_i^{y_0}\|_2$

(c) **Selection of the additional measurements**. Until $n_y = n_u + n_d$ add measurements to the selection process matrix one-by-one

$$\tilde{G}_{j+1,i}^y = \begin{bmatrix} \tilde{G}_j^y \\ \tilde{G}_i^{y_0} \end{bmatrix}$$

for all i and calculate the minimum singular value for all the combinations. Select the new measurement which has the highest minimum singular value and add to the selection process matrix.

(4) **Null space of F and selection of controlled variables**.

(a) Obtain F, for example, numerically from the non-linear equations, $F = \left(\frac{d\mathbf{y_{opt}}}{d\,\mathbf{d}}\right)^*$, or from (24).

(b) Calculate the null space $\mathcal{N}(F^T)$.

(c) Select H such that $H \in \mathcal{N}(F^T)$ and the rows of H form a orthonormal basis. This ensure that $\Delta \mathbf{c_{opt}} = H \Delta \mathbf{y_{opt}} = 0$

5. EXAMPLE: "PETLYUK" (DIVIDING WALL) DISTILLATION COLUMN

The thermally integrated divided wall ("Petlyuk") arrangement has several advantages compared to the traditional arrangements. Smith and Triantafyllou (1992) report typical savings in the order of 30% in *both* energy and capital costs compared to traditional arrangements with two columns in series. The Petlyuk column shown

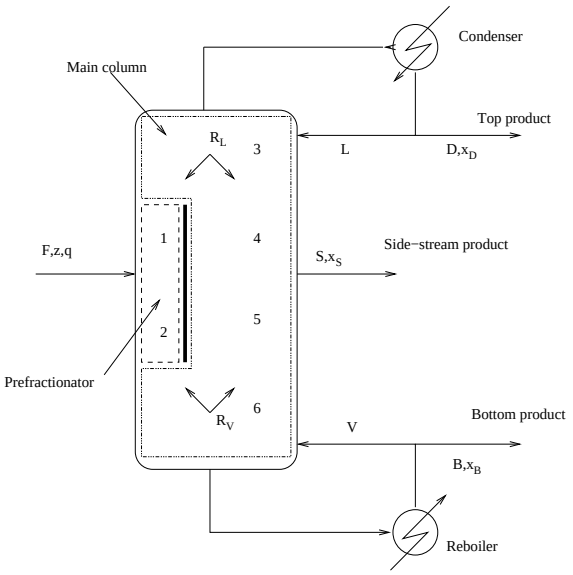

Fig. 2. The Petlyuk Distillation column implemented in a single column shell.

in Figure 2 has at steady state five degrees of freedom, which may be selected as the following inputs $u = [V\ L\ S\ R_l\ R_v]^T$ (boil-up, reflux, mid product side-stream flow, liquid split and vapor split). The product quality constraints correspond to the top purity ($x_{D,A}$), the bottom purity ($x_{B,C}$), the side-stream purity ($x_{S,B}$). The main contribution to the operational cost is the reboiler vapor flow, so the plant objective is to minimize:

$$J(x, u, d) = V \tag{27}$$

The model of the column is based on the assumption of constant relative volalities ($\alpha = [9\ 3\ 1]^T$), constant pressure, equilibrium on all stages, total condenser and constant molar flows. The tray temperatures are calculated based on the compositions using Antoine's equation.

Nominal data: Feed flow $F^* = 1$, liquid fraction $q^* = 0.477$, feed compositions $z_A^* = z_B^* = z_C^* = 1/3$ and product inequality constraints $[x_{D,A}\ x_{S,B}\ x_{B,C}]^T \geq 0.97$. The nominal optimal inputs are $u_{opt}(d*) = [L^*\ V^*\ S^*\ R_l^*\ R_v^*]^T = [0.718\ 0.5810\ 0.3227\ 0.3792\ 0.5123]^T$. Only economically important disturbances should be included in the analysis, which correspond to the disturbances z_A and q, the composition of component A the liquid fraction in the feed respectively, where $\mathbf{d} = [z_A\ q]^T = [z_A^* \pm 0.1\ q^* \pm 0.1]^T$. Since only intensive measurements are considered, feed flow rate is neglected as a disturbance. The implementation error is assumed to be $|\mathbf{n}| = [|n_{T_{j,i}}|\ |n_{R_l}|\ |n_{R_v}|]^T = [0.4\ 0.05\ 0.05]^T$.

For the disturbance range considered here, the optimally active constraints that need to be controlled, are the product compositions for the top, bottom and side-stream ($[x_{D,A}\ x_{S,B}\ x_{B,C}]^T = 0.97$), removing three degrees of freedom (L, V and S).

Based on the observation that the objective function has one "strong" and one "weak" direction, Halvorsen and Skogestad (1999) stated that in order to track the optimal trajectory only one of the remaining degrees of freedom need to be adjusted, so they propose to fix R_v. This is also reasonable from a practical point of view, due to the practical difficulties of implementing the vapor split. For the remaining degrees of freedom, only temperature measurements or combinations thereof are considered.

Halvorsen and Skogestad (1999) studied several candidate controlled variables for good self-optimizing properties, where the most promising control structure was to control DT_S, a measure of the temperature profile symmetry across the dividing wall, and R_v. DT_S is defined as $DT_S = \sum T_{1,i} - T_{4,i} - \sum T_{2,i} - T_{5,i}$, where $T_{j,i}$ is the temperature of tray i in section j.

For the remaining degree of freedom, only temperature measurements or combinations thereof are considered. In addition to the structure (R_v,DT_S) proposed by Halvorsen and Skogestad (1999), several other structures are considered; (R_v, $T_{1,7}$), (R_v, $T_{1,2}$) and the open loop structure (R_v, R_l).

