5) Establishing criteria for stability.
6) Application of estimation theory.

Finally, of a more general nature, the techniques and concepts of optimal control could be extended to the spatial model.

References

[1] A. Rosenfeld, *Picture Processing by Computer*. New York: Academic, 1969.
[2] H. C. Andrews, *Computer Techniques in Image Processing*. New York: Academic, 1970.
[3] B. S. Lipkin and A. Rosenfeld, *Picture Processing and Psychopictorics*. New York: Academic, 1970.
[4] T. S. Huang, W. F. Schreiber, and O. J. Tretiak, "Image processing," *Proc. IEEE*, vol. 59, pp. 1586–1609, Nov. 1971.
[5] W.D. Fryer and G.E. Richmond, "Two-dimensional spatial filtering and computers," in *Proc. Nat. Electron. Conf.*, Oct. 19, 1962.
[6] A. Habibi, "Two-dimensional Bayesian estimate of images," *Proc. IEEE (Special Issue on Digital Picture Processing)*, vol. 60, pp. 878–883, July 1972.
[7] A. Vander Lugt, "Operational notation for the analysis and synthesis of optical data-processing systems," *Proc. IEEE*, vol. 54, pp. 1055–1063, Aug. 1966.
[8] D. D. Givone and R. P. Roesser, "Multidimensional linear iterative circuits—general properties," *IEEE Trans. Comput.*, vol. C-21, pp. 1067–1073, Oct. 1972.
[9] ——, "Minimization of multidimensional linear iterative circuits," *IEEE Trans. Comput.*, vol. C-22, pp. 673–678, July 1973.
[10] B. Vilfan, "Another proof of the two-dimensional Cayley–Hamilton theorem," *IEEE Trans. Comput.* (Corresp.), vol. C-22, p. 1140, Dec. 1973.
[11] F. T. S. Yu, *Introduction to Diffraction, Information Processing, and Holography*. Cambridge, Mass. : M.I.T. Press, 1973.

Robert P. Roesser (S'67—M'69) was born in Bradford, Pa., on August 1, 1942. He received the B.S., MS., and Ph.D. degrees all in electrical engineering from the State University of New York at Buffalo, Buffalo, N.Y., in 1964, 1967, and 1970, respectively.

While at Buffalo he held a Graduate Assistantship, teaching various laboratory courses in the Department of Electrical Engineering. He was employed in 1970 as an Electrical Engineer in the Electronics Division, Houdaille Industries, where he was involved in the design of numerical and computer controls. In 1971 he joined the Department of Electrical Engineering, Wayne State University, Detroit, Mich., as an Assistant Professor where he is teaching courses and is engaged in active research in digital systems.

Estimation for Rotational Processes with One Degree of Freedom—Part I: Introduction and Continuous-Time Processes

JAMES TING-HO LO, MEMBER, IEEE, AND ALAN S. WILLSKY, MEMBER, IEEE

***Abstract*—A class of bilinear estimation problems involving single-degree-of-freedom rotation is formulated and resolved. Continuous-time problems are considered here, and discrete-time analogs will be studied in a second paper. Error criteria, probability densities, and optimal estimates on the circle are studied. An effective synthesis procedure for continuous-time estimation is provided, and a generalization to estimation on arbitrary Abelian Lie groups is included. Applications of these results to a number of practical problems including frequency demodulation will be considered in a third paper.**

Manuscript received July 30, 1973. Paper recommended by H. W. Sorenson, Chairman of the IEEE S-CS Estimation and Identification Committee. The work of A. S. Willsky was performed in part while the author was a Fannie and John Hertz Foundation Fellow in the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Mass., and in part at the Decision and Control Sciences Group, M. I. T. Electronic Systems Laboratory, Cambridge, Mass., under partial support from the AFOSR under Grant 72-2273. The work of J. T.-H. Lo was supported in part by the U. S. Office of Naval Research under the Joint Services Electronics Program by Contract N00014-67-A-0298-0006 while the author was with the Division of Engineering and Applied Physics, Harvard University, Cambridge, Mass.

J. T.-H. Lo is with the Division of Mathematics and Physics, University of Maryland, Baltimore County, Baltimore, Md. 21228.

A. S. Willsky is with the Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

I. Introduction

IN THE past, most optimal estimation problems have been studied in a vector space setting. While these results lend themselves to simple solutions in linear systems [1], [2] and in nonlinear systems with finite dimensional sensor orbits [3], no effective synthesis procedures for optimal estimation have been determined for large classes of nonlinear systems.

It is the purpose of this paper to introduce an alternative to the vector space approach in analyzing the properties of nonlinear stochastic processes. We will study random processes on a different type of space, namely, a differentiable manifold, which is the natural domain for certain nonlinear problems of practical importance. This approach will be shown to be useful both in analyzing the properties of certain stochastic processes and in deriving recursive optimal estimation equations that are easily implemented.

More specifically, we will concern ourselves with the study of random processes on the circle S^1 and its extensions to higher dimensions. Topics such as FM demodulation, frequency stability, and single-degree-of-freedom gyroscoptic analysis are well-known examples in this framework.

As an example of the type of problem that is of interest here, consider the following: suppose a random complex waveform

$$s(t)=e^{i\int_0^t x(s)ds} \tag{1}$$

is transmitted, where x is the (continuous) process to be recovered. As is often the case in optical communication problems [24], the waveform is corrupted by a multiplicative log normal noise—e.g., the received signal might be of the form

$$z(t)=e^{iv(t)}s(t) \tag{2}$$

where v is a Brownian motion process with

$$E\left[dv(t)^2\right]=q(t)dt. \tag{3}$$

Since the set of unit modulus complex numbers can be identified with the unit circle in R^2, we see that the estimation problem suggested by (1)–(3) has S^1 as its natural domain. As we shall see, our results provide the method needed for solving this and a number of related problems (see Part III of this series of papers for more on this specific problem).

It is appropriate to remark that in the paper we will use several representations of the circle interchangeably. A point on the unit circle can be represented either by the angle $\theta\in[-\pi,\pi)$ it makes with a fixed reference point or by the 2×2 orthogonal matrix

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Note that addition of θ_1 and θ_2 modulo 2π corresponds to the multiplication of the two matrices representing the points. Finally, as mentioned above, the set of unit modulus complex numbers $\{e^{i\theta}|\theta\in[-\pi,\pi)\}$ is a representation of S^1, and its relationship to the others is obvious.

The definition of Brownian motion on S^1 [21] provides the basis for our analysis of one-dimensional rotational processes. Consider the situation depicted in Fig. 1. We have a unit circle in R^2 with a straight line of infinite length tangent to it. We allow the line to perform a one-dimensional Brownian motion, fix the center of the circle, and require that there be no slipping at the point of tangency. The line induces a rotation of the circle, and if the line moves a distance x, the circle rotates x rad, and is thus in a position which is $x \bmod 2\pi=\theta$ rad away from its initial position.

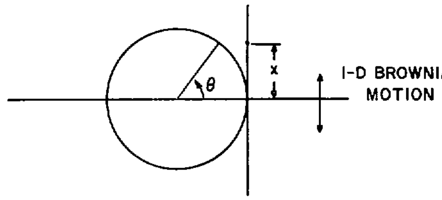

Fig. 1. Brownian motion on S^1.

The probability density function for θ satisfies the classical heat (diffusion, Fokker–Planck) equation on the circle:

$$\frac{\partial p_\theta}{\partial t}-\frac{1}{2}\frac{\partial^2 p_\theta}{\partial \xi^2}=0 \tag{4}$$

with the periodicity condition

$$p_\theta(\xi,t)=p_\theta(\xi+2\pi,t) \tag{5}$$

and initial condition

$$p_\theta(\xi,0)=\delta(\xi-\eta),\text{ a Dirac delta function} \tag{6}$$

where the initial orientation of the circle is η rad from some reference position. The solution of (4)–(6) is widely known, and is given by the two equivalent expressions

$$p_\theta(\xi,t)=\frac{1}{\sqrt{2\pi t}}\sum_{n=-\infty}^{+\infty} e^{-((\xi+2n\pi-\eta)^2/2t)} \tag{7}$$

$$=\frac{1}{2\pi}+\frac{1}{\pi}\sum_{n=1}^{\infty} e^{-n^2t/2}\cos n(\xi-\eta). \tag{8}$$

The density in (7) and (8) will be called the *folded normal density* since, as (7) indicates, it is obtained by "folding" a normal density about the circle. Levy [4] and Perrin [5] have done extensive work with this density.

