2

ESTIMATION FOR BILINEAR
STOCHASTIC SYSTEMS

Alan S. Willsky

Steven I. Marcus

Electronic Systems Laboratory

Decision and Control Sciences Group

MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139

Department of Electrical Engineering

61

ESTIMATION FOR BILINEAR
STOCHASTIC SYSTEMS†

Alan S. Willsky* and Steven I. Marcus**

Electronic Systems Laboratory
Department of Electrical Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

ABSTRACT

In this paper we consider several applications of bilinear stochastic models in which state estimation is an important problem. Bilinear stochastic models occur naturally in many communication problems, including noisy oscillators and phase-lock loops, in which the system evolves on the circle S^1. Similar models arise in the estimation of the position of an orbiting body (in which the state evolves on the 2-sphere S^2) and in the estimation of the orientation of a rotating rigid body (which evolves on SO(3)). The advection-diffusion model of air pollution involves partial differential equations which, when discretized, include bilinear stochastic terms due to the random fluctuations in wind velocity and source rate.

Three techniques for the solution of bilinear estimation problems are presented. First, finite dimensional optimal nonlinear estimators are presented for certain bilinear systems evolving on solvable and nilpotent Lie groups. Then the use of harmonic analysis for estimation problems evolving on spheres and other compact manifolds is investigated. Finally, an approximate estimation technique utilizing cumulants is discussed.

I. Introduction

As is well known, the class of linear dynamical systems with white gaussian driving and observation noises is particularly appealing in that optimal estimation and control systems can be readily determined and are easily implemented (perhaps with the aid of a digital computer). Unfortunately, there exists no such "nice" theory for general finite-dimensional nonlinear systems, and until recently most nonlinear estimation problems were "solved" by various types of linearization and vector space approximation methods.

Recently, a great deal of effort has gone into studying a class of nonlinear systems that possesses a great deal of structure itself -- the class of bilinear systems. Several authors have been able to devise analytical techniques for such systems that are as detailed and as powerful as those for linear systems. Moreover, the mathematical tools behind bilinear system analysis include not only many of the vector space techniques that are so valuable in linear system theory but also a number of tools drawn from the theories of Lie groups and differential geometry. This points out the necessity of viewing the dynamical system of interest in its most natural setting, rather than forcing it into the vector space framework.

Both the Lie theoretic and vector space settings have proven to be useful in the study of bilinear estimation problems, and a number of important and illuminating results have been uncovered. It is the purpose of this paper to explain the practical and mathematical importance of these results. In Section II we view the basic mathematical formulation

of interest to us, and in Sections III-V we discuss several important
practical problems that fall within this framework. These include a
large class of synchronous communication problems, the problems of
altitude estimation and the tracking of an orbiting vehicle, and the
estimation of pollutant concentrations in a diffusive atmosphere subject
to random wind effects and fluctuations in the source rate. These by
no means cover all of the potential applications for our bilinear esti-
mation results, but they do indicate the range of problems that can be
considered. We refer the reader to the references for other applications.

In Sections VI-VIII we review the techniques that have been devel-
oped for bilinear estimation. In Section VI we describe a class of
bilinear systems for which complete analysis is possible, and we display
the optimal, nonlinear, finite-dimensional estimation equations for an
example. The Lie-theoretic significance of these results is also dis-
cussed. In Section VII the use in estimation system design of harmonic
analysis on groups is explored in the context of synchronous communication
and orbital tracking. This represents a potentially powerful tool in the
design of high performance, implementable estimation systems. A second
approximation method, based on the truncation of the cumulants of a
random process, is studied in Section VIII. This approach is more closely
related to the usual vector space techniques.

II. Stochastic Bilinear Systems

In this section we briefly describe the several classes of stochastic equations that will be considered in the remaining sections of this paper. The basic deterministic bilinear equation considered in the literature [1] - [9] is

$$\dot{x}(t) = \left[A_o + \sum_{i=1}^{N} u_i(t) A_i \right] x(t) \tag{2.1}$$

where the A_i are given n x n matrices, the u_i are scalar inputs, and x is either an n-vector or an n x n matrix. As discussed in [1], the additive control model

$$\dot{x}(t) = \left[B_o + \sum_{i=1}^{N} u_i(t) B_i \right] x(t) + Cu(t) \tag{2.2}$$

(here u is the vector of the u_i) can be reduced to the form (2.1) by state augmentation. Also, if we apply the bilinear feedback law

$$u_i(t) = v_i(t) \ell_i(x(t)) + \mu_i(t) \tag{2.3}$$

where v_i and μ_i are scalars, and ℓ_i is a scalar-valued linear function of x, our system equation becomes

$$\dot{x}(t) = \left[A_o + \sum_{i=1}^{N} (v_i(t) \ell_i(x(t)) + \mu_i(t))A_i \right] x(t) \tag{2.4}$$

which involves products of state variables. By including several feedback paths, we can obtain essentially arbitrary polynomials in the state variables.

In this paper we will consider equations such as (2.1) in which the u_i are stochastic processes. Such systems have become considered by several authors [9] - [27]. We refer the reader to [10] - [16] for

-3-

detailed discussions of the properties of such stochastic models.

One must be careful in considering stochastic versions of (2.1).
For instance, if u(t) is a vector zero mean white noise with

$$E[u(t) \; u(s)] = R(t) \; \delta \; (t-s) \qquad (2.5)$$

the Ito stochastic differential analog of (2.1) is

$$dx(t) = \{ [A_o + \frac{1}{2} \sum_{i,j=1}^{N} R_{ij}(t) \; A_i A_j] dt + \sum_{i=1}^{N} A_i dv_i(t) \} \; x(t) \qquad (2.6)$$

where v is the integral of u (i.e., it is a Brownian motion process).

We define $L = \{A_i\}_{LA}$ to be the matrix Lie algebra generated by
$\{A_i\}_{i=o}^{N}$ -- i.e., it is the smallest subspace that contains the $\{A_i\}$ and

is closed under the commutator product

$$[M_1, \; M_2] = M_1 \; M_2 - M_2 \; M_1 \qquad (2.7)$$

One can show that in the deterministic case with x an n x n matrix, if
x(o) is an element of the matrix Lie group

$$G = \{exp \; L\}_G = \{e^{B_1} e^{B_2} \ldots e^{B_m} \; | \; B_i \; \varepsilon \; L\} \qquad (2.8)$$

then x(t) is an element of G (for all t $\geq$ 0). In order to make a similar

statement in the stochastic case when u is a white noise, we must include

a correction term, as indicated in (2.6).

Another case of considerable importance arises if u is generated

by a finite dimensional linear diffusion process

$$d\,\xi\,(t) = F(t)\,\xi\,(t)\,dt + G(t)\,dw(t) + \alpha(t)\,dt \qquad (2.9)$$

$$u(t) = H(t)\,\xi\,(t) \qquad (2.10)$$

where α, F, G, and H are known and w is a standard Brownian motion process $(E\,[dw(t)dw'(t)] = I\,dt)$. In this case, x by itself is not a Markov process, but the pair (x, ξ) is. Augmenting the state with ξ, we obtain a stochastic equation of the form (2.4), where the $v_i = 1$ and some of the μ_i are white noises while the others are zero (see several examples in the following sections). We note that one can show [16], [17] that in this case no correction term need be added to (2.4). Also, the right-hand side of (2.4) does not satisfy the global Lipschitz conditions often assumed in proving the existence of solutions to I to differential equations [28], [29]. Again, one can shown [15], [20], [30], [31] that this causes no problems in the case when (2.4) arises from (2.1) driven by the colored noise (2.9) - (2.10).

In the remaining sections of this paper, we will consider the estimation of processes described by stochastic bilinear equations of the types just discussed. We now briefly describe the various types of measurement processes that will be considered. We also refer the reader to the references [16] - [27] for more on these estimation problems.

One very important measurement process consists of linear measurements corrupted by additive noise

$$dz(t) = L(x(t))dt + dv(t) \qquad (2.11)$$

where L is a linear operator (recall X is either an n-vector or an n x n

matrix) and v is a Brownian motion process. A second observation model is the "multiplicative measurement noise" case

$$Z(t) = X(t) \ V(t) \tag{2.12}$$

in which Z, X and V are all n x n matrices. The third and final measurement process is described by a bilinear dynamical equation

$$dz(t) = \{ [A_o + \frac{1}{2} \sum_{i,j=1}^{N} R_{ij}(t) \ A_i A_j] \ dt$$

$$+ \sum_{i=1}^{N} [\ell_i(x(t)) \ dt + dv_i(t)] \ A_i \} \ z(t) \tag{2.13}$$

Examples of each of these processes will be given in the next few sections.

We close this section by noting that many but not all of our results are motivated by considering these estimation problems in the context of Lie group theory. We refer the reader to [1] - [5], [10], [16], [17], and [20] - [23] for more on this subject.

III. Estimation Problems Arising in Communication Applications

An important problem in a large number of communications appli-
cations [10], [17], [19], [24], [25], [32] - [37] is the processing of
a signal of the form

$$r(t) = A(t) \sin (w_c t + \phi(t) + v(t)) + N(t) \qquad (3.1)$$

where w_c is a known carrier frequency, ϕ is some type of modulating
information, v is a random phase drift, A is the sinusoidal amplitude,
possibly containing modulating information and/or noise, and N is
additive channel noise. As discussed in [10], [17], [19], [24], [25],
and [34], a number of specific problems that fit into the general form
given by (3.1) can be modeled by equations of the type described in the
preceding section. In this section we illustrate these ideas by
considering several specific examples.

Example 1: We consider a phase tracking problem of importance in radio
navigation systems such as Omega [38]. This problem has been studied
in [24], [25], [32] - [34] and [36]. The solution technique developed
in [24], [25] is discussed in Section VII.

Suppose we receive the signal

$$\dot{z}(t) = \sin \theta(t) + r^{1/2} (t) \, \dot{w}(t) \qquad (3.2)$$

where

$$\theta(t) = w_c t + \int_o^t q^{1/2}(s) \, ds + \theta_o \qquad (3.3)$$

and v and w are independent standard Brownian motions, $q(t) \geq 0$,
$r(t) > 0$, and $w_c > 0$. Also θ_o is a random initial condition independent
of v and w. We desire to track the signal phase -- i.e., we wish to

estimate $\theta(t)$ mod 2π given $\{z(s) \mid 0 \le s \le t\}$. Equation (3.2) is, of course, only formal, since w is white noise. The I to differential forms of (3.2) and (3.3) are

$$d\theta(t) = w_c dt + q^{1/2}(t)\, dv(t), \quad \theta(o) = \theta_o \qquad (3.4)$$

$$dz(t) = \sin\theta(t)\, dt + r^{1/2}(t)\, dw(t) \qquad (3.5)$$

and we take as our optimal estimation criterion the minimization of

$$E[(1 - \cos(\theta(t) - \tilde{\theta}(t)) \mid z(s), 0 \le s \le t].$$

Noting that we are essentially tracking a point on the unit circle S^1 in the plane R^2, we can reformulate our problem in Cartesian coordinates. Let

$$x_1 = \sin\theta(t), \qquad x_2 = \cos\theta(t) \qquad (3.6)$$

then

$$\begin{bmatrix} dx_1(t) \\ \\ dx_2(t) \end{bmatrix} = \begin{bmatrix} -q(t)\, dt/2 & w_c dt + q^{1/2}(t)\, dv(t) \\ \\ -(w_c dt + q^{1/2}(t)\, dv(t)) & -q(t)\, dt/2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ \\ x_2(t) \end{bmatrix} \qquad (3.7)$$

$$dz(t) = x_1(t)\, dt + r^{1/2}(t)\, dw(t) \qquad (3.8)$$

which are of the bilinear process - linear measurement type discussed in Section II. Note that (3.7) describes what appears to be a damped oscillator; however the damping terms are the correction terms required by Ito calculus, and one can show that

$$E(x^2(t) + x_2^2(t)) = 1 \tag{3.9}$$

(see [16], [17] for further discussion).

