IDENTIFICATION OF BILINEAR PROCESSES

by

Rodrigo G. Castro

Thesis Submitted to the Faculty of the Graduate School
of the University of Maryland in Partial Fulfillment
of the Requirements for the Degree of
Master of Science
1974

TABLE OF CONTENTS

CHAPTER			PAGE
I.	MOTIVATION AND FORMULATION OF THE PROBLEM: IDENTIFICATION OF BILINEAR SYSTEMS		
	1.1	Motivation	1
	1.2	Discussion of the mathematical model used	2
	1.3	The problem of identification	4
II.	REALIZATION THEORY AND ITS RELATIONS TO THE IDENTIFICATION PROBLEM OF BILINEAR SYSTEMS		
	2.1	The nonparametric model via kernels	9
	2.2	Minimal realizations	12
	2.3	A realization algorithm	19
III.	DETERMINISTIC IDENTIFICATION OF BILINEAR PROCESS		
	3.1	Direct identification of the kernels of the input output representation	24
	3.2	The formulation of the general identification problem for a non-parametric model	31
	3.3	Least squares identification of bilinear processes	33
	3.4	Insufficient runs	37
IV.	STOCHASTIC IDENTIFICATION OF BILINEAR PROCESSES		
	4.1	Identification of bilinear systems via Kalman filtering	39
	4.2	A correlation algorithm	43
V.	FUTURE RESEARCH		46
REFERENCES			48

CHAPTER 1

MOTIVATION AND FORMULATION OF THE PROBLEM: IDENTIFICATION OF BILINEAR SYSTEMS

1.1 Motivation.

Linear models are frequently utilized to approximate the dynamical behavior of non-linear systems. Despite their convenience, linear approximations are often inadequate. So control, realization, estimation and detection problems have been studied for one particular class of non-linear systems ([6] through [12]), these systems have a Lie Group for their natural state space and they have been called "bilinear" because their nonlinearity consists of a multiplicative control term appearing in the equation of their internal dynamics in such a way that they are linear in either, control or state, if one is held fixed. Practical problems of this kind include FM and AM demodulation, frequency stability, gyroscopic analysis, etc. Also of this kind and very common in nature is any problem that obeys the law of mass action such as the population of species process or chemical reactions in which a catalyst is involved. Also cellular processes with enzymes producing bilinear control, transfer between organs with variable diffusivity of membranes causing bilinear control, temperature regulation of the human body with vasomotor control of circulation producing bilinear control, and regulation of carbon dioxide in the lungs. Bilinear nuclear fission, heat transfer, reactor control in the Nuclear and Thermal Control Processes, area and urban dynamics of Socioeconomic Systems can be effectively described with a bilinear model [13].

1.2 Discussion of the Mathematical Model Used.

We will not dwell in this study in too many philosophical questions of the "identification problem". It is rather a straightforward approach that will leave up to the person that chooses to apply the methods and results obtained here, the establishment of his own goals and accuracies.

The first pertinent question to answer in the study of "on the Identification of Bilinear Processes" is indeed: what is a bilinear system? What makes this question such a valid one is the fact that one can see two basically different approaches to the study of Bilinear Systems: the approach that searches for the least linear ambient in which bilinear equations can be embedded [4] and the approach that prefers to study system theory on group manifolds and coset spaces [14].

It is a natural consequence of the first approach to desire to extend some of the interesting results obtained for linear systems (i.e. decomposition of the system into sub-systems of which only one influences the input-output behavior) to the class of bilinear systems.

Some progress is possible towards the accomplishment of this goal, especially when the bilinearity is considered to be produced by the presence of a multiplicative control action <u>besides</u> the additive one. The following equations describe a discrete bilinear system that fits the above description without loss of generality

$$x(k+1) = (F + u(k)G)\, x(k) + b\, u(k)$$

$$y(k) = H\, x(k) \qquad x(0) = x_0 \tag{1.1}$$

where

$y(k)$, $u(k)$ are scalars

$x(k)$, $b \in \mathbb{R}^n$

F, $G \in \mathbb{R}^{nxn}$

$H \in \mathbb{R}^{lxn}$

On the other hand, the approach which prefers to view bilinear systems as a special case of systems which evolve on differentiable manifolds prefers to employ the more differential geometric techniques needed since linear algebra no longer suffices. Although we will make use of the mathematical model that does not contain the additive linear term (for reasons that will become apparent in the next paragraph) it will not be necessary for the purposes of this work to go deep into the relationship between this model and Lie Algebras. We will stress the pertinent consequences on the forms of the system that are relevant to the identification problem.

Because some of the major distinctions between linear and bilinear systems occur mainly when there is no linear term in the mathematical model, the system with additive control will be called "homogeneous in the state". But any inhomogeneous bilinear system can be reduced to a homogeneous in the state by adding an extra component to the state vector [2]. I.e. An input-output map of an inhomogeneous bilinear system which can be realized by

$$z(k+1) = (F + u(k)G)\, z(k) + b\, u(k)$$

$$y(k) = H\, z(k)$$

can be realized by one of the form

$$x(k+1) = (A + u(k)B)\, x(k)$$

$$y(k) = C\, x(k) \qquad (1.2)$$

where

$$A = \begin{bmatrix} 1 & 0 \\ 0 & F \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 0 \\ b & G \end{bmatrix}$$

$$C = (O\ H) \qquad x^T = (1\ z^T)$$

This will be the mathematical model for bilinear systems that we will use. It helps to avoid notational complexities and the results obtained with it are easily extended to the case of multiterminal systems.

The discussion as to whether the initial condition of the system is an equilibrium state or not will be deleted until "controllability" is discussed. An equilibrium state is one for which $Ax(k) = 0$. They come in a natural way in the definition of equivalent realizations. When the system starts from "rest" the initial condition is an equilibrium state.

1.3 The Problem of Identification.

One can think of a particular approach to the identification problem for each circumstance and motivation.

We will basically discuss the situation, where one is able to perform experiments on the system in order to obtain input-output data. In order for these experiments to provide information in a cummulative manner it will be necessary to be able to reset the system every time a new experiment is conducted to the same initial state. This is an assumption which helps in the mathematical formalism of the problem. Otherwise the problem should be treated statistically.

The goal of the identification problem will be that of producing a model that performs in a manner to minimize the value of a performance criterion. That is, the model is to be compared to the original process and the performance criterion will be a function (in general) of the deviation of the output of the model from the output of the plant (under study) given the same intput to both of them. The following diagram should clarify things

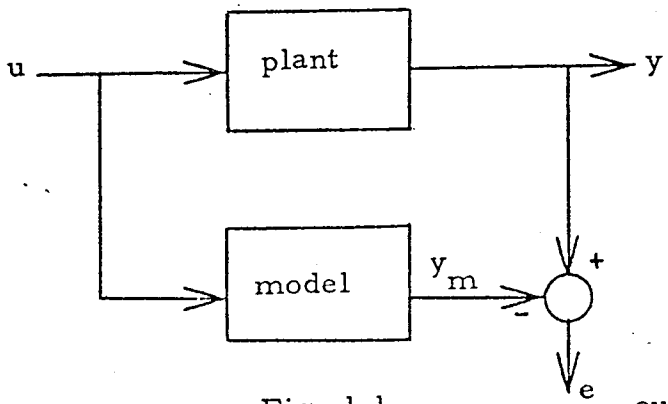


Fig. 1.1 output error

Our goal is basically to minimize the output error of this diagram.

When discussing generalities of System Identification the model is usually characterized by nonparametric representations (impulse responses, transfer functions, Volterra Series) and by parametric representations (a state space description of the model).

For bilinear systems both characterizations are useful. Volterra Series have been used to find regions of stability of non-linear differential equations since the convergence properties of the series are directly related to stability and boundedness of response of a given system [17]. They also have been used, in discrete time, to successfully analyze and synthesize interconnections of linear constant systems and

and zero memory nonlinearities [15]. They are related to the G-functionals of Wiener's methods for the analysis of nonlinear problems [16]. The state space representation has been extensively used in the study of algebraic properties of the system from which very important and useful observations are produced [2].

In linear systems, with scalar inputs and scalar outputs (the importance and weight of the assumption is by no means negligible), the simplicity of the canonical representation together with the well known property of equivalence of realizations allows a very straightforward approach for the identification of a parametric model [18]. Least Squares Identification is even possible with certain nonlinear systems that retain linearity with respect to their parameters [5]. But bilinear systems do not exhibit either linearity on their parameters nor a canonical form amenable to identification. (Brockett [2] has a theorem that explains that every bilinear realization of the form $\dot{x}(t) = (A + \sum_{i=1}^{m} u_i(t)B_i)\, x(t)$; $y = Cx(t)$ where x is an n-tuple y is an m-tuple and A, $\{B_i\}$ and C are matrices of appropriate dimension, is equivalent to one in which the A and B_i matrices are in block triangular form with the diagonal blocks being irreducible. Moreover, it continues, if (A, B_i, C) and (F, G_i, H) are two equivalent realizations in block triangular form with irreducible blocks on the diagonal then there is a permutation π and non-singular matrices P_k such that the diagonal blocks are related by $P_k A_{kk} P_k^{-1} = F_{\pi(k)\pi(k)}$; $P_k B_{kk} P_k^{-1} = G_{\pi(k)\pi(k)}$. We understand by a block being irreducible if there exists no nonsingular

P such that

$$PAP^{-1} = \begin{bmatrix} \tilde{A}_{11} & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}$$

$$PB_iP^{-1} = \begin{bmatrix} \tilde{B}^i_{11} & 0 \\ \tilde{B}^i_{12} & \tilde{B}_{22} \end{bmatrix}$$

where $\tilde{A}_{11}$ and $\tilde{B}^i_{11}$ are square matrices, all of the same dimension.)

