15th Triennial World Congress, Barcelona, Spain

APPLICATION OF MODEL REDUCTION FOR MODEL PREDICTIVE CONTROL

Juergen Hahn[1,3], Uwe Kruger[2], and Thomas F. Edgar[1]

[1]Department of Chemical Engineering University of Texas at Austin Austin, TX 78712-1062 U.S.A.

[2]Intelligent Systems & Control Group Queen's University of Belfast Belfast BT9 5AH United Kingdom

[3]Lehrstuhl für Prozesstechnik RWTH Aachen D - 52056 Aachen Germany

Abstract: In this paper model reduction methods are used to obtain a nonlinear process model for designing a model predictive controller (MPC). The corresponding controller and its closed-loop response is then compared with controllers that are determined from the original model and a linearized version of this model. The reduced dimensional nonlinear MPC controller performs almost as well as the nonlinear MPC controller that is based on the original model and considerably better than the linear MPC controller. *Copyright © 2002 IFAC*

Keywords: Model-based control, Model reduction, Nonlinear systems, Controllability, Observability, System analysis

1. INTRODUCTION

Over the past two decades, model reduction methods have become increasingly popular. Such methods are designed to extract a reduced dimensional state-space model that adequately describes the input-output behavior of the process under study. In fact, the methods are data-driven and attempt to reduce the number of state variables by observing the loss in accuracy for describing the measurable process behavior.

A variety of different model reduction techniques exists for linear and nonlinear models and the aim of almost any reduction technique is to provide a reduced order model that can be used for controller design. However, little work has been done in evaluating the closed-loop performance that is achieved when the controller is based upon the reduced model and used for the original full-order process which can include disturbances and model mismatch.

Reduced dimensional models can be considered for determining an MPC controller, which presents an alternative to using the original process model. The main benefits are a decrease in the effort for computing adjustments of the manipulated variables and identification of the directions in state-space that are meaningful in a control-engineering context. While the former aspect is clearly related to the reduced size of the model, the latter benefit results because in the reduced set of state variables, unobservable and uncontrollable contributions to the process variation have been removed. As a test system, the simulation of two CSTRs that are operating in series is considered. This process exhibits a nonlinear relationship between the input, the state and the output variables and, therefore, is a challenging task for model reduction. An MPC controller is developed for each model and the performance of these controllers subjected to a set point change and an output disturbance provide the basis for comparison. It is demonstrated in this article that the performance of the MPC controller based on the original model is only marginally better than the one based on the reduced dimensional model. In contrast, the computational effort, determined by the number of floating point operations, can be considerably less for the latter one since computations for efficient MPC algorithms grow with the

cubic of the number of states of the model. Additionally, this controller performs significantly better than the linear MPC controller.

The paper is structured as follows. A brief review of existing model reduction methods is presented prior to the introduction of the MPC approach. Then, the example process that is considered for the case study is introduced. This is followed by a presentation of the application study.

2. REVIEW OF PREVIOUS WORK

Balanced model reduction for linear systems was first introduced by Moore (1981) in order to eliminate states that are close to being non-minimal and, therefore, contribute little to the input-output behavior of a system. Scherpen (1993) extended the balancing approach to a specific class of nonlinear systems by introducing energy functions and investigating conditions that guarantee existence of a balanced realization. However, the procedures are only applicable to control-affine systems, present computational difficulties, and in general a closed form solution cannot be obtained. The only numerical implementation of Scherpen's approach is given by Newman and Krishnaprasad (2000) who used a Monte-Carlo approach as an approximation for the computation of the energy functions and tested their algorithm on a pendulum with two states. Their algorithm computes approximations to the energy functions as well as a balancing transformation for the nonlinear system. However, after the coordinate transformation is applied, and even without reducing the model, the transformed system does not exhibit the same input-output behavior as the original system due to the approximations that were applied during the computational procedure.

