Let us write $\dim \mathcal{S}_* = q$. Then generically in A, $\dim A\mathcal{S}_* = q$ also. Using this, we have generically (in A and B) that $\dim(A\mathcal{S}_* + \operatorname{ran} B) = \min(q+m, n)$. Finally, we obtain that generically (in A, B, and $\mathcal{R}$) the following holds.

$$\dim\left(\mathcal{R} \cap \left(A\mathcal{S}_* + \operatorname{ran} B\right)\right) = \max(0, r + \min(q+m, n) - n). \quad (39)$$

From this and (38) we get, using the fact that $q \neq 0$,

$$q = r + \min(q+m, n) - n. \quad (40)$$

Now $\min(q+m, n)$ is equal either to $q+m$ or to n. But in the first case, (40) gives $r+m=n$, which contradicts the given relation $r+m > n+1$. So it must be that $\min(q+m, n) = n$ and then (40) gives $q = r$, implying that $\mathcal{S}_* = \mathcal{R}$ and, thereby, that $\mathcal{R}$ is a controllability subspace. □

The lemma shows, for instance, that the problem of disturbance decoupling with stability (see [7, par. 5.6] also for notation) is generically solvable if the pair (A, B) is stabilizable, $DE = 0$, and $\dim \ker D + \dim \operatorname{ran} B > n+1$. But we can also use it in the present context.

Corollary 5.9: Suppose that (C, A) is observable, (A, B) is controllable, and $\dim \operatorname{ran} B > \dim \ker C + 1$. Then arbitrary pole placement by direct output feedback is generically possible.

Proof: We first use the "observable" version of Lemma 5.3 to obtain a complement $\mathcal{V}_1$ of $\ker C$ such that $\sigma(A + GC: \mathcal{X}/\mathcal{V}_1)$ has an arbitrarily prescribed location for $G \in \boldsymbol{G}(\mathcal{V}_1)$. We have $\dim \mathcal{V}_1 + \dim \operatorname{ran} B > n+1$. So, generically, $\mathcal{V}_1$ is a controllability subspace and we can take $\mathcal{V}_2 = \mathcal{V}_1$ and $F \in \boldsymbol{F}(\mathcal{V}_2)$ such that $\sigma(A + BF: \mathcal{V}_2/\mathcal{O})$ also has an arbitrarily prescribed location. So we can synthesize an observer-based compensator of order 0 with arbitrarily prescribed poles. □

One may note that our method of proof also provides a simple design procedure. We have given the result in terms of genericity. Of course, one may go about to give sufficient conditions for arbitrary pole placement to be possible by direct output feedback, and indeed, several such conditions are available in the literature (see, for instance, [3], [4], and [18]). However, the relations between these results are not quite clear. Perhaps the theory of the present paper could be used to clarify this and also the further results of [18], but we shall not go into these matters here.

VI. Conclusions

We have presented a general symmetric theory of compensator synthesis, which describes the problem as a pairing problem: find a (C, A, B)-pair $(\mathcal{V}_1, \mathcal{V}_2)$ and a pair $(F, G) \in \boldsymbol{P}(\mathcal{V}_1, \mathcal{V}_2)$ such that both $A + BF: \mathcal{V}_2/\mathcal{O}$ and $A + GC: \mathcal{X}/\mathcal{V}_1$ are stable. As this formulation allows many ways to derive the same compensator, one can specialize—with hardly any loss of room, at least theoretically speaking—to the case in which $\mathcal{V}_1$ is a complement of $\ker C$, and so obtain compensators that may be interpreted as observer-based compensators.

Any result on the stable cover problem can now immediately be applied to compensator synthesis. (If there is a restriction on the dimension of the subspace to be covered, one may use a little trick as in the proof of Proposition 5.7.) One should note, however, that in the case of the compensator problem, the subspace to be covered is not really fixed. It is seen from Lemma 5.2 that if $\sigma(A + GC: \mathcal{X}/\mathcal{V}_1)$ is stable for $G \in \boldsymbol{G}(\mathcal{V}_1)$ (where $\mathcal{V}_1$ is a complement of $\ker C$), then $\sigma(A + GC: \mathcal{X}/\mathcal{V}_1)$ will also be stable for $G \in \boldsymbol{G}(\tilde{\mathcal{V}}_1)$ if $\tilde{\mathcal{V}}_1$ is sufficiently close to $\mathcal{V}_1$ (in a suitable topology to be laid on the complements of $\ker C$). The fact that the underlying subspace can be moved a bit is reassuring from a numerical point of view, but can it also be put to any theoretical use?

