Balanced Realization for State-Space Identification and Optimal Output Regulation

W. Fred Ramirez
Dept. of Chemical Engineering, University of Colorado, Boulder, CO 80309

Jan M. Maciejowski
Engineering Dept., University of Cambridge, Cambridge CB2 1PZ, England

A new input–output approach is presented for optimal regulatory control of systems based on balanced state realization. State-space identification is achieved using input–output data and a balanced realization algorithm. Based on the realized model, an optimal controller design is presented to regulate the system outputs at desired levels. The optimal control algorithm is model-predictive in that predictions are made based on measured disturbances. It also compensates for unmeasured disturbances through integral state feedback. This control concept is tested using a continuous stirred-tank chemical reactor model.

Introduction

The goal of this work is to develop an optimal design strategy for process control, using a state-space model, but based only on input–output data. We use the theory of balanced realizations to develop a state-space model and formulate an optimal model-predictive output regulator based on this model. This approach is particularly important for problems where accurate models are not available or too complex for control purposes. Another motivation is the development of reconfigurable control schemes that can cope with major plant failures that affect plant structure and sensor systems.

To meet these goals, the model identification and control design algorithms must work reliably with the minimum of operator intervention. This paramount objective accounts for our choice of approaches to the identification and control problems. In particular, we point out that conventional identification algorithms familiar in the literature do not seem to be suitable for reliable automatic application to complex multivariable processes.

In this article, we demonstrate that the approaches we have selected give a very powerful combination of automatic approximate modeling followed by control design. We apply these approaches to a continuous stirred-tank reactor.

Balanced Realization

System balancing was first introduced by Moore (1981), who showed its importance in model order reduction. System balancing is based upon the simultaneous diagonalization of the controllability and observability Gramians. The controllability Gramian is defined as

$$\boldsymbol{P} = \int_0^\infty e^{\boldsymbol{A}t}\boldsymbol{B}\boldsymbol{B}^T e^{\boldsymbol{A}t}\, dt, \tag{1}$$

where $\boldsymbol{A}$ is the state matrix and $\boldsymbol{B}$ the control matrix for a stable linear system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}. \tag{2}$$

The controllability Gramian satisfies the Liapunov equation

$$\boldsymbol{A}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{B}^T = 0. \tag{3}$$

The observability Gramian is

$$\boldsymbol{T} = \int_0^\infty e^{\boldsymbol{A}t}\boldsymbol{C}^T\boldsymbol{C}e^{\boldsymbol{A}t}\, dt, \tag{4}$$

where $\boldsymbol{C}$ is the observability matrix relating the outputs to the states

$$\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}. \tag{5}$$

The observability Gramian satisfies the following Liapunov equation

$$A^T T + TA + C^T C = 0. \tag{6}$$

A balanced realization is one such that the controllability Gramian, P, and the observability Gramian, T, both have a common diagonal solution to the Liapunov equations, 3 and 6. The common diagonal solution is

$$\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n) \tag{7}$$

with

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0.$$

The diagonal entries of Σ are known as Hankel singular values.

Efficient algorithms for balancing have been given by Laub et al. (1987). The Schur method for model reduction was proposed by Safonov and Chiang (1989). Glover (1984) established error bounds between original plants and reduced order models in terms of Hankel singular values. Imae et al. (1992) proposed an approach for solving balancing problems based upon a Riccati equation, which is good for time-varying problems. Tsai and Shih (1988) present a finite Sarason operator approach rather than the infinite block Hankel matrix approach for balanced realizations. Lam (1991) has presented an algorithm for the balanced realization of Padé approximations for pure dead times.

Maciejowski (1985) suggested that balanced realizations could provide good canonical forms for input–output state identification. This is based on the truncation properties of balanced realizations that make parameter estimation less dependent upon the correct initial choice of model order. Furthermore, model estimation using balanced realization has optimal estimator conditioning in the sense that the distance from singularity of the observability and reachability Gramians is greater than that obtained by any other technique.

Ober (1987) has derived an explicit canonical form for state-space linear models using balanced realization. A canonical form is critical for system identification where a unique parameterization of the system is essential. As an example, the Ober state-space parameterization is contrasted with the traditional controllable canonical form for a two-state single-input single-output model. The traditional canonical form is

$$A = \begin{bmatrix} 0 & 1 \\ a_{21} & a_{22} \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$C = [c_1 \quad c_2] \qquad D = d, \tag{8}$$

whereas the balanced canonical form is

$$A = \begin{bmatrix} \dfrac{-b_1^2}{2\sigma_1} & \dfrac{-b_1 b_2}{\sigma_1 \pm \sigma_2} \\ \dfrac{-b_2 b_1}{\sigma_2 \pm \sigma_1} & \dfrac{-b_2}{2\sigma_2} \end{bmatrix} \qquad B = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$C = [\pm b_1 \quad \pm b_2] \qquad D = d. \tag{9}$$

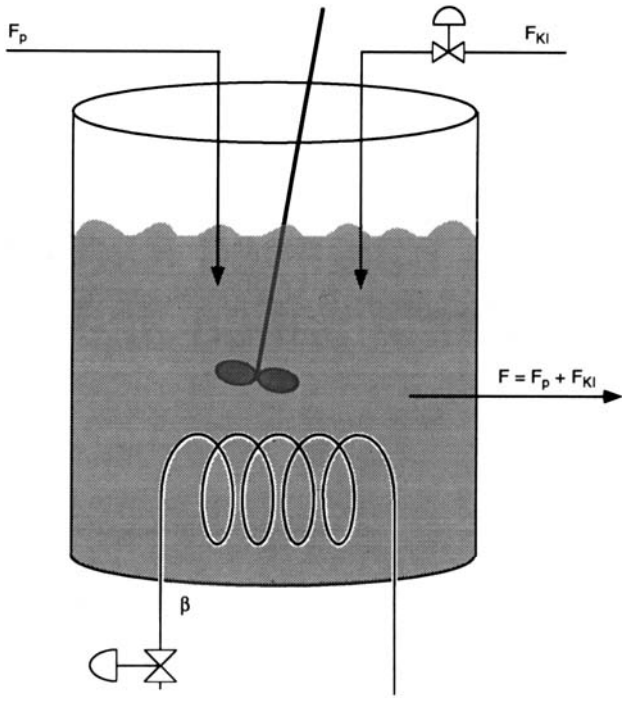


Figure 1. Stirred-tank reactor.

Each of these canonical forms has the same number of real parameters, 5. The controllable canonical form may be neither stable nor minimal, however, whereas the balanced form is guaranteed to be both stable and minimal, providing that the b_i and σ_i parameters are all positive. Parameter estimation is therefore easier if the balanced form is used, assuming that equally good initial estimates are available.

The standard method of obtaining a balanced realization is to use transformation algorithms on linear state-space models such as are found in the Matlab Toolboxes. In this work we obtain balanced realization directly from pulse test data off of the nonlinear system using the discrete-time algorithm of Kung (1979) and the canonical form of Ober (1987).

