A Discrete State-Space Model for Linear Image Processing

ROBERT P. ROESSER, MEMBER, IEEE

***Abstract*—The linear time-discrete state-space model is generalized from single-dimensional time to two-dimensional space. The generalization includes extending certain basic known concepts from one to two dimensions. These concepts include the general response formula, state-transition matrix, Cayley–Hamilton theorem, observability, and controllability.**

I. INTRODUCTION

IMAGE processing by nonoptical means has been receiving extensive attention in the last few years. Several books, e.g., [1]–[3] and many papers, e.g., [4]–[6], have been published that have established nonoptical image processing as a viable area of research. A large portion of this research emphasizes the linear processing of images for two main reasons: 1) Many image processing tasks are linear in nature. These tasks include image enhancement, image restoration, picture coding, linear pattern recognition, and TV bandwidth reduction. 2) There are many known linear techniques that may be brought to bear in the treatment of linear image processing, and therefore simplify such treatment. These techniques include transform theory, matrix theory, superposition, etc.

Several ways are commonly used to represent the operations involved in image processing. These include transfer functions, partial difference (recursive) equations, and convolution summations. For example, VanderLugt [7], [11] has presented an extensive development of linear optics based on transfer functions. The transfer functions relate the two-dimensional Fourier transform of an output image to that of the input image. Complex optical systems are easily described by combinations of transfer functions that correspond to individual components of the optical system.

Partial difference equations are used by Habibi [6] to describe a model for estimating images corrupted by noise. The model corresponds to a two-dimensional extension of Kalman filters.

Convolution summations are discussed by Fryer and Richmond [5] in work that involves simplifying a two-dimensional filter to a single-dimensional filter.

The time-discrete state-space model offers great utility in the formulation and analysis of linear systems. Linear systems that are described by transfer functions, difference equations, or convolution summations are easily formulated into a state-space representation. Once so formulated, many known techniques may be applied to systematically analyze the model. Consequently, the state-space model is a general and powerful tool that is used to unify the research and study of time-discrete linear systems.

This paper develops a discrete model for linear image processing that closely parallels the well-known state-space model for time-discrete systems. Because of this parallel many of the concepts that are known for the temporal model may be carried over to the spatial model. This is done by generalizing from a single coordinate in time to two coordinates in space. The spatial model will hopefully have some of the same utility in unifying the study of two-dimensional linear systems as does the temporal model for one-dimensional linear systems.

Temporal systems are inherently nonanticipatory and are often treated as such for the sake of physical realizability in real time; whereas spatial systems do not have causality as an inherent limitation. That is, an image processor may have right to left dependency as well as left to right dependency. Causality is built into the temporal state-space model if an initial state is assumed to be fully specified. In order to establish a close parallel for the spatial model the same built-in causality will be intentionally assumed despite the fact that causality is not necessary for physical realizability in real space. Such an image processor is said to be unilateral. If the constraint of causality is removed, then an image processor is said to be bilateral.

Concepts that are developed in this paper include 1) formulation of the state-space model, 2) the definition of a state-transition matrix, 3) the derivation of a general response formula, 4) a two-dimensional parallel to the Cayley–Hamilton theorem, 5) observability and controlability, and 6) computation of the state-transition matrix. Some of these concepts are based upon an extension of published material on linear iterative circuits coauthored by this writer [8], [9]. A finite field is assumed in the case of iterative circuits, whereas a real field is assumed for image processing. One particular concept, the two-dimensional Cayley–Hamilton theorem, is treated in [9]. An interesting alternative proof to the theorem is the topic of a paper published by Vilfan [10].

Manuscript received February 5, 1974. Paper recommended by A. S. Morse, Chairman of the IEEE S-CS Linear Systems Committee.
The author is with the Department of Electrical Engineering, Wayne State University, Detroit, Mich. 48202.

II. The Model

An image is a generalization of a temporal signal in that it is defined over two spatial dimensions instead of a single temporal dimension. Consequently, two space coordinates i and j take the place of time t. Also two-state sets are introduced to replace the single-state set. The following definitions are made for the model.

i	An integer-valued vertical coordinate.
j	An integer-valued horizontal coordinate.
$\{R\}$	A set of real n_1-vectors which convey information vertically.
$\{S\}$	A set of real n_2-vectors which convey information horizontally.
$\{u\}$	A set of real p-vectors that act as inputs.
$\{y\}$	A set of real m-vectors that act as outputs.

A specific image processor is then defined as a 6-tuple $<\{R\}, \{S\}, \{u\}, \{y\}, f, g>$ where f is the next state function;

$$f:\{R\}\times\{S\}\times\{u\}\rightarrow\{R\}\times\{S\}$$

and g is the output function;

$$g:\{R\}\times\{S\}\times\{u\}\rightarrow\{y\}.$$

Now, since f and g are to be linear functions they may be represented by the following matrix equations:

$$R(i+1,j)=A_1R(i,j)+A_2S(i,j)+B_1u(i,j)$$

$$S(i,j+1)=A_3R(i,j)+A_4S(i,j)+B_2u(i,j)$$

$$y(i,j)=C_1R(i,j)+C_2S(i,j)+Du(i,j), \qquad i,j\geqslant 0.$$

A_1, A_2, A_3, A_4, B_1, B_2, C_1, C_2, D are matrices of appropriate dimensions. Boundary conditions $R(0,j)$ and $S(i,0)$ and also the input $u(i,j)$ are externally specified. In the next section a computational rule is obtained that uniquely determines the states $R(i,j)$ and $S(i,j)$ and also the output $y(i,j)$ (for $i,j\geqslant 0$) from the boundary conditions and inputs. Thus given values for the boundary conditions (such as all zero) the equations produce an output vector image from an input vector image. This formulation is general so that any discrete linear image process may be so represented. Notation is condensed somewhat by introducing the following matrices and vectors:

$$A=\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \quad B=\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \quad C=[C_1\ C_2]$$

$$T(i,j)=\begin{bmatrix} R(i,j) \\ S(i,j) \end{bmatrix} \quad T'(i,j)=\begin{bmatrix} R(i+1,j) \\ S(i,j+1) \end{bmatrix}.$$

Then

$$T'(i,j)=AT(i,j)+Bu(i,j)$$

$$y(i,j)=CT(i,j)+Du(i,j).$$

III. General Response Formula

A state-transition matrix $A^{i,j}$ is defined as follows.