In Cartesian coordinates our estimation problem is to choose an estimate $(\tilde{x}_1(t), \tilde{x}_2(t))$ on the unit circle - i.e., such that

$$\tilde{x}_1^2(t) + \tilde{x}_2^2(t) = 1 \tag{3.10}$$

If we use the least squares criterion

$$J = 1/2 \, E[(x_1(t) - \tilde{x}_1(t))^2 + (x_2(t) - \tilde{x}_2(t))^2 | z(s), \, 0 \leq s \leq t] \tag{3.11}$$

subject to (3.10), or equivalently subject to

$$\tilde{x}_1(t) = \sin \tilde{\theta}(t), \quad \tilde{x}_2(t) = \cos \tilde{\theta}(t) \tag{3.12}$$

our criterion reduces to

$$J = E[1 - \cos(\theta(t) - \tilde{\theta}(t)) | z(s), \, 0 \leq s \leq t] \tag{3.13}$$

-- i.e. (3.13) represents a contrained least squares criterion. One can show that

$$(\tilde{x}_1(t), \tilde{x}_2(t)) = \frac{(\hat{x}_1(t|t), \hat{x}_2(t|t))}{\sqrt{\hat{x}_1^2(t|t) + \hat{x}_2^2(t|t)}} \tag{3.14}$$

or

$$\tilde{\theta}(t) = \tan^{-1} \frac{\hat{x}_1(t|t)}{\hat{x}_2(t|t)} \tag{3.15}$$

where

$$\hat{x}_i(t|t) = E[x_i(t)|z(s), \ 0 \le s \le t], i = 1,2 \tag{3.16}$$

Referring to Figure 3.1 we can see the geometric significance of this criterion. We note that one can show that

$$P(t) = \sqrt{\hat{x}_1^2(t|t) + \hat{x}_2^2(t|t)} \le 1 \tag{3.17}$$

and the quantity $P(t)$ is a measure of our confidence in our estimate. Specifically if θ is a normal random variable with variance γ, then (see [16] - [18], [24], [25])

$$P = \sqrt{[E(\sin \theta)]^2 + [E(\cos \theta)]^2} = e^{-\gamma/2} \tag{3.18}$$

so $\gamma = 0$ (perfect knowledge of θ) $\implies P = 1$
and $\gamma = \infty$ (no knowledge) $\implies P = 0$.

Example 2: Consider the demodulation of an FM signal in the presence of both phase and additive channel noise. Specifically, suppose the received signal process is

$$dz(t) = \sin \ (w_c t + g \int_0^t x(s) \ ds + \int_0^t e^{1/2}(s) \ df(s)) \ dt + r^{1/2}(t) \ dw(t) \tag{3.19}$$

where x, the modulating information to be recovered, arises from a linear diffusion process

$$dx(t) = a(t) \ x(t)dt + q^{1/2}(t) \ dv(t) \tag{3.20}$$

Here f, w, and v are independent standard Brownian motion processes.

Equations (3.19) and (3.20) can be replaced by equivalanet equations in a manner similar to that used in Example 1. Our state
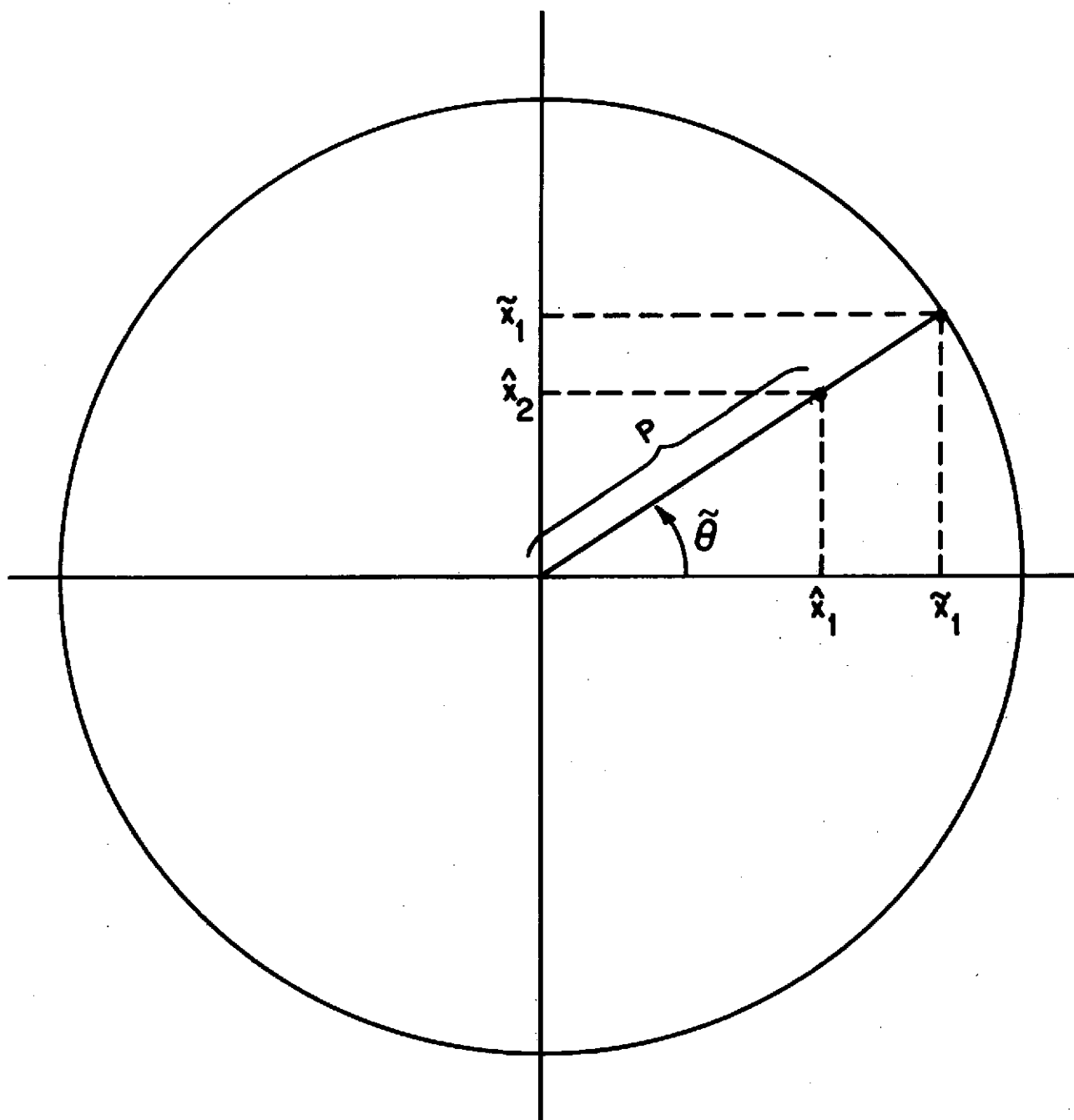
Fig. 3.1   Illustrating the Geometric Interpretation of the Criterion
$$\mathcal{E}\left[1 - \cos(\theta - \widetilde{\theta})\right]$$

equation becomes

$$
\begin{bmatrix} dx_1 \\ dx_2 \\ dx_3 \\ dx_4 \end{bmatrix} = \begin{bmatrix} -edt/2 & (w_c + gx_3)dt + e^{1/2}dt & 0 & 0 \\ -(w_c + gx_3)dt - e^{1/2}dt & -edt/2 & 0 & 0 \\ 0 & 0 & a & q^{1/2}dv \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}
$$

(3.21)

$$
dz = x_1 \, dt + r^{1/2} \, dw(t)
$$

(3.22)

where $x_3 = x$ and $x_4 \equiv 1$ (this is the type of augmentation we use to "bilinearize" linear systems [1]. Note the products $x_1 x_3$ and $x_2 x_3$ in (3.21).

Example 3: In this example we consider an FM problem with phrase noise only. This problem was considered in [16] - [19], and we refer the reader to [19] for further discussions of examples of this type. Suppose we observe the signal

$$
z_1(t) = \sin \left(w_c dt + \int_0^t h(s) \, x(s)ds + \int_0^t q^{1/2}(s) \, dw(s) \right)
$$

(3.23)

where x is given by (3.20) and w, a standard Brownian motion independent of x, represents a random phase drift. A number of physical sources for such noise are discussed in [19]. We note that in standard FM systems involving limiter-discriminators, additive channel noise is processed in such a manner as to yield frequency and phase deviations (see [37], [39]).

Also, such models arise in the problem of tracking drifts in high quality oscillators used as frequency standards or high-precision clocks [19], [40], in the problem of recovering velocity information from Doppler-shifted signals [19], [41].

As discussed in a number of references, given the signal (3.23), there are methods for generating the additional signal

$$
z_2(t) = \cos\left(w_c t + \int_0^t h(s)\, x(s)\, ds + \int_0^t q^{1/2}(s)\, dw(s)\right) \qquad (3.24)
$$

(see the method described in [19] in which the total phase of (3.23) is reconstructed with the aid of cycle counters). Defining

$$
Z(t) = \begin{bmatrix} z_2(t) & z_1(t) \\ -z_1(t) & z_2(t) \end{bmatrix} \qquad (3.25)
$$

$$
X(t) = \begin{bmatrix} \cos\left(w_c t + \int_0^t h(s)\, x(s)\, ds\right) & \sin\left(w_c t + \int_0^t h(s)\, x(s)\, ds\right) \\ -\sin\left(w_c t + \int_0^t h(s)\, x(s)\, ds\right) & \cos\left(w_c t + \int_0^t h(s)\, x(s)\, ds\right) \end{bmatrix} \qquad (3.26)
$$

$$
W(t) = \begin{bmatrix} \cos\int_0^t q^{1/2}(s)\, dw(s) & \sin\int_0^t q^{1/2}(s)\, dw(s) \\ -\sin\int_0^t q^{1/2}(s)\, dw(s) & \cos\int_0^t q^{1/2}(s)\, dw(s) \end{bmatrix} \qquad (3.27)
$$

we find that

$$
dZ = \begin{bmatrix} -q\,dt/2 & (w_c + hx)\,dt + q^{1/2}dw \\ -(w_c + hx)\,dt - q^{1/2}dw & -q\,dt/2 \end{bmatrix} Z \qquad (3.28)
$$

which is a bilinear observation equation. Similarly, X and V satisfy

bilinear stochastic differential equations. Also, we have

$$Z(t) = X(t) \ V(t) \tag{3.29}$$

-- i.e. Z can be thought of as a measurement of X in multiplicative noise.

Note that the estimation of X, instead of x, corresponds to tracking,

rather than demodulating, the signal phase.

Finally, we make several comments about the matrix Lie group

on which X, V and Z evolve. This group denoted by SO(2), is the group

of 2 x 2 orthogonal matrices of determinant +1 -- i.e.,

$$SO(2) = \{X \mid X'X = I, \ \det X = +1\} \tag{3.30}$$

As developed in [16] - [19], each $X \ \varepsilon \ SO(2)$ can be written in the form

$$X \ = \ \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \ , \ \theta \ \varepsilon \ [-\pi, \ \pi] \tag{3.31}$$

and SO(2) is isomorphic (as a Lie group, [42], [43]) to the circle S^1.

Also SO(2) is an abelian Lie group -- i.e.

$$X \ Y \ = \ Y \ X \qquad X, \ Y \ \varepsilon \ SO(2) \tag{3.32}$$

and the multiplication (3.32) corresponds to the mod 2π addition of the

corresponding angles in the representation (3.31). We note that a con-

sequence of the commutativity of SO(2) is the fact that Z can be written

in both the bilinear form (3.28) and the multiplicative form (3.29).

Example 4: As discussed in [17], the real line R^1 and the circle S^1

are essentially the only abelian Lie groups, since any abelian Lie group

G is isomorphic to a direct product of a number of copies of each. For instance, the group D of nonzero complex numbers is isomorphic to $R^1 \times S^1$ under the map

$$(r, \theta) \rightarrow e^{r+i\theta} \qquad r\epsilon R^1, \theta\epsilon[-\pi,\pi) \qquad (3.33)$$

Consider the R^2 process $x' = (x_1, x_2)$

$$dx(t) = A(t) x(t) dt + B(t) dw(t) \qquad (3.34)$$

where w is a standard two-dimensional Brownian motion, and also consider the D-valued signal process

$$z(t) = e^{y_1(t) + iy_2(t)} \qquad (3.35)$$

where

$$dy(t) = x(t) dt + dv(t) \qquad (3.36)$$

Here v is a 2-dimensional Brownian motion process independent of x. We note that z satisfies a (complex) bilinear stochastic differential equation and also that we have the multiplicative form

$$z(t) = \left[e^{v_1(t) + iv_2(t)} \right] \left[e^{\int_0^t (x_1(s) + ix_2(s)) ds} \right] \qquad (3.37)$$

Thus z is both amplitude and angle modulated, and the noise is a multiplicative lognormal process [19], [44], including both phase and amplitude noise.

We note that the multiplicative lognormal noise process in (3.37) is an important model in some optical communication problems [44]. In many cases, changes in the transmission medium -- e.g., turbulence in the

atmosphere -- cause variations in the refractive index of the air. This disturbance can be modeled [44] as a multiplicative lognormal noise process as in (3.37). Also, if v_1 and x_1 are zero, (3.37) is identical to (3.29) if we observe that the set of unit modulus complex numbers, as well as SO(2), is isomorphic to S^1.