State-Space Identification

Let us consider the use of balanced realization for the identification of a state-space model using input–output data for a continuous stirred-tank reactor. We will consider the reactor studied by Lynch and Ramirez (1975). The reactor carries out the reaction of the decomposition of hydrogen peroxide into water and oxygen. This is a highly exothermic reaction that requires cooling control. Also, a homogeneous catalyst of potassium iodide is used. The flow rate of the catalyst feed is a second control on the reactor. A schematic of the system is given in Figure 1. The system dynamics are described by three state variables resulting from a material balance on the hydrogen peroxide, a material balance on the catalyst, and an energy balance. The mathematical model that describes the process dynamics is

Peroxide Material Balance

$$\frac{dC}{dt} = \frac{F_p C_0}{V_x} - \frac{FC}{V_x} - R. \tag{10}$$

Table 1. Parameters Used in Simulations

a	= 1,284.6 cal/min	k_0	= 9.775×10^9
b	= −23.61 cal/min.°C	T_c	= 19.2°C
C_{KI_0}	= 0.01615 mol/L	T_0	= 25°C
C_{KI_s}	= 0.0019 mol/L	T_s	= 30°C
C_0	= 2.25 mol/L	V_H	= 17.36 L
C_p	= 1 cal/g·°C	V_x	= 15.78 L
C_s	= 1.718 mol/L	β_s	= 122.1 cal/min.°C
E/R'	= 6524 cal/mol	ρ	= 1,000 g/L
F_{KI_s}	= 0.100 L/min	(ΔH)	= 22,600 cal/mol
F_p	= 0.750 L/min		

Catalyst Material Balance

$$\frac{dC_{KI}}{dt} = \frac{F_{KI}C_{KI0}}{V_x} - \frac{C_{KI}F}{V_x}. \tag{11}$$

Energy Balance

$$\frac{dT}{dt} = \frac{F(T_0 - T)}{V_x} + \frac{Q_s}{V_x \rho c_p} - \frac{(\Delta H)RV_x}{V_H \rho c_p} - \frac{\beta(T - T_c)}{V_H \rho c_p}. \tag{12}$$

The rate of reaction is

$$R = k_0 C_{KI} C e^{-E/R'(T+273.16)}. \tag{13}$$

The variable Q_s is the rate of conversion of mechanical energy of the stirrer into thermal energy and was modeled as a first-order rate process

$$Q_s = a + bT. \tag{14}$$

Table 1 gives appropriate model parameters for the system. For this system, the state variables are the two-system concentrations and the system temperature

$$\boldsymbol{X} = \begin{pmatrix} C \\ C_{KI} \\ T \end{pmatrix}. \tag{15}$$

The dimensional control vector is the catalyst feed rate, F_{KI}, and the cooling rate, β,

$$\boldsymbol{U} = \begin{pmatrix} F_{KI} \\ \beta \end{pmatrix}. \tag{16}$$

Two system variables are easily measurable. These are the rate of the produced oxygen gas, which is a measure of the rate of reaction, and the system temperature. The measurement vector is, therefore,

$$\boldsymbol{Y} = \begin{pmatrix} R \\ T \end{pmatrix}. \tag{17}$$

There are two classes of disturbances that affect the problem, measurable and unmeasurable. The measurable disturbances are the peroxide feed flow rate, the coolant temperature, and the combined inlet feed temperature of the peroxide and catalyst feeds,

$$\boldsymbol{D}_m = \begin{pmatrix} F_p \\ T_c \\ T_d \end{pmatrix}. \tag{18}$$

The unmeasured disturbances that affect the system are the inlet feed concentration of the peroxide and catalyst,

$$\boldsymbol{D}_u = \begin{pmatrix} C_0 \\ C_{KI_0} \end{pmatrix}. \tag{19}$$

Typical nonlinear system dynamics are illustrated in Figure 2, which shows the response of the system to an initially high catalyst concentration. The catalyst concentration exponentially decays back to its equilibrium value, but the peroxide and temperature response dynamics indicate the nonlinearities involved in the system model.

In order to compare a balanced realization to a traditional linear model of the system, we first linearize the nonlinear system about the steady-state conditions. In terms of normalized deviation variables about the steady-state condition, the linearized model is

$$\dot{\boldsymbol{x}} = \begin{bmatrix} -0.0621 & -0.0082 & -0.0173 \\ 0 & -0.0539 & 0 \\ 0.00978 & 0.00978 & -0.0417 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} -0.00634 & 0 \\ 0.0475 & 0 \\ -0.00101 & -0.00299 \end{bmatrix} \boldsymbol{u} + \begin{bmatrix} 0.0145 & 0 & 0 \\ -0.0475 & 0 & 0 \\ -0.00745 & 0.00454 & 0.0453 \end{bmatrix} \boldsymbol{d}_m \tag{20}$$

$$\boldsymbol{y} = \begin{bmatrix} 1 & 1 & 2.113 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{x}. \tag{21}$$

Figure 3 gives the response of the actual and linearized system outputs to unit-pulse input control and measurement disturbances lasting 5 min. The comparisons are reasonable, especially to pulses in the control variables. There is more discrepancy between the response dynamics of the linear and nonlinear system to disturbance impulses.

Using the algorithm of Kung (1979) and the canonical form of Ober (1987), a balanced realization was obtained from input–output data of pulse-response tests on the nonlinear model. The Kung algorithm is also described by Chen (1984) and Maciejowski and Vines (1984). Since Kung's algorithm assumes discrete time, the output data obtained from the pulse-response tests were sampled, with a sampling interval of 5 minutes, and a discrete-time linear-state-space model was obtained. The discrete-time model can be converted into a continuous-time model using standard transformations.

Kung's algorithm gives a high-order discrete-time balance realization, whose McMillan degree can be reduced simply

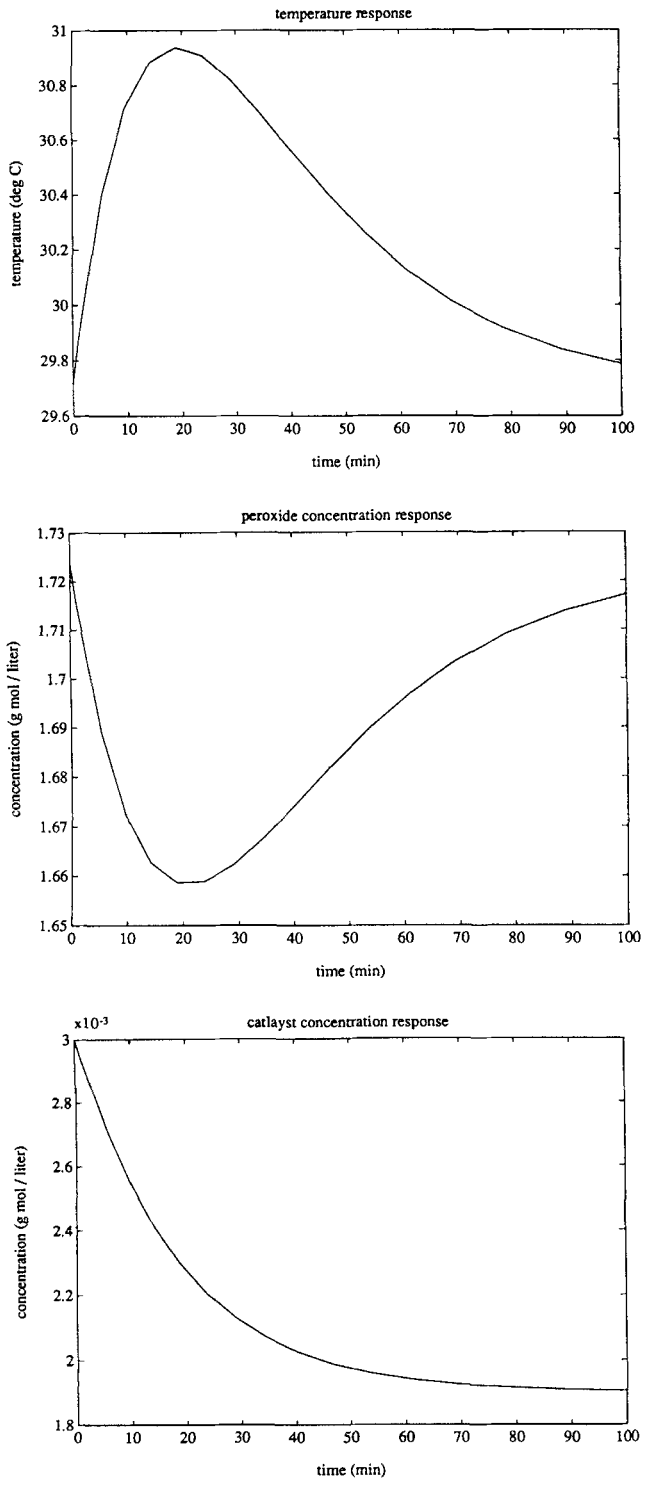


Figure 2. Nonlinear response to an initial catalyst disturbance.