In this paper we deal solely with continuous-time problems. We utilize this concept of "folding" a random process around the circle to define a large class of nonlinear signal and observation processes, and derive easily implemented optimal estimation equations for a class of bilinear problems. In Part II we will consider some discrete-time problems, and will also use Fourier series analysis as an aid in designing discrete-time estimation systems. The discrete-time problem is quite interesting, since it is much more difficult than its continuous-time counterpart, and the difficulty can be explained in a physically appealing manner. In Part III we will discuss some of the possible applications of the techniques to those designed using the optimal methods that we have developed. For more on these subjects, the reader is referred to [25] and [26].

II. Error Criteria and Optimal Estimates

Since the question of optimal estimation will be of central importance in the following sections, it is necessary to study how one uses the knowledge of the probability distribution of the quantity to be estimated to choose an estimate that gives the "best" performance, as measured by some predetermined figure of merit. In this section we present several results on the optimal estimation of random variables on the circle. For the sake of brevity, we state the results of this section without proof. Proofs can be found in [35]. We assume that we are given a random variable θ taking values in $[-\pi,\pi)$ with probability density $p(\theta)$ which is periodic with period 2π. Also, we assume that we have an error function ϕ, also periodic with period 2π, and we wish to choose $\hat{\theta}$ to minimize $E[\phi(\theta-\hat{\theta})]$. We wish to provide simple methods for computing the minimum of the cost and the value of $\hat{\theta}$ that achieves this minimum. To this end, we present a useful result, analogous to that of Sherman [10], [11].

We define the standard distance function (Riemannian metric) on the circle—i.e., the distance ρ between two points on the circle to be the arc length of the shortest path (geodesic line) joining them. If we restrict θ_1 and θ_2 to take values in the range $[-\pi,\pi)$, we have

$$\rho(\theta_1,\theta_2)=\min(|\theta_1-\theta_2|,2\pi-|\theta_1-\theta_2|). \tag{9}$$

The class of error criteria we wish to consider is the class of symmetric, nondecreasing cost functions—i.e., functions $\phi: S^1\rightarrow R$ which satisfy

$$0\leqslant\phi(\theta)=\phi(-\theta),\qquad 0\leqslant\rho(\theta_1,0)\leqslant\rho(\theta_2,0)\Rightarrow\phi(\theta_1)\leqslant\phi(\theta_2). \tag{10}$$

Some examples of cost criteria satisfying (10) are $\rho(\theta)\triangleq\rho(\theta,0)$, $(1-\cos\theta)$, $\rho(\theta)^2$, $(1-\cos\theta)^2$. We note that $\phi=1-\cos\theta$ was used in [34] to design a phase-tracking system. We also wish to consider the special class of unimodal, mode-symmetric probability density functions—i.e., density functions that have unique maxima and are symmetric about their maxima. As the following theorem demonstrates, under these conditions the mode of the density is the optimal estimate.

Theorem 1: Given an error function ϕ that satisfies (10) and a unimodal, mode-symmetric probability density function p, then

$$E(\phi(\theta-\eta))\leqslant E(\phi(\theta-\alpha)),\qquad \forall\alpha \tag{11}$$

where p has its maximum at η.

Note that the symmetry requirements of Theorem 1 are necessary. For instance, if ϕ is not symmetric, the mode of the density need not be the optimal estimate even if all the other assumptions of Theorem 1 do hold (see [35]). In addition, Theorem 1 still holds if a probability density does not exist, but the probability measure is unimodal at and symmetric about some point η (see [10], [25], [26]). Note also that the class of criteria ϕ that satisfies (10) is quite large, but the class of unimodal, mode-symmetric densities is somewhat special. However, in the continuous-time problems discussed in the subsequent sections, the folded normal density plays a central role, and as the next result indicates, Theorem 1 is applicable in this case.

Theorem 2: The folded normal density

$$F(\theta;\eta,\gamma)=\frac{1}{\sqrt{2\pi\gamma}}\sum_{n=-\infty}^{+\infty}e^{-((\theta+2n\pi-\eta)^2/2\gamma)} \tag{12}$$

is unimodal with mode at $\theta=\eta$ and is symmetric about η.

For the case in which $p(\theta)=F(\theta;\eta,\gamma)$, we can say a great deal more about an even larger class of error criteria than those satisfying (10). We remove the symmetry requirement but still require that ϕ be nondecreasing on $[0,\pi]$ and nonincreasing on $[-\pi,0]$ (a physically reasonable assumption). For such a ϕ, the mode η need not be the optimal estimate; however, for this discussion, we will take it as our estimate. The following theorem reveals an important property of the error $E(\phi(\theta-\eta))$.

Theorem 3: For ϕ satisfying the above requirement, and $p(\theta)=F(\theta;\eta,\gamma)$, $E(\phi(\theta-\eta))$ is an increasing function of the variance γ—that is,

$$\frac{d}{d\gamma}E(\phi(\theta-\eta))\geqslant 0. \tag{13}$$

The proof of this result relies on the following lemma, which yields more information about the shape of the folded normal density.

Lemma 1: For an arbitrary but fixed value of $\gamma>0$, there exists $\theta_0\in[0,\pi]$ such that

$$\frac{\partial^2}{\partial\theta^2}F(\theta;0,\gamma)<0,\qquad \theta\in[0,\theta_0) \tag{14}$$

$$\frac{\partial^2}{\partial\theta^2}F(\theta;0,\gamma)=0 \tag{15}$$

$$\frac{\partial^2}{\partial\theta^2}F(\theta;0,\gamma)>0,\qquad \theta\in(\theta_0,\pi] \tag{16}$$

that is, F has a unique inflection point (at θ_0) on $[0,\pi]$.

Note that by symmetry we have that F has a unique inflection point at $-\theta_0$ on the interval $[-\pi,0]$. Theorem 3 tells us that the intuitive notion that we "have more accurate information" for smaller values of γ can be made precise. Also, this theorem is the S^1 analog of a result obtained by Brown [14].

III. Signal Processes and Observation Processes

We now use the projection procedure introduced in Section I to formulate the mathematical models of signal and observation processes to be used in this paper. In doing this, we will find it convenient to use the 2×2 orthogonal matrix representation of S^1. Any element of this group has the direction cosine form displayed in Section I, and for small θ, we have the first-order approximation

$$\exp R\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \approx I + \theta R \tag{17}$$

where the matrix

$$R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{18}$$

is called the *infinitesimal rotation matrix*.

For those familiar with the theory of Lie groups, S^1 is a one-dimensional Abelian Lie group, with the 2×2 orthogonal matrices a representation of the group [sometimes called $SO(2)$]. The infinitesimal rotation R forms a basis for the Lie algebra $L(SO(2))$ of $SO(2)$, and the matrix exponential maps $L(SO(2))$ onto $SO(2)$.

Let $(\Omega, \mathscr{Q}, P)$ be a probability space and s a positive real number. Define C_1^s to be the space of real-valued continuous functions a on $[0,s]$ such that $a(0)=0$, and let C_2^s denote the set of $SO(2)$-valued functions A on $[0,s]$ such that $A(0)=I$. Finally, let $\mathfrak{B}_1^s$ and $\mathfrak{B}_2^s$ denote the Borel σ-fields of C_1^s and C_2^s, respectively. We will use lowercase letters to denote elements in C_1^s and uppercase letters to denote elements in C_2^s.