Another strategy for synthesizing a compensator of order k would be to first select a $(k+p)$-dimensional (A, B)-invariant subspace $\mathcal{V}_2$ such that there exists F with $A + BF: \mathcal{V}_2/\mathcal{O}$ stable, and then look for a suitable complement $\mathcal{V}_1$ of $\ker C$ such that $\mathcal{V}_1 \subset \mathcal{V}_2$. Dualizing this also gives rise to a version of the stable cover problem, but now the cover has to be complementary to a given subspace and, hence, has fixed dimension.

Obviously, there is room for further research here. One could also try to treat some variants of the compensator problem within the above framework (for instance, when only an output function $z(t) = Dz(t)$ needs to be stabilized or when measured or unmeasured disturbances are present). Moreover, it would be of interest to see whether further results in compensator theory as in [18] can be interpreted or even extended with the help of (C, A, B)-pairs. It has already been shown that the concept of (C, A, B)-pairs, when appropriately modified, is instrumental in observer theory [16]. The author is also working on an extension of the theory to the infinite-dimensional context.

References

[1] D. G. Luenberger, "An introduction to observers," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 596–602, 1971.

[2] F. M. Brasch and J. B. Pearson, "Pole placement using dynamic compensators," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 34–43, 1970.

[3] S. H. Wang and E. J. Davison, "On pole assignment in linear multivariable systems using output feedback," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 516–518, 1975.

[4] H. Kimura, "Pole assignment by gain output feedback," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 509–516, 1975.

[5] W. M. Wonham and A. S. Morse, "Feedback invariants of linear multivariable systems," *Automatica*, vol. 8, pp. 93–100, 1972.

[6] ——, "Decoupling and pole assignment in linear multivariable systems," *SIAM J. Contr.*, vol. 12, pp. 1–18, 1970.

[7] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach* (2nd ed). New York: Springer Verlag, 1979.

[8] J. M. Schumacher, "(C, A)-invariant subspaces: Some facts and uses," Wiskundig Seminarium Vrije Universiteit Amsterdam, The Netherlands, Rep. 110, 1979.

[9] F. Hamano and K. Furuta, "Localization of disturbances and output decomposition in decentralized linear multivariable systems." *Int. J. Contr.*, vol. 22, pp. 551–562, 1975.

[10] H. Akashi and H. Imai, "Disturbance localization and output deadbeat control through an observer in discrete-time linear multivariable systems," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 621–627, 1979.

[11] G. Basile and G. Marro, "Controlled and conditioned invariant subspaces in linear system theory," *J. Optimiz. Theory Appl.*, vol. 3, pp. 306–315, 1969.

[12] A. S. Morse, "Structural invariants of linear multivariable systems," *SIAM J. Contr.*, vol. 11, pp. 446–465, 1973.

[13] J. C. Willems and C. Commault, "Disturbance decoupling by measurement feedback with stability or pole placement," *SIAM J. Contr. Optimiz.*, to be published.

[14] R. W. Scott and B. D. O. Anderson, "Least order, stable solution of the exact model matching problem," *Automatica*, vol. 14, pp. 481–492, 1978.

[15] H. R. Sirisena, "Minimal-order observers for linear functions of a state vector," *Int. J. Contr.*, vol. 29, pp. 235–254, 1979.

[16] J. M. Schumacher, "On the minimal stable observer problem," *Int. J. Contr.*, vol. 32, pp. 17–30, 1980.

[17] D. G. Luenberger, "Observers for multivariable systems," *IEEE Trans. Automat. Contr.*, vol. AC-11, pp. 190–197, 1966.