Definition: For

$$A=\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$$

$$A^{i,j}=A^{1,0}A^{i-1,j}+A^{0,1}A^{i,j-1}, \qquad (i,j)>(0,0)$$

$$A^{0,0}=I$$

$$A^{-i,j}=A^{i,-j}=0, \qquad \text{for } j\geqslant 1,\ i\geqslant 1.$$

Examination of this definition bears out that it is an effective recursive definition for integer values of i and j, such that either $i>0$ or $j>0$ or $(i,j)=(0,0)$. It parallels the definition of the time-discrete state-transition matrix $A^t=A^1A^{t-1}$.

It is now to be shown that this state-transition matrix, $A^{i,j}$, may be used in an expression for the response of the model in terms of the inputs and boundary conditions. The term boundary conditions is used here to refer to the states along the edges of the model. Specifically, the set of boundary conditions consist of $R(0,j)$ for $j\geqslant 0$ and $S(i,0)$ for $i\geqslant 0$.

Definition: The following partial ordering is used for integer pairs:

$$(h,k)\leqslant(i,j), \qquad \text{iff } h\leqslant i \text{ and } k\leqslant j$$

$$(h,k)=(1,j), \qquad \text{iff } h=i \text{ and } k=j$$

$$(h,k)<(i,j), \qquad \text{iff } (h,k)\leqslant(i,j) \text{ and } (h,k)\neq(i,j).$$

Lemma: Let the input, $u(i,j)$, for all (i,j) and the boundary conditions, $R(0,j)$ and $S(i,0)$, for $(i,j)\neq(0,0)$ be equal to zero. Then $T(i,j)=A^{i,j}T(0,0)$.

Proof: The proof is accomplished by induction.

First, $T(0,0)=IT(0,0)=A^{0,0}T(0,0)$. This implies the hypothesis is true for $(i,j)=(0,0)$. Now assume the hypothesis is true for all (h,k) such that $(0,0)\leqslant(h,k)<(i,j)$, and show that it is true for (i,j).

$$T(i,j)=\begin{bmatrix} R(i,j) \\ S(i,j) \end{bmatrix}=\begin{bmatrix} A_1R(i-1,j)+A_2S(i-1,j)+B_1\cdot 0 \\ A_3R(i,j-1)+A_4S(i,j-1)+B_2\cdot 0 \end{bmatrix}$$

$$=\begin{bmatrix} A_1 & A_2 \\ 0 & 0 \end{bmatrix}T(i-1,j)+\begin{bmatrix} 0 & 0 \\ A_3 & A_4 \end{bmatrix}T(i,j-1)$$

$$=A^{1,0}A^{i-1,j}T(0,0)+A^{0,1}A^{i,j-1}T(0,0)$$

$$=A^{i,j}T(0,0).$$

This is an effective inductive proof because an enumeration can be found such that all (i,j) are reached but not before all $(h,k)<(i,j)$. Such an enumeration is the diagonal enumeration $(0,0)$, $(0,1)$, $(1,0)$, $(0,2)$, $(1,1)$, $(2,0)$, $\cdots$ Q.E.D.

It is to be noted that since the matrices are not functions of (i,j) the model is spatially invariant. The effect then of $T(h,k)$ on $T(i,j)$ is $A^{i-h,j-k}T(h,k)$. The superposition property of linear systems may be used to obtain a more general expression for $T(i,j)$ in which $u(i,j)$ and the other boundary conditions are not assumed to be zero.

Effect of $u(h,k)$: Assume $u(h,k)$ for some $(h,k)<(i,j)$ is the only nonzero input and all boundary conditions are zero. Then

$$T(h+1,k)=\begin{bmatrix} A_1\cdot 0+A_2\cdot 0+B_1\cdot u(h,k) \\ 0 \end{bmatrix}=\begin{bmatrix} B_1 \\ 0 \end{bmatrix}u(h,k)$$

and

$$T(h,k+1)=\begin{bmatrix} 0 \\ A_3\cdot 0+A_4\cdot 0+B_2\cdot u(h,k) \end{bmatrix}=\begin{bmatrix} 0 \\ B_2 \end{bmatrix}u(h,k).$$

Then

$$T(i,j)=A^{i-(h+1),j-k}T(h+1,k)+A^{i-h,j-(k+1)}T(h,k+1)$$

$$=\left(A^{i-h-1,j-k}\begin{bmatrix} B_1 \\ 0 \end{bmatrix}+A^{i-h,j-k-1}\begin{bmatrix} 0 \\ B_2 \end{bmatrix}\right)u(h,k).$$

Effect of $R(0,k)$: Assume that $R(0,k)$ is the only nonzero boundary condition and that all inputs are zero.

$$T(0,k)=\begin{bmatrix} R(0,k) \\ 0 \end{bmatrix}.$$

Then

$$T(i,j)=A^{i,j-k}T(0,k)$$

$$=A^{i,j-k}\begin{bmatrix} R(0,k) \\ 0 \end{bmatrix}.$$

Effect of $S(h,0)$: Similarly to $R(0,k)$ the effect of $S(h,0)$ is

$$T(i,j)=A^{i-h,j}\begin{bmatrix} 0 \\ S(h,0) \end{bmatrix}.$$

We thus have the following theorem.

Theorem: For all $(i,j)\geqslant 0$.

$$T(i,j)=\sum_{k=0}^{j}A^{i,j-k}\begin{bmatrix} R(0,k) \\ 0 \end{bmatrix}+\sum_{h=0}^{i}A^{i-h,j}\begin{bmatrix} 0 \\ S(h,0) \end{bmatrix}$$

$$+\sum_{(0,0)<(h,k)<(i,j)}\sum$$

$$\cdot\left(A^{i-h-1,j-k}\begin{bmatrix} B_1 \\ 0 \end{bmatrix}+A^{i-h,j-k-1}\begin{bmatrix} 0 \\ B_2 \end{bmatrix}\right)u(h,k).$$

Proof: By superposition of the effects of all inputs and boundary conditions. Q.E.D.