by truncating the state vector. It also gives the Hankel singular values of the realization, and these give an indication of a suitable McMillan degree for the approximate model. From Glover (1984), it is known that the L_∞ error in the input–output response between the original and reduced models is bounded by the sum of the discarded Hankel singular values. The actual error is much smaller than this bound for the great majority of cases. For our balanced realization, the first five Hankel singular values are

$$\Sigma = \text{diag}(1.52, 0.195, 0.0284, 0.0178, 0.0088, \ldots). \quad (22)$$

Only the first five terms are given since all other terms are smaller and less important. The second sigma term is nearly one order of magnitude smaller than the first dominant term. The third term is 50 times smaller than the dominant diagonal element. All other terms continue to decrease in magnitude slowly. A decision is needed on the appropriate order for the balanced realization. Although a second-order model fits the data well, a third-order model was chosen since it maintains one term less than two orders of magnitude smaller than the dominant value. Also, a third-order model more closely reflects the number of actual state variables for this system. Figure 4 compares the nonlinear data and a third-order balanced realization. A comparison of Figures 3 and 4 shows that the balanced realization is a better fit to the nonlinear data than the linear model linearized about the steady state. This is especially true for the output responses due to disturbances in the inlet peroxide flow rate and the inlet feed temperature. The complete third-order discrete-time model is given by the following state-space model:

$$\begin{aligned} x_b(n+1) = &\begin{bmatrix} 0.774 & -0.0493 & -0.0444 \\ 0.0498 & 0.770 & 0.0442 \\ -0.0358 & -0.0193 & 0.789 \end{bmatrix} x_b(n) \\ &+ \begin{bmatrix} 0.218 & -0.0334 \\ -0.223 & -0.0022 \\ 0.0173 & 0.0290 \end{bmatrix} u(n) \\ &+ \begin{bmatrix} -0.266 & 0.0566 & 0.697 \\ 0.150 & 0.0128 & 0.0524 \\ 0.0326 & 0.0064 & 0.0693 \end{bmatrix} d_m(n). \end{aligned} \quad (23)$$

It is interesting to notice that the state transition matrix is diagonally dominant.

We can convert this discrete-time model into an equivalent continuous time model. The result for the state, control, and disturbance matrices are

$$A_b = \begin{bmatrix} -0.0510 & -0.0129 & -0.0110 \\ 0.0131 & -0.0517 & 0.0117 \\ -0.0090 & -0.0052 & -0.0474 \end{bmatrix}$$

$$B_b = \begin{bmatrix} 0.048 & -0.0074 \\ -0.0524 & -0.0004 \\ 0.0043 & 0.00249 \end{bmatrix}$$

$$W_b = \begin{bmatrix} -0.0589 & 0.013 & 0.1588 \\ 0.0358 & 0.0025 & 0.0064 \\ 0.0065 & 0.0017 & 0.0191 \end{bmatrix}$$

$$C_b = \begin{bmatrix} 0.729 & -0.0344 & 0.0816 \\ 0.279 & 0.273 & -0.0246 \end{bmatrix} \quad (24)$$

The eigenvalues of the original linearized A matrix of Eq. 20 are

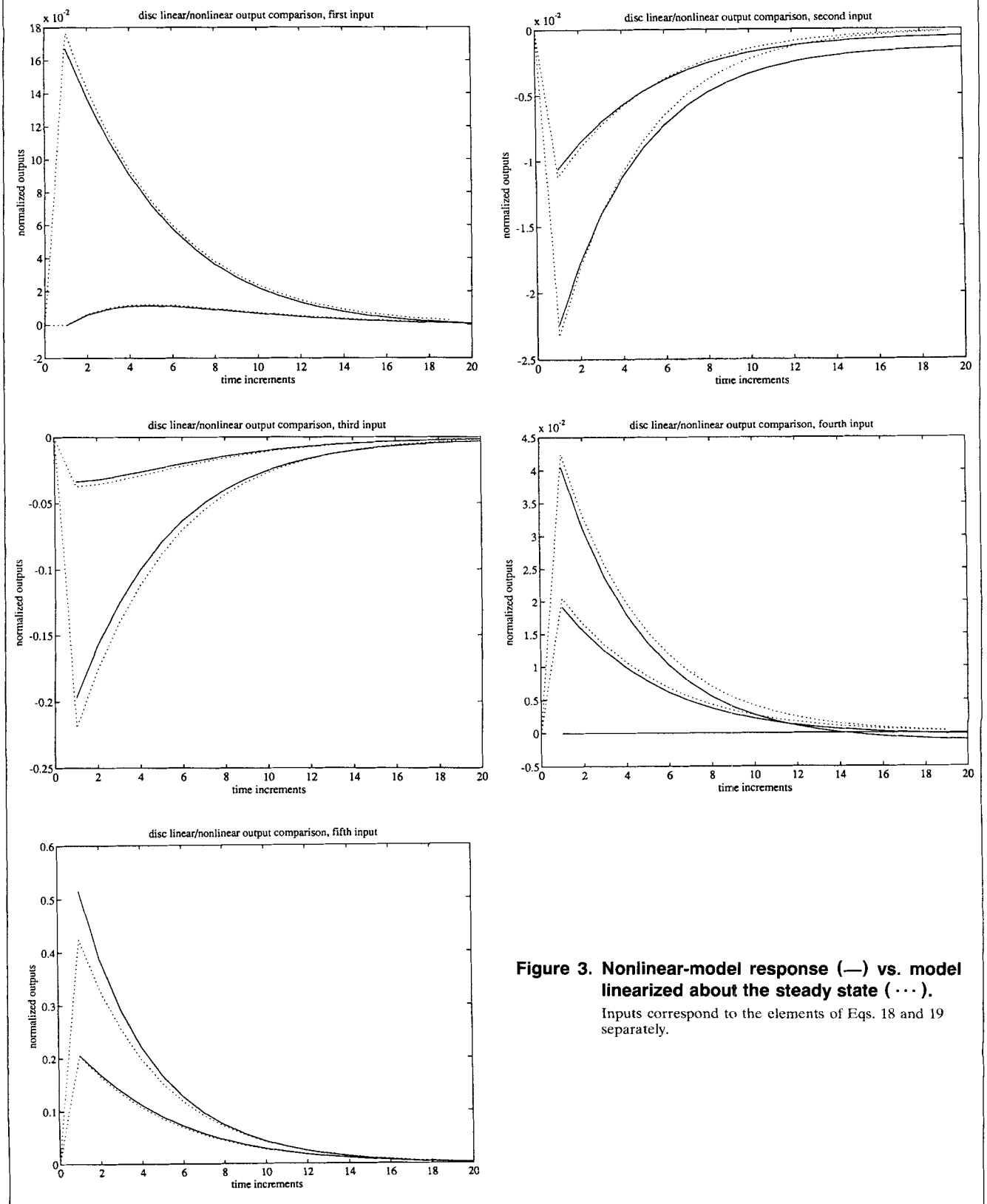


Figure 3. Nonlinear-model response (—) vs. model linearized about the steady state (· · ·).

Inputs correspond to the elements of Eqs. 18 and 19 separately.

