[12] M. C. Yovits and J. L. Jackson, "Linear filter optimization with game theory considerations," in *IRE National Convention Record*, Pt. 4, pp. 193–199, 1955.
[13] M. G. Krein and M. A. Rutman, "Linear operators leaving invariant a cone in a branch space," American Math. Soc. Translation No. 26, 1950.
[14] E. Hille, *Functional Analysis and Semi-Groups*. New York: American Math., 1948.
[15] R. Z. Khas'minskii, "Ergodic properties of recurrent diffusion processes and stabilization of the solution to the Cauchy problem for parabolic equations," *Theory Prob. Appl.* (USSR), vol. 5, pp. 179–196, 1960.
[16] M. Zakai, "A Lyapunov criterion for the existence of stationary probability density functions for systems perturbed by noise," *SIAM J. Contr.*, vol. 7, pp. 390–397, 1969.
[17] S. Karlin and J. McGregor, "Classical diffusion processes and total positivity," *J. Math. Anal. Appl.*, vol. 1, pp. 163–183, 1960.
[18] H. J. Kushner, "The Cauchy problem for a class of degenerate parabolic equations and asymptotic properties of the related diffusion processes," *J. Differential Equations*, vol. 6, pp. 209–231, 1969.
[19] W. M. Wonham, "Some applications of stochastic differential equations to optimal nonlinear filtering," *SIAM J. Contr.*, vol. 12, pp. 347–369, 1965.
[20] G. Lee and J. Komo, "An application of nonlinear filtering theory to PCM bit synchronization," in *National Telemetering Conf. Record*, Washington, D.C., pp. 43–48, Apr. 1969.
[21] D. L. Snyder, "Filtering and detection for doubly stochastic Poisson processes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 91–101, Jan. 1972.

Fourier Series and Estimation on the Circle with Applications to Synchronous Communication—Part I: Analysis

ALAN S. WILLSKY, MEMBER, IEEE

***Abstract*—A wide variety of continuous- and discrete-time estimation problems on the circle S^1 are considered with the aid of Fourier series analysis. Measurement and diffusion update equations are derived for the conditional expectation of certain functions of the parameter to be estimated, and we investigate the use of Fourier series to obtain easily implemented optimal estimation equations. A variety of important examples—phase tracking, frequency demodulation, and phase demodulation in the presence of oscillator instabilities, additive noise, Rayleigh fading, or any combination of these—are considered.**

I. Introduction

As HAS BEEN discussed in the recent literature [1]–[4], [6]–[7], and [9], the circle S^1 provides a great deal of structure for the study of certain stochastic processes. This is reflected in the several new techniques [1]–[9] that have been devised to study and solve a variety of phase tracking and angle demodulation problems.

One of the most important analytical tools related to the circle is Fourier series analysis. Bucy and his associates [6], [7] have used Fourier analysis to derive infinite-dimensional optimal estimation equations for several specific problems, and Willsky and Lo [2], [4] have obtained a general infinite-dimensional discrete-time result. Finite-dimensional approximations were briefly discussed in [2], and an unsuccessful attempt was reported in [6].

Manuscript received July 26, 1973; revised April 3, 1974. This work was done in part while the author was a Fannie and John Hertz Foundation Fellow in the Department of Aeronautics and Astronautics, M.I.T., and in part at the Decision and Control Sciences Group of the M.I.T. Electronic Systems Laboratory with partial support provided by AFOSR under Grant 72-2273.
The author is with the Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

In this set of two papers we extend the Fourier-type results that have been obtained to large classes of discrete- and continuous-time problems of practical importance. In the present paper we will discuss the analytical results that have been obtained. We review the discrete-time result of Willsky and Lo and apply it to several important examples. In addition, we develop (via a series of examples) an approach to solving a number of continuous-time problems, including the problem discussed in [7]. In Part II we discuss the problem of implementation, i.e., of finite-dimensional approximation, and present some numerical results comparing a Fourier-type system to a standard phase-lock loop [10] and a "state-dependent noise" filter [8]. One of the major results of our work is the success we have had in designing high performance approximations to the analytical Fourier series results. We note that the results presented here and in Part II have been reported in part in [4]. The reader is also referred to some analogous results in [11] for spatially discretized versions of some of these problems.

II. Optimal Estimation of S^1 Random Variables Using Fourier Series

In this section we display estimation equations originally derived in [2] and [4]. Let θ be a random variable on S^1 (identified with $[-\pi,\pi)$) with probability density

$$p(\theta) = \sum_{n=-\infty}^{+\infty} c_n e^{in\theta}. \tag{1}$$

Here

$$c_n = \frac{1}{2\pi}\mathscr{E}(e^{-in\theta}) = b_n - ia_n = c_{-n}^* \tag{2}$$

where $\mathscr{E}$ denotes the expectation operator, * the complex conjugate, and

$$a_n = \frac{1}{2\pi}\,\mathscr{E}(\sin n\theta) \qquad b_n = \frac{1}{2\pi}\,\mathscr{E}(\cos n\theta). \tag{3}$$

As we shall see, the Fourier coefficients $\{c_n\}$ provide a useful set of "moments" for θ.