Due to the problems encountered with nonlinear balancing procedures, several methods that perform a Galerkin projection, based upon a linear coordinate transformation, have been developed. Newman and Krishnaprasad (1998) compared models describing chemical vapor deposition that were reduced by principal component analysis (PCA) and balancing, where the balancing transformation was found from the linearized system. Pallaske (1987) investigated a procedure where the linear transformation is found from a covariance matrix that is computed from data collected along system trajectories. These trajectories represent the system behavior under a constant input, but starting from different initial conditions. Löffler and Marquardt (1991) extended this model reduction approach to models described by differential algebraic equation systems. Due to the complexity of the model they investigated the case where the trajectories start at the steady state operating point and are generated by step changes in the inputs to the system. Lee *et al.* (2000) computed the linear coordinate transformation by balancing a system that was generated using subspace identification. This identified system was generated from data collected along system trajectories. Lall *et al.* (1999, 2000) introduced the concept of empirical gramians, which are an extension to the gramians for linear systems. These empirical gramians can be computed for control-affine nonlinear systems and the computation procedure is based upon system trajectories that include changes in the system inputs as well as different initial conditions. Based upon the empirical gramians a linear coordinate transformation can be computed and the model reduced via a Galerkin projection. Hahn and Edgar (1999) showed that the procedure presented by Lall *et al*. (1999, 2000) is limited to control-affine systems and requires modifications when the steady state of the system is different from zero. Additionally, they investigated the extension of balanced residualization to nonlinear systems via the found coordinate transformation. A comprehensive review of model reduction techniques for linear and nonlinear systems was given by Marquardt (2001).

3. MODEL REDUCTION PROCEDURE

The reduction procedure used in this article is an extension of linear balanced truncation to nonlinear systems. The procedure itself can be split up into two steps, the computation of the covariance matrices and the procedure that balances the covariance matrices and reduces the system via a Galerkin projection.

3.1 Covariance matrices

For any stable nonlinear system

$$\dot{x}(t) = f(x(t),u(t)) \qquad (1)$$
$$y(t) = h(x(t))$$

the following sets can be defined for the covariance matrices:

$T^n = \{T_1, \ldots, T_r\ ;\ T_i \in \Re^{n\times n}\ ,\ T_i^T T_i = I\ ,\ i = 1, .., r\}$
$M = \{c_1, \ldots, c_s\ ;\ c_i \in \Re\ ,\ c_i > 0\ ,\ i = 1, \ldots, s\}$
$E^n = \{e_1, \ldots, e_n$ *; standard unit vectors in* $\Re^n\}$
r : number of matrices for excitation/perturbation directions
s : number of different excitation/perturbation sizes for each direction
n : number of inputs/states to the system

Definition 1: Controllability covariance matrix
Let T^p, E^p *and M be given sets as described above, where p is the number of inputs. The controllability covariance matrix is defined by*

$$W_C = \sum_{i=1}^{p}\sum_{l=1}^{r}\sum_{m=1}^{s}\frac{1}{rsc_m^2}\int_0^\infty \Phi^{ilm}(t)dt \qquad (2)$$

where $\Phi^{ilm}(t) \in \Re^{n\times n}$ *is given by* $\Phi^{ilm}(t) = (x^{ilm}(t) - x_{ss}^{ilm}(v(t)))(x^{ilm}(t) - x_{ss}^{ilm}(v(t)))^T$, *and* $x^{ilm}(t)$ *is the state of the nonlinear system corresponding to the input* $u(t) = c_m T_l e_i v(t) + u_{ss}(0)$.

The controllability covariance matrix is computed from data along system trajectories. The input to the system $u(t)$ is defined as above, where the c_m describes the input size, the $T_l e_I$ determines the input direction and $v(t)$ is the shape of the input. The $x_{ss}^{ilm}(v(t))$ represent the desired system trajectory and $u_{ss}(0)$ refers to the input at the original steady state. The input shape should be chosen in such a way that is consistent with typical input behavior of the plant. If impulse inputs are chosen for $v(t)$, then the controllability covariance matrix reduces to the empirical controllability gramian. Due to this, it can be shown that if the system under investigation is linear that the covariance matrix will reduce to the linear controllability gramian as well for the case of impulse inputs.

Definition 2: Observability covariance matrix
Let T^n, E^n and M be given sets as described above, where n is the number of states. The observability covariance matrix is defined by

$$W_O = \sum_{l=1}^{r}\sum_{m=1}^{s}\frac{1}{rsc_m^2}\int_0^{\infty} T_l \Psi^{lm}(t) T_l^T dt \qquad (3)$$

where $\Psi^{lm}(t) \in \Re^{n\times n}$ is given by $\Psi^{lm}_{ij}(t) = (y^{ilm}(t)-y_{ss}^{ilm})^T(y^{jlm}(t)-y_{ss}^{jlm})$, $y^{ilm}(t)$ is the output of the system corresponding to the initial condition $x(0) = c_m T_l e_i + x_{ss}$, and y_{ss}^{ilm} is the steady state that the system will reach after this perturbation.