[18] H. Kimura, "On pole assignment by output feedback," *Int. J. Contr.*, vol. 28, pp. 11–22, 1978.

A Time-Stepping Procedure for $\dot{X} = A_1X + XA_2 + D$, $X(0) = C$

STEVEN M. SERBIN AND CYNTHIA A. SERBIN

***Abstract*—We develop an expression for the exact solution of the matrix differential problem $\dot{X} = A_1X + XA_2 + D$, $X(0) = C$ based on variation of parameters and use this to devise the time-stepping relation $X(t+h) = e^{A_1h}\{X(t) + \int_0^h e^{-A_1s} De^{-A_2s}\, ds\} e^{A_2h}$. We modify a procedure of Van Loan to effect efficient computation of all the terms necessary to advance the solution in time according to this relation. We consider some alternatives when sparsity is a concern. A numerical example of our procedure is included.**

Manuscript received December 6, 1979; revised June 5, 1980. Paper recommended by A. J. Laub, Chairman of the Computational Methods and Discrete Systems Committee. This work was supported by USARO Grant DAAG29-78-C-0024. The Union Carbide Corporation—Nuclear Division is operated for the U.S. Department of Energy under Contract W-7405-eng-26.

S. M. Serbin is with the Department of Mathematics, University of Tennessee, Knoxville, TN 37916.

C. A. Serbin is with the Department of Mathematics and Statistics Research, Computer Sciences Division, Union Carbide Corporation—Nuclear Division, Oak Ridge, TN 37830.

0018-9286/80/1200-1138$00.75 ©1980 IEEE

I. Introduction

Recently, there have been several methods proposed for the numerical solution of the matrix initial-value problem

$$\dot{X}=A_1X+XA_2+D, \qquad X(0)=C \tag{1}$$

where A_i are real square constant matrices of order n_i, $i=1, 2$, *not necessarily invertible*, C and D are constant matrices of size $n_1\times n_2$, and $X(t)$ is the time-varying matrix of size $n_1\times n_2$ for which we wish to solve. Each of these techniques seeks to improve upon the conventional approach of transforming (1) into an equivalent vector-matrix system of order n_1n_2 and is, in some way, based upon the expression of the solution of (1) (cf. Davison [1]) as

$$X(t)=e^{A_1t}(C-E)\,e^{A_2t}+E \tag{2}$$

where E solves the algebraic problem

$$A_1E+A_2=-D. \tag{3}$$

Davison [1] uses Padé approximation for the exponentials in (2) (for $t=h$ sufficiently small) and a clever iterative scheme based upon numerical integration of

$$E=\int_0^\infty e^{A_1\tau}De^{A_2\tau}\,d\tau \tag{4}$$

which can be shown (cf. Bellman [2]) to solve (3). A time-stepping procedure is then employed to approximate $X(kh)$, $k=0, 1, 2\ldots$. Barraud [3] suggests using the Bartels–Stewart algorithm [4] for solving (3) (a recent modification due to Golub, Nash, and Van Loan [5] could also be used), and calculates the Padé approximants for e^{A_ih} as before. The step size h and order of the Padé diagonal approximant may be selected to obtain approximations of any desired accuracy. Hoskins, Meek, and Walton [6], on the other hand, concentrate on a different *iterative* scheme for (3), one involving matrix inversion at each iterative step, to try to speed the convergence process.

Each of these approaches concentrates its effort in the solution of (3); the question of the exponential approximation and time stepping is almost incidental. For dense matrices which are not too large, these approaches would all appear to be quite valid. However, for matrices A_i which are large and/or sparse, the effort in solving (3) in any of the ways proposed would seem to be quite large; moreover, it is not clear that advantage can be taken of the zero structure in sparse problems. Further, the work in time stepping these procedures cannot be ignored.

The approach which we suggest seeks to avoid some of these difficulties. In particular, we never have to solve (3). Instead, we develop a time-stepping procedure wherein the quantity E is not used, but rather

$$B(h)\equiv\int_0^h e^{-A_1s}De^{-A_2s}\,ds. \tag{5}$$

It is most important to note that the integration here proceeds only over the (small) interval $[0, h]$. Hence, numerical integration schemes of high accuracy might be used inexpensively; instead, we shall present a procedure of Van Loan [7] (modified slightly for our purposes) for the calculation of (5) as well as e^{A_ih}.