Suppose we wish to minimize $\mathscr{E}[\phi(\theta - \tilde{\theta})]$ with respect to $\tilde{\theta}$, where $\phi\colon [-\pi,\pi) \to R$ is given by

$$\phi(\theta) = \sum_{n=-\infty}^{+\infty} d_n e^{in\theta}. \tag{4}$$

A simple computation yields

$$J(\tilde{\theta}) \triangleq \mathscr{E}[\phi(\theta - \tilde{\theta})] = \sum_{n=-\infty}^{+\infty} c_n d_n^* e^{in\tilde{\theta}} \tag{5}$$

and necessary conditions for a local minimum are

$$\frac{d}{d\tilde{\theta}} J(\tilde{\theta}) = 0 \qquad \frac{d^2}{d\tilde{\theta}^2} J(\tilde{\theta}) \geq 0 \tag{6}$$

explicit solutions of which is possible only for certain error functions. As an example, consider the error function (also used in [6], [7])

$$\phi(\theta) = 1 - \cos\theta \tag{7}$$

for which we have

$$J(\tilde{\theta}) = 1 - 2\pi(a_1 \sin\tilde{\theta} + b_1 \cos\tilde{\theta}). \tag{8}$$

The optimal estimate and cost are given by

$$\tilde{\theta}_0 = \tan^{-1}\left(\frac{a_1}{b_1}\right) \tag{9}$$

$$J(\tilde{\theta}_0) = 1 - 2\pi\sqrt{a_1^2 + b_1^2} \tag{10}$$

where (9) is to be interpreted as

$$\sin\tilde{\theta}_0 = \frac{a_1}{(a_1^2 + b_1^2)^{1/2}} \qquad \cos\tilde{\theta}_0 = \frac{b_1}{(a_1^2 + b_1^2)^{1/2}}. \tag{11}$$

Since $1 - \cos\theta \approx \theta^2/2$ for small θ, we see that, at least locally, this is a type of least squares criterion. A detailed discussion of this criterion can be found in [2].

We remark that the computations involved in solving (6) become increasingly more difficult as the number of nonzero Fourier coefficients of ϕ increases. The reader is referred to [2], [4], and [7] for more on this subject.

III. General Discrete-Time S^1 Estimation Problems

In this section we will first consider a general single stage S^1 estimation problem (see [2], [4]). Several important examples, extensions to multistage problems, and computational considerations will be discussed later in the section.

Let θ be a random variable on the circle with *a priori* density

$$p_\theta(\xi) = \sum_{n=-\infty}^{\infty} c_n(0) e^{in\xi}. \tag{12}$$

Suppose that we take a single (possibly nonlinear) measurement y of θ and that the noise density $p_{y|\theta}(v \mid \xi)$ exists. Considering this as a function of ξ for fixed v, we have

$$p_{y|\theta}(v \mid \xi) = \sum_{n=-\infty}^{+\infty} d_n(v) e^{in\xi} \tag{13}$$

where the d_n are functions of v (which do *not*, in general, have an interpretation as in (2)). Applying Bayes' rule, we compute the Fourier series form for the conditional density

$$p_{\theta|y}(\xi \mid v) = \sum_{n=-\infty}^{+\infty} c_n(1) e^{in\xi} \tag{14}$$

$$c_n(1) = \frac{\gamma_n(v)}{2\pi\gamma_0(v)} \tag{15}$$

$$\gamma_n(v) = \sum_{k=-\infty}^{+\infty} c_k(0) d_{n-k}(v). \tag{16}$$

Thus the computation of $p_{\theta|y}$ involves the (in general nonlinear) computation of the coefficients $\{d_n(v)\}$ and the evaluation of the convolution (16). We will comment at the end of this section on the computational savings that can be obtained by utilizing the structure of (16). Note that the only restriction on the applicability of these results is that we must be able to write $p_{y|\theta}$ as in (13).

We note that if we use an estimation criterion of the form given in (4), our optimal conditional estimate becomes an explicit function of the $\{c_n(1)\}$ (e.g., see (7)–(9)). It is this observation that provides much of the motivation for our study of the evolution of the Fourier coefficients of probability densities on S^1. The following examples are but a few of the many problems that fall into this framework.

Example 1

Let θ be an S^1 random variable with density given by (12). Let v be a real-valued random variable, independent of θ, with density $p_v(v) = N(v; 0,\gamma)$ (normal density with 0 mean and variance γ). Consider the observation

$$y = \sin\theta + v. \tag{17}$$

In this case

$$p_{y|\theta}(v \mid \xi) = N(v - \sin\xi; 0,\gamma) \tag{18}$$

and we can show [4], [12], [13] that the associated Fourier series (13) is a complicated expression involving Bessel functions. In principle we can use this expression together with the observation value v and the update equations (14)–(16) to compute the conditional density (approximate methods are discussed in Part II). We also note that vast computational simplifications arise if we consider the continuous time analog of (17) (see Example 4).