IV. Estimation of Rotational Processes in Three Dimensions

As we have seen, many communication problems can be placed in the framework of estimation of processes evolving on the circle S^1 -- i.e., processes of rotation in one dimension. In this section we formulate several problems of practical importance involving rotation in three dimensions. As we shall see in this and in later sections, these problems are considerably more difficult than the one dimensional problems, since rotations in three-space do not commute [13], [22], [27], [45], [46].

The problem of estimating and controlling the angular velocity and orientation of a rigid body has been studied by many authors [9], [22], [27], [45] - [48] and is of great importance in many aerospace and inertial navigation applications. Such problems are by no means trivial, and most of the techniques that have been developed are sub-optimal in nature. One feature of the rigid body orientation - angular velocity problem that has received some attention in the past is that the space of possible orientations is a Lie group [22], [27], [48], and the combined orientation - angular velocity space is the "tangent bundle" of the orientation space and thus is a homogeneous space [49]. The framework of differential geometry and Lie theory has proven useful in studying rigid body rotation problems. In fact, there are Lie-theoretic interpretations of four of the most widely used representations of the attitude of a rigid body -- direction cosines, Euler angles, unit quaternions, and Cayley-Klein parameters.

We will consider here only the direction cosine description and refer the reader to [22] and [48] for discussions of the other representations.

-17-

The orientation of a rigid body can be specified in terms of the direction cosines [46] between two sets of orthogonal axes -- one rotating with the body and one an inertial reference frame. The direction cosines are usually given as a 3 x 3 orthogonal matrix X of positive determinant -- i.e.

$$X'X = I \qquad \det X = +1 \qquad (4.1)$$

The set of all direction cosine matrices forms the matrix Lie group SO(3) [1], [16], [22]. If the 3-vector $\xi(t)$ is the (properly coordinatized, [46]) angular velocity vector of the body with respect to inertial space, the time evolution of the orientation of the body can be described by the bilinear equation

$$\dot{X}(t) = (\sum_{i=1}^{3} \xi_i(t) \, R_i) X(t) \qquad (4.2)$$

where $X(t) \in SO(3)$ and the R_i, given by

$$R_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \qquad R_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \qquad R_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$(4.3)$$

form a basis for so(3), the matrix Lie algebra associated with SO(3).

It is well known and easy to check that SO(3) is not an abelian Lie group and, equivalently, that so(3) is not an abelian Lie algebra (i.e., commutator products are not identically zero). In fact SO(3) is

a <u>simple</u> Lie group [16], [22], [50], and it is this fact that makes the study of dynamics on SO(3) so difficult. Recall that in the SO(2) case we were able to represent solutions to bilinear equations by terms of the form

$$X(t) = \exp (R \int_0^t \xi(s) \, ds) \tag{4.4}$$

where

$$R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{4.5}$$

is a basis for the one-dimensional Lie algebra of SO(2). Wei and Norman [51], [52] have shown that one can obtain similar local expressions for the solution to equations of the form (2.1) but that such solutions are global only in certain cases -- i.e., if the underlying Lie algebra is <u>solvable</u> (see [10], [16], [22], and Section VI of this paper for a discussion of the significance of this statement). As simple Lie algebras are not solvable, we obtain no such global representation here, and as we shall see, we must resort to suboptimal methods in the design of attitude estimation systems. We also note that the local Wei-Norman representation of the solution of (4.2) corresponds to the Euler angle description, which is well known to exist only locally (see [22] and [46], where this fact is related to the phenomenon of "gimbal-lock").

Suppose now that the angular velocity vector in (4.2) is stochastic. Specifically, we suppose ξ satisfies

$$d\xi(t) = f(t)dt + A(t) \xi(t)dt + Q^{1/2}(t) \, dw(t) \tag{4.6}$$

where f and $Q \geq 0$ are known, $\xi(0)$ is normally distributed, and w is a standard 3-dimensional Brownian motion process independent of $\xi(0)$. Here f can be thought of as a vector of (known) torques acting on the body, and the Brownian motion term represents random disturbances (e.g., those caused by noisy responses of control devices, such as reaction jets, that are used to implement the desired torque f, or the effect of a gravity gradient).

Note also that the angular velocity equation (4.6) that we have postulated is simpler than the usual nonlinear Euler equations [46]. Equation (4.6), with suitable choice of f, A, and Q, can be viewed as a reasonable approximation if: (1) the rigid body is nearly spherically symmetric (so that the principle moments of inertia are almost equal -- in this case the nonlinear terms in the Euler equations are small and may be lumped into the random disturbance term); or (2) we linearize Euler's equations about a nominal (which might be included in the f(t) term); or (3) we make Q(t) large enough so that the nonlinear effects can be viewed as process noise.

We now can describe two different measurement processes -- one motivated by a strapdown navigation system, and the other by an inertial system in which a platform is to be kept inertially fixed. In a strapdown system, [46] one receives noisy information about either angular velocity or incremental angle changes. Assuming that the size of the increment is small, either type of information can be modeled by the Itô equation

$$dz(t) = C(t)\xi(t)dt + S^{1/2}(t) \, dv(t) \qquad (4.7)$$

where $S > 0$ and v is a standard Brownian motion process, independent of ξ. This equation is the precise analog of the usual (formal) observation equation (for rate gyroscopes)

$$\dot{z}(t) = C(t)\xi(t) + \dot{v}(t) \tag{4.8}$$

Here $\dot{v}$ is white gyro noise. Another possible observation model is provided by integrating gyroscopes, in which case we observe

$$m(t) = C(t) \int_0^t \xi(s)\,ds + N(t) \tag{4.9}$$

Here N represents a gyro drift. Usually when using a model like (4.9), one assumes that the drift is a correlated process. This adds no real difficulty to the problem, since there are simple techniques for handling measurements with additive colored noise [53]. In fact, the consideration of the process (4.9) rather than (4.8) adds no difficulty to the analysis, and thus we will concentrate on (4.8).

In the usual strapdown system the information (4.7) is processed by a "direction cosine computer," which produces $Z(t)$, the solution of

$$dZ(t) = \left[\sum_{i=1}^{3} R_i\,dz_i(t) + \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} S_{ij}(t) R_i R_j\,dt \right] Z(t) \tag{4.10}$$

Our problem is to use the information supplied by z or Z to compute "good" estimates of the angular velocity ξ and the orientation X. Thus, if we take (4.7) as our basic measurement equation, we have linear observations, while (4.10) yields a bilinear measurement process.

A second type of observation process is suggested by an inertial system equipped with a platform that is to "instrument" (i.e. remain fixed

with respect to) the inertial reference frame. Suppose we use the notation b-frame for the body-fixed frame, p-frame for the platform frame, and i-frame for the inertial reference frame. Letting $C_\alpha^\beta(t)$ denote the direction cosine matrix of the β-frame with respect to the α-frame, we have

$$X(t) = C_i^b(t) \tag{4.11}$$

Also, by noting the relative orientation of the platform and the body (perhaps by reading of gimbal angles [46], we can measure

$$M(t) = C_p^b(t) \tag{4.12}$$

Let

$$V(t) = C_i^p(t) \tag{4.13}$$

Then $V(t)$ represents platform misalignment -- a drift of the platform with respect to inertial space due both to drifts in the gyroscopes used to sense rotation of the rigid body and also to inaccuracies in the mechanism that rotates the platform relative to the body in order to keep it inertially fixed. If we model gyro drifts and the other inaccuracies as Brownian motion processes, a reasonable model for V is as an SO(3) Brownian motion, [13], [16], [22], [54]:

$$dV(t) = \left[\sum_{i=1}^{3} R_i dv_i(t) + \frac{1}{2} \sum_{i=1}^{3} \sum_{i=1}^{3} S_{ij}(t) R_i R_j dt \right] V(t) \tag{4.14}$$

where v is the Brownian motion process with $E[dv(t)dv'(t)] = S(t)dt$.

Using elementrary properties of direction cosine matrices [46], we have

$$M(t) = C_p^b(t) = C_i^b(t) \, C_p^i(t) = C_i^b(t) \, [C_i^b(t)]' = X(t)V'(t) \qquad (4.15)$$

(multiplicative observation noise) and, using (4.2) and (4.14) (plus the fact that $R_i' = -R_i$), we have the following stochastic differential equation for M:

$$dM(t) = \left[\sum_{i=1}^{3} R_i \xi_i(t) \right] M(t)dt + M(t) \left[-\sum_{i=1}^{3} R_i dv_i(t) + \frac{1}{2} \sum_{i,j=1}^{3} S_{ij}(t) R_i R_j dt \right]$$

$$(4.16)$$

Again, we wish to consider the problem of estimating ξ and X given the observation process M.

In many inertial systems the type of information that is available is in the form of "pulses" [46] from the gyros. If this is the case, it is logical to take the incremental equation (4.10) as the basic sensor equation. Also, for the second observation process, if we take the incremental gimbal angles as the quantity we observe, equation (4.16) becomes the basic measurement equation.

Now consider (4.2) with the 3 x 3 matrix X replaced by the 3-vector x. Assuming that $x'(0) \, x(0) = 1$, we have that $x'(t) \, x(t) = 1, \forall t$ -- i.e., x evolves on S^2, the unit sphere in 3-space. The study of random processes with constant "energy" is of importance in a number of fields -- including DC to DC conversion [55], statistical mechanics [56], and satellite orbital analysis [16], [57], [58]. We briefly describe a simplified version of

a satellite tracking problem of great practical importance.

Consider a satellite in circular orbit about some celestial body, such as the moon. Because of a variety of effects including anomalies in the gravitational field of the body, effects of the gravitational fields of nearby bodies, and the effects of "solar wind", the orbit of the satellite is perturbed. For simplicity, we assume that the perturbations of the orbit are tangential only -- i.e. the radial effects are unimportant or have been corrected for. In this case, a stochastic version of (4.2) arises. Suppose the angular velocity ξ of the vehicle with respect to body can be written the form

$$\xi(t) = f(t) + \dot{w}(t) \tag{4.17}$$

where f is the nominal orbit ("carrier") angular velocity and $\dot{w}$ is a white perturbation with

$$E[\dot{w}(t)\dot{w}(s)] = Q(t)\delta(t-s) \tag{4.18}$$

The stochastic analog of (4.2) then is

$$dx(t) = \{[\sum_{i=1}^{3} f_i(t)A_i + \sum_{i,j=1}^{3} Q_{ij}(t) A_i A_j]dt$$

$$+ \sum_{i=1}^{3} A_i dw_i(t)\} x(t) \tag{4.19}$$

If we are then given noisy observations of the satellite position

$$dz(t) = H(t) x(t)dt + R^{1/2}(t) dv(t) \tag{4.20}$$

where v is Brownian observation noise independent of x and $R(t) > 0$, our problem is to estimate $x(t)$.

We note that the assumption that the angular velocity perturbations
are white is a simplification. For instance, the anomalies in the moon's
gravitational field are spatially correlated and constitute a random
field [29], [57], [58], which can be estimated [58] from the observed
data (4.20). However, the computations involved in such an estimation
are of such magnitude that they must be computed off-line. The simpler
model (4.19) may lead to simple but accurate on-line tracking schemes
(see Section VII). Also, by including perturbations with time correlation
consistent with the period of the orbit, we may be able to make estimation
systems based on (4.19), (4.20) "smarter" with only minor complications
in tracking system design. We note that (4.19), (4.20) is a direct
s^2 analog of the s^1 incoherent oscillator tracking problem described in
Example 1 (incoherent orbiters?).

In fact, we can again use the constrained least-squares criterion:
minimize

$$E[(x(t) - \widetilde{x}(t))' \, (x(t) - \widetilde{x}(t)) \mid z(s), \, 0 \leq s \leq t] \qquad (4.21)$$

subject to

$$||\widetilde{x}(t)||^2 = \widetilde{x}'(t) \, \widetilde{x}(t) = 1 \qquad (4.22)$$

It can be shown that, as in the s^1 problem,

$$x(t) = \frac{\hat{x}(t|t)}{|| \hat{x}(t|t) ||} \qquad (4.23)$$

V. Estimation of Air Pollution

The problem of estimating the concentration of pollutants in the air is a vital first step toward the goal of maintaining the air pollution levels within safe limits. A recently developed stochastic model for air pollution [63], [64] involves partial differential equations which, when discretized, become discrete-time bilinear equations (see [65], [66] for discussions of discrete-time bilinear systems). The advection-diffusion model of [63], [64], which is a generalization of the widely-used steady-state Gaussian plume model, accounts for the continuous fluctuation of meteorological factors by means of stochastic modeling.

