Minimum Realizations and System Modeling: II. Theoretical and Numerical Extensions

The ideas and techniques discussed in Part I of this Review° are extended to the modeling of nonlinear lumped and distributed systems. Troublesome numerical and computational difficulties are examined, and recommendations for the alleviation of these difficulties are made. Particular emphasis is placed on the stability of system models constructed with the Tether method. Highly successful modeling of realistic systems are presented numerically.

R. H. ROSSEN and L. LAPIDUS

Department of Chemical Engineering
Princeton University
Princeton, New Jersey 08540

SCOPE

Numerous algorithms for the construction of state-space representations of linear systems have appeared in the recent literature. However, with only a few exceptions (Kallina, 1970; Rossen, 1972), the analysis of these procedures has been limited to the area of linear lumped parameter systems, and the reported application of the algorithms has been restricted to abstract systems. Therefore, the difficulties which arise in the realization of real physical processes have received little investigation.

Part I of this Review was devoted to a presentation of the basic concepts underlying the theory of minimum realizations and a summary of the publications relating to that theory and to the algorithms derived from it. In Part II, we will investigate the application of the partial realization procedure developed by Tether to the modeling of processes which have been studied in the literature. Observations will be made concerning the applicability of the procedure to different classes of chemical engineering systems, and techniques will be offered for the alleviation of some of the difficulties relating to that application.

The Tether algorithm is used because it was developed for the construction of the system model directly from the impulse response curve of the system and is therefore easily adapted to the modeling of nonlinear systems or of distributed systems. It has also been determined that the Tether procedure can be superior to the procedures developed by other researchers when the construction of an exact model for the system is impossible and when a partial realization is required.

SIGNIFICANCE AND CONCLUSIONS

The minimum realization algorithm originally proposed by Ho and Kalman (1966) and modified by Tether (1970) is shown to be an excellent tool for generating linear state-space representations of linear and nonlinear systems based solely on their responses to impulse forcing functions. In many respects, it is superior to the frequently encountered modeling techniques based on moment analysis and frequency response analysis. The algorithm requires less input/output data than the other methods and by treating all of the input/output pairs simultaneously, the algorithm is able to extract more information from the available data. Also by not specifying the dimension of the realization a priori, the algorithm retains greater flexibility than the moment methods.

It is suggested here that maximum effectiveness can be gained from the algorithm if the two input sequences of Markov parameters Y_k, $k = 0, 1, 2 \ldots$ (Ho-Kalman) and $k = -1, -2, \ldots$ (Bruni (1968)), are combined. The results presented here indicate that such a combination is superior to either of the sequences individually. The parameters corresponding to $k \geq 0$ ensure accuracy in the neighborhood of a steady state, while the parameters corresponding to $k < 0$ ensure an accurate overall response.

Perhaps the most attractive aspect of the realization algorithm is the simplicity of the required calculations. This is illustrated here with a number of detailed numerical examples on a variety of types of systems. Only linear algebra techniques are involved and the solution does not require an iterative procedure. This feature suggests that the approach may be extremely useful in the on-line modeling of processes.

PRELIMINARIES

The Tether algorithm was discussed in detail in Part I, but here we will outline explicitly the procedure that would be followed if an operating physical process were to be modeled. First, the input/output variables of interest are designated as the input/output variables for the model. Secondly, an impulse is imposed in each of the input variables and the response in each of the output variables is measured. From these response curves, the moments can

° Part I of this paper was published in the July, 1972, *AIChE Journal*.

be computed numerically and the Markov parameters can be obtained by fitting the curve with a polynomial in time. The Markov parameters are the coefficients of a power series representation of this impulse curve, so they can be determined from the experimental data. Recall that the response of a linear system to an impulse of unit area is

$$\mathbf{Y}(t) = \sum_{k=0}^{\infty} \mathbf{H}\mathbf{F}^k\mathbf{G}t^k/k!$$

from which the Markov parameters can be obtained. The moments can be computed by numerical integration over the curve to obtain

$$(-1)^{k+1}\mathbf{H}\mathbf{F}^{-k-1}\mathbf{G}k! = \int_0^{\infty} t^k\mathbf{Y}(t)\,dt$$

These parameters are then used as input to the Ho-Kalman algorithm.

The input to the Tether program will be the sequence of $\mathbf{HF}^k\mathbf{G}$. The $\mathbf{S}_r$ matrix will be formulated as shown in Part I. Construction of the matrices $\mathbf{P}$ and $\mathbf{Q}$ will parallel the procedure, outlined by Andree (1950), for computation of the inverse of a matrix. Andree formed the inverse of a square matrix $\mathbf{S}_r$ by manipulating the rows and columns of the block matrix

$$\begin{bmatrix} \mathbf{I} & \mathbf{S}_r \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

to obtain the matrix

$$\begin{bmatrix} \mathbf{P} & \mathbf{I} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}$$

He then showed that $\mathbf{QP} = \mathbf{S}_r^{-1}$. Note that

$$\mathbf{QP} = \mathbf{S}_r^{-1}$$

$$\mathbf{S}_r\mathbf{QP} = \mathbf{I}$$

$$\mathbf{PS}_r\mathbf{Q} = \mathbf{I}$$

Similarly, if the square matrix $\mathbf{S}_r$ is singular, it can be shown that the row and column calculations that change

$$\begin{bmatrix} \mathbf{I} & \mathbf{S}_r \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{P} & \mathbf{I}_s \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}$$

will generate a $\mathbf{P}$ and $\mathbf{Q}$ so that $\mathbf{PS}_r\mathbf{Q} = \mathbf{I}_s$ is a pseudo-identity matrix of rank s. The matrices $\mathbf{J}$, $\mathbf{E}_m$, $\mathbf{E}_p$ will be defined as in the derivation of the Ho-Kalman algorithm (see Part I).

PARTIAL REALIZATION EXAMPLE

To gain some insight into the advantages and peculiarities of the partial realization method, we will attempt to model a specific linear system. The most popular technique for modeling systems based on their response to the simple impulse forcing function is the moment method of analysis. The basic concepts in this technique can be found in Davison (1968) and Kropholler (1970). Because the moment method and the Tether method treat the same class of problems, it would be helpful to compare them for accuracy and utility in the analysis of a specific system. With this purpose in mind, a numerical problem will be posed here and solved by both methods. The models obtained from the two approaches to the problem will be compared with each other and with the original system. This example should aid us in indentifying some of the important aspects of the use of the partial realization algorithm.

Consider the system $\{\mathbf{F}, \mathbf{G}, \mathbf{H}\}$ such that

$$\mathbf{F} = \begin{bmatrix} -0.5 & 0 & 0 & 0 & 0 \\ 0 & -0.4 & 0 & 0 & 0 \\ 0 & 0 & -0.3 & 0 & 0 \\ 0 & 0 & 0 & -0.2 & 0 \\ 0 & 0 & 0 & 0 & -0.1 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 1.0 & 0 & 0 \\ 1.0 & 0 & 1.0 \\ 1.0 & 0 & 0 \\ 1.0 & 1.0 & 0 \\ 1.0 & 0 & 1.0 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} 1.0 & 0 & 1.0 & 1.0 & 0 \\ 0 & 0 & 0 & 1.0 & 1.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix}$$

For simplicity, $\mathbf{F}$ is chosen to be a diagonal matrix. Any triple can be transformed into a triple with a diagonal $\mathbf{F}$ matrix if the original $\mathbf{F}$ matrix has distinct eigenvalues.