Let $J: C_1^s \to C_2^s$ be defined by

$$Ja(t) = \exp(a(t)R) = \begin{bmatrix} \cos a(t) & \sin a(t) \\ -\sin a(t) & \cos a(t) \end{bmatrix},$$

$$t \in [0,s]. \tag{19}$$

It is clear that J is $\mathfrak{B}_1^s$-measurable. It is claimed that J is also bijective. To show this, we first note that J is the same as the map

$$a(t) \to a(t) \bmod 2\pi \triangleq \theta_a(t) \tag{20}$$

where we identify $SO(2)$ with $[-\pi,\pi)$. The processes of breaking up a to form θ_a and of piecing θ_a together to form a continuous (R^1-valued) function can easily be seen to be inverses of one another (see [35]). Intuitively, what this says is the following: if we continuously watch a rotating object, we know not only its present orientation, but also the number of rotations it has performed. Here $a(t)$ is the total angle swept and $\theta_a(t)$, also represented by $Ja(t) \in SO(2)$, is the present orientation. The inverse of J can be seen to be equal to

$$(J^{-1}(A))(t) = \int_0^t [A'(s)dA(s)]_{12}. \tag{21}$$

Therefore, given a probability space $(\Omega, \mathscr{Q}, P)$ and any continuous $\mathscr{Q}$-measurable random process $Y: \Omega \to C_2^s$, there corresponds a unique $\mathscr{Q}$-measurable random process $y: \Omega \to C_1^s$ such that

$$Y(t) = (Jy)(t), \qquad t \in [0,s]. \tag{22}$$

We now wish to specify our signal process pair (x,X) and observation pair (z,Z) where x and z are real valued and X and Z are obtained from them via (22). For the time being, let (x,X) be any C_1^s, C_2^s pair related by (22). We will consider special cases and nonzero initial conditions later. Depending on the particular problem, we will regard either x or X as the signal process. We remark that the probability distributions of $x(t)$ and $X(t)$ are related by "folding" the distribution for $x(t)$ around the circle. Letting $\theta_x(t)$ be the representation of $X(t)$ as an element of $[-\pi,\pi)$, and choosing any Borel subset A of $[-\pi,\pi)$, we have

$$P(\theta_x(t) \in A) = P\left(x(t) \in \bigcup_{n=-\infty}^{+\infty} (A + 2n\pi)\right)$$

$$= \sum_{n=-\infty}^{+\infty} P(x \in A + 2n\pi). \tag{23}$$

We define a random process $z: \Omega \to C_1^s$ satisfying the Itô differential equation

$$dz(t) = h(x(t),t)dt + r^{1/2}(t)dw(t), \qquad z(0)=0 \tag{24}$$

where $w: \Omega \to C_1^s$ is a standard R^1 Brownian motion process independent of x, $h: R^1 \times R^1 \to R^1$ is Borel measurable, and $r: R^1 \to R^1$ is positive and Borel measurable. Let $Z: \Omega \to C_2^s$ be defined by $Z = J(z)$. Applying the Itô differential rule, we obtain the following matrix Itô differential equation ($Z(0)=I$):

$$dZ(t) = \begin{bmatrix} -\frac{r(t)}{2}dt & dz(t) \\ -dz(t) & -\frac{r(t)}{2}dt \end{bmatrix} Z(t)$$

$$= \left(-\frac{r(t)}{2}Idt + Rdz(t)\right)Z(t) \tag{25}$$

where the diagonal terms $-r(t)dt/2$ are the second-order correction terms given by Itô stochastic calculus. These terms are precisely what is needed to insure that $Z(t)$ remains an orthogonal matrix (i.e., that it stays on the circle). We note that $Z(t)$ can be written in the form

$$Z(t) = \exp\left(R\int_0^t h(x(s),s)ds + R\int_0^t r^{1/2}(s)dw(s)\right)$$

$$= \exp\left(R\int_0^t r^{1/2}(s)dw(s)\right)\exp\left(R\int_0^t h(x(s),s)ds\right). \tag{26}$$

In the second line of (26), we see the inherently multiplicative nature of the observation noise. This will be referred to in Part III in relation to log normal noise in optical communications.

We assume that our sensor observes either $Z(t)$ or $dZ(t)$. In this case, (25) is the Itô differential equation that represents the sensor dynamics. This model deserves some further comment. We first note that we shall find that the differential equations for estimation—either the Kushner-type partial differential equations (41) and (45), or the optimal estimation equations in Section V—explicitly involve dZ. Thus, if we observe Z, we will have to differentiate it to obtain dZ. Also, in (24) and (25) we have assumed that $z(0)=0$ and $Z(0)=I$. However, if dZ is taken as the observation (as it is in devices like integrating gyros, which provide incremental angular change information), the initial condition is unimportant. The reader is referred to Part III, in which we discuss a scheme that avoids differentiating Z but retains the property of being independent of the initial condition. In this case we can drop the assumption that $z(0)=0$ and $Z(0)=I$.

We note that the input to the sensor is not $X(t)$ but $x(t)$, and if we interpret X as the orientation of some object, then x is the total angle swept. Thus, taking x as the input to the sensor reflects the bijectivity of J—i.e., in observing a rotational process $X(t)$, our observation yields information concerning the total rotation $x(t)$. Referring to (24), if we assume that $h(x,s)=x$ and $r=1$, we have

$$z(t)=\int_0^t x(s)ds+w(t). \tag{27}$$

If, as with an integrating gyro or some other angle-detecting sensor, z and Z are measures of angular orientation, then x should be an angular velocity, as opposed to an angle. Thus, it would be physically appealing to consider a problem of the following type: let $x_1=$total angle swept and $x_2=$angular velocity; in this case the dynamical rotation state is $X_1=J(x_1)$. If we take noisy measurements of the total angle swept, our measurement process can be written as

$$z(t)=x_1(t)+w(t)=\int_0^t x_2(s)ds+w(t). \tag{28}$$

As we shall see, the solution of the scalar estimation problem with sensor dynamics given by (24) and (25) will lead *directly* to the solution of multidimensional problems, including the one just described. Also, the scalar results are directly applicable to some practical problems such as frequency demodulation (see Part III). For the multidimensional results, the reader is referred to the part of Section V on the general Abelian Lie group problem and to the examples (in particular, see Example 1).

Finally, we note that an observation equivalent to $Z(t)$ is $y(t)=(\cos z(t),\ \sin z(t))$—i.e., we do not need the full 2×2 matrix. In fact, we will use this measurement pair formulation in the examples of Section VII. However, in the mathematical development of Sections IV–VI, we will continue to write the $SO(2)$ equations.

IV. Conditional Probability Distributions

The problem considered in this section is to determine the conditional probability measures $P(x(\lambda)\in A|Z(\tau),\ \tau\in[0,t])$ and $P(X(\lambda)\in B|Z(\tau),\ \tau\in[0,t])$ where A is a Borel subset of R^1 and B is a Borel subset of $SO(2)$. We remark that the physical motivation for determining the distribution for $x(\lambda)$ comes from such problems as the frequency demodulation problem [28]–[31]. The physical motivation for finding the distribution for $X(\lambda)$ is related to the problems of orientation estimation and phase synchronization and tracking.

In the rest of this section we will use z^t to denote $\{z(\tau),\tau\in[0,t]\}$ and Z^t for $\{Z(\tau),\tau\in[0,t]\}$. In addition, we will sometimes write $x(t)$ as x_t and $X(t)$ as X_t. Let $\mathfrak{L}_1$ be the σ-algebra of Borel subsets of R^1 and $\mathfrak{L}_2$ the Borel subsets of $SO(2)$. The various processes induce probability measures on either $(R^1,\mathfrak{L}_1)$ or $(SO(2),\mathfrak{L}_2)$. For instance,

$$\nu_{x_t}(A)\triangleq P(x_t^{-1}(A)),\qquad A\in\mathfrak{L}_1 \tag{29}$$

$$\nu_{X_t}(B)\triangleq P(X_t^{-1}(B)),\qquad B\in\mathfrak{L}_2. \tag{30}$$

Since J is bijective, the information contained in z^t is the same as that in Z^t. Mathematically, this means that the σ-subfield of $\mathcal{Q}$ generated by z^t is the same as that generated by Z^t, and it will be denoted by $\mathcal{Q}_z^t$. The σ-subalgebra of $\mathcal{Q}$ generated by x_λ will be denoted by $\mathcal{Q}_{x_\lambda}$, and $\mathcal{Q}_{X_\lambda}$ denotes that generated by X_λ.