By this definition the observability covariance matrix is equivalent to the empirical observability gramian if y_{ss}^{ilm} is equal to the measurement at the steady state operating point. Therefore, for this case it will reduce to the linear observability gramian if the system under investigation is linear and is therefore independent of the size of the perturbation c_m for linear systems. These covariance matrices have to be determined from simulation data, collected within a region where the process is to be controlled. The covariance matrices capture part of the nonlinear behavior within the region of operation and are more suitable for determining nonlinear reduced models than gramians of the linearized system.

3.2 Balancing and reduction algorithm

Once the covariance matrices for the system are computed they have to be made equal in the states that are both observable and controllable by a reversible linear coordinate transformation as is shown in equations (4) and (5). (Hahn and Edgar, 2001b)

$$TW_C T^T = \begin{bmatrix} \Sigma_1 & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad (4)$$

$$(T^{-1})^T W_O (T)^{-1} = \begin{bmatrix} \Sigma_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \Sigma_3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad (5)$$

This coordinate transformation can then be used within a Galerkin projection in order to truncate the states x_2 that contribute little to the input-output behavior of the system. (Hahn and Edgar, 2001b)

$$\begin{aligned} \dot{\bar{x}}_1(t) &= PTf(T^{-1}\bar{x}(t), u(t)) = \bar{f}(\bar{x}(t), u(t)) \\ \bar{x}_2(t) &= \bar{x}_{2,ss}(0) \\ y(t) &= h(T^{-1}\bar{x}(t)) = \bar{h}(\bar{x}(t)) \end{aligned} \qquad (6)$$

where

$$\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}, \; P = [I \quad 0], \; rank(P) = n_{reduced} \qquad (7)$$

4. NONLINEAR MODEL PREDICTIVE CONTROL ALGORITHM IMPLEMENTATION

The principles of MPC that are relevant for the implementation are briefly reviewed, since comparisons of the dynamic closed-loop behavior is the main focus of this article.

4.1 Principles of MPC

MPC utilizes a process model to forecast how future process behavior will be influenced by introducing a series of adjustments to the manipulated variables (MVs). Given that the process is sampled at a predefined interval, a series of values for the changes in the manipulated variables are determined which are used for the minimization the following cost function

$$J = \sum_{k=0}^{N-1}(y_k - y_{sp} + d)^T Q(y_k - y_{sp} + d) + \sum_{k=0}^{N_c-1}\Delta u_k^T R \Delta u_k \qquad (8)$$

where J represents the cost function, y_k and y_{sp} are the values of the measurements and their set points at time instance k, respectively. Furthermore, d accounts for the presence of output disturbances, Δu_k are the incremental values of the manipulated variables and N and N_c represent the design horizons for determining a succession of steps of the manipulated variables. $\boldsymbol{Q}$ and $\boldsymbol{R}$ are positive semi-definite weighting matrices that determine how aggressively the controller will attempt to remove any set point error of the controlled variables. The resulting optimization problem is solved using a nonlinear programming approach, since the underlying process model is nonlinear and a quadratic cost function is used.
It should be noted that the usual practice is to implement only the first step by adjusting the MVs accordingly and that the cost function is re-evaluated after each sampling interval.

4.2 Disturbance Estimation

In practice, the mismatch between the measured and the predicted values of the controlled variables (CVs), often referred to as model mismatch, is regarded as the influence of output disturbances. Several approaches have been proposed to estimate out-

put disturbances on the basis of the model mismatch, one of which is presented in equation (9)

$$d = \frac{1}{N_d} \sum_{k=1}^{N_d} y_{k,m} - y_{k,p} \tag{9}$$

where N_d represents a design horizon that is used to smooth out the influence of measurement noise. If the current time horizon is less than N_d steps, then all of the available data is used for the disturbance estimation. It is assumed, similar to an on-line application, that measurements from the real process are only available at discrete measurement points and that the values of the outputs stay constant between two measurements. It should be noted that the variables d and y_k are vectors for models with multiple measurements as will be the case in the case study in the following section.

4.3. MPC based upon reduced-dimensional models

The performance of the MPC controller is clearly dependent upon the degree with which the process model describes the true process behavior. As an example, if a poor process model predicts a transient response of the process that departs considerably from the true response, an inadequate closed-loop response is expected.