An expression for $y(i,j)$, called the *general response formula*, may now be written

$$y(i,j)=[C_1\ C_2]\left(\sum_{k=0}^{j}A^{i,j-k}\begin{bmatrix} R(0,k) \\ 0 \end{bmatrix}\right.$$

$$+\sum_{h=0}^{i}A^{i-h,j}\begin{bmatrix} 0 \\ S(h,0) \end{bmatrix}$$

$$+\sum_{(0,0)<(h,k)<(i,j)}\sum\left(A^{i-h-1,j-k}\begin{bmatrix} B_1 \\ 0 \end{bmatrix}\right.$$

$$\left.\left.+A^{i-h,j-k-1}\begin{bmatrix} 0 \\ B_2 \end{bmatrix}\right)u(h,k)\right)+Du(i,j).$$

IV. Properties of $A^{i,j}$

Some properties of $A^{i,j}$ are

1)

$$A=\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}=\begin{bmatrix} A_1 & A_2 \\ 0 & 0 \end{bmatrix}+\begin{bmatrix} 0 & 0 \\ A_3 & A_4 \end{bmatrix}.$$

Thus $A=A^{1,0}+A^{0,1}$.

2)

$$A^{i,0}=A^{1,0}A^{i-1,0}+A^{0,1}A^{i,-1}=A^{1,0}A^{i-1,0}.$$

Thus $A^{i,0}=(A^{1,0})^i$. Similarly $A^{0,j}=(A^{0,1})^j$.

3)

$$I=\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

where I is the identity matrix with appropriate dimensions. Thus

$$I^{1,0}=\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \text{ and } I^{0,1}=\begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}.$$

4)

$$I^{1,0}A=\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}=\begin{bmatrix} A_1 & A_2 \\ 0 & 0 \end{bmatrix}=A^{1,0}.$$

Briefly $I^{1,0}A=I^{1,0}A^{1,0}=A^{1,0}$. Similarly $I^{0,1}A=I^{0,1}A^{0,1}=A^{0,1}$.

5)

$$I^{0,1}A^{1,0}=\begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} A_1 & A_2 \\ 0 & 0 \end{bmatrix}=0.$$

Briefly $I^{0,1}A^{1,0}=0$. Similarly $I^{1,0}A^{0,1}=0$.

V. Characteristic Function of a Matrix

If the primary inputs and outputs are neglected in the model equations, a representation arises for the state behavior of the circuit, having the form

$$R(i+1,j)=A_1R(i,j)+A_2S(i,j)$$

$$S(i,j+1)=A_3R(i,j)+A_4S(i,j).$$

These equations are useful in the development of a form for a two-dimensional characteristic matrix of A. Operators are first introduced that advance a particular coordinate of their operand.

Definition: Let E be an operator that has the effect of advancing the vertical coordinate or the first subscript of the function upon which it is operating. Likewise, let F be an operator that has the effect of advancing the horizontal coordinate or second subscript of the function upon which it is operating.

The effect of these operators on the state vectors is

$$R(i+1,j)=ER(i,j)$$

$$S(i,j+1)=FS(i,j).$$

The state equations can be rewritten using these advance operators.

$$(EI-A_1)R(i,j)-A_2S(i,j)=0$$

$$-A_3R(i,j)+(FI-A_4)S(i,j)=0.$$

These equations are equivalently represented in the overall matrix form.

$$\begin{bmatrix} (EI-A_1) & -A_2 \\ -A_3 & (FI-A_4) \end{bmatrix}T(i,j)=0.$$

The above equation represents a system of homogeneous equations in the elements of $T(i,j)$. If the system is to have a nontrivial solution for $T(i,j)$, then the transformation represented by the matrix must be singular.

The above matrix is said to be the two-dimensional characteristic matrix of the partitioned matrix A where

$$A=\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}.$$

The characteristic matrix of A is denoted by $cm(A)$ and may be represented as

$$cm(A)=EI^{1,0}+FI^{0,1}-A.$$

Now since $cm(A)$ must be singular, its determinate must be equal to zero.

$$|cm(A)|=0.$$

If E and F are placed in the above by general indeterminates x and y, respectively, the result is an expression called the two-dimensional characteristic equation for A. The determinate of $cm(A)$ with x and y replacing E and F is called the two-dimensional characteristic function of the matrix X, and is denoted by $f(x,y)$.

$$|cm(A)|=f(x,y)=0.$$

$f(x,y)$ will be a monic multinomial in x and y with degree n_1 in x and degree n_2 in y; where n_1 is the dimension of R and n_2 is the dimension of S. $f(x,y)$ has the form

$$f(x,y)=\sum_{(0,0)\leqslant(i,j)\leqslant(n_1,n_2)}\sum a_{i,j}x^iy^j, \qquad \text{where } a_{n_1,n_2}=1.$$

Comparing these concepts to the one-dimensional case, it is observed that they are correspondingly analogous to the one-dimensional characteristic matrix, equation and function of a matrix. $xI-A$ is the one-dimensional characteristic matrix of A and $f(x)=|xI-A|=0$ is the one-dimensional characteristic equation.

The Cayley–Hamilton theorem in the one-dimensional case states that a matrix A satisfies its own characteristic equation, i.e., $f(A)=0$. The following theorem extends this to the two-dimensional case.

Definition: $E^iF^jA=F^jE^iA=A^{i,j}$ for any 2×2 partition of A.

Two-Dimensional Cayley–Hamilton Theorem: Every partitioned matrix

$$A=\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$$

satisfies its own characteristic equation. That is $f(E,F)A=0$.

