



Some Connections Between the Theory of Sufficient Statistics and the Identifiability Problem

Author(s): Giorgio Picci

Source: *SIAM Journal on Applied Mathematics*, Vol. 33, No. 3 (Nov., 1977), pp. 383-398

Published by: [Society for Industrial and Applied Mathematics](#)

Stable URL: <http://www.jstor.org/stable/2100699>

Accessed: 31/10/2014 07:00

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Society for Industrial and Applied Mathematics is collaborating with JSTOR to digitize, preserve and extend access to *SIAM Journal on Applied Mathematics*.

<http://www.jstor.org>

SOME CONNECTIONS BETWEEN THE THEORY OF SUFFICIENT STATISTICS AND THE IDENTIFIABILITY PROBLEM*

GIORGIO PICCI†

Abstract. In this paper the identifiability problem is formulated as a dual of the data reduction problem in statistical inference. Some classical results in the theory of sufficient statistics are dualized in order to obtain criteria for finding “maximal identifiable statistics” in parametric models. Applications to identifiability of linear dynamical systems are discussed.

1. Introduction. The objective of this paper is to describe a remarkable duality which can be established between the concept of sufficient statistics and the problem of identifiability. “Identifiability” is an intrinsic property of a statistical model which makes it possible to distinguish between different parametric structures from observed samples. Though it is especially important in *dynamic* modeling, the problem of identifiability has been explicitly mentioned and studied in classical statistical inference at least since 1950. Rothenberg [16] gives an historical account of the statistical literature as well as the basic ideas in a finite dimensional setting.

Using duality and the concept of sufficiency we prove alternative characterizations of identifiability as well as criteria for finding “maximal identifiable statistics” which are useful in a dynamic context. The derivation of some classical results of the theory of sufficient statistics can also be simplified and put in a more natural light (compare e.g. the characterization of minimal sufficient σ -algebras and the extension of Dynkin’s theorem, for the exponential family).

A few applications are also presented. These are selected from the most commonly encountered identifiability problems in system theory. These examples point out that a “deterministic” approach (implicit for example in [5], [6]) used to find identifiable parametrizations may not be the correct approach to the problem. Somewhat different answers may be obtained when proper recognition is given to the stochastic coupling between observations and parameters. Some related results for discrete time systems can be found in [17] and [18].

2. Sufficient and unresolvable statistics. Let us consider two measurable mappings y and u defined on a common probability space $\{\Omega, \mathcal{A}, \mathbf{P}\}$,

$$(2.1) \quad y: \{\Omega, \mathcal{A}\} \rightarrow \{Y, \mathcal{Y}\},$$

$$(2.2) \quad u: \{\Omega, \mathcal{A}\} \rightarrow \{U, \mathcal{U}\},$$

taking values in complete separable metric spaces Y and U , with relative Borel σ -algebras \mathcal{Y} and \mathcal{U} , respectively. The sub σ -algebras of \mathcal{A} induced by y and u on the basic space Ω , will be denoted by hatted symbols like $\hat{\mathcal{Y}}$ and $\hat{\mathcal{U}}$, respectively.

Referring to the customary setup of statistical inference, we will agree to interpret y as the *observation* variable and u as the *inaccessible* variable (or *parameter*). For applications to identifiability, typically y will be a random process (taking values on an appropriate space of time functions) and u a random vector

* Received by the editors June 7, 1974, and in revised form May 24, 1976.

† Laboratorio per Ricerche di Dinamica dei Sistemi e di Elettronica Biomedica del C.N.R., Padova, Italy.

taking values in a finite dimensional space. Let us suppose that the probabilistic interaction between y and u is specified by assigning a transition probability¹ P on $(\mathcal{Y} \times U)$ and an “a priori” probability measure H_0 on \mathcal{U} satisfying the compatibility conditions:

$$(2.3) \quad P(A, u) = \mathbf{P}(y^{-1}(A) | \hat{\mathcal{U}}), \quad A \in \mathcal{Y},$$

$$(2.4) \quad H_0(B) = \mathbf{P}(u^{-1}(B)), \quad B \in \mathcal{U}.$$

For (2.3) to make sense, one ought to prove that the conditional probability in the second member admits a regular version (see Loeve [13]). This can be done, but, as far as possible, we will try not to get too involved in technicalities of measure theoretical nature. For details we refer the reader to the report [14].

DEFINITION 2.1. The *Bayesian dual* of P is a transition probability H on $(\mathcal{U} \times Y)$ with values $H(B, \eta)$, $B \in \mathcal{U}$, $\eta \in Y$, satisfying

$$(2.5) \quad H(B, y) = \mathbf{P}(u^{-1}(B) | \hat{\mathcal{Y}}) \quad \text{a.s.}$$

In the statistical literature H is usually referred to as the “a posteriori” probability measure. It reflects how the a priori knowledge about u (i.e. the measure H_0 on \mathcal{U}) is modified as a consequence of the observation of the sample value $\eta = y(\omega)$.

The computation of H can be carried out via “Bayes rule” as shown by the following theorem, proved in [10].

THEOREM 2.2. Define the set function P_B , $B \in \mathcal{U}$, as

$$(2.6) \quad P_B(A) = \int_B P(A, \xi) H_0(d\xi), \quad A \in \mathcal{Y}.$$

Then P_B is a finite measure on \mathcal{Y} which is absolutely continuous with respect to P_0 , the probability distribution induced by \mathbf{P} on the space $\{U, \mathcal{Y}\}$. Its Radon–Nikodym derivative satisfies

$$(2.7) \quad H(B, \cdot) = \frac{dP_B}{dP_0} \quad \text{a.s.}$$

where P_0 is given by the formula

$$(2.8) \quad P_0(A) = \int_U P(A, \xi) H_0(d\xi), \quad A \in \mathcal{Y}.$$

Quite often in applications the data provided by the observation of y are *redundant*, in the sense that a smaller σ -algebra than $\hat{\mathcal{Y}}$ may provide the same amount of information about u . The following is an abstract procedure to eliminate this redundancy.

Let us introduce the equivalence relation “ \sim ” on Y , by setting

$$(2.9) \quad \eta_1 \sim \eta_2 \quad \text{if and only if} \quad H(B, \eta_1) = H(B, \eta_2), \quad \forall B \in \mathcal{U}.$$

¹ A transition probability P on $(\mathcal{Y} \times U)$ is a family of measures $\{P(\cdot, \xi), \xi \in U\}$ such that

(i) $P(\cdot, \xi)$ is a probability measure on \mathcal{Y} for each $\xi \in U$,
(ii) $P(A, \cdot)$ is a \mathcal{U} -measurable mapping for each $A \in \mathcal{Y}$.