The transportation in air of a single-species pollutant is approximated by the advection-diffusion equation

$$\frac{\partial C}{\partial t} (x,y,z,t) = - \bar{V} \cdot \nabla C(x,y,z,t) + \nabla (K_c \nabla C(x,y,z,t))$$

$$+ \bar{Q}_c (x,y,z,t) - v \cdot \nabla \bar{C}(x,y,z,t) + q_c (x,y,z,t)$$

$$(5.1)$$

where C is the pollution concentration; $\bar{C}$ is the mean of C; $\bar{V}$ and v are the mean and the zero-mean stochastic component of the wind velocity; $\bar{Q}_c$ and q_c are the mean and the zero-mean stochastic component of the pollution source rate; and K_c is the eddy diffusivity. For the boundary conditions associated with (5.1), see [63], [64].

For practical implementation, the advection-diffusion equation (5.1) is discretized both in space and time, resulting in the finite dimensional system of equations

$$A \; x(t+1) = [2I - A + \hat{E} \; (V_D)] \; x(t) + \overline{D} + L \; \overline{Q}_c + \Gamma(x(t)) \; w(t) \tag{5.2}$$

$$\Gamma(x(t)) = [G(x(t)) \quad L]$$

$$w(t) \quad = \quad \begin{bmatrix} v(t) \\ q_c(t) \end{bmatrix}$$

$$z(t) = H \; x(t) + \xi(t) \tag{5.3}$$

where $x(t)$ is the vector of pollution concentrations in the cells of the spatial discretization, and A, L, $\hat{E}(V_D)$, $\overline{D}$, G, H, and I (= identity) are known constant matrices. The noise processes $v(t)$, $q_c(t)$, and $\xi(t)$ are assumed to be independent, Gaussian, and white. Since Γ $(x(t))$ is <u>linear</u> in $x(t)$, the system (5.2) is obviously of the discrete-time bilinear form, and it is driven by fluctuations in the wind velocity and source rate.

We will not consider specific estimation schemes for the system (5.2) - (5.3); however, we remark that the cumulant method of Section VIII may provide a useful suboptimal approach to the problem. In [63], [64] it is assumed that Γ $(x(t))$ is slowly varying, and a suboptimal filter, which employs a non-Riccati estimation algorithm using incremental covariance [67], [68], is designed for the resulting "linear" system.

Finally, we refer the reader to [80] for a discussion of deterministic bilinear systems described by partial differential equations. Also, if we discretize in space only, we obtain continuous-time bilinear equations, and the techniques of Sections VI and VII may be applicable.

VI. Finite Dimensional Optimal Nonlinear Estimators

In this section we will consider the estimation problem for a class
of systems evolving on nilpotent or solvable Lie groups [10], [16], [22],
[69]. The equations we will consider are motivated by the strapdown
navigation system of System IV --

$$d \xi (t) = F(t) \xi (t) dt + Q^{1/2}(t) dw(t) \qquad (6.1)$$

$$X(t) = (A_o + \sum_{i=1}^{n} \xi_i (t) A_i) X(t); \quad X(o) = I \qquad (6.2)$$

$$dz(t) = H(t) \xi (t)dt + R^{1/2}(t) dv(t) \qquad (6.3)$$

where $\xi (t)$ is an n-vector, $X(t)$ is a k x k matrix, $z(t)$ is a p-vector,
w and v are independent standard Brownian motion processes, $Q \geq 0$,
$R > 0$, and $\xi(0)$ has a Gaussian density independent of w and v.

The criterion for the optimal estimate $\tilde{X}(t|t)$ will be the
minimization of

$$E [tr \{(X(t) - \tilde{X}(t))' (X(t) - \tilde{X}(t))\}|z(s), 0 \leq s \leq t\},$$

where "tr" denotes trace. It is well known [28] that this minimum-variance
estimate is given by the conditional mean

$$\hat{X}(t|t) = E^t[X(t)] \triangleq E[X(t)|z(s), 0 \leq s \leq t\},$$

In general, the computation of $\hat{X}(t|t)$ requires an infinite-dimensional system
of equations, and approximations must be made for practical implementation
(see the related comments in [28] and Sections VII and VIII). Thus it is
of interest to study systems for which the optimal minimum variance estimator
is dimensional (and thus implementable on-line with a digital computer).

Throughout this section, we assume that $\{A_i\}_{LA}$, the smallest Lie algebra which contains $\{A_i, i = 0, 1, \ldots, n\}$, is solvable; this is equivalent to the existance of a complex nonsingular k x k matrix P such that, for all elements A in $\{A_i\}_{LA}$, the matrix $P \, A \, P^{-1}$ is in the upper triangular form where $a_{ij} = 0$ for $i > j$ and the other elements are arbitrary (see [69] for further details on solvable Lie groups and Lie algebras). Thus we will assume that the matrices $\{A_i\}$ in (6.2) are in upper triangular form; this implies that X evolves on the solvable Lie group $G_s(k)$ of nonsingular upper triangular k x k matrices.

First we consider the Lie subgroup $G_N(k)$ of upper triangular k x k matrices with $a_{ij} = 1$, $i = 1, \ldots, k$; this is a nilpotent Lie group. The corresponding nilpotent Lie algebra $L_N(k)$ consists of the <u>strictly</u> upper triangular k x k matrices. In this case, because the Peano-Baker series is finite, X(t) can be expressed in closed form in terms of a <u>finite</u> number of integrals in which the integrands are products of the components of ξ. We can show that, for the system (6.1) – (6.3) evolving on $G_N(k)$, the optimal estimate $\hat{X}(t|t)$ can be computed by a finite dimensional nonlinear filter. The starting point of the derivation is a closed-form expression for X(t).

For simplicity, the filter will only be derived for $n = k = 3$, but the result can also be proved by induction for higher-order systems.

Let $A_o = 0$ and

$$A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad A_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$(6.4)$$

Then $\{A_1, A_2, A_3\}$ is a basis for $L_N(3)$. The solution of (6.2) can be expressed in closed form as

$$X(t) = \begin{bmatrix} 1 & \int_0^t \xi_1(\sigma)\,d\sigma & \int_0^t \xi_2(\sigma)\,d\sigma + \int_0^t\int_0^{\sigma_1}\xi_1(\sigma_1)\xi_3(\sigma_2)\,d\sigma_2\,d\sigma_1 \\ 0 & 1 & \int_0^t \xi_3(\sigma)\,d\sigma \\ 0 & 0 & 1 \end{bmatrix}$$

(6.5)

Before stating the major theorem concerning the computation of $\hat{X}(t|t)$, we need two preliminary results.

Lemma 6.1: Consider the linear system (6.1), (6.3), and define (for $\sigma \le t$)

$$E^t[\xi(\sigma)] = \hat{\xi}(\sigma|t) = E[\xi(\sigma) \mid z(s),\ 0 \le s \le t] \tag{6.6}$$

Then the conditional cross-covariance matrix

$$P(\sigma_1,\ \sigma_2,\ t) = E[\xi(\sigma_1) - \hat{\xi}(\sigma_1|t))\ (\xi(\sigma_2) - \hat{\xi}(\sigma_2|t))'|z(s),\ 0 \le s \le t] \tag{6.7}$$

is nonrandom -- i.e., it is independent of $\{z(s),\ 0 \le s \le t\}$.

The proof of Lemma 6.1 is based upon the fact that the error co-variance matrices for the linear Gaussian estimation and smoothing problems are nonrandom [71]. This lemma allows the off-line computation of $P(\sigma_1,\ \sigma_2,\ t)$ via Kwakernaak's equations [72] (for $\sigma_1 \le \sigma_2$)

$$P(\sigma_1,\ \sigma_2,\ t) = P(\sigma_1)\ \Psi'(\sigma_2,\ \sigma_1)$$

$$-P(\sigma_1)\ [\int_{\sigma_2}^t \Psi'(\tau,\sigma_1)\ H'(\tau)\ R^{-1}(\tau)\ H(\tau)\ \Psi(\tau,\ \sigma_2)\,d\tau]\ P(\sigma_2)$$

(6.8)

$$\frac{d}{dt} \Psi(t,\tau) = [F(t) - P(t)H'(t) \ R^{-1}(t)H(t)] \ \Psi(t,\tau); \ \Psi(\tau,\tau) = I \quad (6.9)$$

where the Kalman filter error covariance matrix $P(t) = P(t, t, t)$ is computed by the Riccati equation

$$\dot{P}(t) = F(t) \ P(t) + P(t) \ F'(t) + Q(t) - P(t)H'(t) \ R^{-1}(t) \ H(t) \ P(t)$$

$$(6.10)$$

Lemma 6.2: The conditional cross-covariance satisfies

$$P(\sigma, t, t) = K(t, \sigma) \ P(t) \quad (6.11)$$

where

$$\frac{d}{dt} K'(t,\sigma) = -[F'(t) + P^{-1}(t) \ Q(t)] \ K'(t, \sigma); \ K'(\sigma, \sigma) = I \quad (6.12)$$

This lemma follows easily from some identities in [73]. Equation (6.12) implies that $K'(t, \sigma)$ is a transition matrix satisfying the semi-group property

$$K'(t, \sigma) = K'(t, s) \ K'(s, \sigma) \quad (6.13)$$

These properties of $P(\sigma, t, t)$ are crucial in the derivation of the next theorem, which states the major result concerning the computation of $X(t|t)$.

Theorem 6.1: Consider the system of equations (6.1) - (6.4). The conditional expectation $\hat{X}(t|t)$ is generated by the following finite-dimensional nonlinear filter. First, augment the state of the linear equation (6.1) by writing

$$
\begin{bmatrix} d\xi(t) \\ d\eta(t) \\ d\alpha(t) \\ d\beta(t) \end{bmatrix} = \begin{bmatrix} F(t) & 0 & 0 & 0 \\ I & 0 & 0 & 0 \\ \int_0^t K_3'(t,\sigma)e_1' d\sigma & 0 & -[F'(t) + P^{-1}(t)Q(t)] & 0 \\ 0 & e_1 e_3' & 0 & -[F(t) + P^{-1}(t)Q(t)] \end{bmatrix} \begin{bmatrix} \xi(t) \\ \eta(t) \\ \alpha(t) \\ \beta(t) \end{bmatrix} dt
$$

$$
+ \begin{bmatrix} Q^{1/2}(t) \\ 0 \\ 0 \\ 0 \end{bmatrix} dw(t) \qquad (6.14)
$$

$$
\eta(0) = \alpha(0) = \beta(0) = 0 \qquad (6.15)
$$

$$
dz(t) = [H(t) \quad 0 \quad 0 \quad 0] \begin{bmatrix} \xi(t) \\ \eta(t) \\ \alpha(t) \\ \beta(t) \end{bmatrix} dt + R^{1/2}(t) \, dv(t) \qquad (6.16)
$$

where $K(t, \sigma)$ is given by (6.12), and $P(t)$ is given by (6.10). Here $K_j(t, \sigma)$ denotes the $j^{\underline{th}}$ row of $K(t, \sigma)$ and e_i is the $i^{\underline{th}}$ unit vector in R^3.