The matrices $\mathbf{G}$ and $\mathbf{H}$ are also carefully selected. The transfer function is $\mathbf{H}(s\mathbf{I} - \mathbf{F})^{-1}\mathbf{G}$. In this instance, therefore, the transfer function, $\overline{\mathbf{Y}}(s)$, is

$$\overline{\mathbf{Y}}(s) = \begin{bmatrix} \dfrac{1}{(s+.5)(s+.3)(s+.2)} & \dfrac{1}{s+.2} & 0 \\ \dfrac{1}{(s+.2)(s+.1)} & \dfrac{1}{s+.2} & \dfrac{1}{s+.1} \\ \dfrac{1}{(s+.5)(s+.4)(s+.3)(s+.2)(s+.1)} & \dfrac{1}{s+.2} & \dfrac{1}{(s+.2)(s+.4)} \end{bmatrix}$$

This transfer function allows the simultaneous study of problems of several different dimensions. The element $\overline{Y}_{11}(s)$ can be viewed as the transfer function of a 1×1 system with three eigenvalues

$$\mathbf{F} = \begin{bmatrix} -0.5 & 0 & 0 \\ 0 & -0.3 & 0 \\ 0 & 0 & -0.2 \end{bmatrix}$$

$$\mathbf{G} = [1.0 \quad 1.0 \quad 1.0]^T$$

$$\mathbf{H} = [1.0 \quad 1.0 \quad 1.0] \ .$$

That is, for this smaller system, $\mathbf{H}(s\mathbf{I} - \mathbf{F})^{-1}\mathbf{G} = \overline{Y}_{11}(s)$. The effect of $u_1(t)$ on the measurement variable $y_1(t)$ in the system described by $\overline{\mathbf{Y}}(s)$ will be identical to the corresponding effect in the three eigenvalue system. Likewise, the element $\overline{Y}_{12}(s)$ can be viewed as a smaller system dependent on only one eigenvalue, while $\overline{Y}_{31}(s)$ can be considered to represent a system of five eigenvalues. In this way, the study of this one system can yield valuable information about systems of various dimensions.

For the purposes of this investigation, it will be assumed initially that the empirical moments and Markov parameters would contain negligible measurement error. Therefore, they would be identical to those computed directly from the triple $\{\mathbf{F}, \mathbf{G}, \mathbf{H}\}$. For this problem and for the other problems treated in the present work, the impulse

response information will be determined analytically, while in the analysis of an existing physical system, experimental methods would generate this information. The Markov parameters are

$$\mathbf{HG} = \begin{bmatrix} 3 & 1 & 0 \\ 2 & 1 & 1 \\ 5 & 1 & 2 \end{bmatrix}$$

$$\mathbf{HFG} = \begin{bmatrix} -1.0 & -0.2 & 0 \\ -0.3 & -0.2 & -0.1 \\ -1.5 & -0.2 & -0.5 \end{bmatrix}$$

$$\mathbf{HF^2G} = \begin{bmatrix} 0.38 & 0.04 & 0.0 \\ 0.05 & 0.04 & 0.01 \\ 0.55 & 0.04 & 0.17 \end{bmatrix}$$

$$\mathbf{HF^3G} = \begin{bmatrix} -0.160 & -0.008 & 0 \\ -0.009 & -0.008 & -0.001 \\ -0.225 & -0.008 & -0.065 \end{bmatrix}$$

$$\mathbf{HF^4G} = \begin{bmatrix} 0.0732 & 0.0016 & 0 \\ 0.0017 & 0.0016 & 0.0001 \\ 0.0989 & 0.0016 & 0.0257 \end{bmatrix}$$

Following Tether's analysis, we see that

$$\rho\mathbf{H}_{2,2} = \rho\mathbf{H}_{3,2} = \rho\mathbf{H}_{2,3} = 6$$

It follows that four parameters are required for an exact Ho solution. Therefore, it will be assumed that only three parameters are available to the Tether algorithm so that an exact solution cannot be constructed.

For the moment method analysis, each of the impulse response elements will be treated separately. The system will be modeled by fitting different common forms of the transfer function to the first three moments (area, mean, and variance). The transfer functions are those utilized frequently in the literature to model chemical engineering processes (Buffham and Gibilaro, 1970; Davison, 1968; Kropholler, 1970; Midoux and Charpentier, 1970). They are, specifically,

$$\begin{aligned} \overline{Y}_A &= \frac{\exp(-\tau_1 s)}{(1+\tau_2 s)}\,\tau_3 \\ \overline{Y}_B &= \frac{\exp(-\tau_1 s)}{(1+\tau_2 s)(1+\tau_3 s)}\,\tau_4 \\ \overline{Y}_C &= \frac{1}{(1+\tau_1 s)(1+\tau_2 s)}\,\tau_3 \\ \overline{Y}_D &= \frac{1.0 + \tau_2\tau_3/(\tau_2+\tau_3)}{(1+\tau_2 s)(1+\tau_3 s)}\,\tau_1 \end{aligned} \qquad (1)$$

where the τ's are constants which must be evaluated in the analysis of the system. Note that the second form $\overline{Y}_B$ requires four moments for the evaluation of its four constants. In each case, the values of the unknown constants are computed by matching the analytically (or empirically) determined moments with the moment expressions derived from each of the estimated transfer functions, in terms of the τ's. For example, in case B, the moments are computed as

$$\begin{aligned} \text{Area} &= \mu_0 = \tau_4 \\ \text{Mean} &= \mu_1 = (\tau_1+\tau_2+\tau_3)\tau_4 \\ \mu_2 &= (\tau_2^2+\tau_3^2+\mu_1^2)\tau_4 \\ \mu_3 &= (3\mu_1\mu_2 + 2\tau_1^3 + 2\tau_2^3 - 2\mu_1^3)\tau_4 \end{aligned}$$

The τ's can then be obtained from the measured moments. In some instances, the moments have values such that an exact real solution for the τ's does not exist. In these cases, the τ's are evaluated so that the sum

$$\begin{aligned} &(\tau_4-\mu_0)^2 + [\mu_1 - (\tau_1+\tau_2+\tau_3)\tau_4]^2 \\ &+ [\mu_2 - (\tau_2^2+\tau_3^2+\mu_1^2)\tau_4]^2 \\ &+ [\mu_3 - (3\mu_1\mu_2 + 2\tau_1^3 + 2\tau_2^3 - 2\mu_1^3)]\tau_4 \end{aligned}$$

is minimized. In this manner, models are obtained for each element in the transfer function.

Three mathematical representations will be constructed via Tether's method. Recall that the partial minimum realizations can be constructed either from the moments or from the Markov parameters of the impulse response. For this example problem, one model will be constructed from the first three moments and two additional models will be constructed from the first three Markov parameters. In this way, the advantages of each procedure can be studied when these three models are compared with each other and with the models obtained from the moment methods. Each of the models formulated by the minimum partial realization approach and each of the models formulated by the moment approach will be used to predict the response of the system to a step input forcing function. The predicted responses will be compared with the actual system response for each element of the transfer function matrix.

The response of a linear system, initially at $\mathbf{x}(0) = \mathbf{0}$, to the forcing function $\mathbf{u}(\tau)$, was given as Equation (6) in Part I

$$\mathbf{y}(t) = \mathbf{H}\int_0^t \exp(\mathbf{F}(t-\tau))\mathbf{G}\mathbf{u}(\tau)\,d\tau \qquad (2)$$

To obtain the response to a step function, the constant 1.0 is substituted for the appropriate element in $\mathbf{u}(\tau)$. If step inputs are introduced in each of the elements of $\mathbf{u}(\tau)$, (2) becomes

$$\begin{aligned} \mathbf{y}(t) &= -\mathbf{H}\,\mathbf{F}^{-1}\exp(\mathbf{F}(t-\tau))\mathbf{G}\Big|_0^t \\ \text{and} \\ \mathbf{y}(t) &= -\mathbf{HF}^{-1}\mathbf{G} + \mathbf{HF}^{-1}\exp(\mathbf{F}t)\mathbf{G} \end{aligned} \qquad (3)$$

The step forcing function is chosen as a test of the accuracy and reliability of each of the models because it corresponds to a frequently encountered input change; precedents for this usage of the step forcing function are found in the chemical engineering literature (Gibilaro and Lees, 1969; Lapidus and Amundson, 1952). Other forcing functions would be expected to give the same qualitative degree of correlation between the models and the real system.

In order to facilitate the presentation of the results of this comparison of modeling techniques, an identifying letter will be assigned to each model. The models obtained from the moment methods will be labeled A, B, C, and D as before. The two partial realizations constructed from the Markov parameters will be designated as E and F and the partial realization constructed from the moments will be designated as G. The actual system step response will be assigned the letter H.