Consider the quotient set $X_0 = Y/\sim$ and the canonical surjection

$$(2.10) \quad \varphi_0: \eta \mapsto [\eta]$$

where $[\eta]$ is the equivalence class of η in the relation.

Consider also the smallest sub σ -algebra of \mathcal{Y} with respect to which the a posteriori probabilities $\{H(B, \cdot), B \in \mathcal{U}\}$ are measurable:

$$(2.11) \quad \mathcal{F}_0 = \sigma\{H(B, \cdot), B \in \mathcal{U}\}.$$

Then the next lemma can be easily checked.

LEMMA 2.3. *If we define, for each $B \in \mathcal{U}$, $\hat{H}(B, \cdot): X_0 \rightarrow [0, 1]$ as*

$$(2.12) \quad \hat{H}(B, x) = H(B, \eta), \quad \text{for } \eta \text{ such that } \varphi_0(\eta) = x,$$

and the σ -algebra \mathcal{X}_0 of subsets of X_0 , as the one induced by the family of mappings $\{\hat{H}(B, \cdot), B \in \mathcal{U}\}$, then:

- (1) *\hat{H} is a transition probability on $(\mathcal{U} \times X_0)$.*
- (2) *The factorization*

$$(2.13) \quad H(B, \eta) = \hat{H}(B, \varphi_0(\eta)),$$

which holds for all $B \in \mathcal{U}$, $\eta \in Y$, is canonical. This means that $H: (Y, \mathcal{Y}) \rightarrow \mathcal{P}(\mathcal{U})$ ($\mathcal{P}(\mathcal{U})$ is the space of all probability measures over $\{U, \mathcal{U}\}$) is factored through a surjective φ_0 and an injective \hat{H} .²

- (3) *With the above defined \mathcal{X}_0 , φ_0 becomes a measurable mapping.*

(4) *The smallest σ -algebra measuring φ_0 is precisely \mathcal{F}_0 , given by (2.11). Actually, φ_0 , considered as a mapping between σ -algebras is a σ -isomorphism of \mathcal{F} onto \mathcal{X}_0 (i.e., φ_0 preserves countable set operations and is one to one and onto).*

The factorization (2.13) emphasizes the role played by φ_0 ; two observed data η_1, η_2 for which $\varphi_0(\eta_1) = \varphi_0(\eta_2)$ provide exactly the same information about the unknown parameter u , the corresponding a posteriori probabilities being identical.

DEFINITION 2.4. Any mapping φ_0 for which the canonical factorization property (2.13) holds is called a *minimal sufficient statistic for u* .

Remarks. (i) An obvious observation to be made is that a minimal sufficient statistic is unique only up to an arbitrary measurable bijective $\gamma: (X_0, \mathcal{X}_0) \rightarrow (X_1, \mathcal{X}_1)$. In fact the commutative diagram formed by the dashed arrows below

$$(2.14) \quad \begin{array}{ccc} \{Y, \mathcal{Y}\} & \xrightarrow{H} & \mathcal{P}(\mathcal{U}) \\ \downarrow \varphi_0 & \nearrow \hat{H} & \uparrow \hat{H}(\cdot, \mu^{-1}) \\ \{X_0, \mathcal{X}_0\} & \xrightarrow{\gamma} & \{X_1, \mathcal{X}_1\} \end{array} \quad \begin{array}{c} \text{dashed arrow } \gamma \circ \varphi_0 \\ \text{dashed arrow } \hat{H}(\cdot, \mu^{-1}) \end{array}$$

² The notation is somewhat inconsistent; here H has to be interpreted as the mapping $\eta \mapsto H(\cdot, \eta)$. The same remark applies to \hat{H} .

represents a new canonical factorization of H (through $\gamma_0\varphi_0$ and $\hat{H}(\cdot, \gamma^{-1})$), for arbitrary bijective γ .

(ii) For any canonical factorization $\hat{H}(\cdot, \varphi_0)$ of H , the σ -algebra $\mathcal{X}_0 = \sigma\{\hat{H}(B, \cdot), B \in \mathcal{U}\}$, induced by \hat{H} on X_0 , is to be regarded as the *natural* family of measurable sets on X_0 . With this choice φ_0 induces precisely \mathcal{F}_0 . We will call \mathcal{F}_0 the *minimal sufficient σ -algebra* for u .

DEFINITION 2.5. Measurable mappings $\varphi: (Y, \mathcal{Y}) \rightarrow (X, \mathcal{X})$ which factorize H for a suitable transition probability \tilde{H} on $(\mathcal{U} \times X)$

$$(2.15) \quad H(B, \eta) = \tilde{H}(B, \varphi(\eta)), \quad \forall B \in \mathcal{U},$$

(not necessarily in a *canonical* way) will be called *sufficient statistics* for u . The induced σ -algebras $\mathcal{F} = \sigma\{\varphi\}$ on $\{Y, \mathcal{Y}\}$ are *sufficient σ -algebras* for u .

From the definition (2.11) of \mathcal{F}_0 it is clear that

$$(2.16) \quad \mathcal{F}_0 \subseteq \mathcal{F} \subseteq \mathcal{Y},$$

for all sufficient σ -algebras \mathcal{F} . Since \mathcal{F}_0 is sufficient by Definition 2.5, we see that sufficient σ -algebras constitute a partially ordered set with minimal element \mathcal{F}_0 . For this reason φ_0 is termed “minimal” in Definition 2.4.

THEOREM 2.6. *The statistic φ is sufficient for u if and only if the conditional distribution $P(A, \xi|\mathcal{F})$ does not depend on the variable $\xi \in U$. More precisely, φ is sufficient for u if and only if*

$$(2.17) \quad P(A, \xi|\mathcal{F}) = P_0(A|\mathcal{F}), \quad \forall A \in \mathcal{Y}, \quad \xi \in U.$$

Proof. By setting $\hat{B} = u^{-1}(B)$, the defining relationship (2.15) for a sufficient statistic φ can be rewritten as

$$(2.18) \quad \mathbf{P}(\hat{B}|\mathcal{Y}) = \mathbf{P}(\hat{B}|\mathcal{F}) \quad \text{a.s.}, \quad \forall \hat{B} \in \hat{\mathcal{U}},$$

which in turn is equivalent to

$$(2.19) \quad \mathbf{P}(\hat{B}|\hat{\mathcal{U}} \vee \hat{\mathcal{F}}) = \mathbf{P}(\hat{B}|\hat{\mathcal{F}}) \quad \text{a.s.}, \quad \forall \hat{B} \in \hat{\mathcal{U}},$$

since $\hat{\mathcal{F}} \subseteq \hat{\mathcal{U}}$. But (2.19) tells that $\hat{\mathcal{U}}$ and $\hat{\mathcal{F}}$ are conditionally independent given $\hat{\mathcal{F}}$; see Loeve [13, p. 351].