For the present, we assume that the times λ and t are fixed, and for simplicity, we drop them as subscripts and superscripts (we simply note that $\lambda<t$ corresponds to a smoothing problem, $\lambda=t$ is the filtering problem, and $\lambda>t$ is the prediction problem). Let P_{xz} be the conditional probability measure on $(\Omega,\mathcal{Q}_x)$ given $\mathcal{Q}_z$:

$$P_{xz}(A,\omega)=P(A|\mathcal{Q}_z)(\omega),\qquad A\in\mathcal{Q}_x,\omega\in\Omega. \tag{31}$$

This measure induces a measure on $(R^1,\mathfrak{L}_1)$:

$$\nu_{xz}(A,\omega)=P_{xz}(x_\lambda^{-1}(A),\omega),\qquad A\in\mathfrak{L}_1. \tag{32}$$

We define P_{Xz} on $(\Omega,\mathcal{Q}_X)$ and ν_{Xz} on $(SO(2),\mathfrak{L}_2)$ analogously. We also note that since ν_{xz} and ν_{Xz} are $\mathcal{Q}_z^t$-measurable, they can be thought of as explicit $\mathfrak{B}_2^t$ measurable functions of Z^t.

The problem is to determine equations for $\nu_{xz}(A,Z^t)$ and $\nu_{Xz}(B,Z^t)$. Because of the equivalence of z^t and Z^t, the equations can be obtained from standard vector space results (see details in [25], [20], [16], [15]). The measure ν_{xz} can be shown [15], [16], [25] to be absolutely continuous with respect to ν_x, $(\nu_{xz}\ll\nu_x)$, and the Radon–Nikodym derivative is given by

$$\frac{d\nu_{xz}}{d\nu_x}(x,Z^t)=\frac{\mathcal{E}(\Theta^t|x_\lambda=x)}{\mathcal{E}(\Theta^t)} \tag{33}$$

where

$$\Theta^t=\exp\left(-\frac{1}{2}\int_0^t\frac{h^2(x(\tau),\tau)}{r(\tau)}d\tau + ⨍_0^t\frac{h(x(\tau),\tau)}{r(\tau)}[Z'(\tau)dZ(\tau)]_{12}\right) \quad (34)$$

where ⨍ denotes an Itô stochastic integral. Here $dz(\tau)=[Z'(\tau)dZ(\tau)]_{12}$ and the expectation is over x^t with Z^t fixed.

Suppose that $\nu_x \ll \mu$ where μ is Lebesgue measure on R^1. In this case, it is clear that ν_{xz} is also absolutely continuous with respect to μ, and denoting this density by $p_x(x,\lambda|Z^t)$, we have

$$p_x(x,\lambda|Z^t)=\frac{\mathcal{E}(\Theta^t|x_\lambda=x)}{\mathcal{E}(\Theta^t)}p_x(x,\Lambda). \quad (35)$$

Instead of ν_{Xz}, we now consider the measure $\nu_{\theta z}$ where θ is the representation of X on $[-\pi,\pi)$. Then, from (23)

$$\nu_{\theta z}(A)=\sum_{n=-\infty}^{+\infty}\nu_{xz}(A+2n\pi). \quad (36)$$

Then, if (35) holds, we have the following result.

Theorem 4: Consider a continuous signal process pair (x,X) related by (22) and a measurement process defined by (24) and (25). If the unconditional density for x_λ, $p_x(x,\lambda)$ exists, then the conditional densities $p_x(x,\lambda|Z^t)$, $p_\theta(\theta,\lambda|Z^t)$ for x_λ and θ_λ, respectively, exist and are given by

$$p_x(x,\lambda|Z^t)=\frac{\mathcal{E}(\Theta^t|x_\lambda=x)}{\mathcal{E}(\Theta^t)}p_x(x,\lambda) \quad (37)$$

$$p_{\theta_x}(\theta,\lambda|Z^t)=\sum_{k=-\infty}^{+\infty}\frac{\mathcal{E}(\Theta^t|x_\lambda=\theta+2k\pi)}{\mathcal{E}(\Theta^t)}p_x(\theta+2k\pi,\lambda) \quad (38)$$

where Θ^t is defined by (34).

We can now consider various values of λ. The one with which we shall mostly concern ourselves is $\lambda=t$—the filtering problem. If we assume that x is the solution of the Itô equation

$$dx(t)=a(x(t),t)dt+b^{1/2}(x(t),t)dv(t) \quad (39)$$

where v is a standard Brownian motion process, then X satisfies

$$dX(t)=[-\tfrac{1}{2}b(x(t),t)Idt+Rdx(t)]X(t). \quad (40)$$

We remark that although x is a Markov process, X need not be. In fact, we can show that X is a Markov process if and only if the right-hand side of (40) is periodic in x with period 2π [i.e., if and only if it depends only on $X(t)$]. Following [19], [20], and [6], we can derive the stochastic partial differential equations for $p_x(x,\lambda|Z^t)$ and $p_\theta(\theta,\lambda|Z^t)$, assuming that the densities are sufficiently smooth:

$$dp_x(x,t|Z^t)=(A^*p_x)(x,t|Z^t)dt + \frac{\left(h(x,t)-\hat{h}_t\right)}{r(t)}\left([Z'(t)dZ(t)]_{12}-\hat{h}_t dt\right)p_x(x,t|Z^t) \quad (41)$$

where

$$A^*(g)=-\frac{\partial(ag)}{\partial x}+\frac{1}{2}\frac{\partial^2(bg)}{\partial x^2} \quad (42)$$

$$\hat{h}_t=\mathcal{E}(h(x,t)|Z^t)=\int_{-\infty}^{+\infty}h(x,t)p_x(x,t|Z^t)dx \quad (43)$$

$$dp_\theta(\theta,t|Z^t)=\sum_{n=-\infty}^{+\infty}dp_x(\theta+2n\pi,t|Z^t). \quad (44)$$

In general, we cannot simplify (44) any further. However, suppose a, b, and h are periodic in x with period 2π. Then we have the equation

$$dp_\theta(\theta,t|Z^t)=(B^*p_\theta)(\theta,t|Z^t)dt + \frac{\left(h(\theta,t)-\hat{h}_t\right)}{r(t)}\left([Z'(t)dZ(t)]_{12}-\hat{h}_t dt\right)p_\theta(\theta,t|Z^t) \quad (45)$$

where

$$(B^*g)=-\frac{\partial(ag)}{\partial\theta}+\frac{1}{2}\frac{\partial^2(bg)}{\partial\theta^2}. \quad (46)$$

For $\lambda>t$, we have the prediction problem. Assume x is a Markov process and let $p_x(x,\lambda|y,t)$ be the transition density for x_λ given $x_t=y$. Then one has (see [17])

$$p_x(x,\lambda|Z^t)=\int_{-\infty}^{+\infty}p_x(x,\lambda|y,t)p_x(y,t|Z^t)dy. \quad (47)$$

Thus, the prediction result can be expressed explicitly in terms of the filtering solution.

For the smoothing problem $\lambda<t$, we can also express the conditional densities in terms of filtering (see [17], [18], [25]):

$$p_x(x,\lambda|Z^t)=p_x(x,\lambda|Z^\lambda)\cdot\exp\left(⨍_\lambda^t\frac{\alpha_s}{r(s)}dy(s)-\tfrac{1}{2}\int_\lambda^t\frac{\alpha_s^2}{r(s)}ds\right) \quad (48)$$

where

$$\alpha_s=\mathcal{E}(h(x(s),s)|Z^s,x_\lambda=x)-\mathcal{E}(h(x(s),s)|Z^s) \quad (49)$$

$$dy(s)=[Z'(s)dZ(s)]_{12}-\mathcal{E}(h(x(s),s)|Z^s). \tag{50}$$

For both the smoothing and filtering problems, we have

$$p_\theta(\theta,\lambda|Z^t)=\sum_{n=-\infty}^{+\infty}p_x(\theta+2n\pi,\lambda|Z^t). \tag{51}$$

Thus, we have obtained differential equations for the conditional densities of the R^1 process x and the S^1 process θ.