The complete set of transfer function models is not presented here but can be found in Rossen (1972). One representative model, solution A, is

$$\overline{\mathbf{Y}}_A = \begin{bmatrix} \dfrac{31/3\,\exp(1.84s)}{(1+4.151s)} & \dfrac{5.0}{1+5s} & 0 \\ \dfrac{15\,\exp(0.6416s)}{(1+8.975s)} & \dfrac{5.0}{1+5s} & \dfrac{10}{1+10s} \\ \dfrac{22.8\,\exp(1.51s)}{(1+0.792s)} & \dfrac{5.0}{1+5s} & \dfrac{12.5\,\exp(0.98s)}{(1+9.48s)} \end{bmatrix}$$

Likewise, a representative model found by Tether's approach is the partial realization found by method G,

$$\mathbf{HF}^{-1} = \begin{bmatrix} -10.333 & 0 & 0 & 0 & 0 \\ -15.0 & 2.2581 & 0 & 0 & 0 \\ -22.833 & 6.0484 & 14.286 & 0 & 0 \end{bmatrix}$$

$$= \mathbf{F} \begin{bmatrix} -0.27796 & 0.037722 & 0.16705 & 0.22203 & -0.019068 \\ -0.95771 & 0.263411 & 1.60939 & 2.1393 & -0.18395 \\ 0.30370 & -0.14695 & -0.69828 & -0.39643 & 0.034033 \\ -0.044562 & 0.021563 & 0.052556 & -0.36499 & 0.031398 \\ -0.041178 & 0.019925 & 0.048238 & -0.34003 & 0.028294 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 1.0 & 0.48387 & 0 \\ 0 & 1.0 & -4.4286 \\ 0 & 0 & 1.0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

When a step forcing function is applied to each of the models, it is observed that the accuracy of the predicted responses varies widely between the input/output elements $\overline{Y}_{ij}(s)$ in $\overline{\mathbf{Y}}(s)$. Each element in $\overline{\mathbf{Y}}(s)$ is equivalent to a 1×1 system. In each model constructed for this example system, those elements which are equivalent to systems with fewer eigenvalues are more accurately represented. For example, all seven of the models give an accurate prediction for those elements, such as $\overline{Y}_{12}(s)$, which are equivalent to systems of one eigenvalue. For these input/output elements, the appropriate choice of the unknowns in any of the moment methods will produce an exact solution. That is, each of the selected transfer functions has, as a special case, the precise transfer function for a linear system with one eigenvalue. The three Tether models also accurately predict the step response. Thus, it can be concluded that, for this simple one-eigenvalue system, there is no advantage to either type of approach.

The predictions of the step responses of input/output elements which depend on larger numbers of eigenvalues are less accurate. The element $\overline{Y}_{21}$, in $\overline{Y}(s)$, is equivalent to a system of dimension two. The results of the study on $\overline{Y}_{21}$ are illustrated graphically in Figure 1. The Tether method based on three moments and the moment method D again accurately predict the response. The Tether method based on Markov parameters is less effective in the later portion of the curve but approximates the response well in the neighborhood of the origin. The other moment methods are clearly unacceptable, even for this small uncomplicated system. Method D provides a good fit of the data because, once again, it can exactly represent any two-dimensional linear system. Note also that the Tether methods all provide models which describe the system very well in the neighborhood of the origin (for the first half of the response curve), but that the Tether method based on moments is clearly superior for response prediction at long times, as the new steady state is approached.

Element $\overline{Y}_{31}$, representing the five-eigenvalue system, provides the most useful comparison because it is the most complex problem. The predicted step responses for this element have been plotted in Figure 2 and are tabulated in Table 1. Moment method D can no longer give an accurate prediction of the step response because, with only three parameters, it is not able to handle this more complex system. However, it is still superior to the moment techniques which led to models A, B, and C whose predictions are totally unacceptable. As in the previous cases, the early response was most successfully handled by the Tether solutions formulated from the Markov parameters. The best overall model for the most complex system appears to be the Tether method G which utilizes the moments as suggested in the paper by Bruni (see Part I).

Several interesting conclusions can be drawn from this study of a linear system. The most important of these will be summarized here. First, if sufficient information is available for an exact minimum realization solution, that solution technique is, of course, superior to the moment

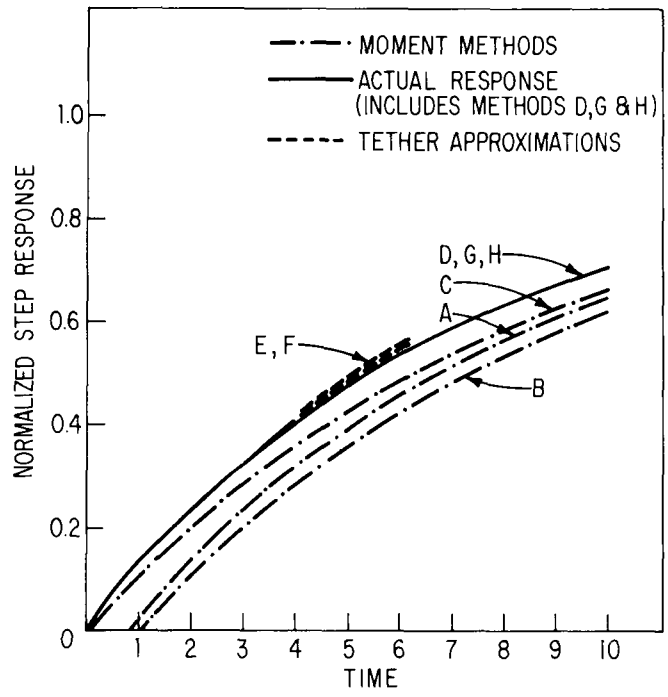

Fig. 1. Two-eigenvalue linear problem.

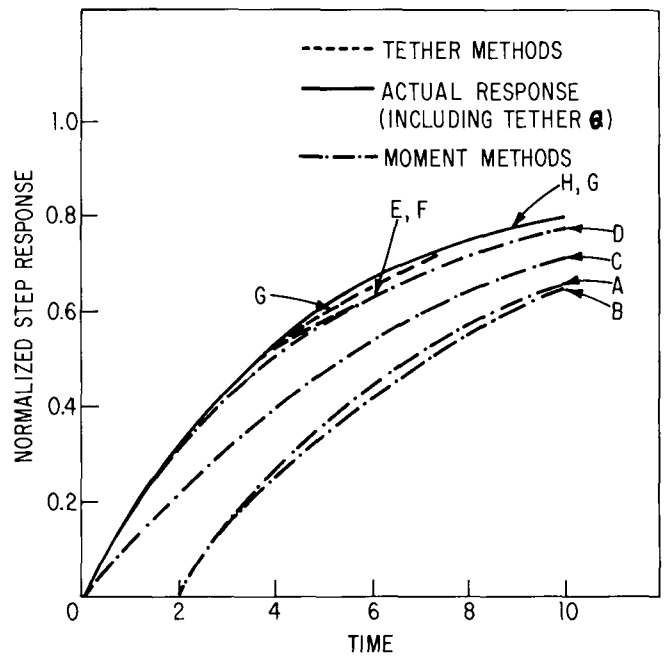

Fig. 2. Five-eigenvalue linear problem.

Table 1. Step Response for $\overline{Y}_{31}(s)$ in Linear Example System

Time	Solutions A	B	C	D	E	F	G	H
0	0	0	0	0	0	0	0	0
1	0	0	0.117	0.173	0.189	0.189	0.189	0.190
2	0.060	0.040	0.220	0.304	0.331	0.331	0.331	0.334
3	0.171	0.154	0.310	0.407	0.437	0.437	0.442	0.443
4	0.270	0.252	0.390	0.490	0.519	0.519	0.530	0.530
5	0.432	0.417	0.524	0.613	0.622	0.624	0.656	0.659
10	0.657	0.645	0.710	0.774	—	—	0.798	0.796

methods. At the same time, it is a numerically simple way of obtaining an exact model. Second, if an exact solution is impossible, and if the original system is of sufficiently small dimension, the moment method and the minimum partial realization method are both effective and there is no substantial superiority to either technique. However, as the dimension of the system increases, the moment techniques lose their effectiveness and the Tether techniques give rise to more useful models. For linear systems, as would be expected, the most advantageous choice for a transfer function to fit the data by moment analysis is of the form,

$$\overline{Y}(s) = \sum_{i=1}^{k} \frac{\alpha_i}{(1 + \tau_i s)}$$

This is the general form of the linear transfer function. The constants must be determined empirically. It is, however, exact only for systems of dimension $n \leq k$ and such an exact solution would require the accurate measurement of $2n$ moments. The partial minimum realization solution, however, can be used to construct an exact model using as few as $2n/\xi$ moments, where ξ is defined as the minimum of p and m, the number of input and output variables respectively. The number $2n$ is an upper bound for the number of moments which are required for a solution. It is clear that $2n/\xi$ is a lower bound because if $2n/\xi$ moments are used, then the smaller dimension (number of rows or columns) of the Hankel matrix is $2n/\xi \times \xi/2 = n$. Specifically, if N_0 is the number of moments required for an exact realization, then

$$2n/\xi \leq N_0 \leq 2n \qquad (4)$$

The Tether solution, then, requires less information than the moment method. This is because the partial minimum realization technique operates on all of the input/output pairs simultaneously. In this way, valuable information concerning the interaction of these pairs is utilized—information which is not utilized by the moment methods. For linear systems, this additional information means that fewer moments are required for an exact model or that a more comprehensive and a more accurate model can be constructed from a given number of moments.