The role of $\hat{\mathcal{U}}$ and $\hat{\mathcal{F}}$ can be reversed, yielding

$$(2.20) \quad \mathbf{P}(\hat{A}|\hat{\mathcal{U}} \vee \hat{\mathcal{F}}) = \mathbf{P}(\hat{A}|\hat{\mathcal{F}}) \quad \text{a.s.}, \quad \forall \hat{A} \in \hat{\mathcal{V}},$$

where we have used again the notation $\hat{A} = y^{-1}(A)$, $A \in \mathcal{Y}$. Now, it is not hard to show [14, Lemma 1.3] that $\mathbf{P}(y^{-1}(A)|\hat{\mathcal{U}} \vee \hat{\mathcal{F}}) = P(A, \xi|\mathcal{F})$ a.s., where the symbol $P(\cdot, \xi|\mathcal{F})$ stands for the conditional probability given \mathcal{F} , which can directly be defined over the space $\{Y, \mathcal{Y}, P(\cdot, \xi)\}$. Thus (2.19) gives immediately (2.16). Q.E.D.

Remarks. Theorem 2.6 is usually stated as a *definition* of a sufficient statistic (or σ -algebra); see Bahadur [2], Zacks [20]. It seems that the approach followed here is more direct and intuitive. At least it has the merit of clarifying the essential equivalence between sufficiency and the concept of conditional independence (c.f. the above proof). This does not seem to have been noticed before.

The main motivation for the different approach we have taken comes, however, from the *duality* which can be established between the data reduction

problem (origin of the concept of sufficiency) and the identifiability problem (which will produce the dual concept of *unresolvable statistic*).

Let us define another equivalence relation on the parameter space U , by defining

$$(2.21) \quad \xi_1 \sim \xi_2 \text{ if and only if } P(A, \xi_1) = P(A, \xi_2), \quad \forall A \in \mathcal{Y}$$

(since in the future we will refer only to the equivalence (2.21), we continue to use the same symbol \sim).

We denote by Σ_0 the quotient set U/\sim and consider the canonical surjection

$$(2.22) \quad \theta_0: \xi \mapsto [\xi],$$

mapping U onto Σ_0 . As before we consider also the smallest sub σ -algebra of \mathcal{U} with respect to which all random variables $\{P(A, \cdot), A \in \mathcal{Y}\}$ are measurable:

$$(2.23) \quad \mathcal{C}_0 = \sigma\{P(A, \cdot), A \in \mathcal{Y}\}.$$

Then, in perfect duality to Lemma 2.3, we have

LEMMA 2.7. Let us define, for each $A \in \mathcal{Y}$, $\hat{P}(A, \cdot): \Sigma_0 \rightarrow [0, 1]$ as

$$(2.24) \quad \hat{P}(A, \sigma) = P(A, \xi), \quad \text{for } \xi \text{ such that } \theta_0(\xi) = \sigma,$$

and the σ -algebra \mathcal{S}_0 as the one induced on Σ_0 by the family of mappings $\{\hat{P}(A, \cdot), A \in \mathcal{Y}\}$. Then:

- (1) \hat{P} is a transition probability on $(\mathcal{Y} \times \Sigma_0)$.
- (2) The factorization

$$(2.25) \quad P(A, \xi) = \hat{P}(A, \theta_0(\xi)),$$

holding for all $A \in \mathcal{Y}$, $\xi \in U$ is canonical (i.e., θ_0 is surjective and \hat{P} is injective).

- (3) θ_0 is a $(\mathcal{U}, \mathcal{S}_0)$ measurable mapping,

(4) The smallest sub σ -algebra of \mathcal{U} with respect to which θ_0 is measurable is precisely $\mathcal{C}_0 \cdot \mathcal{C}_0$ and \mathcal{S}_0 are σ -isomorphic.

The canonical factorization property (2.25) brings into evidence the way the probability law governing the observations is affected by the parameter values ξ . The dependence is not in general one to one (unless θ_0 reduces to the identity); thus, if we are to infer about u , we cannot expect that the observation of y would permit an arbitrarily fine discrimination on the set U .

The equivalence classes $[\xi]$ might be called *maximally unresolvable for y*. Actually, if two parameter values ξ_1, ξ_2 belong to the same equivalence class then $(\theta_0(\xi_1) = \theta_0(\xi_2))$ and so $P(A, \xi_1) = P(A, \xi_2)$ for all $A \in \mathcal{Y}$. Thus we cannot hope to distinguish ξ_1 from ξ_2 by observing y . Conversely, each observation class corresponds to a different probability distribution of the variable y ; i.e. ξ_1 and ξ_2 belong to different equivalence classes if for at least one event $A \in \mathcal{Y}$, $P(A, \xi_1) \neq P(A, \xi_2)$. This means that we can, in principle, devise some testing procedure in order to distinguish ξ_1 from ξ_2 .

DEFINITION 2.8. Any measurable mapping θ_0 for which the canonical factorization property (2.25) holds will be called a *maximal unresolvable statistic for y*.

Again a maximal unresolvable statistic is only unique up to an arbitrary measurable bijection $\beta: (\Sigma_0, \mathcal{S}_0) \rightarrow (\Sigma_1, \mathcal{S}_1)$, the composition rules

$$(2.26) \quad \theta_1 = \beta \circ \theta_0, \quad \hat{P}_1(\cdot, \sigma_1) = \hat{P}(\cdot, \beta^{-1}(\sigma_1))$$

producing new canonical factorizations of P . By duality we are also led to

DEFINITION 2.9. Any measurable mapping $\theta: (U, \mathcal{U}) \rightarrow (\Sigma, \mathcal{S})$ which factorizes P , for a suitable transition probability \tilde{P} on $(\mathcal{Y} \times \Sigma)$,

$$(2.27) \quad P(A, \xi) = \tilde{P}(A, \theta(\xi)), \quad \forall A \in \mathcal{Y},$$

(not necessarily in a canonical way) is an *unresolvable statistic* for y . The induced σ -algebra $\mathcal{C} = \sigma(\theta)$ on (U, \mathcal{U}) will be called an *unresolvable σ -algebra* for y .

