TABLE I

Method		Cost: Multiplications	Divisions
Riccati equation		$n^3 + \frac{1}{2} mn^2 + \frac{3}{2} pn^2$	—
Square root methods	Andrews	$\frac{7}{6} n^3 + \frac{3}{2} mn^2 + 2pn^2$	n
	Tapley and Choe	$\frac{4}{3} n^3 + mn^2 + 2pn^2$	n
	equation (12)	$\frac{4}{3} n^3 + \frac{3}{2} mn^2 + \frac{3}{2} pn^2$	n

view: for example, by using the backward Markovian model for $x(s)$ described in [20], and by propagating the square roots $\mathcal{U}$ and $\mathcal{L}$ backwards in time, we can derive a set of backward-smoothing formulas similar to those presented in [14] from the scattering point of view. In fact, this connection even suggests some possible relations between scattering theory and square-root methods, though we shall not pursue these here.

IV. Conclusions

The method used in this paper to obtain the square-root equations has been to substitute directly $P=SS^T$ inside the usual Riccati equation for the error variance. An alternative approach would be to extend to continuous-time the square-root array methods used in discrete-time by Bierman [7] and Morf and Kailath [8]. These methods do not require the introduction of the Riccati equation and have the general advantage of reducing the filtering problem to a state-estimation problem.

However, while discrete-array techniques depend on a simple Gram–Schmidt orthonormalization of the input–output variables (see [8]), the continuous arrays require more sophisticated tools. For example, the Gram–Schmidt orthonormalization technique has to be replaced by the so-called Doob–Meyer decomposition of quasi-martingales, and orthonormal operators have to be substituted for orthonormal matrices. The details will be omitted here.

Appendix
Computational Aspects

It was shown in Section II that (12) has the advantage over the square-root algorithms of Andrews [9] and Tapley and Choe [10] of giving an explicit differential equation for the square root S, a property that simplifies the analog simulation of S. However, if implemented digitally, (12) requires approximately the same number of operations as the algorithms of Andrews and Tapley and Choe, as is indicated in Table I which gives the number of operations per step of integration. Here, n is the number of states, m the number of inputs, p the number of outputs, and only the square-root updates are considered. This comparison also shows that the computational complexity of the square-root algorithms is only slightly greater than the one of the Riccati equation.

Moreover, we note that the square-root algorithms considered here require a similar amount of storage. However, the number of operations and the volume of storage needed are not the only relevant computational aspects, and some further studies would be needed to compare the numerical stability properties of these algorithms (we have shown in Section II that the stability of the Riccati equation implies the stability of (12)) and to see if they require the same step size to retain a predetermined accuracy.

References

[1] J. E. Potter and R. G. Stern, "Statistical filtering of space navigation measurements," in *Proc. 1963 AIAA Guidance and Control Conf.*

[2] G. H. Golub, "Numerical methods for solving linear least-squares problems," *Numer. Math.*, vol. 7, pp. 206–216, 1965.

[3] S. F. Schmidt, "Computational techniques in Kalman filtering," NATO AGARD-139, C. T. Leondes, Ed., pp. 67–103, Feb. 1970.

[4] P. G. Kaminski, "Square root filtering and smoothing for discrete processes," Ph.D. dissertation, Dep. Aero. & Astro., Stanford University, Stanford, CA, Sept. 1971.

[5] P. G. Kaminski and A. E. Bryson, "Discrete square-root smoothing," *Proc. AIAA Guidance and Control Conf.*, Paper 72-844, 1972.

[6] G. J. Bierman, "Sequential square-root filtering and smoothing of discrete linear systems," *Automatica*, vol. 10, pp. 147–158, Mar. 1974.

[7] ——, *Factorization Algorithms for Discrete Sequential Estimation.* New York: Academic, 1977.

[8] M. Morf and T. Kailath, "Square-root algorithms for least-squares estimation," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 487–497, Aug. 1975.

[9] A. Andrews, "A square root formulation of the Kalman covariance equations," *AIAA J.*, vol. 6, pp. 1165–1166, June 1968.

[10] B. D. Tapley and C. Y. Choe, "An algorithm for propagating the square-root covariance matrix in triangular form," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 122–123, Feb. 1976.

[11] G. J. Bierman, "A square-root data array solution of the continuous-discrete filtering problem," *IEEE Trans. Automat. Contr.*, vol. AC-18, pp. 675–676, Dec. 1973.

[12] T. Kailath, "Some new algorithms for recursive estimation in constant linear systems," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 750–760, 1973.

[13] L. Zachrisson, "On optimal smoothing of continuous-time Kalman processes," *Inform. Sci.*, vol. 1, pp. 143–172, 1969.

[14] L. Ljung and T. Kailath, "A unified approach to smoothing formulas," *Automatica*, vol. 12, pp. 147–157, 1976.

[15] R. E. Kalman, "New methods in Wiener filtering theory," in *Proc. First Symposium on Engineering Applications of Random Functions Theory*, J. Bogdanoff and F. Kozin, Eds. New York: Wiley, 1963, pp. 270–388.

[16] D. G. Lainiotis, "Optimal linear smoothing; continuous data case," *Int. J. Control*, vol. 17, pp. 921–930, 1973.

[17] ——, "Partitioned estimation algorithms, II: Linear estimation," *Inform. Sci.*, vol. 7, pp. 317–340, 1974.

[18] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, pp. 1445–1450, Aug. 1965.

[19] T. Kailath and P. Frost, "An innovations approach to least-squares estimation, pt. II: Linear smoothing in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 655–660, Dec. 1968.

[20] L. Ljung and T. Kailath, "Backwards Markovian models for second-order stochastic processes," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 488–491, July 1976.

The Use of Harmonic Analysis in Suboptimal Estimator Design

STEVEN I. MARCUS, MEMBER, IEEE, ALAN S. WILLSKY, MEMBER, IEEE, AND KAI HSU, STUDENT MEMBER, IEEE

***Abstract*—The state estimation problem for bilinear stochastic systems evolving on the spheres S^n, the special orthogonal groups $SO(n)$, and certain other compact Lie groups and homogeneous spaces is considered. The problem is motivated by some applications involving rotational processes in three dimensions. The theory of harmonic analysis on compact Lie groups is used to define assumed density approximations which result in implementable suboptimal estimators for the state of the bilinear system. The results of Monte Carlo simulations are reported; these indicate that simple filters designed by these techniques perform well as compared to other filters.**

I. Introduction

Fourier series analysis has been applied in several recent studies [1]–[4] to estimation problems for stochastic processes evolving on the circle S^1. Willsky [4] used Fourier series methods to define "assumed density" approximations for certain phase tracking and demodulation problems. In fact, a system designed using these techniques performed better than other estimators, including an optimal phase-lock loop.