The study of linear systems also indicates that the two varieties of partial realizations have different properties and that either can be of value depending on the desired model characteristics. If the model is constructed from the sequence of Markov parameters, it will be a very accurate representation of the system in the neighborhood of the origin. Responses to other forcing functions, predicted from such a model, will be very accurate for short times. This result is not unexpected since the Markov parameters are essentially the local time derivatives around $t = 0$. The predicted responses are less dependable at later times, and the predicted step response can be expected to yield an erroneous value for the new steady state. This method is then most advantageous when the model will be required to predict accurately the initial slope of the response to various forcing inputs.

The second variety of minimum partial realization consists of models constructed from the moments of the impulse response. This type of solution is less accurate in the prediction of responses near the origin but can provide a more useful overall prediction of the system response because it is a better fit of the overall impulse response. With this technique, there will in general be no steady state error in the step response. In the sample problem, the method formulated models which could predict the entire step response with a maximum error of less than 2% over the range of the curve.

For linear systems, then, the theory of minimum realization can be very useful. There are, however, difficulties which arise in the application of these techniques. In the earlier sections, it was noted that the Tether method is not a precise algorithm and does not yield a unique solution. Instead, it provides a set of mathematical guidelines for the solution of the realization problem. Many different solutions to the same problem are therefore often possible using the Tether approach. All of the solutions will fit the given matrices exactly, but some of these solutions may be found to be unstable.

Moments with Measurement Error

Assume now that the moments in the previous problem contain some measurement error. This situation can be approximated by randomly altering the moments by approximately 1%. The solution of this new problem is an accurate mathematical description of the erroneous moments and thus is found to differ from the previous Tether model by about 1%. The accuracy of the Tether solution, of course, depends on the accuracy of the moments. This is preferable, however, to the accuracy which can be expected from the moment methods. We have illustrated that the Tether models are generally more accurate than the moment method representations when a linear system is modeled from exact moments. Therefore, they would be expected to be more accurate in the modeling of a partial sequence of erroneous moments. The moment approach also attempts to fit the erroneous moments but does so less effectively. The resulting model is an approximate fit to approximate data, which is less desirable than an accurate fit of approximate data.

It should also be noted that, since the measurement of moments and Markov parameters will always be susceptible to error, there will seldom be any cancellation in the rows and columns of the Hankel matrix, defined in Part I. That is, if the available information consists of $N_0 n \times n$ moments, the Tether solution will, in general, contain $nN_0/2$ eigenvalues. There are two important consequences of this result. First, the advantage enjoyed by partial realization techniques over the moment methods can be expected to be large in real systems. Second, the choice of matrices to fill out the Hankel matrix, if any are required, will be highly arbitrary.

Observations

From this investigation into the modeling of linear systems, several observations can be made concerning the solution of linear partial realization problems. These observations derive from the mechanics of the Tether algorithm and they will be stated briefly here. As before, p, m, and n are defined to be the number of input, output, and state variables in the model. It is found that if $p = m$, any even number of moments can be used to construct a minimum partial realization without the need to estimate additional moments. Alternately, if the number of moments is odd, any one additional moment matrix is sufficient to construct a partial minimum realization and the choice of the elements of this matrix is completely arbitrary.

Similarly, if $p > m$ or if $p < m$, the solution methods will be mirror images. The same amount of information is available in both cases and the mechanics of the solution method does not differentiate mathematically between the two solutions.

It has been shown that Tether's method is, in general, superior to moment techniques and that the superiority increases with increasing system complexity. For systems with measurement error, either the Ho-Kalman algorithm or the Tether extension to the algorithm will fit the approximate moments exactly.

NONLINEAR SYSTEMS

The procedure for modeling nonlinear systems would be identical to that used for linear systems. Once again, a partial sequence of $\mathbf{HF}^k\mathbf{G}$ would be determined from the impulse response and the realization would be constructed from this partial sequence. The models would be essentially linearizations of the system around the steady state and would therefore be expected to be less accurate than models of linear systems. The literature has few examples of realization models for nonlinear chemical engineering systems, but the models have been found to be quite satisfactory in the cases tested. We will present one interesting example of a nonlinear system here.

A Numerical Example

Consider a nonlinear process, taken from a paper by Kallina (1970): a nonisothermal stirred tank reactor with a first-order irreversible exothermic reaction $A \rightarrow B$. Energy and mass balances on the contents of the reactor yield the equations:

$$V\rho\, c_{fr} dT/dt = G_f c_f (T_{in} - T) + V\rho\gamma(\Delta H) k_0 \exp(-E/RT)$$

$$V\rho\gamma dC/dt = G_f(C_{in} - C) - V\rho\gamma k_0 \exp(-E/RT)$$

where T and C are the reactor temperature and concentration of species A and T_{in} and C_{in} are the inlet temperature and concentration of species A. Also, G_f is the inlet mass flow rate, c_f and c_{fr} are the specific heats of the inlet stream and of the reactor contents, γ is the fraction of the reactor volume occupied by the fluid, and ρ is the density of the inlet stream. Note that the equations are highly nonlinear. The object is to replace this nonlinear mathematical description with a linear model which will describe the system in the neighborhood of the initial steady state. Kallina estimated the values of the parameters and, with a numerical integration procedure, obtained the steady state operating conditions of $T_{SS} = 1293.3$ and $C_{SS} = 0.169$, corresponding to the inlet conditions of $T^\circ = 1109.7$ and $G^\circ = 0.189$. Deviation variables are defined to be

$$y_1(t) = T(t) - T_{SS}$$

$$y_2(t) = C(t) - C_{SS}$$

and

$$u_1(t) = T_{in} - T^\circ$$

$$u_2(t) = G_f - G^\circ$$

To construct a transfer function, Kallina applied an impulse of $u_1(t) = 10^0$ for $0 \leq t \leq 2$ to the system which was originally at $x_1(0) = x_2(0) = 0$. The problem was analyzed for constant flow rate ($u_2(t) = 0$). Techniques developed by Liu (see Kallina, 1970) were used to construct the transfer function

$$\overline{\mathbf{Y}}(s) = \begin{bmatrix} \dfrac{35.5}{(s+0.747)(s+22.4)} \\ \dfrac{-0.00831}{(s+0.747)(s+4.52)} \end{bmatrix}$$

The minimum realization method of Gilbert (see Part I) was applied to $\overline{\mathbf{Y}}(s)$ to obtain a realization:

$$\mathbf{F} = \begin{bmatrix} -22.4 & 0 & 0 \\ 0 & -4.52 & 0 \\ 0 & 0 & -0.747 \end{bmatrix}$$

$$\mathbf{G} = [-1.64 \quad 0.0022 \quad -0.0022]^T$$

$$\mathbf{H} = \begin{bmatrix} 1.0 & 0 & -745.0 \\ 0 & 1.0 & 1.0 \end{bmatrix}$$

In the present analysis, the transfer function step will be eliminated, and a partial minimum realization will be constructed directly from the impulse response.

Kallina assumed that the flow rate G_f was constant, but in the following analysis it will be treated as another control variable. This modification is made in order to derive maximum benefit from the Tether method's simultaneous treatment of all input/output pairs. The same impulse in $u_1(t)$ will be applied to the system but, in addition, an impulse in the flow rate $u_2(t) = 0.01$ will be applied. The impulse response will be computed by Euler's forward difference method with small time steps. The moments of this response are used to construct a partial realization. The details can be found in Rossen (1972) and the resulting model is

$$\mathbf{F} = \begin{bmatrix} -1.3697 & -0.13086 & 0.006372 & 0 \\ 0.69713 & -0.21411 & -0.020473 & 0.000714 \\ 1.0 & -0.20417 & -1.0842 & -0.11631 \\ 0 & 1.0 & -1.7695 & -4.6128 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 0.05 & -31.3 \\ 0 & 1.0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{H} = \begin{bmatrix} -21.255 & 0 & 0 & 0 \\ 0.02529 & -0.006883 & 0 & 0 \end{bmatrix}$$

The responses of the reactor temperature y_1 and the concentration of A, y_2, to a step change in the inlet temperature u_1 are estimated from this model and the one obtained by Kallina. The estimated responses and the actual response are illustrated in Table 2. Obviously, the construction of the partial realization directly from the impulse data is superior, in this case, to the two-step realization procedure involving the determination of a transfer function and the subsequent construction of a model to fit the transfer function.