Then the Kalman filter [28] for (6.14) - (6.16) yields the conditional expectations $\hat{\xi}(t|t)$, $\hat{\eta}(t|t)$, $\hat{\alpha}(t|t)$, and $\hat{\beta}(t|t)$. The conditional expectation $\hat{X}(t|t)$ is computed by

$$
\hat{X}(t|t) = \begin{bmatrix} 1 & \hat{\eta}_1(t|t) & \hat{\eta}_2(t|t) + \hat{\gamma}(t|t) \\ 0 & 1 & \hat{\eta}_3(t|t) \\ 0 & 0 & 1 \end{bmatrix} \qquad (6.17)
$$

where

$$d\hat{\gamma}(t|t) = [\hat{\xi}_1(t|t) \ \hat{\eta}_3(t|t) + \int_0^t P_{13}(t, \sigma, t) \ d\sigma] \ dt$$

$$+ [\hat{\alpha}'(t|t) + \hat{\beta}'(t|t)] \ P(t) \ H'(t) \ R^{-1}(t) \ [dz(t) - H(t) \ \hat{\xi}(t|t)dt]$$

$$\gamma(0|0) = 0 \tag{6.18}$$

Proof: All the terms in (6.17) result from <u>linear</u> filtering theory [28], except for $\hat{\gamma}(t|t)$. In the derivation of (6.18) we will frequently use a version of Fubinis theorem [74] which allows us to interchange integrals with conditional expectations. Notice that

$$\hat{\gamma}(t|t) = E^t[\int_0^t \int_0^{\sigma_1} \xi_1(\sigma_1) \ \xi_3(\sigma_2) \ d\sigma_2 \ d\sigma_1] \tag{6.19}$$

Thus if we define γ by

$$d\gamma(t) = [\xi_1(t) \int_0^t \xi_3(\sigma) \ d\sigma] \ dt$$

$$= \xi_1(t) \ \eta_3(t) \ dt \tag{6.20}$$

$$\gamma(0) = 0$$

then Kushner's equation [28] yields the conditional expectation

$$d\hat{\gamma}(t|t) = E^t[\xi_1(t) \ \eta_3(t)] \ dt$$

$$+ \{E^t[\gamma(t)\xi'(t)] - \hat{\gamma}(t|t) \ \hat{\xi}(t|t)\} \ H'(t)R^{-1}(t) \ [dz(t) - H(t) \ \hat{\xi}(t|t) \ dt] \tag{6.21}$$

From the definition of $P(\sigma_1, \sigma_2, t)$ it follows that

$$E^t[\xi_1(t)\; \eta_3(t)] = \hat{\xi}_1(t|t)\; \hat{\eta}_3(t|t) + \int_0^t P_{13}\;(t,\; \sigma,\; t)\; d\sigma \qquad (6.22)$$

Considering the second term in (6.21),

$$E^t[\gamma(t)\; \xi_i(t)] - \hat{\gamma}(t|t)\; \hat{\xi}_i(t|t)$$

$$= \int_0^t \int_0^{\sigma_1} \{E^t[\xi_1(\sigma_1)\; \xi_3(\sigma_2)\; \xi_i(t)] - \hat{\xi}_i(t|t)\; E^t[\xi_1(\sigma_1)\; \xi_3(\sigma_2)]\}\; d\sigma_2 d\sigma_1$$

$$= \int_0^t \int_0^{\sigma} \{P_{1i}(\sigma_1,\; t,\; t)\; E^t[\xi_3(\sigma_2)] + P_{3i}(\sigma_2,\; t,\; t)\; E^t[\xi_1(\sigma_1)]\}\; d\sigma_2 d\sigma_1$$

$$(6.23)$$

where the last term in (6.23) is a result of the definition of $P(\sigma_1,\; \sigma_2,\; t)$ and the expansion of the third order moment of a Gaussian distribution [75].

Thus by Lemmas 6.1 and 6.2

$$E^t[\gamma(t)\xi'(t)] - \hat{\gamma}(t|t)\; \hat{\xi}'(t|t)$$

$$= \int_0^t \int_0^{\sigma_1} \{K_1(t,\; \sigma_1)\; E^t[\xi_3(\sigma_2)] + K_3(t,\; \sigma_2)\; E^t[\xi_1(\sigma_1)]\}\; P(t)\; d\sigma_2 d\sigma_1$$

$$= E^t[\int_0^t K_1(t,\; \sigma_1)\; \eta_3(\sigma_1)\; d\sigma_1 + \int_0^t \xi_1(\sigma_1) \int_0^{\sigma_1} K_3(t,\; \sigma_2)\; d\sigma_2 d\sigma_1]\; P(t)$$

$$(6.24)$$

$$= (E^t[\beta'(t)] + E^t[\sigma'(t)])\; P(t) \qquad (6.25)$$

The fact that equation (6.14) for α and β provides a realization of the argument in (6.24) is a direct consequence of (6.12)

Further insight into the structure of the optimal nonlinear filter of Theorem 6.1 is provided by the equivalent formulation

$$d\hat{X}(t|t) = \{(\sum_{i=1}^{3} A_i \hat{\xi}_i(t|t)) \; \hat{X}(t|t) + E_{13} \int_0^t P_{13}(t, \sigma, t) \; d\sigma\} \; dt$$

$$+ \left\{ \int_0^t \begin{bmatrix} 0 & K_1(t, \sigma) & K_2(t, \sigma) \\ 0 & 0 & K_3(t, \sigma) \\ 0 & 0 & 0 \end{bmatrix} d\sigma + (\hat{\alpha}'(t|t) + \hat{\beta}'(t|t))E_{13} \right\}$$

$$\cdot P(t)H'(t) \; R^{-1}(t) \; [dz(t) - H(t) \; \hat{\xi}(t|t)dt] \tag{6.26}$$

where E_{ij} has a 1 as its (i, j) th element and zeros elsewhere. Thus the filter for $\hat{X}(t|t)$ contains a model of the original system (6.2) driven by the innovations process $d\nu(t) = dz(t) - H(t) \; \hat{\xi}(t|t)$ and the outputs of a linear filter (see the block diagram in Figure 6.1).

In addition to providing the optimal minimum variance estimate, it can be shown that the filter of Theorem 6.1 contains fewer states than the extended Kalman filter [28] for the same system. This is due to the on-line computation of the "approximate" error covariance matrix in the extended Kalman filter.

Theorem 6.1 can be extended in various ways. First, it can be extended to systems of the form (6.1), (6.3), and

$$\dot{Y}(t) = \begin{bmatrix} a_1(t) & \xi_1(t) & \xi_2(t) \\ 0 & a_2(t) & \xi_3(t) \\ 0 & 0 & a_3(t) \end{bmatrix} Y(t); \quad Y(0) = I \tag{6.27}$$

with nonrandom diagonal terms a_i ; such systems evolve on the solvable Lie group $G_s(3)$. Second, the results are valid for systems evolving on
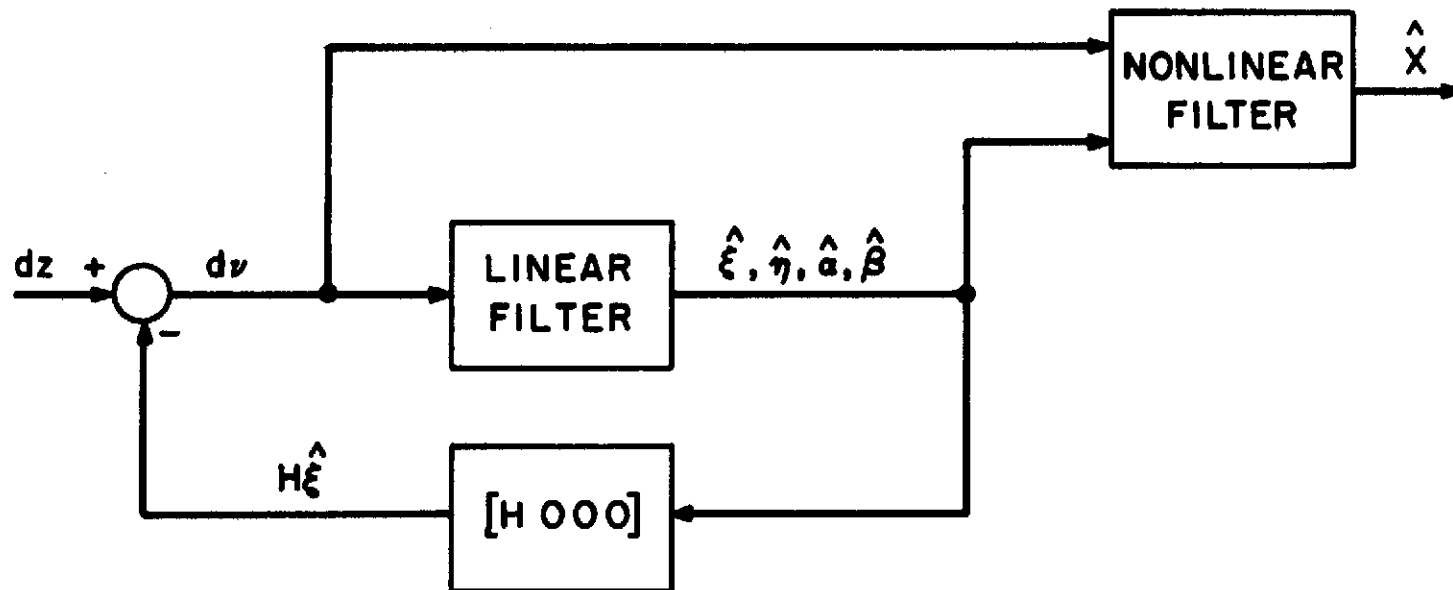
Fig. 6.1   Block Diagram for Optimal Nonlinear Filtering on $G_N(3)$

$G_N(k)$ and for systems with nonrandom diagonal terms evolving on $G_s(k)$. Finally, the results may be extended to observations of the form of (2.12) or (2.13); in these cases, it can be shown [20], [26] that the observations are, in a certain sense, "conditionally" linear. For all of these extensions of Theorem 6.1, it can be shown that the optimal minimum variance filter is finite dimensional.

VII. The Use of Harmonic Analysis in Suboptimal Filter Design

A very important result of harmonic analysis states that the eigenfunctions of the Laplace—Betrami operator on a compact manifold M are a complete set of functions in $L^2(M, \mu)$ (where μ = Haar measure) [43], [76]. In this section we will develop suboptimal estimation techniques for bilinear systems evolving on compact manifolds by employing an approximation to the conditional density which is based on these eigenfunctions.

First consider Example 1 of Section III which evolves on the Lie group S^1, described by (3.4) - (3.5) or (3.7) - (3.8). As discussed in [24], the optimal (constrained least-squares) filter is described as follows. The conditional probability density of θ given $\{z(s), 0 \leq s \leq t\}$ may be expanded in the Fourier series (notice that the trigonometric polynomials are eigenfunctions of the Laplacian on S^1)

$$p(\theta, t) = \sum_{n=-\infty}^{+\infty} c_n(t) \, e^{in\theta} \tag{7.1}$$

where

$$c_n(t) = \frac{1}{2\pi} E[e^{-in\theta(t)} | z(s), 0 \leq s \leq t]$$

$$= b_n(t) - ia_n(t) \tag{7.2}$$

Then

$$dc_n(t) = -[inw_c + \frac{n^2}{2} q(t)]c_n(t)dt$$

$$+ \left[\frac{(c_{n-1}(t) - c_{n+1}(t))}{2i} + 2\pi c_n(t) \operatorname{Im}(c_1(t))\right] \left[\frac{dz(t) + 2\pi \operatorname{Im}(c_1(t))dt}{r(t)}\right]$$

$$\tag{7.3}$$

$$\tilde{\theta}(t) = \tan^{-1}(a_1(t)/b_1(t))$$

(7.4)

Since $c_o = \frac{1}{2\pi}$ and $c_{-n} = c_n^*$ (where * denotes the complex conjugate), we need only solve (7.3) for $n \geq 1$. The structure of the optimal filter, which is illustrated in Figures 7.1 and 7.2, deserves further comment (recall that $c_n = b_n - ia_n$). The filter consists of an infinite bank of filters, the nth of which is essentially a damped oscillator, with oscillator frequency nw_c, together with nonlinear couplings to the other filters and to the received signal. Notice, however, that the equation for c_n is coupled only to the filters for c_1, c_{n-1}, and c_{n+1}. This fact will play an important part in our approximation.