Manuscript received April 28, 1977; revised October 13, 1977, March 3, 1978, and May 15, 1978. Paper recommended by A. H. Haddad, Chairman of the Estimation Committee. This work was supported by the NSF under Grants GK-42090 and ENG-11106; by the AFOSR under Grant 72-2273; by an NSF Fellowship (the work of S. I. Marcus); and by the Joint Services Electronics Program under Contract F44620-71-C-0091. This paper is based in part on a dissertation by S. I. Marcus, submitted to the Massachusetts Institute of Technology, Cambridge, MA, in partial fulfillment of the requirements for the Ph.D. degree.

S. I. Marcus and K. Hsu are with the Department of Electrical Engineering, University of Texas at Austin, Austin, TX 78712.

A. S. Willsky is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139.

0018-9286/78/1000-0911$00.75

In this paper we study bilinear systems evolving on the n-spheres S^n, the special orthogonal groups $SO(n)$, and certain other compact Lie groups or homogeneous spaces [29], to be described below. The optimal estimator is in general infinite dimensional [7], and our approach to the design of suboptimal estimators is a generalization of that of Willsky [4]. The basic approach involves the definition of an "assumed density" form for the conditional density of the system state at time t given observations up to time t. These densities are defined via the techniques of harmonic analysis on compact Lie groups [5], [6] (which generalize the Fourier series on the Lie group S^1). Our method differs from most previous assumed density approximations in that our approximation is defined on the appropriate compact manifold (as opposed to the usual Gaussian approximations, for example, which are defined on R^n [7]). This method also avoids the problem of merely truncating higher order terms in a harmonic expansion; as pointed out by Lo [18] and Willsky [4], such higher order terms will not be negligible, especially if the filter is performing well. For an alternative approach to discrete-time estimation problems on Lie groups and homogeneous spaces, see the work of Lo and Eshleman [18]–[20], who use exponential Fourier densities to avoid the truncation problem.

In Section II we review some general properties of stochastic bilinear systems and discuss the estimation problem for systems evolving on compact Lie groups and homogeneous spaces. Section III contains the application of the technique to systems evolving on S^n, while Section IV contains the application to systems on $SO(n)$. Results of Monte Carlo simulations of the S^2 estimator are presented in Section V. Some of the concepts of this paper were introduced in [21] and [22], but no simulation results were presented.

II. Estimation for Stochastic Bilinear Systems

The basic stochastic bilinear system (or linear system with state-dependent noise) considered here is described by the Ito stochastic differential equation [4], [9]–[17], [21], [25]

$$dx(t)=\left\{\left[A_0+\tfrac{1}{2}\sum_{i,j=1}^{N}Q_{ij}(t)A_iA_j\right]dt+\sum_{i=1}^{N}A_i dw_i(t)\right\}x(t) \qquad (1)$$

where x is an n-vector or an $n\times n$ matrix, the A_i are $n\times n$ matrices, Q_{ij} is the (i,j)th element of Q, and w is a Brownian motion (Wiener) process with strength $Q(t)$ such that $E[w(t)w'(s)]=\int_0^{\min(t,s)}Q(\tau)d\tau$. Following the notation of [8]–[11], we define $\mathfrak{L}=\{A_0,A_1,\cdots,A_N\}_{LA}$ to be the smallest Lie algebra containing these matrices. The corresponding connected matrix Lie group is denoted by $G=\{\exp\mathfrak{L}\}_G$. Then, if x is an $n\times n$ matrix and $x(t_0)\in G$, the solution $x(t)$ of (1) evolves on G (i.e., $x(t)\in G$ for all $t\geq 0$) in the mean-square sense and almost surely [15]–[17]. If x is an n-vector, then the solution of (1) evolves on the homogeneous space $G\cdot x(t_0)$.

Associated with the Ito equation (1) is a sequence of equations for the powers of the state $x(t)$ (see Brockett [9], [10]). If $N(n,p)$ denotes the binomial coefficient $\binom{n+p-1}{p}$, then given an n-vector x, we define $x^{[p]}$ to be the $N(n,p)$-vector with components equal to the monomials (homogeneous polynomials) of degree p in $x_1,\cdots,x_n$, the components of x, scaled so that $\|x\|^p=\|x^{[p]}\|$. Given an $m\times n$ matrix A, we denote by $A^{[p]}$ the unique matrix which verifies

$$y=Ax\Rightarrow y^{[p]}=A^{[p]}x^{[p]}. \qquad (2)$$

$A^{[p]}$ can be interpreted as a linear operator on symmetric tensors of degree p [9], and is known as the symmetrized Kronecker pth power of A [26]. It is clear that if x satisfies the linear differential equation

$$\dot{x}(t)=Ax(t) \qquad (3)$$

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

$$\dot{x}^{[p]}(t)=A_{[p]}x^{[p]}(t). \qquad (4)$$

We regard this as the definition of $A_{[p]}$, which is the infinitesimal version of $A^{[p]}$. In fact, $A_{[p]}$ can be easily computed from A [25].

It can easily be shown that if x satisfies (1), then $x^{[p]}$ satisfies the Ito equation

$$dx^{[p]}(t)=\left\{A_{0_{[p]}}+\frac{1}{2}\sum_{i,j=1}^{N}Q_{ij}(t)A_{i_{[p]}}A_{j_{[p]}}\right\}x^{[p]}(t)dt + \sum_{i=1}^{N}A_{i_{[p]}}x^{[p]}(t)dw_i(t). \qquad (5)$$

In addition, if the $n\times n$ matrix X satisfies (1), it is easy to show that $X^{[p]}$ also satisfies (5). As we shall see later in the section, this sequence of equations is a valuable tool in the study of state estimation.

The observation model considered in this paper consists of linear observations of the state corrupted by additive white noise, or

$$dz(t)=L(x(t))dt+dv(t) \qquad (6)$$

where L is a linear operator and v is a Wiener process. This bilinear system-linear observation model is useful in the study of certain practical problems, such as the S^2 satellite tracking and $SO(3)$ rigid body orientation estimation problems discussed in [13, ch. 4] and [21, sec. IV].

The remainder of the paper is devoted to the study of the estimation problem for two classes of systems of the form (1), (6), which are motivated by the aforementioned examples. The first system consists of the bilinear state equation

$$dX(t)=\left[A_0+\frac{1}{2}\sum_{i,j=1}^{N}Q_{ij}(t)A_iA_j\right]X(t)dt+\sum_{i=1}^{N}A_iX(t)dw_i(t) \qquad (7)$$

with linear measurements

$$dz_1(t)=X(t)h(t)dt+R^{\frac{1}{2}}(t)dv(t) \qquad (8)$$

where $X(t)$ and $\{A_i\}$ are $n\times n$ matrices, $z_1(t)$ is a p-vector, w is a Wiener process with strength $Q(t)\geq 0$, v is a standard Wiener process independent of w, and $R>0$. More general linear measurements can obviously be considered, but for simplicity of notation we restrict our attention to (8), which arises in the star tracking example of [13, ch. 4]. We will make the crucial assumption that the Lie group $G=\{\exp\mathfrak{L}\}_G$ *is compact*; in particular, we will usually assume that (7) evolves on the special orthogonal group $SO(n)\triangleq\{X\in R^{n\times n}|X'X=I,\ \det X=+1\}$, and thus that $\{A_i\}$ are skew-symmetric.