In this article, reduced dimensional process models are utilized within a MPC framework. The main benefit of using reduced dimensional models is that a reduction of the computational effort can be achieved. Furthermore, reduced dimensional process models are designed to omit only insignificant contributions from individual states. Hence, it is expected that the performance of controllers using such models departs only marginally from controllers that take advantage of an accurate process model. In order to use different models for the controller design the computation integrates the system model over the prediction horizon by calling a different file that contains the new model. It does not matter if the model that the optimization routine tries to invert is of the form of equation (1), a linearized version of equation (1) resulting in linear MPC on a nonlinear model or a reduced model of the form of equation (6). This makes performance tests easy to evaluate, because no additional parameters need to be adapted for a meaningful comparison.

5. APPLICATION STUDY

In this section, the performance of the MPC controller based upon a model derived from first principles, two reduced dimensional models and a model based upon a linear approximation are compared. The example study relates to the simulation of two serially operating CSTRs. At first, the open loop behavior of the example process is compared with each of the considered models. This is to demonstrate the nonlinear character of this process and to justify the use of a nonlinear model. This is then followed by contrasting the performance achieved by the different MPC controllers.

5.1 Description of the process

The example is based upon the model of two CSTRs operating in series with one reaction A->B that has been used extensively by other researchers for nonlinear controller design (Henson and Seborg, 1997). The original model consisted of four differential equations representing the energy and component balances for each reactor, where the coolant flow rate is the only input and the temperature measurement the only output of the model. The model has been augmented with volume balances for each reactor and one additional measurement (volume in the second reactor) and one extra input (valve position at the outflow of the second reactor). This results in a model with six states, two inputs and two outputs. For a more detailed description of the model see Henson and Seborg (1997) and Hahn and Edgar (1999). For this model a nonlinear controller was found to offer much improved performance over a linear controller due to the nonlinearity of the model (Henson and Seborg, 1997). In fact for step changes of more than 8% in the valve position at the outlet of the second reactor, the system will move to a vastly different steady state corresponding to lower conversion. For the application study, the sampling time is 6 seconds and the control horizon is 12 seconds, i.e. a control horizon of two control moves for each optimization step. Furthermore, the prediction horizon is selected to be 36 seconds, equaling six time steps into the future. It should be noted that the underlying process model is a continuous model. Thus, each optimization step requires integration of the model over the prediction horizon in order to predict the process response. While this is computationally expensive and may lead to difficulties for on-line applications, we believe it gives a meaningful comparison of the controller performance for different types of models. In order to reduce the computational burden, each optimization step is initiated by incorporating the previously obtained solution for that step. For estimating the impact of unmeasured disturbances, the application of equation (9) is considered with a design horizon of $N = 10$ (1 minute).

5.2 Open-loop behavior

Based upon earlier investigations (Hahn and Edgar, 2001a) the process becomes highly nonlinear for excitations around 7-8% for step changes in the manipulated variables for the system or when some of the states of the system move away from their steady state value by more than 2-3%. The open-loop responses of the full-order nonlinear system with six

states, a linearized system derived from linearization at the original operating point, and two reduced order systems each of which contains four states were computed. The only difference between the two reduced models is that one was reduced based upon covariance matrices computed for step inputs and the other one from covariance matrices computed for impulse inputs. In order to achieve a meaningful comparison, the excitation sizes for the impulse inputs were chosen such that they result in a similar degree of nonlinearity as for step changes of 7%. Figures 1 and 2 show the changes of the volume and temperature for a 7% change in the valve position at the outlet of the second reactor.

