

# A Geometric Approach to the Singular Filtering Problem

JOHANNES M. SCHUMACHER, MEMBER, IEEE

**Abstract**—We consider the least-squares filtering problem for a stationary Gaussian process when the observation is not fully corrupted by white noise, the so-called “singular” case. An optimal estimator is constructed consisting of an integrating part, which is, as in the regular case, computed from a spectral factorization or an equivalent matrix problem, and a differentiating part whose parameters are computed from a single matrix equation. This improves on older results which either work under restrictive assumptions, or describe the solution only as the result of some nested algorithm.

## I. INTRODUCTION

THIS paper is concerned with stationary least-squares estimation for linear Gaussian systems in continuous time, with particular emphasis on the singular case (partially noise-free observations). In contrast to previous works in which either the solution is only sketched as the result of some nested algorithm or restrictive assumptions are made, we present an explicit state-space solution for the general situation. An optimal estimator will be constructed that consists of an integrating part, which is computed from a spectral factorization or an equivalent matrix problem (as in the regular case), and a differentiating part whose parameters are computed from a single matrix equation. The construction leads to a unique result if and only if the spectral density matrix of the observed signal is full rank. Moreover, in the nonfull rank case the freedom of design can be described precisely.

The method that we use in the present paper is based on the observation that the problem of finding the optimal estimator is the same for state-space models that represent the same spectral density. We identify a number of invariants shared by all systems that are related in this way. Earlier results in this direction were obtained in the dual context of optimal control by Hautus and Silverman in [1], which has been a major source of inspiration for the present work. In contrast to [1], however, we do not use the “structure algorithm” but rather the “geometric approach” of [2] and [3], which, in this author’s view, leads to a more transparent picture.

The singular filtering problem occurs when all or part of the observations are modeled as being noise-free. As explained in [4], this will happen, for instance, when the actual observations are corrupted by colored noise. It is also made plausible in [4] that the solution should be to differentiate the observations (or linear combinations of these) until white noise appears, and to use a standard Kalman–Bucy filter for the estimation of those state variables that have not already been obtained in the differentiation process. What procedure one should follow exactly is suggested in [4] by the presentation of a few worked examples. Some explicit formulas are given in [5, pp. 352–356], but these soon become unwieldy as the differentiation proceeds step-by-step. A considerable part of the literature treats the problem under the simplifying

but (except for the single-output case) restrictive assumption that the number of differentiation steps needed in order to reach white noise is the same for all outputs (see, for instance, [6]); or the even stronger assumption is made, such as in [7] and [8], that white noise is obtained in all channels after one differentiation step. The author is unaware of any previous publication to give a completely explicit state-space solution to the singular filtering problem without imposing such restrictive conditions.

An outline of the paper is as follows. In Section II we define “signal model equivalence” (SME). This is followed by a quick review of some of the notions from the geometric theory that will be needed, along with some new material. Section IV is devoted to the uncovering of a number of invariants under SME; in the author’s opinion, these are crucial for a complete understanding of the singular filtering problem. The application to filtering is made in Section V, and the conclusions are summarized in the final Section VI.

## II. SIGNAL MODEL EQUIVALENCE (SME)

In this section, we mainly review some basic facts from linear mean-square estimation; see, for instance, [5] and [9]–[11]. Let  $y(t)$  be a stationary, zero-mean, Gaussian,  $\mathbb{R}^p$ -valued process with a rational spectral density, defined on  $[0, \infty)$ . Such a process can always be represented in the form

$$\dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0 \quad (2.1)$$

$$y(t) = Cx(t) + Du(t). \quad (2.2)$$

Here, the matrices  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times m}$ ,  $C \in \mathbb{R}^{p \times n}$ , and  $D \in \mathbb{R}^{p \times m}$  can be chosen to satisfy the following requirements:

- i) all eigenvalues of  $A$  are in the open left half plane;
- ii) the pair  $(A, B)$  is controllable;
- iii) the pair  $(C, A)$  is observable;
- iv) the matrix  $[B' \ D']'$  is full column rank.

To have a short phrase, systems  $\Sigma(A, B, C, D)$  that satisfy i)–iv) will henceforth be called *signal models*. Further properties satisfied by (2.1)–(2.2) are that the process  $u(t)$  is Gaussian white noise of unit covariance, and the stochastic vector  $x_0$  is normally distributed with zero mean and covariance matrix  $Q$  satisfying the Lyapunov equation

$$QA' + AQ + BB' = 0. \quad (2.3)$$

Moreover, the initial condition  $x_0$  and the white noise input  $u(t)$  are uncorrelated.

The representation of the given process  $y(t)$  by (2.1)–(2.2) is in the sense that the output process of (2.1)–(2.2) is equal a.s. for all  $t$  to the given process; this requires not only that the matrices  $A$ ,  $B$ ,  $C$ , and  $D$  are chosen correctly, but also that the white noise process  $u(t)$  is appropriately selected (see [12]).

Suppose now that we want to compute, for every  $t \geq 0$ , the conditional expectation of  $x(t)$  given the observations  $y(s)$ ,  $0 \leq s \leq t$ . The principle of orthogonality allows us to formulate this problem in terms of an integral equation

$$\int_0^t K(t; s) R_{yy}(s - \tau) ds = R_{xy}(t - \tau) \quad \forall \tau \in [0, t]. \quad (2.4)$$

Manuscript received December 12, 1983; revised March 27, 1985. Paper recommended by Past Associate Editor, A. S. Willsky. This work was supported by The Netherlands Organization for the Advancement of Pure Research.

The author is with the Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

Here,  $K(t; s)$  is the unknown integral kernel, and  $R_{yy}$  and  $R_{xy}$  are covariance functions given, with  $Q$  as in (2.3), by

$$R_{yy}(t) = H(t)Ce^{At}(QC' + BD') + \delta(t)DD' \\ + (1 - H(t))(CQ + DB')e^{A't}C' \quad (2.5)$$

where  $H(t)$  and  $\delta(t)$  denote the functions of Heaviside and Dirac, respectively, and

$$R_{xy}(t) = e^{At}(QC' + BD') \quad (t \geq 0). \quad (2.6)$$

The equation (2.4) is, of course, the starting point of Wiener's filtering theory, which, by the way, was effectively used also in the "singular" case (see [13, p. 66]). We will not work directly on the integral equation (2.4), but we will use it to draw one important conclusion, which is that the crucial parameters that determine the solution of the optimal filtering problem are not the system parameters  $A$ ,  $B$ ,  $C$ , and  $D$  plus the specific white noise process  $u(t)$  in (2.1), (2.2), which are needed to represent the signal process  $y(t)$  up to equality a.s. for all  $t$ , but only the parameters  $C$ ,  $A$ ,  $QC' + BD'$ , and  $DD'$ . To be certain, the conditional expectation process  $\hat{x}(t)$  will depend on the choice of  $u(t)$ , etc., however, we are interested in the *estimator*, i.e., the mapping which takes the observations  $\{y(s), 0 \leq s \leq t\}$  to the estimate  $\hat{x}(t)$ . For a discussion which puts great emphasis on the difference between estimates and estimators, see [14, pp. 28–29]. The fact that the optimal estimator can be found in terms of covariances is a basic property of Gaussian random variables and does not depend on the stationarity of the given processes; this is emphasized in [15] and [16].

