NONLINEAR ESTIMATION THEORY

AND PHASE-LOCK LOOPS

by

JOHN STEVEN ETERNO

B.S., Case Western Reserve University, 1971

S.M., Massachusetts Institute of Technology, 1974

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1976

Signature of Author ______________________
Department of Aeronautics and Astronautics
September 8, 1976

Certified by ______________________
Thesis Supervisor

Certified by ______________________
Thesis Supervisor

Certified by ______________________
Thesis Supervisor

Certified by ______________________
Thesis Supervisor

Accepted by ______________________
Chairman, Departmental Graduate Committee

This report was prepared at The Charles Stark Draper Laboratory, Inc. under Independent Research and Development Grant 18703.

NONLINEAR ESTIMATION THEORY AND PHASE-LOCK LOOPS

by

John Steven Eterno

Submitted to the Department of Aeronautics and Astronautics on August 23, 1976, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The basic problem of phase estimation for sinusoids of Gaussian phase processes transmitted over channels with additive white Gaussian measurement noise is considered. In high signal-to-noise ratio (SNR) regions, where the classic phase-lock loop (PLL) is the optimal tracking filter, a unique frequency acquisition scheme is developed.

In low SNR environments, where the class PLL is not optimal, Bucy's representation theorem is used to motivate novel approximations to the exact conditional probability density function. This approximation method is quite general and may be used in other nonlinear problems of low dimension. The technique has the advantage of producing positive approximate density functions which converge to the correct density as the process driving noise strength goes to zero, or as the order of the approximation becomes infinite. The approximation method is applied to the design of phase estimators for first - and second - order PLL problems. For a high-noise first-order problem simulated, the first term in the approximation outperformed the classic PLL -- the extended Kalman filter for the problem.

Thesis Supervisor: Wallace E. Vander Velde, Sc. D.
Title: Professor of Aeronautics and Astronautics

Thesis Supervisor: Alan S. Willsky, Ph. D.
Title: Associate Professor of Electrical Engineering

Thesis Supervisor: Donald E. Gustafson, Ph. D.
Title: Technical Staff, C. S. Draper Laboratory

Thesis Supervisor: Jason L. Speyer, Ph. D.
Title: Professor of Aerospace Engineering and Engineering Mechanics, University of Texas

ACKNOWLEDGEMENTS

I would like to thank many people for their help, guidance and friendship during my research, and only regret that the following list is unavoidably incomplete.

First, thanks go to my advisory committee -- Wallace Vander Velde, chairman, Jason Speyer, Don Gustafson, and especially Alan Willsky, who did much to make my research both rigorous and rewarding. I would also like to thank Sheldon W. Buck of the Draper Laboratry for giving my education the practical perspective I missed at the Institute. Thanks are due to the Draper Laboratory for their financial assistance and faith through the years, in particular for their unselfish funding (through an independent research and development grant) of the research which led to this thesis. In addition, thanks to Margaret Flaherty for making the final preparation of this report as painless as possible.

I also wish to thank my family and close friends for their support throughout my marathon education. And finally, thanks to Becky, for making it all worthwhile.

TABLE OF CONTENTS

TABLE OF CONTENTS (con't)

TABLE OF CONTENTS (Con't)

TABLE OF CONTENTS (con't)

TABLE OF CONTENTS (con't)

CHAPTER I

INTRODUCTION AND MOTIVATION

1.1 Introduction

This thesis is concerned with nonlinear estimation theory in general and phase-lock loops (PLL's) in particular. We have concentrated on those areas where phase-lock loops perform poorly (during acquisition and periods of high noise) and have examined methods for the design of filters with improved performance. Some of our results are quite general, however, and have possible uses in other nonlinear filtering problems.

For a general class of nonlinear problems, we have developed an approximation technique for the conditional probability density function that makes no a priori assumptions about the shape or moments of the density. The result is an approximate density function that can be constructed from a finite set of statistics which are functionals of the measurements. We have applied this technique to the design of phase estimators without making the high signal-to-noise ratio assumptions inherent in phase-lock loops.

We have also investigated the acquisition behavior of phase-lock loops in the very-high signal-to-noise ratio area, and have found that significant improvement is possible without degrading the filtering properties of the loop. Our technique uses a compound PLL to move the "small sine" approximation in PLL design from the narrow band filtering loop to a wider-bandwidth "phase-detector" loop. The linearized noise analysis is barely influenced by this change, but the redistribution of filtering tasks greatly improves acquisition performance, as discussed in the next chapter.

1.2 History and Motivation

Perhaps the first person to propose a "phase-lock loop" was Bellescize in 1932 [3], who applied the idea to the synchronous reception of radio signals. After a rather slow beginning, phase-lock loops have steadily grown in importance and today are some of the most widely used nonlinear devices in this world of linear engineering.

The past analysis of "classic" phase-lock loops centered on the optimization of the given PLL structure for a class of communications problems. An excellent history of these efforts is contained in Ho's dissertation [18], while Klapper and Frankle include a good, brief history in their more accessible book [24]. As a general introduction to PLL's, Viterbi's excellent book [41] is recommended.

In recent years a number of investigators have used estimation theory to attack the phase-lock loop problem. Motivated in part by the growing use of PLL's in noncommunications areas (eg. motor speed control), researchers have been interested in improving upon PLL performance in the high-noise regions not usually encountered in communications.

Mallinckrodt, et. al. [29] were perhaps the first of these researchers. In 1970 they obtained some numerical results for the optimal nonlinear filter for a Browian motion phase process transmitted over a channel with additive Gaussian measurement noise. Their results provide a useful benchmark for evaluating sub-optimal filters, but the incredible complexity of their filter poses no threat to the simple PLL now available as an integrated circuit.

Subsequent work by Mallinckrodt, et. al. and their students [8, 10, 12, 16] has centered on extending the numerical results to a second-order PLL problem and improving the computational speed of their point-mass density approximation. Independent of their numerical studies, however, they [29] proposed a suboptimal filter - the "static phase" filter - for the first-order problem that turns out to be one of the filters that we derive from density approximations. This will be discussed in Chapter 5. Mallinckrodt, et. al. also first pointed out that the phase-lock loop can be considered an extended Kalman filter, although it was not designed as such. We demonstrate this in section 1.4.2.

In an attempt to obtain more useful sub-optimal filters, Gustafson and Speyer [15] developed a linear filter that minimized the error variance in the measurement space. This work resulted in a filter, a substitute for the first-order PLL, that works quite well at all signal-to-noise ratios. Moreover, upon closer examination, it turns out that this linear minimum-variance filter is a type of "static phase" filter and that it converges to the optimal filter as the process noise strength goes to zero. We also discuss this in Chapter 5.