In addition, two structures based on the methods proposed in this paper, are compared for self-optimizing properties.

1. (R_v,$c_{LC,3}$) with R_v fixed and the implementation error in R_v is added as a disturbance. Thus, $d = [z_A\ q\ n_{R_v}]$ and the number of measurements needed is $n_y = n_u + n_d = 1 + 3 = 4$. Maximizing the minimum singular value of the scaled augmented plant give that the subset $T_{5,5}, T_{2,2}, T_{4,2}, T_{2.1}$ of the temperature measurements should be combined. Selecting $\mathbf{c} = H\mathbf{y}$ such that H is in the left null space of F, result in $c_{LC,3} = -0.959T_{5,5} + 0.1969T_{2,2} + 0.00956T_{4,2} + 0.1770T_{6,3}$.

2. ($c_{LC,1}$,$c_{LC,2}$). This case was included in order to check if there is any additional economic advantage of using both degrees of freedom as inputs. Here we have two unconstrained degrees of freedom and the number of required measurements is

$n_y = n_u + n_d = 2 + 2 = 4$. The temperatures $T_{5,5}, T_{2,3}, T_{4,2}, T_{2,1}$ minimized the singular value of the augmented plant and the corresponding optimal measurement combinations are $c_{LC,1} = 0.222T_{5,5} - 0.7052T_{2,3} + 0.490T_{4,2} + 0.462T_{2,1}$ and $c_{LC,2} = -0.946T_{5,5} - 0.003T_{2,3} + 0.1592T_{4,2} + 0.2821T_{2,1}$. For the controlled variables that are linear combinations of the measurements it is assumed that the implementation error is $n_{c,i} = \|H_i W_y\|_2$.

In calculating the loss in Table 1, it is assumed that the combined implementation and disturbance vector is 2-norm bounded and that there is no implementation error in enforcing the active constraints. The average loss is calculated as a weighted sum of all combinations of the implementation and disturbance vector in which each disturbance and implementation error has a low, nominal and high value, with equal weighting.

Table 1. Loss for the different controlled variables in the Petlyuk Column case

c_1	c_2	Average loss (%)	Worst case loss (%)
$c_{LC,1}$	$c_{LC,2}$	0.01	0.02
R_v	$c_{LC,3}$	0.16	0.87
R_v	DT_s	2.40	11.9
R_v	R_l	18.0	123.0
R_v	$T_{1,7}$	22.7	118.4
R_v	$T_{5,2}$	infeasible	infeasible

As seen from Table 1, control structures ($c_{LC,1}$, $c_{LC,2}$) and (R_v, $c_{LC,3}$) track the optimal trajectory and give near-optimal operation. Controlling structure (R_v,DT_s) also give acceptable operation, while controlling the single temperatures ($T_{1,7}$ and $T_{5,2}$) give a very high loss or infeasible operation, which is expected since a change in the inflow composition is one of the disturbances. From Table 1 it is evident that fixing R_v gives only a small increase in the loss, but this is necessarily not true for all liquid fractions in the feed.

6. CONCLUDING REMARKS

Selecting the right variable to control is of great importance to overcome uncertainty in operation. A new method for selecting controlled variables as linear combinations of a subset of the available measurements has been proposed in addition to a method for selecting the subset of measurements. The idea is to find a linear combination of the measurements such that $\Delta c_{opt} = H\Delta y_{opt} = 0$ by using as many measurements as there are unconstrained inputs and disturbances. From a linear point of view, the proposed method guaranty perfect self-optimizing properties if we neglect implementation error. The proposed method has been illustrated on a simulated Petlyuk distillation column, which show that the proposed method give controlled variables with good self-optimizing properties.

7. ACKNOWLEDGMENTS

The financial support of The Research Council of Norway, ABB and Norsk Hydro is gratefully acknowledged.

REFERENCES

Ganesh, N. and L. T. Biegler (1987). A reduced hessian strategy for sensitivity analysis of optimal flowsheets. *AIChE Journal* pp. 282–296.

Govatsmark, M. S. and S. Skogestad (2002). Selection of controlled variables and robust setpoints. In: *Proc. of IFAC World Congress*. IFAC. Paper T-Mo-M-11-4.

Halvorsen, I.J. and S. Skogestad (1999). Optimal operation of petlyuk distillation: steady-state behavior. *J. Proc. Control* pp. 407–424.

Horn, R. A. and C. R. Johnson (1991). *Topics in Matrix Analysis*. Cambridge.

Maarleveld, A. and J.E. Rijnsdorp (1970). Constraint control on distillation columns. *Automatica* **6**(1), 51–58.

Mahajanam, R.V., A. Zheng and J.M. Douglas (2001). A shortcut method for controlled variable selection and its application to the butane alkylation process. *Industrial and Engineering Chemistry Research* **40**(14), 3208–3216.

Morari, M., G. Stephanopoulos and Y. Arkun (1980). Studies in the synthesis of control structures for chemical processes. part i: Formulation of the problem. process decomposition and the classification of the controller task. analysis of the optimizing control structures. *AIChE Journal* **26**(2), 220–232.

Skogestad, S. (2000). Plantwide control: the search for the self-optimizing control structure. *J. Proc. Control* **10**, 487–507.

Skogestad, S. and I. Postlethwaite (1996). Multivariable feedback control. *John Wiley & Sons*.

Skogestad, S., I. Halvorsen, J.C. Morud and V.Alstad (2003). Self-optimizing control: Local taylor series analysis. *Ind. Eng. Chem. Res.* Submitted.

Smith, R. and C. Triantafyllou (1992). The design and operation of fully thermally coupled distillaton columns. *Trans. IChemE.* pp. 118–132.

Strang, G. (1988). *Linear Algebra and its Applications*. 3 ed.. Harcourt Brace & Company.