So there is good reason to call two signal models  $\Sigma_1 (A, B_1, C, D_1)$  and  $\Sigma_2 (A, B_2, C, D_2)$  "equivalent" if

$$Q_1 C_1' + B_1 D_1' = Q_2 C_2' + B_2 D_2' \quad (2.7)$$

where  $Q_1$  and  $Q_2$  are each determined by the appropriate Lyapunov equation (2.3), and

$$D_1 D_1' = D_2 D_2'. \quad (2.8)$$

This type of equivalence will be called "signal model equivalence" (SME). It corresponds to the " $W$ -transformations" introduced by Popov [17] for the dual problem of optimal control. The equivalence can be characterized as follows.

**Proposition 2.1:** Two signal models  $\Sigma_1 (A, B_1, C, D_1)$  and  $\Sigma_2 (A, B_2, C, D_2)$  are related by SME if and only if there exists a symmetric matrix  $\bar{Q}$  such that

$$\begin{pmatrix} \bar{Q}A' + A\bar{Q} + B_1 B_1' & \bar{Q}C' + B_1 D_1' \\ C\bar{Q} + D_1 B_1' & D_1 D_1' \end{pmatrix} = \begin{pmatrix} B_2 \\ D_2 \end{pmatrix} \begin{pmatrix} B_2 \\ D_2 \end{pmatrix}'. \quad (2.9)$$

*Proof:* Let  $Q_i (i = 1, 2)$  denote the solution to the Lyapunov equation

$$Q_i A' + A Q_i + B_i B_i' = 0 \quad (2.10)$$

and write  $\bar{Q} = Q_1 - Q_2$ . If  $\Sigma_1$  and  $\Sigma_2$  are equivalent under SME, then it is easy to verify that (2.9) holds with  $\bar{Q} = \bar{Q}$ . On the other hand, if (2.9) holds, then we must have  $\bar{Q} = \bar{Q}$ , because both matrices satisfy (2.10), and this equation has a unique solution. The equivalence of  $\Sigma_1$  and  $\Sigma_2$  is now immediate from (2.9).

The dualized version of the matrix on the left-hand side of (2.9) has been termed the "dissipation matrix" in optimal control theory [18]. It is well known that if we extend SME by also allowing change of basis in state space, then the resulting equivalence is the one that relates two systems if their corresponding spectral densities are equal (see [19]).

Our strategy for solving the singular filtering problem will be to transform the given system under SME into a system for which it is fairly obvious what the optimal filter will be. This is basically also the strategy which is used in the standard methods for solving

the regular problem (transformation determined by the positive definite solution of the Riccati equation, or by the minimal spectral factor), although this is not always made explicit.

### III. PRELIMINARIES

For a system  $\Sigma (A, B, C, D)$  with state-space  $X = \mathbb{R}^n$ , input space  $U = \mathbb{R}^m$ , and output space  $Y = \mathbb{R}^p$ , consider the two sequences of subspaces that are defined as follows:

$$V^0(\Sigma) = X \quad (3.1)$$

$$V^{k+1}(\Sigma) = \{x \in X \mid \exists u \in U \text{ s.t. } Ax + \\ + Bu \in V^k(\Sigma) \text{ and } Cx + Du = 0\} \quad (k \geq 0) \quad (3.2)$$

$$T_0(\Sigma) = \{0\} \quad (3.3)$$

$$T^{k+1}(\Sigma) = \{x \in X \mid \exists w \in T^k(\Sigma), u \in U \text{ s.t. } Aw + Bu \\ = x \text{ and } Cw + Du = 0\} \quad (k \geq 0) \quad (3.4)$$

(cf. [20], [2], [3], [21]). The sequence  $\{V^k(\Sigma)\}$  is decreasing and will, after a finite number of steps, reach a limit which is denoted by  $V^*(\Sigma)$ . The sequence  $\{T^k(\Sigma)\}$  is increasing and converges to a subspace denoted by  $T^*(\Sigma)$ . The two limit subspaces can be characterized as follows (see [1], [22]).

**Proposition 3.1:** For a given system  $\Sigma (A, B, C, D)$ , consider the set of all subspaces  $T$  of  $X$  for which there exists a mapping  $G: Y \rightarrow X$  such that

$$(A + GC)T \subset T, \text{ Im } (B + GD) \subset T. \quad (3.5)$$

This set has a minimal element, which is equal to  $T^*(\Sigma)$ . Also, consider the set of all subspaces  $V$  for which there exists a mapping  $F: X \rightarrow U$  such that

$$(A + BF)V \subset V, V \subset \ker (C + DF). \quad (3.6)$$

This set has a maximal element, which is equal to  $V^*(\Sigma)$ .

To give another characterization of the sequences  $\{V^k(\Sigma)\}$  and  $\{T^k(\Sigma)\}$ , consider the following definitions from [23]. For a given system  $\Sigma (A, B, C, D)$ , write (for  $k \geq 1$ )

$$B_k = [A^{k-1}B \ \cdots \ AB \ B] \quad (3.7)$$

$$C_k = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{bmatrix}, D_k = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{k-2}B & \cdots & CB & D \end{bmatrix}. \quad (3.8)$$

The following proposition is due to Silverman [22, p. 356]. For a quick proof, see also [24]. Note that (3.10) is dual to (3.9).

**Proposition 3.2:** Let a system  $\Sigma (A, B, C, D)$  be given, and let  $k \geq 1$ . We then have

$$V^k(\Sigma) = C_k^{-1} [\text{Im } D_k] \quad (3.9)$$

$$T^k(\Sigma) = B_k [\ker D_k]. \quad (3.10)$$

This characterization can be used to derive results about membership of  $V^k(\Sigma) + T^l(\Sigma)$  in terms of certain rational equations. The  $d$ -dimensional space of real rational vectors will be denoted by  $\mathbb{R}^d(s)$ ; its submodule of proper rational vectors is written  $\mathbb{R}_+^d(s)$ .

**Proposition 3.3:** Let a system  $\Sigma (A, B, C, D)$  be given, and let  $G(s) = C(sI - A)^{-1}B + D$  be its transfer matrix. Then, for each  $k \geq 0$  and  $l \geq 0$ , an element  $x \in X$  belongs to  $V^k(\Sigma) + T^l(\Sigma)$  if and only if

$$s^{k+1}C(sI - A)^{-1}x \in s^{k+1}G(s)\mathbb{R}_+^m(s) + \mathbb{R}_+^p(s). \quad (3.11)$$

*Proof:* First assume that  $k \geq 1, l \geq 1$ . Expansion around

infinity shows that (3.11) is equivalent to

$$\begin{bmatrix} 0 \\ C_k \end{bmatrix} x \in \text{Im} \begin{bmatrix} D_l & 0 \\ C_k B_l & D_k \end{bmatrix}. \quad (3.12)$$

But this is true if and only if

$$C_k x \in C_k B_l [\ker D_l] + \text{Im } D_k \quad (3.13)$$

or in other words,

$$\begin{aligned} x &\in B_l [\ker D_l] + C_k^{-1} [\text{Im } D_k] \\ &= T^l(\Sigma) + V^k(\Sigma). \end{aligned} \quad (3.14)$$

For  $l = 0$ , (3.12) is replaced by

$$C_k \in \text{Im } D_k \quad (3.15)$$

which is equivalent to  $x \in V^k(\Sigma)$  as it should be, since  $T^0(\Sigma) = \{0\}$ . Finally, if  $k = 0$ , then (3.11) holds for all  $x$  and this is also exactly what we need, since  $V^0(\Sigma) = X$ .