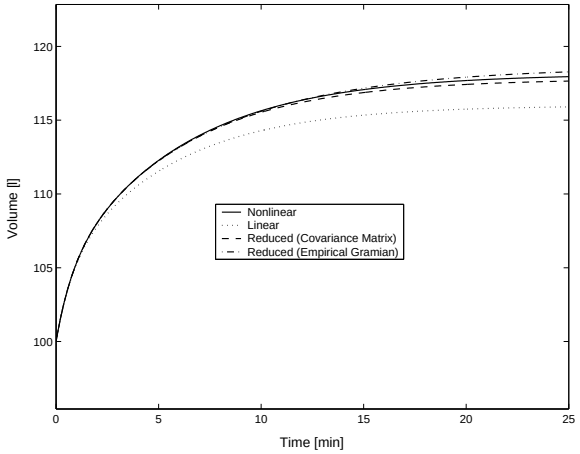

Fig. 1. Response of the volume to a 7% change in the valve position

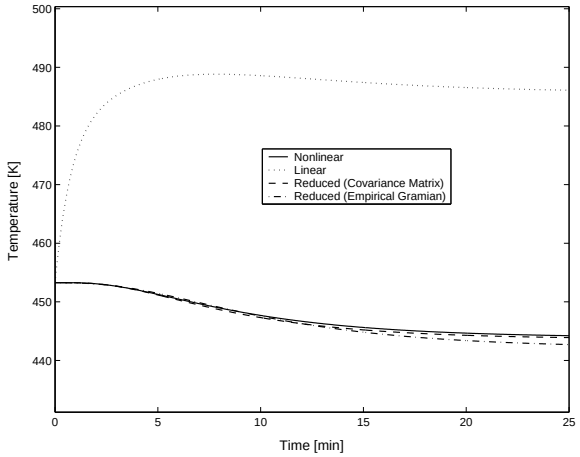

Fig. 2. Response of the temperature to a 7% change in the valve position

It can be concluded that all four models correctly predict the dynamic behavior of the volume. However, the change in temperature is not predicted accurately by the linear model due to the strong nonlinearities of the process within this operating region. In all cases that were evaluated both reduced nonlinear models provided a closer approximation to the system behavior than the linear model. Additionally, the reduced model that was based upon covariance matrices computed for step changes performed slightly better in the open-loop comparison than the one computed for impulse inputs. This is due to the fact that the models are evaluated for step input changes in this case study and the reduced model should reflect the behavior under study to obtain the best possible result.

5.3 Closed-loop behavior

The main reason for model reduction is to develop smaller models for on-line control applications. Therefore, the reduced models described in the previous subsection have been used within the model predictive control scheme described in section 4 for set-point changes and disturbance rejection. Additionally, in the case of the reduced models and for the linear model, model mismatch also exists between the model used to derive the controller and the real process.

Different controllers based upon the four models are used to control the process for a set point change as well as disturbance rejection and the results compared. The set point change occurs immediately, where the process is at its steady state and the set point is changed to *110 l* for the volume and *445 K* for the temperature of the second reactor. After reaching steady state 10 minutes into the simulation, the system is subjected to an output disturbance of *5 K* in the temperature measurement of the second reactor. The dynamic responses to these conditions are shown in Figures 3 and 4. It can be concluded from the results that all of the models result in good performance for the set point change in the volume of the reactor. This is not surprising, since all four open–loop responses described the dynamics of the volume fairly well. However, major differences can be seen for the temperature. The two reduced models are essentially indistinguishable, staying very close to the behavior of the full-order system that does not exhibit model mismatch between the controller and the plant model. This contrasts with the behavior of the linear model that takes much longer to reach its new set point and exhibits overshoot and oscillations.

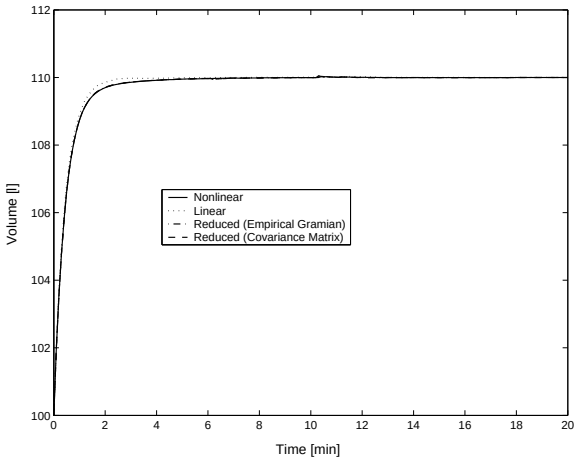

Fig. 3. Closed-loop responses of the volume for set point change

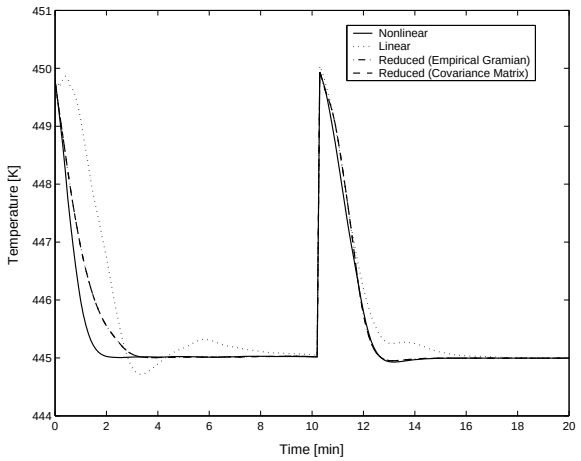

Fig. 4. Closed-loop responses of the temperature for set point change and disturbance rejection

Figure 5 compares the optimal input trajectories for the systems. Since all four models gave good results in describing the dynamic volume behavior they were virtually identical and only the normalized cooling rate is shown. Furthermore, since both reduced models resulted in identical closed-loop trajectories, they are lumped together in the dashed line in Figure 5. It can be seen that the input trajectory for the linear MPC controller is quite different from any of the nonlinear ones, especially in the beginning, or right after the disturbance affects the process.