In order to construct a finite-dimensional suboptimal filter, we wish to approximate the conditional density (7.1) by a density determined by a finite set of parameters. Several examples of "assumed density" approximations for this problem are discussed in [16], [25], but the most useful involves the assumption that $p(\theta, t)$ is a folded normal density (see [16], [17], [24]) with mode $\eta(t)$ and "variance" $\gamma(t)$:

$$p(\theta, t) = \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} e^{-n^2\gamma(t)/2} e^{in(\theta-\eta(t))} = F(\theta; \eta(t), \gamma(t))$$

(7.5)

This density is related to the normal density in the following way: if x is a real random variable with density $N(\alpha;\eta,\gamma)$, then $\theta = x \mod 2\pi$ has the density

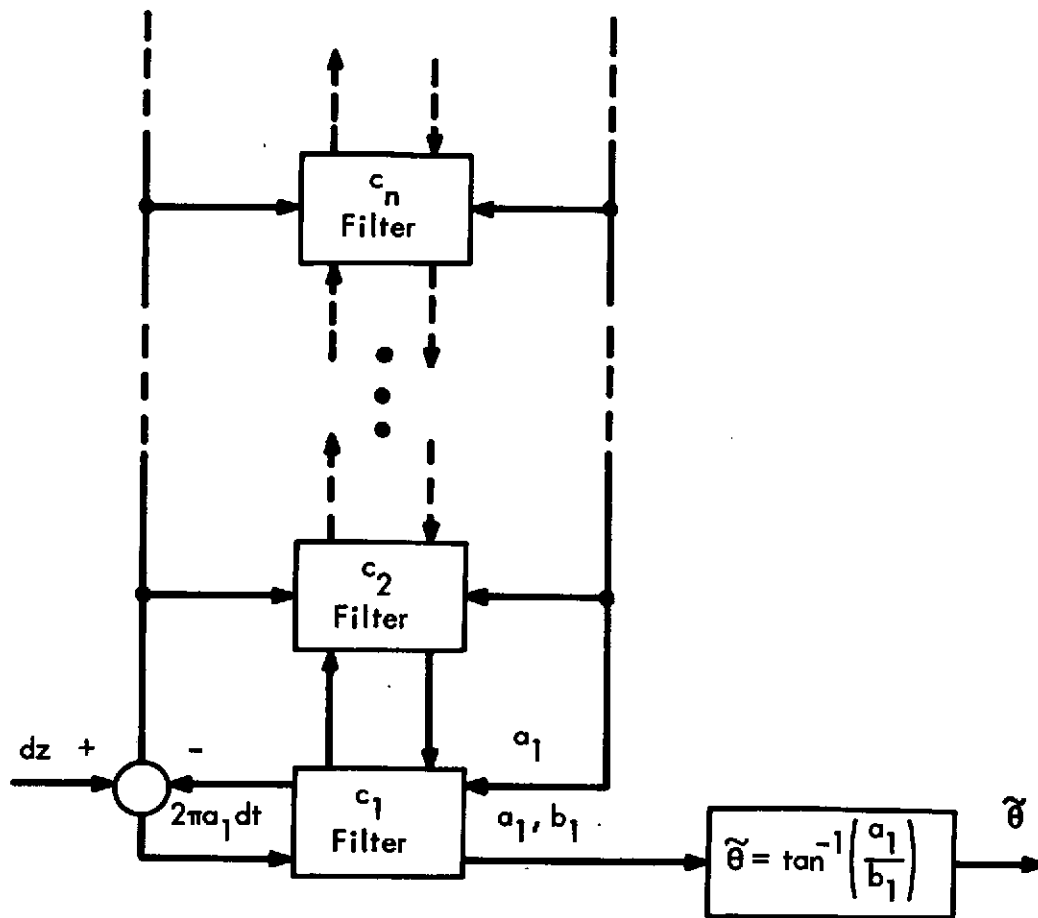$$F(\alpha;\eta,\gamma) = \sum_{n=-\infty}^{+\infty} N(\alpha + 2n\pi;\eta,\gamma)$$

(7.6)

Fig. 7.1   Illustrating the Form of the Infinite Dimensional Optimal
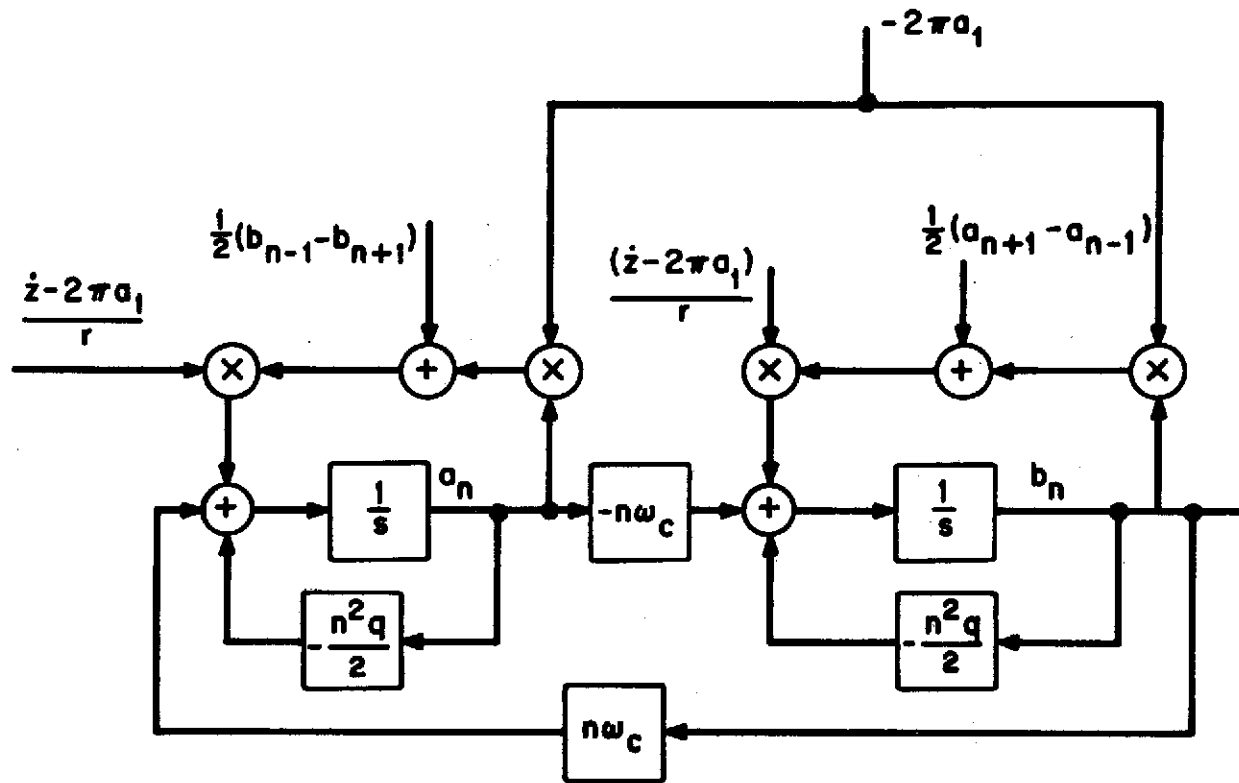Filter of Example 1, Section III

Fig. 7.2   Diagram of the c_n Filter shown in Fig. 7.1

--i.e., we "fold" $N(\alpha;\eta,\gamma)$ around the circle. We note that the folded normal density is the solution of the standard diffusion equation on the circle (i.e., it is the density for S^1 Brownian motion processes) and is as important a density on S^1 as the normal is on R^1.

In this case, if c_1 has been computed and if $p(\theta, t)$ satisfies (7.5), then c_{N+1} can be computed (for any N) from the equation

$$c_{N+1} = (2\pi)^{(N+1)^2-1} |c_1|^{N(N+1)} c_1^{(N+1)} \qquad (7.7)$$

Thus we can truncate the bank of filters described by (7.3) by approximating c_{N+1} by (7.7) and substituting this approximation into the equation for c_N. This was done for N=1 in [25]; the resulting suboptimal filter equations are

$$\dot{a}_1 = (w_c b_1 - \frac{q}{2} a_1) + \frac{(\dot{z}-2\pi a_1)}{r} [\frac{1}{2} (\frac{1}{2\pi} - 8\pi^3 (b_1^4 - a_1^4)) - 2\pi a_1^2]$$

$$(7.8)$$

$$\dot{b}_1 = -(w_c a_1 + \frac{q}{2} b_1) + \frac{(\dot{z}-2\pi a_1)}{r} [8\pi^3 a_1 b_1 (a_1^2 + b_1^2) - 2\pi a_1 b_1] \quad (7.9)$$

$$\tilde{\theta} = \tan^{-1}(a_1/b_1) \qquad (7.10)$$

In [25], this Fourier coefficient filter (FCF) was compared to a phase-lock loop [32] and to the Gustafson-Speyer "state-dependent noise filter" (SDNF) [34]. The FCF performed consistently better than the other systems, although the SDNF performance was quite close.

Similar analyses and "assumed density" approximations can be applied to the other examples in Section III; the reader is referred to

[25], [28], [35], and Section VIII of this paper.

Motivated by the success of the previous example evolving on S^1, we will now extend these results to the system (4.19) - (4.20) evolving on the sphere S^2. Following [3], [9], [10], we define the $\binom{p+2}{p}$-vector $x^{[p]}$ consisting of the pth order moments (homogeneous polynomials) in (x_1, x_2, x_3):

$$\sqrt{\binom{p}{p_1}\binom{p-p_1}{p_2}\binom{p-p_1-p_2}{p_3}} \ x_1^{p_1} x_2^{p_2} x_3^{p_3} \ ; \quad \sum_{i=1}^{3} p_i = p; \quad p_i \geq 0 \qquad (7.11)$$

If y satisfies the linear differential equation

$$\dot{y}(t) = Ay(t) \qquad (7.12)$$

then $y^{[p]}$ satisfies a linear differential equation

$$\dot{y}^{[p]}(t) = A^{[p]} y^{[p]}(t) \qquad (7.13)$$

We regard this as the definition of $A^{[p]}$. It can be shown that if x satisfies (4.19), then $x^{[p]}$ satisfies

$$dx^{[p]}(t) = \left\{ \sum_{i=1}^{3} f_i(t) \ A_i^{[p]} + \sum_{i,j=1}^{3} Q_{ij}(t) \ A_i^{[p]} A_j^{[p]} \right\} x^{[p]}(t) \ dt$$

$$+ \sum_{i=1}^{3} A_i^{[p]} \ x^{[p]}(t) \ dw_i(t) \qquad (7.14)$$

The optimal (constrained least-squares) filter is given by the infinite set of coupled equations

$$dE^t[x^{[p]}(t)] = \left\{ \sum_{i=1}^{3} f_i(t) A_i^{[p]} + \sum_{i,j=1}^{3} Q_{ij}(t) A_i^{[p]} A_j^{[p]} \right\} E^t[x^{[p]}(t)] dt$$

$$+ \{E^t[x^{[p]}(t)x'(t)] - E^t[x^{[p]}(t)] E^t[x'(t)]\}$$

$$\cdot H'(t) R^{-1}(t) [dz(t) - H(t) E^t[x(t)] dt \qquad (7.15)$$

$$\tilde{x}(t) = \frac{E^{(t)}[x^{[1]}(t)]}{||E^t[x^{[1]}(t)]||} = \frac{\hat{x}(t|t)}{||\hat{x}(t|t)||} \qquad (7.16)$$

The structure of this filter is quite similar to that of (7.3) – (7.4) —
i.e., it consists of an infinite bank of filters, and the filter for $x^{[p]}$
is coupled only to those for x and $x^{[p+1]}$.

The similarities with the previous example are further illuminated
by considering the spherical harmonics [77], [78], which constitute an
orthonormal basis for the eigenfunctions of the Laplacian on S^2. We
introduce polar coordinates (θ, ϕ) on S^2, where $0 \le \theta \le \pi$, $0 \le \phi \le 2\pi$,
by defining

$$x_1 = \cos \theta; \quad x_2 = \sin \theta \cos \phi; \quad x_3 = \sin \theta \sin \phi \qquad (7.17)$$

The normalized spherical harmonics $Y_{\ell m}(\theta, \phi)$ are defined by [77]

$$Y_{\ell m}(\theta, \phi) = (-1)^m \left[\frac{(\ell-m)!}{(\ell+m)!} \frac{(2\ell+1)}{4\pi} \right]^{1/2} P_{\ell m}(\cos \theta) e^{im\phi} \qquad (7.18)$$

$$Y_{\ell,-m}(\theta, \phi) = (-1)^m Y_{\ell m}^*(\theta,\phi) \qquad (7.19)$$

for $\ell = 0, 1, \ldots$ and $m = 0, 1, \ldots, \ell$, where $P_{\ell m}(\cos \theta)$ are the associated

Legendre functions.

According to the remark at the beginning of this section, any square-integrable function on S^2 can be expanded in a series of the form

$$f(\theta,\phi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} C_{\ell m} Y_{\ell m}(\theta, \phi) \qquad (7.20)$$

where

$$C_{\ell m} = \int_{0}^{2\pi} \int_{0}^{\pi} Y_{\ell m}^{*}(\theta,\phi) \ f(\theta,\phi) \ \sin\theta d\theta d\phi \qquad (7.21)$$

In particular, the conditional probability density function for our estimation problem on S^2 can be expanded via the series

$$p(\theta,\phi,t) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} C_{\ell m}(t) \ Y_{\ell m}(\theta,\phi) \qquad (7.22)$$

where

$$C_{\ell m}(t) = E[Y_{\ell m}^{*}(\theta(t), \phi(t)) \mid z(s), 0 \leq s \leq t] \qquad (7.23)$$

The set of coefficients $\{C_{\ell m}(t)\}$ for $\ell = 1, \ldots, k$ and $m = -\ell, \ldots, \ell$ is equivalent to the set of conditional moments $\{E^t[x^{[p]}(t)]\}$ for $p = 1, \ldots, k$ (see [3] for a detailed description of this equivalence). Thus a filter which generates $\{C_{\ell m}(t)\}$ is equivalent to (7.15); in fact, it can be shown that such a filter consists of an infinite set of equations (similar in form to (7.15)), and the equation for C_{LM} is coupled only to those for $\{C_{\ell m}; \ \ell=L-1, L, L+1; \ m=-\ell,\ldots,\ell\}$. The optimal estimate is

$$\tilde{x}'(t) = \frac{[\sqrt{2}\ C_{10}(t), \ -(C_{11}(t) + C_{11}^{*}(t)), \ i(C_{11}^{*}(t) - C_{11}(t)]}{\sqrt{2}\ (C_{10}^{2}(t) + 2 \mid C_{11}(t) \mid^{2})^{1/2}} \qquad (7.24)$$

This filter is completely analogous to the S^1 filter (7.3) - (7.4).