The second system consists of the bilinear state equation (1) with linear measurements

$$dz_2(t)=H(t)x(t)dt+R^{\frac{1}{2}}(t)dv(t) \qquad (9)$$

where $x(t)$ is an n-vector, A_i are $n\times n$ matrices, and z_2, v and w are as above. It will be assumed that x evolves on a *compact homogeneous space* [8], [13], [29]—i.e., the solution of (1) is

$$x(t)=X(t)x(0) \qquad (10)$$

where X satisfies (7) with $X(0)=I$ and evolves on the compact Lie group $G=\{\exp\mathfrak{L}\}_G$; in particular, we will usually assume that (1) evolves on the $(n-1)$-sphere $S^{n-1}\triangleq\{x\in R^n|x'x=1\}$, and thus that $\{A_i\}$ are skew-symmetric.

It is shown in [13] that, by a linear change of basis on the state space, (7) and (1) evolving on other compact Lie groups can be transformed into equations which evolve on $SO(n)$ and S^{n-1}, respectively. Hence, the results presented here can also be extended to certain other compact Lie groups and homogeneous spaces.[1]

The estimation criterion which will be used for these two problems on $SO(n)$ and S^n is the constrained least-squares estimator, which is analogous to the criterion used in [1], [4], and [21] for the phase estimation problem. That is, for (7)–(8) we wish to find $\tilde{X}(t|t)$ which minimizes the conditional error covariance

$$J_1=E\left[\operatorname{tr}\{(X(t)-\tilde{X}(t|t))'(X(t)-\tilde{X}(t|t))\}|z_1^t\right] \qquad (11)$$

subject to the $SO(n)$ constraint $\tilde{X}(t|t)'\tilde{X}(t|t)=I$, where the notation (11)

[1] In fact, the system (7) may evolve on some subgroup of $SO(n)$, in which case this constraint can also be taken into account in the analysis. If, for example, the system evolves on the torus $T^n\cong SO(2)x\cdots xSO(2)$ (n times), a subgroup of $SO(2n)$, the system can be decoupled into n systems of the form (7) on $SO(2)$.

denotes the conditional expectation given the σ-field $\sigma\{z_1^t\}$ generated by the observed process $z_1^t \triangleq \{z_1(s), 0 \leq s \leq t\}$ up to time t. For (1), (9) we seek $\tilde{x}(t|t)$ which minimizes

$$J_2 = E\left[(x(t)-\tilde{x}(t|t))'(x(t)-\tilde{x}(t|t))|z_2^t\right] \tag{12}$$

subject to the S^{n-1} constraint $\|\tilde{x}(t|t)\|^2 = \tilde{x}(t|t)'\tilde{x}(t|t)=1$. It is easily shown [13] that the optimal estimates are, respectively,

$$\tilde{X}(t|t) = \pm \hat{X}(t|t)\left[\hat{X}(t|t)'\hat{X}(t|t)\right]^{-\frac{1}{2}} \tag{13}$$

$$\tilde{x}(t|t) = \frac{\hat{x}(t|t)}{\|\hat{x}(t|t)\|} \tag{14}$$

where the conditional expectation is denoted by the equivalent notations

$$\hat{x}(t|t) \triangleq E\left[x(t)|z_2^t\right] \triangleq E^t[x(t)]. \tag{15}$$

The sign in (13) is chosen to ensure that $\det \tilde{X}(t|t) = +1$ [23].[2] Thus in both cases we must compute the conditional expectation of the state ($X(t)$ or $x(t)$) given the past observations (z_1^t or z_2^t).

The equations for computing the conditional expectation can be derived from the general nonlinear filtering equation [7] and the moment equation (5). The resultant equations for the $SO(n)$ system (7)–(8) are

$$\begin{aligned} dE^t[X_c^{[p]}(t)] &= \left[\left(A_{0_{[p]}} + \frac{1}{2}\sum_{i,j=1}^{N} Q_{ij}(t)A_{i_{[p]}}A_{j_{[p]}}\right)\otimes I\right]E^t[X_c^{[p]}(t)]\,dt \\ &\quad + \{E^t[X_c^{[p]}(t)h'(t)X(t)] \\ &\quad - E^t[X_c^{[p]}(t)]h'(t)E^t[X(t)]\}R^{-1}(t)\,d\nu_1(t) \end{aligned} \tag{16}$$

$$d\nu_1(t) = dz_1(t) - \hat{X}(t|t)h(t)\,dt \tag{17}$$

where $\otimes$ denotes Kronecker product and $X_c^{[p]}$ is the vector containing the elements of the matrix $X^{[p]}$ in lexicographic order [26], [32, p. 64]. For the S^{n-1} system (1), (9), we have

$$\begin{aligned} dE^t[x^{[p]}(t)] &= \left[A_{0_{[p]}} + \frac{1}{2}\sum_{i,j=1}^{N} Q_{ij}(t)A_{i_{[p]}}A_{j_{[p]}}\right]E^t[x^{[p]}(t)]\,dt \\ &\quad + \{E^t[x^{[p]}(t)x'(t)] - E^t[x^{[p]}(t)]E^t[x'(t)]\} \\ &\quad \cdot H'(t)R^{-1}(t)\,d\nu_2(t) \end{aligned} \tag{18}$$

$$d\nu_2(t) = dz_2(t) - H(t)\hat{x}(t|t)\,dt. \tag{19}$$

The structure of these equations is quite similar to that of [4]—i.e., each estimator consists of an infinite band of filters, and the filter for the pth moment is coupled only to those for the first and $(p+1)$st moments. Therefore, we are led to the design of suboptimal estimators. The technique proposed here is motivated by the highly successful use of folded normal assumed density approximations in the phase tracking problem [4]; filters designed using this technique performed very well as compared with other suboptimal estimators. We will describe similar techniques for the design of suboptimal estimators on S^n and $SO(n)$.