THEOREM 2.10. θ is an unresolvable statistic if and only if the conditional distribution $H(B, \eta | \mathcal{C})$ does not depend on $\eta \in Y$. In fact θ is unresolvable if and only if

$$(2.28) \quad H(B, \eta | \mathcal{C}) = H_0(B | \mathcal{C}), \quad \forall B \in \mathcal{U}, \quad \eta \in Y.$$

A criterion for unresolvability is supplied by the following result, dual of the classical factorization theorem.

THEOREM 2.11. Assume that the family $\{H(\cdot, \eta), \eta \in Y\}$ is dominated on U , by some fixed σ -finite measure μ ; let the density

$$(2.29) \quad h(\cdot, \eta) = dH(\cdot, \eta) / d\mu$$

be jointly measurable on $\{U \times Y, \mathcal{U} \otimes \mathcal{Y}\}$. Then θ is unresolvable if and only if the density $h(\xi, \eta)$ factorizes according to

$$(2.30) \quad h(\xi, \eta) = g(\xi) \hat{h}(\theta(\xi), \eta), \quad \mu\text{-a.s.} \quad \forall \eta \in Y,$$

where g is a random variable which is μ -a.s. nonnegative, $\hat{h}(\theta, \eta)$, $\eta \in Y$, is a family of random variables a.s. nonnegative with respect to the measure $\int g d\mu$ on \mathcal{U} .

Remarks. We may interpret Theorem 2.10 as follows. Suppose the unresolvable statistic θ is known. Then formula (2.28) says that the observed sample $\eta = y(\omega)$ does not help in improving our a priori knowledge of u . This is because the a posteriori probability given θ and $y(\omega) = \eta$ is actually independent of η . Thus an unresolvable statistic gives a resolution on the parameter space which cannot be improved (refined) by the observations. Accordingly, a maximal unresolvable statistic θ_0 , defines the coarsest partition of U which is left "invariant" (in the above specified sense) by the observation of y .

As previously pointed out, a *maximal* unresolvable statistic, θ_0 , is the unique (modulo isomorphisms) unresolvable statistic for which \tilde{P} in (2.27) is one to one. This is also equivalent to saying that θ_0 defines the best possible discrimination on U allowed by the observational scheme. Of course there is a whole class of statistics on $\{U, \mathcal{U}\}$ giving a "worse" resolution on the parameter values than θ_0 . For an estimation problem on u to be well-posed one has to restrict oneself to this class.

DEFINITION 2.12. Let $\{S, \mathcal{B}\}$ be a measurable space. An *identifiable statistic of u* is any measurable mapping

$$(2.31) \quad \psi: (U, \mathcal{U}) \rightarrow (S, \mathcal{B})$$

whose induced σ -algebra is coarser than \mathcal{C}_0 (i.e., is a sub σ -algebra of \mathcal{C}_0).

Equivalently we could say that ψ is identifiable if it depends on ξ only through the maximal unresolvable statistic θ_0 , i.e. if $\psi(\xi) = \hat{\psi}(\theta_0(\xi))$ for a suitable $\hat{\psi}$.

By definition the identifiable statistic which gives the maximal attainable information about u is precisely θ_0 . (On the other hand, "trivial" statistics exist which can always be identified, e.g. any constant function on U).

Identifiability and unresolvability correspond to dual concepts of necessity and sufficiency in the classical theory of sufficient statistics (Bahadur [2]); just as φ_0 is both necessary and sufficient, similarly θ_0 is at the same time identifiable and unresolvable. θ_0 will also be called *maximal identifiable statistic*.

In order to determine the maximal identifiable statistic we need a sharper form of the factorization theorem 2.11. The following lemma plays a basic role.

LEMMA 2.13. Let the family $\{P(\cdot, \xi), \xi \in U\}$ be dominated by a σ -finite measure W , then $\{P(\cdot, \xi), \xi \in U\}, \{H(\cdot, \eta), \eta \in Y\}$ are dominated by P_0 and H_0 respectively; i.e.

$$(2.32) \quad P_0(A) = 0 \Rightarrow P(A, \xi) = 0 \quad \text{for } H_0 \text{ almost all } \xi \in U,$$

$$(2.33) \quad H_0(B) = 0 \Rightarrow H(B, \eta) = 0 \quad \text{for } P_0 \text{ almost all } \eta \in Y.$$

There exists a fixed H_0 null set $N_1 \in \mathcal{U}$, such that $P_0(A) = 0$ implies $P(A, \xi) = 0$ for all $\xi \in U - N_1$ and N_1 is independent of A . Similarly, a fixed null set $N_2 \in \mathcal{Y}$ can be found such that $H_0(B) = 0 \Rightarrow H(B, \eta) = 0$ for all $\eta \in Y - N_2$, independently of B .

The proof is somewhat lengthy and will be omitted. It can be found in [14].

Notice that $\{P(\cdot, \xi), \xi \in U\}$ and $\{H(\cdot, \eta), \eta \in Y\}$ are then *equivalent* (almost surely) to P_0 and H_0 respectively. This follows from (2.7) and from the (dual) relationship

$$(2.34) \quad H_0(B) = \int_Y H(B, \eta) P_0(d\eta), \quad B \in \mathcal{U}.$$

Since $P(\cdot, \xi)$ can be modified to an arbitrary probability measure for all ξ belonging to a H_0 null set, there is the possibility of selecting a version for which the equivalence to P_0 holds without exceptional sets. From now on we will always assume that we are working with such a version of P . The same applies to $H(\cdot, \eta)$.

By Lemma 2.13 there exist densities m', m'' such that

$$(2.35) \quad P(A, \xi) = \int_A m'(\eta, \xi) P_0(d\eta), \quad \xi \in U,$$

for all $A \in \mathcal{Y}$, and

$$(2.36) \quad H(B, \eta) = \int_B m''(\eta, \xi) H_0(d\xi), \quad \eta \in Y,$$

for all $B \in \mathcal{U}$.

THEOREM 2.14. *The densities m' , m'' coincide $P_0 \times H_0$ a.s. Every version mod $P_0 \times H_0$, of $m = m' = m''$ possesses the following properties:*

(i) $\{m(\cdot, \xi), \xi \in U\}, \{m(\eta, \cdot), \eta \in Y\}$ induce minimal sufficient and maximally unresolvable σ -algebras. More precisely,

$$(2.37) \quad \sigma\{m(\cdot, \xi), \xi \in U - N_1\} = \mathcal{F}_0,$$

$$(2.38) \quad \sigma\{m(\eta, \cdot), \eta \in Y - N_2\} = \mathcal{C}_0,$$

the sets N_1 and N_2 being H_0 and P_0 null (respectively), depending on the particular version of m being chosen.