*Corollary 3.4:* For a system  $\Sigma(A, B, C, D)$  with transfer matrix  $G(s)$ , one has  $x \in V^k(\Sigma) + T^*(\Sigma)$  if and only if

$$s^{k+1} C(sI - A)^{-1} x \in G(s)\mathbb{R}^m(s) + \mathbb{R}_+^p(s). \quad (3.16)$$

*Proof:* This is an immediate consequence of Proposition 3.3.

*Corollary 3.5:* For a system  $\Sigma(A, B, C, D)$  with transfer matrix  $G(s)$ , one has  $x \in V^*(\Sigma) + T^l(\Sigma)$  if and only if

$$C(sI - A)^{-1} x \in s^{l-1} G(s)\mathbb{R}_+^m(s). \quad (3.17)$$

*Proof:* The sufficiency of (3.17) is clear from Proposition 3.3. Now, suppose first that  $x \in V^*(\Sigma)$ . Take  $F: X \rightarrow U$  such that  $V^*(\Sigma)$  is  $(A + BF)$ -invariant, and  $V^*(\Sigma) \subset \ker(C + DF)$ . For such  $F$ , computation shows that

$$\begin{aligned} C(sI - A)^{-1} x &= G(s)F(sI - (A + BF))^{-1} x \\ &\in s^{-1} G(s)\mathbb{R}_+^m(s). \end{aligned} \quad (3.18)$$

Next, let  $x$  be an element of  $T^l(\Sigma)$ . By Proposition 3.2, there exists  $u^l = (u_1^l \cdots u_l^l)' \in \ker D_l$  such that  $B_l u^l = x$ . Define  $u(s) = u_1 s^{l-1} + \cdots + u_l$ . For every  $k \in \mathbb{N}$ , the first  $l + k$  coefficients in the Laurent expansion around infinity of  $G(s)u(s)$ , starting with the coefficient of  $s^{l-1}$ , are given by the vector

$$\begin{bmatrix} D_l & 0 \\ C_k B_l & D_k \end{bmatrix} \begin{bmatrix} u^l \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ C_k x \end{bmatrix}. \quad (3.19)$$

This shows that

$$C(sI - A)^{-1} x = G(s)u(s) \in s^{l-1} G(s)\mathbb{R}_+^m(s). \quad (3.20)$$

The proof is complete, since the set of all  $x \in X$  satisfying (3.17) is clearly a linear space.

A different proof of this result is outlined in [44] (Theorem 3.3). The  $l = 0$  case is older (see [45], [46], [25]).

*Corollary 3.6:* For a system  $\Sigma(A, B, C, D)$  with transfer matrix  $G(s)$ , one has  $x \in V^*(\Sigma) + T^*(\Sigma)$  if and only if

$$C(sI - A)^{-1} x \in G(s)\mathbb{R}^m(s). \quad (3.21)$$

*Proof:* This is immediate from Corollary 3.5.

We close this section by introducing some notation that will be needed below. The *degree*  $\pi(f)$  of an element  $f(s)$  of the rational vector space  $\mathbb{R}^d(s)$  is defined by

$$\pi(f) = \min \{k \in \mathbb{Z} \mid s^{-k} f(s) \in \mathbb{R}_+^d(s)\}. \quad (3.22)$$

On  $\mathbb{R}^d(s)$ , a Hermitian form (cf. [26, ch. XIV]) is defined by

$$\langle f, g \rangle = \sum_{i=1}^d f_i(s) g_i(-s) \quad (f(s), g(s) \in \mathbb{R}^d(s)). \quad (3.23)$$

For a rational matrix  $W(s)$ , define

$$W^*(s) = W'(-s). \quad (3.24)$$

Note that this implies

$$\langle Wf, g \rangle = \langle f, W^*g \rangle. \quad (3.25)$$

Two elements  $f(s)$  and  $g(s)$  of  $\mathbb{R}^d(s)$  are said to be *orthogonal* if  $\langle f, g \rangle = 0$ . If  $E$  is a subspace of  $\mathbb{R}^d(s)$ , we write

$$E^\perp = \{ \phi(s) \in \mathbb{R}^d(s) \mid \langle \phi, g \rangle = 0 \forall g \in E \}. \quad (3.26)$$

Using (3.25) and the fact that the Hermitian form (3.23) is definite (i.e.,  $\langle f, f \rangle = 0$  implies  $f = 0$ ), one easily proves that the following holds for any matrix  $W(s)$  over the field of rational functions:

$$(\text{Im } W^*)^\perp = \ker W. \quad (3.27)$$

Finally, we also note that

$$\pi(\langle f, f \rangle) = 2\pi(f) \quad (3.28)$$

for all  $f(s) \in \mathbb{R}^d(s)$ .

#### IV. IDENTIFICATION OF INVARIANTS

In this section, we will indicate some invariants under signal model equivalence. This will help to clarify the meaning of SME in a state-space framework, and it also provides information that is useful in the filtering context. It has already been noted in Section II that the spectral density  $\Phi(s) = G(s)G'(-s)$  of a system  $\Sigma(A, B, C, D)$  with transfer matrix  $G(s)$  is an invariant under SME. An easy consequence of this is the following (see [27, Lemma 2]).

*Proposition 4.1:* If two signal models  $\Sigma_1$  and  $\Sigma_2$  are related by SME, then

$$G_1(s)\mathbb{R}^{m_1}(s) = G_2(s)\mathbb{R}^{m_2}(s) \quad (4.1)$$

where  $G_i(s)$  denotes the transfer matrix of  $\Sigma_i$ , and  $m_i$  is the number of inputs of  $\Sigma_i$  ( $i = 1, 2$ ).

*Corollary 4.2:* If two signal models  $\Sigma_1$  and  $\Sigma_2$  are related by SME, then, for all  $k \geq 0$ ,

$$T^*(\Sigma_1) + V^k(\Sigma_1) = T^*(\Sigma_2) + V^k(\Sigma_2). \quad (4.2)$$

*Proof:* It follows from Corollary 3.4 that the subspaces  $T^*(\Sigma) + V^k(\Sigma)$  ( $k \geq 0$ ) are determined completely by the matrices  $C$  and  $A$  and by the image of the transfer matrix of  $\Sigma$ . Therefore, the statement is an immediate consequence of Proposition 4.1.

The following result is more powerful. We will use the Hermitian form that has been introduced at the end of the previous section.

*Theorem 4.3:* If two signal models  $\Sigma_1$  and  $\Sigma_2$  are related by SME, then

$$G_1(s)\mathbb{R}_+^{m_1}(s) = G_2(s)\mathbb{R}_+^{m_2}(s) \quad (4.3)$$

where  $G_i(s)$  denotes the transfer matrix of  $\Sigma_i$ , and  $m_i$  is the number of inputs of  $\Sigma_i$  ( $i = 1, 2$ ).