This superiority is due to a greater flexibility in the direct minimum partial realization technique. By a priori assuming a form for the transfer function, Kallina restricted

TABLE 2. NONLINEAR CSTR SYSTEM (STEP RESPONSE)

	Time	System	Minimum realization	Kallina
Temperature	0.01	1.17	1.21	0.93
	0.2	2.39	2.44	2.32
	0.3	3.61	3.66	3.68
	0.4	4.81	4.84	4.94
	0.5	5.95	5.97	6.11
	0.6	7.04	7.05	7.20
	0.8	9.03	9.02	9.15
	1.0	10.75	10.74	10.82
Concentration	0.2	−0.00111	−0.00123	−0.00119
	0.4	−0.00343	−0.00351	−0.00354
	0.6	−0.00604	−0.00603	−0.00609
	0.8	−0.00855	−0.00849	−0.00851
	1.0	−0.01080	−0.01075	−0.01068
	1.2	−0.01274	−0.01275	−0.01259

the domain from which the mathematical model could be drawn. The direct construction of the model allows greater flexibility in its form and dimension.

SPECIAL NONLINEAR SYSTEMS

The partial realization approach has been shown to be a valuable tool in the modeling of frequently encountered chemical engineering systems. At this point, we will briefly discuss the modeling of some unique and interesting special classes of nonlinear systems. This discussion should help to give a better understanding of the applications and drawbacks of the approach. Some techniques which serve to minimize the computational difficulties involved in the procedure will be offered. A detailed analysis of these aspects can be found in Rossen (1972), and they will be only touched on here.

The most important restriction on the use of the minimum partial realization algorithm is that it is essentially a linearization. As such, it is subject to all of the limitations placed on linearizations. It is strictly accurate only in the neighborhood of the steady state. The algorithm is not applicable to systems operating at unstable steady states because any perturbation in the inlet variables will result in a change in the steady state. In addition, if the system contains multiple steady states, the algorithm is not able to predict system movements between steady states.

Another drawback of the algorithm is the direct dependence of the Markov parameters on the coefficients of the polynomial used to fit the impulse response. These coefficients vary with the order of the polynomial. Since the Tether algorithm defines a model which exactly fits a specific set of parameters, each set of parameters would generate a different realization. It has been shown that the use of Chebyshev (orthogonal) polynomials to fit the impulse curve can significantly improve the accuracy and reproducibility of the parameters and thus of the model (Rossen, 1972).

We should say a few words about the modeling of systems operating in limit cycles. A limit cycle does not satisfy the realization requirement of an initial steady state $x(0) = 0$. Therefore, it cannot be analyzed by the usual procedure. However, if a new variable is defined such that $\xi[x_1(0), x_2(0) \ldots] = 0$, the new variable can be analyzed by partial realization techniques to predict future behavior of the limit cycle. Then, while precise variations in the state could not be predicted, variations in the limit cycle containing the state could be predicted.

INSTABILITY CONSIDERATIONS

The most serious weakness in the minimum partial realization method, as introduced by Tether, is the possibility that the resulting mathematical model will contain a positive eigenvalue. The model would then be unstable and would be inadequate for the prediction of future system behavior. For example, the step response for a linear constant-coefficient system [Equation (3)] is

$$\mathbf{HF}^{-1}\mathbf{G} + \mathbf{HF}^{-1} \exp(\mathbf{F}t)\mathbf{G} \ .$$

If the $\mathbf{F}$ matrix has a positive eigenvalue, the second term will not decay to zero but will grow exponentially. Therefore, a linear model which contains an eigenvalue with a positive real part can never predict a stable response to the step input. The same is found for other common input forcing functions, including the impulse, indicating that such a model would predict an unstable response to any perturbation, regardless of size.

In order to construct a realization of a given sequence of moments or Markov parameters $\mathbf{Y}_k$, the Tether method requires the selection of an extension sequence of $\mathbf{Y}_k$ to fill the Hankel matrix. Because the method generates a solution which is not unique, and because the method merely provides guidelines for the selection of the extension sequence, the subsequent model is often determined by the arbitrary selection of this sequence and, in many instances, the resulting model is unstable. Indeed, if the model is of high dimension, containing numerous distinct eigenvalues, it will be more likely to be unstable than to be stable. It would be advantageous, then, to be able to ensure a priori that the model will be stable or to be able to alter the Tether model so that a stable system representation can be constructed.

Since the mathematical model is dependent on the unspecified Markov parameters which must be selected in adherence with the Tether guidelines, it would seem logical to attempt to restrict the choice of these parameters so that the model will be stable. Recall that the $\mathbf{F}$ matrix is defined in the Ho-Kalman algorithm to be

$$\mathbf{U}_S[\mathbf{JP}(\tau\mathbf{S}_r)\mathbf{QJ}]\mathbf{U}_S^T \ .$$

The matrices $\mathbf{U}_S$, $\mathbf{J}$, and $\mathbf{U}_S^T$ are added to eliminate extraneous rows and columns from the matrix, so the essential element of the $\mathbf{F}$ matrix is $\mathbf{P}(\tau\mathbf{S}_r)\mathbf{Q}$. Remember, also, that $\mathbf{S}_r$ is a block matrix formed by the Markov parameters, which are $p \times m$ matrices, in this manner:

$$\mathbf{S}_r = \begin{bmatrix} \mathbf{Y}_0 & \mathbf{Y}_1 & \cdots & \mathbf{Y}_N \\ \mathbf{Y}_1 & \mathbf{Y}_2 & \cdots & \mathbf{Y}_{N+1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{Y}_{N'} & \mathbf{Y}_{N'+1} & \cdots & \mathbf{Y}_{N+N'} \end{bmatrix}$$

Note that the form of $\mathbf{S}_r$ depends on the values of p, m, N, and N' and that this form can vary widely from problem to problem. Therefore, the effect of the selected matrices on the eigenvalues of $\mathbf{F}$ will differ greatly between problems and it would be very difficult, if not impossible, to develop a single formula for determining the needed sequence of matrices $\mathbf{Y}_k$ for all situations. In this discussion, a method which can be used to construct a stable model from the Tether unstable model and which will ensure that the new model will still fit the given $\mathbf{Y}_k$ exactly will thus be offered.

The initial step in the treatment of an unstable system is the isolation of the positive eigenvalue and the analysis of a smaller system with only one eigenvalue. The original model is partitioned into a stable segment, containing the eigenvalues with negative real parts, and an unstable seg-

ment, containing the eigenvalues with positive real parts. Only the unstable segment is of interest, so it is analyzed independently. The stable segment is neglected in this treatment. Assume that the number of given Markov parameters is ϕ. The positive eigenvalue will then be replaced by ϕ negative eigenvalues without altering the corresponding sequence of parameters. The negative eigenvalues are totally arbitrary and can be selected at random. Define the vector $\mathbf{h}_i$ to be the ith column of $\mathbf{H}$ and the vector $\mathbf{g}_i$ to be the ith row of $\mathbf{G}$. Also, assume that λ_j is the positive eigenvalue and that the original system is of dimension n. The stable system representation will be of dimension $(n + \phi - 1)$. The new $\mathbf{F}$ will be a diagonal matrix consisting of the old and new negative eigenvalues. The new $\mathbf{H}$ matrix will consist of the original columns plus $(\phi - 1)$ additional columns, each identical to $\mathbf{h}_j$. The new $\mathbf{G}$ matrix will consist of the original $(n - 1)$ rows, excluding the jth row and will have ϕ new rows of the form, $\alpha_i \mathbf{g}_j$, $1 \leq i \leq \phi$.

When the α's have been evaluated, the problem will be solved. The first ϕ Markov parameters of this new model must match the given sequence of matrices. This requirement will be met if:

$$\sum_{i=1}^{\phi} \alpha_i \bar{\lambda}_i^k = \lambda_j^k \qquad 0 \leq k \leq -1$$

where $\bar{\lambda}_i$ is the ith negative replacement eigenvalue.