Example 2

Let θ and v be as in the previous example and suppose our observation is

$$y = (\theta + v) \bmod 2\pi. \tag{19}$$

This type of problem was studied in [2] and [4] in a different way, using the mod 2π equivalence of points on the real line. We have

$$p_{y|\theta}(v \mid \xi) = F(v - \xi; 0,\gamma) \tag{20}$$

where $F(\alpha; \eta,\gamma)$ is the "folded normal" density (evaluated at α) with mode η and "variance" γ. This density is related to the normal density in the following way: if x is a real random variable with density $N(\alpha; \eta,\gamma)$, then $\theta = x \bmod 2\pi$ has the density

$$F(\alpha; \eta,\gamma) = \sum_{n=-\infty}^{+\infty} N(\alpha + 2n\pi; \eta,\gamma) \tag{21}$$

[i.e., we "fold" $N(\alpha; \eta,\gamma)$ around the circle]. We note that the folded normal density is the solution of the standard diffusion equation on the circle (i.e., it is the density for S^1 Brownian motion processes) and is as important a density on S^1 as the normal is on R^1. We will find this density to be most useful in devising suboptimal schemes in Part II. For further discussions on the folded normal, see [1], [4], [14]–[16].

One can show [1], [4] that the Fourier series form for $p_{y|\theta}$ in (20) is

$$p_{y|\theta}(\nu \mid \xi) = \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} e^{-n^2\gamma/2} e^{in(\nu-\xi)}. \tag{22}$$

We can then compute the Fourier series form of $p_{\theta|y}$ from (14)–(16).

We can also handle the problem of computing conditional densities of random processes given a series of discrete measurements. Assuming that the measurement noises of the various measurements are independent of each other and of the process θ that is to be estimated, we can process each measurement as in (14)–(16) and can propagate the density between measurements via some sort of "diffusion" equation. Of course, if we are using the Fourier series representation for probability densities, our diffusion update equations should be in terms of the Fourier coefficients. We consider an example of a random process on S^1 to show how the Fourier series approach leads to simple diffusion update equations.

Example 3

Consider a discrete-time random process θ_k that satisfies

$$\theta_{k+1} = (\theta_k + w_k) \bmod 2\pi \tag{23}$$

where the w_k are independent random variables on S^1. Given a sequence of (possibly nonlinear) noisy observations y_k of θ_k, we wish to compute the conditional distributions

$$p_\theta(\xi; k,k) = p(\theta_k = \xi \mid y_1,\cdots,y_k) = \sum_{n=-\infty}^{+\infty} c_n(k \mid k) e^{in\xi} \tag{24}$$

$$\begin{aligned} p_\theta(\xi; k+1, k) &= p(\theta_{k+1} = \xi \mid y_1,\cdots,y_k) \\ &= \sum_{n=-\infty}^{+\infty} c_n(k+1 \mid k) e^{in\xi}. \end{aligned} \tag{25}$$

The computation of $p_\theta(\xi; k,k)$ from $p_\theta(\xi; k, k-1)$ and the measurement y_k proceeds as discussed previously. We now consider the "diffusion update"—the computation of $p_\theta(\xi; k+1, k)$ from $p_\theta(\xi; k,k)$. Assuming that the density for w_k can be written as

$$p_w(\alpha,k) = \sum_{n=-\infty}^{+\infty} d_n(k) e^{in\alpha} \tag{26}$$

we have the convolution form of the diffusion update

$$p_\theta(\xi; k+1, k) = \int_0^{2\pi} p_\theta(\xi - \alpha; k,k) p_w(\alpha,k)\, d\alpha \tag{27}$$

which leads to the Fourier series update equation

$$c_n(k+1 \mid k) = 2\pi c_n(k \mid k) d_n(k). \tag{28}$$

The fact that the convolution (27) transforms into the pointwise multiplication (28), together with the interpretation of c_n in (2), indicates the close relationship between the Fourier coefficients and the characteristic function of a real random variable [17]. We also note a type of "dual" result to the diffusion update pair (27), (28). Looking at the measurement update computation, we see that, except for the normalization factor, Bayes' rule involves pointwise multiplication of functions ($p_\theta(\xi)$ times $p_{y|\theta}(\nu \mid \xi)$; regard the latter as a function of ξ), but when we transform to the Fourier coefficient framework, we obtain a convolution (16). This observation is important in understanding the computational aspects of our results. Since we cannot avoid a convolution (we obtain one in either the diffusion or measurement update depending on whether we use the density or its Fourier coefficients), the high-speed convolution and fast Fourier transform techniques discussed in [26] and [19] may be of value in performing the calculations efficiently. (See [4] for a further discussion and [11] for a related discussion on quantized versions of these problems.)

Note that (28) can be used to study the properties of stable distributions on S^1, i.e., classes of densities that are closed under convolution, or equivalently, classes of densities that are closed under the multiplication of Fourier coefficients, as in (28). We will not go into this problem here except to remark that one can show that the class of folded normal densities is a stable class. (See Lemma 2 of [1] and the discussion in [4].)