*Proof:* We shall show that for every  $u_1(s) \in \mathbb{R}_+^{m_1}(s)$  there exists a  $u_2(s) \in \mathbb{R}_+^{m_2}(s)$  such that

$$G_1(s)u_1(s) = G_2(s)u_2(s). \quad (4.4)$$

The results will then follow by symmetry. So, take  $u_1(s) \in \mathbb{R}_+^{m_1}(s)$ . Let  $\{G_1^* y_1, \dots, G_1^* y_k\}$  be an orthogonal basis for  $\text{Im}$

$G_1^*$ . (Such bases exist by [26, p. 358, p. 374].) Define

$$y = \sum_{i=1}^k \langle\langle G_1^* y_i, G_1^* y_i \rangle\rangle^{-1} \langle u_1, G_1^* y_i \rangle y_i. \quad (4.5)$$

For all  $j = 1, \dots, k$ , one has

$$\begin{aligned} \langle u_1 + G_1^* y, G_1^* y_j \rangle &= \\ &= \langle u_1, G_1^* y_j \rangle - \sum_{i=1}^k \langle\langle G_1^* y_i, G_1^* y_i \rangle\rangle^{-1} \langle u_1, G_1^* y_i \rangle \langle G_1^* y_i, G_1^* y_j \rangle \\ &= \langle u_1, G_1^* y_j \rangle - \langle u_1, G_1^* y_j \rangle = 0. \end{aligned} \quad (4.6)$$

Using (3.27), we conclude that

$$u_1 - G_1^* y \in \ker G_1. \quad (4.7)$$

In other words, we have

$$G_1 u_1 = G_1 G_1^* y. \quad (4.8)$$

Now, define  $u_2$  by

$$u_2 = G_2^* y = \sum_{i=1}^k \langle\langle G_1^* y_i, G_1^* y_i \rangle\rangle^{-1} \langle u_1, G_1^* y_i \rangle G_2^* y_i. \quad (4.9)$$

Since  $G_1 G_1^* = G_2 G_2^*$ , we get

$$G_2 u_2 = G_2 G_2^* y = G_1 G_1^* y = G_1 u_1. \quad (4.10)$$

It remains to prove that  $u_2$  is proper. Note that

$$\langle G_1^* y_i, G_1^* y_i \rangle = \langle G_2^* y_i, G_2^* y_i \rangle \quad (4.11)$$

for all  $i = 1, \dots, k$ , so that [see (3.28)]

$$\pi(G_1^* y_i) = \pi(G_2^* y_i) \quad (i=1, \dots, k). \quad (4.12)$$

Using this, we can derive an upper bound for the degree of each term in (4.9)

$$\begin{aligned} \pi(\langle\langle G_1^* y_i, G_1^* y_i \rangle\rangle^{-1} \langle u_1, G_1^* y_i \rangle G_2^* y_i) \\ &= \pi(\langle u_1, G_1^* y_i \rangle) + \pi(G_2^* y_i) - 2\pi(G_1^* y_i) \\ &\leq \pi(u_1) \leq 0. \end{aligned} \quad (4.13)$$

We see that  $u_2$  is a sum of proper terms; consequently,  $u_2$  is proper itself.

**Corollary 4.4:** If two signal models  $\Sigma_1$  and  $\Sigma_2$  are related by SME, then

$$V^k(\Sigma_1) + T^l(\Sigma_1) = V^k(\Sigma_2) + T^l(\Sigma_2) \quad (4.14)$$

for all  $k \geq 0$  and  $l \geq 0$ .

*Proof:* This is immediate from Proposition 3.3.

Taking  $l = 0$  and  $k$  large in Corollary 4.4, we see in particular that  $V^*(\Sigma)$  is an invariant under SME. This has been proved earlier by Hautus and Silverman in [1] (Lemma 6.21), who used techniques from functional analysis for the proof. Another immediate consequence of the theorem is the following.

**Corollary 4.5:** Let  $\Sigma_1$  and  $\Sigma_2$  be related by SME. Then the transfer matrices of  $\Sigma_1$  and  $\Sigma_2$  have the same zero structure at infinity.

*Proof:* It has been shown in [28] (see also [2], [29]) that the number of zeros at infinity and their orders can be determined completely from the numbers  $\dim(V^*(\Sigma) + T^l(\Sigma))$ ,  $l \geq 0$ . Therefore, the statement is clear from Corollary 4.4.

Recall the definition of the matrix  $D_k$  in (3.8). If  $u^k = (u_0', \dots, u_{k-1}')'$  gives the first  $k$  terms in the Taylor series development around infinity of  $u(s) \in \mathbb{R}_+^m(s)$ , then  $D_k u^k$  gives

the first  $k$  terms in the corresponding series for  $y(s) = G(s)u(s)$ . In conjunction with Theorem 4.3, this observation yields the following.

**Corollary 4.6:** If  $\Sigma_1$  and  $\Sigma_2$  are related by SME, then one has

$$\text{Im } D_k(\Sigma_1) = \text{Im } D_k(\Sigma_2) \quad (4.15)$$

for all  $k \geq 1$ .

**Remark 1:** One can consider a discrete-time version of SME, in which two signal models are called equivalent if they have the same spectral density

$$\Phi(z) = G(z)G' \left( \frac{1}{z} \right). \quad (4.16)$$

It is known how to formulate this equivalence in state space [14, ch. 9]. It should be emphasized that Theorem 4.3 and its corollaries are *not true* for the equivalence induced by (4.16), as can be seen from simple examples. (Proposition 4.1 will still hold, however, with essentially the same proof.) The reason is basically that the point at infinity is fixed under the transformation  $s \rightarrow -s$  but not under  $z \rightarrow z^{-1}$ . In the author's opinion, this makes the continuous-time singular filtering problem fundamentally different from its discrete-time version; in fact, the vanishing of observation noise presents no essential difficulties in the discrete-time case. This point of view, although recently disputed [30], [31], is by no means new (see [32]).

**Remark 2:** Also invariant under SME are the zeros of the transfer matrix on the imaginary axis. This is easily shown because  $G'(-s)$  is the Hermitian adjoint of  $G(s)$  for  $s = i\omega$ ,  $\omega \in \mathbb{R}$  fixed. When they are present, purely imaginary zeros cause serious difficulties for filtering.

## V. CONSTRUCTION OF THE OPTIMAL FILTER

In this section, we shall describe how to compute an optimal filter for a system of the form (2.1), (2.2) whether one has full observation noise or not. The idea is to transform the given signal model under SME into a form for which the optimal estimator becomes easy to write down. This form will be that of a model having a stable left inverse. First, we have to discuss the concept of left invertibility in state-space terms. A well-known result from [33] says that the transfer matrix of a system  $\Sigma(A, B, C, D)$  is left invertible (as a matrix over the field of rational functions) if and only if the matrix  $[B' \ D']'$  has full column rank and  $T^*(\Sigma) \cap V^*(\Sigma) = \{0\}$ . To formulate a criterion for the transfer matrix to have a stable left inverse, let  $T_g^*(\Sigma)$  denote the minimal element in the set of all subspaces  $T$  of  $X$  for which there exists a mapping  $G: Y \rightarrow X$  such that the eigenvalues of  $A + GC$  are in the closed left half plane, and (3.5) holds. The existence of such a minimal element can be proved (constructively) by combining the arguments in [3, p. 114] and [21]. It is clear from Proposition 3.1 that we always have  $T_g^*(\Sigma) \supset T^*(\Sigma)$ . The following proposition was proved (in dual form) in [34]; see also [35].

