Dynamic Realizations of Sufficient Sequences

BRADLEY W. DICKINSON, SENIOR MEMBER, IEEE, AND EDUARDO D. SONTAG

Abstract—Let $(U_1, U_2, \cdots)$ be a sequence of observed random variables and $(T_1(U_1), T_2(U_1, U_2), \cdots)$ be a corresponding sequence of sufficient statistics (a sufficient sequence). Under certain regularity conditions, the sufficient sequence defines the input/output map of a time-varying, discrete-time nonlinear system. This system provides a recursive way of updating the sufficient statistic as new observations are made. Conditions are provided assuring that such a system evolves in a state space of minimal dimension. Several examples are provided to illustrate how this notion of dimensional minimality is related to other properties of sufficient sequences. The results can be used to verify the form of the minimum dimension (discrete-time) nonlinear filter associated with the autoregressive parameter estimation problem.

I. Introduction

VARIOUS inference problems in time series analysis (i.e., discrete-time signal processing) involve growing subsequences of an infinite sequence of random variables, say $(U_1, U_2, \cdots, U_N, \cdots)$. The theory of sufficient statistics has been applied to this sort of situation in the following way. Let $\Omega^{(k)}$, $k \geq 1$ be the sample space for the first k random variables, $(U_1, U_2, \cdots, U_k)$, and let $F^{(k)}$ designate the Borel sigma field on $\Omega^{(k)}$. We suppose that $P^{(k)}$ is a family of probability measures on $(\Omega^{(k)}, F^{(k)})$ and that for each $k \geq 1$ the family $P^{(k)}$ admits a sufficient statistic, say $T_k(U_1, U_2, \cdots, U_k)$. The sequence $(T_1, T_2, \cdots, T_N, \cdots)$ is called a *sufficient sequence*. T_k summarizes all of the information contained in the first k observations in the sense that statistical inferences concerning $P^{(k)}$ based on $(U_1, \cdots, U_k)$ can be no better than ones based on T_k.

In the applications that provide the motivation for this work certain simplifying assumptions may be added to this general framework. We assume that each random variable U_i takes its values in $\boldsymbol{R}^m$ and that the sufficient sequence $(T_1, T_2, \cdots)$ obeys the following regularity conditions.

Condition 1: T_k takes values in $\boldsymbol{R}^d$ for some fixed d, independent of k.

Condition 2: The functions T_k: $\boldsymbol{R}^{km} \to \boldsymbol{R}^d$ are continuously differentiable.

A final assumption concerns the sequential structure of the sufficient sequence. We are concerned with statistical models where recursive inference techniques may be exploited, so we introduce a *dynamic realizability* assumption to be made precise later. Informally, we assume that the sufficient sequence may be generated as the output of a discrete-time system whose input is the sequence of observed random variables. If its state vector is finite dimensional, such a system provides a practical means of computing the sufficient statistic from the observations, eliminating the need to store a growing observation sequence by exploiting a recursive implementation. The purpose of this paper is to present a suitable formulation of realizability and then to characterize minimal dimension systems associated with realizable sufficient sequences. The tools employed are those of mathematical system theory. The results will be illustrated by examples, and the notion of realizability will be related to a property of sequences of statistics known as *transitivity*, introduced by Bahadur [1]. We will also describe a connection with nonlinear filtering.

Manuscript received August 20, 1984; revised October 19, 1984. This work was supported by the National Science Foundation under Grants ECS82-05772 and ECS84-05460, and by the US Air Force under Grants AFOSR-84-0381 and AFOSR-80-0196.

B. W. Dickinson is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA.

E. D. Sontag is with the Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA.

II. A Motivating Example

Before presenting out technical results, we will introduce an example that led to our study. We consider the problem of parameter estimation for a pth-order autoregressive process. Let $(U_1, U_2, \cdots)$ be successive random variables from the stationary, zero-mean, Gaussian process that satisfies the stochastic difference equation

$$U_t + a_1 U_{t-1} + \cdots + a_p U_{t-p} = W_t, \qquad t \in \boldsymbol{Z}, \quad (1)$$

where $\{W_t\}$ is a sequence of independent, identically distributed Gaussian random variables with mean 0 and variance σ^2. The parameters $\vartheta = (a_1, a_2, \cdots, a_p, \sigma^2)$ belong to the open set $\Theta \subseteq \boldsymbol{R}^{p+1}$ defined by $\sigma^2 > 0$ and the requirement that the polynomial $1 + a_1\lambda + \cdots + a_p\lambda^p$ has all of its zeros outside the unit circle in the complex λ-plane.

The joint density function for $(U_1, \cdots, U_k)$ may be written, for $k > 2p$, in the form [2]

$$p(u_1, \cdots, u_k | \vartheta) = c_k(\vartheta) \exp\left(-A S_k A^* / 2\sigma^2\right), \quad (2)$$

where the superscript "*" indicates transposition, $c_k(\vartheta)$ is a normalizing constant, independent of $(u_1, \cdots, u_k)$, $A = (1, a_1, \cdots, a_p)$, and the $(p+1)$ by $(p+1)$ matrix S_k has elements $(S_k)_{i,j}$, $0 \leq i, j \leq p$, given by

$$(S_k)_{i,j} = \sum_{t=1}^{k-i-j} u_{t+i} u_{t+j}. \quad (3)$$

0018-9448/85/0900-0670$01.00

It follows from the Fisher–Neyman factorization theorem [1], [3] that the statistic composed of the $(p+2)(p+1)/2$ distinct elements of the symmetric matrix S_k is sufficient. This statistic is also necessary in the sense of Dynkin [4], as noted by Arato [5]; hence it is a minimal sufficient statistic, meaning that it is a function of every other sufficient statistic. (We refer the reader to Zacks [6] and Lehmann [7] for a discussion of these concepts.)