(ii) $m(\eta, \xi)$ factorizes according to

$$(2.39) \quad m(\eta, \xi) = \hat{m}(\varphi_0(\eta), \theta_0(\xi)),$$

with $\hat{m}: X_0 \times \Sigma \rightarrow R_+$ measurable and (φ_0, θ_0) minimal sufficient and maximal unresolvable statistics.

(iii) The mappings $x \mapsto \hat{m}(x, \cdot)$, $x \in X_0$ and $\sigma \mapsto \hat{m}(\cdot, \sigma)$, $\sigma \in \Sigma_0$ are one to one (a.s.). As a consequence the factorization (2.39) is essentially unique.

The proof of this theorem is in [14].

Remarks. Formula (2.37) is reminiscent of the well known recipe of Halmos-Savage [7] (see also Bahadur [2, Thm. 6.2]) for generating the minimal sufficient σ -algebra \mathcal{F}_0 . The interesting point for our purposes is that, by (2.38), the maximal unresolvable σ -algebra is induced by the same density m . Thus \mathcal{C}_0 can be determined directly from the original description of the inference problem (i.e. P and H_0) without computing the dual measure.

We now give an application of this result which will be used in the sequel.

Assume that $P(\cdot, \xi)$ is absolutely continuous with respect to some σ -finite measure W , for all $\xi \in U$. Let further the density $f := dP/dW$ have the structure

$$(2.40) \quad f(\eta, \xi) = q(\eta)g(\xi) \exp \sum_1^n c_i(\xi)s_i(\eta),$$

with real valued random variables q, g, c_i, s_i , $i = 1, \dots, n$; q and g positive P_0 and H_0 almost surely, respectively.³ If we compute m , starting from (2.40), we find

$$(2.41) \quad m(\eta, \xi) = \frac{dP(\cdot, \xi)/dW}{dP_0/dW} = r(\eta)g(\xi) \exp \sum_1^n c_i(\xi)s_i(\eta),$$

where

$$(2.42) \quad r(\eta) = 1 / \int_U g(\xi) \exp \sum_1^n c_i(\xi)s_i(\eta) H_0(d\xi).$$

Since q is positive r is also positive (P_0 a.s.). Notice that r depends on η only through s_1, \dots, s_n (Lehmann has shown that this dependence is also "smooth" [12, pp. 52–53]).

THEOREM 2.15. *Let f have the structure (2.40). Then if $(c_1, \dots, c_n, 1)$ and*

³ Densities of the form (2.40) are said to belong to an *exponential family*.

$(s_1, \dots, s_n, 1)$ are linearly independent systems of functions,

(s_1, \dots, s_n) form a minimal sufficient statistic for u ,

(c_1, \dots, c_n) form a maximal unresolvable statistic for y .

Proof. Define $c_0(\xi) = \log g(\xi)$, $s_0(\eta) = \log r(\eta)$, and rewrite $m(\eta, \xi)$ as

$$(2.43) \quad m(\eta, \xi) = \exp \left\{ \sum_1^n c_i(\xi) s_i(\eta) + c_0(\xi) + s_0(\eta) \right\}.$$

Let now $x_i = s_i(\eta)$, $i = 1, \dots, n$; then s_0 can be rewritten as $s_0 = \hat{s}(x_1, \dots, x_n)$ for a suitable \hat{s} . Consider now the mapping $\hat{m}: (x_1, \dots, x_n) \mapsto \exp \{ \sum_1^n c_i(\cdot) x_i + c_0(\cdot) + \hat{s}(x_1, \dots, x_n) \}$. If $(c_1, \dots, c_n, 1)$ are linearly independent the above defined \hat{m} is one to one. For $\hat{m}(x'_1, \dots, x'_n, \cdot) = \hat{m}(x''_1, \dots, x''_n, \cdot)$ is equivalent to

$$(2.44) \quad \sum_1^n c_i(\cdot)(x'_i - x''_i) + [\hat{s}(x'_1, \dots, x'_n) - \hat{s}(x''_1, \dots, x''_n)] = 0,$$

and since no nontrivial linear combination of the c_i can be a constant independent of ξ (like the term between square brackets), we have $x'_i = x''_i$, $i = 1, \dots, n$, as required.

By Theorem 2.14 (iii), (s_1, \dots, s_n) form a minimal sufficient statistic.

The dual result for (c_1, \dots, c_n) can be obtained following the same line of reasoning. Q.E.D.

This result generalizes a classical characterization of the minimal sufficient statistic for a distribution belonging to the exponential family, due to Dynkin [4, Thm. 3.a].

3. Applications to identifiability. The applications we shall deal with in this section will be concerned with particularly simple (but important) cases where the observation process y is of the type “signal plus white noise”.

Let Y be the space of continuous R^m -valued functions of time $\eta: I \rightarrow R^m$. (I is an interval, not necessarily bounded). If this space is equipped with the topology of uniform convergence on compact subsets of I , we obtain a well known separable Banach space

$$(3.1) \quad Y = C^m(I).$$

Let \mathcal{Y}_t be the σ -algebra of subsets of Y induced by the coordinate maps, $\pi_s: \eta \mapsto \eta(s)$, for all $s \leq t$. We define \mathcal{Y} as the union $\bigvee_{t \in I} \mathcal{Y}_t$. It is well known that \mathcal{Y} is generated by the open sets in the metric topology of Y .

Assume that the observation process $y = \{y(t)\}_{t \in I}$ is generated by an equation of the following type:

$$(3.2) \quad dy(t) = h(u, t) dt + dw(t), \quad t \in I,$$

where:

1. h is an R^m -valued function (in general random), jointly measurable in (t, ω) and adapted to a given increasing family of sub σ -algebras of \mathcal{A} , $(\mathcal{A}_t)_{t \in I}$. We

assume h to be square integrable over I , i.e.

$$(3.3) \quad \int_I |h(u, t)|^2 dt < \infty, \quad \text{a.s.}$$

2. w is a R^m valued separable standard Wiener process, with respect to the family (\mathcal{A}_t) and the measure \mathbf{P} .

3. u is the random parameter ranging over some nice topological manifold U (the relative Borel σ -algebra will be denoted by the symbol \mathcal{U}). We assume u to be independent of $(\mathcal{A}_t)_{t \in I}$.

We are interested in computing the conditional probability measure induced by y on (Y, \mathcal{Y}) , given u . According to the general notations introduced at the beginning of § 2 we shall denote its values by $P(A, \xi)$.