Finally, we note that if $\theta(t) = w(t) \bmod 2\pi$, where $w(t)$ is a standard Brownian motion process, we obtain an equation of the form (23) if we take $\theta_k = \theta(k\Delta)$, $w_k = w((k+1)\Delta) - w(k\Delta)$. In this case, we can show [1], [4] that $p_\theta(\xi; k+1, k)$ is the solution of the diffusion equation

$$\frac{\partial p}{\partial t} - \frac{1}{2}\frac{\partial^2 p}{\partial \theta^2} = 0 \tag{29}$$

at $t = (k+1)\Delta$ when we start at time t with $p_\theta(\xi; k,k)$. Also

$$p_w(\alpha,k) = \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} e^{-n^2\Delta/2} e^{in\alpha}. \tag{30}$$

IV. Fourier Analysis and Continuous-Time Phase Tracking and Demodulation

In this section we consider a class of estimation problems of importance in a number of communication applications. We wish to investigate the processing of a signal of the form

$$r(t) = \sin(\omega_c t + \phi(t) + v(t)) + \dot{w}(t) \tag{31}$$

where ω_c is a carrier frequency, ϕ is some type of modulating information, v is a phase drift, and $\dot{w}$ is additive channel noise. With this model we can consider phase tracking, phase demodulation, and frequency demodulation. Standard techniques for such problems involve a system called a phase-lock loop (PLL) [10]. The reader is referred to Part II, where we review the basic ideas behind the PLL and present some results comparing the performance of PLL systems with the performance of systems designed using the techniques developed here.

In the following discussion we utilize the tool of Fourier series analysis to design tracking and demodulation systems. The goal of our approach and also of the similar techniques discussed in [6] and [7] is to devise methods for designing filters that utilize the inherent structure of the problem at hand and that are of practical value. In this light, one of the main contributions of our work is that the Fourier series formulation allows us to understand the structure of optimal (infinite-dimensional) trackers and receivers and to devise high performance finite-dimensional approximations. Instead of describing a general method, we illustrate our technique by discussing several important examples.

Example 4

We consider a phase tracking problem that is essentially the same as that studied in [7]. This problem is the continuous time analog of Example 1. Suppose we receive the signal

$$\dot{z}(t) = \sin \theta(t) + r^{1/2}(t)\dot{w}(t) \tag{32}$$

where

$$\theta(t) = \omega_c t + \int_0^t q^{1/2}(s)\, dv(s) + \theta_0 \tag{33}$$

and v and w are independent Brownian motions, $q(t) \geq 0$, $r(t) > 0$, and $\omega_c > 0$. Also θ_0 is a random initial condition independent of v and w.

Suppose we wish to estimate $\theta(t)$ mod 2π given $\{\dot{z}(s) \mid 0 \leq s \leq t\}$; i.e., we wish to filter out $r^{1/2}\dot{w}$ and track the phase. Equation (32) is, of course, only formal since $\dot{w}$ is white noise. The Itô differential forms of (32) and (33) are

$$d\theta(t) = \omega_c\, dt + q^{1/2}(t)\, dv(t), \qquad \theta(0) = \theta_0 \tag{34}$$

$$dz(t) = \sin \theta(t)\, dt + r^{1/2}(t)\, dw(t). \tag{35}$$

We take as our optimal estimation criterion the minimization of $\mathscr{E}[(1 - \cos(\theta(t) - \hat{\theta}(t)) \mid z(s), 0 \leq s \leq t]$. As discussed previously, it then makes sense to compute the Fourier coefficients $\{c_n\}$ of the distribution for $\theta(t)$ conditioned on $z(s)$, $0 \leq s \leq t$.

The stochastic differential equation for $c_n(t)$ can be obtained using the results of Kushner [21], [22]

$$\begin{aligned} dc_n(t) = &-\left[in\omega_c + \frac{n^2}{2} q(t)\right] c_n(t)\, dt \\ &+ \left[\frac{(c_{n-1}(t) - c_{n+1}(t))}{2i} + 2\pi c_n(t) \operatorname{Im}(c_1(t))\right] \\ &\cdot \left[\frac{dz(t) + 2\pi \operatorname{Im}(c_1(t))\, dt}{r(t)}\right]. \end{aligned} \tag{36}$$

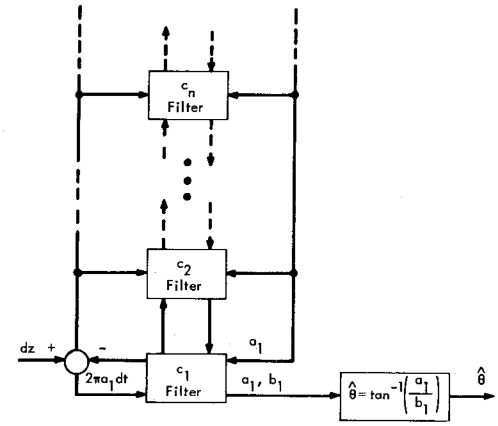

Fig. 1. Form of infinite-dimensional optimal filter of Example 4.