In general, this equation can be expressed in the form:

$$\mathbf{Aa} = \mathbf{\Lambda} \tag{5}$$

where

$$\mathbf{a} = [\alpha_1 \quad \alpha_2 \quad \ldots \quad \alpha_\phi]^T$$

$$\mathbf{\Lambda} = [1 \quad \lambda_j \quad \ldots \quad \lambda_j^{\phi-1}]^T$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ \bar{\lambda}_1 & \bar{\lambda}_2 & \bar{\lambda}_\phi \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \bar{\lambda}_1^{\phi-1} & \bar{\lambda}_2^{\phi-1} & \bar{\lambda}_\phi^{\phi-1} \end{bmatrix}$$

This matrix equation can be solved by a number of numerical methods, including Gaussian elimination, Gaussian elimination with pivoting, and numerous matrix inversion routines. However, because of the unique form of the $\mathbf{A}$ matrix, the quickest and most accurate solution technique is based on the determinant of $\mathbf{A}$.

Equation (5) has the solution

$$\alpha_k = \frac{D_k(\mathbf{A})}{\text{Det } \mathbf{A}}$$

where $D_k(\mathbf{A})$ is the determinant of the matrix $\mathbf{A}$ when the kth column is replaced by $\mathbf{\Lambda}$. In this particular instance, because of the form of $\mathbf{A}$, the solution reduces to

$$\alpha_k = (-1)^{\phi-1} \left[\prod_{i \neq k} (\bar{\lambda}_i - \bar{\lambda}_k) \right]^{-1}$$

$$\left[\alpha_\phi - \left(\sum_{i \neq k} \bar{\lambda}_i \right) \alpha_{\phi-1} + \left(\sum_{\substack{i>j \\ i,j \neq k}} \bar{\lambda}_i \bar{\lambda}_j \right) \alpha_{\phi-2} + \ldots + \left(\prod_{i \neq k} \bar{\lambda}_i \right) \alpha_1 \right] \tag{6}$$

The proof is found in Rossen (1972).

The stable model obtained by the method outlined here will satisfy one requirement placed on the minimum partial realization in that it will be controllable and observable. It will, however, necessarily be of higher dimension than the Tether solution. It is not a minimum solution, in the sense of Tether's algorithm, for the given sequence of Markov parameters. Minimality, in that sense, has been given up for the sake of stability. If a model of smaller dimension is desired, the techniques of Davison discussed earlier can be applied. Since the choice of $\bar{\lambda}_i$ is completely arbitrary, it is possible to select eigenvalues with widely differing orders of magnitude. The model can then be reduced in size by Davison's technique without significantly changing the properties of the model.

DISTRIBUTED SYSTEMS

The partial realization approach can also be useful in the modeling of distributed systems. It is important to note that the models constructed by this method are lumped systems depending only on time while the distributed systems depend also on one or more spatial variables. Therefore, in order to model the distributed system with a lumped model, we must fix the spatial coordinates. That is, we must select a finite number of specific locations in space at which to measure the output variables. The resulting model will describe the interaction between the output variables at those specific locations and the input variables.

The most important application of minimum partial realization analysis to distributed parameter systems is the formulation of a mathematical expression for the relationship between the function $\psi(x, t)$ at any finite number of positions x_i and the control variables for the system. To construct this expression, an impulse is imposed on each of the control variables and the corresponding response of $\psi(x,t)$ is measured at the desired x_i.

As an example, consider a tubular reactor operating at some set of steady state conditions. Assume it is necessary to estimate, a priori, the effect of variations in the inlet conditions on the function $\psi(x, t)$ at the end or in the center of the reactor. The response of the system to an impulse in the inlet conditions (that is, concentration or temperature) would then be monitored at these specific positions. The inlet conditions would be the control variables $u_i(t)$ and the function $\psi(x, t)$ evaluated at the locations x_i would be the measurement variables $y_i(t)$. A mathematical model relating $\mathbf{y}(t)$ and $\mathbf{u}(t)$

$$d\mathbf{y}(t)/dt = \mathbf{HFH}^{-1}\mathbf{y}(t) + \mathbf{HGu}(t)$$

would be formed, just as it was in the modeling of lumped parameter systems. From this model, the response of the function $\psi(x, t)$ at the x_i could be predicted for any other forcing function in the input variables. This method of modeling is especially attractive because of its simplicity.

It should be noted that for the construction of a partial realization of the problem, the initial and boundary conditions assumed in the development of the Ho algorithm must be met. This requires that

1. the system be at steady state for some time before the impulse,
2. the steady state be stable,
3. the impulse generate a response that can be modeled by a power series in the variable time, and
4. the measured variables be selected so that they are at zero at the steady state conditions.

At this point, it is important to recall that the Ho algorithm and the Tether algorithm construct minimum realizations based on the partial sequence $\mathbf{HF}^k\mathbf{G}$, $k =$

$\alpha \ldots \beta$, where α and β are the values of k in the first and last terms in the given partial sequence. Previously, and in the publications available in the literature, only solutions using $\beta > \alpha = 0$, $\beta < \alpha = 0$, and $\beta < \alpha = 1$ have been attempted. In the theory of partial minimum realizations, however, there is no inherent restriction on the values of α and β. The same basic algorithm can be used to construct realizations which match any possible sequence of Y_k. It would be profitable, then, to attempt to model the sequence of Y_k which will provide the most accurate description of the original system. In some instances (Rossen, 1972), the partial sequence of moments (especially a few moments) might not give a sufficiently accurate representation of a distributed parameter system. Earlier, it was observed that the models based on a partial sequence of Markov parameters provided excellent initial response predictions, but that the accuracy of the predictions decreased with increasing time. There were often steady state errors associated with these predictions. Therefore, it is recommended here that a combination of the Bruni approach and the Ho-Kalman approach be used. In this way, the advantages of both techniques could be combined. The steady state error in the Markov parameter approach can be alleviated by adding the first few moments, notably $\mathbf{HF}^{-1}\mathbf{G}$, to the partial sequence. The best partial sequence for the modeling of chemical processes might be

$$\{\mathbf{HF}^k\mathbf{G}\} \qquad k = \ldots -2, -1, 0, 1, 2, \ldots$$

The matrices corresponding to $k \geq 0$ will assure accurate representation in the neighborhood of the origin, while the matrices corresponding to $k < 0$ should improve the accuracy of the overall response curve.

A NUMERICAL DISTRIBUTED EXAMPLE

One example of a parabolic system will be discussed in detail in order to illustrate the application of some of these techniques. This particular problem is analyzed in a paper by Lapidus and Amundson (1952). Consider a packed bed adsorption column with axial diffusion but with negligible radial concentration and velocity gradients. Equilibrium is assumed to be established instantaneously at each point in the bed. The mathematical equations representing the system are then

$$D\partial^2 C/\partial z^2 = V\partial C/\partial z + \partial C/\partial t + (1/\alpha)\partial n/\partial t \qquad (7)$$

$$n = k_1 C + k_2 \qquad (8)$$

where C is the concentration of the adsorbate in the liquid, n is the concentration in the solid, α is the fractional void volume of the bed, and z is the spatial coordinate in the axial direction. V and D are the velocity of the fluid based on the void volume and the diffusivity of the adsorbate in the fluid, respectively. The boundary conditions for a step change in the inlet concentration are

$$\begin{aligned} C &= C_0(t) = C_0 && \text{at } z = 0, \quad t > 0 \\ C &= C_i(z) = C_i && \text{at } t = 0, \quad z > 0 \\ n &= n_i(z) = k_i C_i + k_2 && \text{at } t = 0, \quad z > 0 \end{aligned} \qquad (9)$$

A partial realization of the system will be constructed and will be tested, as in the previous examples, on a step forcing function.

Lapidus and Amundson solved (7) and (8), subject to (9), and found that the response, at any point z in the bed, to a step change in the inlet concentration is

$$H(m) = \frac{C - C_i}{C_0 - C_i}$$

$$= \frac{1}{2}\left[1 + \operatorname{erf}\left(\sqrt{\frac{mV}{4\alpha\gamma D}} - z\sqrt{\frac{\alpha\gamma V}{4mD}}\right) + \exp(Vz/D)\operatorname{erfc}\left(\sqrt{\frac{mV}{4\alpha\gamma D}} + z\sqrt{\frac{\alpha\gamma V}{4mD}}\right)\right] \qquad (10)$$

where $m = Vt$ and $\gamma = 1 + k_1/\alpha$. It is also assumed that the impulse response can be represented by the superposition of the responses to a positive step input at $t = 0$ and an equal negative step input at $t = \bar{t}$. The impulse response is then

$$C(z, t) = C_0[H(m) - H(m - V\alpha\bar{t})] \qquad (11)$$

where $\bar{t}$ is the duration of the square wave input, C_0.