Proof: Let $B=(xI^{1,0}+yI^{0,1}-A)$, so that $f(x,y)=\det B$. Cramer's rule for computing the inverse of a matrix, in this case matrix B, states

$$\operatorname{adj} B\cdot(xI^{1,0}+yI^{0,1}-A)=\det B\cdot I=f(x,y)I \tag{1}$$

where adj B is the transpose of the cofactor matrix of B. The elements of adj B transpose are computed by taking the determinate of the matrix formed by deleting the row and column containing the corresponding element in B. The elements of adj B will consequently be multinomials in x and y having degrees in x and y not greater than n_1 and n_2, respectively, where n_1 is the rank of $I^{1,0}$ and n_2 is the rank of $I^{0,1}$. Therefore, adj B may be written in the form of a matrix multinomial.

$$\operatorname{adj} B=\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} B_{i,j}x^iy^j. \tag{2}$$

Represent the characteristic multinomial $f(x,y)$ as

$$f(x,y)=\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} b_{i,j}x^iy^j, \qquad \text{where } b_{n_1,n_2}=1. \tag{3}$$

Substituting (2) and (3) into (1), we have,

$$\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} B_{i,j}x^iy^j(xI^{1,0}+yI^{0,1}-A)=\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} b_{i,j}x^iy^jI.$$

Expand the left side and adjust i and j.

$$\sum_{i=1}^{n_1+1}\sum_{j=0}^{n_2} B_{i-1,j}I^{1,0}x^iy^j+\sum_{i=0}^{n_1}\sum_{j=1}^{n_2+1} B_{i,j-1}I^{0,1}x^iy^j - \sum_{i=0}^{n_1}\sum_{j=0}^{n_2} B_{i,j}Ax^iy^j=\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} b_{i,j}x^iy^jI.$$

The coefficients of each term x^iy^j on both sides of the equation must be equal. Equating these coefficients yields the following:

$$B_{n_1,j}I^{1,0}=0, \qquad \text{for } 0\leqslant j\leqslant n_2 \tag{4}$$

$$B_{i,n_2}I^{0,1}=0, \qquad \text{for } 0\leqslant i\leqslant n_1 \tag{5}$$

$$B_{i-1,j}I^{1,0}+B_{i,j-1}I^{0,1}-B_{i,j}A=b_{i,j}I, \qquad \text{for } i\neq 0\ j\neq 0 \tag{6}$$

$$B_{0,j-1}I^{0,1}-B_{0,j}A=b_{0,j}I, \qquad \text{for } j\neq 0 \tag{7}$$

$$B_{i-1,j}I^{1,0}-B_{i,0}A=b_{i,0}I, \qquad \text{for } i\neq 0 \tag{8}$$

$$-B_{0,0}A=b_{0,0}I. \tag{9}$$

Multiply each of the equations (6)–(8) on the right by $A^{i,j}$ and sum equations (6)–(9) over all i,j.

$$\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}(B_{i-1,j}I^{1,0}+B_{i,j-1}I^{0,1}-B_{i,j}A)A^{i,j} + \sum_{j=1}^{n_2}(B_{0,j-1}I^{0,1}-B_{0,j}A)A^{0,j} + \sum_{i=1}^{n_1}(B_{i-1,0}I^{1,0}-B_{i,0}A)A^{i,0}-B_{0,0}A = \sum_{i=0}^{n_1}\sum_{j=0}^{n_2} b_{i,j}IA^{i,j}.$$

Collect coefficients of $B_{i,j}$ on the left side of the equation.

$$B_{i,j}(I^{1,0}A^{i+1,j}+I^{0,1}A^{i,j+1}-AA^{i,j}), \qquad \text{for } 1\leqslant i\leqslant n_1\ 1\leqslant j\leqslant n_2 \tag{10}$$

$$B_{n_1,j}(I^{0,1}A^{n_1,j+1}-AA^{n_1,j}), \qquad \text{for } j\neq 0\ j\neq n_2 \tag{11}$$

$$B_{i,n_2}(I^{1,0}A^{i+1,n_2}-AA^{i,n_2}), \qquad \text{for } i\neq 0\ i\neq n_1 \tag{12}$$

$$B_{0,j}(I^{1,0}A^{1,j}+I^{0,1}A^{0,j+1}-AA^{0,j}), \qquad \text{for } j\neq 0\ j\neq n_2 \tag{13}$$

$$B_{i,0}(I^{0,1}A^{i,1}+I^{1,0}A^{i+1,0}-AA^{i,0}), \qquad \text{for } i\neq 0\ i\neq n_1 \tag{14}$$

$$B_{n_1,n_2}(AA^{n_1,n_2}) \tag{15}$$

$$B_{0,0}(I^{0,1}A^{0,1}+I^{1,0}A^{1,0}-A). \tag{16}$$

Expressions (10)–(16) exhaust all i,j combinations. We will now evaluate each in turn.

$$I^{1,0}A^{i+1,j}=I^{1,0}A^{1,0}A^{i,j}+I^{1,0}A^{0,1}A^{i+1,j-1}$$
$$I^{0,1}A^{i,j+1}=I^{0,1}A^{1,0}A^{i-1,j+1}+I^{0,1}A^{0,1}A^{i,j} \tag{10}$$

but $I^{1,0}A^{0,1}=0$ and $I^{0,1}A^{1,0}=0$.

Thus,

$$I^{1,0}A^{i+1,j}+I^{0,1}A^{i,j+1}-AA^{i,j}=A^{1,0}A^{i,j}+A^{0,1}A^{i,j}-AA^{i,j} =AA^{i,j}-AA^{i,j}=\mathbf{0}.$$

$$I^{0,1}A^{n_1,j+1}=I^{0,1}A^{1,0}A^{n_1-1,j+1}+I^{0,1}A^{0,1}A^{n_1,j} =A^{0,1}A^{n_1,j}. \tag{11}$$

Thus,

$$B_{n_1,j}(I^{0,1}A^{n_1,j+1}-AA^{n_1,j}) =B_{n_1,j}(A^{0,1}A^{n_1,j}-A^{0,1}A^{n_1,j}-A^{1,0}A^{n_1,j}) =-B_{n_1,j}A^{1,0}A^{n_1,j}=-B_{n_1,j}I^{1,0}A^{1,0}A^{n_1,j}=\mathbf{0}$$

since $B_{n_1,j}I^{1,0}=0$ from (4).

$$\text{Similarly}=\mathbf{0} \text{ as in (11)}. \tag{12}$$

$$AA^{0,j}=A^{1,0}A^{0,j}+A^{0,1}A^{0,j}=A^{1,0}A^{0,j}+A^{0,j+1}$$
$$I^{1,0}A^{1,j}=I^{1,0}A^{1,0}A^{0,j}+I^{1,0}A^{0,1}A^{1,j-1}=A^{1,0}A^{0,j}. \tag{13}$$