The structure of the matrix S_k may be exploited to simplify this statistic [5], [8], [9]. From (3) we have

$$(S_k)_{i,j} = (S_k)_{i-1,j-1} - u_i u_j - u_{k+1-i} u_{k+1-j}, \quad (4)$$

for $1 \le i, j \le p$. Thus the statistic composed of the first row of S_k together with the $(p+1)p/2$ quantities $(U_i U_j + U_{k+1-i} U_{k+1-j})$, $1 \le i, j \le p$, is also (minimal) sufficient. For later use, we give this statistic explicitly:

$$\begin{aligned} T_k = \big((S_k)_{0,0}, \cdots, (S_k)_{0,p}, \\ & U_1^2 + U_k^2, U_1U_2 + U_kU_{k-1}, \cdots, \\ \times \quad & U_1U_p + U_kU_{k+1-p}, U_2^2 + U_{k-1}^2, \\ & U_1U_3 + U_{k-1}U_{k-2}, \cdots, \cdots, \\ & U_p^2 + U_{k+1-p}^2\big). \end{aligned} \quad (5)$$

However, it is also clear that the statistic composed of the first row of S_k and the end segments of the observation sequence,

$$T_k' = \big((S_k)_{0,0}, \cdots, (S_k)_{0,p}, U_1, \cdots, U_p, U_k, \cdots, U_{k-p+1}\big), \quad (6)$$

is also sufficient. Notice that T_k' is not minimal; for example, it is not a function of T_k.

The importance of the statistic T_k' arises from its use in constructing a system whose output is the sufficient sequence $(T_1, T_2, \cdots)$ when its input is the sequence $(U_1, U_2, \cdots)$. (T_k and T_k' are defined in the following way for $k \le 2p$. For $p+1 \le k \le 2p$, (3), (5), and (6) are used directly. For $1 \le k \le p$, (5) and (6) are used with the understanding that variables with subscripts outside the range of 1 to k are set equal to zero, and (3) is used with the understanding that sums whose upper limits are nonpositive are equated to zero.) The desired system is easily constructed by choosing a state vector $x_k = T_{k-1}'$ and using (3) and (5) to obtain recursive equations of the form

$$x_{k+1} = p_k(x_k, U_k), \qquad k \ge 1, \ x_1 = 0 \quad (7a)$$

$$T_k = q_k(x_k, U_k). \quad (7b)$$

We will return to this example after presenting our main result in order to show that no well-behaved system having smaller state dimension can duplicate the input/output mapping of the system of (7).

III. Some System Theory

In this section we will develop the system-theoretic results that will enable us to make the discussion in the introductory section precise. We start with some definitions. Throughout we assume that the space of input values U and the space of output values Y are fixed finite-dimensional differentiable manifolds.

Definition 1: A *(finite-dimensional) system* Σ is a tuple $(X, \{p_k\}_{k\ge1}, \{q_k\}_{k\ge1}, 0)$, where: X is a finite-dimensional differentiable manifold (of states); 0, the initial state, is an element of X; p_k: $X \times U \to X$ is a continuously differentiable map for all $k \ge 1$; and q_k: $X \times U \to Y$ is a continuously differentiable map for all $k \ge 1$.

This definition specifies precisely the nature of the nonlinear, time-varying, discrete-time systems under consideration. The state equations are given by

$$x_{k+1} = p_k(x_k, u_k), \qquad k \ge 1, \ x_1 = 0 \quad (8a)$$

$$y_k = q_k(x_k, u_k). \quad (8b)$$

Some related maps will also be useful. Let $P_{t,j}$: $X \times U^j \to X$ denote the j-step state transition map for the system starting in state x at time t. This is defined for $j \ge 0$ by

$$P_{t,0}(x) = x, \quad (9a)$$

$$P_{t,1}(x,u) = p_t(x,u), \quad (9b)$$

$$P_{t,j+1}(x, u_1, \cdots, u_{j+1}) = p_{t+j}\big(P_{t,j}(x, u_1, \cdots, u_j), u_{j+1}\big), \qquad j \ge 1. \quad (9c)$$

The corresponding output map is $Q_{t,j}$: $X \times U^j \to Y$, defined for $j \ge 1$ by

$$Q_{t,j}(x, u_1, \cdots, u_j) = q_{t+j-1}\big(P_{t,j-1}(x, u_1, \cdots, u_{j-1}), u_j\big), \qquad j \ge 1. \quad (10)$$

The family of maps, $f_k(u_1, \cdots, u_k) = Q_{1,k}(0, u_1, \cdots, u_k)$, constitutes an external, or input/output, description of the system.

Definition 2: The *input/output map* of the system Σ, denoted f_Σ, is the family of continuously differentiable maps $\{f_k(u_1, \cdots, u_k),\ k \ge 1\}$. We say that Σ is a *realization* of the input/output map f_Σ.

We now introduce our definition of realizability for sufficient sequences, using the system theoretic concepts defined previously.

Definition 3: A sufficient sequence $(T_1, T_2, \cdots)$ is *realizable* if the functions $(T_1(u_1), T_2(u_1, u_2), \cdots)$ comprise the input/output map of some system Σ.