II. The Time Stepping Relation

The change of variables $Y(t)=X(t)\,e^{-A_2t}$ reduces (1) to the nonhomogeneous equation

$$\dot{Y}=A_1Y+De^{-A_2t}, \qquad Y(0)=C \tag{6}$$

to which variation of parameters can be easily applied. Returning to the original variables, it follows that

$$X(t)=e^{A_1t}\left\{C+\int_0^t e^{-A_1\tau}De^{-A_2\tau}\,d\tau\right\}e^{A_2t}, \tag{7}$$

which serves as an alternative to (2). From this, it is easily discerned that

$$X(t+h)=e^{A_1h}\left\{X(t)+\int_t^{t+h}e^{A_1(t-\tau)}De^{A_2(t-\tau)}\,d\tau\right\}e^{A_2h} \tag{8}$$

and so by changing variables $s=\tau-t$ and employing the definition (5) we have

$$X(t+h)=e^{A_1h}\{X(t)+B(h)\}\,e^{A_2h}. \tag{9}$$

We see that to advance the solution from time t to $t+h$, we still require pre- and postmultiplication by e^{A_1h}, e^{A_2h}, respectively, but now only need compute (once and for all) $B(h)$ or some appropriate approximation.

III. Time-Stepping Approximate Scheme

Suppose, for the moment, that we have selected a time step h at which we wish to observe the solution and for which $B(h)$, e^{A_ih}, $i=1, 2$ can be accurately approximated by $\tilde{B}(h)$, $\tilde{E}_i(h)$, $i=1, 2$, respectively. Then, setting $X_k\approx X(kh)$, with $X_0=X(0)=C$, our approximate scheme is just

$$X_{k+1}=\tilde{E}_1(h)\{X_k+\tilde{B}(h)\}\,\tilde{E}_2(h) \tag{10}$$

which, if implemented in this form, requires just two multiplications per time step.

Now, the way that we choose to make our approximations and compute in (10) is influenced by sparsity considerations. First, we shall discuss the general situation in which no special structure is assumed. As mentioned previously, we could compute our approximation to $B(h)$ by numerical integration and use matrix exponential routines for $\tilde{E}_i(h)$, thus applying (10) directly. We wish to suggest another alternative, which modifies a procedure of Van Loan [7], to approximate $B(h)\approx\tilde{B}(h)$, $e^{A_1h}\approx\tilde{E}_1(h)$, and $e^{-A_2h}\approx\tilde{F}_2(h)$. This procedure affords a convenient, if not less expensive, organization of the computations so that all approximations are obtained via one matrix exponential evaluation at the expense of obtaining $\tilde{F}_2(h)$ instead of $\tilde{E}_2(h)$ for (10). We shall accomodate this by rewriting (10) as

$$X_{k+1}\tilde{F}_2(h)=\tilde{E}_1(h)\{X_k+\tilde{B}(h)\} \tag{11}$$

and solving the linear system at each step. As $\tilde{F}_2(h)$ can be factored once and for all as the product of lower and upper triangular matrices, at an expense of about $1/3n_2^3$ multiplications (where $n_2\leqslant n_1$ can always be arranged by transposing (1) if necessary), the work thereafter in time stepping (11) is the same as in (10).