Thus,

$$I^{1,0}A^{1,j}+I^{0,1}A^{0,j+1}-AA^{0,j} =A^{1,0}A^{0,j}+A^{0,j+1}-A^{1,0}A^{0,j}-A^{0,j+1}=\mathbf{0}.$$

$$\text{Similarly}=\mathbf{0} \text{ as in (13)}. \tag{14}$$

$$B_{n1,n2}=\mathbf{0}, \text{ since adj } B \text{ can have no term in } x^{n_1}y^{n_2}. \tag{15}$$

$$A=A^{1,0}+A^{0,1}, \text{ thus, } I^{0,1}A^{0,1}+I^{1,0}A^{1,0}-A=\mathbf{0}. \tag{16}$$

Briefly, the left side of the equation being considered is zero, giving

$$0=\sum_{i=0}^{n_1}\sum_{j=0}^{n_2} b_{i,j}A^{i,j}=f(E,F)A$$

which completes the proof.

VI. Observability and Controllability

The notions of observability and controllability for time-discrete systems carry over to parallel notions for discrete-space image processors.

Definition: A state T_0 is observable iff whenever it appears as the initial state and all other boundary conditions are zero, there exists a pattern of inputs and a pair $(i,j) \geqslant (0,0)$ such that $y(i,j)$ is not the same as when the initial state is zero, and the same pattern of inputs is applied.

Definition: A state T_0 is controllable iff when all boundary conditions are zero there exists some pair $(i,j) \geqslant 0$ and some input pattern such that $T(i,j) = T_0$.

An image processing model is said to be observable (controllable) iff all states are observable (controllable).

It is often desirable to reduce a model to an equivalent model that is observable and controllable. Here equivalence between models will be taken to mean no pattern of inputs exist so that the output from one model is different at some pair $(i,j) \geqslant (0,0)$ than the output from the other model at (i,j) when the boundary conditions of both models are zero.

To test for observability the output $y(i,j)$ for each possible initial state is compared with the output $y_0(i,j)$ with zero initial state, for all $(i,j) \geqslant (0,0)$ and all input patterns. If $y(i,j) = y_0(i,j)$ for all $(i,j) \geqslant 0$ and all input conditions, then the model is not observable. Using the general response formula this condition reduces to

$$CA^{i,j}T(0,0) = 0, \qquad \text{for all } (i,j) \geqslant 0.$$

The two-dimensional Cayley–Hamilton theorem implies that any two-tuple power of A is linearly dependent upon those $A^{i,j}$ for which $(0,0) \leqslant (i,j) < (n_1,n_2)$ so that the condition for nonobservability can be limited to those (i,j) such that $(0,0) \leqslant (i,j) < (n_1,n_2)$. The condition may then be put into matrix form $KT = 0$, where K is the diagnostic matrix defined as follows:

$$K = \begin{bmatrix} C \\ CA^{0,1} \\ CA^{0,2} \\ \vdots \\ CA^{0,n_2} \\ CA^{1,0} \\ CA^{1,1} \\ \vdots \\ CA^{n_1,0} \\ CA^{n_1,1} \\ \vdots \\ CA^{n_1,n_2-1} \end{bmatrix}.$$

If $KT = 0$ then the model is not observable, but may be reduced to an observable model. Let K_1 be a matrix consisting of the first n_1 columns of K and let K_2 be a matrix consisting of the last n_2 columns of K. The conditions $KT = 0$ may then be split into the two conditions $K_1R = 0$ and $K_2S = 0$. A reduced model may then be formed by using the equivalence classes of $\{R\}$ modulo the null space of K_1 as the new vertical state set, and the equivalence classes of $\{S\}$ modulo the null space of K_2 as the new horizontal state set. As a result only the new zero vector will satisfy $K_1R = 0$ and $K_2S = 0$, implying that the new model is observable. It remains to find the characterizing matrices for the reduced model. First vector representation for the equivalence classes of vertical and horizontal states is found, then the original characterizing matrices are modified so that the behavior of the model remains the same. K_1 and K_2 may be reduced so that their new dimension agrees with their rank. This is done by forming matrices G_1 and G_2 from the first complete set of linearly independent rows of K_1 and K_2, respectively. $G_{(\,)}$ and $K_{(\,)}$ have the same null space. Furthermore each equivalence class of $\{R\}$ and $\{S\}$ will correspond to a single vector G_1R and G_2S, respectively. The equivalence classes of $\{T\}$ may then be represented as

$$\overline{T} = GT, \qquad \text{where } G = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}.$$

Now let

$$H = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix}$$

be a right inverse to G, i.e., $GH = I$. The reduced model will then have the following characterizing matrices:

$$\overline{A} = GAH \qquad \overline{B} = GB \qquad \overline{C} = CH \qquad \overline{D} = D.$$

To test for controllability the state $T(i,j)$ is examined with all boundary conditions equal to zero for all input patterns. If $T(i,j)$ doesn't equal T_0 for some T_0, for any $(i,j) \geqslant (0,0)$ and any input pattern, then the model is uncontrollable. Using the general formula for $T(i,j)$ this condition becomes

$$\sum_{(0,0) \leqslant (h,k) < (i,j)} \sum \left(A^{i-h-1,j-k} \begin{bmatrix} B_1 \\ 0 \end{bmatrix} + A^{i-h,j-k-1} \begin{bmatrix} 0 \\ B_2 \end{bmatrix} \right) u(k,k) \neq T_0$$

for all (i,j) and all $u(h,k)$. As for observability, the two-dimension Cayley–Hamilton theorem allows us to limit the condition for noncontrollability to those (i,j) such that $(0,0) \leqslant (i,j) < (n_1,n_2)$. The condition may then be put into matrix form.