In [44], Willsky used the technique of assumed density filtering to truncate the infinitely-coupled set of differential equations for the Fourier coefficients of the conditional desnity in the first-order problem. His analysis resulted in a filter that slightly out performed that of Gustafson and Speyer, but with a sizeable increase in complexity.

Recently, Tam and Moore [39] have applied the Gaussian-sum technique of Sorenson and Alspach [2, 36] to produce a collection of filters that

that converges to the optimal as the number of filters increases. Their results have also been quite good, although their higher-order filters become very complex, with occasional ad-hoc re-initialization required.

Thus, some of the biggest guns of nonlinear estimation theory have been brought to bear on the PLL problem. They have shown that carefully-designed sub-optimal filters can offer improved performance over classic PLL's. However, no general method has been developed for examining and analytically approximating the conditional density. In this thesis we describe a method for approximating the conditional density that is fundamentally different from the numerical approximations of [11] and the Gaussian sum approximation of [2, 36]. We then apply this method to the design of sub-optimal filters for the first- and second-order PLL problems.

1.3 General Nonlinear Problem

1.3.1 Problem Statement

The central problem that we consider involves a Gaussian phase process $\theta(t)$ transmitted over a noisy channel and received as $\dot{z}(t)$

$$\dot{z}(t) = A \sin(\omega_c t + \theta(t)) + \dot{n}(t) \tag{1.1}$$

where A is a known amplitude, ω_c is a known carrier frequency, and $\dot{n}(t)$ is a white Guassian noise process. Our problem is to estimate $\theta(t)$ given measurements $\dot{z}(s)$ for $0 \leq s \leq t$.

This model is a very good one for a wide variety of communications problems. A Guassian phase process is a suitable representation for a number of modulation techniques and information processes, and the additive channel noise is an excellent model for most radio receivers. It should be noted that we consider the Gaussian message spectrum to refer to the process we are trying to estimate rather than "oscillator jitter" or some other noise that corrups the process we are trying to estimate. The message (θ) will be modelled as filtered white noise, with the filter chosen to provide the desired spectrum. This concept is familiar to control engineers from the Kalman filter problem formulation and to communications engineers as a shaping filter.

This problem belongs to the general class of problems where a Gaussian state vector (x) propagates through the differential equation

$$\dot{x} = Fx + G\dot{u} \tag{1.2}$$

where $\dot{u}(t)$ is a zero-mean white Guassian noise process with covariance

$$E[\dot{u}(t)\dot{u}^T(\tau)] = Q\delta(t-\tau) \tag{1.3}$$

where δ is the Dirac delta function, and θ is positive semidefinite.

Several comments about notation are in order. We do not distinguish between vectors and scalars, in general. We also have suppressed the time dependence of all of our functions, except where necessary. Thus

$$x = x(t) = x_t$$

where the subscript notation will be used whenever there is no danger of confusion with vector components. We have also called "$\dot{u}$" white noise, in place of the (perhaps) more familiar "u". Our notation will facilitate later conversion to an Ito calculus framework.

Thus, we will also write equation (1.2) as

$$dx = Fx\ dt + G\ du \tag{1.4}$$

where

$$E[du\ du^T] = Q\ dt \tag{1.5}$$

and u is a Brownian motion process.

We assume that the initial density for x is Gaussian with mean $\bar{x}_0$ and covariance P_0, which we denote by

$$x_0 \sim N(x_0, P_0)$$

Then the mean $(\bar{x})$ and covariance (S) of x propagate through the equations

$$\dot{\bar{x}} = F\bar{x} \tag{1.6}$$

$$\dot{S} = FS + SF^T + GQG^T \tag{1.7}$$

We also assume that there is a measurement $(\dot{z})$ of x available:

$$\dot{z} = h(x) + \dot{n} \tag{1.8}$$

where h is a vector-valued function of x, and $\dot{n}$ is a zero-mean, white Gaussian noise process with

$$E[\dot{u}(t)\dot{u}^T(z)] = R\delta(t-\tau) \tag{1.9}$$

where R is positive definite.

We also write

$$dz = h(x)dt + dn \tag{1.10}$$

Finally, we assume that x_0, u_t and n_t are independent for all t.

Given all of these conditions, our problem is to estimate x_t given all of the measurements up to time t.

Thus, we want

$$p(x,t|z_0^t)$$

where

$$z_0^t = \{z_s;\ 0 \leq s \leq t\}$$

This problem is quite straightforward: a finite number of parameters totally specify the system and measurement. Only the solution is difficult.

If the measurement were linear in x, that is, if $h(x) = Cx$, we would have a linear filtering problem and the solution would be given by the Kalman Filter (see, e.g., Bryson and Ho [6])

$$p(x,t|z_0^t) = N(\hat{x}, P)$$

where

$$\dot{\hat{x}} = F\hat{x} + PC^TR^{-1}(\dot{z}-C\hat{x}) \tag{1.11}$$

$$\dot{P} = FP + PF^T + GQG^T - PC^TR^{-1}CP \tag{1.12}$$

Since h is <u>not</u> a linear function of x, however, things get more complicated. The conditional density is <u>not</u> Gaussian, and no standard method exists for finding even a finite number of sufficient statistics with which to create the conditional density.

1.3.2 Extended Kalman Filter

One standard way to obtain an approximate answer is to create a Kalman Filter linearized about the current best estimate of the state x. That is, define

$$H = \left. \frac{\partial h(x)^T}{\partial x} \right|_{x=\hat{x}_t} \tag{1.13}$$

and use the filter

$$\dot{\hat{x}} = F\hat{x} + PH^T R^{-1}(\dot{z} - h(\hat{x})) \tag{1.14}$$

$$\dot{P} = FP + PF^T + GQG^T - PH^T R^{-1} HP \tag{1.15}$$

This is the so-called "extended Kalman filter" for this problem. It is particularly relevant to us because it happens that phase-lock loops <u>are</u> extended Kalman filters, as we demonstrate later in this chapter. There is also a marked similarity between extended Kalman filters and a linearized (realizable) version of the "MAP" estimator of Van Trees [40, sec. 2.4] for general nonlinear filtering problems.

One often-mentioned "drawback" to extended Kalman filters is that the gains (P) depend on the data ($\dot{z}$) through H, a function of $\hat{x}$. Thus P is not precomputable, as it is for the regular Kalman filter. Actually, in nonlinear problems the gains <u>must</u> depend on the data, and if there is a drawback to this filter, it is that the dependence is through $\hat{x}$ rather than $(\dot{z}-h(\hat{x}))$, as would be the case for the real covariance equation (see [19]). Extended Kalman filters occasionally "diverge" (the error $(x-\hat{x})$ increases) because the gains depend on $\hat{x}$ and not $(\dot{z}-h(\hat{x}))$. Gains which are inappropriate for the real state make the filter respond incorrectly to the data, causing the error to increase and the gains to become even worse.