If a sufficient sequence is realizable, then there is a finite-dimensional recursive procedure for computing the sufficient statistic as data are sequentially processed. In cases where a fixed, finite-dimensional sufficient statistic can be found, it is usually easy to see how the sufficient sequence can be realized as the input/output map of some system; recall the example already described. There is no completely general realization theory for the class of systems described above yet available in the systems literature. However, there are results that seem adequate for applications like those discussed in this paper. If the input/output map takes the following homogeneous form, given here for scalar inputs

$$y_k = \sum_{i_1=0}^{k} \sum_{i_2=0}^{k} \cdots \sum_{i_m=0}^{k} h_m(k, i_1, \cdots, i_m) \cdot g_m(u_{i_1}) \cdots g_m(u_{i_m}) \quad (11)$$

(or takes a polynomial form, a linear combination of homogeneous terms of various orders), then a certain separable structure of the so-called Volterra kernel h_m will assure existence of a state-affine realization, which takes the form

$$x_{k+1} = \sum_{i=0}^{m-1} A_i(k) x_k g_m^i(u_k) + \sum_{i=1}^{m} b_i(k) g_m^i(u_k) \quad (12a)$$

$$y_k = \sum_{i=0}^{m-1} c_i(k) x_k g_m^i(u_k) + d_m(k) g_m^m(u_k). \quad (12b)$$

Conditions assuring state-affine realizability of input/output maps described by Volterra series have been given in various forms (see [10]–[12] for stationary series and [13] for the time-varying case). Linear analytic realizability of stationary input/output maps has been discussed in [14].

Our goal is to provide conditions assuring that a given system Σ is the most succinct realization of its input/output map, in the sense that its state space X has minimum dimension. This is a standard problem in system theory, although no result in the literature is directly applicable to the class of systems defined above. (Specifically, we need a result that applies to *time-varying*, nonlinear systems). Therefore, we formulate conditions and prove the appropriate theorem here. Roughly speaking, a state space of minimum dimension is one consisting of only "reachable" and "observable" states. These fundamental system-theoretic concepts must be suitably framed for the class of systems under consideration. For a system Σ the image of the map $P_{1,j-1}(0, u_1, \cdots, u_{j-1})$: $U^{j-1} \to X$, denoted by $\text{REACH}_j(\Sigma)$, is called the *reachable set* at time j. Observability, which reflects how the state space X and the output space Y are coupled, will be expressed in terms the Jacobian of the map $Q_t^{\omega_1, \omega_2, \cdots, \omega_r}$: $X \to Y^r$, sending x to the vector $(Q_{t,l_1}, \cdots, Q_{t,l_r})$ evaluated at the points $(x, \omega_1), \cdots, (x, \omega_r)$, respectively. Here the fixed input sequences ω_i have lengths l_i, i.e., $\omega_i \in U^{l_i}$. We make a final definition to combine the appropriate reachability and observability conditions.

Definition 4: A system Σ is *weakly canonical* if and only if there exists an element $x_0 \in X$ and a time $t_0 > 1$, such that: (1) x_0 belongs to the interior of $\text{REACH}_{t_0}(\Sigma)$; (2) there exists an integer r and input sequences $\omega_1 \in U^{l_1}, \cdots, \omega_r \in U^{l_r}$, such that the Jacobian of the map $Q_{t_0}^{\omega_1, \cdots, \omega_r}$ has rank $n = \dim X$ at x_0.

Our main result states that a weakly canonical system has minimal state dimension. As we will see, this is a useful result because it allows this conclusion to be based on straightforward Jacobian calculations.

Theorem 1: Let Σ be a weakly canonical realization of the input/output map $f = f_\Sigma$. Then $\dim X$ is minimal among all possible realizations of f_Σ.

Proof: We must establish that if Σ' is any system with $f_{\Sigma'} = f$, and if $\dim \Sigma' = n'$, then $n' \geq n$. By assumption, Σ is weakly canonical, so there exist x_0, t_0, r and input sequences $\omega_1, \cdots, \omega_r$, as above. Let Q denote the map $Q_{t_0}^{\omega_1, \cdots, \omega_r}$. The observability assumption implies that x_0 is a regular point of Q, so Q is an immersion when restricted to a small enough neighborhood of x_0. Thus, locally Q is equivalent to a standard injection $\boldsymbol{R}^n \to \boldsymbol{R}^n \times \boldsymbol{R}^s$. This implies the existence of an open neighborhood V of x_0, an open subset W of Y^r, and a continuously differentiable map R: $W \to V$, such that $R \circ Q$ restricts to the identity map on V. Since x_0 is in the interior of the reachable set at time t_0, we may assume (taking a smaller U if necessary) that U is included in $\text{REACH}_{t_0}(\Sigma)$.

Now let Q' be the map $Q_{t_0}^{\omega_1, \cdots, \omega_r}$ obtained for the system Σ' when using the same inputs $\omega_1, \cdots, \omega_r$, as above. Let V' be the inverse image of W under Q'; this is open (in X') by continuity of Q'. Denote by T the composition of maps $R \circ Q'$: $V' \to V$. Since V' is a manifold of dimension n', V of dimension n, and T is smooth, it will be enough to show that T is onto in order to conclude that $n' \geq n$.

Pick any x in V. We need to find an x' in V' with $Tx' = x$. Since x is in REACH_{t_0}, there is some input ω of length $t_0 - 1$ with $x = P_{1,t_0-1}(0, \omega)$. Consider the state $x' = P'_{1,t_0-1}(0, \omega)$, obtained by applying the same input to the system Σ'. We first claim that x' is in V' For any i, $1 \leq i \leq r$,

$$Q'_{t_0, l_i}(x', \omega_i) = Q'_{1, t_0+l_i-1}(0, \omega : \omega_i) = Q_{1, t+l_i-1}(0, \omega : \omega_i), \quad (13)$$

where the colon denotes concatenation of the input strings, and the second equality follows because $f_{\Sigma'} = f$. From the definition of ω the last expression equals $Q_{t_0, l_i}(x, \omega_i)$. Thus $Q'(x') = Q(x)$, so in particular $Q'(x')$ is in $Q(V)$, and hence in W, and we have verified the claim that x' is in V'. Then $T(x') = R \circ Q'(x') = R \circ Q(x) = x$, and the proof is complete.