We first review the notions of Brownian motion and Gaussian densities on Lie groups and homogeneous spaces. Yosida [28] proved that the fundamental solution of

$$\frac{\partial p(x,t)}{\partial t} - \gamma \Delta p(x,t) = 0 \tag{20}$$

where $\gamma > 0$ and Δ is the Laplace–Beltrami operator (Laplacian) on a Riemannian homogeneous space M [5],[13],[29], is the density (with respect to the Riemannian measure) of a Brownian motion on M.[3]

According to [5], the fundamental solution of (20) is given by

$$p(x,t;x_0,t_0) = \sum_i \phi_i(x)\phi_i(x_0)e^{-\lambda_i(t-t_0)\gamma} \tag{21}$$

where λ_i and ϕ_i are the eigenvalues and the corresponding eigenfunctions of the Laplacian. The function $p(x,t;\,x_0,t_0)$ is the solution to (20) with initial condition equal to the singular distribution concentrated at $x=x_0$. Also, Grenander [27] defines a Gaussian (normal) density to be the solution of (20) for some t.

The folded normal density $F(\theta;\,\eta,\gamma)$ used by Willsky as an assumed density approximation for the phase tracking problem is indeed a normal density on S^1 in the sense of Grenander [4]; in fact, the trigonometric polynomials $e^{-in\theta}$ are eigenfunctions of the Laplacian on S^1. Motivated by the success of Willsky's suboptimal filter, we will design suboptimal estimators for the $SO(n)$ and S^n bilinear systems by employing normal assumed conditional densities of the form

$$p(x,t) = \sum_i \phi_i(x)\phi_i(\eta(t))e^{-\lambda_i\gamma(t)} \tag{22}$$

where $\eta(t)$ and $\gamma(t)$ are parameters of the density which are to be estimated.[4]

III. Estimation on S^n

In this section the suboptimal estimation technique discussed in the previous section will be used in order to design filters for the S^n estimation problem (1), (9). The optimal constrained least-squares estimator is described by (14) and (18)–(19). First, the suboptimal estimator for S^2 will be described in detail; then the generalization to S^n will be discussed.

In our discussion of estimation on S^2, we will refer to a point on S^2 in terms of the Cartesian coordinates $x \triangleq (x_1,x_2,x_3)$ or the polar coordinates (θ,ϕ), where $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$.

The normalized spherical harmonics of degree l on S^2 are defined by [6]

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

$$Y_{l,-m}(\theta,\phi) = (-1)^m Y_{lm}^*(\theta,\phi)$$

for $l=0,1,\cdots$ and $m=0,1,\cdots,l$, where $P_{lm}(\cos\theta)$ are the associated Legendre functions and $*$ denotes complex conjugate. It is shown in [30] that the $\{Y_{lm}\}$ are the eigenfunctions of the Laplacian

$$\Delta_{S^2} = \frac{1}{\sin\theta}\left[\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin\theta}\frac{\partial^2}{\partial\phi^2}\right],$$

and all spherical harmonics of degree l have the same eigenvalue $-l(l+1)$. Thus, the assumed density approximation is a normal density on S^2 of the form (22), as discussed in the previous section:

$$p(\theta,\phi,t) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} Y_{lm}(\theta,\phi)Y_{lm}^*(\eta(t),\lambda(t))e^{-l(l+1)\gamma(t)}. \tag{23}$$

In other words,

$$c_{lm}(t) \triangleq E^t[Y_{lm}^*(\theta(t),\phi(t))] = Y_{lm}^*(\eta(t),\lambda(t))e^{-l(l+1)\gamma(t)}. \tag{24}$$

In order to truncate the optimal estimator (18)–(19) after the $\hat{x}^{[N]}(t|t)$ equation using the assumed density (23), we must compute $E^t[x^{[N]}(t)x'(t)]$, or equivalently, $\hat{x}^{[N+1]}(t|t)$, in terms of $\hat{x}^{[p]}(t|t), p=1,2,\cdots,N$. However, if $\hat{x}(t|t)$ is known, so are $c_{10}(t)$ and $c_{11}(t)$. A simple computation [13], [22], then shows that $\{c_{N+1,m}, m=-(N+1),\cdots,N+1\}$ can be computed from

[2] The optimal estimates are slightly different for other compact Lie groups and homogeneous spaces. For example, if $x(t)$ belongs to the ellipsoid $\{x|x'Px=1\}$, then the estimate which minimizes $E[(x(t)-\tilde{x}(t|t))'P(x(t)-\tilde{x}(t|t))|z_2^t]$ subject to $\tilde{x}(t|t)'P\tilde{x}(t|t)=1$, is $\tilde{x}(t|t) = \hat{x}(t|t)/(\hat{x}(t|t)'P\hat{x}(t|t))^{\frac{1}{2}}$. If $X(t)$ belongs to the Lie group $\{X|X'PX=P\}$, then the estimate which minimizes $E[\mathrm{tr}[(X(t)-\tilde{X}(t|t))'P(X(t)-\tilde{X}(t|t))]|z_1^t]$ subject to $\tilde{X}'(t|t)\cdot P\tilde{X}(t|t)=P$, is given by $\tilde{X}(t|t) = \pm\hat{X}(t|t)[\hat{X}'(t|t)P\hat{X}(t|t)]^{-\frac{1}{2}}P^{\frac{1}{2}}$. Other cases can be derived similarly using Lagrange multipliers.

If the system evolves on a subgroup of $SO(n)$, the estimate should be further constrained. For example, if the state space $T^n \cong SO(2)x\cdots xSO(2)$, an estimate $\tilde{X}_i(t|t)$ of the form (13) is formed on each component $SO(2)$ of the direct product, and the resulting estimate $(\tilde{X}_1(t|t),\cdots,\tilde{X}_n(t|t))$ will belong to the subgroup $SO(2)x\cdots xSO(2)$.

[3] Yosida defines a Brownian motion process to be a temporally and spatially homogeneous Markov process on M which satisfies a continuity condition of Lindeberg's type.

[4] In order to assure the existence of a conditional density, it is sufficient to assume that the system is "controllable from the noise" [10], [13], [17].

$$c_{N+1,m}(t)=Y^*_{N+1,m}(\eta(t),\lambda(t))e^{-(N+1)(N+2)\gamma(t)}$$
$$=(-1)^m\left[\frac{(N+1-m)!}{(N+1+m)!}\frac{2N+3}{4\pi}\right]^{\frac{1}{2}}P_{N+1,m}\left(\frac{c_{10}(t)}{(c_{10}^2(t)+2|c_{11}(t)|^2)^{\frac{1}{2}}}\right)$$
$$\cdot\left(\frac{c_{11}(t)}{c_{11}^*(t)}\right)^{m/2}\left[\frac{4\pi}{3}(c_{10}^2(t)+2|c_{11}(t)|^2)\right]^{\frac{1}{4}(N+1)(N+2)}. \quad (25)$$

Finally, it is shown in [13] that there exists a nonsingular matrix T such that, if Y_l is the $(2l+1)$-vector with components $\{Y_{lm}, -l\leqslant m\leqslant l\}$, then

$$Tx^{[N+1]}=\begin{bmatrix} Y_{N+1}(x)\\ x^{[N-1]}\end{bmatrix}. \quad (26)$$

Thus, $\hat{x}^{[N+1]}(t|t)$ can be computed from $\{c_{N+1,m}, -(N+1)\leqslant m\leqslant N+1\}$ and $\hat{x}^{[N-1]}(t|t)$. The optimal estimator (18) is truncated by substituting this approximation for $\hat{x}^{[N+1]}(t|t)$ into the equation for $\hat{x}^{[N]}(t|t)$. Notice that the entire procedure for truncating the optimal estimator can equivalently be performed on the infinite set of coupled equations for the generalized Fourier coefficients $c_{lm}(t)$, using the approximation (24).