The present computer capabilities and existing numerical methods compel us to discuss very superficially at this point a possible solution to the problem of direct identification of a parametric model for discrete bilinear systems. Because the equation describing the system's dynamics is of the form

$$x(k+1) = (A + u(k)B)\, x(k)$$

one can substitute

$$x(k) = (A + u(k-1)B)\, x(k-1)$$

and so on until we get an expression of the form

$$x(k+1) = A^{k+1}x_0 + u(0)A^kBx_0 + u(1)A^{k-1}B^2x_0 + \ldots + u(0)u(1)\ldots u(k)B^{k+1}x_0 \qquad (1.1)$$

the input-output map can then be expressed

$$\hat{y}(k+1) = C\, x(k+1)$$
$$= CA^{k+1}x_0 + u(0)CA^kBx_0 + u(1)\, CA^{k-1}B^2x_0 + \ldots + u(0)u(1)\ldots u(k)\, CB^{k+1}x_0$$

and an algorithm can be provided to determine the components of matrices A, B and C that would minimize a performance criterion of the

form $F(y(k+1), \hat{y}(k+1))$.

We will not dwell on this formulation of the problem any further mainly because a realization of a state variable model has been produced parting from a representation via kernels of the input-ouput map [1]. This representation is closely related to the classical Volterra Series Expansion of non-linear input-output maps. Identification of these kernels is then important per-se and also because a parametric model can be obtained from them if so desired. The rest of our work deals with the identification of these kernels in a deterministic and a stochastic situation. A chapter on the realization theory applicable to the identification problem at hand follows in order to provide a selfcontained treatment and to adjust the existing theory to our needs.

CHAPTER II

REALIZATION THEORY AND ITS RELATIONS TO THE IDENTIFICATION PROBLEM OF BILINEAR SYSTEMS

2.1 The nonparametric model via kernels.

As it was mentioned earlier, very recently an abstract realization theory was obtained for input-output maps, characterizing bilinear systems [1].

Our approach to the identification problem will be the identification of the kernels in the representation of the input-output map data to a certain degree of accuracy. The aforementioned realization theory will be used subsequently for the construction of a parametric model from these kernels. We will also examine the advantages and disadvantages between obtaining a parametric model in this indirect way and a direct identification of a parametric model. Question concerning the order of the model as it relates to the order of the system will have to be resolved. The order is not specified explicitly in nonparametric representations and because nonparametric representations are intrinsically infinite dimentional it is (in principle) possible to obtain a model whose outpuf agrees exactly with the process output. Moreover, the results should be interpreted very carefully.

Understanding, clarifying, and adapting to our needs the realization theory will answer many of these questions. Therefore, our first step is to develop completely the realization theory for homogeneous in the state bilinear systems. This involves adaptation of the realization theory

of Isidory [1] to our problem. We would like to note that the realization theory described here generalizes slightly and simplifies (especially the notation) that of [1] . The state space model we use is repeated for convenience.

$$x(k+1) = (A + u(k)B)\, x(k) \tag{2.1}$$

$$y(k) = C\, x(k) \qquad x(0) = x_0$$

where

$y(k),\ u(k) \in \mathbb{R}$

$x(k) \in \mathbb{R}^n$

$A, B \in \mathbb{R}^{nxn}$

$C \in \mathbb{R}^{lxn}$

The first difference that must be noted between this model and (1.1) is the fact that the initial condition cannot be the zero vector. Zero is an isolated state of the system: $X \subseteq \mathbb{R}^n - \{0\}$.

A very simple expression for the input-output map can be developed departing from (2.1). Let $P_0 = x_0$ and generate recursively

$$P_i = (A\, P_{i-1}\ ;\ B\, P_{i-1}) \qquad i = 1, 2, \ldots \tag{2.2}$$

also generate the 2^{i+1} x 1 vectors

$$u_0 = \begin{bmatrix} 1 \\ u(0) \end{bmatrix}$$

$$u_i = \begin{bmatrix} u_{i-1} \\ u(i)\ u_{i-1} \end{bmatrix} \tag{2.3}$$

and observe

Notice that the w_k are related to the Volterra Series Expansion kernels (for more on that see reference[4]).

It is because of this relation that we believe it is justified to call the $w_{k's}$ in our representation kernels. Therefore knowledge of the kernels w_k, $k = 0, 1, 2, \ldots$ provides a complete description of the model via (2.6). By a nonparametric model of bilinear system we understand the description given in equation (2.7).

2.2 Minimal realizations.

Let us recall that when identifying a certain model of a Plant one of the major concerns is whether the model is unique. For linear systems, it is a well known fact that only the completely controllable and completely observable part of the system is identifiable via input-output observations and this guarantees a minimal model which is unique modulo a choice of basis of the state space. It is not hard to show that a similar result can be produced for bilinear systems described by (2.1) provided an appropriate "controllability" condition is satisfied. The term "reachability" has been preferred to " controllability" and is mainly concerned with whether the associated state space is spanned by the states reachable from a given initial state [2].

We will then call the bilinear system reachable if and only if the reachable set is not confined to a subspace. It has been proved [2] that the reachable set of (2.1) is confined to subspace if and only if there exists T nonsingular such that

$$TAT^{-1} = \begin{bmatrix} \tilde{A}_{11} & 0 \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix}$$

$$TBT^{-1} = \begin{bmatrix} \tilde{B}_{11} & 0 \\ \tilde{B}_{21} & \tilde{B}_{22} \end{bmatrix}$$

$$Tx_0 = \begin{bmatrix} 0 \\ \tilde{x}_0 \end{bmatrix} \qquad (2.8)$$

where the zero blocks are all of the same dimension.

The definition of observability used in connection with bilinear systems is not different from the definition by which we understand the concept in linear systems. A system is said to be observable if none of its states are indistinguishable. We understand by x_0 being indistinguishable from x_1 if for all inputs u, the response y is the same [2].

By a realization of an input-output map in bilinear systems we understand the construction of the triple of matrices plus the initial condition $\{A, B, x_0, C\}$ respective by nxn, nxn, nxl, lxn from an infinite sequence of row vectors $\{\omega_0, \omega_1, \ldots \omega_j, \ldots\}$, respectively 1×2^j, in such a way that (2.6) holds for all j.

Two realizations will be called equivalent if they have the same input-output map. A realization is minimal if and only if the associated state space is both reachable from any initial condition (except $x = 0$) and observable. These minimality conditions can be checked by the algebraic criteria developed below (see also [1]). To this end observe that $\{A, B\}$ can be simultaneously tringularized if and only if one can simultaneously tringularize the associative algebra $\{A, B\}_{AA}$, where by

associative algebra we understand the set of matrices that can be expressed as products and sum of products of the matrices A and B[3]. Denote the least subspace of $\mathbb{R}^n$ invariant under $\{A, B\}$ containing x_0 by

$$\text{gen}_{\{A, B\}}(x_0) \ .$$

Observe that (2.8) and the above comments about the associative algebra imply

$$\text{gen}_{\{A, B\}}(x_0) = \text{gen}_{\{A, B\}_{AA}}(x_0) \ .$$

With the $P_{i's}$ as previously defined (equation (2.2)) construct the $n \times 2^n - 1$ matrix

$$\mathfrak{P}_n \triangleq [P_0 : P_1 : P_2 : \ldots : P_{n-1}]$$

the following lemma has been proved [4]:

The least subspace of $\mathbb{R}^n$ invariant under $\{A, B\}$ and containing x_0 can be expressed as

$$\text{gen}_{\{A, B\}}(x_0) = R(\mathfrak{P}_n) \ .$$

Where $R(\mathfrak{P}_n)$ is the range of the matrix $\mathfrak{P}_n$. Hence

$$\text{gen}\{A, B\}_{AA}(x_0) = R(\mathfrak{P}_n) \ .$$

Since the associative algebra is a larger set containing all linear combinations, products of any two elements, products of products, etc., we know because of the Cayley-Hamilton Theorem that this generation of linearly independent elements terminates once $\{A, B\}$ start appearing at the power of their order in the elements of $\{A, B\}_{AA}$. These comments produce a very important observation: if the system is of order n and we express x(n) as a function of the initial condition x_0 and of the inputs

$$x(n) = A^n x_0 + u(0)A^{n-1} Bx_0 + u(1)A^{n-2} BAx_0 + \ldots$$

$$\ldots + u(1)\ldots u(n-1)B^{n-1}Ax_0 + u(0)u(1)\ldots u(n-1)B^n x_0$$

we see that $x(n)$ is given by x_0 multiplied by linear combinations of elements of the associative algebra $\{A, B\}_{AA}$ where the n-th power already appears. This implies that if one is going to reach n linearly independent states one will be guaranteed to have done so in n-1 steps (if the system is of order n and non-zero controls have been used). The following theorem can now be proved:

Theorem 2.2.1: A bilinear system (represented by (2.1)) of order n has a reachable set which is not confined to a subspace of $\mathbb{R}^n$ iff rank $\mathcal{P}_n = n$.