IV. Applications

The simplest application of the theorem proved above is to a sequence of independent random variables, $(U_1, U_2, \cdots)$, identically distributed according to an n-parameter exponential family. The form of the common density function is

$$p(u | \vartheta_1, \cdots, \vartheta_n) = h(u) c(\vartheta_1, \cdots, \vartheta_n) \exp - \sum_{i=1}^{n} \vartheta_i g_i(u). \quad (14)$$

We assume that the parameter space contains n linearly independent vectors, that the density function is strictly positive, and that the functions $g_i(u)$ are continuously differentiable. Then

$$T_j(U_1, \cdots, U_j) = \left(\sum_{i=1}^{j} g_1(U_i), \cdots, \sum_{i=1}^{j} g_n(U_i) \right) \quad (15)$$

is a sufficient statistic, and $(T_1, T_2, \cdots)$ is a sufficient sequence. Choosing as a state vector $x_k = T_{k-1}$, we obtain the obvious corresponding realization

$$x_{k+1} = x_k + (g_1(u_k),\cdots,g_n(u_k)) \tag{16a}$$

$$y_k = x_k + (g_1(u_k),\cdots,g_n(u_k)). \tag{16b}$$

Notice that this realization is time-invariant; the families $\{p_k(x,u)\}$ and $\{q_k(x,u)\}$ do not depend on k.

We suppose that the Jacobian of the functions $(g_1(u_1),\cdots,g_n(u_n))$ has rank n at some point $(u_1^0,\cdots,u_n^0) \in \mathbf{R}^n$. This is the Jacobian of the map $P_{1,n}(0,u_1,\cdots,u_n)$, and by the implicit function theorem, $(u_1^0,\cdots,u_n^0)$ is in the interior of the image of this map, $\text{REACH}_{n+1}(\Sigma)$. The observability condition is trivially satisfied by choosing a single input u of length 1 and noting that for every $t \geq 1$, $Q_{t,1}(x,u) = x + (g_1(u),\cdots,g_n(u))$, whose Jacobian is the identity matrix. Thus, the system is weakly canonical and so has minimum dimension by Theorem 1.

This example is universal, since for independent, identically distributed random variables with a marginal density function satisfying certain regularity properties, a fixed-dimension sufficient statistic will exist only when the density belongs to an exponential family [4], [15]–[18]. The sufficient statistic for such a class is minimal sufficient, and under the Jacobian condition described above, it may be used as the state vector in a minimal-dimension realization of the sufficient sequence.

We now complete the example introduced earlier concerning autoregressive processes to show that these two properties do not always coincide when we have a sequence of dependent random variables. We first describe the realization in (7) in greater detail so that Theorem 1 may be applied. The state at time $k > 2p$, $x_k \in \mathbf{R}^{3p+1}$, takes the form

$$x_k = (\alpha_0,\cdots,\alpha_p,\beta_1,\cdots,\beta_p,\gamma_1,\cdots,\gamma_p), \tag{17}$$

where

$$\alpha_i = \sum_{t=1}^{k-i-1} u_t u_{t+i}, \qquad 0 \leq i \leq p \tag{18a}$$

$$\beta_j = u_j, \qquad 1 \leq j \leq p \tag{18b}$$

$$\gamma_j = u_{k-j}, \qquad 1 \leq j \leq p. \tag{18c}$$

The state equations will be written with superscript "+" denoting time instant $k+1$. From (3), (4), and (6)

$$\alpha_0^+ = \alpha_0 + u_k^2, \alpha_1^+ = \alpha_1 + u_k\gamma_1,\cdots,\alpha_p^+ = \alpha_p + u_k\gamma_p \tag{19a}$$

$$\beta_1^+ = \beta_1,\cdots,\beta_p^+ = \beta_p \tag{19b}$$

$$\gamma_1^+ = \gamma_2,\cdots,\gamma_{p-1}^+ = \gamma_p, \gamma_p^+ = u_k. \tag{19c}$$

The reachability condition will now be verified. Let $t_0 = 3p+2$. The map $P_{1,3p+1}$ sends $(u_1,\cdots,u_{3p+1})$ to the point

$$\left(\sum_{t=1}^{3p+1} u_t^2, \sum_{t=1}^{3p} u_t u_{t+1},\cdots, \sum_{t=1}^{2p+1} u_t u_{t+p}, u_1,\cdots, u_p, u_{3p+1},\cdots,u_{2p+2}\right),$$

and by the implicit function theorem, it suffices to show that the Jacobian of this map has rank $3p+1$ for some $(\bar{u}_1,\cdots,\bar{u}_{3p+1})$, taking x_0 as the image of this point under $P_{1,3p+1}$. The Jacobian is easily calculated, and the task reduces to showing that the following (square) matrix has rank $p+1$ for some $(\bar{u}_1,\cdots,\bar{u}_{3p+1})$:

$$\Delta = \begin{bmatrix} u_{p+1} & u_{p+2} & \cdots & u_{2p+1} \\ u_p + u_{p+2} & u_{p+1} + u_{p+3} & & \cdot \\ \vdots & \vdots & & \vdots \\ u_1 + u_{2p+1} & u_2 + u_{2p+2} & \cdots & u_{p+1} + u_{3p+1} \end{bmatrix}. \tag{20}$$

It is sufficient to show that the determinant of Δ, a polynomial, is not identically zero; this is easily done by noting that the variable u_{p+1} appears only in the diagonal entries of Δ and that its determinant thus takes the form u_{p+1}^{p+1} plus lower degree terms in u_{p+1}.