1.3.3 Threshold and Acquisition Problems

There are two primary causes of divergence. The first we will call the "threshold phenomenon", after experience with nonlinear filters in communications. If the noise strengths are low, the filter may work quite well, since the feedback nature of the design will tend to keep the error small and the linearization valid. As the noise gets stronger, however, the performance may abruptly degrade when the filter can no longer reinforce its own linearization assumption. The "threshold" is that value of the noise at which the performance suddenly degrades.

The second area of poor performance is during "acquisition." Even though the filter would work if the error <u>became</u> small (the noise is weak enough to justify the linearization), the filter may not be able to reduce a large initial error by itself.

When extended Kalman filters don't work well, what can be done? For the PLL acquisition problem, we propose (in the next chapter) a method of improving acquisition performance without sacrificing the noise-filtering properties of the loop. We believe that this technique offers significant advantages over the external acquisition aids that are usually used.

For the threshold problem, we propose a filtering technique that avoids the linearization usually used to make the problem tractable. Our method approximates the exact answer for our problem, rather than developing an exact answer for an "approximate" problem. In Chapter 3, we develop the exact answer that we will need — Bucy's Representation Theorem.

In Chapter 4, we derive our approximation method. In Chapter 5, we investigate the 1st order PLL problem, comparing results from our method with those of other researchers. Chapter 6 discusses the 2nd order PLL problem, and Chapter 7 contains a summary and conclusion.

We begin by describing the specific phase-lock loop problem which will dominate our investigation. We examine the classic PLL, demonstrate that it _is_ an extended Kalman Filter, and show that for the 1st order loop one has to change the filter structure to improve performance.

1.4 General PLL Problem

1.4.1 Introduction

The general phase-lock loop problem we consider involves a Gaussian state vector x_t as in equation (1.2). The received signal is assumed to be

$$\dot{z}' = A \sin(\omega_c t + \theta_t) + \dot{n}' \tag{1.16}$$

where $\dot{z}'$ is a scalar, A is a known amplitude, ω_c is a known carrier frequency, θ_t is the first component of the vector x_t, and $\dot{n}'$ is a zero mean white Gaussian noise process of two-sided spectral height "r."

Since A is known, we may assume "without loss of generality" that $A = 1$. (If $A \neq 1$, we may obtain a measurement equivalent to $\dot{z}'$ by dividing by A, thus rescaling the noise strength by $1/A^2$.)

We now need to define carefully what we mean by "white Gaussian noise." Specifically, we assume that the spectral density of $\dot{n}'$ is flat, of height r, from $\omega_c - \omega$ to $\omega_c + \omega$ (where $\omega < \omega_c$) and negligible outside that region. Then we can decompose $\dot{n}'$ (see Viterbi [41] Chapter II or Van Trees [40] Chapter 2) into

$$\dot{n}' = \dot{n}_1 \cos \omega_c t + \dot{n}_2 \sin \omega_c t \tag{1.17}$$

where $\dot{n}_1$ and $\dot{n}_2$ are independent, zero mean white Gaussian noise processes of strength "2r" (that is, a flat spectral density of height 2r from $-\omega$ to ω). Then we may form $\dot{z}_1$ and $\dot{z}_2$ by multiplying ("heterodyning") $\dot{z}'$ by $2 \cos \omega_c t$ and $2 \sin \omega_c t$ (respectively) and then low-pass filtering (to remove the "$2\,\omega_c$" terms)

$$\begin{pmatrix} \dot{z}_1 \\ \dot{z}_2 \end{pmatrix} = \begin{pmatrix} \sin \theta \\ \cos \theta \end{pmatrix} + \begin{pmatrix} \dot{n}_1 \\ \dot{n}_2 \end{pmatrix} \tag{1.18}$$

We refer to the vectors $\dot{z}$ and $\dot{n}$ as the baseband signal and noise respectively

$$\dot{z} = \begin{pmatrix} \dot{z}_1 \\ \dot{z}_2 \end{pmatrix} \quad \text{and} \quad \dot{n} = \begin{pmatrix} \dot{n}_1 \\ \dot{n}_2 \end{pmatrix} \qquad (1.19),\ (1.20)$$

Also, for some function $\tilde{\theta}$, we may heterodyne by 2 $\cos(\omega_c t + \tilde{\theta})$ and 2 $\sin(\omega_c t + \tilde{\theta})$ to obtain $\dot{z}_I$ and $\dot{z}_Q$ respectively*, where

$$\begin{pmatrix} \dot{z}_I \\ \dot{z}_Q \end{pmatrix} = \begin{pmatrix} \sin(\theta-\tilde{\theta}) \\ \cos(\theta-\tilde{\theta}) \end{pmatrix} + \begin{pmatrix} \dot{n}_I \\ \dot{n}_Q \end{pmatrix} \tag{1.21}$$

where

$$\dot{n}_I = \dot{n}_1 \cos\tilde{\theta} - \dot{n}_2 \sin\tilde{\theta} \tag{1.22}$$

$$\dot{n}_Q = \dot{n}_1 \sin\tilde{\theta} + \dot{n}_2 \cos\tilde{\theta} \tag{1.23}$$

If $\tilde{\theta}$ is "slower" than white noise, that is, at least one integration removed from $\dot{z}$ (e.g., $\dot{\tilde{\theta}} = f(\dot{z})$ but <u>not</u> $\tilde{\theta} = f(\dot{z})$), then (see again Viterbi [41] or Van Trees [40]) $\dot{n}_I$ and $\dot{n}_Q$ may be considered zero mean white Gaussian processes independent of each other <u>and</u> $\tilde{\theta}$, with 2 sided spectral height "2r." We will use these relations throughout this work.

It is worth stressing that the baseband measurements (1.18) fit neatly into our formula for the general nonlinear problem (1.8), with

$$h(x_t) = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix}, \qquad \theta = (x_t)_1$$

We now want to restrict our phase processes to those realizable with a voltage-controlled oscillator (VCO). This is a device whose

---

*The I and Q subscripts refer to "In phase" and "Quadrature", and will assume more meaing in later chapters.

output is a sinusoid at an instantaneous frequency proportional to the input voltage. The output frequency for no input is called the "quiescent" frequency, which we will assume to be equal to the carrier frequency. The main reason for our constraint is to allow "optimal" phase-lock loops to be constructed for our signals. It is, moreover, a very reasonable restriction, since most transmitters involve a VCO.