Proof: Necessity, assume rank $\mathcal{P}_n = n$ form $\mathcal{U}_n$ with u_i's previously defined

$$\mathcal{U}_n = \begin{bmatrix} 1 \\ u_0 \\ u_1 \\ \vdots \\ u_{n-2} \end{bmatrix}$$

examine the product $\mathcal{P}_n \mathcal{U}_n$

$$\mathcal{P}_n \mathcal{U}_n = [P_0 : P_1 : \ldots : P_{n-1}] \begin{bmatrix} 1 \\ u_0 \\ \vdots \\ u_{n-2} \end{bmatrix}$$

$$= P_0 + P_1 u_0 + P_2 u_1 + \ldots + P_{n-1} u_{n-2}$$

$$= x_0 + x(1) + x(2) + \ldots + x(n-1)$$

This vector belongs in the span of reachable states. Therefore rank $\mathcal{P}_n = n$ implies the reachable states span $\mathbb{R}^n$.

Sufficiency: Assume the reachable set spans $\mathbb{R}^n$ or in otherwords that we can reach n linearly independent states. Recall that this can be done in the first n - 1 steps. This implies that these states are in the range of $\mathcal{P}_n$. Therefore rank $\mathcal{P}_n = n$.

Theorem 2.2.1 provides a very useful algebraic criterion for reachability of bilinear systems. Notice that the initial condition is incorporated in the criterion. Similarly a dual observability criterion can be developed. To this end let $Q_0 = C$ and generate recursively

$$Q_i = \begin{bmatrix} Q_{i-1} A \\ Q_{i-1} B \end{bmatrix} \qquad i = 1, 2, 3, \ldots\ldots$$

now form

$$\mathcal{Q}_n \equiv \begin{bmatrix} Q_0 \\ Q_1 \\ \vdots \\ Q_{n-1} \end{bmatrix}$$

Theorem 2.2.2: The bilinear system (4) has no indistinguishable states iff rank $\mathcal{Q}_n = n$.

Proof: Necessity; assume rank $\mathcal{Q}_n = n$. Observe that

$$y(k) = u_{k-1}^T Q_k x_0$$

define the functional

$$\mathcal{Q}_n(\mathcal{U}_n) : \mathbb{R}^n \to \mathbb{R} \quad \text{by}$$

$$x_i \longmapsto \mathcal{U}_n^T \mathcal{Q}_n x_i \equiv \mathcal{Y}_i$$

$\mathcal{Y}_i$ has the form $\mathcal{Y}_i = \sum_{k=1}^{n} y_i(k)$ (the i in y_i reflects the dependence on x_i) suppose now that x_i, x_j are indistinguishable. Then $\mathcal{Y}_i = \mathcal{Y}_j$, i.e.

$$\mathcal{U}_n^T \mathcal{Q}_n x_i = \mathcal{U}_n^T \mathcal{Q}_n x_j$$

$$\mathcal{U}_n^T \mathcal{Q}_n (x_i - x_j) = 0$$

$$\langle \mathcal{U}_n, \mathcal{Q}_n (x_i - x_j) \rangle = 0 \quad .$$

Since $\mathcal{U}_n$ is arbitrary the only vector that is orthogonal to every other vector is the zero vector. $\mathcal{Q}_n (x_i - x_j) = 0$. So $(x_i - x_j) \in \ker \mathcal{Q}_n$. But since it has been assumed that rank $\mathcal{Q}_n = n$ the nullity of $\mathcal{Q}_n = 0 \therefore x_i = x_j$.

Sufficiency: Let X_q denote the subspace of all indistinguishable states. Because a state x is indistinguishable even after we let the system propagate with zero control or with any control u (i.e. Ax, (A + uB)x, etc. are indistinguishable also) X_q is the largest subspace of $\mathbb{R}^n$ invariant under {A, B} and contained in the kernel of C (anihilated by the row vector C).

Denote $X_q = \overline{\text{gen}}_{\{A, B\}}(C)$

now that

$$\mathcal{Q}_n x = 0 \quad \text{for} \quad x \in X_q$$

is obvious from the construction of $\mathcal{Q}_n$ therefore $X_q \subset \ker \mathcal{Q}_n$ to show $\ker \mathcal{Q}_n \subset X_q$ notice that $\mathcal{Q}_n x$ is the multiplication of x by linear combinations of elements of the associative algebra $\{A, B\}_{AA}$ up to the n^{th} power of A and B premultiplied by C. Again, the Cayley-Hamilton theorem indicates that any element of the associative algebra containing

A or B to the $(n+i)^{th}$ power for $i = 0, 1, 2\ldots$ can be expressed as linear combinations of the elements containing A and B to the n-1 power (this indicates that if a state is not distinguishable in n steps it will not be so in $n+i$ steps for $i = 1, 2, 3, \ldots$). Therefore

$$\overline{\text{gen}}_{\{A,B\}}(C) = \text{kernel}\,(\mathcal{Q}_n)$$

and clearly, observability implies $\dim X_q = 0$ and therefore rank $\mathcal{Q}_n = n$.

Having verified these minimality conditions the question of equivalence still requires the verification of the following: "the reachable sets are reached from respective equilibrium states of the models under investigation." Since we do not want to introduce additional restrictions in the identification problem we will not require the initial state (which will be identified along with matrices A, B and C) to be an equilibrium state. To establish equivalence it will be then necessary to allow the system to propagate uncontrolled until it reaches an equilibrium state from the initial condition identified. This realization is equivalent to any other one of the same input-output map that satisfies the minimality conditions from its respective equilibrium state, i.e.

$$(A, B, C, x_0) \xrightarrow{u=0} (A, B, C, x_e) \sim (F, G, H, z_e)$$

where

$$A = PFP^{-1} \qquad C = HP^{-1}$$

$$B = PGP^{-1} \qquad x_e = Pz_e$$

2.3 A realization algorithm.

We procede to describe the Hankel Matrix for bilinear systems modeled by(2.1)and to describe the criteria and the algorithm with which we can obtain a realization, given the kernels that describe the input-output map. The development parallels completely Isidoris' development [1], the difference being that Isidory obtains, a bilinear realization which contains a linear term (an additive control) and whose initial state is the origin of the state space whereas our realization assumes an arbitrary but non-zero initial condition and contains no linear term (we gain a little generality).We will then prove only the 'Realization Lemma" that shows that the difference in essence is that an nxl linear term coefficient has been replaced by an nxl initial state vector. The proofs of the rest of the theorems are then straight-forward with the appropriate substitution, and can be found in [1]. Begin by letting

$$S_{ij} \equiv \omega_{j-1} \qquad j = 1, 2, 3, \ldots$$

and construct $S_{ij}(i = 2, 3, \ldots, j = 1, 2, \ldots)$ recursively: Partition $S_{i-1, j+1}$ as

$$S_{i-1, j+1} = [\, S^1_{i-1, j+1} \quad S^2_{i-1, j+1} \,]$$

where both blocks on the right hand side have the same number of columns. Form then

$$S_{ij} = \begin{bmatrix} S^1_{i-1, j+1} \\ S^2_{i-1, j+1} \end{bmatrix}$$

notice that S_{ij} are $2^{i-1} \times 2^{j-1}$ matrices. The infinite block Hankel matrix is formed from these

$$\mathcal{S} = \begin{bmatrix} S_{11} & S_{12} & \cdots \\ S_{21} & S_{22} & \cdots \\ \vdots & \vdots & \cdots \end{bmatrix}$$

The finite Hankel matrix $\mathcal{S}_{M'M}$ is the blockwise M' x M submatrix in the upper left hand corner of $\mathcal{S}$ this matrix is $(2^{M'}-1) \times (2^{M}-1)$ and contains the same information as the finite sequence of kernels $\{\omega_0, \omega_1, \omega_2, \ldots, \omega_{M'+M-2}\}$.

Now form the following special submatrices:

$\mathcal{S}^1_{M'M}$ = take the first half of all the columns (blockwise) except the first of $\mathcal{S}_{M', M+1}$

$\mathcal{S}^2_{M'M}$ = take the second half of all the columns (blockwise) except the first of $\mathcal{S}_{M', M+1}$

Observe that $\mathcal{S}^1_{M'M}$ and $\mathcal{S}^2_{M'M}$ have the same dimension as $\mathcal{S}_{M'M}$.