To check observability note that the output maps into the space $Y = \mathbf{R}^{(p+2)(p+1)/2}$. From (5) the coordinates of y_k are given by

$$\big(\alpha_0 + u_k^2, \alpha_1 + u_k\gamma_1,\cdots,\alpha_p + u_k\gamma_p, \beta_1^2 + u_k^2,$$
$$\beta_1\beta_2 + u_k\gamma_1,\cdots,\beta_1\beta_p + u_k\gamma_{p-1},$$
$$\beta_2^2 + \gamma_1^2, \beta_2\beta_3 + \gamma_1\gamma_2,\cdots,\cdots,\beta_p^2 + \gamma_{p-1}^2\big). \tag{21}$$

As before, let $t_0 = 3p+2$. Choose two input sequences of length 1: $\omega_1 = 0$ and $\omega_2 = 1$. Then consider only the following $3p+1$ coordinates of the map $Q_{t_0}^{\omega_1,\omega_2}$: $\{\alpha_0,\alpha_1,\cdots,\alpha_p,\beta_1^2,\beta_1^2+\gamma_1^2,\cdots,\beta_1^2+\gamma_{p-1}^2\}$ and $\{\alpha_1+\gamma_1, \alpha_2+\gamma_2,\cdots,\alpha_p+\gamma_p\}$. The first group corresponds to ω_1 and the second to ω_2. Writing these coordinates as a column vector and computing the Jacobian (with respect to $\alpha_0,\cdots,\alpha_p$, $\beta_1,\cdots,\beta_p$, $\gamma_1,\cdots,\gamma_p$) gives

$$J = \begin{bmatrix} I & O & O \\ O & B & G \\ S & O & I \end{bmatrix}, \tag{22}$$

where O denotes a matrix of all zeros, I is an identity matrix of appropriate size, and B is a diagonal matrix having β_i as its ith diagonal entry. The exact forms of matrices G and S are not important since it is clear that the matrix J has rank $3p+1$ if and only if $\beta_i \neq 0$ for $1 \leq i \leq p$, or, equivalently, $u_i \neq 0$, $1 \leq i \leq p$. This condition is easily incorporated in the construction of x_0 as a point in the image of the map $P_{1,3p+1}$, as above. This completes our verification that the system in (7), defined by (17)–(19), is weakly canonical and hence, by our theorem, it has a state space of minimum dimension.

V. Realizability and Transitivity

In sequential decision problems involving the models $(\Omega^{(k)}, F^{(k)}, P^{(k)})$, $k \geq 1$, both a stopping rule and an action rule must be determined for each $k \geq 1$. Bahadur [1] showed that it suffices to consider sequential decision rules based only on a sequence of statistics, provided that the sequence is both sufficient and *transitive*. A sequence $(H_1(U_1), H_2(U_1,U_2),\cdots)$, where H_k is $F^{(k)}$-measurable for each k, is called transitive if for every $k \geq 1$ and for any function ψ on $U^{(k)}$, $P^{(k)}$-integrable, and measurable with respect to the sigma field $F_H^{(k)}$ induced by H_k,

$$E_P\{\psi(U_1,\cdots,U_k)|F^{(k-1)}\} = E_P\{\psi(U_1,\cdots,U_k)|F_H^{(k-1)}\}, \quad \text{a.s. } [P], \tag{23}$$

for all $P \in P^{(k)}$. In words the statistic H_k depends on the past, $U^{(k-1)}$, only through H_{k-1}. As illustrated by our autoregressive process example, a sufficient sequence consisting of minimal sufficient statistics in generally not transitive. For instance, the component $U_1U_2 + U_kU_{k-1}$ of T_k from (5) does not depend on $U^{(k-1)}$ only through T_{k-1}.

When a realization for a sufficient sequence is given in the form of (7), a realizable, sufficient sequence is easily obtained by appending the observations ($H_1 = (x_1, U_1)$, $H_2 = (x_2, U_2), \cdots$); this sequence will be transitive if $E_P\{U_k|F^{(k-1)}\} = E_P\{U_k|F_H^{(k-1)}\}$, a.s. $[P]$. This condition ensures that the realization incorporates the innovations representation [19] for the process $\{U_i, i \geq 1\}$ in the sense that the state must contain all past information in the process needed to compute the minimum mean-square error one-step prediction of U_k, given any parametric model. For example, $E_P\{U_k|F^{(k-1)}\} = E_P\{U_k|U_{k-1},\cdots,U_{k-p}\}$, a.s. $[P]$ for a pth-order autoregressive process, and the transitivity condition holds in this case (cf. (18c)).

This discussion suggests an example showing that not all finite-dimensional sufficient sequences are realizable by a finite-dimensional systems. Let $(U_i, i \geq 1)$ be a sequence of observations from a purely nondeterministic, stationary, Gaussian random process with unknown mean μ and known, smooth spectral density function. The joint density function for $(U_1,\cdots,U_n)$ takes the form

$$p(u_1,\cdots,u_n) = h(u_1,\cdots,u_n)c(u) \cdot \exp -\{\mu g_n(u_1,\cdots,u_n)\}, \tag{24}$$

where the sufficient statistic is given explicitly by

$$g_n(u_1,\cdots,u_n) = (u_1,\cdots,u_n)R_n^{-1}\chi_n, \tag{25}$$

where R_n is the covariance matrix of $(U_1,\cdots,U_n)$ and χ_n denotes the n-vector whose entries are all 1.