Since u is independent of (\mathcal{A}_t) , we can compute $P(\cdot, \xi)$ for each $\xi \in U$, simply by considering the measure induced on (Y, \mathcal{Y}) by the process

$$(3.4) \quad dy(t) = h(\xi, t) dt + dw(t), \quad t \in I,$$

corresponding to fixed value ξ in U .

Let W denote the Wiener measure on (Y, \mathcal{Y}) , then it is well known [8], [9] that, under (3.3),

$$(3.5) \quad P(\cdot, \xi) \text{ is absolutely continuous with respect to } W.$$

In order to give an explicit expression for the Radon–Nikodym derivative $f = dP(\cdot, \xi)/dW$ we need some supplementary assumptions.

LEMMA 3.1 [8], [9]. *If, in addition to the preceding assumptions, $h(\xi, t)$ is measurable with respect to \mathcal{Y}_t (the σ -algebra induced by $y(s)$, $s \leq t$) for each $t \in I$, and w is a Wiener process with respect to $(\mathcal{Y}_t)_{t \in I}$, then*

$$(3.6) \quad \frac{dP(\cdot, \xi)}{dW} = \exp \left\{ \int_I h(\xi, t)' d\eta(t) - \frac{1}{2} \int_I |h(\xi, t)|^2 dt \right\}.$$

We can obtain an important conclusion from this formula. Since $f(\cdot, \xi_1) = f(\cdot, \xi_2)$ a.s., iff $h(\xi_1, t) = h(\xi_2, t)$ a.s. for all $t \in I$, we deduce

PROPOSITION 3.2. *Under the same hypotheses as in Lemma 3.1, $\xi_1 \sim \xi_2$ if and only if $h(\xi_1, t) = h(\xi_2, t)$, a.s., for all $t \in I$.*

Linear dynamical systems. Suppose we observe a *linear dynamical system* over the interval $[0, T]$,

$$(3.7) \quad \dot{x}(t) = Fx(t) + Gv(t), \quad x(0) = x_0,$$

$$(3.8) \quad dy(t) = Hx(t) dt + dw(t).$$

The *input path* is assumed to be accessible to the experimenter, i.e. v is a known function of time with values in R^p which we assume bounded and measurable. The state $x(t) \in R^n$ is an n dimensional vector and $w(t)$ an m dimensional standard Wiener process. The matrices F, G, H , are of appropriate dimension.

A. Stochastic observability. Let H, F, G , be known matrices and suppose that the initial state x_0 is an unknown random vector, independent of w , which we would like to reconstruct from the noisy observations (3.8). The problem is to

determine the best possible reconstruction of x_0 allowed by the measurements.

The obvious identifications to be made are $U = R^n$, $\mathcal{U} = \mathcal{R}^n$ (the Borel σ -algebra on R^n), $u = x_0$. The function h can be easily written down as

$$(3.9) \quad h(\xi, t) = H e^{Ft} \xi + h_0(t), \quad 0 \leq t \leq T,$$

where h_0 is a zero state response, which does not depend on ξ . Since h is a nonrandom function we may apply Lemma 3.1, getting the following expression for $f(\eta, \xi)$:

$$(3.10) \quad f(\eta, \xi) = \exp \left\{ \int_0^T [H e^{Ft} \xi]' d\eta(t) + \int_0^T h_0'(t) d\eta(t) - \frac{1}{2} \int_0^T |h(\xi, t)|^2 dt \right\}.$$

We show that f is of the exponential family type. First, notice that only the first term between square brackets depends jointly on η and ξ .

By standard linear algebra we can write

$$(3.11) \quad \int_0^T [H e^{Ft} \xi]' d\eta(t) = \sum_0^{n-1} (HF^i \xi)' \int_0^T \alpha_i(t) d\eta(t)$$

where $\alpha_0, \dots, \alpha_{n-1}$ are linearly independent scalar functions of time. Let r'_1, \dots, r'_m, \dots be the row vectors forming the 1st, \dots , m th row of the matrices H, HF, \dots, HF^{n-1} , taken in that order, and let us define nm real functions c_k by putting

$$(3.12) \quad c_k: \xi \mapsto r'_k \xi, \quad k = 1, \dots, nm.$$

Define also

$$(3.13) \quad s_k: \eta \mapsto \int_0^T \alpha_i(t) d\eta_j(t), \quad k = (i+1)j = 1, \dots, nm,$$

so that the first member in (3.11) can be expressed as a sum, $\sum_1^{nm} c_k(\xi) s_k(\eta)$. With the obvious identifications, we see that f is precisely of the form (2.40).

We can check at once that the nm functions s_k are linearly independent (almost surely) on (Y, \mathcal{U}, W) . Hence, the system of functions (3.12) form a maximal identifiable statistic θ_0 .

Since everything is linear, Σ_0 can be written as $U/\ker \theta_0$ and $\xi_1 \sim \xi_2$ if and only if $\xi_1 - \xi_2$ belongs to the linear subspace, $\ker \theta_0$, of U . On the other hand θ_0 , as a mapping between the vector spaces R^n and R^{nm} , is represented by the *observability matrix*.

$$(3.14) \quad O = \begin{bmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{bmatrix};$$

thus the equivalence classes reduce to points in U , if and only if $\ker \theta_0 = \ker O$ is zero, in other words if and only if $\text{rank } O = n$. Let us agree to call *stochastically completely observable* the system (3.7), (3.8), if the equivalence classes reduce to points in U . Then we may summarize the above discussion in the following proposition.

PROPOSITION 3.3. *The maximal identifiable statistic for the initial state of the linear system described by (3.7), (3.8), is the linear mapping defined by the observability matrix O . The system is completely observable if and only if $\text{rank } O = n$.*

It is of some interest to give an interpretation of the above discussion from the point of view of *invariant theory* (see Popov [15] or Denham [3] for the basic concepts).

From the general definition (2.22) we see that θ_0 is always a *complete invariant* for the equivalence relation \sim , defined by (2.21).⁴ Here, using equivalent terminology we might say that the family of mappings c_k , $k = 1, \dots, nm$, defined by (3.12) is a *complete system of invariants* for our problem.

Recall that a set of invariants is said to be *independent* if “none of the functions of the family can be expressed as a function of the others and hence eliminated”. This rather loose definition can be made precise in the present linear setting by requiring *linearly independent* invariants. But, a trivial check shows that the c_k defined before do not meet this condition, in general. Even though in the present context, it is rather obvious how to extract an independent family out from the c_k ’s, the general situation is that a maximal identifiable statistic is defined by a highly “redundant” family of functions. This may happen even if θ_0 has linearly independent components, as we will see below.

B. Identifiability with known input. We consider again the linear system (3.7), (3.8), but we now assume $x_0 = 0$ and H, F, G are *unknown matrices*.