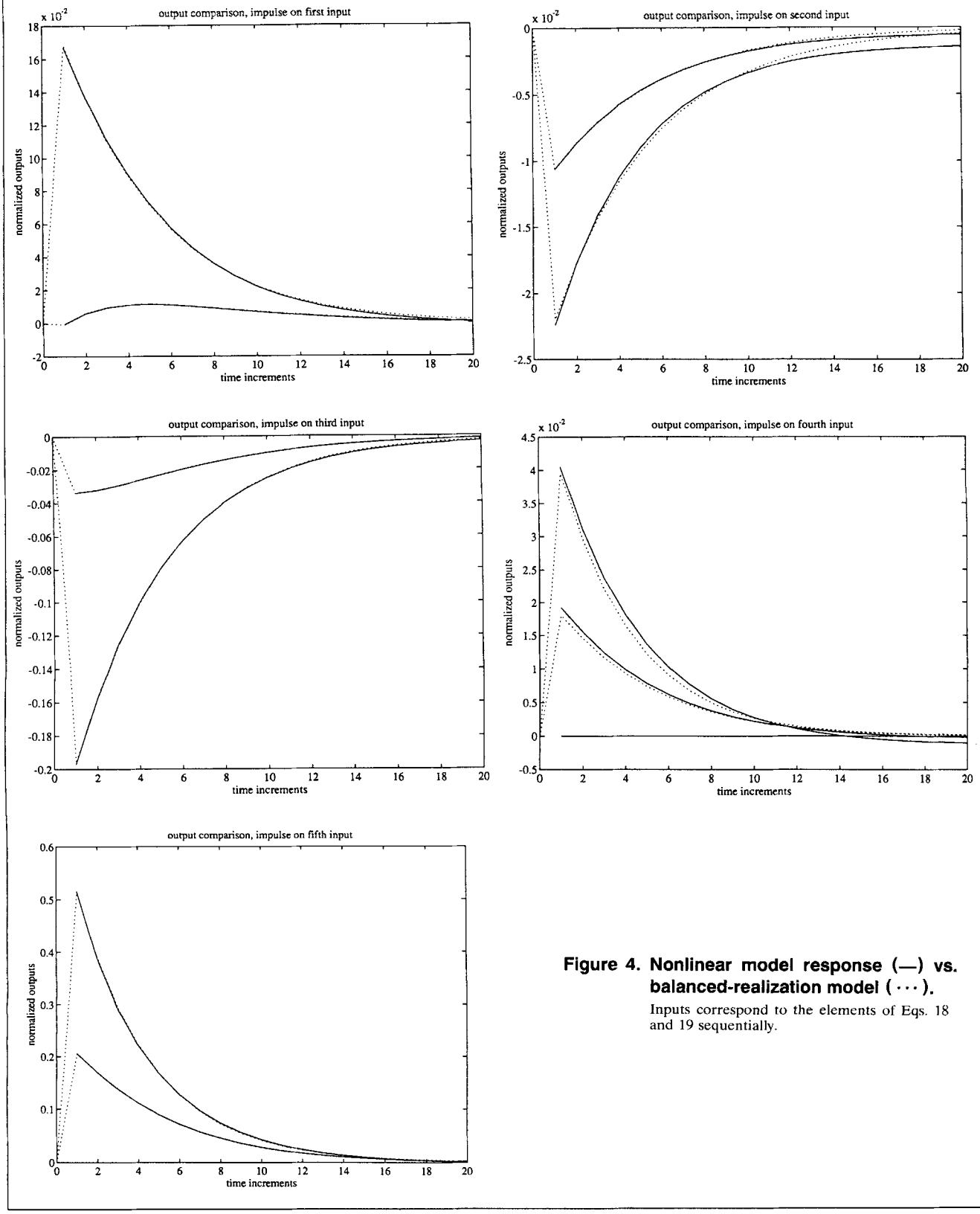


Figure 4. Nonlinear model response (—) vs. balanced-realization model (· · ·).

Inputs correspond to the elements of Eqs. 18 and 19 sequentially.

$$\lambda = \begin{pmatrix} -0.0519 + 0.00812\,i \\ -0.0519 + 0.00812\,i \\ -0.0539 \end{pmatrix}, \tag{25}$$

whereas the eigenvalues of the state matrix for the balanced realization are

$$\lambda_b = \begin{pmatrix} -0.0555 + 0.01472\,i \\ -0.0555 + 0.01472\,i \\ -0.0392 \end{pmatrix}. \tag{26}$$

There is good agreement between both sets of eigenvalues. A small adjustment to the realization dynamics and proper control and measurement matrices have resulted in a significantly improved fit to the nonlinear data. Since the balanced realization state variables are not the actual physical state variables, we do not expect a one-to-one correspondence between the system matrices of the linearized system (Eqs. 20 and 21) and the balanced system (Eq. 24). It is interesting to note the similarity of the two state matrices, however. Also it is clear from the C_b matrix that there is a reduced importance of the third balanced state variable in capturing the output system dynamics. All other balanced states are even less important due to the property of Eq. 7.

We have shown that the balanced realization approach yields very satisfactory state-space models using input–output data generated from a nonlinear model. In fact, the results are better than those obtained by conventional linearization. The explanation for this is that the data supplied to the realization algorithm contain information on the nonlinearity of the model for perturbed signals of finite amplitude. Conventional linearization assumes infinitesimally small perturbations. Consequently, the balanced realization approach gives a more accurate model. This is a valuable advantage, if amplitudes of typical perturbations are known.

Optimal Output Controller Design

We now consider the design of an optimal model predictive controller for output regulation. The goal is to use the balanced realization model derived solely from input–output data to determine a controller that minimizes the quadratic Gaussian performance functional

$$J = \mathcal{E}\left\{\frac{1}{2}\int_0^\infty (\boldsymbol{y}^T\boldsymbol{Q}\boldsymbol{y} + \boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u})\,dt\right\}. \tag{27}$$

This performance index can be transformed into a state performance functional, quadratic in the balanced realized state vector, $\boldsymbol{x}_b$

$$J = \mathcal{E}\left\{\frac{1}{2}\int_0^\infty (\boldsymbol{x}_b^T\boldsymbol{Q}_b\boldsymbol{x}_b + \boldsymbol{u}^T\boldsymbol{R}\boldsymbol{u})\,dt\right\}, \tag{28}$$

where

$$\boldsymbol{Q}_b = \boldsymbol{C}_b^T\boldsymbol{Q}\boldsymbol{C}_b. \tag{29}$$

Since both feedforward model predictive elements using the measured disturbances, $\boldsymbol{d}_m$, and integral elements compensating for unmeasured disturbances, $\boldsymbol{d}_u$, should be incorporated into the system design, an approach similar to that developed by Ramirez et al. (1987) and Park and Ramirez (1990) is taken.

For the system

$$\dot{\boldsymbol{x}}_b = \boldsymbol{A}_b\boldsymbol{x}_b + \boldsymbol{B}_b\boldsymbol{u} + \boldsymbol{W}_b\boldsymbol{d}_m + \boldsymbol{W}_u\boldsymbol{d}_u + \boldsymbol{v}, \tag{30}$$

we first design a predictive controller to handle the measured disturbances $\boldsymbol{d}_m$ based upon the model

$$\dot{\boldsymbol{x}}_b = \boldsymbol{A}_b\boldsymbol{x}_p + \boldsymbol{B}_b\boldsymbol{u}_p + \boldsymbol{W}_b\boldsymbol{d}_m. \tag{31}$$

The predictive states, $\boldsymbol{x}_p$, and predictive controls, $\boldsymbol{u}_p$, are computed in order that the system maintains as many outputs $\boldsymbol{y}$ at desired levels as the controls can allow. This is equivalent to forming the partitioned set

$$\begin{pmatrix} \boldsymbol{0} \\ \dot{\boldsymbol{x}}_{fp} \end{pmatrix} = \begin{bmatrix} \boldsymbol{A}_{b11} & \boldsymbol{A}_{b12} \\ \boldsymbol{A}_{b21} & \boldsymbol{A}_{b22} \end{bmatrix} \begin{pmatrix} \boldsymbol{x}_{sp} \\ \boldsymbol{x}_{fp} \end{pmatrix} + \begin{bmatrix} \boldsymbol{B}_{b1} \\ \boldsymbol{B}_{b2} \end{bmatrix} \boldsymbol{u}_p + \begin{bmatrix} \boldsymbol{W}_{b1} \\ \boldsymbol{W}_{b2} \end{bmatrix} \boldsymbol{d}_m, \tag{32}$$

where $\boldsymbol{x}_{sp}$ are specified states and $\boldsymbol{x}_{fp}$ are free states. Usually, if there are m controls, then m outputs $\boldsymbol{y}_s$ can be maintained at desired levels even with load disturbances, $\boldsymbol{d}_m$. The specified outputs $\boldsymbol{y}_s$ then allow a similar number of balanced states to be specified, $\boldsymbol{x}_{sp}$, through the equation

$$\boldsymbol{y}_s = \boldsymbol{C}_{b1}\boldsymbol{x}_{sp} + \boldsymbol{C}_{b2}\boldsymbol{x}_{fp}. \tag{33}$$

The specified predicted balanced states are chosen such that $\boldsymbol{C}_{b1}$ is square and its inverse exists. Due to the property that each balanced state is of less importance in describing the output dynamics than the preceding one, the $\boldsymbol{C}_{b1}$ matrix will usually contain the first m columns of $\boldsymbol{C}_b$. Therefore,

$$\boldsymbol{x}_{sp} = \boldsymbol{C}_{b1}^{-1}\boldsymbol{y}_s - \boldsymbol{C}_{b1}^{-1}\boldsymbol{C}_{b2}\boldsymbol{x}_{fp}. \tag{34}$$

Normally it is desired to maintain the specified outputs at the original equilibrium condition, $\boldsymbol{y}_s = \boldsymbol{0}$. The free predicted states, $\boldsymbol{x}_{fp}$, and the predictive controls, $\boldsymbol{u}_p$, are computed from the algebraic and differential relations of Eq. 32. As discussed by Ramirez et al. (1987), there are some conditions on the $\boldsymbol{A}_b$ and $\boldsymbol{B}_b$ partitioned matrices that guarantee the existence of a stable solution to the set of relations of Eq. 32.