For analytical purposes, it is convenient to use the innovations representation of the process $\{W_i, i \geq 1\}$, where $w_i = U_i - \mu$, $i \geq 1$. Specifically, for each $n \geq 1$ the sequence $(W_1,\cdots,W_n)$ may be represented as an invertible linear transformation of the first n terms of a process of independent Gaussian random variables, say $\{v_i, i \geq 1\}$. The innovations sequence is obtained by applying the Gram–Schmidt procedure to the original sequence $\{W_i\}$. The form of the representation is [19], [20]:

$$\nu_1 = W_1 \tag{26a}$$

$$\nu_i = W_i + \sum_{j=1}^{i-1} a_{i,j}W_{i-j}, \qquad i > 1. \tag{26b}$$

It follows that the innovations sequence is obtained as the output sequence of a time-varying invertible linear system whose input sequence is $\{W_i\}$. The coefficients in (26b) and the variances of the ν_i random variables are determined by the projection theorem of linear least-squares estimation theory and may be efficiently computed from the covariance sequence using the Levinson–Durin recursions (see [21] for further discussion).

Let $\{\bar{\nu}_i, i \geq 1\}$ be the sequence of random variables obtained from (26) when the input sequence is changed from $\{W_i\}$ to $\{U_i\}$. Using some simple algebra as in [9], it may be shown that the sufficient statistic g_n of (25) is given by

$$g_n(U_1,\cdots,U_n) = \sum_{i=1}^{n} \gamma_i\bar{\nu}_i. \tag{27}$$

The γ_i coefficients are all positive and may be given explicitly in terms of the partial correlation sequence of the process; again see [9]. Here it suffices to note that the computation required to update the sufficient statistic in a recursive way as more observations are made is equivalent to computation of the output of a system that generates the process $\{\bar{\nu}_i\}$. The system is the innovations representation for $\{W_i\}$, a process having the same spectral density as the process $\{U_i\}$. When this spectral density function does not determine a finite-dimensional innovations representation, the sufficient sequence $\{g_n\}$ does not admit a finite-dimensional realization. A necessary condition for the existence of a finite-dimensional system (linear or nonlinear) that provides an innovations representation is that the process have a rational spectral density function [22]. A particular example arises from the choice of the covariance function $E\{(U_k - \mu)(U_j - \mu)\} = \rho^{|k-j|^2}$, for ρ satisfying $-1 < \rho < 1$.

Lauritzen [23] has introduced another notion, related to transitivity, in order to study prediction problems in stochastic processes whose probabilistic structure is unknown. A sufficient sequence $(G_1(U_1), G_2(U_1, U_2), \cdots)$ is called *totally sufficient* if for every $k \geq 1$ the (marginal) sigma field induced by U_k, say $\sigma(U_k)$, and $F^{(k-1)}$ are conditionally independent, given $F_G^{(k-1)}$, the sigma field induced by G_{k-1}, for all $P \in P^{(k)}$. This condition implies that $E_P\{U_k|F^{(k-1)}\} = E_P\{U_k|F_G^{(k-1)}\}$, a.s. $[P]$, which is the condition introduced above to obtain transitivity of the sufficient sequence of augmented states from a realization.

These ideas are illustrated in the following example, due to D. Basu [24], where it turns out to be necessary to increase the dimension of a transitive sufficient sequence in order to achieve totality; as a result, the dimension of the corresponding realization must also increase. Let U_1 and U_2 be independent, identically distributed Gaussian random variables with unknown mean μ and variance 1. For $j > 2$ let $U_j = U_2 + W_j$, where the $\{W_i\}$ is a sequence of independent, identically distributed Gaussian random variables with mean 0 and variance 1. The sequence $(U_1, U_1 + U_2, U_1 + U_2, \cdots)$ is a transitive sufficient sequence with an obvious 1-dimensional realization: $x_1 = 0$; $x_2 = U_1$; $x_3 = x_2 + U_2$; and $x_{k+1} = x_k$ for $k > 3$. The sequence is not totally sufficient because U_j is not independent of $(U_1,\cdots,U_{j-1})$ given $U_1 + U_2$ for $j > 2$. However, U_j is independent of $(U_1,\cdots,U_{j-1})$ given U_2. Hence the sequence $((U_1, 0), (U_1, U_2), (U_1, U_2), \cdots)$ is totally sufficient (and transitive). This sequence requires a realization of dimension 2. Viewed another way, the one-dimensional realization of the original sufficient sequence does not incorporate the innovations representation of the U_i process, because $E\{U_k|F^{(k-1)}\} = U_2 \neq U_1 + U_2 = x_k$ for $k > 2$. After expanding the state space to include U_2, a totally sufficient sequence is obtained.

VI. Realizability and Discrete-Time Nonlinear Filtering Theory

Our autoregressive process example is also a convenient setting to illustrate a connection with (discrete-time) *nonlinear filtering theory*, by which we mean the theory of state estimation in nonlinear stochastic dynamical systems of the form

$$\zeta_{k+1} = f(\zeta_k, w_k) \tag{28a}$$

$$\eta_k = h(\zeta_k, \omega_k), \tag{28b}$$

where ζ_1 and ω_k, $k \geq 1$, are random variables with known statistics. It is desired to find a recursion for computing the sequence of conditional densities $p(\zeta_{k+1}|\eta_1, \cdots, \eta_k)$, $k \geq 1$. The autoregressive process of (1) is transformed into this framework by adopting a Bayesian viewpoint and regarding the parameter vector s as a random quantity. Then by choosing the $2p+1$ dimensional state vector $\zeta_k = (U_{k-1}, \cdots, U_{k-p}, \vartheta)$, and observations $\eta_k = U_k$, state equations are found by using (1) and the time invariance of ϑ. Explicitly,

$$\zeta_{k+1} = \zeta_k \begin{bmatrix} Z & O \\ O & I \end{bmatrix} + \eta_k e_1 \tag{29a}$$

$$\eta_k = -\vartheta(U_{k-1}, \cdots, U_{k-p}, 0)^* + W_k, \tag{29b}$$

where O denotes a matrix of zeros of appropriate dimensions, I is the $(p+1)$-dimensional identity matrix, e_1 is the first unit vector of dimension $2p+1$, and the matrix $Z = (z_{i,j})$, $1 \leq i, j \leq p$ is given by $z_{i,j} = \delta_{i+1-j}$, where δ is the Kronecker symbol.