For the phase process θ, where x_t is an n-vector, we have

$$\theta = (x_t)_1 \tag{1.24}$$

$$\dot{x} = Fx + G\dot{u} \tag{1.25}$$

From the VCO restriction F must have zeros in the first column:

$$F = \begin{bmatrix} 0 & B \\ \cdot & \\ \cdot & \\ \cdot & F' \\ \cdot & \\ 0 & \end{bmatrix} \tag{1.26}$$

where B is a $1 \times (n-1)$ row vector and $\mathbf{F}'$ is an $(n-1) \times (n-1)$ matrix. Thus, no function of θ is fed back to the higher-order derivatives (since θ is not availble at the VCO output, only $\sin(\omega_c t + \theta)$ or $\cos(\omega_c t + \theta)$).

We remark that this model allows frequency modulation (FM) signal forms, where

$$x = \left(\frac{\theta}{\mathbf{x}'}\right) \quad \text{and} \quad \dot{u} = \left(\frac{0}{\dot{u}'}\right)$$

with

$$\dot{\theta} = \omega = (x)_2 = (x')_1$$

Then

$$\dot{x}' = F'x' + G'\dot{u}' \tag{1.27}$$

Given the measurement (1.18) and signal (1.25), we now construct a PLL for our problem. Examining the in-phase measurement $\dot{z}_I$, we note that if

$$\tilde{\theta} \simeq \theta$$

then

$$\sin(\theta-\tilde{\theta}) \simeq \theta-\tilde{\theta} \tag{1.28}$$

and we may consider the linear problem with "pseudo-measurement"

$$\dot{z}_p = \theta + \dot{n}_I \tag{1.29}$$

Now we take $\tilde{\theta}=\hat{\theta}$, where $\hat{\theta}$ is the conditional mean from a Kalman filter designed for the linear (pseudo) problem. Thus

$$\hat{\theta} = (\hat{x})_1 \tag{1.30}$$

$$\dot{\hat{x}} = F\hat{x} + PC^T \frac{1}{2r} (\dot{z}_P-\hat{\theta}) \tag{1.31}$$

$$\dot{P} = FP + PF^T + GQG^T - \frac{1}{2r} PC^T CP \tag{1.32}$$

where

$$C = (1, 0, \ldots, 0) \tag{1.33}$$

In reality, of course,

$$\dot{z}_P - \hat{\theta} = \dot{z}_I = \sin(\theta-\hat{\theta}) + \dot{n}_I \tag{1.34}$$

We can now implement this filter, because of our restriction on F, in a classic VCO loop as follows: We first form the gain vector K,

$$K = \frac{1}{2r} PC^T = \frac{1}{2r} \begin{pmatrix} P_{11} \\ P_{12} \\ \vdots \\ P_{1n} \end{pmatrix} \tag{1.35}$$

and partition it into K_1 and K',

$$K = \begin{bmatrix} K_1 \\ \hline K' \end{bmatrix} = \frac{1}{2r} \begin{pmatrix} P_{11} \\ \hline P_{12} \\ \vdots \\ P_{1n} \end{pmatrix} \tag{1.36}$$

We also define

$$\theta = x_1$$

$$x' = \begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Then the PLL (1.30 - 1.32) becomes

$$\dot{\hat{\theta}} = B\hat{x}' + K_1 \dot{z}_I \tag{1.37}$$

$$\dot{\hat{x}}' = F'\hat{x}' + K'\dot{z}_I \tag{1.38}$$

We construct this filter in figure 1.1.

If the system matrix (F) is not time-dependent and we let the filter gains (K) go to their steady state values, we can represent the "Linear" Filter" in Figure 1.1 by the Laplace Transform of its transfer function (A(s)), arriving at the diagram in Figure 1.2.

Using the low-pass filtering assumptions described above, we may use the equivalent "Baseband" representation of the PLL, as in Figure 1.3.

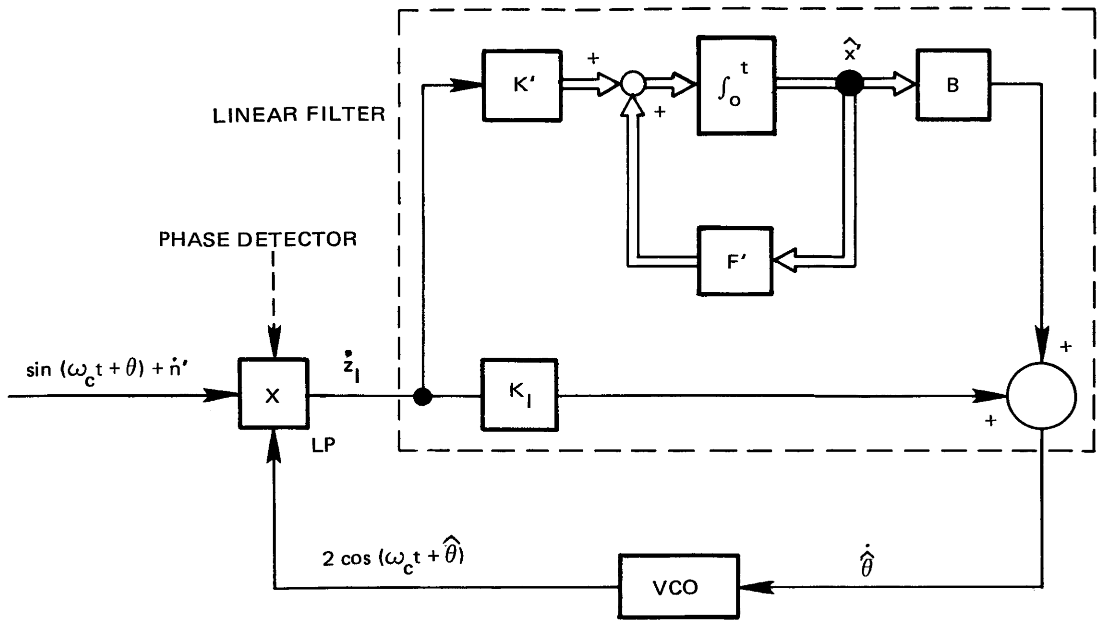


Figure 1.1 Phase-Lock Loop*

We have added the labels "phase detector" and "linear filter" in their usual places.

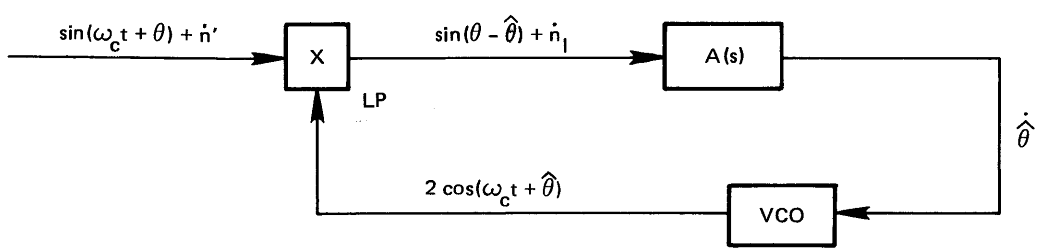


Figure 1.2 Laplace Form of PLL with VCO

*The "LP" after the multiplier refers to the low-pass filtering required to remove the "$2\omega_c$" terms. The double lines in the diagram indicate vector signals.