Defining the variables

$$\bar{\boldsymbol{x}} = \boldsymbol{x}_b - \boldsymbol{x}_p \quad \bar{\boldsymbol{u}} = \boldsymbol{u} - \boldsymbol{u}_p, \tag{35}$$

the system dynamics of Eq. 30 becomes

$$\dot{\bar{\boldsymbol{x}}} = \boldsymbol{A}_b\bar{\boldsymbol{x}} + \boldsymbol{B}_b\bar{\boldsymbol{u}} + \boldsymbol{W}_u\boldsymbol{d}_u + \boldsymbol{v}. \tag{36}$$

This is now the unmeasurable disturbance problem of Johnson (1972), which has an optimal control solution

$$\boldsymbol{u} = \boldsymbol{u}_p + \boldsymbol{K}_2(\hat{\boldsymbol{x}} - \boldsymbol{x}_p) - [\boldsymbol{K}_1 - \boldsymbol{K}_2\boldsymbol{A}_b]\int(\hat{\boldsymbol{x}}_b - \boldsymbol{x}_p)\,dt \tag{37}$$

minimizing the regulator performance index

$$J = \mathcal{E}\left\{\frac{1}{2}\int_0^\infty \left(\tilde{x}^T\tilde{Q}\tilde{x} + \tilde{u}^T\tilde{R}\tilde{u}\right)\,dt\right\}, \tag{38}$$

where

$$\tilde{x} = \begin{pmatrix}\bar{x}\\ z\end{pmatrix}\dot{\tilde{u}} = \dot{\bar{u}} \tag{39}$$

with

$$z = B_b\bar{u} + W_u d_u \tag{40}$$

and

$$\tilde{Q} = \begin{bmatrix}Q_b & 0\\ 0 & 0\end{bmatrix}\tilde{R} = R. \tag{41}$$

Note that the optimal solution has used the separation principle of stochastic control. The solution to the stochastic LQG control problem is to use the result of the optimal LQR problem except that the control is feedback in the optimal state estimates obtained from the use of a Kalman filter, $\tilde{x}_b$, rather than the instantaneous values themselves, x_b.

A final consideration is the introduction of control constraints into the optimal control law

$$u_{\min} \le u \le u_{\max}. \tag{42}$$

Often control saturation can be minimized by the proper choice of the Q/R ratio.

As discussed by Ramirez (1994), a practical suboptimal algorithm for control saturation, which is a close approximation to the complete iterative algorithm, is to use the relations of Eq. 37, imposing the constraints of Eq. 42 on the actual implemented controls. Whenever the constraints are active, the integral term of Eq. 37 is not calculated in order to prevent integral windup. Campo and Morari (1990) have discussed a complete two-step design technique for saturating systems. First the controller is designed neglecting saturation as was shown earlier. Then a saturation compensation scheme is designed that provides graceful degradation of the closed-loop performance in the face of saturation.

The model predictive nature of the control design is shown clearly in Figure 5 since the controller functions on the error, e, which is the difference between the actual estimated balanced state vector, $\tilde{x}_b$, and the predicted state, x_p, which compensates for measurable disturbances. Also, the figure illustrates the need for a Kalman filter in order to estimate the balanced states for the system outputs using the balanced realization state, control, and disturbance matrices (A_b, B_b, C_b, W_b). The Kalman filter is capable of filtering out measurement noise and explicitly considers white Gaussian model uncertainty in the state variables. For this application, we desire the filter to track the system outputs well so the outputs should be prefiltered so that measurement noise is small compared to the model uncertainty. Accurate and rapid output tracking means that the estimated states, $\hat{x}_b$, will reflect well the true output dynamic response of the system.

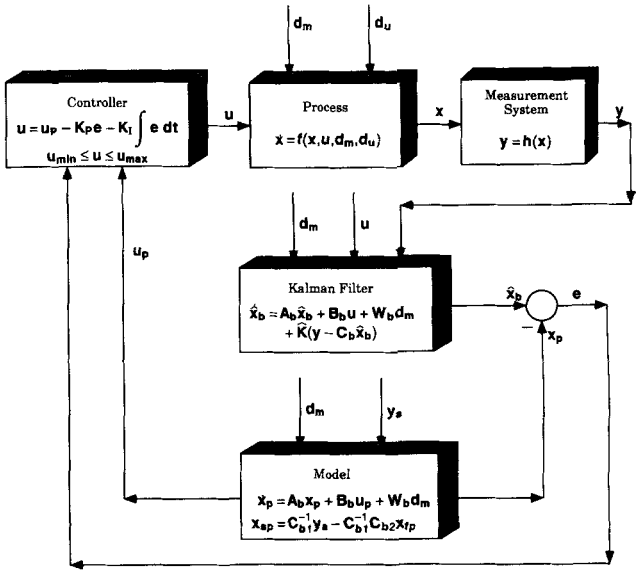


Figure 5. Balanced-realization optimal controller.

The control structure of Figure 5 is new in the sense that it combines a particular form of predictive control with an output regulation form of LQG control. Using LQG for output regulation is well established (Maciejowski, 1989), and proposals for combining this with predictive control have appeared (Bitmead et al., 1990). The new features are the particular form of the predictive control based upon Ramirez et al. (1987) and the use of a balanced model based upon actual response data to give an obvious choice of a well-conditioned submatrix C_{b1}, for use in Eq. 34.

We have designed the control algorithms using a continuous time development since the transformations to handle explicitly both measurable and unmeasurable disturbances appear naturally using a continuous formulation rather than a discrete formulation (Ramirez, 1994). The LQG formulation of the problem contrasts with other model predictive control methods (Prett and Garcia, 1988; Eaton and Rawlings, 1992) that use receding horizon concepts and direct quadratic function minimization to handle control and state constraints. Often unfeasible solutions appear in the direct minimization approach that lead to temporary but large violations of constraints rather than the longer but smaller violations resulting from the LQG method.

Optimal CSTR Output Control

Let us apply the optimal output control algorithm to the Lynch-Ramirez reactor. For this problem we have two controls (Eq. 16) and two outputs (Eq. 17); therefore, we can specify that both normalized outputs be specified outputs, $y = 0$, and Eq. 34 becomes,

$$x_{sp} = \begin{pmatrix}x_{1p}\\ x_{2p}\end{pmatrix} = -\begin{bmatrix}0.729 & -0.0344\\ 0.279 & 0.273\end{bmatrix}^{-1}\begin{bmatrix}0.0816\\ -0.0246\end{bmatrix}x_{3p} \tag{43}$$

$$\begin{pmatrix}x_{1p}\\ x_{2p}\end{pmatrix} = \begin{pmatrix}-0.103\\ 0.195\end{pmatrix}x_{3p}. \tag{44}$$