Lemma 2.3.1. (Realization Lemma): The finite sequence $\{\omega_0, \omega_1, \omega_2, \ldots, \omega_{M'+M-1}\}$ is realized by the quadruplet (A, B, C, x_0) if and only if

$$\mathcal{S}_{M'M} = \mathcal{Q}_{M'} \mathcal{P}_M \qquad \text{(i)}$$

$$\mathcal{S}^1_{M'M} = \mathcal{Q}_{M'} A \mathcal{P}_M \qquad \text{(ii)}$$

$$\mathcal{S}^2_{M'M} = \mathcal{Q}_{M'} B \mathcal{P}_M \qquad \text{(iii)}$$

where $\mathcal{Q}_{M'}$ and $\mathcal{P}_M$ are previously defined.

Proof: necessity:

$$[S_{11}\ S_{12}\ S_{13} \cdots S_{1M}] = [\omega_0\ \omega_1\ \omega_2 \cdots \omega_{M-1}]$$

$$= [CP_0\ CP_1\ CP_2 \cdots CP_{M-1}]$$

$$= Q_0\ \mathfrak{P}_M$$

$$[S_{21}\ S_{22} \cdots S_{2M}] = \begin{bmatrix} CAP_0 & CAP_1 & \cdots & CAP_{M-1} \\ CBP_0 & CBP_1 & \cdots & CBP_{M-1} \end{bmatrix}$$

$$= Q_1\ \mathfrak{P}_M$$

$$[S_{31}\ S_{32}\ S_{33} \cdots S_{3M}] = \begin{bmatrix} Q_1AP_0 & Q_1AP_1 & \cdots & Q_1AP_{M-1} \\ Q_1BP_0 & Q_1BP_1 & \cdots & Q_1BP_{M-1} \end{bmatrix}$$

$$= Q_2\ \mathfrak{P}_M$$

and so on until (i) is verified. Using this result we can write

$$S_{M',M+1} = \mathfrak{Q}_{M'}\ \mathfrak{P}_{M+1} = [\mathfrak{Q}_{M'}P_0\ \mathfrak{Q}_{M'}P_1 \cdots \mathfrak{Q}_{M'}P_M]$$

$$= [\mathfrak{Q}_{M'}P_0\ \boxed{\mathfrak{Q}_{M'}AP_0}\ \mathfrak{Q}_{M'}BP_0 \cdots \boxed{\mathfrak{Q}_{M'}AP_{M-1}}\ \mathfrak{Q}_{M'}BP_{M-1}].$$

Observe that the submatrix formed from the dotted blocks is by definition $\mathfrak{S}^1_{M'M}$. Therefore we have shown that

$$\mathfrak{S}^1_{M'M} = \mathfrak{Q}_{M'}\ A\ \mathfrak{P}_M$$

Taking the undotted blocks except the first one we can show that

$$\mathfrak{S}^2_{M'M} = \mathfrak{Q}_{M'}\ B\ \mathfrak{P}_M$$

To prove sufficiency it is only necessary to reverse the above arguments. Observe that x_0 and C are respectively the first column of $\mathfrak{P}_M$ and the first row of $\mathfrak{Q}_{M'}$.

The rest of the Theorems are included for the sake of completeness (with the appropriate substitution) without proof (for proofs see [1]).

Theorem (3) (Rank Condition): If for some M' and M

$$\text{rank } \mathcal{S}_{M'M} = \text{rank } \mathcal{S}_{M',M+1} = \text{rank } \mathcal{S}_{M'+1,M}$$

then the finite sequence $\{\omega_0, \omega_1, \omega_2, \ldots, \omega_{M'+M-1}\}$ admits a partial bilinear realization.

And the very important theorems concerning minimal realizations:

Theorem 2.3.2: Let $\{\omega_0, \omega_1, \omega_2, \ldots, \omega_{M0}\}$ be an arbitrary finite sequence of 1×2^j row vectors ($j = 1, 2, \ldots, M_0$). If there exist positive integers M and $M' = M_0 - M + 1$ such that the rank condition is satisfied, then the minimal partial bilinear realization problem has a unique solution (modulo the choice of basis in the state space). The dimension of the realization is equal to rank $\mathcal{S}_{M'M}$ and an actual realization is provided by the quadruplet

$$A = (\mathcal{Q}_{M'}{}^T \mathcal{Q}_{M'})^{-1} \mathcal{Q}_{M'}{}^T \mathcal{S}^1_{M'M} \mathcal{P}_M{}^T (\mathcal{P}_M \mathcal{P}_M{}^T)^{-1}$$

$$B = (\mathcal{Q}_{M'}{}^T \mathcal{Q}_{M'})^{-1} \mathcal{Q}_{M'}{}^T \mathcal{S}^2_{M'M} \mathcal{P}_M{}^T (\mathcal{P}_M \mathcal{P}_M{}^T)^{-1}$$

x_0 = first column of $\mathcal{P}_M$

C = first row of $\mathcal{Q}_{M'}$

where $\mathcal{Q}_{M'}$ and $\mathcal{P}_M$ are $(2^{M'}-1) \times n$ and $n \times (2^M-1)$ matrices respectively such that

$$\mathcal{Q}_{M'} \mathcal{P}_M = \mathcal{S}_{M'M}$$

Theorem 2.2.3. An arbitrary infinite sequence $\{\omega_0, \omega_1, \omega_2, \ldots, \omega_j, \ldots\}$ of 1×2^j row vectors admits a bilinear complete realization if and only if the infinite matrix $\mathcal{S}$ has finite rank n (i.e. there exist M and M' such that rank $\mathcal{S}_{M'+i, M+j}$ = rank $\mathcal{S}_{M'M}$ = n for all $i, j \geq 0$). The dimension of the minimal realization is equal to n and an actual minimal realization is provided by the quadruplet (A, B, x_0, C) as expressed in the previous theoerem.

CHAPTER III

DETERMINISTIC IDENTIFICATION OF BILINEAR PROCESS

3.1 Direct Identification of the Kernels of the Input-Ouput Representation

We obtained in the previous Chapter the following input-output representation of the model portrayed by equation (2.1)

$$y(k) = \omega_k u_{k-1} \qquad k = 0, 1, 2, \ldots \tag{3.1}$$

where $u_{-1} \equiv 1$

$$u_0 = \begin{bmatrix} 1 \\ u(0) \end{bmatrix}$$

$$u_i = \begin{bmatrix} u_{i-1} \\ u_{i-1}\, u(i) \end{bmatrix} \qquad i = 1, 2, 3 \ldots \tag{3.2}$$

and the $\omega_k = CP_k$ (3.3)

where

$$P_0 = x_0$$

$$P_i = [AP_{i-1},\ BP_{i-1}] \quad i = 1, 2, \ldots \tag{3.4}$$

The realization theory examined indicated how to obtain the quadruplet (A, B, x_0, C) from the ω_k's such that these matrices depicted the correct order of the model if the "Rank Condition" was satisfied for all i.

Observe from the two last theorems in the previous Chapter that although they refer to the complete-minimal realization for an infinite sequence $\{\omega_0, \omega_1, \ldots, \omega_j, \ldots\}$ of 1×2^j row vectors, if S has finite

rank n and at this n the rank condition is satisfied, it is only necessary to obtain $M_0 = M' + M - 1$ row vectors. $\mathcal{S}_{M'M}$ is formed from these such that it contains the minimum M' and M necessary to satisfy rank $\mathcal{S}_{M'M} = \text{rank } \mathcal{S}_{M'+i, M+j} = n$ for all $i, j \geq 0$.

These are very useful observations for the identification problem, when one has a priori knowledge of the order of the system and of its bilinearity. This is not a very restrictive assumption. Often one is faced with a bilinear model of known order, which is the result of the underlying theories of the process (see also Chapter I). If under these circumstances we are allowed to perform experiments then it is possible with an appropriate selection of inputs to actually obtain the kernels from the measurements and the input-output representations that we displayed in (3.1). The realization algorithm would produce a canonical model and it would be global in the sense that it would serve as a canonical model for the continuation of input-output pairs from $k = n$ on

If we have no such a priori knowledge of the order of the system and of its bilinearity then we must realize that the rank condition is impossible to check for all $i, j > 0$ because we require an infinite amount of data. It is possible for the $\mathcal{S}$ matrix to have stationary rank n for $M' + i$, $M + j$ where $0 < i, j < q$, a finite number and for this rank to increase for $i, j > q$. Therefore one should not be mislead to believe that one has identified the canonical part of a bilinear system whose order was not known a priorir with the realistically finite amount of checks and computations possible. We will comment on this problem

further after we explain how the special purpose inputs can be generated in order to obtain the kernels exactly. It is necessary to remember, that the assumption of previous knowledge of order and bilinearity of the system provides us automatically with the knowledge of an upper bound number of kernel vectors needed for the complete nonparametric model of the system.