$QU \neq T_0$ for all U and some T_0 where U and Q are defined as follows:

$$U = \begin{bmatrix} u(0,0) \\ u(0,1) \\ \vdots \\ u(0,n_2) \\ u(1,0) \\ \vdots \\ u(n_1,n_2-1) \end{bmatrix}$$

$$Q = [M(0,0), M(0,1), \ldots, M(0,n_2), M(1,0), \ldots, M(n_1,n_2-1)]$$

$$M(i,j) = A^{i-1,j}\begin{bmatrix} B_1 \\ 0 \end{bmatrix} + A^{i,j-1}\begin{bmatrix} 0 \\ B_2 \end{bmatrix}.$$

The maximum rank of Q is equal to n_1+n_2 (the number of rows), which is equal to the dimension of $\{T\}$. If Q has n_1+n_2 as its rank, then its range space must equal $\{T\}$ so that there would be no T_0 such that $QU \neq T_0$ for all U. The model would then be controllable. If however the rank of Q is less than n_1+n_2, then the model may be reduced to a controllable model. Let Q_1 be formed from the first n_1 rows of Q and Q_2 be formed from the last n_2 rows of Q.

Q_1 and Q_2 may be reduced so that their column dimension agrees with their rank. This is done by forming matrices G_1 and G_2 from the first complete set of linearly independent columns of Q_1 and Q_2, respectively.

The controllable states T will be those formed from the direct sum of the vectors in the range spaces of G_1 and G_2. A state set for the reduced model may now be specified as a set of vectors

$$\left\{ \overline{T} = \begin{bmatrix} \overline{R} \\ \overline{S} \end{bmatrix} \right\}$$

which are mapped into the controllable states under the direct sum of G_1 and G_2. That is $\{\overline{T}\}$ is the domain of the linear mapping

$$T = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}\begin{bmatrix} \overline{R} \\ \overline{S} \end{bmatrix} = G\overline{T}.$$

Let

$$H = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix}$$

be a left inverse of G, i.e., $HG = I$. Then,

$$\overline{T} = HG\overline{T} = HT.$$

Note that the states in the range space of G are controllable and that the dimension of $\{\overline{T}\}$ is equal to the rank of G. This implies that each $\overline{T}$ is mapped by G uniquely into a controllable T. Therefore, each $\overline{T}$ of the reduced model will be controllable. The characterizing matrices of the reduced model can now be found by noting the effect of the mappings G and H on the original characterizing matrices

$$\overline{A} = HAG \qquad \overline{B} = HB$$

$$\overline{C} = CG \qquad \overline{D} = D.$$

VIII. Computation of the Transition Matrix

Computing the transition matrix $A^{i,j}$ using the recursive definition becomes quite tedious as i and j become large. It is therefore desirable to extend the techniques that are known for computing a single power of a matrix, such as the Cayley-Hamilton technique or Sylvester's theorem, to parallel techniques for $A^{i,j}$. The Cayley–Hamilton technique will be treated in what follows. Other methods may be extended in a similar fashion.

From the two-dimensional Cayley–Hamilton theorem we have

$$f(E,F)A = 0.$$

Letting $a_{h,k}$ be the coefficient of E^hF^k in $f(E,F)$ the last equation becomes

$$\sum_{(0,0) \leqslant (h,k) \leqslant (n_1,n_2)} \sum a_{h,k} A^{h,k} = 0.$$

If both sides of this equation are operated upon by E^iF^j it becomes

$$\sum_{(0,0) \leqslant (h,k) \leqslant (n_1,n_2)} \sum a_{h,k} A^{h+i,k+j} = 0.$$

Consider this equation for the set of pairs $\{(i,j) | (-n_1 \leqslant i \leqslant 0 \text{ and } j=1) \text{ or } (-n_2 \leqslant j \leqslant 0 \text{ and } i=1)\}$. These pairs have the property that whenever one exponent of $A^{h+i,k+j}$ is negative the other will be positive for $(0,0) \leqslant (h,k) \leqslant (n_1,n_2)$. From the definition of $A^{i,j}$ each such $A^{h+i,k+j}$ will be equal to zero. The following set of equations is produced, each corresponding to a pair (i,j) in the set above:

$$\sum \sum_{(i,0)<(h,k)<(n_1,n_2)} a_{h,k} A^{h-i,k+1} = 0, \qquad 0 \leq i \leq n_1$$

and

$$\sum \sum_{(0,j)<(h,k)<(n_1,n_2)} a_{h,k} A^{h+1,k-j}, \qquad 0 \leq j \leq n_2.$$

Each of these equations can be written in terms of a function of E and F obtained by modifying the two-dimensional characteristic function $f(E,F)$. The modification in the first case consists of multiplying $f(E,F)$ by $E^{-i}F$ and then deleting all terms involving negative exponents of E. The modification for the second case consists of multiplying $f(E,F)$ by EF^{-j} and then deleting all terms involving negative exponents of F. Let

$$f_{i,j}(E,F) = E^{-i}F^{-j}f(E,F)$$

(with deletion of terms having negative exponents). Matrix A then must satisfy the following set of equations:

$$f_{i,-1}(E,F)A = 0, \qquad 0 \leq i \leq n_1 - 1$$

$$f_{-1,j}(E,F)A = 0, \qquad 0 \leq j \leq n_2 - 1.$$

That is, not only must A satisfy the two-dimension characteristic equation, but it must also satisfy the above set of $n_1 + n_2$ equations as well. It will be shown that these equations may be used to reduce $A^{i,j}$ to a linear combination of those $A^{h,k}$ where $(0,0) \leq (h,k) < (n_1,n_2)$. This leads to the definition of two-dimensional eigenvalues.

Definition: The two-dimensional eigenvalues of a partitioned matrix

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$$

are the pairs (x,y) that simultaneously solve the following set of equations:

$$f_{i,-1}(x,y) = 0, \qquad 0 \leq i \leq n_1 - 1$$

$$f_{-1,j}(x,y) = 0, \qquad 0 \leq j \leq n_2 - 1$$

$$f(x,y) = 0.$$

The equations in the above definition may be used to reduce a multinomial in x,y to one of degree $(h,k) < (n_1,n_2)$. The next theorem establishes this result.

Theorem: Two-dimensional division algorithm. Any multinomial $g(x,y)$ of degree $\geq (0,0)$ may be expressed as follows:

$$g(x,y) = \sum_{j=0}^{n_2} p_j(x) f_{-1,j}(x,y) + \sum_{i=0}^{n_1} q_i(y) f_{i,-1}(x,y) + m(x,y)f(x,y) + r(x,y)$$

where the degree (h,k) of $r(x,y)$ is less than (n_1,n_2).