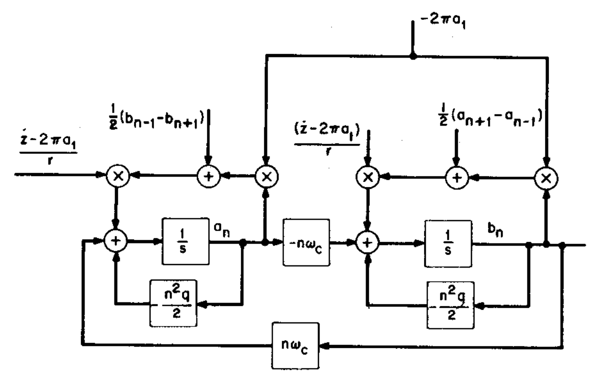

Fig. 2. Diagram of the c_n filter shown in Fig. 1.

Here Im $(c_n) = (c_n - c_n^*)/2i$. Recalling that $c_0 = \frac{1}{2}\pi$ and $c_{-n} = c_n^*$, we see that we need only solve (36) for $n \geq 1$. Note that these equations are simpler than the discrete-time equations of Example 1, since the equation for c_n depends only on c_1, c_{n-1}, c_n, and c_{n+1} instead of on all of the coefficients as in (16) (also we do not have to compute Bessel functions).

The structure of the optimal filter, which is illustrated in Figs. 1 and 2, deserves further comment (recall that $c_n = b_n - ia_n$). The filter consists of an infinite bank of filters, the nth of which is essentially a damped oscillator, with oscillator frequency $n\omega_c$, together with nonlinear couplings to the other filters and to the received signal. Note that without measurements but with the oscillator phase noise $q^{1/2}(t)\, dv(t)$ present, the steady state density for θ is uniform; i.e., $c_n = 0$, for all $n \neq 0$. The purpose of the damping term in the c_n filter is to account for this diffusive effect. We will have more to say about the structure of this filter in Part II.

We note one very appealing feature of our filter; i.e., it is time-invariant (see Figs. 1 and 2) if q and r are constant and is equipped to handle any initial conditions. If we are interested in the pure synchronization problem after we have acquired the signal, we might impose the initial condition $c_n(0) = \frac{1}{2}\pi$, for all n, which corresponds to our knowing that the initial phase is 0. On the other hand, setting $c_n(0) = 0$, for all $n \neq 0$, corresponds to the assumption that the initial phase is uniformly distributed; i.e., this is the acquisition problem. Thus, suppose we can build a finite-dimensional approximation to the infinite dimensional filter given by (36). Then, assuming we retain the time-invariant nature of the original system, we see that the same filter can handle either synchronization or acquisition merely by the proper choice of initial conditions. The reader is referred to Part II, where we present and discuss some simulation results for the problem discussed in Example 4.

Example 5

In this example we will consider a phase demodulation problem. Consider the R^1 signal process $x(t)$ satisfying the Itô differential equation

$$dx(t) = a(t)x(t)\,dt + q^{1/2}(t)\,dv(t) \tag{37}$$

where v is a standard Brownian motion process independent of the initial condition $x(0)$. We wish to estimate $x(t)$ given the phase-modulated observation process

$$dz(t) = \sin(\omega_c t + x(t))\,dt + r^{1/2}(t)\,dw(t) \tag{38}$$

where w is a Brownian motion independent of v and $x(0)$.

Suppose we want the minimum variance estimate of $x(t)$, given $\{z(s) \mid 0 \le s \le t\}$. It is known that the desired estimate is the conditional expectation $\mathscr{E}[x(t) \mid z(s), 0 \le s \le t] \triangleq \hat{x}(t \mid t)$. As in Example 4, we write the optimal demodulator equations in terms of some very special functions. The form of the optimal demodulator, which is infinite-dimensional, is appealing physically and suggests some new suboptimal demodulation techniques. The exponential time correlation of $x(t)$ (the presence of the term $a(t)x(t)\,dt$ on the right side of (37)) makes this problem somewhat more complicated than the preceding example.

We now write the stochastic differential equations satisfied by

$$c_{nm}(t) = \frac{1}{2\pi}\mathscr{E}[x^n(t)e^{-im\theta(t)} \mid z(s), 0 \le s \le t] \tag{39}$$

where $\theta(t)$ is the signal phase

$$\theta(t) = (\omega_c t + x(t)) \bmod 2\pi. \tag{40}$$

Again using Kushner's results, we find that

$$\begin{aligned} dc_{nm}(t) = \Bigg\{ &\left[na(t) - \frac{m^2 q(t)}{2} - im\omega_c\right] c_{nm}(t) \\ &- ima(t)c_{n+1,m}(t) + \frac{n(n-1)q(t)}{2} \\ &\cdot c_{n-2,m}(t) - iq(t)nmc_{n-1,m}(t)\Bigg\}\,dt \\ &+ \left[\frac{c_{n,m-1}(t) - c_{n,m+1}(t)}{2i} + 2\pi c_{nm}(t)\,\mathrm{Im}\,(c_{01}(t))\right] \\ &\cdot \frac{(dz(t) + 2\pi\,\mathrm{Im}\,(c_{01}(t)\,dt))}{r(t)} \end{aligned} \tag{41}$$