We note that one can show that

$$\alpha(t)\triangleq\sqrt{\|\hat{x}(t|t)\|}\leqslant 1$$

and this quantity can be used as a measure of our confidence in our estimate. If $\hat{x}(t|t)$ satisfies the assumed density (23),

$$\alpha(t)=\|\hat{x}(t|t)\|=e^{-\gamma(t)} \quad (27)$$

so $\gamma=0$ (zero "variance") implies $\alpha=1$, and $\gamma=\infty$ (infinite "variance") implies $\alpha=0$ (see [4] for the S^1 analog).

Example 1: Suppose that we truncate the optimal S^2 estimator (18) after $N=1$—i.e., we approximate $\hat{x}^{[2]}(t|t)$ using the above approximation. Assume that $Q(t)=I$ and $\{A_i, i=1,2,3\}$ are given. Then the resulting suboptimal estimator is (for $Q(t)=I$)

$$d\hat{x}(t|t)=\left[A_0+\frac{1}{2}\sum_{i=1}^{3}A_i^2\right]\hat{x}(t|t)\,dt$$
$$+P(t)H'(t)R^{-1}(t)[dz_2(t)-H(t)\hat{x}(t|t)\,dt] \quad (28)$$

where the "covariance" matrix $P(t)$ is given by

$$P_{ii}(t)=\hat{x}_i^2(t|t)\left(\frac{2}{3}\|\hat{x}(t|t)\|-1\right)$$
$$-\frac{1}{3}\left(\hat{x}_j^2(t|t)+\hat{x}_k^2(t|t)\right)\|\hat{x}(t|t)\|+\frac{1}{3} \quad (29)$$

for $i\neq j$, $i\neq k$, $j\neq k$, and

$$P_{ij}(t)=\hat{x}_i(t|t)\hat{x}_j(t|t)(\|\hat{x}(t|t)\|-1) \quad (30)$$

for $i\neq j$. It is shown in [37] that the matrix $P(t)$ of (29)–(30) is positive semidefinite, and thus can be viewed as a covariance matrix. The results of Monte Carlo simulations to evaluate the performance of this estimator are presented in Section VII.

The extension to S^n of this technique for constructing suboptimal estimators is straightforward. The procedure uses the spherical harmonics on S^n. In polar coordinates, a point on S^n can be described by $(\theta_1,\theta_2,\cdots,\theta_{n-1},\phi)\triangleq(\theta,\phi)$, where $0\leqslant\theta_j\leqslant\pi$ and $0\leqslant\phi\leqslant 2\pi$. Also, the spherical harmonics are denoted by

$$Y_{l,(m)}(\theta,\phi)\triangleq Y_{l,m_1,\cdots,m_{n-1}}(\theta_1,\cdots,\theta_{n-1},\pm\phi)$$
$$=e^{\pm im_{n-1}\phi}\prod_{k=0}^{n-2}(\sin\theta_{k+1})^{m_{k+1}}C^{m_{k+1}+\frac{1}{2}(n-k-1)}_{m_k-m_{k+1}}(\cos\theta_{k+1}) \quad (31)$$

where $l\geqslant m_1\geqslant\cdots\geqslant m_{n-1}\geqslant 0$ and C_j^i are the Gegenbauer polynomials [33] (that is, the functions $Y_{l,(m)}$ are eigenfunctions of the Laplace–Beltrami operator with eigenvalue $-l(n+l-1)$). Hence, the assumed density approximation on S^n is

$$p(\theta,\phi,t)=\sum_{l,(m)}Y_{l,(m)}(\theta,\phi)Y^*_{l,(m)}(\eta(t),\lambda(t))e^{-l(l+n-1)\gamma(t)}. \quad (32)$$

That is,

$$c_{l,(m)}(t)\triangleq E^t[Y^*_{l,(m)}(\theta(t),\phi(t))]$$

is assumed to be

$$c_{l,(m)}(t)=Y^*_{l,(m)}(\eta(t),\lambda(t))e^{-l(l+n-1)\gamma(t)}. \quad (33)$$

The procedure for truncating the filter (18) is identical to the S^2 case. If $\hat{x}(t|t)$ is known, so are $c_{1,(m)}(t)$, and these can be used to compute $\gamma(t)$, $\eta(t)$, and $\lambda(t)$. Then $\{c_{N+1,(m)}(t)\}$ can be computed from (33), and $\hat{x}^{[N+1]}(t|t)$ can be computed from $\{c_{N+1,(m)}(t)\}$ and $\hat{x}^{[N-1]}(t|t)$. The estimator is truncated by substituting this approximate expression for $\hat{x}^{[N+1]}(t|t)$ into the equation (18) for $\hat{x}^{[N]}(t|t)$.

IV. Estimation on $SO(n)$

In this section we discuss the construction of suboptimal estimators for the $SO(n)$ estimation problem (7)–(8). We will only consider the $SO(3)$ problem; the results are extended to $SO(n)$ in [13]. The concepts of harmonic analysis on $SO(3)$ presented in [5], [6], [13, Appendix], [29], and [30] will be used extensively.

Any matrix R in $SO(3)$ can be described in local coordinates in terms of the *Euler angles* ϕ, θ, ψ, where $0\leqslant\phi<2\pi$, $0\leqslant\theta\leqslant 2\pi$, $0\leqslant\psi<2\pi$. Talman [6] computes a sequence $D^l(\phi,\theta,\psi)$, $l=0,1,\cdots$, of unitary irreducible representations of $SO(3)$; its matrix elements are given by

$$D^l_{mn}(\phi,\theta,\psi)=i^{m-n}e^{-im\phi}d^l_{mn}(\theta)e^{-in\psi}$$

where

$$d^l_{mn}(\theta)=\sum_t(-1)^t\frac{[(l+m)!(l-m)!(l+n)!(l-n)!]^{\frac{1}{2}}}{(l+m-t)!(t+n-m)!t!(l-n-t)!}$$
$$\cdot\cos^{2l+m-n-2t}\left(\frac{\theta}{2}\right)\sin^{2t+n-m}\left(\frac{\theta}{2}\right)$$

for $-l\leqslant m$, $n\leqslant l$. Here t is summed over all nonnegative integers such that