**Proposition 5.1:** The transfer matrix of a system  $\Sigma(A, B, C, D)$  has a left inverse having all its poles in the closed left half plane if and only if the matrix  $[B' \ D']'$  has full column rank and

$$T_g^*(\Sigma) \cap V^*(\Sigma) = \{0\}. \quad (5.1)$$

We can now formulate the basic result that will be needed.

**Theorem 5.2:** Let  $\Sigma(A, B, C, D)$  be a signal model. Then there exists a unique model  $\tilde{\Sigma}(A, \tilde{B}, C, \tilde{D})$  which is equivalent to  $\Sigma$  under SME, and which satisfies  $T_g^*(\tilde{\Sigma}) \cap V^*(\tilde{\Sigma}) = \{0\}$ .

*Proof:* This is the dual of [27, Theorem 4] (which is, in its turn, a version of Youla's classical theorem on spectral factorization [11], with a proof that uses ideas in [18] and is based essentially on calculations in [34]).

The condition  $T_g^* \cap V^* = \{0\}$  means that a vector  $x \in X$  is determined completely if it is known modulo  $T_g^*$  and modulo  $V^*$ .

Therefore, this condition suggests that the optimal estimator should consist of an estimation modulo  $T_g^*$  and an estimation modulo  $V^*$ . This is what we shall do, be it that a certain redundancy may be present which allows us to use subspaces that are even larger than  $T_g^*$  or  $V^*$ , but still intersect only at zero. The precise results are given in the next two theorems.

**Theorem 5.3:** Let  $\Sigma(A, B, C, D)$  be a signal model with state space  $X = \mathbb{R}^n$  and output space  $Y = \mathbb{R}^p$ , and let  $H \in \mathbb{R}^{l \times n}$  be such that  $\ker H \supset V^k(\Sigma)$ . Then there exist matrices  $R_i \in \mathbb{R}^{l \times p}$  ( $i = 1, \dots, k$ ) such that

$$[R_k \cdots R_1][C_k \ D_k] = [H \ 0] \quad (5.2)$$

[note the order of the  $R_i$ 's;  $C_k$  and  $D_k$  are as in (3.8)]. If we define a series of processes  $z_j(t)$  ( $j = 1, \dots, k$ ) by

$$z_1(t) \stackrel{\text{def}}{=} R_1 y(t) \quad (5.3)$$

$$z_{j+1}(t) \stackrel{\text{def}}{=} \frac{d}{dt} z_j(t) + R_{j+1} y(t) \quad (j = 1, \dots, k-1) \quad (5.4)$$

then none of the processes  $z_j(t)$  has a white noise component, so that the differentiation is justified for each step, and moreover

$$Hx(t) = z_k(t) \quad (5.5)$$

a.s. for all  $t$ .

**Theorem 5.4:** Let  $\Sigma(A, B, C, D)$  be a signal model, with state space  $X = \mathbb{R}^n$  and output space  $Y = \mathbb{R}^p$ , and let  $T$  be a subspace of  $X$  for which there exists a mapping  $G: Y \rightarrow X$  such that the eigenvalues of  $A + GC$  are in the open left half plane, and (3.5) holds. Decompose  $X = X_1 \oplus X_2$  with  $X_2 = T$ , and write the differential equations accordingly

$$\dot{x}_1(t) = A_{11}x_1(t) + A_{12}x_2(t) + B_1u(t) \quad (5.6)$$

$$\dot{x}_2(t) = A_{21}x_1(t) + A_{22}x_2(t) + B_2u(t) \quad (5.7)$$

$$y(t) = C_1x_1(t) + C_2x_2(t) + Du(t). \quad (5.8)$$

Then there exists a mapping  $G_1: Y \rightarrow X_1$  such that

$$A_{12} + G_1C_2 = 0, \quad B_1 + G_1D = 0 \quad (5.9)$$

and the eigenvalues of  $A_{11} + G_1C_1$  are in the open left half plane. Moreover, the equation

$$\dot{w}(t) = (A_{11} + G_1C_1)w(t) - G_1y(t), \quad w(0) = x_1(0) \quad (5.10)$$

defines a stationary process  $w(t)$  having the property

$$x_1(t) = w(t) \quad (5.11)$$

a.s. for all  $t$ .

Before going into the proofs of the theorems, let us note how these results can be used to set up an optimal estimator. First of all, we have to remark that what is needed in Theorems 5.3 and 5.4 is a little more than what is provided by Theorem 5.2. While we had stability with respect to the closed left half plane in Theorem 5.2, we require stability with respect to the open left half plane in Theorem 5.4. It follows from the dual version of the result of [36] that  $T_g^*(\Sigma)$  does not qualify as a subspace  $T$  in Theorem 5.4 if and only if the transfer matrix of  $\Sigma$  has one or more zeros on the imaginary axis. Transformation under SME will not help in such a situation, since these zeros are invariant (see the remark at the end of the previous section). The problem that is encountered here is not specific for singular filtering or for continuous-time systems, and a suboptimal solution is suggested in [14, p. 253]. We will bypass the issue and proceed under the assumption that the given model (3.1), (3.2) does not have zeros on the imaginary axis.

Under this assumption, we can apply Theorems 5.3 and 5.4 to

solve the filtering problem. First, one transforms the given model into one for which  $T_g^*(\Sigma) \cap V^*(\Sigma) = \{0\}$ . This can be done either by a spectral factorization or by solving a certain matrix equation (cf. [27]). Now, take  $k \geq 0$  and a subspace  $T \subset X$  satisfying the conditions of Theorem 5.4 such that  $T \cap V^k(\Sigma) = \{0\}$ . It may be necessary to take  $T = T_g^*(\Sigma)$  and  $k$  such that  $V^k(\Sigma) = V^*(\Sigma)$ , or it may be possible to let  $T$  and/or  $V^k(\Sigma)$  be larger than this. Decompose  $X = X_1 \oplus X_2$ , with  $X_2 = T$  and  $X_1 \supset V^k(\Sigma)$ . Let  $H$  be the projection along  $X_1$  onto  $X_2$ ; then  $x_2(t) = Hx(t)$ , and  $\ker H \supset V^k(\Sigma)$ . From Theorem 5.3, we get that  $x_2(t)$  can be recovered by applying the differentiation scheme (5.3) and (5.4), the parameters for which are found by solving the matrix equation (5.2). From Theorem 5.4 it is seen that  $x_1(t)$  is obtained by integration; the parameter  $G_1$  can be found by the procedure described in [3, p. 88], which comes down to solving a linear matrix equation. So we have constructed an estimator which in fact reproduces the state of the (transformed) model exactly, up to sets of zero measure, so that it certainly qualifies as a least-squares estimator for this model. The first property is of course not preserved under SME but the second property is, so that we have indeed constructed a least-squares estimator for the original model. The procedure we have sketched is just intended to give the basic idea of the method; it is not optimized with respect to the amount of computation. Now, we proceed to prove the two theorems.