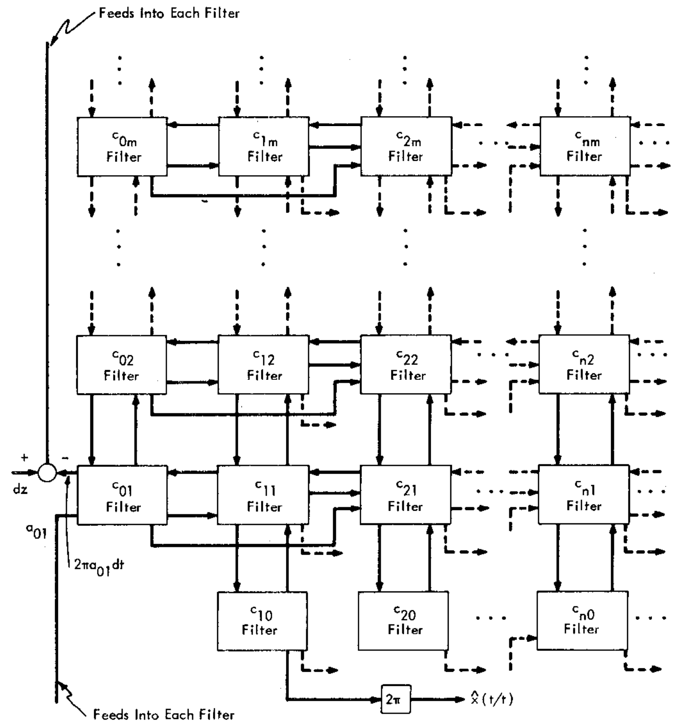

Fig. 3. Form of optimal phase demodulator of Example 5.

where $c_{nk}(t) \triangleq 0$, for all $n < 0$, and $c_{nm}(t) = c^*_{n,-m}(t)$. Note that our optimal estimate is

$$\hat{x}(t \mid t) = 2\pi c_{10}(t). \tag{42}$$

We will make a few comments about this filter, which is illustrated in Fig. 3. The demodulator consists of a doubly infinite bank of filters. The c_{nm} filter is directly connected only to the c_{01}, $c_{n+1,m}$, $c_{n-1,m}$, $c_{n-2,m}$, $c_{n,m-1}$, and $c_{n,m+1}$ filters. Referring to (41), we see that the c_{nm} filter for $m > 0$ resembles the damped oscillator filter in Fig. 2. The c_{n0} filters are much simpler in form.

The reader is referred to Part II, in which we discuss a number of systematic methods for truncating this doubly infinite bank of filters. We also note that in this same framework, one can consider phase modulating more complicated signals and the filtering out of phase drift noise or a random initial phase (i.e., the acquisition problem). The next example, in which we consider an FM problem, illustrates the versatility of this conceptual approach.

Example 6

Consider the received signal process

$$dz(t) = \sin\left(\omega_c t + g\int_0^t x(s)\,ds + \int_0^t e^{1/2}(s)\,df(s)\right)dt + r^{1/2}(t)\,dw(t) \tag{43}$$

where x satisfies (37), and f and w are independent standard one-dimensional Brownian motion processes, both independent of x. The term $\int_0^t e^{1/2}(s)\,df(s)$ represents random phase drift, and the $dw(t)$ term is additive channel noise. We note that $x(t)$ can be considered to be an error in our knowledge of ω_c (e.g., $dx \equiv 0$ with $x(0)$ unknown corresponds to a constant offset in the carrier frequency or a constant Doppler shift [23]). Let us define the two-dimensional signal process

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} x(t) \\ g \int_0^t x(s)\, ds + \int_0^t e^{1/2}(s)\, df(s) \end{bmatrix}. \quad (44)$$

We then have

$$dy(t) = A(t)y(t)\, dt + Q^{1/2}(t)\, du(t) \quad (45)$$

where

$$u(t) = \begin{bmatrix} v(t) \\ f(t) \end{bmatrix} \qquad A(t) = \begin{bmatrix} a(t) & 0 \\ g & 0 \end{bmatrix}$$

$$Q^{1/2}(t) = \begin{bmatrix} q^{1/2}(t) & 0 \\ 0 & e^{1/2}(t) \end{bmatrix}. \quad (46)$$

Also

$$dz(t) = \sin(\omega_c t + h'y(t))\, dt + r^{1/2}(t)\, dw(t) \quad (47)$$

$$h' = [0, 1]. \quad (48)$$

As in the last example, suppose we want the minimum variance estimate of $x(t)$. To do this, we write the stochastic differential equations for

$$c_{nm} = \frac{1}{2\pi} \mathscr{E}[y_1^n(t) e^{-im(\omega_c t + y_2(t))} \mid z(s), 0 \le s \le t]. \quad (49)$$