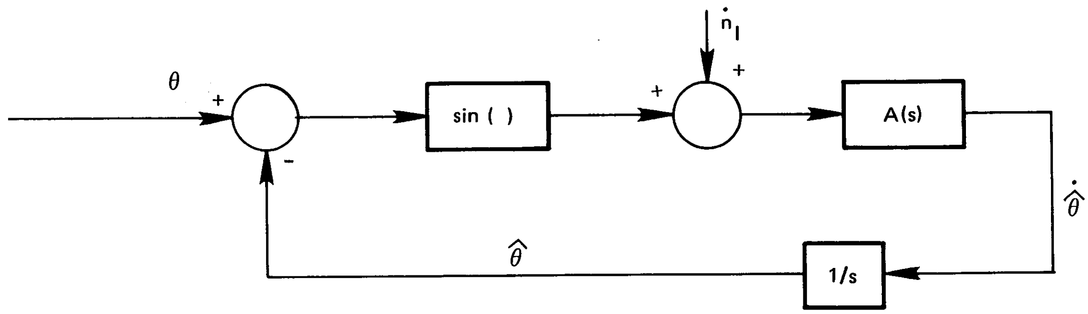


Figure 1.3 Baseband PLL

We note that the "order" of a PLL is the number of integrators in the linearized model or, for our use, the dimension (n) of x. The order of A(s) is seen to be (n-1), with the VCO supplying the final integration.

1.4.2 PLL-Extended Kalman Filter Equivalence

Although it may not be obvious at this point, when the linearized PLL is constructed as a Kalman filter (as above), the actual PLL may be regarded as an extended Kalman filter. This result was first noted in Mallinckrodt, et. al. [29] and Bucy and Mallinckrodt [10]. It is not widely recognized, however, so we give a brief demonstration here.

Using the baseband measurement (1.18), we define

$$H(\hat{\theta}) = \left.\frac{\partial h^T}{\partial x}\right|_{x = \hat{x}} = \begin{bmatrix} \cos\hat{\theta}, 0, \ldots, 0 \\ -\sin\hat{\theta}, 0, \ldots, 0 \end{bmatrix} \tag{1.39}$$

with

$$R = \begin{pmatrix} 2r & 0 \\ 0 & 2r \end{pmatrix} \tag{1.40}$$

Then for the extended Kalman filter gain equation (1.15), we need

$$PH^TR^{-1}HP = \frac{1}{2r} P \begin{bmatrix} 1, & 0, & \dots, & 0 \\ 0 & 0 & & \vdots \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 0 \end{bmatrix} P \tag{1.41}$$

Thus the "covariance" equation <u>decouples</u> from the estimate equation, and in fact the gains become precisely those of the linearized filter (eq. 1.32) since

$$PH^TR^{-1}HP = \frac{1}{2r} PC^TCP \tag{1.42}$$

The estimate equation for this extended Kalman filter is given by

$$\dot{\hat{x}} = F\hat{x} + PH^TR^{-1} [\dot{z} - h(\hat{\theta})] \tag{1.43}$$

but

$$H^TR^{-1}h(\hat{\theta}) = \frac{1}{2r} \begin{bmatrix} \cos\hat{\theta}\sin\hat{\theta} - \sin\hat{\theta}\cos\hat{\theta} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{1.44}$$

$$= 0$$

and

$$H^TR^{-1}\dot{z} = \frac{1}{2r} \begin{bmatrix} \cos\hat{\theta}\,\dot{z}_1 - \sin\hat{\theta}\,\dot{z}_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{1.45}$$

$$= \frac{1}{2r} \begin{bmatrix} \dot{z}_I \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \frac{1}{2r} C^T \dot{z}_I$$

Thus

$$\dot{\hat{x}} = F\hat{x} + \frac{1}{2r} \; PC^T \; \dot{z}_I \tag{1.46}$$

$$\dot{P} = FP + PF^T + GQG^T - PC^T CP \frac{1}{2r} \tag{1.47}$$

just as in the PLL definitions. Thus, the PLL is an extended Kalman filter for our broad class of problems (1.26).

One of the most interesting implications of this is that for the 1st-order PLL problem (which we discuss in the next section), the PLL represents one of the few cases where the error density of an extended Kalman filter is known. A second intriguing idea is that well-documented PLL acquisition and threshold problems may explain extended-Kalman filter divergence under conditions of poor initial conditions or too much noise.

1.4.3 Brownian Motion Phase Process

1.4.3.1 First-Order PLL

In this section we wish to investigate the 1st-order PLL in detail. This filtering problem is one of the most analyzed nonlinear problems and indeed contains all of the basics with few added complexities. (It is simple but unsolvable.)

The signal form we assume is a Brownian motion phase process:

$$\dot{\theta} = \dot{u} \tag{1.48}$$

$$E[\dot{u}(t) \; \dot{u}(\tau)] = q \; \delta(t-\tau) \tag{1.49}$$

with an initial density:

$$p(\theta) = \frac{1}{2\pi} \qquad -\pi < \theta \leq \pi \tag{1.50}$$

We also assume that we have the baseband phase measurements (1.18) and (1.21).

The Kalman filter for the linearized problem has an error variance that propagates as

$$\dot{P} = q - P^2/2r \tag{1.51}$$

so that in steady-state

$$P_{\theta\ell} = \sqrt{2qr} \tag{1.52}$$

where the "$\theta\ell$" subscript is somewhat standard, and refers to the linearized analysis.