We already mentioned this upper bound at the beginning of the chapter but because of its relevance at this point lets understand how to obtain this upper bound from the definition of the kernel vectors as they are obtained from the state space model:

$$\omega_k = CP_k$$

where $P_0 = x_0$

$$P_i = [AP_{i-1}, BP_{i-1}]$$

the comments about associative algebra made in chapter II and the Cailey-Hamilton Theorem as they were used to prove the "Controllability" and observability Theorems (2.2.1 and 2.2.2) should allow us to understand very clearly that for a bilinear system of order n ω_k, where $k \geq n - 1$, contains redundant information. To assure ourselves that this information is included in the matrix $S_{M'M}$ we should form the $k - 1 \times k - 1$ block order matrix which has the form

$$\begin{bmatrix} \omega_0 & \omega_1 & \omega_2 \cdots & \omega_{k-1} \\ \omega_1 & \omega_2 & \omega_3 \cdots & \omega_k \\ \omega_2 & \omega_3 & \cdots\cdots & \cdot \\ \vdots & \vdots & & \\ \omega_{k-1} & \omega_k & \cdots\cdots & \omega_{2k-1} \end{bmatrix}$$

An upper bound in the number of kernels that one must identify for a system of known order n is then 2n-1.

It is apparent from the realization theory that there exists a unique sequence of kernels associated with a system that we know to be bilinear. Since our known model may not be canonical the rank of $S_{M'M}$ say q, will be smaller or equal to its maximum possible rank n. So then for the first algorithm the assumptions are: a) previous knowledge of order and bilinearity and b) that it is possible to reset and run the test as many times as necessary. Since for the observation at the 2n-1 vector one needs to identify a $1 \times 2^{2n-1}$ vector we need 2^{2n-1} runs. The way of obtaining the components of the kernel vectors is obviously not unique but we propose the following scheme: Arrange the input-output map as indicated in Eq. 3.5 on the next page, where ω_{ki} is the i^{th} component of the k^{th} kernel vector.

For illustration purposes lets obtain four kernel vectors $\omega_0, \omega_1, \omega_2$, and ω_3.

On your first run let

$$u(0) = u(1) = u(2) = 0$$

from (3.5) we obtain

$$y(0) = \omega_0$$

$$y(1) = \omega_{11}$$

$$y(2) = \omega_{21}$$

$$y(3) = \omega_{31}$$

$$
\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ y(3) \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & u(0) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & u(0) & u(1) & u(0)u(1) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & u(0) & u(1) & u(0)u(1) & u(2) & u(2)u(0) & u(2)u(1) & u(0)u(1)u(2) & \\ \vdots & & & & & & & & & & & & & & & \end{bmatrix} \begin{bmatrix} \omega_0 \\ \omega_{11} \\ \omega_{12} \\ \omega_{21} \\ \omega_{22} \\ \omega_{23} \\ \omega_{24} \\ \omega_{31} \\ \omega_{32} \\ \omega_{33} \\ \omega_{34} \\ \omega_{35} \\ \omega_{36} \\ \omega_{37} \\ \omega_{38} \\ \vdots \end{bmatrix}
$$

Equation (3.5)

Now obtain a second run of input-output pairs but this second time let

$$u(0) \neq 0 \quad , \quad u(1) = u(2) = 0$$

Since this is just an illustration, let $u(0) = 1$. We obtain

$$y(1) = \omega_{11} + \omega_{12}$$

$$y(2) = \omega_{21} + \omega_{22}$$

$$y(3) = \omega_{31} + \omega_{32}$$

Since one already has the values of ω_{11}, ω_{21} and ω_{31} one can solve for ω_{12}, ω_{22} and ω_{32}.

Obtain a third run of input-output pairs letting $u(1) \neq 0$ (say $u(1) = 1$), $u(0) = u(2) = 0$ now

$$y(2) = \omega_{21} + \omega_{23}$$

$$y(3) = \omega_{31} + \omega_{33}$$

again one can solve for $\omega_{23} + \omega_{33}$. The fourth time let $u(0) \neq 0$ and $u(1) \neq 0$ (say $u(0) = u(1) = 1$) and $u(2) = 0$

$$y(2) = \omega_{21} + \omega_{23} + \omega_{24}$$

$$y(3) = \omega_{31} + \omega_{32} + \omega_{34}$$

solve for ω_{24} and ω_{34}. The fifth time let $u(2) \neq 0$ (say $u(2) = 1$), $u(0) = u(1) = 0$

$$y(3) = \omega_{31} + \omega_{35}$$

solve for ω_{35}. The sixth time $u(0) \neq 0, u(2) \neq 0$ (say $u(0) = u(2) = 1$) and $u(1) = 0$

$$y(3) = \omega_{31} + \omega_{32} + \omega_{35} + \omega_{36}$$

solve for ω_{36}. The seventh time $u(1) \neq 0$, $u(2) \neq 0$ (say $u(1) = u(2) = 1$) and $u(0) = 0$

$$y(3) = \omega_{31} + \omega_{33} + \omega_{35} + \omega_{37}$$

solve for ω_{37}.

Finally run the tests through an eighth time using all non-zero controls (say $u(0) = u(1) = u(2) = 1$)

obtain $$y(3) = \omega_{31} + \omega_{32} + \omega_{33} + \omega_{34} + \omega_{35} + \omega_{36} + \omega_{37} + \omega_{38}$$

solve for ω_{38}.

It is interesting to make a chart indicating in which running the components of the kernel vectors were obtained and the values of the controls at each run

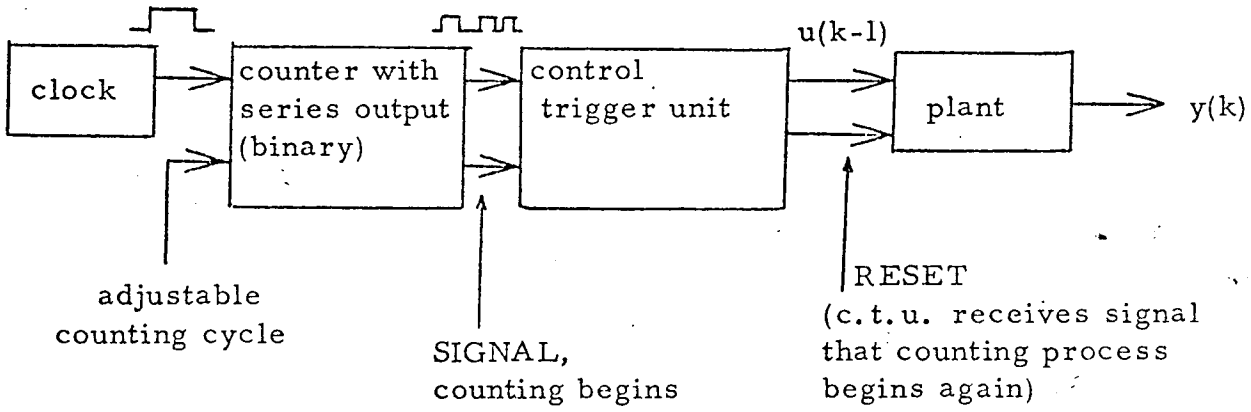
running	controls u(2)	u(1)	u(0)	kernel vector comp.
1	0	0	0	ω_0, ω_{11}, ω_{21}, ω_{31}
2	0	0	1	ω_{12}, ω_{22}, ω_{32}
3	0	1	0	ω_{23}, ω_{33}
4	0	1	1	ω_{24}, ω_{34}
5	1	0	0	ω_{35}
6	1	0	1	ω_{36}
7	1	1	0	ω_{37}
8	1	1	1	ω_{38}

Notice that under the j^{th} running we have j-1 written in binary form. The following general observations can be made:

i) The order in which one obtains the kernel vector components is not unique but the order that we propose here is sequential in the sense that on each running one obtains a new set of components using

data from the previous runnings and is not necessary to wait until all the runnings have been completed to obtain some of the kernel vectors. As a matter of fact, one obtains the j^{th} component of the kernel vector when one conducts the j^{th} running.

ii) Obtaining the kernel vector components in the order that I propose is facilitated because of the existence of binary digital equipment. The following diagram depicts a possible scheme



3.2 The Formulation of the General Identification Problem for a Non-Parametric Model.

If one doesn't know the order of the system it is not possible to assert that any model produced via identification of the kernel vectors is of the correct rank because of the impossibility to check the "rank condition" (Ch. II) for the $\mathcal{S}$ matrix. Instead we suggest to obtain a model identifying a reasonably large number of kernel vectors. If one doesn't obtain a model of the correct order then every time we run the iterations with a different set of inputs we are going to get a different model. This leads to the question: do we want a different model per iteration set or should we form a bigger iteration concatenating the runs

of the smaller iteration sets and produce one model that minimizes the error with respect to this larger data set?