The optimal estimate is $2\pi c_{10}(t)$, and the filter equations are

$$\begin{aligned} dc_{nm}(t) = {} & \left\{ \left[na(t) - \frac{m^2}{2} e(t) - im\omega_c \right] c_{nm}(t) \right. \\ & \left. - imgc_{n+1,m}(t) + \frac{n(n-1)q(t)}{2} c_{n-2,m}(t) \right\} dt \\ & + \left[\frac{(c_{n,m-1}(t) - c_{n,m+1}(t))}{2i} + 2\pi c_{nm}(t) \operatorname{Im}(c_{01}(t)) \right] \\ & \cdot \left[\frac{dz(t) + 2\pi \operatorname{Im}(c_{01}(t))\, dt}{r(t)} \right]. \end{aligned} \quad (50)$$

Again $c_{nk} \triangleq 0$, for all $n < 0$.

In a similar manner we can consider the general problem

$$dy(t) = A(t)y(t)\, dt + Q^{1/2}(t)\, dv(t) \quad (51)$$

$$dz(t) = \sin(\omega_c t + h'y(t)) + r^{1/2}\, dw(t) \quad (52)$$

where y is an n-vector, v is an m-vector, and the functions $v_1(t), \cdots, v_m(t)$, $w(t)$ are independent standard Brownian motion processes, all independent of $y(0)$. Given (52) we can consider finding the minimum variance estimate (i.e., the conditional mean) of $Cy(t)$ given $\{z(s) \mid 0 \le s \le t\}$. (Note that if we allow $y(0)$ to be random, we can consider a random initial phase problem.) Examples 4, 5, and 6 indicate that this model includes the phase tracking, phase demodulation, and frequency demodulation problems in the presence of both additive channel noise and phase drift noise. (In the phase tracking problem, our estimation criterion may be something other than the minimum variance estimate, e.g., $1 - \cos(\theta - \hat{\theta})$.)

We also note that the general problem (51), (52) is conceptually no more difficult than the problems considered in the examples and, as in these cases, yields a time-invariant solution if A, Q, and r are constant. However, the bookkeeping becomes more complicated. Some straightforward calculations indicate that, in general, we must compute

$$\mathscr{E}[y_1^{k_1} y_2^{k_2} \cdots y_n^{k_n} e^{-im(\omega_c t + h'y(t))} \mid z(s), 0 \le s \le t].$$

Thus the general demodulator–tracker is a multidimensional version of the systems considered in the examples. As a final indication of how these techniques can be used, we consider a phase tracking problem in which the amplitude of the sinusoidal signal is unknown.

Example 7

Suppose we receive the signal

$$dz(t) = A \sin \theta(t)\, dt + r^{1/2}(t)\, dw(t) \quad (53)$$

where

$$\theta(t) = \omega_c t + \int_0^t q^{1/2}(s)\, dv(s) + \theta_0. \quad (54)$$

We assume that the signal amplitude is constant but unknown with *a priori* probability distribution $p_A(\alpha)$. Also, we let v and w be independent Brownian motions, which are both independent of A. We wish to devise a technique for tracking $\theta(t)$. We use the criterion $\min \mathscr{E}[1 - \cos(\theta(t) - \hat{\theta}(t)) \mid z(s), s \le t]$. Adapting the techniques used in the preceding examples, we write the differential equations for

$$c_{nm}(t) = \frac{1}{2\pi} \mathscr{E}[A^n e^{-im\theta(t)} \mid z(s), 0 \le s \le t]. \quad (55)$$

We can show that

$$\begin{aligned} dc_{nm}(t) = {} & - \left[im\omega_c + \frac{m^2 q(t)}{2} \right] c_{nm}(t)\, dt \\ & + \left[\frac{(c_{n+1,m-1}(t) - c_{n+1,m+1}(t))}{2i} \right. \\ & \left. + 2\pi c_{nm}(t) \operatorname{Im}(c_{11}(t)) \right] \left[\frac{dz(t) + 2\pi \operatorname{Im}(c_{11}(t))\, dt}{r(t)} \right] \end{aligned} \quad (56)$$

and our estimate is

$$\hat{\theta}(t) = \tan^{-1} \left(\frac{a_{01}(t)}{b_{01}(t)} \right). \quad (57)$$

We note that in acquiring a sinusoidal signal such as in (53), we often do not know the transmitted signal power; i.e., A can be taken as a random variable. Thus we might use a filter of the type described here to determine A (equivalently, to determine the signal to noise ratio). Once we have "determined" A, i.e., reduced the variance in our estimate ($2\pi c_{10}$) of A to an acceptable level, e.g.,

$$\sqrt{2\pi c_{20} - 4\pi^2 c_{10}^2} < 0.02\pi c_{10},$$

we can simplify the tracking filter. That is, we use the approximation $A \approx 2\pi c_{10}$, and the problem reduces to the type considered in Example 4. In relation to (56), we then need only compute c_{0m}. We note that the only c_{nm} with $n \geq 1$ that directly enters the equations for dc_{0m} is c_{1m}, and we can approximate it by

$$c_{1m} \approx 2\pi c_{10} c_{0m}. \tag{58}$$

We can also consider the problem in which A is time-varying. One important problem is the tracking of a sinusoidal signal that is transmitted through a Rayleigh channel [10], [24]. The received signal for such a channel can be modeled [10], [25] as

$$dz(t) = x_1(t) \sin \theta(t)\, dt + x_2(t) \cos \theta(t)\, dt + r^{1/2}(t)\, dw(t) \tag{59}$$

where x_1 and x_2 are zero mean, independent, identically distributed Gaussian random processes. If we know that x_1 and x_2 satisfy a particular linear Itô differential equation, we can use the techniques developed in this section to track $\theta(t)$. In this case, the quantities to be estimated are $(1/\pi)x_1^{\,l}x_2^{\,m}e^{-in\theta}$.