Our version of Van Loan's procedure is as follows. We let $\delta=h/2^m$ for some nonnegative integer m to be selected (very possibly $m=0$). Defining the block upper triangular matrix T by

$$T=\begin{bmatrix}A_1 & D\\ 0 & -A_2\end{bmatrix} \tag{12}$$

then it can be shown easily as in [7], that with $\hat{T}\equiv T\delta$, $\hat{A}_i=A_i\delta$

$$e^{\hat{T}}=\begin{bmatrix}e^{\hat{A}_1} & G(\delta)\\ 0 & e^{-\hat{A}_2}\end{bmatrix} \tag{13}$$

where

$$G(\delta)=e^{A_1\delta}\int_0^\delta e^{-A_1\tau}De^{-A_2\tau}\,d\tau. \tag{14}$$

Now, the well-known exponential property yields

$$e^{Th}=(e^{Th/2^m})^{2^m}=\begin{bmatrix}e^{A_1h} & G(h)\\ 0 & e^{-A_2h}\end{bmatrix}. \tag{15}$$

[Indeed, one can prove inductively that the block form in (15) is obtained by the m successive squarings of (13).] Noting that (9) can be rewritten as

$$X(t+h)=\{e^{A_1h}X(t)+e^{A_1h}B(h)\}\,e^{A_2h} \tag{16}$$

we see from (5) and (14) that $e^{A_1h}B(h)=G(h)$, so that the approxima-

tion (10) can be rewritten as

$$X_{k+1}=\left\{\tilde{E}_1(h)X_k+\tilde{G}(h)\right\}\tilde{E}_2(h) \tag{17}$$

for $\tilde{G}(h)$ an approximation to $G(h)$.

Correspondingly, (11) becomes

$$X_{k+1}\tilde{F}_2(h)=\left\{\tilde{E}_1(h)X_k+\tilde{G}(h)\right\}. \tag{18}$$

For simplicity in the sequel, let us deal only with (18) and further denote $\tilde{E}_1(h)\equiv\tilde{F}_1(h)$. In order to obtain the required approximations, we could, in principle, just appeal to a good existing code, such as the scaling and squaring algorithm of Ward [8], which approximates $e^{\hat{T}}\approx R(\hat{T})$, some appropriate rational approximation (in Ward's program, $R=R_{qq}$, the qth diagonal Padé approximation), and then (15) is approximated by $e^{Th}\approx R(\hat{T})^{2^m}$. However, T is of size $(n_1+n_2)\times(n_1+n_2)$, so this approach is not economical. Fortunately, the block upper triangularity of T saves us. Indeed, it is never necessary to work with a matrix of size $(n_1+n_2)\times(n_1+n_2)$ at all. For if the rational approximation $R(\hat{T})\equiv Q^{-1}(\hat{T})P(\hat{T})$, then it can be seen easily that

$$Q(\hat{T})=\begin{bmatrix} Q(\hat{A}_1) & V \\ 0 & Q(-\hat{A}_2)\end{bmatrix},\ P(\hat{T})=\begin{bmatrix} P(\hat{A}_1) & U \\ 0 & P(-\hat{A}_2)\end{bmatrix} \tag{19}$$

where U and V depend on A_i, D, δ and the specific approximation being used and would normally require a few matrix multiplications to determine. Then, if we set

$$R(\hat{T})=R(T\delta)=\begin{bmatrix} R_{11}(\delta) & R_{12}(\delta) \\ 0 & R_{22}(\delta)\end{bmatrix}$$

we see that the blocks of $R(\hat{T})$ are determined by solving

$$Q(-\hat{A}_2)R_{22}(\delta)=P(-\hat{A}_2) \tag{20}$$

$$Q(\hat{A}_1)R_{11}(\delta)=P(\hat{A}_1) \tag{21}$$

$$Q(\hat{A}_1)R_{12}(\delta)=U-VR_{22}(\delta). \tag{22}$$

Thus, the only additional work here over just finding approximations to $e^{A_1\delta}$ and $e^{-A_2\delta}$ is one extra matrix multiplication and back substitution in (22), which costs $n_1n_2(n_1+n_2)$ multiplications.