The optimal gain is then

$$K = P_{\theta\ell}\big/2r = \sqrt{q/2r} \tag{1.53}$$

and the "linear filter" in figure 1.3 is simply

$$A(s) = K \tag{1.54}$$

We note that a PLL with <u>any</u> (positive) gain K_s will specify a linearized error equation of

$$\dot{\varepsilon} = -K_s\varepsilon - K_s\dot{n}_I + \dot{u} \tag{1.55}$$

for

$$\varepsilon = \theta-\hat{\theta} \tag{1.56}$$

The steady-state (linear-predicted) error variance (P) is then given by

$$P = \frac{\sigma^2}{2K_s} \tag{1.57}$$

where

$$\sigma^2 = 2rK_s^2 + q \tag{1.58}$$

The actual error propagates according to the differential equation

$$\dot{\varepsilon} = -K_s \sin \varepsilon - K_s \dot{n}_I + \dot{u} \tag{1.59}$$

so that the density for ε satisfies the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial \varepsilon} [K_s \sin \varepsilon] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial \varepsilon^2} P \tag{1.60}$$

In steady-state, this equation can be solved (see Viterbi [41]), for a modulo-2π error ε, to yield

$$P(\varepsilon) = \frac{e^{\alpha \cos \varepsilon}}{2\pi I_0(\alpha)} \qquad -\pi < \varepsilon \leq \pi \tag{1.61}$$

where

$$\alpha = \frac{2K_s}{\sigma^2} = \frac{1}{P} \tag{1.62}$$

and $I_0(\alpha)$ is a modified Bessel function (see Appendix A).

The actual error variance may be found from the identity (see, e.g., [1], p. 376)

$$e^{\alpha \cos \varepsilon} = I_0(\alpha) + 2 \sum_{K=1}^{\infty} I_K(\alpha) \cos K\varepsilon \tag{1.63}$$

to yield

$$E[\varepsilon^2] = \frac{\pi^2}{3} + 4 \sum_{K=1}^{\infty} \frac{(-1)^K}{K^2} \frac{I_K(\alpha)}{I_0(\alpha)} \tag{1.64}$$

which is tabulated in Van Trees [40].

1.4.3.2 Optimal PLL Gain

In this section we will show that the optimal gain from the linearized analysis is the overall optimal constant gain for the 1st-order PLL. Thus, if one wants to improve performance over that of the PLL, the filter structure must be changed.

We consider the cost function $f(\varepsilon)$, where ε is the phase error and f is any positive, symmetric (about the origin) function that is monotonically increasing on the interval $[0, \pi]$. In particular, such common cost functions as

$$f(\varepsilon) = \varepsilon^2 \tag{1.65}$$

and

$$f(\varepsilon) = 2(1-\cos \varepsilon) \tag{1.66}$$

are included.* Then, using the error density from equation (1.61), we have

$$E[f(\varepsilon)] = 2 \int_0^{\pi} f(\varepsilon)p(\varepsilon)\, d\varepsilon \tag{1.67}$$

*The cosine cost function resembles ε^2 for small ε, but is insensitive to modulo-2π errors. We will encounter it again in the next section and in Chapter 5.

we note that

$$\frac{\partial}{\partial \varepsilon} p(\varepsilon) = -\alpha \sin \varepsilon \, p(\varepsilon) \tag{1.68}$$

so that the density is a strictly decreasing function of ε on $[0, \pi]$ for $\alpha > 0$.

We first examine the behavior of the center point, $p(0)$, as α varies. We have:

$$p(0) = \frac{e^{\alpha}}{2\pi I_0(\alpha)} \tag{1.69}$$

Thus

$$\frac{\partial p(0)}{\partial \alpha} = \frac{e^{\alpha}}{2\pi I_0(\alpha)} \left[\frac{I_0(\alpha) - I_1(\alpha)}{I_0(\alpha)}\right] \tag{1.70}$$

where we have used the Bessel function relations in Appendix A. Now since

$$\int_0^{\pi} \cos(\varepsilon) e^{\alpha \cos \varepsilon} \, d\varepsilon < \int_0^{\pi} e^{\alpha \cos \varepsilon} \, d\varepsilon$$

because

$$\cos(\varepsilon) < 1 \qquad 0 < e < \pi$$

We have that

$$I_0(\alpha) > I_1(\alpha)$$

and therefore

$$\frac{\partial p(0)}{\partial \alpha} > 0$$

We next examine the behavior of $p(\varepsilon)$ at the endpoints $\pm\pi$. Specifically

$$\frac{\partial p(\pm\pi)}{\partial \alpha} = \frac{e^{-\alpha}}{2\pi I_0(\alpha)} \left(\frac{-I_0(\alpha) - I_1(\alpha)}{I_0(\alpha)} \right) < 0 \tag{1.71}$$

Thus, for $\alpha_1 > \alpha_2$, we have

$$p(\varepsilon = 0, \alpha_1) > p(\varepsilon = 0, \alpha_2) \tag{1.72}$$

and

$$p(\varepsilon = \pm\pi, \alpha_1) < p(\varepsilon = \pm\pi, \alpha_2) \tag{1.73}$$

We also claim that there is a unique point $s \varepsilon (0, \pi)$ where

$$p(\varepsilon = s, \alpha_1) = p(\varepsilon = s, \alpha_2) \tag{1.74}$$

There is at least one such "s" since the densities are continuous and relationships 1.72 and 1.73 hold. We may solve for s explicitly as

$$\frac{e^{\alpha_1 \cos s}}{2\pi\ I_0(\alpha_1)} = \frac{e^{\alpha_2 \cos s}}{2\pi\ I_0(\alpha_2)} \tag{1.75}$$

so that

$$s = \cos^{-1} \left[\frac{\ln\ (I_0(\alpha_1)/I_0(\alpha_2))}{(\alpha_1 - \alpha_2)} \right] \tag{1.76}$$

$$s \varepsilon (0, \pi)$$

The relationship between the densities is shown in figure 1.4, where we have defined the areas A, B, and C as shown.

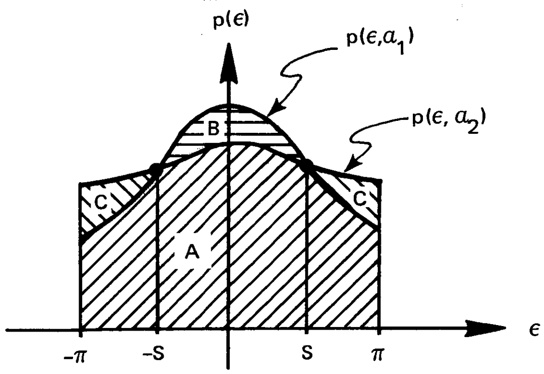


Figure 1.4 Relationship between $p(\varepsilon, \alpha_1)$ and $p(\varepsilon, \alpha_2)$ for $\alpha_1 > \alpha_2$