In the next section we describe how to obtain a model that minimizes the error of the bigger data set in the least-squares sense (deterministic). Here I formalize the problem at hand:

Situation: Order not known

Possibility 1):

Obtain $\{\omega_0, \omega_1, \ldots, \omega_{M_0}\}_j$ from the iteration set $\{u_j^1(0), u_j^1(1), u_j^2(1), \ldots, u_j^1(k), u_j^2(k), u_j^3(k), \ldots, u_j^{2k}(k), \ldots, u_j^1(M_0-1), u_j^2(M_0-1), \ldots, u_j^{2^{M_0}}(M_0-1), y_j^1(0), y_j^1(1), y_j^2(1), \ldots, y_j^1(k), y_j^2(k), y_j^3(k), \ldots, y_j(k), \ldots, y_j^1(M_0), y_j^2(M_0), \ldots, y_j^{M_0}(M_0)\}$ and obtain $R_j \equiv \{A, B, \chi_0, C\}_j$ for each iteration set. Each R_j models the set of input-output pairs used in that particular iteration exactly but $R_j \neq R_\ell$ (including a possible difference between the order of R_j and R_ℓ also).

Possibility 2):

Obtain $\{\omega_0, \omega_1, \ldots, \omega_{M_0}\}_{unique}$ from the bigger iteration set formed by concatenating all the input-output pairs of each iteration set of possibility 1) and obtain a model $R_{u(unique)}$ such that the error between the output-sets obtained by running an input set corresponding to the j^{th} iteration through R_u and through its repsective R_j is minimized. Notice that the error between the output sets indicated above can vary if the orders of the models corresponding to different iteration sets are different.

3.3 Least-Squares Identification of Bilinear Processes

Instead of going through the steps indicated in the last section to obtain the bigger iteration set we simply form observation vectors like the following

$$\begin{bmatrix} y^1(k) \\ y^2(k) \\ \vdots \\ y^{2^k}(k) \\ y^{2k+1}(k) \\ \vdots \\ y^m(k) \end{bmatrix}$$

The number of runs m is such that the matrix made up of the controls in our matrix formulation of the relations $y(k) = u_{k-1}^T \omega_k^T$ has more rows than columns.

The assumption that at each time we can find a set of controls that produce control submatrices of column rank $= n_k$ where n_k is the number of components of the kernel vector corresponding to the k^{th} time (i.e. $n_k = 2^k$) will allow us to find a unique solution that minimizes the error between the outputs that make up our observation vector and the outputs obtained when the same inputs are fed through our model.

The matrix formulation of the problem has the following form indicated on the next page. To obtain a solution we adopt the following matrix notation

$$\underset{\sim}{y}_m = \underset{\sim}{U}_{mn} \underset{\sim}{\omega}_n$$

Equation(3.6)

$$
\begin{bmatrix}
y^1(0) \\ \vdots \\ y^{m_k}(0) \\ y^1(1) \\ \vdots \\ y^{m_k}(1) \\ y^1(2) \\ \vdots \\ y^{m_k}(2) \\ \vdots \\ y^{(1)}{}_k \\ y^2(k) \\ \vdots \\ y^{m_k}(k)
\end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\
0 & 1 & u^1(0) & 0 & 0 & 0 & 0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\
0 & 1 & u^{m_k}(0) & 0 & 0 & 0 & 0 & \cdots \\
0 & 0 & 0 & 1 & u^1(0) & u^1(1) & u^1(0)u^1(1) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\
0 & 0 & 0 & 1 & u^1(0) & u^1(1) & u^{m_k}(0)u^{m_k}(1) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \\
 & & & & & & & u^{1\,T}_{k-1} \\
 & & & & & & & u^{2\,T}_{k-1} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \vdots \\
 & & & & & & & u^{m_k\,T}_{k-1}
\end{bmatrix}
\begin{bmatrix}
\omega_0 \\ \omega_{11} \\ \omega_{12} \\ \omega_{21} \\ \omega_{22} \\ \omega_{23} \\ \omega_{24} \\ \vdots \\ \bar{\omega}_k
\end{bmatrix}
$$

where

$$m = (k+1)m_k \qquad (m_k > k)$$

$$n = 2^{k+1} - 1 \qquad (m > n)$$

i.e. y_m is an m vector, U_{mn} is an mxn matrix of rank n and ω_n is an n vector.

The Least Squares Estimate Theorem indicates that if y_m is an m vector and U_{mn} an mxn matrix with linearly independent columns, there is a unique n vector $\hat{\omega}_n$ which minimizes $||y_m - U_{mn}\,\omega_n||$ over all ω_n (the norm takes as the Euclidean m-space norm). Furthermore

$$\hat{\omega}_n = (U_{mn}^T U_{mn})^{-1} U_{mn}^T y_m$$

Notice that the form of U_{mn} allows a complete decoupling of the estimation of each kernel vector, i.e. we could solve the least-squares problem for each kernel vector independently forming a partial observation vector

$$\begin{bmatrix} y^1(j) \\ y^2(j) \\ y^3(j) \\ \vdots \\ y^{m_k}(j) \end{bmatrix} = \begin{bmatrix} u_{j-1}^{1\,T} \\ u_{j-1}^{2\,T} \\ u_{j-1}^{3\,T} \\ \vdots \\ u_{j-1}^{m_k\,T} \end{bmatrix} [\omega_j]$$

We are assured that the columns of the matrix formed with the controls are linearly independent because they are assumed linearly independent in the matrix U_{mn}. Denote the matrices in this individual kernel estimation problem respectively by

$$y(j) = U_{m_k n_j} \omega_j \qquad (n_j = 2^j)$$

and obtain a least-squares estimate for each kernel vector

$$\hat{\omega}_j = (U_{m_k n_j}^T U_{m_k n_j})^{-1} U_{m_k n_j}^T y(j)$$

It is then evident that the total error in the estimation of the complete observation vector is equal to the sum of the errors of the estimation of the partial observation vectors.

One can pose another interesting question at this point: is it implied that if we estimate more kernel vectors the error is going to keep increasing? There obviously is a trade-off between the number of kernel vectors that one identifies and the error because by increasing the order of the model one is bound to obtain a model of the correct order corresponding to the system under identification.

Lets remember that for a bilinear system of order n, ω_k where $k \geq n-1$, contains redundant information. The Least Square Error should be interpreted carefully because if one has already identified more kernel vectors than the order of the system one is adding redundant error to the estimate. It is convenient therefore to obtain a state space realization each time that one chooses to identify more kernel vectors to see if the state space model increases in order. If the order of the state space model reaches a stationary point the error in the Least Squares Estimate should be taken to be the sum of the errors of the least squares estimate of the minimum amount of kernel vectors necessary to obtain such realization. Recall, however, that this stationing in the order

order of the state space model implies that the rank of $\mathcal{H}_{M'M}$ is also stationary and if we don't know the order of the system a priori it is impossible to assert that the rank condition is satisfied for all i. A lot of judgement has to be used when deciding to adopt a particular model for a system.

3.4 Insufficient Runs.

If in a particular circumstance it is not possible to run sufficient set of controls through a system for the purpose of making the 2^j necessary observations to identify the j first kernel vectors then the columns of the matrix made up by the controls are obviously not going to be linearly independent. This implies that when we try to calculate ω_j, $(U^T_{m.n_j} U_{m.n_j})$ is not going to be an invertible matrix.

The Least Squares Theory tells us that under those circumstances the ω_j that solves the least-squares problem is not unique. All the ω_j under those circumstances form a linear variety of E^{n_j}. Since in every linear variety there is always a vector of smallest norm, in order to provide our problem with a unique solution we are going to define this particular $\hat{\omega}_j$. But before we indicate what the form of the solution is, some important observations concerning least-squares theory are at hand.

From the vector space point of view, given $y = Ax$ where y is an m vector, A is an mxn matrix and x is an n vector, if $m > n$ and A has linearly independent columns, $A\hat{x} = A(A^TA)^{-1}A^Ty$ is the perpendicular projection of the vector y on the space spanned by the columns of A.

The error which is minimum is the difference between y and this projection. Therefore, from this vector space point of view, the situation $m < n$ implies that it is possible to find a linear combination of the n x vectors that fits y exactly (that is if out of these n x vectors these are at least m linearly independent). Assuming that a set of controls can be found to form a control matrix that has at least column rank then the least square solution that produces a vector of minimum norm is given by

$$\hat{x}_m = A^T(AA^T)^{-1} y \ .$$

Zero error is evidenced when we substitute in the error expression

$$e = \| y - \hat{y} \|$$
$$= \| y - AA^T(AA^T)^{-1} y \| = 0$$

One might inquire whether it is not advantageous to obtain this zero error? The disadvantage lies in the restrictiveness of the solution. The model obtained is only good to handle the limited amount of data that was used to determine it. This is in sharp contrast with the larger data handling possibilities provided by models identified with larger amounts of data.

CHAPTER IV

STOCHASTIC IDENTIFICATION OF BILINEAR PROCESSES

4.1 Identification of Bilinear Systems via Kalman Filtering

One of the advantages of presenting a least-squares solution for the identification of the kernel vectors is the fact that the problem can very easily be imbedded in a probabilistic framework.

This becomes apparent when we examine Kalman's filtering theory. A dynamical model of a random process consists of

i) A vector difference equation

$$x(k+1) = \phi\, x(k) + v(k)$$

$$y(k) = C\, x(k) + e(k)$$

where $\{ v(k),\ k = 1, 2, \ldots \}$ and $\{ e(k),\ k = 1, 2, \ldots \}$ are sequences of independent equally distributed random vectors with zero mean values and covariance matrices R_1 and R_2 respectively.

ii) An initial random vector $x(0)$ together with an initial estimate $\hat{x}(0)$ having covariance $E[(\hat{x}(0)-x(0))(\hat{x}(0)-x(0))^T] = R_0$. In addition it is assumed that the random vectors $x(0)$, $v(j)$ and $e(k)$ are uncorrelated for $j \geq 0$, $k \geq 0$.