In the deterministic case (i.e., with no additive noise in 3.8), the maximum we can hope to reconstruct from observations of y and v is the transfer function of the system, that is, only the “completely reachable and completely observable” part.

For this reason, we assume from the beginning that our unknown parameter u ranges over the manifold U of completely reachable and completely observable triples (H, F, G) . This manifold can be partitioned into disjoint (finite dimensional) subsets, U_n , including only those triples (H, F, G) for which F has dimension $n \times n$. Let us denote by

$$(3.15) \quad M(\xi, t) = H e^{Ft} G,$$

the impulse response matrix of the linear system, corresponding to a particular value $\xi = (H, F, G)$ taken by the random vector u . We identify the function h in (3.4) with the output,

$$(3.16) \quad h(\xi, t) = \int_0^t M(\xi, t-s) v(s) ds,$$

and (since h is nonrandom) use formula (3.6) to obtain the density of $P(\cdot, \xi)$

⁴ Thus, by Definition 2.12 all identifiable statistics are (measurable) invariants.

with respect to the Wiener measure, i.e.

$$(3.17) \quad f(\eta, \xi) = \exp \left\{ \int_0^T \left(\int_0^t M(\xi, t-s)v(s) ds \right)' d\eta(t) - \frac{1}{2} \int_0^T \left| \int_0^t M(\xi, t-s)v(s) ds \right|^2 dt \right\}.$$

Define now

$$(3.18) \quad M_k: U \mapsto R^{m \times p}, \quad M_k(\xi) = HF^k G, \quad k = 0, 1, \dots,$$

(the *Markov parameters* relative to M) and

$$(3.19) \quad i_k(t) = \int_0^t \frac{(t-s)^k}{k!} v(s) ds, \quad k = 0, 1, \dots$$

With a little effort we may now check that the following expansion holds:

$$(3.20) \quad \int_0^T \left(\int_0^t M(\xi, t-s)v(s) ds \right)' d\eta(t) = \sum_{k=0}^{\infty} \left\langle M_k(\xi), \int_0^T i_k'(t) d\eta(t) \right\rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes inner product in $R^{m \times p}$ and the series at the second member is convergent in $L^2(Y, \mathcal{Y}, W)$, for all $\xi \in U$. We observe that the sequence M_k , $k = 0, 1, \dots$, form a linearly independent set in the real vector space $[R^{m \times p}]^U$, since no (finite) sum of the form

$$(3.21) \quad \sum \alpha_i HF^{k_i} G, \quad \alpha_i \in R,$$

can be the zero matrix for *all* triples $\xi = (H, F, G)$. Likewise, the real valued entries of the M_k 's form a linearly independent set and no linear combination of them can be constant independent of ξ .

If the p components of the input v are linearly independent over $[0, T]$, the functions i_k , defined by (3.19) can also be shown to be componentwise linearly independent. This implies linear independence of the family of functions $\{s_k, k = 0, 1, \dots\}$ defined as

$$(3.22) \quad s_k(\eta) = \int_0^T i_k'(t) d\eta(t).$$

Now, from (3.20), we can transform $f(\eta, \xi)$ to an exponential type density. By Theorem 2.15 we have

PROPOSITION 3.4. *If the input function v has linearly independent components over $[0, T]$, the maximal identifiable and minimal sufficient statistics for u are given, respectively, by*

the sequence of Markov parameters $M_k: \xi \mapsto HF^k G$, $k = 0, 1, \dots$;

the sequence of "empirical cross correlations" $s_k: \eta \mapsto \int_0^T i_k(t)' d\eta(t)$, $k = 0, 1, \dots$.

This result tells us that, in principle, we can reconstruct the transfer function exactly, as in the deterministic case. The equivalence classes $[\xi]$, in U are easily obtained.⁵ Take any point in the image of θ_0 , i.e. a matrix sequence

⁵ Of course (by Proposition 3.2) there is no need for computing θ_0 . We could just use (3.16).

$\sigma = \{\sigma_0, \sigma_1, \dots\}$, $\sigma_i \in R^{m \times p}$ and construct the Hankel matrix

$$(3.23) \quad \mathcal{H} = \begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_2 & \cdots \\ \sigma_1 & \sigma_2 & \cdots & \\ \sigma_2 & & & \\ \vdots & & & \end{bmatrix};$$

then \mathcal{H} must necessarily have finite rank, say n . This integer determines an element of the partition $\{U_n\}$ of U . In each U_n the equivalence classes are the orbits under the action of the general linear group $\text{Gl}(n)$; that is, if $\xi = (H_0 F_0, G_0) \in U_n$, then

$$(3.24) \quad [\xi] = \{H, F, G \mid H = H_0 T^{-1}, F = T F_0 T^{-1}, G = T G_0; T \in \text{Gl}(n)\}$$

These last considerations are standard; see e.g. [11].

Remark. If in (3.4) h is nonrandom then (y_t) is a *Gaussian process* with mean $\int_0^t h(\xi, s) ds$ and covariance $I(t \wedge s)$. In this case $P(\cdot, \xi)$ is uniquely determined by these functions (in the sense that measures corresponding to different means and/or covariances are different). Thus for a Gaussian process having mean $m(\xi, t)$ and covariance $\Sigma(\xi, t, s)$, $\xi_1 \sim \xi_2$ if and only if

$$(3.25a) \quad m(\xi_1, t) = m(\xi_2, t) \quad \text{for all } t \in I$$

$$(3.25b) \quad \Sigma(\xi_1, t, s) = \Sigma(\xi_2, t, s) \quad \text{for all } (t, s) \in I \times I.$$

The conclusions of examples *A* and *B* could also have been derived starting from these considerations.

We note another consequence of this last remark. For an observation scheme of the type

$$(3.26) \quad dy_t = h(H, F, G; t) dt + J dw_t,$$

where $\Sigma = JJ' > 0$ is the *unknown covariance* matrix of the additive noise, a maximal identifiable statistic is the pair $(H, F, G) \mapsto \{H F^k G, k = 0, 1, \dots\}$ and $J \mapsto JJ'$. Thus Σ is “identifiable”.