$$m-n\leqslant t\leqslant l+m, \qquad 0\leqslant t\leqslant l-n.$$

The Peter–Weyl Theorem [5, p. 40], [29, p. 257] implies that, for fixed l, the matrix elements $\{D^l_{mn};\ -l\leqslant m,\ n\leqslant l\}$ are eigenfunctions of the Laplacian [31]

$$\Delta_{SO(3)}=\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\left(\frac{\partial^2}{\partial\phi^2}-2\cos\theta\frac{\partial^2}{\partial\phi\partial\psi}+\frac{\partial^2}{\partial\psi^2}\right)$$

with the same eigenvalue λ_l; also, each eigenfunction of the Laplacian can be written as a linear combination of the $\{D^l_{mn}\}$. Hence, the assumed density which will be used to truncate the optimal estimator (16)–(17) is a normal density on $SO(3)$ of the form (22):

$$p(R,t)=\sum_{l=0}^{\infty}\sum_{m,n=-l}^{l}D^l_{m,n}(R)D^l_{m,n}(\eta(t))^*e^{-\lambda_l\gamma(t)} \quad (34)$$

where R, $\eta(t)\in SO(3)$ and $\gamma(t)$ is a scalar. That is,

$$c^l_{mn}(t)\triangleq E^t[D^l_{mn}(\eta(t))^*] \quad (35)$$

is assumed to be

$$c^l_{mn}(t)=D^l_{mn}(\eta(t))^*e^{-\lambda_l\gamma(t)}. \quad (36)$$

The procedure for truncating the filter (16) is similar to the S^n case, although we make use of some additional concepts from representation theory. If $\hat{X}(t|t)$ is known, so are $\{c^1_{mn}(t);\ -1\leqslant m,n\leqslant 1\}$, since D^1 is equivalent to the self-representation of $SO(3)$. Define the matrix $C^l(t)$ with elements $c^l_{mn}(t)$, $-l\leqslant m,n\leqslant l$; then

$$A(t)\triangleq\bar{C}^1(t)C^1(t)=[D^1(\eta(t))]'[D^1(\eta(t))]^*e^{-2\lambda_1\gamma(t)}$$
$$=I\cdot e^{-2\lambda_1\gamma(t)} \quad (37)$$

since D^1 is unitary (here $\bar{C}$ is the Hermitian transpose of C). Thus, $\gamma(t)$

can be computed from

$$\gamma(t) = -\frac{1}{2\lambda_1}\log\left[\frac{1}{3}\operatorname{tr}A(t)\right]. \tag{38}$$

Then the elements of $\eta(t)$ can be computed from (36) and (38), since $D^1(\eta(t))$ is similar to $\eta(t)$. Once $\gamma(t)$ and $\eta(t)$ have been computed, $\{c_{mn}^{N+1};\ -(N+1)\leq m,n\leq N+1\}$ are computed from the formula (36).

In order to truncate (16) after the Nth moment equation, we must approximate $E^t[X_v^{[N]}(t)h'(t)X(t)]$; however, this matrix consists of time-varying deterministic functions multiplying elements of $\hat{X}^{[N+1]}(t|t)$, so we will show how to approximate this matrix. The symmetrized Kronecker pth power $X^{[p]}$ operating on the symmetric tensors $x^{[p]}$ such that $\|x^{[p]}\|=\|x\|^p=1$ furnishes a representation of $SO(3)$ which is reducible [26]. It is shown in [13] that there is a nonsingular matrix such that

$$TX^{[p]}T^{-1} = \begin{bmatrix} D^p(X) & 0 \\ 0 & X^{[p-2]} \end{bmatrix}. \tag{39}$$

The matrix T is related to the Clebsch–Gordan coefficients [6], but T can also be computed by the method of Gantmacher [34, p. 160]. It is clear from the decomposition (39) that $\hat{X}^{[N+1]}(t|t)$ can be computed from $C^{N+1}(t)$ and $\hat{X}^{[N-1]}(t|t)$. The optimal estimator (16) is truncated by substituting this approximation into the equation for $\hat{X}^{[N]}(t|t)$.

We note here that, due to the decomposition (39), the estimation equations and the truncation procedure could have been expressed solely in terms of the irreducible representations $D^p(X(t))$. However, we have chosen to work with the $X^{[p]}$ equations primarily for ease of notation. For large N, the D^p equations would provide significant computational savings over the $X^{[p]}$ equations, as these are redundant; however, the practical implementation of this technique will probably be limited to small values of N.

V. Simulation Results

As an illustration of the techniques presented in the previous sections, the first order filter (FOF) of Example 1 (Section III) was evaluated by means of digital Monte Carlo simulations. It was compared to both the extended Kalman filter (EKF) [7] and the Gustafson–Speyer linear, minimum-variance quadrature filter (LQF) [24]. Identical noise sequences were used to allow direct comparisons.

The system considered was the S^2 system, i.e.,

$$dx(t) = Fx(t)\,dt + \sum_{i=1}^{3} A_i x(t)\,dw_i(t) \tag{40}$$

$$dz(t) = x(t)\,dt + r^{1/2}dv(t) \tag{41}$$

where $F=\Sigma_{i=1}^3 f_iA_i + \frac{1}{2}q\Sigma_{i=1}^3 A_i^2$, and $\{A_i, i=1,2,3\}$ are the skew-symmetric matrices

$$A_1=\begin{bmatrix} 0 & 0 & 0\\ 0 & 0 & -1\\ 0 & 1 & 0\end{bmatrix} A_2=\begin{bmatrix} 0 & 0 & 1\\ 0 & 0 & 0\\ -1 & 0 & 0\end{bmatrix} A_3=\begin{bmatrix} 0 & -1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}. \tag{42}$$

Also, $w(t)$ has strength qI; and v has strength I. In this experiment, the nominal angular velocities $\{f_i, i=1,2,3\}$ were chosen to be 100.0, and q and r were varied.