Note that after p observations are made, i.e., $k \geq p$, the filter need generate only the posterior density function for ϑ, namely $p(\vartheta|u_1, \cdots, u_k)$, or by sufficiency of the statistic T_k from (5) $p(\vartheta|T_k(u_1, \cdots, u_k))$. We can give a useful form for this family of density functions for ϑ by employing natural conjugate density functions [25] or any reproducing family of densities [26] corresponding to the conditional density function (2). Specifically, we rewrite (2) using the functions comprising the sufficient statistic T_k:

$$p(u_1, \cdots, u_k|\vartheta) = g_k(T_k(u_1, \cdots, u_k), \vartheta). \tag{30}$$

Let y_k be the value of T_k obtained from observations of $(U_1, \cdots, U_k)$ as the output of the system (7). A family of densities for ϑ is obtained by taking

$$p_k(\vartheta|y) = g_k(y, \vartheta) \Big/ \int_\Theta g_k(y, \varphi)\, d\varphi, \tag{31}$$

for all y in the image of the map $T_k(\cdot)$. By the reproducing property of this family [26], $\{p_k(\vartheta|y_k)\}$ is a sequence of density functions obeying the sequential form of Bayes' rule required for compatability:

$$p_{k+1}(\vartheta|y_{k+1}) = c(\vartheta)p(u_{k+1}|y_k, \vartheta)p_k(\vartheta|y_k), \tag{32}$$

where $c(\vartheta)$ is the appropriate normalizing constant. Thus, the system (7) is naturally regarded as a nonlinear filter generating the conditional density functions for the state of the system (29). From our previous analysis we may conclude that the minimum-dimension nonlinear filter associated with the autoregressive parameter estimation problem has dimension $3p+1$. For applications of reproducing densities to sequential decision problems involving simultaneous estimation and hypothesis testing, see the work of Birdsall and Gobien [26], who clearly recognized the importance of realizable sufficient sequences.

Our results on minimum-dimension realizations for sufficient sequences may be applied to a more general class of discrete-time nonlinear filtering problems that includes the autoregressive parameter estimation problem. Suppose that there is a underlying Markovian state process and that the observation at each instant is conditionally independent of past states and observations given the present state. Then under suitable regularity conditions, there exists a finite-dimensional transitive sufficient sequence for the conditional distribution of the state given the observation sequence, if and only if this distribution and the conditional distribution of an observation given the state both have the exponential family form [27], [28]. (This is a significant generalization of the situation in the case of independent observations mentioned above in connection with the first example of Section IV.) Sawitzki [29] considered the construction of realizations using a transitive sufficient sequence as a state, but he did not attempt to study any of the associated system theoretic issues such as minimum dimensionality.

VII. Concluding Remarks

As a final point, we will compare our result with the local results on minimum-dimension sufficient statistics of Barankin and Katz [30], [31]. Thanks to our assumptions about the regularity of the functions that comprise a sufficient sequence, realizability is divorced from probabilistic considerations. However, various probabilistic properties of the state space of a stochastic system of the form (7) may be investigated. A particular example from [30], [31] (see also [6]) illustrates the character of our notion of system dimension. Consider the bivariate random variables $U_k = (u', u'')_k$ belonging to the 2-parameter exponential family of density functions

$$f(u', u''|\vartheta_1, \vartheta_2) = c(\vartheta_1, \vartheta_2) \exp -\{\vartheta_1\varphi(u') + \vartheta_2\varphi(u'')\}, \tag{33}$$

where the parameters ϑ_1 and ϑ_2 are positive, $c(\vartheta_1, \vartheta_2)$ is a normalizing constant, and the function φ is defined by

$$\varphi(w) = \begin{cases} w^2, & \text{if } w < 0 \\ 0, & \text{if } 0 \leq w \leq 1. \\ (w-1)^2, & \text{if } w > 1 \end{cases} \tag{34}$$

In the notation of Section IV we have $g_1(u) = \varphi(u')$ and $g_2(u) = \varphi(u'')$. The Jacobian of these functions vanishes only on the set given by $\{(u', u''): 0 \leq u' \leq 1 \text{ or } 0 \leq u'' \leq 1\}$, so the corresponding system (16) has minimum dimension 2. Yet for every $j \geq 1$, there is a subset of $\Omega^{(j)}$ having nonzero probability where the sufficient statistic T_j from (15) is identically zero.

Such *local* phenomena are incorporated into our theory in a straightforward way and will be important in a formulation of a realization theory for stochastic nonlinear systems. However, we believe our result will be adequate in many statistical applications, because the conditions required for a system to be weakly canonical will turn out to be generic, in a suitable sense. In the strongest sense, for all k sufficiently large, they will be satisfied P-almost surely for all $P \in P^{(k)}$. Even in contrived examples like the one of (33 and 34), the probability of observing an exceptional sequence decreases geometrically, and our system-theoretic approach gives generic results in the weak (Chebyshev) sense.