The areas are defined positive as

$$A = 2\int_0^s p(\varepsilon, \alpha_2)d\varepsilon + 2\int_s^\pi p(\varepsilon, \alpha_1)d\varepsilon \tag{1.77}$$

$$B = 2\int_0^s [p(\varepsilon, \alpha_1) - p(\varepsilon, \alpha_2)]d\varepsilon \tag{1.78}$$

$$C = 2\int_s^\pi [p(\varepsilon, \alpha_2) - p(\varepsilon, \alpha_1)]d\varepsilon \tag{1.79}$$

and the unit mass in each density implies that

$$A + B = A + C = 1$$

so that

$$B = C \tag{1.80}$$

We can now show the monotonicity of $E[f(\varepsilon)]$ with respect to α.* We define

$$\Delta = E[f(\varepsilon)|\alpha_1] - E[f(\varepsilon)|\alpha_2] \tag{1.81}$$

$$= 2\int_0^{\pi} f(\varepsilon)\ [p(\varepsilon,\alpha_1) - p(\varepsilon,\alpha_2)]d\varepsilon$$

for $\alpha_1 > \alpha_2$

Then

$$\Delta = 2\int_0^{s} f(\varepsilon)[p(\varepsilon,\alpha_1) - p(\varepsilon,\alpha_2)]d\varepsilon - 2\int_s^{\pi} f(\varepsilon)\ [p(\varepsilon,\alpha_2) - p(\varepsilon,\alpha_1)]d\varepsilon \tag{1.82}$$

We see that

$$2\int_0^{s} f(\varepsilon)[p(\varepsilon,\alpha_1) - p(\varepsilon,\alpha_2)]d\varepsilon < f(s)B \tag{1.83}$$

and

$$2\int_s^{\pi} f(\varepsilon)\ [p(\varepsilon,\alpha_2) - p(\varepsilon,\alpha_1)]d\varepsilon > f(s)C \tag{1.84}$$

because of our restrictions on $f(\varepsilon)$.

Thus

$$\Delta < f(s)\ [B-C] = 0 \tag{1.85}$$

---

*This approach was suggested by A. Willsky which means that

$$\alpha_1 > \alpha_2 \Longrightarrow E[f(\varepsilon)|\alpha_1] < E[f(\varepsilon)|\alpha_2] \tag{1.86}$$

And since $\alpha = 1/P$ (equation 1.62), where P is the linear-predicted error variance, we conclude that

$$P_1 < P_2 \Longrightarrow E[f(t)|P_1] < E[f(\varepsilon)|P_2] \tag{1.87}$$

Thus, the K_s that minimizes the linear predicted variance (the Kalman filter gain K with minimum variance $P_{\theta\ell}$) also minimizes $E[f(\varepsilon)]$ for the wide class of cost functions we consider.

1.4.3.3 Cosine Cost Function

We now present a different proof of the monotonicity of the expected value of the cosine cost function. We include this section, even though the general case was proven in the last section, because it demonstrates the type of manipulation of Bessel functions which we will find useful. These functions arise naturally in phase measurement problems, and we will encounter them again in later chapters.

We consider the function $f(\varepsilon)$

$$f(\varepsilon) = 2(1 - \cos \varepsilon) \tag{1.88}$$

which is periodic (of period 2π) and resembles the error-squared criterion for small values of ε. We begin by noting that

$$E[f(\varepsilon)] = 2\left[1 - \frac{I_1(\alpha)}{I_0(\alpha)}\right] \triangleq F(\alpha) \tag{1.89}$$

where we have used the following relationship from the Bessel function appendix (A).

$$E[\cos n\,\varepsilon] = \frac{I_n(\alpha)}{I_0(\alpha)} \qquad (1.90)$$

for

$$p(\varepsilon) = \frac{e^{\alpha \cos \varepsilon}}{2\pi I_0(\alpha)}$$

Then we have that

$$\frac{dF}{d\alpha} = -\frac{I_0^2(\alpha) + I_0(\alpha)I_2(\alpha) - 2I_1^2(\alpha)}{I_0^2(\alpha)} \qquad (1.91)$$

Now, using Schwarz's inequality

$$E[\cos \varepsilon]^2 \leq E[\cos^2 \varepsilon] \qquad (1.92)$$

with

$$E[\cos^2 \varepsilon] = \frac{1}{2}\,(1 + E[\cos 2\,\varepsilon]) \qquad (1.93)$$

we have

$$\left(\frac{I_1(\alpha)}{I_0(\alpha)}\right)^2 \leq \frac{1}{2}\left(1 + \frac{I_2(\alpha)}{I_0(\alpha)}\right) \qquad (1.94)$$

which becomes

$$I_0^2(\alpha) + I_0(\alpha)I_2(\alpha) - 2I_1^2(\alpha) \geq 0 \qquad (1.95)$$

so that

$$\frac{dF}{d\alpha} \leq 0 \qquad (\alpha \geq 0) \qquad (1.96)$$

Thus, F is a decreasing function of α, and an increasing function of $P(= 1/\alpha)$, and therefore the gain that minimizes P will also minimize F.

1.4.3.4 Concluding Remarks

We have shown that the expected value of a cost function is a monotone function of the linear predicted error variance in a first order phase-lock loop for a large class of cost functions.

Mallinckrodt, Bucy and Cheng have previously claimed that the actual error variance is a monotone function of the linear-predicted variance, but they offered no proof [29], and Van Trees has plotted the densities for different values of α (his Λ_m) without discussing monotonicity [40].

1.5 Summary and Synopsis

In this chapter we have defined the general nonlinear filtering problem of interest to us. We have discussed why the usual "linearized" filtering techniques fail and where we hope to improve upon their performance. We have also described the "phase-lock loop problem" and examined some of the interesting features of phase-lock loops.

The second chapter discusses the PLL "acquisition problem" (see section 1.3.3) in low-noise environments. We develop a "compound" phase-lock loop which dramatically improves the acquisition performance of a classic PLL without degrading its noise attenuation properties. This chapter "stands alone" as the only chapter concerned solely with high-SNR (signal-to-noise ratio) PLL acquisition and not with the general, all-SNR, nonlinear filtering problem.

The third chapter analyzes the "threshold problem" where extended Kalman filters (or other linearized filters) do not work well. This chapter discusses the nature of the complete solution to any nonlinear filtering problem - the conditional density function. We outline a derivation of one representation of this density (Bucy's representation theorem) which we will approximate in Chapter 4. Chapter 3 closes with two examples of the problem for which the representation theorem can be solved completely - the no-process-noise case.

Chapter 4 proposes a new method for approximating the conditional density function, as expressed in the third chapter, when there is process noise. This method generates approximate densities which (like the real density) are functions of the state x, time t, and measurement history z_0^t. The convergence of these approximate densities is discussed, and one of the approximations (the cumulant expansion) is shown to converge to the corrent density as the process-noise strength goes to zero (for any order approximation) and as the number of terms in the approximation becomes infinite (for any process-noise strength).