For all three filters, the normalized estimate $\tilde{x}(t)=\hat{x}(t|t)/\|\hat{x}(t|t)\|$ was used. The filters have an identical structure for the approximate $\hat{x}$ equation:

$$d\hat{x}(t|t) = F\hat{x}(t|t) + \frac{1}{r}P(t)[dz(t) - \hat{x}(t|t)\,dt]. \tag{42}$$

However, for the FOF, $P(t)$ is given by the highly nonlinear memoryless equations (29)–(30). In the EKF, $P(t)$ satisfies the Riccati equation

$$\frac{d}{dt}P(t) = FP(t) + P(t)F' + qG(\hat{x}(t|t))G'(\hat{x}(t|t)) - \frac{1}{r}P(t)P'(t) \tag{43}$$

where

$$G(\hat{x}) = [A_1\hat{x}, A_2\hat{x}, A_3\hat{x}]. \tag{44}$$

Since the Riccati equation (43) is a function of $\hat{x}(t|t)$, the $P(t)$ calculation in the EKF requires *extensive on-line computation*, which represents a considerable burden. In the LQF, $P(t)$ is given as the solution of the coupled dynamic equations

$$\frac{d}{dt}P(t) = FP(t) + P(t)F' + \Delta(X(t),t) - \frac{1}{r}P(t)P'(t) \tag{45}$$

$$\frac{d}{dt}X(t) = FX(t) + X(t)F' + \Delta(X(t),t) \tag{46}$$

where $\Delta(X(t),t)$ is a diagonal matrix with ith component

$$\Delta(X(t),t)_i = q\sum_{k,l,m=1}^{3}(A_m)_{ik}(A_m)_{il}X_{kl}(t). \tag{47}$$

Notice that these equations for $P(t)$ and $X(t)$ can be calculated off-line, but the LQF thus has a considerable storage requirement. Because $P(t)$ in the FOF is only given by a memoryless nonlinearity, this filter requires considerably less storage than the LQF and less on-line computation than the EKF.

Our approach to the statistical analysis of the Monte Carlo simulations closely parallels that of Bucy and his associates [1], [3], [35]. The steady-state mean-squared error

$$\mu_2 = E[\|x(t)-\tilde{x}(t)\|^2] = \sum_{i=1}^{3} E\left[(x_i(t)-\tilde{x}_i(t))^2\right] \tag{48}$$

where $\tilde{x}_i(t)$ denotes the estimate of the ith component of the state $x_i(t)$, was used as the performance criterion. If $\{x^n\}$ and $\{\tilde{x}^n\}$, $n=1,\cdots,N$, are sequences of independent realizations of $x(t)$ and $\tilde{x}(t)$, respectively, then the statistic

$$\mu_2 = \frac{1}{N}\sum_{n=1}^{N}\|x^n - \tilde{x}^n\|^2 \tag{49}$$

is an approximation to μ_2 for sufficiently large N. In fact, by the Central Limit Theorem [36, p. 278], $\hat{\mu}_2$ is asymptotically normal with

$$E[\hat{\mu}_2] = \mu_2 \tag{50}$$

$$\operatorname{var}[\hat{\mu}_2] = \frac{1}{N}\left\{\sum_{i=1}^{3}(\mu_4)_i + 2(\mu_4)_{12} + 2(\mu_4)_{13} + 2(\mu_4)_{23} - (\mu_2)^2\right\} \tag{51}$$

where $(\mu_4)_i \triangleq E[(x_i(t)-\tilde{x}_i(t))^4]$ and

$$(\mu_4)_{ij} = E\left[(x_i(t)-\tilde{x}_i(t))^2(x_j(t)-\tilde{x}_j(t))^2\right].$$

Thus, for large N, a consideration of the 3σ confidence interval implies that

$$\Pr\left\{\hat{\mu}_2 - 3\sqrt{\operatorname{var}(\hat{\mu}_2)} \leq \mu_2 \leq \hat{\mu}_2 + 3\sqrt{\operatorname{var}(\hat{\mu}_2)}\right\} = 0.9974. \tag{52}$$

In the Monte Carlo simulations, var ($\hat{\mu}_2$) was estimated from the samples (using sample means to estimate $(\mu_4)_i$ and $(\mu_4)_{ij}$ as in (49)), and approximate confidence intervals were thus computed.

In the experiment, 15 sample paths, each of which contained 1000 steps of length 0.001 s, were run in each simulation. The first 200 samples in each sample path were discarded to allow the transients to decay, so the remaining 800 samples represented steady-state. If all the steady-state errors were averaged as in (49), this would lead to 12000 samples of the steady-state error. However, as noted in [4], [24], and [35], adjacent errors in each path are correlated, so the effective Monte Carlo length is somewhere in the range between $N=1200$ and $N=12000$. The three standard deviation confidence intervals were calculated for both values of N.

The results of the simulations are presented in Table I. The 3σ confidence intervals I_1 (for $N=12\,000$) and I_2 (for $N=1200$) are shown. The results of this approximate statistical analysis of the Monte Carlo simulations indicate that, for this simple example, the FOF performs comparably to the LQF, and better than the EKF. The FOF seems to perform better in comparison to the other filters as q increases, due to the increasing dominance of the bilinear noise term in the system

TABLE I
MONTE CARLO M. S. ESTIMATION ERRORS ($\times 10^{-3}$)

		M.S. Error $\hat{u}_2$	I_1 (N=12000)	I_2 (N=1200)
q = 0.01, r = 0.01	FOF	9.65	[9.41,9.91]	[8.92,10.52]
	EKF	10.21	[9.96,10.48]	[9.45,11.11]
	LQF	9.49	[9.26,9.75]	[8.78,10.34]
q = 0.01, r = 1.00	FOF	11.69	[11.39,12.01]	[10.79,12.77]
	EKF	11.75	[11.44,12.07]	[10.84,12.82]
	LQF	11.69	[11.39,12.01]	[10.79,12.77]
q = 1.00, r = 0.01	FOF	170.38	[165.26,175.82]	[155.19,188.87]
	EKF	193.75	[188.74,199.04]	[178.74,211.51]
	LQF	170.65	[165.58,176.03]	[155.60,188.92]

equation (40). These results are significant, due to the fact that the FOF designed here requires considerably less storage and computation than the other filters (no additional differential equations or storage for $P(t)$ are required).

VI. CONCLUSIONS

The state estimation problem for bilinear stochastic systems evolving on compact Lie groups and homogeneous space has been considered. The techniques of harmonic analysis on compact Lie groups have been applied to the design of suboptimal estimators for such systems. Monte Carlo simulations of a simple example indicate that a computationally simple filter designed by these methods performs favorably as compared to two other filters.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for several helpful suggestions and clarifications.