Proof: Each term in $g(x,y)$ of degree $(i,j) \geq (n_1,n_2)$, having the coefficient $b_{i,j}$, may be reduced to a sum of terms of degree less than (k,j) by subtracting $b_{i,j}x^{i-n_1}y^{j-n_2}f(x,y)$ from it. This is repeated until there are no terms of degree greater than (n_1,n_2). The remainder will have the form

$$g(x,y) - m(x,y)f(x,y).$$

Each term of this remainder of degree $(i,n_2-1) \geq (n_1,n_2-1)$ having the coefficient $c_{i,j}$, may then be reduced to a sum of terms of degree less than (i,n_2-1) by subtracting $c_{i,j}f_{-1,n_2-1}(x,y)$ from it. This is repeated until there are no terms $(h,n_2-1) \geq (n_1,n_2-1)$. A similar process is used for the other functions $f_{-1,j}$ and $f_{i,-1}$. The result will be a remainder of the form

$$r(x,y) = g(x,y) - \sum_{j=0}^{n_2} p_j(x) f_{-1,j}(x,y) + \sum_{i=0}^{n_1} q_i(y) f_{i,-1}(x,y) + m(x,y)f(x,y)$$

where the degree of $r(x,y)$ is less than (n_1,n_2). Q.E.D.

If the equation $f_{i,1}(E,F)\,A = 0$ is operated upon by F^j no additional terms are generated so that the result will be a valid equation. Likewise, $E^i f_{1,j}(E,F)A = 0$ is a valid equation. Consequently, the two-dimensional division algorithm may be applied to matrix multinomials and in particular to $A^{i,j}$. Thus,

$$A^{i,j} = \sum_{j=0}^{n_2} p_j(E) f_{-1,j}(E,F)A + \sum_{i=0}^{n_1} q_i(F) f_{i,-1}(E,F)A + M(E,F)f(E,F)A + r(E,F)A$$

where the degree of $r(E,F)A$ is less than (n_1,n_2). The first three terms are zero because A satisfies the previously-mentioned equations. Then $A^{i,j} = r(E,F)A$. Likewise, if (x,y) is an eigenvalue of A, then from the division algorithm $g(x,y) = r(x,y)$.

Since the operations are the same for obtaining $r(x,y)$ and $r(E,F)A$ they will have the same coefficients. $A^{i,j}$ may then be computed by determining $r(x,y)$ from x^iy^j. $r(x,y)$ has $(n_1+1)(n_2+1)-1$ coefficients. These are determined by solving the set of simultaneous equations resulting from $x^iy^j = r(x,y)$, using different eigenvalues (x,y). The same number of eigenvalues as coefficients are necessary.

The determination of two-dimensional eigenvalues of a matrix A is not, in general, obvious. However, if the characteristic function $f(x,y)$ is factorable into linear factors then the two-dimensional eigenvalues are easily identified.

Theorem: Suppose the two-dimensional characteristic function of a matrix A is factorable into linear factors as follows:

$$f(x,y)=(x-a_1)(x-a_2)\cdots(x-a_{n_1})(y-b_1)(y-b_2)\cdots(y-b_{n_2}).$$

Let $a_0=b_0=0$. The set of two-dimensional eigenvalues is

$$\{(a_i,b_j)|(0,0)<(i,j)\leqslant(n_1,n_2)\}.$$

Proof: From the hypothesis $f(x,y)$ may be expressed as the product $p(x)q(y)$. Consequently, $f_{i,-1}(x,y)$ may be expressed as $p_i(x)yq(y)$, where $p_i(x)=x^{-i}p(x)$ (with negative powers of x deleted). Thus, $f_{i,-1}(x,y)$ contains each $(y-b_j)$ as a factor, which implies that $f_{i,-1}(x,b_j)=0$. Likewise $f_{-1,j}(a_i,y)=0$. Thus, for $(x,y)=(a_i,b_j)$ all functions, including f, will be zero. Q.E.D.

If the factors of $f(x,y)$ in the previous theorem are distinct then there will be a total of $(n_1+1)(n_2+1)-1$ eigenvalues, which is the same as the number of coefficients in $r(x,y)$. There is, therefore, just enough equations $a_h^i b_k^j=r(a_h,b_k)$ to solve for the coefficients of $r(x,y)$.

Example: Let

$$A=\left[\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4\end{array}\right]=\left[\begin{array}{c|c} a & b \\ \hline 0 & d\end{array}\right];$$

then

$$cm(A)=\begin{bmatrix} x-a & -b \\ 0 & y-d\end{bmatrix}$$

$$f(x,y)=|cm(A)|=(x-a)(y-d).$$

The two-dimensional eigenvalues are $(0,d)$, $(a,0)$, and (a,d).

$$A^{i,j}=r(E,F)A \quad \text{where deg } r(E,F)<(1,1)$$
$$=a_{0,0}I+a_{1,0}A^{1,0}+a_{0,1}A^{0,1}.$$

To find the coefficients substitute the eigenvalues into

$$x^iy^j=r(x,y)=a_{1,1}xy+a_{1,0}x+a_{0,1}y$$

$$0^id^j=a_{0,0}+a_{1,0}0+a_{0,1}d$$

$$a^i0^j=a_{0,0}+a_{1,0}a+a_{0,1}0$$

$$a^id^j=a_{0,0}+a_{1,0}a+a_{0,1}d.$$

The solution to these equations is

$$a_{0,0}=-a^id^j \quad a_{1,0}=a^{i-1}d^j \quad a_{0,1}=a^id^{j-1}$$

Note that

$$a^{1,0}=\begin{bmatrix} a & b \\ 0 & 0\end{bmatrix}$$

$$a^{0,1}=\begin{bmatrix} 0 & 0 \\ 0 & d\end{bmatrix}.$$

Then

$$a^{i,j}=-a^id^j\begin{bmatrix} 1 & 0 \\ 0 & 1\end{bmatrix}+a^{i-1}d^j\begin{bmatrix} a & b \\ 0 & 0\end{bmatrix}$$

$$+a^id^{j-1}\begin{bmatrix} 0 & 0 \\ 0 & d\end{bmatrix}$$

$$=\begin{bmatrix} 0 & a^{i-1}d^jb \\ 0 & 0\end{bmatrix}.$$

To check:

$$A^{1,1}=A^{1,0}A^{0,1}+A^{0,1}A^{1,0}$$

$$=\begin{bmatrix} a & b \\ 0 & 0\end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & d\end{bmatrix}+\begin{bmatrix} 0 & 0 \\ 0 & d\end{bmatrix}\begin{bmatrix} a & b \\ 0 & 0\end{bmatrix}$$

$$=\begin{bmatrix} 0 & bd \\ 0 & 0\end{bmatrix}+\begin{bmatrix} 0 & 0 \\ 0 & 0\end{bmatrix}=\begin{bmatrix} 0 & bd \\ 0 & 0\end{bmatrix}.$$

This checks with the previous result.