V. Optimal Sequential Bilinear Estimation

In this section we will consider a subclass of the class of systems described in the previous section, and will use the tools of linear estimation theory to solve an optimal nonlinear estimation problem. Specifically, we consider the signal process pair (x,X) given by (22), (39), and (40) with the restriction that (39) be a linear Itô equation:

$$dx(t)=a(t)x(t)dt+b^{1/2}(t)dv(t), \qquad x(0)=0. \tag{52}$$

We assume that $b(t)>0 \forall t\in[0,T]$. In this case, we can write the bilinear Itô equation satisfied by X in two ways:

$$dX(t)=[-\tfrac{1}{2}b(t)I dt+R dx(t)]X(t) \tag{53}$$

or

$$dX(t)=-\tfrac{1}{2}b(t)X(t)dt$$
$$+\left\{a(t)\left[\int_0^t\left[\exp\int_s^t a(\tau)d\tau\right]b^{1/2}(s)dv(s)\right]dt + b^{1/2}(t)dv(t)\right\}RX(t) \tag{54}$$

$$x(t)=\int_0^t\left[\exp\int_s^t a(\tau)d\tau\right]b^{1/2}(s)dv(s). \tag{55}$$

The observation process to be used is

$$dZ(t)=\left[-\frac{r(t)}{2}I+c(t)x(t)R\right]Z(t)dt+r^{1/2}(t)RZ(t)dw(t) \tag{56}$$

with $Z(0)=I$ and where $r(t)>0$ $\forall t\in[0,T]$. As shown in Section III, $z=J^{-1}(Z)$ satisfies

$$dz(t)=c(t)x(t)dt+r^{1/2}(t)dw(t), \qquad z(0)=0. \tag{57}$$

Note that like (53), (56) is bilinear.

The problem now is to determine equations for the optimal filtering estimates of $x(t)$ and $X(t)$ given Z^t (the smoothing and prediction results are a straightforward extension [35]). Our choice of criteria will be the following: let $\phi_1:S^1\to R^1$ be any error function satisfying (10), and let $\phi_2:R^1\to R^1$ be symmetric about 0 and nondecreasing on the positive half-line. Our optimal estimates $\hat{x}(t|t)$ and $\hat{X}(t|t)$ are taken to be, respectively, the $\mathcal{B}_1^t$- and $\mathcal{B}_2^t$-measurable functions of Z^t such that

$$\mathcal{E}\left(\Phi_1(X(t),\hat{X}(t|t))|Z^t\right)\leq\mathcal{E}(\Phi_1(X(t),M)|Z^t) \tag{58}$$

$$\mathcal{E}(\phi_2(x(t)-\hat{x}(t|t))|Z^t)\leq\mathcal{E}(\phi_2(x(t)-\mu)|Z^t) \tag{59}$$

for all $\mathcal{Q}_z^t$-measurable random 2×2 orthogonal matrices M and all $\mathcal{Q}_z^t$-measurable real random variables μ where

$$\Phi_1(X_1,X_2)=\Phi_1(X_2,X_1)=\Phi_1(X_1X_2^{-1},I)=\phi_1(\theta). \tag{60}$$

Here θ is the $[-\pi,\pi)$ representation of $X_1^{-1}X_2$.

We first solve a well-known problem. Since z^t and Z^t generate the same σ-algebra $\mathcal{Q}_z^t$, $\mathcal{E}(x(t)|\mathcal{Q}_z^t)$ is both a $\mathcal{B}_1^t$-measurable function f_1 of z^t and a $\mathcal{B}_2^t$-measurable function f_2 of Z^t, and $f_2=f_1\circ J^{-1}$. In terms of z^t, the solution of (59) for $x(t)$ and $z(t)$ given by (52) and (57) is well known [1], [2], [6]. The conditional density $p_x(x,t|z^t)$ is a normal density

$$p_x(x,t|z^t)=\frac{1}{\sqrt{2\pi P(t)}}\exp\left[-\frac{(x-\hat{x}_{t|t})^2}{2P(t)}\right] \tag{61}$$

where Kalman–Bucy linear filtering theory yields the equations

$$d\hat{x}_{t|t}=a(t)\hat{x}_{t|t}dt+P(t)c(t)r^{-1}(t)(dz(t)-c(t)\hat{x}_{t|t}dt) \tag{62}$$

$$\hat{x}_{0|0}=0 \tag{63}$$

$$\dot{P}(t)=2a(t)P(t)-c^2(t)r^{-1}(t)P^2(t)+b(t), \qquad P(0)=0. \tag{64}$$

Then by the R^1 version of Theorem 1, since the normal density is unimodal and symmetric,

$$\mathcal{E}(\phi_2(x(t)-\hat{x}_{t|t})|z^t)\leq\mathcal{E}(\phi_2(x(t)-\mu)|z^t) \tag{65}$$

for all $\mathcal{Q}_z^t$-measurable μ. Thus, the optimal estimate in terms of z^t is $\hat{x}_{t|t}=f_1(z^t)$, so the optimal estimate as a function of Z^t is just $\hat{x}(t|t)=f_1(J^{-1}(Z^t))$, and we have proven the following.

Theorem 5: Let x, the signal process, be given by (52) and let (56) be the observation process. Then the optimal filtering equations are

$$d\hat{x}(t|t)=a(t)\hat{x}(t|t)dt$$
$$+P(t)c(t)r^{-1}(t)([Z'(t)dZ(t)]_{12}b\text{-}c(t)\hat{x}(t|t)dt) \tag{66}$$

$$\hat{x}(0|0)=0 \tag{67}$$

where P is defined by (64).

We now turn to the problem of estimating $X(t)$ using the criterion (58). An alternative representation for $X(t)$ is $\theta_x(t)=x(t)\bmod 2\pi$, and using (61) and (38), we have that the conditional density $p_\theta(\theta,t|Z^t)$ for $\theta_x(t)$ given Z^t is the folded normal density $F(\theta;\hat{x}(t|t),P(t))$. Then, from Theorems 1 and 2, we have that the estimate $\hat{\theta}_x(t|t)$ that minimizes $\mathcal{E}(\phi_1(\theta_x(t)-\hat{\theta}_x(t|t))|Z^t)$ (over the class of all $\mathcal{Q}_z^t$-measurable functions) is

$$\hat{\theta}_x(t|t)=\hat{x}(t|t)\bmod 2\pi \tag{68}$$

and the estimate $\hat{X}(t|t)$ that is optimal with respect to (58) is

$$\hat{X}(t|t)=\exp[\hat{x}(t|t)R]. \tag{69}$$

Note that

$$\mathcal{E}(\exp[x(t)R]|Z^t)=e^{-P(t)/2}\exp[\hat{x}(t|t)R] \tag{70}$$

so the optimal estimate in the sense of (58) is *not* the conditional expectation unless $P^2(t)=0$—i.e., unless we know exactly where we are on the circle and the line.

We can write a stochastic differential equation for $\hat{X}(t|t)$, and we include this in the following theorem which has just been proven.

Theorem 6: Let X, the signal process, be given by (54), and let (56) be the measurement process. Then, with respect to (58), the optimal filtering equations are

$$\begin{aligned} d\hat{X}(t|t)= &-\tfrac{1}{2}P^2(t)c^2(t)r^{-1}(t)\hat{X}(t|t)dt \\ &+\Bigg\{a(t)-P(t)c^2(t)r^{-1}(t) \\ &\cdot\left[\int_0^t\left[\exp\int_s^t(a(\tau)-P(\tau)c^2(\tau)r^{-1}(\tau))d\tau\right]\right. \\ &\left.\cdot P(s)c(s)r^{-1}(s)[Z'(s)dZ(s)]_{12}\right]dt \\ &+P(t)c(t)r^{-1}(t)[Z'(t)dZ(t)]_{12}\Bigg\}R\hat{X}(t|t) \end{aligned} \tag{71}$$

$$\hat{X}(0|0)=I. \tag{72}$$

The expected error $\mathcal{E}(\Phi_1(X(t),\hat{X}(t|t))|Z^t)$ of the optimal estimate $\hat{X}(t|t)$ can be obtained by straightforward computation. For example, if $\phi_1(\theta)=1-\cos\theta$, we have

$$\mathcal{E}\big(\Phi_1\big(X(t),\hat{X}(t|t)\big)\big)=1-\exp(-\tfrac{1}{2}P(t)). \tag{73}$$

See [35] for other examples.

The filter equation (71) is quite complex; however, using its relationship with $\hat{x}(t|t)$, we have the easily implemented optimal filter depicted in the block diagram of Fig. 2.