*Proof of Theorem 5.3:* The matrix equation (5.2) is solvable if and only if

$$\ker [C_k \ D_k] \subset \ker [H \ 0]. \quad (5.12)$$

This is equivalent to

$$C_k^{-1}[\text{Im } D_k] \subset H \quad (5.13)$$

which is true by assumption (cf. Proposition 3.2). Next, we use the definition (3.8) to write out (5.2) explicitly

$$\sum_{i=1}^k R_i C A^{k-i} = H \quad (5.14)$$

$$\sum_{i=1}^{j-1} R_i C A^{j-i-1} B + R_j D = 0 \quad (j = 1 \cdots k). \quad (5.15)$$

For the processes  $z_j(t)$  defined by (5.3), (5.4) we shall show that

$$z_j(t) = \sum_{i=1}^j R_i C A^{j-i} x(t) \quad (5.16)$$

a.s. for all  $t$ . The proof of (5.16) is by induction. For  $j = 1$ , we have from (5.3)

$$z_1(t) = R_1 y(t) = R_1 (Cx(t) + Du(t)) = R_1 Cx(t) \quad (5.17)$$

where the last equality holds because  $R_1 D = 0$  [see (5.15)]. Now assume that (5.16) holds for some  $j \leq k-1$ . Then  $z_j(t)$  does not contain a white noise component, and we can write

$$\begin{aligned} z_{j+1}(t) &= \sum_{i=1}^j R_i C A^{j-i} (Ax(t) + Bu(t)) + R_{j+1} (Cx(t) + Du(t)) \\ &= \sum_{i=1}^{j+1} R_i C A^{j+1-i} x(t) + \left( \sum_{i=1}^j R_i C A^{j-i} B + R_{j+1} D \right) u(t) \\ &= \sum_{i=1}^{j+1} R_i C A^{j+1-i} x(t) \end{aligned} \quad (5.18)$$

according to (5.15). This proves our claim. We see that, indeed,

all processes  $z_j(t)$  ( $j = 1, \dots, k$ ) do not contain white noise components, and furthermore the equality (5.5) is immediate from (5.14) and (5.16).

*Proof of Theorem 5.4:* Let  $G: Y \rightarrow X$  be a mapping as described in the theorem statement. When  $X$  is decomposed as  $X = X_1 \oplus X_2$ ,  $G$  is decomposed as  $G = [G_1' \ G_2']'$ . It then follows from (3.5) that  $G_1$  has the required properties. Now, we rewrite (5.6) using (5.9) and (5.8)

$$\dot{x}_1(t) = A_{11}x_1(t) - G_1C_2x_2(t) - G_1Du(t) = (A_{11} + G_1C_1)x(t) + G_1y(t). \quad (5.19)$$

We see that the process  $x_1(t)$  and  $w(t)$  satisfy the same differential equation (in particular, the driving process is the same); by definition [see (5.10)], they also have the same initial condition. Since the eigenvalues of  $A_{11} + G_1C_1$  are in the open left half plane, the solutions of (5.10) and (5.19) are well-defined stationary processes which must be equal a.s. for all  $t$ .

*Remark 1:* The solvability of the matrix equation (5.2) under the condition  $\ker H \supset V^k(\Sigma)$  was proved earlier in [37] for the case  $H = I$  (so  $V^k(\Sigma) = \{0\}$ ). The idea of obtaining information about the state modulo  $V^k(\Sigma)$ , irrespective of the behavior or the disturbance  $u(t)$ , is in fact one of the earliest results in the geometric approach to linear systems: see [38], where, however, the problem is treated in a deterministic context and no connection is made with signal model equivalence or with left invertibility. Of course, our construction of an optimal estimator is intimately related to the construction of a left inverse. The version presented here seems to be new in that it uses a differentiating and an integrating part in parallel rather than in series as in [23] and [39] or in a hybrid form as in [37].

*Remark 2:* The equation (5.2) is equivalent to the two conditions below

$$[R_k \ \cdots \ R_1]C_k = H \quad (5.20)$$

$$\text{Im } D_k \subset \ker [R_k \ \cdots \ R_1]. \quad (5.21)$$

Therefore, we see from Corollary 4.6 that the set of solutions of (5.2) is invariant under SME. This means that a given model (2.1)–(2.2) does not have to be transformed under SME in order to compute the differentiating scheme (5.3)–(5.4). Note that this provides a justification for the method of “differentiating until you get white noise” proposed in [4].

In general, there is no guarantee that the parameters which determine the optimal estimator as constructed above (the matrices  $R_1, \dots, R_k$  of Theorem 5.3, and the matrix  $G_1$  of Theorem 5.4) are determined uniquely. However, we do get a uniqueness result in an important special case.

*Theorem 5.5:* Let a signal model  $\Sigma$  be given by (2.1), (2.2) and let  $\bar{\Sigma}$  be an equivalent signal model such that  $T_g^*(\bar{\Sigma}) \cap V^k(\bar{\Sigma}) = \{0\}$ . If the spectral density matrix  $\Phi(s) = G(s)G'(-s)$  has full rank, then:

- i) if  $T \subset X$  is a subspace for which the conditions of Theorem 5.4 are satisfied and if  $k$  is a nonnegative integer, then  $T \cap V^k(\bar{\Sigma}) = \{0\}$  implies  $T = T_g^*(\bar{\Sigma})$ ,  $V^k(\bar{\Sigma}) = V^*(\bar{\Sigma})$ ;
- ii) there is only one decomposition  $X = X_1 \oplus X_2$  with  $X_2 = T_g^*(\bar{\Sigma})$ ,  $X_1 \supset V^*(\bar{\Sigma})$ ;
- iii) for any  $H$  such that  $\ker H \supset V^k(\Sigma)$ , the equation (5.2) has only one solution; and
- iv) there is only one matrix  $G_1$  that satisfies (5.9), where the parameters are taken with respect to  $\bar{\Sigma}$  and the decomposition is as in ii).

Moreover, the matrix  $G_1$  in iv) will automatically be such that the eigenvalues of  $A_{11} + G_1C_1$  are in the closed left half plane (open left half plane, if the transfer matrix of  $\Sigma$  has no zeros on the imaginary axis).

*Proof:* First, we note that the spectral density matrix is full rank if and only if the transfer matrix is surjective. This follows from the fact that the rank of  $\Phi(i\omega)$  (as a matrix over  $\mathbb{C}$ ) is equal to

the row rank of  $G(i\omega)$  (over  $\mathbb{C}$ ), for all  $\omega \in \mathbb{R}$ . Moreover, it is known that the transfer matrix of a system  $\Sigma$  ( $A, B, C, D$ ) is right invertible if and only if the matrix  $[C \ D]$  is surjective and  $T^*(\Sigma) + V^*(\Sigma) = X$ . This is, in fact, just the dual of the result from [33] mentioned at the beginning of this section.