C. Identifiability of a stationary time series. Assume that we observe a Gaussian m dimensional stationary process (z_t) plus an additive white noise term,

$$(3.27) \quad dy_t = z_t dt + dw_t.$$

Let (z_t) admit a finite dimensional realization of the type

$$(3.28) \quad \begin{aligned} dx_t &= Fx_t + dv_t, \\ z_t &= Hx_t, \quad t \in R, \end{aligned}$$

where (v_t) is a p dimensional orthogonal increments process, in general correlated with (w_t) , with

$$(3.29) \quad E \begin{bmatrix} v_t \\ w_t \end{bmatrix} \begin{bmatrix} v_s' & w_s' \end{bmatrix} = \frac{1}{2} \begin{bmatrix} Q & S \\ S' & I \end{bmatrix} (|t| + |s| - |t - s|).$$

The identifiability problem is to determine how much of the “internal structure”

of z can be inferred from the observations (3.27). Let $\xi = (H, F, Q, S)$ and write z_t as $z(\xi, t)$. We may assume F strictly stable, $\Sigma := \int_0^\infty e^{Ft} Q e^{F't} dt$, positive definite and (H, F) completely observable (these are the so-called “globally minimal” realizations [1]).

Let us define $\hat{z}_t = E(z_t | \hat{\mathcal{Y}}_t)$ and

$$(3.30) \quad \nu_t = y_t - \int_0^t \hat{z}_s ds,$$

then ν_t is a standard Wiener process and we can rewrite (3.27) as

$$(3.31) \quad dy_t = \hat{z}(\xi, t) dt + d\nu_t,$$

with \hat{z} adapted to $(\hat{\mathcal{Y}}_t)$. From Proposition 3.2 the equivalence classes $[\xi]$ are determined as the subsets of all ξ' for which the conditional means $\hat{z}(\xi, t)$ and $\hat{z}(\xi', t)$ coincide a.s. over the observation interval I .

If we take $I = (-\infty, T)$ (i.e. we have been observing z from the “remote past”), then \hat{z}_t can be explicitly computed as the output of the “steady state” Kalman filter

$$(3.32) \quad d\hat{x}_t = F\hat{x}_t dt + K d\nu_t, \quad \hat{z}_t = H\hat{x}_t,$$

where $K = S + PH'$ and P is the positive solution of the algebraic Riccati equation:

$$(3.33) \quad FP + PF' - KK' + Q = 0.$$

Thus $\xi_1 = (H_1, F_1, Q_1, S_1)$ and $\xi_2 = (H_2, F_2, Q_2, S_2)$ are equivalent if and only if $H_1 F_1^k K_1 = H_2 F_2^k K_2$, $k = 0, 1, \dots$. (In particular two equivalent globally minimal parametrizations have the same dimension n .) By Lemma 2 in [1] we can get a more explicit description of the equivalence, namely,

PROPOSITION 3.5. $\xi_1 = (H_1, F_1, Q_1, S_1)$ and $\xi_2 = (H_2, F_2, Q_2, S_2)$ are equivalent parametrizations of (z_t) with respect to the observations (3.27) if and only if there exists $T \in \text{Gl}(n)$ and a symmetric $(n \times n)$ matrix Δ such that

$$(3.34) \quad \begin{aligned} H_1 &= H_2 T^{-1}, & F_1 &= T F_2 T^{-1}, \\ S_1 + \Delta H_1' &= T S_2, \\ F_1 \Delta + \Delta F_1' &= -Q_1 + T Q_2 T'. \end{aligned}$$

Notice that the solution is now significantly different from the “deterministic” case, in which we observe (z_t) directly. In the latter case, by using (3.25b), we can check that $\xi_1 \sim \xi_2$ if and only if (3.34) hold with $\Delta = 0$. This corresponds to a much better resolution.

REFERENCES

- [1] B. D. O. ANDERSON, *The inverse problem of stationary covariance generation*, J. Statist. Phys., 1 (1969), pp. 133–147.
- [2] R. R. BAHADUR, *Sufficiency and statistical decision functions*, Ann. Math. Statist., 25 (1954), pp. 423–462.
- [3] M. J. DENHAM, *Canonical forms for the identification of multivariable linear systems*, IEEE Trans. Automatic Control, AC-19 (1974), pp. 646–656.

- [4] E. B. DYNKIN, *Necessary and sufficient statistics for a family of probability distributions*, Selected Translations on Mathematical Statistics and Probability, vol. 1, American Mathematical Society, Providence, RI, 1961, pp. 23–40.
- [5] K. GLOVER, *Structural aspects of system identification*, Ph.D. thesis, Dept. of Electr. Engrg. Mass. Inst. of Tech., Cambridge, MA, 1973.
- [6] K. GLOVER AND J. C. WILLEMS, *Parametrizations of linear dynamical systems: Canonical forms and identifiability*, IEEE Trans. Automatic Control, AC-19 (1974), p. 640.
- [7] P. R. HALMOS AND L. J. SAVAGE, *Application of the Radon–Nikodym theorem to the theory of sufficient statistics*, Ann. Math. Statist., 20 (1949), pp. 225–241.
- [8] T. T. KADOTA AND L. A. SHEPP, *Conditions for absolute continuity between a certain pair of probability measures*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 16 (1970), pp. 250–260.
- [9] T. KAILATH AND M. ZAKAI, *Absolute continuity and Radon–Nikodym derivatives for certain measures relative to Wiener measure*, Ann. Math. Statist., 42 (1971), pp. 130–140.
- [10] G. KALLIANPUR AND C. STRIEBEL, *Estimation of stochastic processes: Arbitrary system process with additive noise observation error*, Ibid., 39 (1968), pp. 785–801.
- [11] R. E. KALMAN, P. L. FALB AND M. A. ARBIB, *Topics in Mathematical Systems Theory*, McGraw-Hill, New York, 1969.
- [12] E. L. LEHMANN, *Testing Statistical Hypotheses*, John Wiley, New York, 1959.
- [13] M. LOEVE, *Probability Theory*, Van Nostrand, Princeton, NJ, 1963.
- [14] G. PICCI, *Theory of sufficient statistics and structure analysis of transition probabilities with applications to identifiability*, Rep. C.N.R.-LADSEB 74/06, Padova, Italy, 1974.
- [15] V. M. POPOV, *Invariant description of linear time-invariant controllable systems*, SIAM J. Control, 10 (1972), pp. 252–264.
- [16] T. J. ROTHENBERG, *Identification in parametric models*, Econometrica, 39 (1971), pp. 577–591.
- [17] E. TSE AND J. ANTON, *On the identifiability of parameters*, IEEE Trans. Automatic Control, AC-17 (1972), pp. 637–646.
- [18] E. TSE AND H. WEINERT, *Correction and extension of “On the identifiability of parameters”*, Ibid., AC-18 (1973), p. 687.
- [19] ———, *Structure determination and parameter identification for multivariable stochastic linear systems*, Proc. J. A.C.C., (1973).
- [20] S. ZACKS, *The Theory of Statistical Inference*, John Wiley, New York, 1971.