The measurement process dZ is processed by a nonlinear transformer that yields $dz=[Z'dZ]_{12}$ as its output. This process then goes through a Kalman–Bucy filter that computes $\hat{x}(t|t)$, which is then injected into S^1 via the map J to produce the desired estimate $\hat{X}(t|t)$. The concept

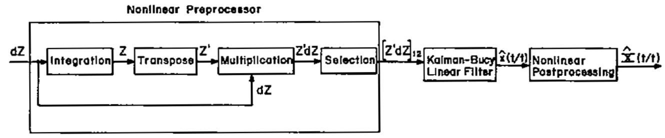

Fig. 2. Block diagram for optimal filtering.

of preprocessing and postprocessing a signal with a linear filter in the middle has also been discussed by Oppenheim *et al.* [32]. Also, as mentioned in the previous section, the differential nature of the observation means that the results are unchanged if $Z(0)\neq I$. This is of importance, for instance, in the frequency demodulation problem where we observe only $\sin z(t)$ and must construct $\cos z(t)$ from it. This will be discussed in detail in Part III. In addition, in Part III we will discuss an alternative nonlinear preprocessor that avoids differentiating $Z(t)$.

The smoothing and prediction problems can be solved in a manner analogous to the solution of the filtering problem—i.e., we can sandwich the solution to the linear smoothing or prediction problem between the nonlinear pre- and postprocessors of Fig. 2. For details, see [35].

Before closing this section, we make some comments concerning random initial conditions for x and X. If we do not specify initial conditions, the map J defined in (29) is *not* invertible. In fact, one can show that $J(a)=J(b)$ if and only if

$$b(t)-a(t)=b(0)-a(0)=2n\pi, \qquad \forall t. \tag{74}$$

Thus, we cannot uniquely recover x from X if we allow nonzero initial conditions. However, as long as we assume that we are given a, the process $J(a)$ is uniquely defined. For instance, in the results of this section, we can assume that $x(0)$ has the density $N(x;\hat{x}(0|0),P(0))$, and it is clear that the initial conditions for $\hat{x}(t|t)$, $P(t)$, $\hat{X}(t|t)$ in Theorems 5 and 6 can be replaced with $\hat{x}(0|0)$, $P(0)$, $\exp[R\hat{x}(0|0)]$. If we are given X instead of x, we may be able to make an assumption about $x(0)$ so that we can apply the results of this section.

An alternative to this approach is the following. Let Y denote the dynamical state with random initial condition Y_0, independent of all other random processes. We observe that the input to the observation process (56) at time t is the angle that the rotational process represented by the signal has swept over the time interval $[0,t]$. Taking this viewpoint, our present problem can be solved with some modification to the previous results.

Let $y(t)$ denote the angle that the signal Y has swept over $[0,t]$. It is easily seen that

$$y(t)=\int_0^t[Y'(s)dY(s)]_{12}. \tag{75}$$

Let $X(t)=Y_0^{-1}Y(t)$. Then $X(0)=I$ and, as before, we may define

$$x(t)=(J^{-1}(X))(t)=\int_0^t[X'(s)dX(s)]_{12}. \tag{76}$$

We note that $x(t)=y(t)$. In other words, the angles swept by X and by Y over $[0,t]$ are the same. Hence, (56) can also be used as the observation process for our present problem. The conditional distribution of $X(\lambda)$ given observation Z^t of the form given in (56) can be determined by the application of the previous results.

We note that Y_0 and $X(\lambda)$ are conditionally independent given Z^t. If the distribution of Y_0 and the conditional distribution of $X(\lambda)$ given Z^t are both folded normal, then the following lemma easily leads to the conclusion that the optimal estimate $\hat{Y}(\lambda|t)$ of $Y(\lambda)$ given Z^t is equal to $\hat{Y}_0\hat{Y}(\lambda|t)$ where $\hat{Y}_0$ is the mode of the distribution of Y_0 and $\hat{X}(\lambda|t)$ is the mode of the conditional distribution of $X(\lambda)$ given Z^t.

Lemma 2: Let A and B be two independent 2×2 orthogonal random matrices which have folded normal distributions with modes $\hat{A}$ and $\hat{B}$, respectively. Then AB is a 2×2 orthogonal random matrix which has a folded normal distribution with mode equal to $\hat{A}\hat{B}$.

Proof: See [35].

VI. Multichannel Estimation

The results of the previous subsections can be extended to a large class of problems—those involving processes evolving on Abelian Lie groups. It is well known [23] that a given Abelian Lie group G is isomorphic to the direct product of a number of copies of the circle and a number of copies of the real line, i.e., $G \approx R^n \times (S^1)^m$ where $(S^1)^m$ is usually called a "torus." The diffusion processes on this type of space have been used to model some interesting satellite and pendulum systems in [33]. Analogous to (19), a bijective mapping $J_{nm}:(C_1^s)^{n+m}\rightarrow(C_1^s)^n\times(C_2^s)^m$ is defined by

$$(J_{nm}a)(t)=[a_1(t),\cdots,a_n(t),(Ja_{n+1})(t),\cdots,(Ja_{n+m})(t)] \tag{77}$$

for $a\in(C_1^s)^{n+m}$, a_i being the ith component of a. Then a continuous $\mathcal{C}$-measurable random process $X:\Omega\rightarrow(C_1^T)^n\times(C_2^T)^m$ on G corresponds to a unique continuous $\mathcal{C}$-measurable random process $x:\Omega\rightarrow(C_1^T)^{n+m}$ on R^{n+m} via the identification $X=J_{nm}(x)$. Let $x=(x_1,\cdots,x_{n+m})'$ be a normally distributed $(n+m)$-dimensional random variable with density $N_{n+m}(x;\eta,P)$ where $\eta=\mathcal{E}(x)$ and $P=\mathcal{E}[(x-\eta)(x-\eta)']$. Then the density $F_{nm}(y;\eta,P)$ for y, which is defined by

$$y=(x_1,\cdots,x_n,x_{n+1}\bmod 2\pi,\cdots,x_{n+m}\bmod 2\pi)', \tag{78}$$

is called the (n,m)-folded normal density. Note that n of the marginal densities (those for x_1 through x_n) are normal and the other m are folded normal.

We can now consider the vector analog of the processes described by (22) and (39). We can model our observation process analogously. Let z be the $(l+k)$-dimensional solution of the vector Itô equation

$$dz(t)=h(x(t),t)dt+Q^{1/2}(t)dv(t), \qquad z(0)=0 \tag{79}$$

and define our observation process $Z\in(C_1^T)^l x(C_2^T)^k$ by $Z=J_{lk}(z)$. We can write stochastic differential equations for X and Z analogous to (40) and (25) by means of the Itô differential rule. The calculation is straightforward and will not be displayed. By the objectiveness of J_{nm}, it is clear that the preceding analysis extends directly to the multidimensional case. Furthermore, if the Itô equations for x and z are linear, the block diagram of Fig. 2 is conceptually correct. The nonlinear preprocessor becomes $(J_{lk})^{-1}$, the postprocessor is J_{nm}, and we use the vector version of the Kalman–Bucy filter.

Thus, we have shown that the domain of the Kalman–Bucy filter includes estimation on Abelian Lie groups. Also, since R^{n+m} can be identified with the Lie algebra of $R^n\times(S^1)^m$, we see that the filtering technique involves the processing of the observations so that the actual filtering is done in the Lie algebra—a vector space.

Note that the isomorphism between G and $R^n\times(S^1)^m$ may be such that the actual noise processes are quite interesting. For example, the Lie group of positive real numbers under multiplication is isomorphic to R^1 under addition by the map $a\rightarrow e^a$. Thus, multiplicative amplitude changes can be looked at as additive noises. We will have more to say about this type of problem in Part III.