Now, let a signal model  $\Sigma$  be given and let  $\bar{\Sigma}$  be an equivalent model as described in the theorem statement. It follows from the arguments above (and also from Corollary 4.2) that  $T^*(\bar{\Sigma}) + V^*(\bar{\Sigma}) = T^*(\Sigma) + V^*(\Sigma) = X$ . So if  $T \subset X$  is a subspace satisfying the conditions of Theorem 5.4 and  $k$  is a nonnegative integer such that  $T \cap V^k(\bar{\Sigma}) = \{0\}$ , then we must have  $T = T^*(\bar{\Sigma})$  and  $V^k(\bar{\Sigma}) = V^*(\bar{\Sigma})$  since it is always true that  $T \supset T^*(\bar{\Sigma})$  and  $V^k(\bar{\Sigma}) \supset V^*(\bar{\Sigma})$ . This proves i). To prove ii), note that it follows from the above argument that  $T_g^*(\bar{\Sigma}) = T^*(\bar{\Sigma})$ . Therefore, we have  $T_g^*(\bar{\Sigma}) \oplus V^*(\bar{\Sigma}) = X$  and the decomposition of the state space is indeed unique, with  $X_1 = V^*(\bar{\Sigma})$  and  $X_2 = T_g^*(\bar{\Sigma})$ . The claim iv) is basically a standard result in the geometric theory of linear systems: see [3, p. 109], use [2, Lemma 1.1], use also [21], and dualize. The fact that the eigenvalues of  $A_{11} + G_1C_1$ , where  $G_1$  is the unique solution of (5.9), are automatically in the closed (or open) left half plane also follows essentially from [3, p. 109].

It remains to prove iii). One easily sees that the equation (5.2) has a unique solution if and only if the matrix  $[C_k \ D_k]$  is surjective. To show that this is the case, we use Theorem 2 in [23] which says that the transfer matrix  $G(s)$  of a system  $\Sigma$  ( $A, B, C, D$ ) with  $p$  outputs is right invertible if and only if

$$\text{rank } D_{l+1}(\Sigma) - \text{rank } D_l(\Sigma) = p \quad (5.22)$$

for all sufficiently large  $l$ . We apply this to the system  $\Sigma$  ( $A^k, B_k, C_k, D_k$ ) which we shall denote by  $\Sigma_k$ . It is readily checked that for all  $l \geq 1$  and  $k \geq 1$ , one has

$$D_l(\Sigma_k) = D_{lk}(\Sigma). \quad (5.23)$$

So if  $\Sigma$  has a right invertible transfer matrix, then, for all sufficiently large  $l$ ,

$$\begin{aligned} & \text{rank } (D_{l+1}(\Sigma_k)) - \text{rank } (D_l(\Sigma_k)) \\ &= \text{rank } (D_{kl+k}(\Sigma)) - \text{rank } (D_{kl}(\Sigma)) \\ &= \sum_{i=1}^k \text{rank } (D_{kl+i}(\Sigma)) - \text{rank } (D_{kl+i-1}(\Sigma)) = kp \end{aligned} \quad (5.24)$$

which shows that  $\Sigma_k$  is right invertible, since this system has  $kp$  outputs. In particular, it follows that  $[C_k \ D_k]$  is surjective for all  $k \geq 1$ . The proof is complete.

*Remark 3:* Suppose now that the spectral density of the given model  $\Sigma$  does not have full rank. In the transformed model  $\bar{\Sigma}$  one will have  $T_g^*(\bar{\Sigma}) \cap V^*(\bar{\Sigma}) = \{0\}$  but  $T_g^*(\bar{\Sigma}) + V^*(\bar{\Sigma}) \neq X$ . Consequently, there will be many ways to decompose the state space  $X = X_1 \oplus X_2$  where  $X_2 = T_g^*(\bar{\Sigma})$  and  $X_1 \supset V^*(\bar{\Sigma})$ . Also, there will be many subspaces  $T$  larger than  $T_g^*(\bar{\Sigma})$  that satisfy the conditions of Theorem 5.4 and that are such that  $T \cap V^*(\bar{\Sigma}) = \{0\}$ . It may also be possible to replace  $V^*(\bar{\Sigma})$  by a larger member of the sequence  $\{V^k(\bar{\Sigma}) \mid k = 0, 1, 2, \dots\}$ . In general, there is a tradeoff between differentiation and integration: taking  $T$  larger than  $T_g^*(\bar{\Sigma})$  means less differentiation steps, but one has to make sure that  $T \cap V^k(\bar{\Sigma}) = \{0\}$ . For a fixed decomposition, the nonuniqueness in solving (5.2) is described by the standard methods of linear algebra, whereas the nonuniqueness of the integration gain  $G_1$  (pole placement) is described in [3, pp. 111–112].

Of course, the fact that we get a uniqueness result for the case of a full rank spectral density matrix is not entirely surprising. After all, we have seen that the spectral density matrix is full rank if and only if the transfer matrix is right invertible, and this is the necessary and sufficient condition for the left inverse (if it exists) to be unique.

## VI. CONCLUSIONS

The purpose of this paper has been to utilize the geometric theory of linear systems for a better understanding of the singular filtering problem. It turns out that this framework enables us to treat the "singular" and the "regular" problem much at the same level. The conclusions can be summarized as follows. If the spectral density matrix for a given signal is full rank at infinity, then we have the "regular" case and the solution takes the form of the traditional Kalman-Bucy filter with an integration gain that can be determined after an algebraic Riccati equation has been solved (or, equivalently, a spectral factorization has been done). If the spectral density is full rank as a rational matrix but has a singularity at infinity, then we are in the full rank singular case and the optimal estimator takes the form of a differentiator and an integrator in parallel. The differentiation parameters are found by solving a single matrix equation, whereas the integration gain is found, as in the regular case, either by doing a spectral factorization or by solving a problem stated in terms of constant matrices (see [27]). If, finally, the spectral density is not even full rank as a rational matrix (this happens, for instance, always when the number of outputs in the model (2.1), (2.2) exceeds the number of white noise inputs), then the same structure of the optimal estimator is obtained but the uniqueness of the construction breaks down. We are still, however, able to describe exactly what the alternative forms for the optimal estimator are, and what freedom one has in selecting the differentiation parameters and the integration gain.

In this paper, we only discussed basic principles, and we have not addressed the issue of efficient and reliable numerical procedures to obtain the optimal estimator. Also, we did not discuss the relation with the "nearly singular" filtering problem (see, e.g., [40] and [41]), although it is expected that the approach of this paper will be helpful in this case too. The theory here can be extended to estimation of linear functions of the state; this would bring us even closer to Wiener filtering, and it would also enable us to recover the Bucy filter [42] for problems with full, but colored, observation noise. Finally, another connection that deserves further study is the one with Wiener-Hopf integral equations (cf. [43]).

## ACKNOWLEDGMENT

The work leading to this paper was begun while the author stayed at the Laboratory for Information and Decision Systems, M.I.T. The author would like to take the opportunity to express his gratitude to the staff and personnel of LIDS for their interest and helpfulness. In particular, it is a pleasure to thank Prof. S. K. Mitter for his gracious hospitality and his stimulation, especially in connection with the subject of the present paper. The author also profited from discussions with Dr. J. H. van Schuppen (CMCS, Amsterdam).