V. Conclusion

In this paper we have investigated a number of phase estimation and demodulation problems with the aid of Fourier series analysis. We have found that this approach exposes the rich structure inherent in quite general classes of estimation problems on the circle. We have considered a number of important practical problems as examples and have derived infinite-dimensional optimal estimation equations. In Part II we will consider approximating these sets of equations and will present some numerical results that indicate the worth of our Fourier series approach.

Acknowledgment

The author wishes to thank Roger W. Brockett of Harvard University, Cambridge, Mass., for his advice during the development of these results and wishes to express his appreciation to the reviewers of this Transactions and to Associate Editor T. Kailath of Stanford University, Stanford, Calif., for their helpful suggestions.

References

[1] J. T. Lo and A. S. Willsky, "Estimation for rotational processes with one degree of freedom—I: Introduction and continuous time processes," to appear.

[2] A. S. Willsky and J. T. Lo, "Estimation for rotational processes with one degree of freedom—II: Discrete time processes," to appear.

[3] ——, "Estimation for rotational processes with one degree of freedom—III: Applications and implementation," to appear.

[4] A. S. Willsky, "Dynamical systems defined on groups: Structural properties and estimation," Ph.D. dissertation, Dep. Aeronautics and Astronautics, M.I.T., Cambridge, Mass., June 1973.

[5] J. T. Lo and A. S. Willsky, "Stochastic control of rotational processes with one degree of freedom," to appear.

[6] 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.

[7] A. 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, Greenbelt, Md., NASA Contract NAS5-10789, 1970.

[8] D. E. Gustafson and J. L. Speyer, "Linear and asymptotic minimum variance filters applied to phase-lock loops," to appear.

[9] R. W. Brockett, "Lie theory and control systems on spheres," *SIAM J. Appl. Math.*, vol. 25, pp. 213–225, Sept. 1973.

[10] H. L. Van Trees, *Detection, Estimation, and Modulation Theory—Part II: Nonlinear Modulation Theory.* New York: Wiley, 1971.

[11] A. S. Willsky, "Estimation on finite cyclic groups with applications to coding," to appear.

[12] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions* (Applied Mathematics Series 55). Washington, D.C.: NBS, 1964.

[13] F. B. Hildebrand, *Advanced Calculus for Applications.* Englewood Cliffs, N.J.: Prentice-Hall, 1962.

[14] F. Perrin, "Étude mathématique du mouvement Brownien de rotation," *Ann. Ecole Norm. Sup.*, vol. 45, pp. 1–51, 1928.

[15] P. Lévy, "L'addition des variables aléatoires définies sur une circonférence," *Bull. Soc. Math. Fr.*, vol. 67, pp. 1–41, 1939.

[16] R. Bellman, *A Brief Introduction to Theta Functions.* New York: Holt, Rinehart, and Winston, 1961.

[17] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 1965.

[18] B. W. Licht, "Approximations in optimal nonlinear filtering," Systems Research Center, Case Western Reserve Univ., Cleveland, Ohio, Rep. no. SRC 70-1, Apr. 1970.

[19] P. J. Nicholson, "Algebraic theory of finite Fourier transforms," *J. Comput. Syst. Sci.*, vol. 5, pp. 524–547, Oct. 1971.

[20] J. A. Pierce, W. Palmer, A. D. Watt, and R. H. Woodward, "Omega—A world-wide navigational system specification and implementation—Second revision," Pickard and Burns Electronics, Waltham, Mass., May 1966.

[21] 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, pp. 332–344, 1964.

[22] ——, "Dynamical equations for optimal nonlinear filtering," *J. Differential Equations*, vol. 3, pp. 179–190, 1967.

[23] M. Kayton and W. R. Fried, Eds., *Avionics Navigation Systems.* New York: Wiley, 1969.

[24] H. L. Van Trees, *Detection, Estimation, and Modulation Theory—Part I: Detection, Estimation, and Linear Modulation Theory.* New York: Wiley, 1968.

[25] W. B. Davenport and W. L. Root, *An Introduction to the Theory of Random Signals and Noise.* New York: McGraw-Hill, 1958.

[26] T. G. Stockham, Jr., "High-speed convolution and correlation," in *1966 Spring Joint Computer Conf., AFIPS Conf. Proc.*, vol. 28. Washington, D.C.: Spartan, 1966, pp. 229–233.