Assume that the only variables of interest are the conditions in the stream leaving the bed at $z = L$. The problem is then to construct a model of the bed by which the effect of the inlet condition (concentration) on the outlet concentration can be predicted. The initial step in the modeling is, once again, the measurement of the appropriate moments and Markov parameters at $z = L$ for the impulse response. It will be assumed that the constants have the following values:

$$V = 20 \qquad V/D = 1$$

$$\alpha = \gamma = 5 \qquad \bar{t} = 0.1$$

$$C_i = 0 \qquad C_0 = 0.1$$

The Y_k, $k \geq 0$, are computed by a least squares time fit of a 26 term polynomial to the impulse curve. The first two moments are measured from this curve. Models for the system will be constructed from the two sequences of parameters

$$A \rightarrow \{\mathbf{HF}^k\mathbf{G}\}, \quad k = -1, \quad 0, 1, \ldots, 6$$

$$B \rightarrow \{\mathbf{HF}^k\mathbf{G}\}, \quad k = -2, -1, 0, \ldots, 7$$

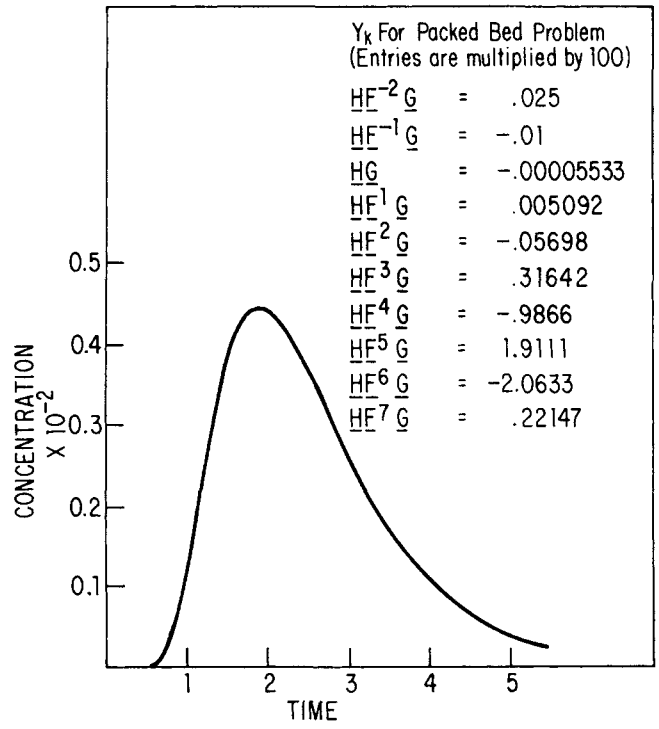

Fig. 3. Impulse response and partial sequence data for one-dimensional packed bed problem.

TABLE 3. REALIZATIONS FOR PACKED BED SYSTEM

Solution A

$$\mathbf{HF}^{-1} = \begin{bmatrix} -1.0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} 0.005533 & -0.509224 & 0 & 0 \\ 1.0 & -11.20127 & -62.697 & 0 \\ 0 & 1.0 & 3.26974 & -8.11716 \\ 0 & 0 & 1.0 & 2.65142 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 1.0 & 0 & 0 & 0 \end{bmatrix}^T$$

Solution B

$$\mathbf{HF}^{-2} = \begin{bmatrix} 2.5 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} -0.4 & -0.162213 & 0 & 0 & 0 \\ 1.0 & -0.85016 & 12.4886 & 0 & 0 \\ 0 & 1.0 & -3.6416 & -19.1281 & 0 \\ 0 & 0 & 1.0 & 0.35912 & -0.0635 \\ 0 & 0 & 0 & 1.0 & 6.2318 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$$

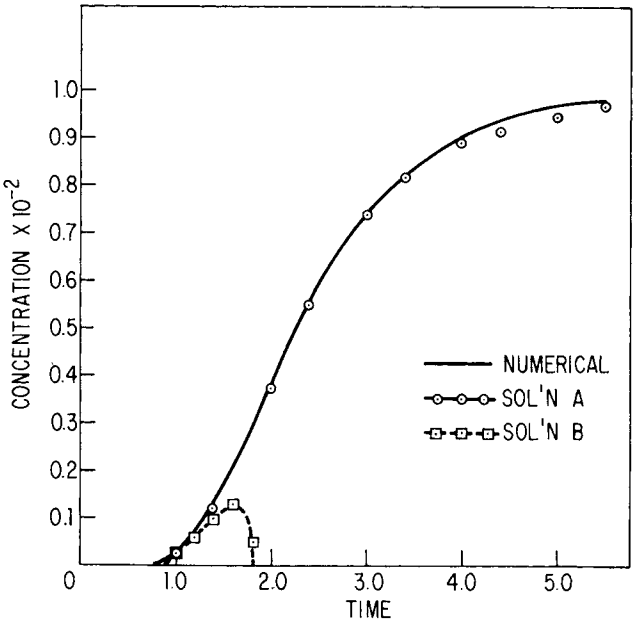

Fig. 4. One-dimensional packed bed responses. Tether models.

TABLE 4. STEP RESPONSE FOR PACKED BED SYSTEM

Time	Numerical	Solution A	Solution B	Stabilized solution B
0	0	0	0	0
0.5	0	0.00002	0.00002	0.000028
1.0	0.00025	0.000215	0.000206	0.000223
1.5	0.0016	0.00155	0.00114	0.00157
2.0	0.00383	0.00371	−0.00460	0.00381
2.5	0.00585	0.00582	−0.175	0.00602
3.0	0.00737	0.00738	−4.32	0.00761
3.5	0.00838	0.00830	—	0.00855
4.0	0.00903	0.00884	—	0.00906
4.5	0.00943	0.00918	—	0.00936
5.0	0.00966	0.00943	—	0.00956
5.5	0.00980	0.00961	—	0.00971
6.0	0.00988	0.00974	—	0.00982

Analysis of the impulse response yields the sequence presented in Figure 3, which includes all of the data necessary to construct the model. The problem will be treated as a system with one input variable and one measurement variable. The resulting realizations of the two partial sequences are shown in Table 3. These models are used to predict the response in the effluent concentration to a step in the inlet concentration. The plots of the predicted and analytical responses are in Figure 4 and are tabulated in Table 4.

From these plots it can be seen that solution *A*, based on eight parameters, is a stable representation of the system. In this case, the partial minimum realization approach has again generated an acceptable model. Solution *B*, however, blows up almost immediately; its utility as a model is therefore poor. The cause of the instability in solution *B* is the presence of a positive eigenvalue in the **F** matrix.

The effectiveness of the stabilization procedure detailed in this paper will be illustrated by applying the procedure to this problem. The positive eigenvalue will be replaced by a set of negative eigenvalues to generate a stable model which will be far superior to the unstable model constructed directly from the Tether method.

The first step in the treatment of solution *B* is the evaluation of the **F** matrix. Then, the corresponding eigenvectors are computed from the matrix equation

$$\mathbf{FT} = \mathbf{TD}$$

where **D** is the diagonal eigenvalue matrix and **T** is the eigenvector matrix. The **F** matrix is then replaced by $\mathbf{TDT}^{-1}$ and the partial realization can be transformed into $\{\mathbf{D}, \mathbf{T}^{-1}\mathbf{G}, \mathbf{HT}\}$. For this problem, the matrices are

$$\mathbf{D} = \begin{bmatrix} \begin{bmatrix} -1.18197 \\ +1.8667i \end{bmatrix} & 0 & 0 & 0 & 0 \\ 0 & \begin{bmatrix} -1.18197 \\ -1.8667i \end{bmatrix} & 0 & 0 & 0 \\ 0 & 0 & -2.01398 & 0 & 0 \\ 0 & 0 & 0 & -0.86583 & 0 \\ 0 & 0 & 0 & 0 & 6.22473 \end{bmatrix}$$

and

$$\mathbf{T} = \begin{bmatrix} \begin{bmatrix} -3.5124 & -3.5124 \\ -7.7851i & +7.7851i \end{bmatrix} & -13.6098 & -43.739 & -0.007216 \\ \begin{bmatrix} -106.5 & -106.5 \\ +2.8907i & -2.8907i \end{bmatrix} & -135.4 & -125.6 & 0.02947 \\ \begin{bmatrix} 2.6793 & 2.6793 \\ -15.375i & +15.375i \end{bmatrix} & 13.7092 & 3.6599 & 0.016751 \\ \begin{bmatrix} 7.4138 & 7.4138 \\ +1.8667i & -1.8667i \end{bmatrix} & -8.2458 & -7.0977 & -0.0070999 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix}$$

At this point, the positive eigenvalue, 6.22473, will be replaced by a set of negative eigenvalues. The selection of the negative eigenvalues is completely arbitrary, but it might be expected that eigenvalues whose magnitudes lie on both sides of the positive eigenvalue should be included in the solution set. The set of negative eigenvalues will be [−1.0, −2.0, −2.5, −50, −51, −52, −53, and −54]. The α's are computed with the method of determinants. For this set of eigenvalues, the magnitudes of the α's are between 10 and 10,000. It will be seen that this model is an accurate representation of the packed bed, approximating the actual response to within three percent over the range of the curve.