The same kind of local phenomena were treated by Barankin and Katz [30], [31] in their study of minimum-dimension, continuously differentiable sufficient statistics. Viewed in terms of sufficient sequences and realizations, the Barankin–Katz work deals with constructing an output space $Y^{(k)}$ of minimum dimension at a fixed time instant k. This *static* problem is much earlier than the problem we have addressed. Indeed, Jacobian conditions provide existence results by application of the implicit function theorem.

References

[1] R. R. Bahadur, "Sufficiency and statistical decision functions," *Ann. Math. Statist.*, vol. 25, pp. 423–462, 1954.

[2] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, rev. ed. San Francisco: Holden-Day, 1976.

[3] J. Neyman, "Su un teorema concernente le cosiddette statistiche sufficienti," *Giornale Inst. Attauri*, vol. 6, pp. 320–344, 1935.

[4] E. B. Dynkin, "Necessary and sufficient statistics for a family of probability distributions," in *Selected Translations in Mathematical Statistics and Probability*, vol. 1. Providence, RI: Amer. Math. Soc., 1961, pp. 17–40.

[5] M. Arato, "On the sufficient statistics for stationary Gaussian random processes," *Theory Probab. Appl.*, vol. 6, pp. 199–210, 1961.

[6] S. Zacks, *The Theory of Statistical Inference*. New York: Wiley, 1971.

[7] E. L. Lehmann, *Testing Statistical Hypotheses*. New York: Wiley, 1959.

[8] T. W. Anderson, "Determination of the order of dependence in normally distributed time series," in *Proc. Symp. Time Series Analysis, 1962*, M. Rosenblatt, Ed. New York: Wiley, 1963.

[9] B. W. Dickinson, "Theory and application of Gaussian autoregressive processes in detection theory," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 343–347, 1981.

[10] E. D. Sontag, *Polynomial Response Maps*, vol. 13, Lecture Notes in Control and Information Sciences. Berlin: Springer-Verlag, 1979.

[11] ——, "Realization theory of discrete time nonlinear systems: Part I—The bounded case," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 342–356, 1979.

[12] S. J. Clancy and W. J. Rugh, "The regular transfer function and bilinear and state-affine realizations for stationary, homogeneous, discrete-time systems," in *Proc. Conf. Information Sciences and Systems*, Johns Hopkins University, Baltimore, MD, 1978, pp. 167–172.

[13] C. A. Schwartz, "Topics in finite dimensional realization theory," Ph.D. dissertation, Princeton University, Princeton, NJ, 1984.

[14] S. Monaco and D. Normand-Cyrot, "On the realization of nonlinear discrete-time systems," *Syst. Contr. Lett.*, vol. 5, pp. 145–152, 1984.

[15] L. D. Brown, "Sufficient statistics in the case of independent random variables," *Ann. Math. Statist.*, vol. 35, pp. 1456–1474, 1964.

[16] J. L. Denny, "Sufficient conditions for a family of distributions to be exponential," in *Proc. Nat. Acad. Sci.*, vol. 57, 1967, pp. 1184–1187.

[17] ——, "Note on a theorem of Dynkin on the dimension of sufficient statistics," *Ann. Math. Statist.*, vol. 40, pp. 1474–1476, 1968.

[18] R. Shimizu, "Remarks on sufficient statistics," *Ann. Inst. Statist. Math.*, vol. 18, pp. 49–66, 1966.

[19] M.R. Gevers and T. Kailath, "An innovations approach to least-squares estimation—Part VI: Discrete-time innovations representations and recursive estimation," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 588–600, 1973.

[20] P. Whittle, *Prediction and the Regulation by Linear Least-Squares Methods*. New York: Van Nostrand Reinhold, 1963.

[21] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 146–181, 1974.

[22] C. A. Schwartz, B. W. Dickinson, and E. D. Sontag, "Characterizing innovations realizations for random processes," *Stochastics*, vol. 11, pp. 159–172, 1984.

[23] S. L. Lauritzen, "Sufficiency, prediction and extreme models," *Scand. J. Statist.*, vol. 1, pp. 128–134, 1974.

[24] D. Basu, Comments on [23], in *Proc. Conf. Foundational Questions in Statistical Inference*, O. Barndorff-Nielsen, P. Blaesild, and G. Schou, Eds. Aarhus, Denmark; also *Memoirs*, vol. 1, Dept. of Theoretical Statistics, Univ. of Aarhus, 1974, p. 269.

[25] H. Raiffa and R. Schlaiffer, *Applied Statistical Decision Theory*, Cambridge, MA: Mass. Inst. Tech., 1961.

[26] T. G. Birdsall and J. O. Gobien, "Sufficient statistics and reproducing densities in simultaneous sequential detection and estimation," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 760–768, 1973.

[27] J. A. Bather, "Invariant conditional distributions," *Ann. Math. Statist.*, vol. 36, pp. 829–846, 1975.

[28] G. Sawitzki, "Finite dimensional filter systems in discrete time," *Stochastics*, vol. 5, pp. 107–114, 1981.

[29] G. Sawitzki, "Exact filtering in exponential families: discrete time," in *Stochastic Control Theory and Stochastic Differential Systems*, vol. 16, Lecture Notes in Control and Information Sciences, M. Kohlmann, W. Vogel, Eds. Berlin: Springer-Verlag, 1979, pp. 554–558.

[30] E. W. Barankin and M. Katz, "Sufficient statistics of minimal dimension," *Sankhyā*, vol. 21, pp. 217–246, 1959.

[31] E. W. Barankin, "Application to exponential families of the solution of the minimal dimension problem of sufficient statistics," *Bull. Inst. Int. Statist.*, vol. 38, pp. 141–150, 1961.