VII. Examples

Example 1: Consider a cylindrical shaft of unit radius being spun about its longitudinal axis by an electric motor. We assume that the total rotation of the shaft x_1 is related to the driving force u by the differential equation

$$\ddot{x}_1+\dot{x}_1+x_1=u \tag{80}$$

with both $x_1(0)$ and $\dot{x}_1(0)$ equal to zero. The driving force consists of a known force and a disturbance. The known force adds neither difficulty to the analysis nor complexity to the solution. Thus, for simplicity, we assume that the known force is zero and that the disturbance is white Gaussian noise $\dot{v}$ with $\mathcal{E}(\dot{v}(t)\dot{v}(\tau))=\delta(t-\tau)$. Setting $x_2=\dot{x}_1$, we obtain the vector differential equation

$$dx(t)=Ax(t)dt+Bdv(t), \qquad x(0)=0 \tag{81}$$

where

$$x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad A=\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}, \qquad B=\begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{82}$$

Suppose we wish to determine the orientation of the shaft—i.e., $X_1(t)=(Jx_1)(t)$. Then our estimation state space is $R^1\times S^1$ with

$$(J_{1,1}(x_1,x_2))(t)=(x_2(t),X_1(t)). \tag{83}$$

The orientation is determined by

$$[X_1(t)]_{11}=\cos x_1(t)=\cos\int_0^t x_2(\tau)d\tau \tag{84}$$

$$[X_1(t)]_{12}=\sin x_1(t)=\sin\int_0^t x_2(\tau)d\tau. \tag{85}$$

Suppose we measure these quantities, but that noise corrupts our observations, so we actually observe

$$z_1(t)=\cos(x_1(t)+w(t)),\ z_2(t)=\sin(x_1(t)+w(t)) \tag{86}$$

where w is a Brownian motion of unit strength independent of v. Note that z_1 and z_2 are the $(1,1)$ and $(1,2)$ components of

$$Z(t)=\exp R\left[\int_0^t x_2(\tau)d\tau+w(t)\right] \tag{87}$$

and satisfy the equations

$$dz_1(t)=-\tfrac{1}{2}z_1(t)dt-x_2(t)z_2(t)dt-z_2(t)dw(t),\quad z_1(0)=1 \tag{88}$$

$$dz_2(t)=-\tfrac{1}{2}z_2(t)dt+x_2(t)z_1(t)dt+z_1(t)dw(t),\quad z_2(0)=0. \tag{89}$$

Using the results of Section VI, we have the following optimal filter equations:

$$d\hat{x}(t|t)=A\hat{x}(t|t)dt+P(t)c'[z_1(t)dz_2(t)-z_2(t)dz_1(t)-c\hat{x}(t|t)dt] \tag{90}$$

where $c=[0,1]$, $\hat{x}(0|0)=0$, and P is the solution of the Riccati equation $(P(0)=0)$.

$$\dot{P}(t)=AP(t)+P(t)A'-P(t)c'cP(t)+BB'. \tag{91}$$

The optimal estimate of the orientation is

$$\hat{X}_1(t|t)=\exp(\hat{x}_1(t|t)R). \tag{92}$$

Note that, as in the standard linear case, we can compute an optimal steady-state filter by replacing $P(t)$ in (90) with the positive definite solution of the algebraic Riccati equation.

Example 2: In this example, we will take a set of nonlinear signal and observation equations and show that by interpreting them as signals on $S^1\times R^2$ and $S^1\times R^1$, respectively, we can use the results of Sections V and VI to solve the optimal filtering problem. The signal process is specified by the following equations:

$$dx_1=-\tfrac{1}{2}x_1dt-x_2x_4dt-x_2dv,\qquad x_1(0)=1 \tag{93}$$

$$dx_2=-\tfrac{1}{2}x_2dt+x_1x_4dt+x_1dv,\qquad x_2(0)=0 \tag{94}$$

$$dx_3=x_4dt+dv,\qquad x_3(0)=0 \tag{95}$$

$$\dot{x}_4=-x_3,\qquad x_4(0)=0 \tag{96}$$

$$\dot{x}_5=x_4+x_5,\qquad x_5(0)=0 \tag{97}$$

where v is a unit strength Brownian motion process. The measurement equations are

$$dz_1=-\tfrac{1}{2}z_1dt-\left(2x_5+x_4-\int_0^t x_1(s)dx_2(s)+\int_0^t x_2(s)dx_1(s)\right)z_2(t)dt-z_2dw_1 \tag{98}$$

$$dz_2=-\tfrac{1}{2}z_2dt+\left(2x_5+x_4-\int_0^t x_1(s)dx_2(s)+\int_0^t x_2(s)dx_1(s)\right)z_1dt+z_1dw_1 \tag{99}$$

$$dz_3=\left(\int_0^t x_1(s)dx_2(s)-\int_0^t x_2(s)dx_1(s)\right)dt+dw_2 \tag{100}$$

$$z(0)=1\qquad z_2(0)=z_3(0)=0 \tag{101}$$

where w_1 and w_2 are standard Brownian motions, independent of each other and of v. The problem is to find the least squares estimates $\hat{x}_1$ and $\hat{x}_2$ for x_1 and x_2 under the constraint $\hat{x}_1^2+\hat{x}_2^2=1$.

It is easy to show that $x_1=\cos x_3, x_2=\sin x_3$, and therefore, our optimal filter need only estimate x_3, x_4, and x_5 to get estimates of x_1 and x_2. Also, (98) and (99) yield

$$\begin{bmatrix} dz_1 & dz_2 \\ -dz_2 & dz_1 \end{bmatrix}=-\frac{1}{2}\begin{bmatrix} z_1 & z_2 \\ -z_2 & z_1 \end{bmatrix}dt+\begin{bmatrix} z_1 & z_2 \\ -z_2 & z_1 \end{bmatrix}R[(2x_5+x_4-x_3)dt+dw_1]. \tag{102}$$

Note that the system is not observable with just the S^1 observation pair $\{z_1,z_2\}$ or with just the R^1 observation z_3, but the system is observable with both observations.

Following the approach developed in Sections V–VI, we have the optimal filter equations

$$\begin{bmatrix} d\hat{x}_3 \\ d\hat{x}_4 \\ d\hat{x}_5 \end{bmatrix}=A\begin{bmatrix} \hat{x}_3 \\ \hat{x}_4 \\ \hat{x}_5 \end{bmatrix}dt+PC'\left\{\begin{bmatrix} z_1dz_2-z_2dz_1 \\ dz_3 \end{bmatrix}-C\begin{bmatrix} \hat{x}_3 \\ \hat{x}_4 \\ \hat{x}_5 \end{bmatrix}\right\}dt,$$

$$[\hat{x}_3(0),\hat{x}_4(0),\hat{x}_5(0)]=0 \tag{103}$$

where

$$\dot{P}=AP+PA'-PC'CP+BB' \qquad P(0)=0 \quad (104)$$

$$A=\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad B=\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad C=\begin{bmatrix} -1 & 1 & 2 \\ 1 & 0 & 0 \end{bmatrix}. \quad (105)$$

With the help of previous results,

$$\mathcal{E}(1-\cos(x_3-\hat{x}_3)) \leqslant \mathcal{E}(1-\cos(x_3-\xi)) \quad (106)$$

for all z^t-measurable ξ. Hence,

$$\begin{aligned}\tfrac{1}{2}\mathcal{E}\left[(\cos x_3-\cos\hat{x}_3)^2+(\sin x_3-\sin\hat{x}_3)^2\right] \\ = 1-\mathcal{E}[\cos(x_3-\hat{x}_3)] \leqslant 1-\mathcal{E}[\cos(x_3-\xi)] \\ = \tfrac{1}{2}\mathcal{E}\left[(\cos x_3-\cos\xi)^2+(\sin x_3-\sin\xi)^2\right] \quad (107)\end{aligned}$$

and thus, the optimal constrained least squares estimates are

$$\hat{x}_1=\cos\hat{x}_3, \qquad \hat{x}_2=\sin\hat{x}_3. \quad (108)$$

VIII. Conclusions

In this paper, a class of estimation problems on the unit circle is formulated and resolved. The signal and observation processes on the circle are constructed by taking the projection modulo 2π of the corresponding standard one-dimensional processes. The stochastic differential equations which govern their evolution are bilinear in form. The observational noise can be viewed as entering multiplicatively.

Error criteria, probability distributions, and optimal estimates on the circle are studied. In particular, various properties of the folded normal density in connection with estimation are discussed in detail.

An effective and physically appealing synthesis procedure for continuous-time estimation is provided. The measurement data are first processed through a nonlinear transformation. The transformed process then goes through an ordinary estimator, such as the Kalman–Bucy filter. After another nonlinear processing of the output of the ordinary estimator, the desired estimate is yielded. Filtering, smoothing, and prediction can all be treated in this manner, and the generalization to estimation on an arbitrary Abelian Lie group is straightforward.