Note, however, that this solution is of dimension 12 whereas the original unstable system was of dimension 5. It would be advantageous to reduce the dimension of the model if the reduction could be accomplished without great loss of accuracy. This is the reason for the selection of eigenvalues that differ widely in magnitude. Davison's work suggests that the eigenvalues $\lambda_i \geqq 50$ could be ignored without significant loss of accuracy. The system is thus reduced to a system of dimension 7. The **H** matrix is premultiplied by a Davison factor of $d_i = [\mathbf{HF}^{-1}\mathbf{G}]_i / [\mathbf{H}^*\mathbf{F}^{-1*}\mathbf{G}^*]_i$ to adjust the steady state prediction to eliminate the small offset error. The *'s denote matrices in the seven eigenvalue stable model. The numerator contains matrices from the twelve eigenvalue model. The model's step response is compared with the step responses of the unstable Tether model and the analytical solution in Figure 5 and Table 4. The final seven dimensional model is shown in Table 5. The predicted response of the 7-eigenvalue system is seen to be vastly superior to the response predicted from the unstable model. It is accurate to within three percent of the steady state response over the range of the curve.

From this illustration and others not shown here, it is clear that the partial minimum realization approach can generate useful models for parabolic systems and that the stabilization method is a very powerful technique in the analysis of such systems.

ACKNOWLEDGMENT

The authors wish to acknowledge the support of this work from National Science Foundation Grant NSF-GP-2858. Furthermore, this work made use of Princeton University Computer facilities supported in part by National Science Foundation Grant NSF-GJ-34 and GU-3157.

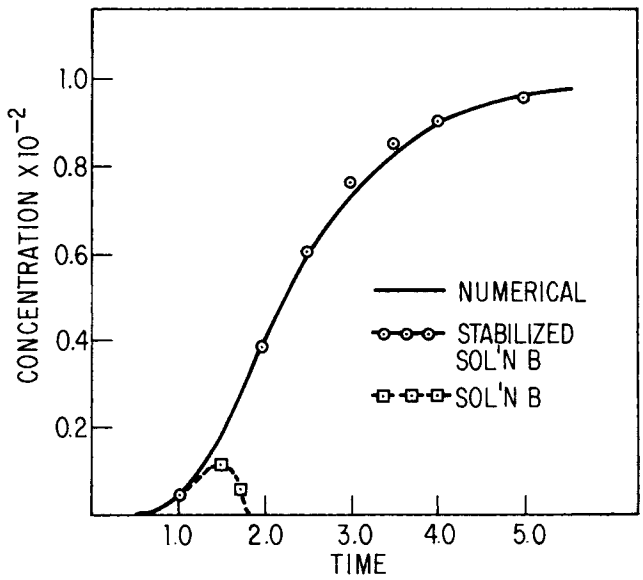

Fig. 5. One-dimensional packed bed responses. Stabilized model.

TABLE 5. STABLE 7-EIGENVALUE MODEL OF PACKED BED SYSTEM

$$\mathbf{F} = \begin{bmatrix} -1.1820 \\ 1.8667i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \begin{matrix} -1.1820 \\ -1.8667i \end{matrix} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2.01398 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.86583 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2.5 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 0.00433 & -0.00796i \\ 0.00433 & +0.00796i \\ 0.0253 \\ 0.0343 \\ 0.002934 \\ 0.002934 \\ 0.002934 \end{bmatrix} \quad ;\mathbf{HF}^{-2} = \begin{bmatrix} -8.76 & -19.4i \\ -8.76 & +19.4i \\ -34.0 \\ -109.2 \\ -0.001803\alpha_1 \\ -0.001803\alpha_2 \\ -0.001803\alpha_3 \end{bmatrix} \quad \begin{matrix} \alpha_1 = 92.82 \\ \alpha_2 = -270.1 \\ \alpha_3 = 178.5 \end{matrix}$$

LITERATURE CITED

Andree, R. V., "Computation of the Inverse of a Matrix," *Monthly*, **58**, 87 (1951).

Buffham, B. A., and L. G. Gibilaro, "A Unified Time Delay Model for Dispersion in Flowing Media," *Chem. Eng. J.*, **1**, 31 (1970).

Gibilaro, L. G, and F. P. Lees, "The Reduction of Complex Transfer Function Models to Simple Models Using the Method of Moments," *Chem. Eng. Sci.*, **24**, 55 (1969).

Davison, E. J., "A Method for Simplifying Linear Dynamic Systems," *IEEE Trans. Auto. Cont.*, **11**, 93 (1966).

———, "A New Method for Simplifying Large Linear Dynamic Systems," *ibid.*, **13**, 214 (1968).

Kallina, C., "An Approach to the Regulation Problem for Nonlinear Chemical Processes," *JACC Preprints*, **11**, 329 (1970).

Kropholler, H. W., "Determination of Relative Variance and Other Moments from Generalized Flow Networks on System Transfer Functions," *Ind. Eng. Chem. Fundamentals*, **9**, 329 (1970).

Lapidus, L., and N. R. Amundson, "Mathematics of Adsorption in Beds," *J. Phys. Chem.*, **56**, 948 (1952).

Midoux, N., and J. C. Charpentier, "Note on a Method of Fitting Two-Parameter Distribution Models to an Experimental Residence Time Distribution," *Chem. Eng. J.*, **1**, 163 (1970).

Rossen, R. H., Ph.D. thesis, Princeton Univ., New Jersey (1972).

THE AUTHORS

Leon Lapidus received his Ph.D. from the University of Minnesota, Class of 1943. He is now Professor and Chairman, Department of Chemical Engineering, Princeton University. He is the winner of AIChE Professional Progress Award in 1966, the author of over 100 articles and three textbooks, and a consultant to many industrial organizations.

Robert H. Rossen received his Ph.D. from Princeton University; he has received National Science Foundation Fellowships for graduate study. He is now with Esso Production Research Company in Houston, Texas.

JOURNAL REVIEW

Losses of Mercury from Chlorine Plants: A Review of a Pollution Problem

The problem of mercury losses from chlorine plants is reviewed. The mercury material balance is difficult to establish because mercury escapes from the process by a number of routes. Some of the reasons for the mercury emissions are explored, and some problems involved in the measurement of mercury in various effluents are discussed.

A number of mercury containment processes for specific streams have already been proposed, but the need for an integrated mercury recycle policy still exists.

REINALDO CABAN and THOMAS W. CHAPMAN

Department of Chemical Engineering
University of Wisconsin
Madison, Wisconsin 53706

SCOPE

Over the past two years there has been much concern and discussion over environmental pollution by industrial discharges of mercury. This article attempts to review the mercury problem from the process engineering point of view. The particular situation that we examine in detail is that of a mercury-cell chlorine plant. Such plants have been the largest contributors to mercury pollution in the chemical industry.

Although chlorine plants should not consume any mercury in the process, steady losses from the mercury inventory have been the universal experience over the years. One reason this situation persisted so long is that a direct determination of the mercury material balance is extremely difficult. Although a complete and detailed measurement of the mercury losses is not available, we present some estimates of mercury flows with various effluent streams. We discuss some of the causes of these losses as well as possible reasons that they have eluded quantitative detection.

In any case, it is clear that mercury loss from a chlorine process does occur by various routes and in several forms. The containment and recovery of this mercury presents a challenging problem to the chemical engineer. A number of schemes for mercury separation from effluent streams have been proposed. We present a short review and evaluation of these processes.

We hope that this exposition will not only put the mercury problem into perspective but also provide an illustrative example that can help solve or prevent similar loss problems in other industries.