The predicted relations given by Eq. 32 become

$$\mathbf{0} = \begin{bmatrix} -0.0510 & -0.0129 \\ 0.0131 & -0.0517 \end{bmatrix} \begin{pmatrix} -0.103 \\ 0.195 \end{pmatrix} x_{3p} + \begin{pmatrix} -0.0110 \\ 0.0117 \end{pmatrix} x_{3p} + \begin{bmatrix} 0.048 & -0.0074 \\ -0.0524 & -0.0004 \end{bmatrix} \boldsymbol{u}_p + \begin{bmatrix} -0.0589 & 0.013 & 0.1588 \\ 0.0348 & 0.0025 & 0.0064 \end{bmatrix} \boldsymbol{d}_m \quad (45)$$

$$\dot{x}_{3p} = (-0.0090 - 0.0052) \begin{pmatrix} -0.103 \\ 0.195 \end{pmatrix} x_{3p} - 0.0474 x_{sp} + (0.0043 \quad 0.00249) \boldsymbol{u}_p + (0.0065 \quad 0.0017 \quad 0.0191) \boldsymbol{d}_m. \quad (46)$$

Equation 45 is solved for the predicted controls as

$$\boldsymbol{u}_p = \begin{pmatrix} -0.013 \\ 1.034 \end{pmatrix} x_{3p} + \begin{bmatrix} 0.709 & 0.0327 & -0.0397 \\ -3.361 & 1.969 & 21.202 \end{bmatrix} \boldsymbol{d}_m. \quad (47)$$

Substituting Eq. 47 into Eq. 46 gives

$$\dot{x}_{3p} = -0.0501 x_{3p} + (0.0012 \quad 0.0067 \quad 0.0717)\, \boldsymbol{d}_m. \quad (48)$$

Equation 48 shows that the free predicted state equation is stable since the homogeneous part is described by a negative semidefinite matrix (in this case, the scalar -0.0501). Equations 47 and 48 give the relations for the predicted control, $\boldsymbol{u}_p$, and the free predicted state $\boldsymbol{x}_{3p}$. The specified predicted states are calculated using Eq. 44.

To calculate the proportional and integral feedback gains, we use the weighting matrices

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \boldsymbol{R} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}. \quad (49)$$

These weighting matrices give an order of magnitude more importance to regulating the system outputs rather than the cost of control. The resulting feedback gains are

$$\boldsymbol{K}_p = \begin{bmatrix} 7.612 & -1.371 & 0.8379 \\ -10.39 & -8.416 & 0.9665 \end{bmatrix}$$

$$\boldsymbol{K}_I = \begin{bmatrix} 2.238 & -0.1692 & 0.2629 \\ -1.042 & -0.8535 & 0.0595 \end{bmatrix}. \quad (50)$$

The equal diagonal weightings in the $\boldsymbol{Q}$ matrix reflect the fact that with two controls, both output variables, temperature and reaction rate, can be regulated equally well about their equilibrium conditions.

For the design of the Kalman filter, we assume that the state uncertainty is one order of magnitude larger than that of the filtered measurement noise. If there is a 1% measurement noise and a 10% model uncertainty, then

$$\tilde{\boldsymbol{Q}} = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix} \quad \tilde{\boldsymbol{R}} = \begin{bmatrix} 0.0001 & 0 \\ 0 & 0.0001 \end{bmatrix}. \quad (51)$$

The resulting Kalman filter gain is

$$\tilde{\boldsymbol{K}} = \begin{bmatrix} 9.557 & 3.036 \\ -3.136 & 9.296 \\ -0.054 & -1.009 \end{bmatrix} \quad (52)$$

To test the effectiveness of this optimal output control algorithm based upon balanced state realization, we first investigated the response characteristics with no disturbances present. Model uncertainty is included in the simulations since the actual system is nonlinear whereas the balanced realized model is linear. Gaussian white noise was added to the system measurements. Since the system is nonlinear and the balanced realization model and model predictive control is linear, there should be some small dynamic perturbations in the closed-loop response. Figure 6 shows the temperature response of the system with the feedback gains of Eq. 50 and the Kalman filter gain of Eq. 52. There is an expected small transient, but also a small steady-state error. The steady-state error is due to the fact that there is an inaccurate state prediction in the Kalman filter so that at equilibrium, both the model prediction term and the output error compensation term of the filter are not zero but equal to each other. This means that the dynamic filter

$$\dot{\hat{\boldsymbol{x}}}_b = \boldsymbol{A}_b \hat{\boldsymbol{x}}_b + \boldsymbol{B}_b \boldsymbol{u} + \boldsymbol{W}_b \boldsymbol{d}_m + \hat{\boldsymbol{K}}(\boldsymbol{y} - \boldsymbol{C}_b \hat{\boldsymbol{x}}_b) \quad (53)$$

becomes at equilibrium

$$\boldsymbol{A}_b \hat{\boldsymbol{x}}_b + \boldsymbol{B}_b \boldsymbol{u} + \boldsymbol{W}_b \boldsymbol{d}_m = -\hat{\boldsymbol{K}}(\boldsymbol{y} - \boldsymbol{C}_b \hat{\boldsymbol{x}}_b), \quad (54)$$

yielding a steady-state error in the output variables $\boldsymbol{y}$.

Figure 7 shows the balanced state response. It is significant to notice that state one is quite noisy, with a high-frequency oscillation about its equilibrium condition. This oscillation is due to the large gain matrix used by the Kalman filter.

We would like to eliminate the steady-state output error as well as reduce the filter gain. To reduce the filter gain in order to eliminate the balanced state oscillations, we design

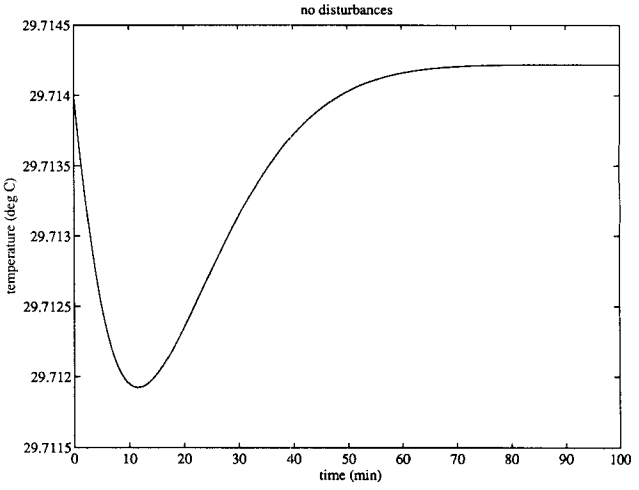


Figure 6. Temperature response, no disturbances.

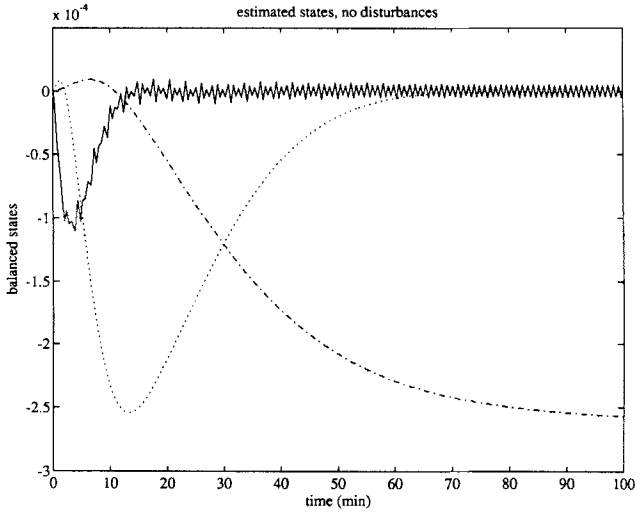


Figure 7. Balanced-state response, no disturbances.

the filter gain based upon equal uncertainties in the states and measurements,

$$\hat{Q} = \text{diag}(0.0001)\,\hat{R} = \text{diag}(0.0001), \tag{55}$$

which results in a filter gain of

$$\hat{K} = \begin{bmatrix} 0.9014 & 0.292 \\ -0.2704 & 0.777 \\ -0.0101 & -0.1002 \end{bmatrix}. \tag{56}$$

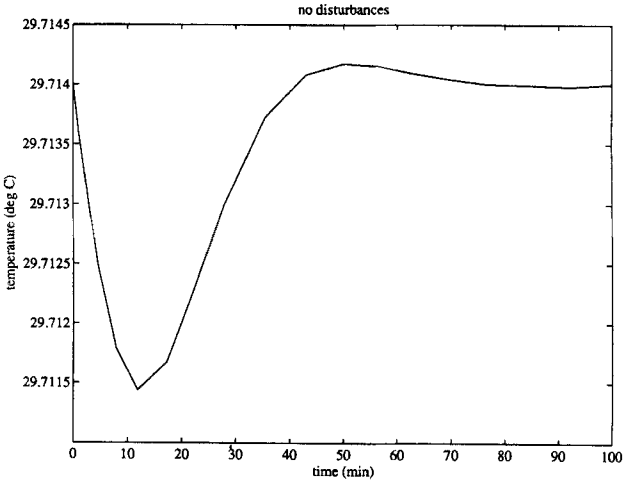
To eliminate the steady-state output errors, we eliminate the model prediction part from the Kalman filter and use a modified state estimator algorithm

$$\dot{\hat{x}} = \hat{K}(y - C_b\hat{x}_b). \tag{57}$$

The stability of the estimation is given by the eigenvalues of $(-\hat{K}C_b)$.