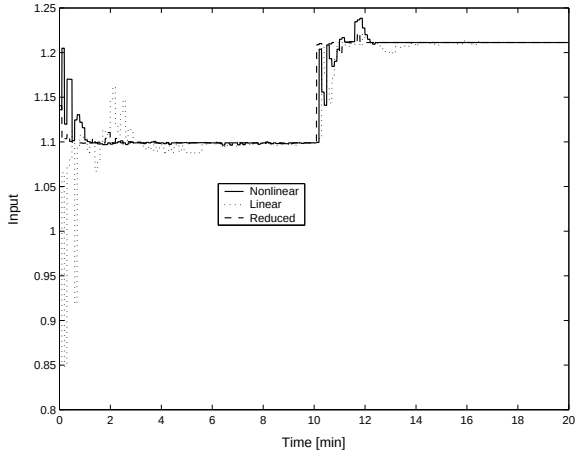

Fig. 5. Comparison of inputs to the system

6. CONCLUSIONS

While the main purpose of model reduction is to achieve smaller models that can be used for the design of controllers, surprisingly few results exist that actually compare the performance achieved with the controllers to the ones for the full-order system. This paper describes a nonlinear MPC scheme in which the closed-loop performance of different controllers has been evaluated. It was found that nonlinear controllers based upon reduced models can result in significantly better performance than linear controllers for some models. This was illustrated with a detailed example that included open-loop as well as closed-loop studies of four different controllers.

REFERENCES

Hahn, J. and T.F. Edgar (1999). An improved method for nonlinear model reduction using balancing of empirical gramians. submitted to *Comp. Chem. Eng*.

Hahn, J. and T.F. Edgar (2001a). A Gramian Based Approach to Nonlinearity Quantification and Model Classification. *Indust. & Eng. Chem. Research*., 40, pp. 5724-5731.

Hahn, J. and T.F. Edgar (2001b). An Balancing Approach to Minimal Realization and Model Reduction of Nonlinear Systems. In press *Indust. & Eng. Chem. Research*.

Henson, M.A. and D.E. Seborg (1997, Ed). *Nonlinear Process Control*; Prentice Hall, Englewood Cliffs, NJ.

Lall, S., J.E. Marsden. and S. Glavaski (1999). Empirical model reduction of controlled nonlinear systems, *14th IFAC World Congress*, Beijing, China.

Lall, S., J.E. Marsden and S. Glavaski (2000). A subspace approach to balanced truncation for model reduction of nonlinear control systems. submitted to *Intern. J. on Robust and Nonlinear Control*.

Lee, K.S., Y. Eom, J.W. Chung, J. Choi and D. Yang (2000). A control-relevant model reduction technique for nonlinear systems. *Comp. Chem. Eng*., 24, pp. 309-315.

Löffler, H.-P. and W. Marquardt (1991). Order reduction of nonlinear differential-algebraic process models. *Journal of Process Control*, pp. 32-40.

Marquardt, W. (2001). Nonlinear Model Reduction for Optimization Based Control of Transient Chemical Processes. *Proc. CPC VI*, pp. 30-60.

Moore, B.C. (1981). Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Automatic Control*, 26, pp. 17-32

Newman, A. and P.S. Krishnaprasad (2000). Computing balanced realizations for nonlinear systems. *14th Int. Symp. Math. Theory Networks and Systems*, Perpignan, France.

Newman, A. and P.S. Krishnaprasad. (1998). Nonlinear model reduction for RTCVD. *IEEE Proc.32nd Conference on Information Sciences and Systems*, Princeton, NJ.

Pallaske, U. (1987). Ein Verfahren zur Ordnungsreduktion mathematischer Prozessmodelle. *Chemie-Ingenieur-Technik*, pp. 604-605.

Scherpen, J.M.A. (1993). Balancing for nonlinear systems. *Systems and Control Letters*, 21**,** pp. 143-153.