We will again employ an "assumed density" approximation in order to obtain a finite-dimensional filter. An S^2-analog of the folded normal density is obtained as the solution to the standard diffusion equation on S^2:

$$\frac{\partial p(\theta,\phi,t)}{\partial t} - \frac{1}{2} \beta(t) \nabla^2 p(\theta,\phi,t) = 0 \qquad (7.25)$$

The Green's function for (7.25) is given by the "bilinear series" [79]

$$G(\theta,\phi,t; \eta,\nu,\tau) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell m}(\theta,\phi) \; Y_{\ell m}^*(\eta,\nu) \; e^{-\ell(\ell+1)\int_{\tau}^{t} \beta(s)ds} \qquad (7.26)$$

This is the solution to (7.25) with initial condition equal to the singular distribution concentrated at (η,ν).

Thus by analogy with (7.5), we assume that the conditional density for the S^2 estimation problem is of the form

$$p(\theta,\phi,t) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell m}(\theta,\phi) \; Y_{\ell m}^*(\eta(t),\nu(t)) \; e^{-\frac{1}{2}\ell(\ell+1)\gamma(t)} \qquad (7.27)$$

In order words, $C_{\ell m}(t)$ (as defined in (7.22) - (7.23) is assumed to be

$$C_{\ell m}(t) = Y_{\ell m}^*(\eta(t), \nu(t)) \; e^{-\frac{1}{2}\ell(\ell+1)\gamma(t)} \qquad (7.28)$$

In this case, if C_{10} and C_{11} have been computed, then

$$\gamma(t) = -\frac{1}{2} \log \left[\frac{4\pi}{3} \left(C_{10}^2(t) + 2|C_{11}(t)|^2 \right) \right] \qquad (7.29)$$

$$\cos \eta(t) = \frac{c_{10}(t)}{[c_{10}^2(t) + 2|c_{11}(t)|^2]^{1/2}} \qquad (7.30)$$

$$\sin \eta(t) = \frac{\sqrt{2} \ | \ c_{11}(t)|}{[c_{10}^2(t) + 2|c_{11}(t)|^2]^{1/2}} \qquad (7.31)$$

If $c_{11}(t) = 0$, then the density is independent of $\nu(t)$; otherwise,

$$e^{2i\nu(t)} = \frac{c_{11}^*(t)}{c_{11}(t)} \qquad (7.32)$$

Then $\{c_{N+1,m}, \ m=-(N+1), \ \ldots, \ N+1\}$ can be computed (for any N) from

$$c_{N+1,m}(t) = Y_{N+1,m}^*(\eta(t), \ \nu(t))e^{-\frac{1}{2}(N+1)(N+2)\gamma(t)} \qquad (7.33)$$

Thus we can truncate the bank of filters for $\{c_{\ell m}\}$ by approximating $\{c_{N+1,m}\}$ via (7.29) - (7.33) and substituting these approximations into the equations for $\{c_{N,m}\}$.

The performance of this suboptimal filter will be tested by simulation, and the results will be presented in a future report. Other "assumed densities" will also be studied, including those resulting from degenerate diffusions or diffusions with unequal drifts around the three axes [3].

The techniques of this section can in theory be extended to any compact manifold by employing the eigenfunctions of the Laplace-Beltrami operator. For example, on the n-sphere S^n these functions are the n-dimensional spherical harmonics [78]. Since SO(3) is isomorphic (as a Lie

group) to $S^3/\{\pm I\}$ (see [22], [43]), suboptimal filters for some of the rotational estimation problems of Section IV can be constructed by means of spherical harmonics on S^3.

VIII. A Cumulant Method for Suboptimal Filter Design

As we have seen, we are able to obtain finite dimensional optimal filters only for certain bilinear estimation problems. In the preceding section we described how harmonic analysis could be used to design high quality estimation systems for processes evolving on spheres and, more generally, on compact manifolds. In this section we describe a design technique which we call the cumulants method. This approach has been considered by several authors [10], [59], [60] and is related to statistical linearization techniques [60] - [62].

We wish to consider the estimation of x(t) given the observation process z, where these processes satisfy

$$dx(t) = a(x(t), t)dt + B(x(t), t)dw(t) \qquad (8.1)$$

$$dz(t) = H(t) x(t)dt + R^{1/2}(t) dv(t) \qquad (8.2)$$

where a is an n-vector and B an nxm matrix of polynomials in the components of x, w and v are independent standard vector Brownian motion processes, and R > 0. All of the random processes considered in the preceding sections are of the form (8.1) (see Section II). We consider only the linear observation process (8.2), but the analysis of this section can also be carried out for bilinear and multiplicative processes of the forms given by (2.12) and (2.13).

We wish to compute the conditional moments

$$m_{k_1 \ldots k_n}(t|t) = E\left[x_1^{k_1}(t) \ldots x_n^{k_n}(t) | z(s), 0 \le s \le t\right] \qquad (8.3)$$

Recall that these quantities were of direct interest in the various estimation problems described earlier. For any twice continuously

differentiable real-valued function ϕ of x, let

$$\hat{\phi}(t|t) = E[\phi(x(t)) \mid z(s), 0 \leq s \leq t] \tag{8.4}$$

We then have Kushner's stochastic differential equation for $\hat{\phi}$ [63], [64]:

$$d\hat{\phi} = \{\widehat{\phi_x a} + \frac{1}{2} \text{ tr } [\widehat{BB'\phi_{xx}}]\} \, dt$$

$$+ [\widehat{\phi Hx} - \hat{\phi}\hat{Hx}]' \, R^{-1}[dz - H\hat{x}] \tag{8.5}$$

Note that if ϕ is of the form $x_1^{k_1}...x_n^{k_n}$, the right-hand side of (8.5) consists solely of various conditional moments as in (8.3) (see for example (7.15)). This is a direct consequence of the fact that the right-hand sides of (8.1) and (8.2) contain only polynomial functions of x.

A major complication with these equations (when $\hat{\phi}$ is a moment (8.3)) is that they are all coupled together (as in (7.15)). The reason for this is the following: let the order of a moment $m_{k_1...k_n}$ be the sum of the k_i. Then, because of the polynomial nature of the various co-efficients, the orders of some of the terms on the right-hand side of (8.5) are at least one higher than the order of ϕ. Consider the scalar example

$$dx(t) = \alpha x^2(t)dt + \beta x(t)dw(t) \tag{8.6}$$

$$dz(t) = x(t)dt + dv(t) \tag{8.7}$$

Then

$$dm_1 = \alpha m_2 dt + [m_2 - m_1^2] \, [dz - m_1 dt] \tag{8.8}$$

$$dm_2 = [2\alpha m_3 + \beta^2 m_2]dt + [m_3 - m_2 m_1] \, [dz - m_1 dt] \tag{8.9}$$

$$dm_3 = [3\alpha m_4 + 3\beta^2 m_3]dt + [m_4 - m_3 m_1] \; [dz - m_1 dt] \qquad (8.10)$$

From the preceding comments, it is clear that the implementation of these equations must involve an approximation -- i.e., a truncation of the infinite set of equations. For several reasons, the direct truncation method -- setting to zero all moments greater than some given order -- can cause difficulties. First of all, there is no reason to expect the higher moments to be small, and in many cases (such as the Gaussian case) the sequence of moments is unbounded. In addition, if we use the fact that the moments are the coefficients of the Taylor series expansion of the characteristic function of x, the assumption that the higher moments are zero corresponds to assuming that the density for x is a sum of derivatives of Dirac delta functions [59].

As suggested in [59] and [60], a more useful set of variables is the set of cumulants, which are the coefficients of the Taylor series expansion of the logarithm of the characteristic function. For the present discussion we limit ourselves to the scalar case, although the vector case can be handled similarly. For a scalar variable the kth cumulant is a polynomial in the first k moments and, in fact, the first k moments and the first k cumulants contain precisely the same information. For example, the first 4 cumulants of the scalar process x(t) in (8.6) are

$$k_1(t) = m_1(t) \qquad (8.11)$$

$$k_2(t) = m_2(t) - m_1^2(t) \qquad (8.12)$$

$$k_3(t) = m_3(t) - 3m_1(t)m_2(t) + 2m_1^3(t) \qquad (8.13)$$

$$k_4(t) = m_4(t) - 3m_2^2(t) - 4m_1(t)m_3(t) + 12m_1^2(t)m_2(t) - 6m_1^4(t) \qquad (8.14)$$

and we see that k_1 is the mean and k_2 is the variance.

As discussed in [59], a reasonable procedure is to set to zero all cumulants of order higher than some given number. Note that assuming $k_i = 0 \, \forall \, i \geq 3$ is <u>equivalent</u> to assuming that $x(t)$ is Gaussian! Note also that if we take $k(t) = 0$, $i \geq M$, we can obtain equations for the corresponding higher moments of $x(t)$. In this way, we can effectively truncate the infinite set of moment equations. For instance, if we return to the scalar example (8.6), (8.7), and if we assume $k_i(t) = 0 \, \forall \, i \geq 4$ equations (8.8) and (8.9) are unchanged and (8.10) is replaced by

$$dm_3 = [3\alpha\eta(m_1, m_2, m_3) + 3\beta^2 m_3]dt$$

$$+ [\eta(m_1, m_2, m_3) - m_3 m_1] [dz - m_1 dt] \qquad (8.15)$$

$$\eta(m_1, m_2, m_3) = 3m_2^2 + 4m_1 m_3 - 12m_1^2 m_2 + 6m_1^4 \qquad (8.16)$$

We note that this technique can be extended to the general vector problem (8.1), (8.2) with no conceptual, but some bookkeeping, difficulties. For instance, the cumulants approach provides an alternative to the Fourier series methods described in Section VII for the design of phase tracking and demodulation systems. An open question related to the cumulants method is that of performance analysis -- e.g., how does performance improve as we retain more cumulants and do we achieve the optimal performance in the limit?

IX. Conclusions

In this paper we have discussed the practical importance and the mathematical analysis of several classes of bilinear estimation problems. We have seen that such mathematical models arise in a wide variety of applications, and we refer the reader to the references for further verification of this fact. We have also indicated how such estimation problems can be solved. In some cases, best explained in a Lie-theoretic framework, we have seen that finite-dimensional optimal estimation equations can be derived. For other problems the tools of harmonic analysis have turned out to be most appropriate and extremely useful. Finally, a general but untried approximation method based on the truncation of the cumulants of a random process has been described. In conclusion, the class of stochastic bilinear systems is not only an appealing class of systems from an applications point of view, but it also is a highly structured class of systems for which analysis nearly as detailed and successful as that for linear systems is possible.

References

1. R.W. Brockett, "System Theory on Groups Manifolds and Coset Spaces," SIAM J. Control, Vol. 10, No. 2, May 1972, pp. 265-284.

2. R.W. Brockett, "On the Algebraic Structure of Bilinear Systems," in Theory and Applications of Variable Structure Systems, R. Mohler and A. Ruberti, eds., Academic Press, New York, 1972.

3. R.W. Brockett, "Lie Algebras and Lie Groups in Control Theory," in Geometric Methods in System Theory, D.Q. Mayne and R.W. Brockett, eds., Reidel Pub. Co., The Netherlands, 1973.

4. V. Jurdjevic and H.J. Sussmann, "Control Systems on Lie Groups," J. Diff. Eqns., Vol. 12, No. 2, 1972, pp. 313-329.

5. R.M. Hirschorn, "Controllability in Nonlinear Systems," in Geometric Methods in System Theory, D.Q. Mayne and R.W. Brockett, Eds., Reidel Pub. Co., The Netherlands, 1973.

6. A. Isidori and A. Ruberti, "Realization Theory of Bilinear Systems," in Geometric Methods in System Theory, D.Q. Mayne and R.W. Brockett, eds., Reidel Pub. Co., The Netherlands, 1973.

7. R. Mohler, "Bilinear Structures and Man," in Theory and Applications of Variable Structure Systems, R. Mohler and A. Ruberti, eds., Academic Press, New York, 1972.