Figure 8 shows the outputs using this modified state estimator. Indeed, there is no steady-state output error. Figure 9 gives the balanced state response, which shows that the high-frequency state oscillations have been removed by designing for a reduced filter gain. The sharp turning points in Figures 8 and 9 are due to the expanded scales used to illustrate the fact that there are no offsets or significant dynamic effects.

We now test the proposed closed-loop algorithm of Figure 5 with the modified state estimator of Eq. 57 by applying 20% step changes to both measured and unmeasured disturbances. The output response of the system is given in Figure 10. It shows that both outputs are well regulated without any steady-state errors and small dynamic excursions from desired values. Figure 11 gives the response of the other two actual but nonmeasurable physical states, the reactant concentration and the catalyst concentration. These variables do not return to their original values as expected with load disturbances present. The control action is shown in Figure 12.



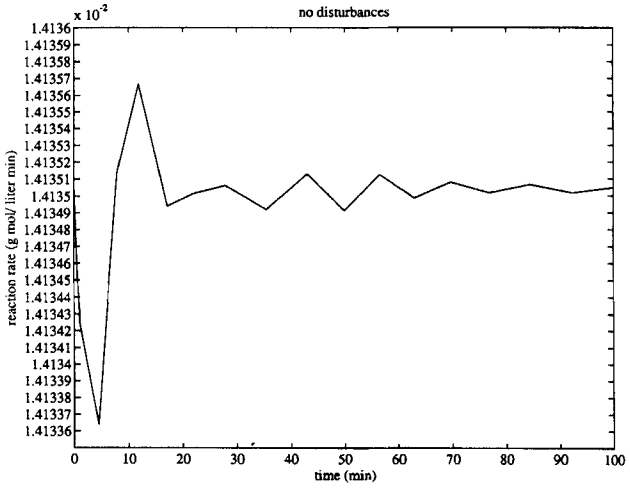


Figure 8. Output response.

Normalized deviation controls are plotted that are deviations from their nominal values and normalized by the same nomi-

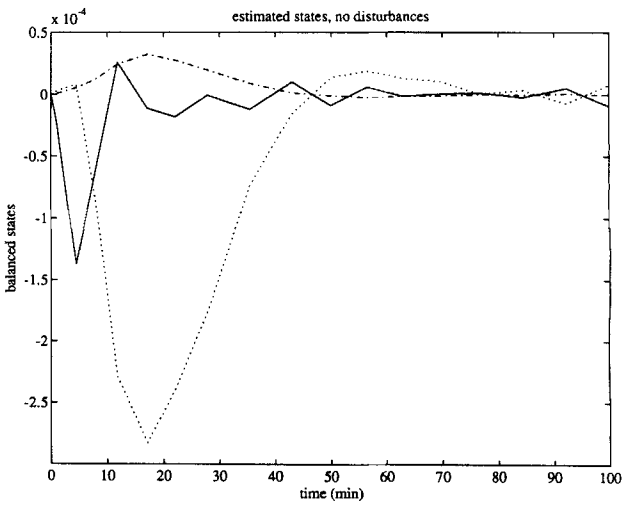
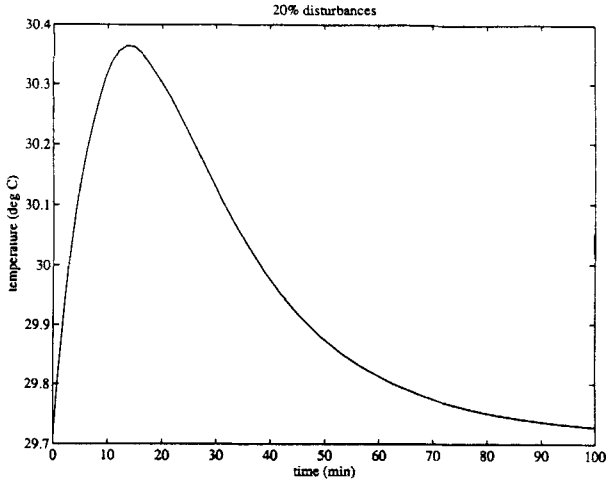


Figure 9. Balanced states, no disturbances.



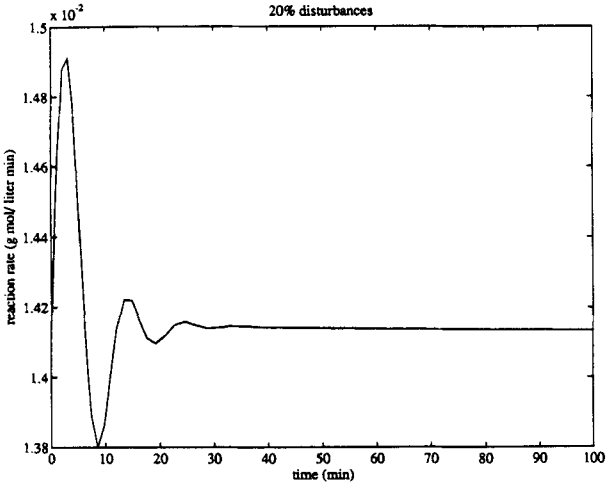
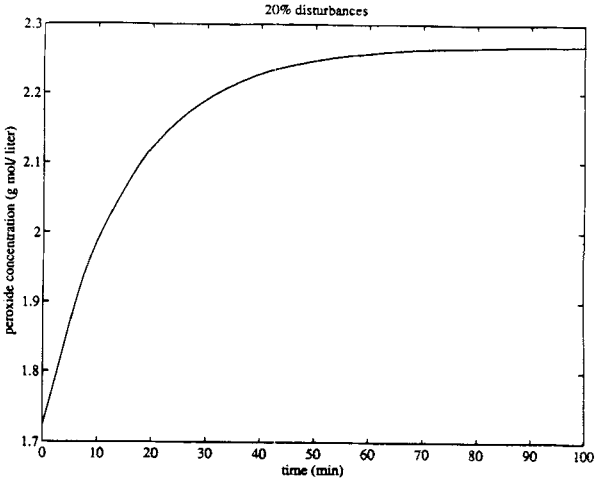


Figure 10. Output response.



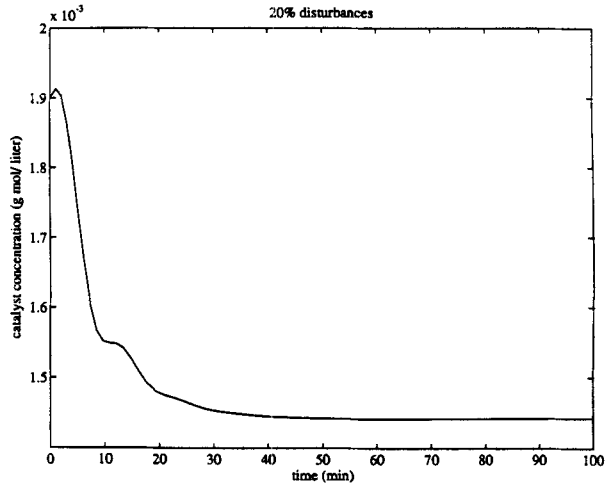


Figure 11. State-variable response.

nal values. The effect of feedforward model prediction is seen in the fact that there is an immediate movement of the controls based upon measured disturbances. It should be noted that rather large control action is necessary in order to compensate for the rather severe disturbances applied to the system. The balanced states are given in Figure 13. They show little resemblance to the actual states except that balanced state two, x_{b2}, resembles the temperature response that is one of the output variables. Balanced state one, x_{b1}, resembles closely the reaction rate output response. It is significant to note that although these balanced states do not mimic the physical states, they were successfully used to provide excellent output control of the process. Indeed, these balanced states were created in order to accurately model the input–output behavior of the system.