## REFERENCES

- [1] M. L. J. Hautus and L. M. Silverman, "System structure and singular control," *Linear Algebra Appl.*, vol. 50, pp. 369-402, 1983.
- [2] A. S. Morse, "Structural invariants of linear multivariable systems," *SIAM J. Contr.*, vol. 11, pp. 446-465, 1973.
- [3] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach*. New York: Springer-Verlag, 1979.
- [4] A. E. Bryson, Jr. and D. E. Johansen, "Linear filtering for time-varying systems using measurements containing colored noise," *IEEE Trans. Automat. Contr.*, vol. AC-10, pp. 4-10, 1965.
- [5] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*. New York: Wiley, 1972.
- [6] T. Kailath and R. A. Geesey, "An innovations approach to least-squares estimation—Part V: Innovations representations and recursive estimation in colored noise," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 435-453, 1973.
- [7] E. Tse and M. Athans, "Observer theory for continuous-time linear systems," *Inform. Contr.*, vol. 22, pp. 405-434, 1973.
- [8] J. J. Deyst, "A derivation of the optimum continuous linear estimator for systems with correlated measurement noise," *AIAA J.*, vol. 7, pp. 2116-2119, 1969.
- [9] P. Faurre, M. Clerget, and F. Germain, *Opérateurs Rationnels Positifs*. Paris, France: Dunod, 1978.
- [10] B. D. O. Anderson, "The inverse problem of stationary covariance generation," *J. Statist. Phys.*, vol. 1, pp. 133-147, 1969.
- [11] D. C. Youla, "On the factorization of rational matrices," *IRE Trans. Inform. Theory*, vol. IT-7, pp. 172-189, 1961.
- [12] A. Lindquist and G. Picci, "On the stochastic realization problem," *SIAM J. Contr. Optimiz.*, vol. 17, pp. 365-389, 1979.
- [13] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. Cambridge, MA: M.I.T. Press, 1949.
- [14] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.
- [15] —, "Solution of a time varying Wiener filtering problem," *Electron. Lett.*, vol. 3, pp. 562-563, 1967.
- [16] T. Kailath and R. A. Geesey, "An innovations approach to least-squares estimation—Part IV: Recursive estimation given the covariance functions," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 720-727, 1971.
- [17] V. M. Popov, "Hyperstability and optimality of automatic systems with several control functions," *Rev. Roum. Sci. Technol.—Electrotechnol. et Energ.*, vol. 9, pp. 629-690, 1964.
- [18] J. C. Willems, "Least squares stationary optimal control and the algebraic Riccati equation," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 621-634, 1971.
- [19] K. Glover and J. C. Willems, "Parametrizations of linear dynamical systems: Canonical forms and identifiability," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 640-645, 1974.
- [20] G. Basile and G. Marro, "Controlled and conditioned invariant subspaces in linear systems theory," *J. Optimiz. Theory Appl.*, vol. 3, pp. 306-315, 1969.
- [21] B. D. O. Anderson, "Output-nulling invariant and controllability subspaces," in *6th Triennial World Congress IFAC*, Boston, Cambridge, 1975, part 1B, paper 45.6.
- [22] L. M. Silverman, "Discrete Riccati equations: Alternative algorithms, asymptotic properties, and system theory interpretations," *Control and Dynamic Systems*, C. T. Leondes, Ed. vol. 12, pp. 313-386, 1976.
- [23] M. K. Sain and J. L. Massey, "Invertibility of linear time-invariant dynamical systems," *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 141-149, 1969.
- [24] H. Nijmeijer and J. M. Schumacher, "On the inherent integration structure of nonlinear systems," *IMA J. Math. Contr. Inform.*, 1985.
- [25] M. L. J. Hautus, "(A, B)-invariant and stabilization subspaces, A frequency domain description," *Automatica*, vol. 16, pp. 703-707, 1980.
- [26] S. Lang, *Algebra*, Reading, MA: Addison-Wesley, 1965.
- [27] J. M. Schumacher, "The role of the dissipation matrix in singular optimal control," *Syst. Contr. Lett.*, vol. 2, pp. 262-266, 1983.
- [28] C. Commault and J. M. Dion, "Structure at infinity of linear multivariable systems: A geometric approach," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 693-696, 1982.
- [29] A. S. Morse, "System invariants under feedback and cascade control," in *Mathematical Systems Theory* (Proc. Int. Symp., Udine, Italy, 1975, Lecture Notes in Economics and Math. Systems). G. Marchesini and S. K. Mitter, Eds. New York: Springer-Verlag, 1976, pp. 61-74.
- [30] J. O'Reilly, "Comments on two recent papers on reduced-order optimal state estimation for linear systems with partially noise corrupted measurements," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 280-283, 1982.
- [31] V. B. Haas, "Minimal order Wiener filter for a linear system with exact measurements," in *Proc. Conf. Decision Contr.*, Las Vegas, NV, 1984, pp. 1257-1261.
- [32] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng., Trans. ASME, Ser. D*, vol. 82, pp. 35-45, 1960.
- [33] A. S. Morse and W. M. Wonham, "Status of noninteracting control," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 568-581, 1971.
- [34] B. A. Francis, "The optimal linear-quadratic time invariant regulator with cheap control," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 616-621, 1979.
- [35] G. Bengtsson, "Feedforward control of linear multivariable systems—The non-causal case," *Dep. Elec. Eng., Univ. Toronto, Toronto, Ont., Canada, Contr. Syst. Rep. 7506*, 1975.
- [36] B. D. O. Anderson, "A note on transmission zeros of a transfer matrix," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 589-591, 1976.
- [37] G. Bengtsson, "Minimal system inverses for linear multivariable systems," *J. Math. Anal. Appl.*, vol. 46, pp. 261-274, 1974.
- [38] G. Basile and G. Marro, "On the observability of linear, time-invariant systems with unknown inputs," *J. Opt. Theory Appl.*, vol. 3, pp. 410-415, 1969.
- [39] L. M. Silverman, "Inversion of multivariable linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 270-276, 1969.
- [40] P. Sannuti and H. Wason, "Singular perturbation analysis of cheap control problems," in *Proc. Conf. Decision Contr.*, 1983, pp. 231-236.

- [41] U. Shaked, "Nearly singular filtering for uniform and non-uniform rank linear continuous systems," *Int. J. Contr.*, vol. 38, pp. 811-830, 1983.
- [42] R. S. Bucy, "Optimal filtering with correlated noise," *J. Math. Anal. Appl.*, vol. 20, pp. 1-8, 1967.
- [43] A. G. Ramm, *Theory and Applications of Some New Classes of Integral Equations*. New York: Springer-Verlag, 1980.
- [44] H. Trentelman and J. C. Willems, "Guaranteed roll-off in a class of high-gain feedback design problems," *Syst. Contr. Lett.*, vol. 3, pp. 361-369, 1983.
- [45] G. Bengtsson, "On the solvability of an algebraic equation," Dep. Elec. Eng., Univ. Toronto, Toronto, Ont., Canada, Contr. Syst. Rep. 7507, 1975.
- [46] E. Emre and M. L. J. Hautus, "A polynomial characterization of  $(A, B)$ -invariant and reachability subspaces," *SIAM J. Contr. Optimiz.*, vol. 18, pp. 420-436, 1980.

**Johannes M. Schumacher** (M'80), for a photograph and biography, see p. 573 of the June 1985 issue of this TRANSACTIONS.