8. C. Bruni, G. DiPillo, and G. Koch, Bilinear Systems: An Appealing Class of "Nearly Linear" Systems in Theory and Applications, Universita'di Roma, Istituto di Automatica, R. 2-29, December, 1972

9. R.W. Brockett, "Lie Theory and Control Systems on Spheres," SIAM J. Appl. Math., Vol. 25, No. 2, Sept. 1973, pp. 213-225.

10. A.S. Willsky and S.I. Marcus, "Analysis of Bilinear Noise Models in Circuits and Devices," Monograph of the Colloquium on the Application of Lie Group Theory to Nonlinear Network Problems, 1974 IEEE International Symposium on Circuits and Systems, San Francisco, Calif., April 1974.

11. G.L. Blankenship, "Perturbation Theory for Stochastic Ordinary Differential Equations with Applications to Optical Waveguide Analysis," Monograph of the Colloquium on the Application of Lie Groups Theory to Nonlinear Network Problems 1974 IEEE International Symposium on Circuits and Systems, San Francisco, Calif., April 1974.

12. H.P. McKean, Jr., Stochastic Integrals, Academic Press, N.Y., 1969.

13. H.P. McKean, Jr., "Brownian Motion on the 3-Dimensional Rotation Groups," Mem. Coll. Sci. Kyoto Univ., Vol. 33, 1960, pp. 25-38.

14. J.T. Lo, "Bilinear Stochastic Systems and Finite Dimensional Sensor Orbits," Eighth Princeton Conference on Information Sciences and Systems, Princeton, New Jersey, March 28-29, 1974.

15. G. Koch, "Some Results on Stochastic Bilinear Systems," Fourth Symposium on Nonlinear Estimation and Its Applications," San Diego, Calif., Sept. 10-12, 1973.

16. A.S. Willsky, Dynamical Systems Defined on Groups: Structural Properties and Estimation, Ph.D. Thesis, Dept. of Aeronautics and Astronautics, M.I.T., Cambridge, Mass., June 1973.

17. J.T. Lo and A.S. Willsky, "Estimation for Rotational Processes with One Degree of Freedom I: Introduction and Continuous Time Processes," IEEE Trans. on Aut. Cont., to appear.

18. A.S. Willsky and J.T. Lo, "Estimation for Rotational Processes with One Degree of Freedom II: Discrete Time Processes," IEEE Trans. on Aut. Control, to appear.

19. A.S. Willsky and J.T. Lo, "Estimation for Rotational Processes with One Degree of Freedom III: Applications and Implementation," IEEE Trans. on Aut. Control, to appear.

20. A.S. Willsky, "Estimation and Detection of Signals in Multiplicative Noise," submitted to IEEE Trans. on Aut. Control.

21. A.S. Willsky, "Some Estimation Problems on Lie Groups," in Geometric Methods in System Theory, R.W. Brockett and D.Q. Mayne, eds., Reidel Pub. Co., The Netherlands, 1973.

22. A.S. Willsky, "Some Results on the Estimation of the Angular Velocity and Orientation of a Rigid Body," submitted to Automatica.

23. J.T. Lo and A.S. Willsky, "Stochastic Control of Rotational Processes with One Degree of Freedom," submitted to SIAM J. Control.

24. A.S. Willsky, "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication I: Analysis," IEEE Trans. on Inf. Theory, to appear.

25. A.S. Willsky, "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication II: Implementation," IEEE Trans. on Inf. Theory, to appear.

26. J.T. Lo, "Signal Detection on Lie Groups," IEEE Trans. on Inf. Theory, to appear.

27. B.W. Stuck, Space Satellite Dynamics with Applications to Sunlight Pressure Attitude Control, Ph.D. Thesis, Dept. of Electrical Engineering, M.I.T., Cambridge, Mass., June 1972.

28. A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.

29. E. Wong, Stochastic Processes in Information and Dynamical Systems, McGraw-Hill Book Co., New York, 1971.

30. E.J. McShane, "Toward a Stochastic Calculus, I," Proc. N.A.S., Vol. 63, 1969, pp. 275-280.

31. E.J. McShane, "Toward a Stochastic Calculus, II," Proc. N.A.S., Vol. 63, 1969, 1084-1087.

32. H.L. Van Trees, Detection, Estimation, and Modulation Theory, Part II Nonlinear Modulation Theory, John Wiley and Sons, Inc., New York, 1971

33. A.J. Viterbi, Principles of Coherent Communication, McGraw-Hill Book Co., New York, 1966.

34. D.E. Gustafson and J.L. Speyer, "Linear and Asymptotic Minimum Variance Filters Applied to Phase-Lock Loops," IEEE Trans. on Comm., to appear.

35. R.S. Bucy, C. Hecht, and K.D. Senne, "New Methods for Nonlinear Filtering," Revue Francaise d'Automatique, Informatique et de Recherche Operationnelle, Feb. 1973, pp. 3-54.

36. J.J. Mallinckrodt, R.S. Bucy, and S.Y. Chang, Final Project Report for a Design Study for an Optimal Non-Linear Receiver/Demodulator, NASA Contract NAS 5-10789, Goddard Space Flight Center, Maryland, 1970.

37. J.M. Wozencraft and I.M. Jacobs, Principles of Communication Engineering, John Wiley and Sons, Inc., New York, 1965.

38. J.A. Pierce, W. Palmer, A.D. Watt, and R.H. Woodward, Omega -- A World-Wide Navigational System Specification and Implementation -- Second Revision, Pickard and Burns Electronics, Waltham, Mass., May 1966.

39. M.G. Crosby, "Frequency-Modulation Noise Characteristics," Proc. of the IRE, Vol. 25, April 1937, pp. 472-514.

40. L.S. Cutler and C.L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards, Proc. of the IEEE, Vol. 54, No. 2, Feb. 1966, pp. 136-154.

41. M. Kayton and W.R. Fried, eds., _Avionics Navigation Systems_, John Wiley and Sons, Inc., New York, 1969.

42. C. Chevalley, _Theory of Lie Groups_, Princeton University Press, Princeton, 1946.

43. F. Warner, _Foundations of Differential Manifolds and Lie Groups_, Scott, Foresman and Co., Glenview, Ill., 1971.

44. E.V. Hoversten, R.O. Harger, and S.J. Halme, "Communication Theory for the Turbulent Atmosphere," _Proc. of the IEEE_, Vol. 58, No. 10, Oct. 1970, pp. 1626-1650.

45. R.L. Halfman, _Dynamics Vol. I_, Addison-Wesley Pub. Co., Inc., Reading, Mass., 1962.

46. W. Wrigley, W. Hollister, and W. Denhard, _Gyroscopic Theory, Design, and Instrumentation_, The M.I.T. Press, Cambridge, Mass., 1969.

47. A.E. Bryson, Jr. and W. Kortum, "Estimation of the Local Attitude of Orbiting Spacecraft," _Automatica_, Vol. 7, 1971, pp. 163-180.

48. B.W. Stuck, "A New Method for Attitude Estimation," presented at the AAS/AIAA Astrodynamics Conference, Vail, Colorado, July 16-18, 1973.

49. R.W. Brockett and H.J. Sussmann, "Tangent Bundles of Homogeneous Spaces are Homogeneous Spaces," _Proc. Amer. Math. Soc._, Vol. 35, No. 2, Oct. 1972, pp. 550-551.

50. H. Samelson, _Notes on Lie Algebras_, Van Nostrand Reinhold Co., New York, 1969.

51. J. Wei and E. Norman, "Lie Algebraic Solution of Linear Differential Equations," _J. Mathematical Physics_, Vol. 4, 1963, pp. 575-581.

52. J. Wei and E. Norman, "On Global Representations of the Solutions of Linear Differential Equations as a Product of Exponentials," _Proc. Amer. Math. Soc._, Vol. 15, 1964, pp. 327-334.

53. A.E. Bryson, Jr. and D.E. Johansen, "Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise," _IEEE Trans. Aut. Control_, Vol. AC-10, 1965, pp. 4-10.

54. J.T. Lo, "Signal Detection of Rotational Processes and Frequency Demodulation," _Proc. of the 1973 IEEE Conference on Decision and Control_, San Diego, December 1973.

55. R.W. Brockett and J.R. Wood, "Electrical Networks Containing Controlled Switches," _Monograph of the Colloquium on the Application of Lie Group Theory to Nonlinear Network Problems_, 1974 IEEE International Symposium on Circuits and Systems, San Francisco, Calif., April 1974.

56. T.G. Dankel, Jr., "Mechanics on Manifolds and the Incorporation of Spin into Nelson's Stochastic Mechanics," Arch. Rat. Mech. Anal., Vol. 37, No. 3, 1970, pp. 192-221.

57. J.E. Potter and E.J. Frey, Rotation Invariant Probability Distributions on the Surface of a Sphere, with Applications to Geodosy, M.I.T. Experimental Astronomy Lab. Report No. RE-27, May 1967.

58. B.D. Tapley and D.S. Ingram, "Orbit Determination in the Presence of Unmodeled Accelerations," IEEE Trans. on Aut. Control, Vol. AC-18, August 1973, pp. 369-373.

59. R.L. Stratonovich, Topics in the Theory of Random Noise, Volume I, Translated by R.A. Silverman, Gordon and Breach, New York, 1963.

60. T. Nakamizo, "On the State Estimation for Non-Linear Dynamic Systems," Int. J. Control, Vol. 11, No. 4, 1970, pp. 683-695.

61. R.J. Phaneuf, Approximate Nonlinear Estimation, Ph.D. Thesis, Dept. of Aeronautics and Astronautics, M.I.T., Cambridge, Mass., May 1968.

62. Y. Sunahara, Technical Report 67-8, Center for Dynamical Systems, Div. of Applied Math., Brown Univ., Providence, R.I., 1967.

63. A.A. Desalu, Stochastic Modeling and Dynamic Estimation of Air Quality in an Advective-Diffusive Atmosphere, Ph.D. Thesis, Dept. of Electrical Engineering, M.I.T., Cambridge, Mass., 1974.

64. A.A. Desalu, L.A. Gould, and F.C. Schweppe, "Dynamic Estimation of Air Pollution," submitted to IEEE Trans. on Aut. Control.

65. T.J. Tarn, D.L. Elliott, and T. Goka, "Controllability of Discrete Bilinear Systems with Bounded Control," IEEE Trans. on Aut. Control, Vol. AC-18, 1973, pp. 298-301.

66. A. Isidori, "Direct Construction of Minimal Bilinear Realizations from Nonlinear Input-Output Maps," IEEE Trans. on Aut. Control, Vol. AC-18, 1973, pp. 626-631.

67. T. Kailath, "Some Chandrasekar-Type Algorithms for Linear Quadratic Regulators," Proc. of the 1972 IEEE Conf. on Decision and Control, New Orleans, December 1972.

68. A. Lindquist, "A New Algorithm for Optimal Filtering of Discrete-Time Stationary Processes," SIAM J. Control, to appear.

69. A.A. Sagle and R.E. Walde, Introduction to Lie Groups and Lie Algebras, New York, Academic Press, 1973.

70. R.W. Brockett, _Finite Dimensional Linear Systems_, New York, John Wiley, 1970.

71. C.T. Leondes, J.B. Peller, and E.B. Stear, "Nonlinear Smoothing Theory," _IEEE Trans. Syst. Sci. Cyber._, Vol. SSC-6, 1970, p.63.

72. H. Kwakernaak, "Optimal Filtering in Linear Systems with Time Delays," _IEEE Trans. on Aut. Control_, Vol. AC-12, 1967, pp. 169-173.

73. T. Kailath and P. Frost," An Innovations Approach to Least-Squares Estimation, Part II: Linear Smoothing in Additive White Noise," _IEEE Trans. on Aut. Control_, Vol. AC-13, 1968, pp. 655-660.

74. W. Rudin, _Real and Complex Analysis_, New York, McGraw-Hill, 1966.

75. K.S. Miller, _Multidimensional Gaussian Distributions_, New York, John Wiley, 1965.

76. L.H. Loomis, _An Introduction to Abstract Harmonic Analysis_, Princeton, Van Nostrand, 1953.

77. J.D. Talman, _Special Functions -- A Group Theoretic Approach_, New York, Benjamin, 1968.

78. A. Erdelyi, W. Magnus, F.Oberhettinger, and F. Tricomi, _Higher Transcendental Functions_, Vol. II, New York, McGraw-Hill, 1953.

79. R. Dennemeyer, _Introduction to Partial Differential Equations and Boundary Value Problems_, New York, McGraw-Hill, 1968.

80. W. Kohn, T.L. Johnson, and A.S. Willsky, "Analysis of a Mathematical Model for Transport Phenomena," to appear.