The estimation problem is that of obtaining the linear minimum-variance estimate of the state vector x from the measurements y.

From the vector space point of view $\hat{x}(k|j)$, which is the optimal estimate of $x(k)$ given the observations y up to instant j, is the appropriate projection of $x(k)$ onto the space generated by the random vectors $y(0), y(1), \ldots, y(j)$.

We will be concerned exclusively with the case $k = j + 1$ in $\hat{x}(k/j)$. Kalman's theorem [19] providing a solution to the recursive estimation problem states that the best estimate of $x(k)$, in the sense of least squares, given the observed outputs $y(1), y(2) \ldots y(k)$ of the dynamical model of a random process described above, is given by the recursive equations: ($\hat{x}(k/k)$ is denoted $\hat{x}(k)$ for simplicity)

$$\hat{x}(k) = \phi\, \hat{x}(k-1) + \Gamma(k)\, [\, y(k) - C\, \phi \hat{x}(k-1)]$$

$$\hat{x}(0) = m$$

where

$$m = E[x(0)]$$

$$\Gamma(k) = S(k)\, C^T [C\, S(k) C^T + R_2]^{-1}$$

$$S(k) = \phi\, P(k-1)\, \phi^T + R_1$$

$$P(k) = S(k) - \Gamma(k)\, C\, S(k)$$

$$S(0) = R_0$$

$S(k)$ is the covariance matrix of the a priori estimate of $x(k)$ given $y(1), \ldots, y(k-1)$ and the matrix $P(k)$ is the covariance matrix of the posterior estimate of $x(k)$ given $y(1), \ldots, y(k)$.

In order to provide a state variable formulation to the kernel vector problem we will take advantage of the decoupling observed. That is the fact that each kernel vector estimation can be formulated as a separate problem and incorporate each in a state variable model. It will only be necessary to concatenate them to give a state variable model to the overall problem.

Since the kernel vectors of the system we are dealing with are assumed

to be time invariant and deterministic in nature the difference equation involving a particular kernel-vector can be expressed as

$$\omega_k(j+1) = \omega_k(j)$$

and

$$y^j(k) = (u^j_{k-1})^T \omega_k{}^T(j) + e^j$$

will depict the observation at time k obtained at the j^{th} running, each of which is contaminated with white measurement noise (or could very will be Gaussian $(0, \lambda)$ random noise).

The least squares identification problem of the kernel vectors can then be stated as a Kalman filtering problem with $\Phi = 1$, $R_1 = 0$ and $R_2 = \lambda^2$. Observe that when dealing with a particular kernel vector the time (k in $y^j(k)$) is fixed. Updating is provided by each run or iteration (j in $y^j(k)$).

Substituting our values in Kalmans Filtering and Ricatti equations we obtain.

$$\hat{\omega}_k{}^T(j) = \hat{\omega}_k{}^T(j-1) + \Gamma(j)\,[\, y^j(k) - (u^j_{k-1})^T \omega_k{}^T(j-1)]$$

where $\hat{\omega}_k(0) = E[\,\omega_k(0)]$

$$\Gamma(j) = S(j)u^j_{k-1}((u^j_{k-1})^T S(j) u^j_{k-1} + \lambda^2)^{-1}$$

$$S(j) = P(j-1)$$

$$P(j) = P(j-1) - \Gamma(j)\,(u^j_{k-1})^T\, P(j-1)$$

$$S(0) = R_0$$

The greatest advantage in using this filtering approach is the fact that one doesn't have to wait to complete an iteration vector to begin the calculations. This approach provides the possibility of doing the computation on line which in turn allows the possibility of weighting the value of further iterations before one has actually performed them.

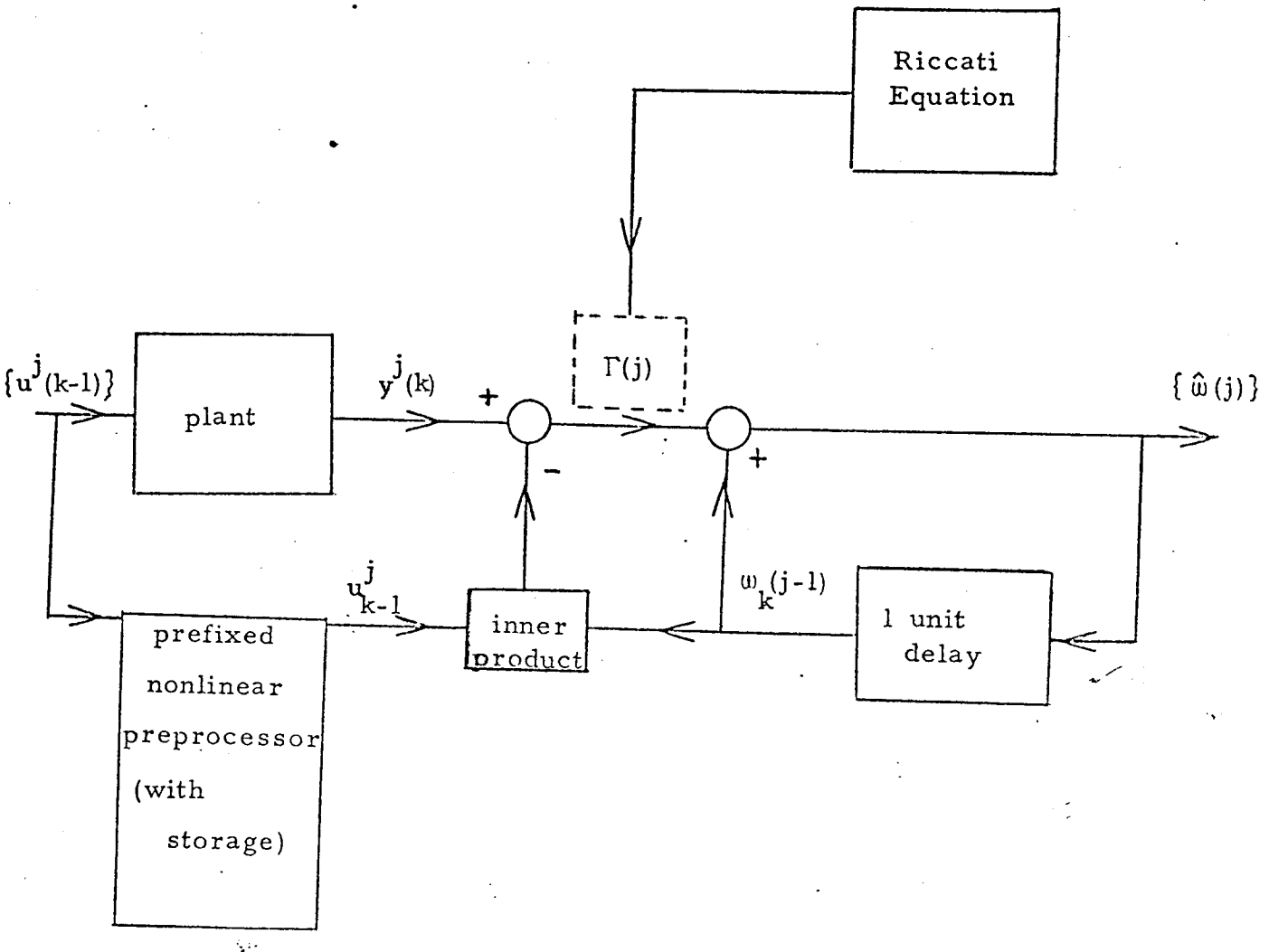

Riccati Equation
Γ(j)
{u^j(k-1)}
plant
y^j(k)
+
-
+
{ ŵ(j)}
u^j_{k-1}
inner product
w_k(j-1)
1 unit delay
prefixed nonlinear preprocessor (with storage)

Figure 4.1

Lets remember that the iterations necessary to obtain ω_k are more than the necessary to obtain ω_q where $q < k$. Therefore in the process of obtaining ω_k one can obtain all the "previous" ω_i. The block diagram of figure 4.1 explains how the overall identification is accomplished on line.

4.2 A Correlation Algorithm.

The input-output representation via kernels allows us to develope the formulation of an algorithm inspired by the Crosscorrelation Algorithms which are so popular in the identification of linear systems. Crosscorrelation algorithms allow us to deal with the situation of input measurement error or for that matter, we can apply them when it is possible to provide stochastic inputs to the system undergoing identification.

At this point we only attempt to provide a formulation of the problem. So, in order to maintain certain consistency in this chapter we will again develope the formulation identifying one kernel vector at the time.