In addition, if $m>0$, the m successive squarings needed to ultimately obtain $\tilde{F}_i(h)$, $\tilde{G}(h)$ can also be developed by appealing to the structure of T, as is done similarly by Van Loan [7]. Initializing $\tilde{F}_{1,0}=R_{11}(\delta)$, $\tilde{F}_{2,0}=R_{22}(\delta)$, $\tilde{G}_0=R_{12}(\delta)$, we perform for $j=0,\cdots,m-1$

$$\begin{cases}\tilde{F}_{i,j+1}=\tilde{F}_{i,j}^2, & i=1,2\\ \tilde{G}_{j+1}=\tilde{F}_{1,j}\tilde{G}_j+\tilde{G}_j\tilde{F}_{2,j}\end{cases} \tag{23}$$

and finally $\tilde{F}_i(h)\equiv\tilde{F}_{i,m}$ and $\tilde{G}(h)\equiv\tilde{G}_m$. Again, the number of operations needed to calculate $\tilde{G}(h)$ from $\tilde{G}_0$ is about $mn_1n_2(n_1+n_2)$.

This procedure (20)–(23) can be accomplished, in particular, with the qth diagonal Padé approximation $R=R_{qq}$ (cf. Van Loan [7] for a discussion of the Padé routine). The choice of δ and q for accurate approximation rests on standard error analysis for Padé approximation for the exponential; Barraud [3] states an applicable criterion. Normally, though, we would not want q to become too large, as the work involved just in forming Q_{qq} and P_{qq} becomes appreciable.

IV. Numerical Example and Ideas for Further Exploration

In order to prepare to judge the merits of our time-stepping procedure, our first effort has been to test the method in the form (17). We have used Ward's scaling and squaring program to generate the approximations $\tilde{E}_1(h)$, $\tilde{E}_2(h)$, and $\tilde{G}(h)$ almost to machine accuracy (in double precision on the DEC-10 computer). Thus, errors in the matrix approximations are essentially avoided and only the time stepping is assessed.

TABLE I
Relative Error at $t=1$

N	h	$\epsilon(h)$
2	0.2	0.254 (−17)
	0.1	0.283 (−17)
5	0.2	0.257 (−17)
	0.1	0.274 (−16)
10	0.2	0.241 (−16)
	0.1	0.166 (−16)

Specifically, we have constructed test problems with $n_1=n_2=N$, $2<N<10$. For diagonal matrices Λ, Γ, Φ, Ψ, we construct A_1, A_2, C, D by similarity transformation, i.e., $A_1=S^{-1}\Lambda S$, $A_2=S^{-1}\Gamma S$, $C=S^{-1}\Phi S$, $D=S^{-1}\Psi S$ where

$$S_{ij}=\begin{cases}N-i+1, & i\geq j\\ N-j+1, & i<j\end{cases} \tag{24}$$

$$S_{ij}^{-1}=\begin{cases}1, & i=j=1\\ 2, & i=j\neq 1\\ -1, & i=j-1,\ i=j+1\\ 0, & \text{otherwise.}\end{cases} \tag{25}$$

Then, it can be shown easily that

$$X(t)=S^{-1}\left\{\operatorname{diag}\left[\left(\phi_i+\frac{\psi_i}{\lambda_i+\gamma_i}\right)e^{(\lambda_i+\gamma_i)t}-\frac{\psi_i}{\lambda_i+\gamma_i}\right]\right\}S \tag{26}$$

solves (1).

Since the solution will, in general, contain both increasing and decreasing exponential terms, we measure the error (at $t=1$ for convenience) in the relative sense:

$$\epsilon(h)\equiv\left[\sum_{i,j=1}^{N}\left(X_{ij}(1)-X_{M,ij}\right)^2\Big/\sum_{i,j=1}^{N}X_{ij}^2(1)\right]^{1/2} \tag{27}$$

where $Mh=1$. We chose (arbitrarily) $\Lambda=\operatorname{diag}(1,2,\cdots,N)$, $\Gamma=\operatorname{diag}(0.5, 1.5,\cdots,N-0.5)$, $\Phi=\operatorname{diag}(2,4,\cdots,2N)$, and $\Psi=\operatorname{diag}(1,3,\cdots,2N-1)$. Some sample results are shown in Table I.

Clearly, the procedure we have tested obtains approximations to essentially within machine precision of the true solution in the relative sense if we compute the required matrix approximations to great accuracy; this points out the effectiveness of using high quality software for the matrix exponential in conjunction with the time-stepping procedure. The effect of lower accuracy approximations and the concomitant error analysis of the time-stepping scheme is a subject of future concern; also, a comparison of timings of this approach with that of a conventional one, e.g., Davison's, is clearly indicated.