Conclusions

We have presented a new input–output approach for the optimal regulatory control of systems based upon balanced-state realization. State-space identification is achieved using

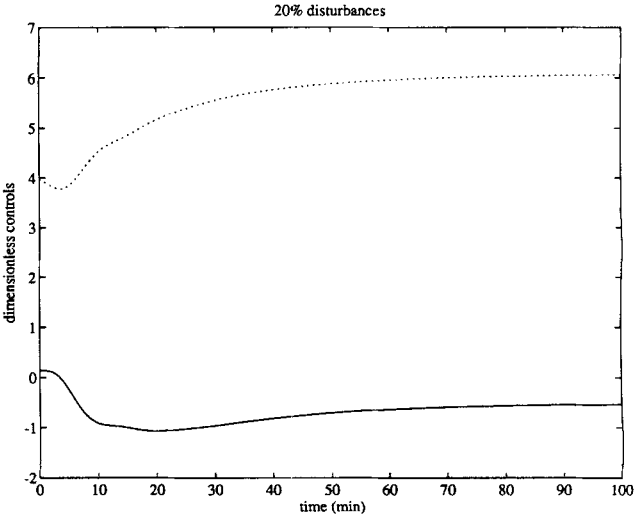


Figure 12. Control variables (—) u_1, ($\cdots$) u_2.

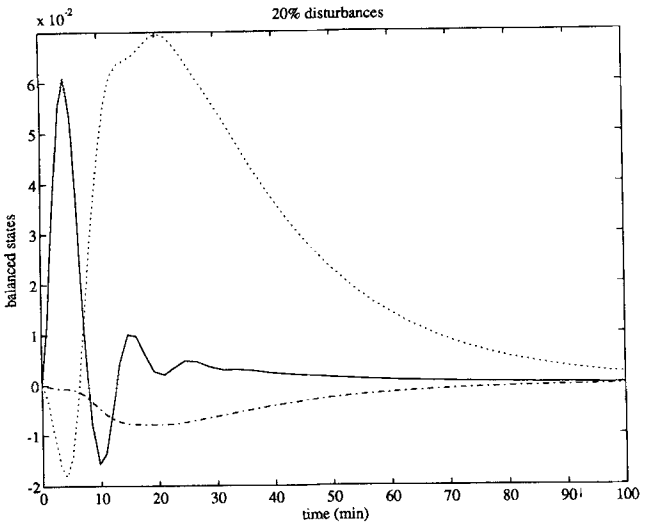


Figure 13. Balanced states (—) x_1, (···) x_2, (– · –) x_3.

input–output data and a balanced realization algorithm. The balanced-realization state-space model more accurately captured the nonlinear dynamics for a CSTR system. Based upon the realized model, an optimal controller design is quite general since it is based upon a standard structure of a linear balanced-state realization description of input–output response characteristics. The optimal control algorithm is model predictive in that predictions are made based upon measured disturbances. It also compensates for unmeasured disturbances through integral state feedback. There is a need for state identification. Simulations showed that a standard Kalman filter can lead to steady-state errors in the output variables because of the adopted control structure. A modified filter that drops the model prediction term in the Kalman filter is shown to eliminate these output steady-state errors.

The design concept was shown to lead to excellent control of a continuous stirred-tank reactor process. The combination of model representation and optimal control design is very powerful and should prove particularly useful for those applications in which on-line redesign of controllers for complex processes is required.

Acknowledgments

The first author gratefully acknowledges the financial support of a Faculty Fellowship granted by the Committee on Research and Creative Work of the University of Colorado that allowed this work to be carried out at Cambridge University.

Notation

C = hydrogen peroxide concentration, mol/L
C_{KI} = catalyst concentration, mol/L
C_0 = inlet peroxide concentration, mol/L
C_{KI_0} = inlet catalyst concentration, mol/L
D = disturbance matrix
R = reaction rate, mol/L·min
t = time, min
T = temperature, °C
v = white Gaussian noise
V_x = reactor volume, L
V_H = thermal reactor volume, L

Literature Cited

Bitmead, R. R., M. Gevers, and V. Wertz, *Optimal Adaptive Control—The Thinking Man's CPC*, Prentice Hall, Englewood Cliffs, NJ (1990).

Chen, C. T., *Linear System Theory and Design*, Holt, Rinehart & Winston (1984).

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

Eaton, J. W., and J. B. Rawlings, "Model Predictive Control of Chemical Processes," *Chem. Eng. Sci.*, **47**, 705 (1992).

Glover, K., "All Optimal Hankel-Norm Approximations of Linear Multivariable Systems and Their L Infinity Error Bounds," *Int. J. Contr.*, **39**, 1115 (1984).

Imae, J., J. E. Perkins, and J. B. Moore, "Toward Time-Varying Balanced Realization via Riccati Equations," *Math. Contr. Signals Syst.*, **5**, 313 (1992).

Johnson, C. D., "Accommodation of Disturbances in Optimal Control Problems," *Int. J. Cont.*, **15**, 209 (1972).

Kung, S. Y., "A New Low-Order Approximation Algorithm via Singular Value Decomposition," *Proc. Asilomar Conf. on Circuits, Systems and Computers* (1979).

Lam, J., "Balanced Realization of Padé Approximations of Dead Time," *IEEE Trans. Automat. Contr.*, **AC-36**, 1096 (1991).

Laub, A. J., C. C. Heath, C. C. Paige, and R. C. Ward, "Computation of System Balancing Transformations and Other Applications of Simultaneous Diagonalization Algorithms," *IEEE Trans. Automat. Contr.*, **AC-32**, 115 (1987).

Lynch, E. B., and W. F. Ramirez, "Realtime Time Optimal Control of a Stirred Tank Reactor Using Kalman Filtering for State Estimation," *AIChE J.*, **21**, 799 (1975).

Maciejowski, J. M., "Balanced Realizations in System Identification," *Proc. IFAC Symp. on System Identification*, York, England (1985).

Maciejowski, J. M., *Multivariable Feedback Design*, Addison-Wesley, Reading, MA (1989).

Maciejowski, J. M., and D. A. Vines, "Decoupled Control of a Macroeconomic Model Using Frequency Domain Methods," *J. Econ. Dyn. Contr.*, **1**, 55 (1984).

Moore, B. C., "Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction," *IEEE Trans. Automat. Contr.*, **AC-26**, 17 (1981).

Ober, R. J., "Balanced Realizations: Canonical Form, Parametrization, Model Reduction," *Int. J. Contr.*, **46**, 643 (1987).

Park, S., and W. F. Ramirez, "Optimal Regulatory Control of Bioreactor Nutrient Concentration Incorporating System Identification," *Chem. Eng. Sci.*, **45**, 3467 (1990).

Prett, D. M., and C. E. Garcia, *Fundamental Process Control*, Butterworths, London (1988).

Ramirez, W. F., *Process Control and Identification*, Academic Press, New York (1994).

Ramirez, W. F., K. P. Beyer, and H. Abedi, "Optimal Regulator Control for Systems with Time-Varying Measurable and Partially Measurable Disturbances," *Optim. Control Appl. Methods*, **8**, 159 (1987).

Safonov, M. G., and R. Y. Chiang, "A Schur Method for Balanced Truncation Model Reduction," *IEEE Trans. Automat. Contr.*, AC-**34**, 729 (1989).

Tsai, M., and Y. Shih, "Balanced Minimal Realization via Singular Value Decomposition of Sarason Operator," *Automatica*, **24**, 701 (1988).

Manuscript received Dec. 16, 1993, and revision received July 5, 1994.