REFERENCES

[1] R. S. Bucy, C. Hecht, and K. D. Senne, "New methods for nonlinear filtering," *Revue Francaise d'Automatique, Informatique et de Recherche Operationnelle*, pp. 3–54, Feb. 1973.
[2] R. S. Bucy and H. Youssef, "Fourier realization of the optimal phase demodulator," presented at 4th Symposium on Nonlinear Estimation and Its Applications, San Diego, CA, 1973.
[3] J. J. Mallinckrodt, R. S. Bucy, and S. Y. Chang, "Final project report for a design study for an optimal non-linear receiver/demodulator," Goddard Space Flight Center, NASA Contract NAS 5-10789, MD, 1970.
[4] A. S. Willsky, "Fourier series and estimation on the circle with applications to synchronous communication, parts I, II," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 577–590, Sept. 1974.
[5] E. M. Stein, *Topics in Harmonic Analysis*. Princeton: Princeton University Press, 1970.
[6] J. D. Talman, *Special Functions—A Group Theoretic Approach*. New York: Benjamin, 1968.
[7] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.
[8] R. W. Brockett, "System theory on group manifolds and coset spaces," *SIAM J. Contr.*, vol. 10, no. 2, pp. 265–284, May 1972.
[9] ——, "Lie algebras and Lie groups in control theory," in *Geometric Methods in System Theory*, D. Q. Mayne and R. W. Brockett, Eds. The Netherlands: Reidel, 1973, pp. 43–82.
[10] R. W. Brockett, "Lie theory and control systems on spheres," *SIAM J. Appl. Math.*, vol. 25, no. 2, pp. 213–225, Sept. 1973.
[11] A. S. Willsky, S. I. Marcus, and D. N. Martin, "On the stochastic stability of linear systems containing colored multiplicative noise," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 711–713, Oct. 1975.
[12] A. S. Willsky and S. I. Marcus, "Analysis of bilinear noise models in circuits and devices," *J. Franklin Inst.*, vol. 301, no. 1, pp. 103–122, Jan. 1976.
[13] S. I. Marcus, "Estimation and analysis of nonlinear stochastic systems," Ph.D. dissertation, Dep. of Elec. Eng., M.I.T., Cambridge, MA, June 1975; also M.I.T. Electronic Systems Lab. Rep. ESL-R-601.
[14] S. I. Marcus and A. S. Willsky, "Algebraic structure and finite dimensional nonlinear estimation," *SIAM J. Math. Anal.*, Apr. 1978.
[15] J. T. Lo, "Signal detection on Lie groups," in *Geometric Methods in System Theory*, D. Q. Mayne and R. W. Brockett, Eds. The Netherlands: Reidel, 1973, pp. 295–304.
[16] J. M. C. Clark, "An introduction to stochastic differential equations on manifolds," in *Geometric Methods in System Theory*, D. Q. Mayne and R. W. Brockett, Eds. The Netherlands: Reidel, 1973, pp. 131–149.
[17] D. L. Elliott, "Diffusions on manifolds arising from controllable systems," in *Geometric Methods in System Theory*, D. Q. Mayne and R. W. Brockett, Eds. The Netherlands: Reidel, 1973, pp. 285–294.
[18] J. T. H. Lo, "Exponential Fourier densities and optimal estimation and detection on the circle," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 110–116, Jan. 1977.
[19] J. T. H. Lo and L. R. Eshleman, "Exponential Fourier densities on S^2 and optimal estimation and detection for directional processes," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 321–336, May 1977.
[20] ——, "Exponential Fourier densities on $SO(3)$ and optimal estimation and detection for rotation processes," in *Proc. 1976 IEEE Conf. on Decision and Control*, pp. 134–142.
[21] A. S. Willsky and S. I. Marcus, "Estimation for bilinear stochastic systems," in *Variable Structure Systems with Applications to Economics and Biology*, A. Ruberti and R. R. Mohler, Eds. New York: Springer-Verlag, 1975, pp. 116–137.
[22] S. I. Marcus and A. S. Willsky, "The use of spherical harmonics in suboptimal estimator design," in *Proc. 1976 IEEE Conf. on Decision and Control*, pp. 199–203.
[23] I. Y. Bar-Itzhack, "Iterative optimal orthogonalization of the strapdown matrix," *IEEE Trans. Aero. and Elec. Systems*, vol. AES-11, pp. 30–37, Jan. 1975.
[24] D. E. Gustafson and J. L. Speyer, "Linear minimum variance filters applied to carrier tracking," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 65–73, Feb. 1976.
[25] G. L. Blankenship, "Lie theory and the moment stability problem in stochastic differential equations," presented at Triennial World Congress of IFAC, Cambridge, MA, Aug. 24–30, 1975.
[26] F. D. Murnaghan, *Theory of Group Representations*. Baltimore: Johns Hopkins, 1938.
[27] U. Grenander, *Probabilities on Algebraic Structures*. New York: Wiley, 1963.
[28] K. Yosida, *Functional Analysis*. New York: Springer-Verlag, 1974, pp. 398–403.
[29] F. Warner, *Foundations of Differential Manifolds and Lie Groups*. Glenview, IL: Scott, Foresman, 1971.
[30] C. F. Dunkl and D. E. Ramirez, *Topics in Harmonic Analysis*. New York: Appleton-Century-Crofts, 1971.
[31] R. Burridge, G. C. Papanicolaou, and D. McLaughlin, "The solution by descent of certain partial differential equations on the sphere and Lobachevsky plane," *Comm. Pure Appl. Math.*, vol. XXVI, pp. 105–129, 1973.
[32] R. W. Brockett, *Finite Dimensional Linear Systems*. New York: Wiley, 1970.
[33] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. Tricomi, *Higher Transcendental Functions*, vol. II. New York: McGraw-Hill, 1953.
[34] F. R. Gantmacher, *The Theory of Matrices*, vol. I. New York: Chelsea, 1960.
[35] R. S. Bucy, C. Hecht, and K. D. Senne, *An Engineers' Guide to Building Nonlinear Filters*, F. J. Seiler Research Lab. Rep. SRL-TR-72-0004, May 1972.
[36] H. Bauer, *Probability Theory and Elements of Measure Theory*. New York: Holt, Rinehart, and Winston, 1972.
[37] K. Hsu, *The Evaluation of Several Suboptimal Filters for Bilinear Systems Evolving on Spheres*, M. S. thesis, Dep. of Elec. Eng., Univ. Texas at Austin, Austin, TX, May 1977.

Applying a Smoothing Criterion to the Kalman Filter

PER HEDELIN AND INGVAR JÖNSSON

***Abstract*—A performance measure is suggested for evaluating the performance of a given optimal estimator at other lags than the design lag. Applying this idea, suboptimal smoothers are found for both continuous- and discrete-time systems, combining low complexity and good performance. Several examples are considered. Suboptimal-smoothing improvement is related to optimal improvement and interpreted in terms of input–output transfer-function properties.**

A special class of discrete-time systems is also discussed where the optimal smoother is of the same complexity as the zero-lag filter.

I. INTRODUCTION

The classical Kalman filter is the optimal solution to the following momentary estimation problem: given a noisy measurement, compute recursively an estimate of the "present" message. Due to its robustness and relative simplicity the Kalman filter has enjoyed much attention, as shown in the survey, Kailath [1].

An estimation similar to that of the Kalman theory can be posed: given a noisy measurement, compute recursively an estimate of a "past" message. This is the smoothing problem. Meditch [2] presents the major

Manuscript received September 8, 1977; revised February 7, 1978 and April 17, 1978. Paper recommended by A. H. Haddad, Chairman of the Estimation Committee.

The authors are with the Division of Information Theory, Chalmers University of Technology, Göteborg, Sweden.

0018-9286/78/1000-0916$00.75 ©1978 IEEE