We should also like to offer a brief suggestion for some future investigation, particularly with regard to large sparse problems. Rather than form the rational approximations of (20)–(22) directly, after which advantage of the zero structure in the A_i may be lost, there are alternative procedures which have been suggested in the literature in a different setting which might be employed here. One is a computational version of the (2, 2) Padé approximation suggested by Fairweather [9], which entails the solution of complex linear systems which maintain the sparsity structure of the A_i. Another alternative would be to introduce a different class of rational approximations to the exponential more suited to the sparsity considerations. Nørsett [10] has studied a class of "restricted" Padé approximants; the important feature here is that these rational approximations can be formed by repeated solution of matrix problems with the same coefficient matrix which is linear in $\hat{T}$ and therefore has the same sparsity structure. Finally, we remark that if we would choose $h=\delta$, then (18) can be rewritten as

$$Q(A_1h)X_{k+1}P(-A_2h)=\left\{P(A_1h)X_k+\tilde{G}(h)\right\}Q(A_2h) \tag{28}$$

and thus we can deal with the sparse systems in (28), back solving at each time step.

V. Conclusion

A new numerical scheme for the initial value problem (1) has been proposed. This scheme is based upon approximating the exact recursive formula (9) (or (16) in another form) by the discrete time-stepping procedure (10) [respectively, (17)]. We suggest implementing the scheme in the form (18). Solution of any algebraic problem of the form (3) is avoided. Van Loan's procedure is adapted in (20)–(22) and (23) to economically compute approximations required for (18). We allow the time step for the differential equation to be of the form $h=2^m\delta$, so that the work needed to compute the rational approximations is effected by the choice of m, but appears to be considerably less than that of current methods.

References

[1] E. J. Davison, "The numerical solution of $\dot{X}=A_1X+XA_2+D$, $X(0)=C$," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 566–567, Aug. 1975.

[2] R. Bellman, *Introduction to Matrix Analysis*. New York: McGraw-Hill, 1960, p. 231.

[3] A. Y. Barraud, "A new numerical solution of $\dot{X}=A_1X+XA_2+D$, $X(0)=C$," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 976–977, Dec. 1977.

[4] R. H. Bartels and G. W. Stewart, "Algorithm 432, Solution of the matrix equation $AX+XB=C$," *Commun. Ass. Comput. Mach.*, vol. 15, pp. 820–826, Sept. 1972.

[5] G. H. Golub, S. Nash, and C. F. Van Loan, "A Hessenberg-Schur method for the problem $AX+XB=C$," Dep. Comput. Sci., Cornell Univ., Ithaca, NY, TR 78-354.

[6] W. D. Hoskins, D. S. Meek, and D. J. Walton, "The numerical solution of $\dot{X}=A_1X+XA_2+D$, $X(0)=C$," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 881–882, Oct. 1977.

[7] C. F. Van Loan, "Computing integrals involving the matrix exponential," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 395–404, June 1978.

[8] R. C. Ward, "Numerical computation of the matrix exponential with accuracy estimate," *SIAM J. Numer. Anal.*, vol. 14, pp. 600–610, 1977.

[9] G. Fairweather, "A note on the efficient implementation of certain Padé methods for linear parabolic problems," *BIT*, vol. 18, pp. 101–109, 1978.

[10] S. P. Nørsett, "Restricted Padé approximations to the exponential function," *SIAM J. Numer. Anal.*, vol. 15, pp. 1008–1029, 1978.