The reader is referred to Part II of this series of papers for a discussion of discrete-time S^1 estimation problems and to Part III for a discussion of some applications and some numerical results. In addition, [27] contains some optimal stochastic control schemes for S^1 processes, and [36] and [37] contain some likelihood formulas for processes on more general Lie groups.

Acknowledgment

The authors wish to thank Prof. R. W. Brockett of Harvard University, Cambridge, Mass., who provided the motivation and direction for this work, and Prof. J. C. Willems of the University of Groningen, Groningen, The Netherlands, and Prof. J. J. Deyst of M.I.T., Cambridge, Mass., for their criticism and advice during the preparation of this paper.

References

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME* (J. Basic Eng.), ser. D, vol. 82, pp. 35–45, 1960.
[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME* (J. Basic Eng.), ser. D, pp. 95–108, 1961.
[3] J. T.-H. Lo, "Finite dimensional sensor orbits and optimal nonlinear filtering," Ph.D. dissertation, Univ. Southern California, Los Angeles, Tech. Rep. USCAE 114, Aug. 1969; also, *IEEE Trans. Inform. Theory*, vol. IT-8, pp. 583–588, Sept. 1972.
[4] P. Levy, "L'addition des variables aléatoires définies sur une circonference," *Bull. Soc. Math. France*, vol. 67, pp. 1–41, 1939.
[5] F. Perrin, "Étude mathematique de mouvement Brownien de rotation," *Ann. Ecole Norm. Sup.*, vol. 45, pp. 1–51, 1928.
[6] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance.* New York: Wiley-Interscience, 1968.
[7] A. H. Jazwinski, *Stochastic Processes and Filtering Theory.* New York: Academic, 1970.
[8] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities.* Cambridge, Mass.: Cambridge Univ. Press, 1952.
[9] G. Szegö and G. Pólya, *Isoperimetric Inequalities in Mathematical Physics.* Princeton, N. J.: Princeton Univ. Press, 1951.
[10] S. Sherman, "Non-mean-square error criteria," *IRE Trans. Inform. Theory* (Corresp.), vol. IT-4, pp. 125–126, Sept. 1958.
[11] ——, "A theorem on convex sets with applications," *Ann. Math. Stat.*, vol. 26, pp. 763–767, 1955.
[12] R. Bellman, *A Brief Introduction to Theta Functions.* New York: Holt, Rinehart, and Winston, 1961.
[13] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis.* New York: MacMillan, 1945.
[14] J. L. Brown, "Asymmetric non-mean-square error criteria," *IRE Trans. Automat. Contr.*, vol. AC-7, pp. 64–66, Jan. 1962.
[15] T. T. Kadota and L. A. Shepp, "Conditions for absolute continuity between a certain pair of probability measures," *Z. Wahrscheinlichkeits theorie Verw.*, vol. 16, pp. 250–260, 1970.
[16] J. T.-H. Lo, "A general Bayes rule and its application to estimation," in *Proc. 7th Annu. Princeton Conf. Inform. Sci. and Syst.*, Mar. 1973; also *Inform. Sci.*, vol. 7, 1974.
[17] J. T.-H Lo, "On optimal nonlinear estimation—Part 1: Continuous observation," in *Proc. 8th Annu. Allerton Conf. Circuits and Syst.*, 1970; also, *Inform. Sci.*, vol. 6, 1973.
[18] P. A. Frost and T. Kailath, "An innovations approach to least-squares estimation—Part III: Nonlinear estimation in white Gaussian noise," *IEEE Trans. Automat. Contr.*, vol. AC-16, pp. 217–226, June 1971.
[19] R. L. Stratonovich, "Conditional Markov processes," *Theory of Prob. and Applications*, vol. 5, 1962.
[20] H. J. Kushner, "On the dynamic equations of conditional probability density functions, with applications to optimal stochastic control theory," *J. Math. Anal. Appl.*, vol. 8, 1964.

[21] K. Itô and H. P. McKean, Jr., *Diffusion Processes and Their Sample Paths*. New York: Academic, 1965.
[22] T. Kailath and P. Frost, "An innovations approach to least squares estimation—Part II: Linear smoothing in additive white noise," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 655–660, Dec. 1968.
[23] F. W. Warner, *Foundations of Differentiable Manifolds and Lie Groups*. Glenview, Ill.: Scott, Foresman and Co., 1971.
[24] E. V. Hoversten, R. O. Harger, and S. J. Holme, "Communication theory for the turbulent atmosphere," *Proc. IEEE*, vol. 58, pp. 1626–1650, Oct. 1970.
[25] J. T.-H. Lo and A. S. Willsky, "Estimation for rotational processes with one degree of freedom," Div. Eng. Appl. Phys., Harvard Univ., Cambridge, Mass., Tech. Rep. 635, July 1972.
[26] A. S. Willsky, "Dynamical systems defined on groups: Structural properties and estimation," Ph.D. dissertation, Dep. Aeronautics and Astronautics, Massachusetts Inst. Technol., Cambridge, June 1973.
[27] J. T.-H. Lo and A. S. Willsky, "Stochastic control of rotational processes with one degree of freedom," *SIAM J. Contr.*, to be published.
[28] D. Middleton, *An Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1960.
[29] A. J. Viterbi, *Principles of Coherent Communication*. New York: McGraw-Hill, 1966.
[30] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part II Nonlinear Modulation Theory*. New York: Wiley, 1971.
[31] D. L. Snyder, *The State-Variable Approach to Continuous Estimation with Applications to Analog Communication Theory*. Cambridge, Mass., M.I.T. Press, 1969.
[32] A. V. Oppenheim, R. W. Schafer, and T. G. Stockham, Jr., "Nonlinear filtering of multiplied and convolved signals," *Proc. IEEE*, vol. 56, pp. 1264–1291, Aug. 1968.
[33] Y. H. Ku and L. P. Sheporaitis, "Global properties of diffusion processes on cylindrical type phase space," *J. Franklin Inst.*, vol. 289, pp. 1087–1103, Feb. 1970.
[34] R. S. Bucy and A. J. Mallinckrodt, "An optimal phase demodulator," *Stochastics*, vol. 1, pp. 3–23, 1973.
[35] J. T.-H. Lo and A. S. Willsky, "Estimation for rotational processes with one degree of freedom—I: Introduction and continuous-time processes," Electron. Syst. Lab., Massachusetts Inst. Technol., Cambridge, Rep. ESL-P-506, July 1973.
[36] J. T.-H. Lo, "Signal detection of rotational processes and frequency demodulation," *Inform. Contr.*, vol. 26, pp. 99–115, 1974.
[37] ——, "Signal detection for bilinear processes," Univ. Maryland, Baltimore County, Math. Res. Rep. 73–7, Baltimore, Md., Nov. 1973.

James Ting-Ho Lo (M'73) was born in Kwangtung, China, on March 6, 1943. He received the B.S. degree in mechanical engineering from the National Taiwan University, Taipei, Taiwan, China, in 1964, and the Ph.D. degree in aerospace engineering from the University of Southern California, Los Angeles, in 1969.

From 1969 to 1971 he was a Research Associate at Stanford University, Stanford, Calif. From 1971 to 1972 he was a Research Fellow in Applied Mathematics at Harvard University, Cambridge, Mass. Since September 1972 he has been an Assistant Professor at the University of Maryland, Baltimore County, Baltimore, Md. His current research interests include detection, estimation, and control on Lie groups and algebraic system theory.

Alan S. Willsky (S'70–M'73) was born in Newark, N. J., on March 16, 1948. He received the S.B. degree in aeronautics and astronautics and the Ph.D. degree in instrumentation and control from the Massachusetts Institute of Technology, Cambridge, in 1969 and 1973, respectively.

From 1969 through 1973 he held a Fannie and John Hertz Foundation Doctoral Fellowship. He is presently an Assistant Professor of Electrical Engineering at M.I.T. and Assistant Director of the M.I.T. Electronic Systems Laboratory. His interests are in estimation techniques for nonlinear systems, especially with the use of harmonic analysis and Lie group theory, algebraic system theory and its impact on the design of dynamical systems for estimation, control, and computation, and the design of failure detection and self-reorganization systems for estimation and control.

Dr. Willsky is a member of Sigma Xi, Sigma Gamma Tau, and Tau Beta Pi.