We will adopt the following notation which is more appropriate for the purpose at hand:

$$y(k) = \omega_k \, u_{k-1} = \sum_{i=1}^{2^k} u_{(k-1)i} \, \omega_{ki}$$

where $u_{(k-1)i}$ and ω_{ki} are the i^{th} component of the k-1 control vector and k^{th} kernel vector respectively.

With this notation it is clear that $y(k)$ is a weighted sum of all the controls used up to time k-1 and of products of these controls made up of two, three, etc., up to all k-1 controls. i.e.,

$$y(k) = \omega_{k1} + \omega_{k2} u(0) + \omega_{k3} u(1) + \omega_{k4} u(0)\, u(1) +$$
$$\omega_{k5} u(2) + \omega_{k6} u(0)\, u(2) + \omega_{k7} u(1)\, u(2) + \omega_{k8} u(0)\, u(1)\, u(2)$$
$$+ \ldots + \omega_{k2^k} u(0)\, u(1)\, u(2) \ldots u(k-1)$$

y(k) is quite a non-linear function of the controls and this is why we qualify the fact that the algorithm presented here is inspired by the "Modified Discrete Crosscorrelation Algorithm" [20] and not a straightforward application of it.

We procede as follows: take y(k) and crosscorrelate it with each one of the inputs (which we assume to be a stochastic process) and with all the input combinations that appear in u_{k-1} (notice that the latter ones will be higher order correlations). We obtain the following matrix equation

$$\begin{bmatrix} R(y(k), u_{(k-1)1}) \\ R(y(k), u_{(k-1)2}) \\ R(y(k), u_{(k-1)3}) \\ \vdots \\ R(y(k), u_{(k-1)2^k}) \end{bmatrix} = \begin{bmatrix} R(u_{(k-1)1}, u_{(k-1)1}) & R(u_{(k-1)2}, u_{(k-1)1}) \ldots R(u_{(k-1)2^k}, u_{(k-1)1}) \\ R(u_{(k-1)1}, u_{(k-1)2}) & \vdots \\ \vdots & \\ R(u_{(k-1)1}, u_{(k-1)2^k}) & \ldots\ldots\ldots\ldots \quad R(u_{(k-1)2^k}, u_{(k-1)2^k}) \end{bmatrix} \begin{bmatrix} \omega_{k1} \\ \omega_{k2} \\ \\ \\ \omega_{k2^k} \end{bmatrix} \tag{4.1}$$

where $R(u_{(k-1)i}, u_{(k-1)j}) \equiv E[u_{(k-1)i} u_{(k-1)j}]$ (4.2)

and $R(y(k), u_{(k-1)i}) \equiv E[y(k) u_{(k-1)i}]$ (4.3)

Notice that equation (4.2) warns about the care with which the stochastic inputs have to be chosen. Very few stochastic process have high order correlations expressible in closed form (the Gaussian process is one of them

but we will not dwell in this particular question). This is of considerable importance for computations, and for applications we must have explicit values for the elements of the matrix of the right hand side of equation (4.1). The components of the left-hand side of equation (4.1) cannot be computed exactly but several iterations will allow us to update them in the following way

$$R^M(y^M(k), u^M_{(k-1)i}) \equiv \frac{1}{M+1} \sum_{j=0}^{M} y^j(k)\, u^j_{(k-1)i} \tag{4.4}$$

where $y^j(k)$ and $u^j_{(k-1)i}$ are respectively the k^{th} output at the j^{th} running and the i^{th} component of the u_{k-1} vector used in the j^{th} running.

Observe that the number of iterations is not established by the number of components of the kernel vector being identified but by the desired accuracy of the left hand side of (4.1).

(4.4) can be written in updating form:

$$R^M(y^M(k),\ u^M_{(k-1)i}) = R^{M-1}(y^{M-1}(k),\ u^{M-1}_{(k-1)i}) + \frac{1}{M+1}\left[y^M(k) u^M_{(k-1)i} - R^{M-1}(y^{M-1}(k),\ u^{M-1}_{(k-1)i}) \right] \tag{4.5}$$

expressing eq. (4.1) in the following compact form

$$F(y(k), u_{k-1})^M = R(u_{k-1}, u_{k-1})\omega_k^M \tag{4.6}$$

we can do the updating directly on (4.6)

$$R^M(u^M_{k-1},\ u^M_{k-1})\,\omega_k^M = R^{M-1}(u^{M-1}_{k-1},\ u^{M-1}_{k-1})\,(\omega_k^{M-1} + \frac{1}{M+1}\left[y^M(k) u^M_{k-1} - R^{M-1}(u^{M-1}_{k-1},\ u^{M-1}_{k-1})\,\omega_k^{M-1} \right] \tag{4.7}$$

CHAPTER V

FUTURE RESEARCH

The study of bilinear systems is by no means complete. This first attempt to provide a closed form solution to the identification problem of bilinear systems can provide us with a criterion regarding the applicability and usefulness of the existing theory.

The main problem encountered with the state space representation of bilinear system (as far as identification) was the nonlinearity observed with respect to the parameters. Identification of a parametric model takes the form of a nonlinear optimization problem. The known canonical forms of the state space representation do not provide any simplification to this problem. The mathematics necessary to understand the natural ambient of the state space have not been excrutiated enough. It is known for example that for continuous time representation of bilinear systems the state space is contained in a Lie Group and that the matrices A and B_i in $\dot{x}(t) = (A + \sum_{i=1}^{v} n(t) B_i)X(t)$ belong to the Lie Algebra associated with the group but for discrete bilinear representations the differential-geometric point of view has not been emphasized.

For discrete bilinear systems the state at each time can be expressed as a linear combination of the associative algebra generated by A and B (equation (1.1)) and we believe that further understanding of the representation theory of associative algebras as they apply to discrete bilinear systems would yield some of the answers necessary to simplify the parametric identification problem. Perhaps even a better

understanding of the numerical techniques in existence would not make the problem appear so formidable.

Other representations of input-output maps via non-parametric models are also necessary. The limitations of the existing Volterra-Series representation and treated in this work are apparent: bilinearity cannot be tested for if unknown and even if known very careful judgement is necessary to choose the order of the model identified because of the impossibility of checking the "rank condition". Realistic considerations also render impossible the obtainment of models of order higher than 8 or 9 because of the amount of iterations necessary (2^{15} or 2^{17} respectively for orders 8 or 9).

It is hoped that this work will serve as a valuable guideline or departing point for interested researchers in this field.

References

1. A. ISIDORI, Direct construction of minimal bilinear realizations from nonlinear input-output maps, IEEE Trans. Automatic Control, AC 18 (1973).

2. R. W. BROCKETT, On the algebraic structure of bilinear systems, Theory and Applications of Variable Structure Systems (R. R. Mohler and A. Ruberti, eds.), Academic Press, New York (1972), pp. 153-168.

3. H. SAMELSON, Notes on Lie Algebras, Van Nostrand Co. 1969.

4. A. ISIDORI, A. RUBERTI, Realization theory of bilinear systems, Instituto di Automatica, Universita di Roma

5. K. J. ASTROM and P. EYKOFF, System Identification - A Survey, Automatica, Vol. 7, pp. 123-162. Pergamon Press, 1971.

6. H. J. SUSSMANN, " The Bang-Bang Problem for Linear Control Systems in Lie Groups," SIAM J. on Control.

7. V. J. URDJEVIC and H. J. SUSSMANN, 'Control Systems in Lie Groups," J. Differential Equations, 12, No. 2., 1972.

8. J. T. LO and A. S. WILLSKY, "Stochastic Control of Rotational Process with one Degree of Freedom," submitted for publication.

9. J. T. LO and A. S. WILLSKY, "Estimation for Rotational Processes with One Degree of Freedom," Harvard University Technical Report No. 635, July 1972.

10. J. T. LO, 'Signal Detection of Rotational Processes and Frequency Demodulation," presented at the 7th Annual Princeton Conference on Information Sciences and Systems, March 1973.

11. F. S. GRODINS, Control Theory and Biological Systems. New York: Columbia University Press, 1963, pp. 113-117.

12. J. D. HARDY, " Physiology of temperature regulation," Physiol. Rev., Vol. 41, p. 521, 1961.

13. R. R. MOHLER, Bilinear control process with applications to engineering, ecology and medicine, Academic Press, New York (1973).

14. R. W. BROCKETT, " System Theory on Group Manifolds and Coset Spaces," SIAM J. on Control, May 1972.

15. P. ALPER, A consideration of the discrete volterra series. IEEE Trans. Aut. Contr. AC-10, 322-327 (1965).

16. N. WIENER, " Non-Linear Problems in Random Theory," Technology Press of MIT, 1958.

17. J. F. BARRET, "The Use of Volterra Series to Find Region of Stability of a Non-Linear Differential Equation".

18. B. L. HO and R. E. KALMAN, Effective construction of linear state-variable models from input/output functions. Regelungstechnick 14, 545-548 (1966).

19. D. G. LUENBERGER, "Optimization by Vector Space Methods," Wiley 1969.

20. GEORGE N. SARIDIS, "Comparison of Five Popular Identification Algorithms: A Survey," Proc. of the IEEE Conference on Decision and Control, p. 40, 1972.