Second-Order Correlation Method for Bilinear System Identification

R. S. BAHETI, MEMBER, IEEE, R. R. MOHLER, MEMBER, IEEE, AND H. A. SPANG, III, SENIOR MEMBER, IEEE

***Abstract*—An algorithm is developed to estimate parameters of a class of discrete-time bilinear systems using second-order correlations. It is shown that for pseudorandom binary and ternary input signals the computations in the estimation algorithm can be simplified. The method is compared with least-squares algorithm for a model of nuclear fission. The proposed algorithm gives a computationally simple and effective way of parameter estimation.**

I. Introduction

The parameter identification of linear systems from input–output measurements has received considerable attention in the literature [1], [2]. In comparison, bilinear system identification is in its infancy. Balakrishnan [3] proposed a maximum-likelihood algorithm for bilinear system estimation. Bruni and DiPillo [4] identified deterministic bilinear systems by the solution of a nonlinear multipoint boundary value problem. Koch [5] suggested the state and parameter identification of the stochastic bilinear system by augmentation of the state vector and solving the nonlinear filtering problem. Recently, Beghelli and Guidorzi [6] have presented an identification algorithm based on an input–output model of the bilinear system with a finite number of terms linear in parameters.

Mehra [7] and Isermann and Bauer [8] have developed a two-step identification approach for linear systems with correlation analysis and least-squares estimation. In this paper, the method is extended to discrete-time bilinear systems. A canonical state-variable model is transformed to an equivalent input–output relation. The first- and second-order correlations are used to directly estimate the unknown system parameters. Least-squares method is used as a second step to improve the estimates. The algorithm is simplified for pseudorandom binary and ternary input signals. The results are applied to a nuclear fission model.

Manuscript received May 3, 1977; revised October 12, 1978 and June 23, 1980. Paper recommended by P. E. Caines, Past Chairman of the Identification Committee. This work was supported in part by the National Science Foundation under Grant ENG 77-07027.

R. S. Baheti and H. A. Spang, III are with the Research and Development Center, General Electric Company, Schenectady, NY 12301.

R. R. Mohler is with the Department of Electrical and Computer Engineering, Oregon State University, Corvallis, OR 97331.

II. Problem Formulation

A scalar-input scalar-output discrete-time stochastic bilinear state variable model is represented by

$$x(k+1)=Ax(k)+Bx(k)u(k)+cu(k)+gw(k)$$
$$y(k)=dx(k)+v(k) \qquad k=0,1,\cdots \tag{1}$$

where $x(k)$ is the n dimensional state vector. The input $\{u(k)\}$ and output $\{y(k)\}$ are assumed stationary stochastic sequences. A, B, c, d, and g are constant matrices of dimensions $n\times n$, $n\times n$, $n\times 1$, $1\times n$, and $n\times 1$, respectively. $\{w(k)\}$ and $\{v(k)\}$ are assumed Gaussian white noise sequences with mean and variances given by

$$E\{v(k)\}=0, \qquad E\{v(k)v(j)\}=\gamma\delta_{kj}$$
$$E\{w(k)\}=0, \qquad E\{w(k)w(j)\}=q\delta_{kj}$$
$$E\{w(k)v(j)\}=0 \qquad k,j=0,1,\cdots, \tag{2}$$

γ and q are assumed finite. δ_{kj} denotes the Kronecker delta function. The input signal $\{u(k)\}$ is assumed uncorrelated with the state noise $\{w(k)\}$ and the measurement noise $\{v(k)\}$. The aim of an identification experiment is to estimate the unknown but constant parameters in A, B, c, and d from the observations of the input and output of the system. If the linear part of the system is completely observable, then it can be shown [6], [9] that the system (1) is equivalent to a bilinear system, where A, B, c, d have the following form:

$$A=\begin{bmatrix} 0 & 1 & 0 & \cdots & \\ 0 & 0 & 1 & \cdots & \\ \vdots & & & & \\ 0 & & \cdots & 0 & 1 \\ \hline a_1 & a_2 & \cdots & & a_n \end{bmatrix}$$

$$B=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & & & \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}$$

$$c=\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}, \qquad d^T=\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{3}$$

For the identification algorithm presented in this paper, it is necessary to transform the state-variable model (internal description) to an input–output relation (external description). For a bilinear system with A, B, c, d defined by (3), the complexity in the external description increases with increase in the order of the system [6], [10]. To obtain a simple input–output model, a specific class of bilinear systems is considered in the following work. For this class of systems only the first- and second-order correlations are needed to estimate the unknown parameters.

0018-9286/80/1200-1141$00.75