VIII. Conclusion

This paper is an attempt to establish a parallel of the linear discrete-time state-space model for linear discrete-space image processing. However, it can only be assumed to be an initial attempt. Only the more basic and well-known concepts have been extended. Thus, there is much room for future research to be done along this line. Specifically this research should include:

1) Generalization to bilateral models.
2) Methods for programming a spatial transfer function into a state-space model.
3) Discovery of a general method for factoring multinomials and a method for finding two-dimensional eigenvalues.
4) Finding methods for obtaining canonical forms.

5) Establishing criteria for stability.
6) Application of estimation theory.

Finally, of a more general nature, the techniques and concepts of optimal control could be extended to the spatial model.

References

[1] A. Rosenfeld, *Picture Processing by Computer*. New York: Academic, 1969.
[2] H. C. Andrews, *Computer Techniques in Image Processing*. New York: Academic, 1970.
[3] B. S. Lipkin and A. Rosenfeld, *Picture Processing and Psychopictorics*. New York: Academic, 1970.
[4] T. S. Huang, W. F. Schreiber, and O. J. Tretiak, "Image processing," *Proc. IEEE*, vol. 59, pp. 1586–1609, Nov. 1971.
[5] W.D. Fryer and G.E. Richmond, "Two-dimensional spatial filtering and computers," in *Proc. Nat. Electron. Conf.*, Oct. 19, 1962.
[6] A. Habibi, "Two-dimensional Bayesian estimate of images," *Proc. IEEE (Special Issue on Digital Picture Processing)*, vol. 60, pp. 878–883, July 1972.
[7] A. Vander Lugt, "Operational notation for the analysis and synthesis of optical data-processing systems," *Proc. IEEE*, vol. 54, pp. 1055–1063, Aug. 1966.
[8] D. D. Givone and R. P. Roesser, "Multidimensional linear iterative circuits—general properties," *IEEE Trans. Comput.*, vol. C-21, pp. 1067–1073, Oct. 1972.
[9] ——, "Minimization of multidimensional linear iterative circuits," *IEEE Trans. Comput.*, vol. C-22, pp. 673–678, July 1973.
[10] B. Vilfan, "Another proof of the two-dimensional Cayley–Hamilton theorem," *IEEE Trans. Comput.* (Corresp.), vol. C-22, p. 1140, Dec. 1973.
[11] F. T. S. Yu, *Introduction to Diffraction, Information Processing, and Holography*. Cambridge, Mass. : M.I.T. Press, 1973.

Robert P. Roesser (S'67—M'69) was born in Bradford, Pa., on August 1, 1942. He received the B.S., MS., and Ph.D. degrees all in electrical engineering from the State University of New York at Buffalo, Buffalo, N.Y., in 1964, 1967, and 1970, respectively.

While at Buffalo he held a Graduate Assistantship, teaching various laboratory courses in the Department of Electrical Engineering. He was employed in 1970 as an Electrical Engineer in the Electronics Division, Houdaille Industries, where he was involved in the design of numerical and computer controls. In 1971 he joined the Department of Electrical Engineering, Wayne State University, Detroit, Mich., as an Assistant Professor where he is teaching courses and is engaged in active research in digital systems.

Estimation for Rotational Processes with One Degree of Freedom—Part I: Introduction and Continuous-Time Processes

JAMES TING-HO LO, MEMBER, IEEE, AND ALAN S. WILLSKY, MEMBER, IEEE

***Abstract*—A class of bilinear estimation problems involving single-degree-of-freedom rotation is formulated and resolved. Continuous-time problems are considered here, and discrete-time analogs will be studied in a second paper. Error criteria, probability densities, and optimal estimates on the circle are studied. An effective synthesis procedure for continuous-time estimation is provided, and a generalization to estimation on arbitrary Abelian Lie groups is included. Applications of these results to a number of practical problems including frequency demodulation will be considered in a third paper.**

Manuscript received July 30, 1973. Paper recommended by H. W. Sorenson, Chairman of the IEEE S-CS Estimation and Identification Committee. The work of A. S. Willsky was performed in part while the author was a Fannie and John Hertz Foundation Fellow in the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Mass., and in part at the Decision and Control Sciences Group, M. I. T. Electronic Systems Laboratory, Cambridge, Mass., under partial support from the AFOSR under Grant 72-2273. The work of J. T.-H. Lo was supported in part by the U. S. Office of Naval Research under the Joint Services Electronics Program by Contract N00014-67-A-0298-0006 while the author was with the Division of Engineering and Applied Physics, Harvard University, Cambridge, Mass.

J. T.-H. Lo is with the Division of Mathematics and Physics, University of Maryland, Baltimore County, Baltimore, Md. 21228.
A. S. Willsky is with the Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

I. Introduction

IN THE past, most optimal estimation problems have been studied in a vector space setting. While these results lend themselves to simple solutions in linear systems [1], [2] and in nonlinear systems with finite dimensional sensor orbits [3], no effective synthesis procedures for optimal estimation have been determined for large classes of nonlinear systems.

It is the purpose of this paper to introduce an alternative to the vector space approach in analyzing the properties of nonlinear stochastic processes. We will study random processes on a different type of space, namely, a differentiable manifold, which is the natural domain for certain nonlinear problems of practical importance. This approach will be shown to be useful both in analyzing the properties of certain stochastic processes and in deriving recursive optimal estimation equations that are easily implemented.

More specifically, we will concern ourselves with the