In the fifth chapter we consider the Brownian motion phase process (first-order PLL problem) introduced in section 1.4.3. The chapter begins by analyzing sub-optimal filters proposed by recent researchers in the area and noting "hidden" filter equivalences and high-SNR convergence properties. Principally, however, this chapter demonstrates the application of the approximate-density filtering technique developed in Chapter 4. Several approximate-density filters are designed and compared to the PLL and the other sub-optimal filters. Computer simulations of several of of these filters are described for the high-noise (low SNR) area where the PLL performance is poor. Finally, a modification of one of the approximate-density filters is derived and is shown to offer increased performance with a modest increase in implementation complexity.

Chapter 6 applies the approximation method of Chapter 4 to the design of sub-optimal filters for three phase-measurement problems usually solved by second-order PLL's. The filters are shown to resemble the (infinite-dimensional) optimal filter when the process noise strength is zero. Implementation and approximation techniques are suggested for the filter, and simplifications are discussed for cases which require only phase <u>or</u> frequency estimation.

The last chapter (7) summarizes the unique contributions of the thesis and suggests several areas for future research. A number of appendices are also included for reference. In addition to the appendices containing computational details, Appendix A discusses some interesting properties of Bessel functions and Appendix C summarizes useful results in Ito (stochastic) calculus.

CHAPTER 2

ACQUISITION IMPROVEMENT FOR PLL'S IN HIGH-SNR APPLICATIONS

2.1 Introduction

The design of classic phase-lock loops for high signal-to-noise ratio (SNR) regions is often frustrating because the procedure used to reduce the steady-state error covariance also degrades the acquisition performance. In general, the "noise bandwidth" of the loop [41], that bandwidth related to the measurement noise that passes through to the phase estimate, is proportional to the acquisition bandwidth. Wideband loops tend to acquire better and pass more noise, narrowband loops tend to filter better but take longer to acquire, or in some cases, acquire only a narrower range of frequencies.

Classic PLL's often are designed, therefore, by compromising acquisition and filtering performance. When the frequency uncertainty is large enough so that a PLL that could acquire the signal would pass an intolerable amount of noise, however, no single PLL can do the job, and various acquisition aids must be employed. One method is to slew the VCO with a voltage ramp until "phase-lock" (no cycle-skipping) is detected. A second method is to change the bandwidth of the loop by changing loop components. When the frequency uncertainty is quite large, it is reasonable to use a bank of frequency detectors operating in parallel to estimate the carrier frequency, then slew the VCO to the estimated value.

If the frequency ever shifts, however (e.g. if the receiving or transmitting stations ever change velocities causing a doppler frequency shift), these schemes require "loss of lock" detection, the energizing of the external acquisition aid or the switching of loop components, and the subsequent lock detection and de-energizing of the acquisition system. This takes time and makes the receiver quite complex.

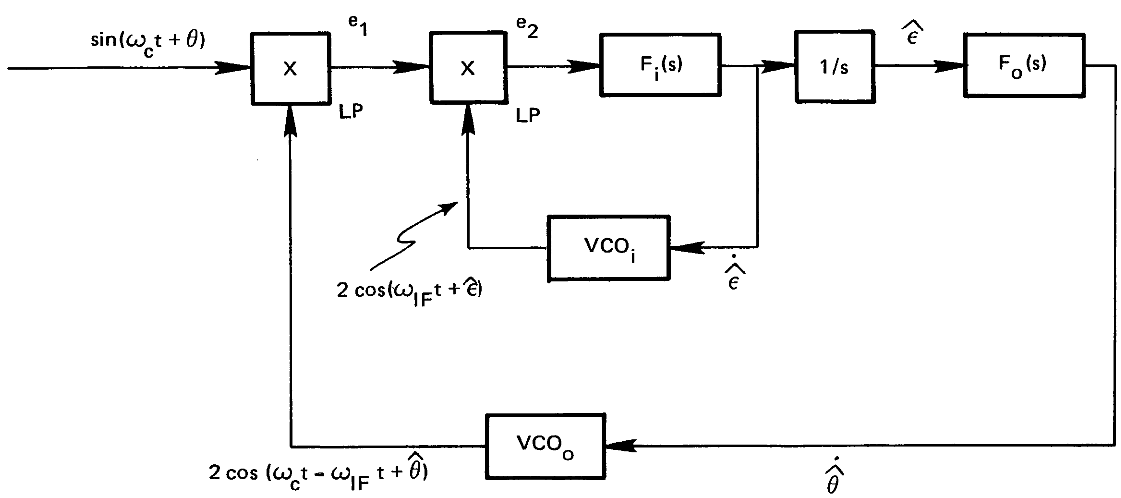
We have investigated the possibility of using one wideband PLL as the phase-detector for a narrowband PLL, so that the acquisition bandwidth of the wideband loop can be combined with the filtering properties of the narrowband PLL. This technique, which we describe in this chapter, offers a significant advantage over other acquisition schemes for high SNR applications. The improved acquisition performance is always available, without component switching or time delay, and without degrading the noise attenuation properties of the narrow-band loop.

2.2 Compound PLL

2.2.1 General Description

The simplest implementation of our idea involves a modification to the linear filter (A(s)) of the classic PLL (Figure 1.3), and a consequent reassignment of variable locations (the phase estimate is not fed back alone). The particular design philosophy, however, resulted from an attempt to improve the phase-detector section of the classic PLL and extend the linear operating region to phase errors of more than π radians.

We thus may explain our concept by describing a "compound" PLL, where an inner broadband phase-lock loop is used as an extended-range phase-detector for an outer, narrow-band loop. If the initial frequency offset is within the capture range of the inner loop, it will track the signal fed to it (the outer loop phase error) over a nearly unlimited range of phase errors. Then the whole combination will operate linearly, even though the frequency offset may be too great for initial acquisition by the narrow-band loop alone (with a sinusoidal phase detector). Figure 2.1 demonstrates one realization of a compound PLL.



$\epsilon = \theta - \hat{\theta}$ = "phase error"

$e_1 = \sin(\omega_{IF}t + \epsilon)$

$e_2 = \sin(\epsilon - \hat{\epsilon})$

Fig. 2.1 Compound Phase-Lock Loop

We have made two low-pass filtering assumptions here. The signal out of the first phase detector is really

$$e_1 + \sin(2\omega_c t - \omega_{IF} t + \theta + \hat{\theta})$$

We assume that we can low-pass-filter this signal to remove the high frequency term (or we can design the inner loop to respond only to the lower frequency component e_1).

The signal out of the second phase detector is really

$$e_2 + \sin(2\omega_{IF} t + \varepsilon + \hat{\varepsilon})$$

We assume that this signal also may be low-pass-filtered to remove the $2\omega_{IF}$ term. (The need to remove this $\varepsilon + \hat{\varepsilon}$ term is the reason that ω_{if} cannot be zero.) If these assumptions are justified, we may proceed to the baseband model of Figure 2.2.

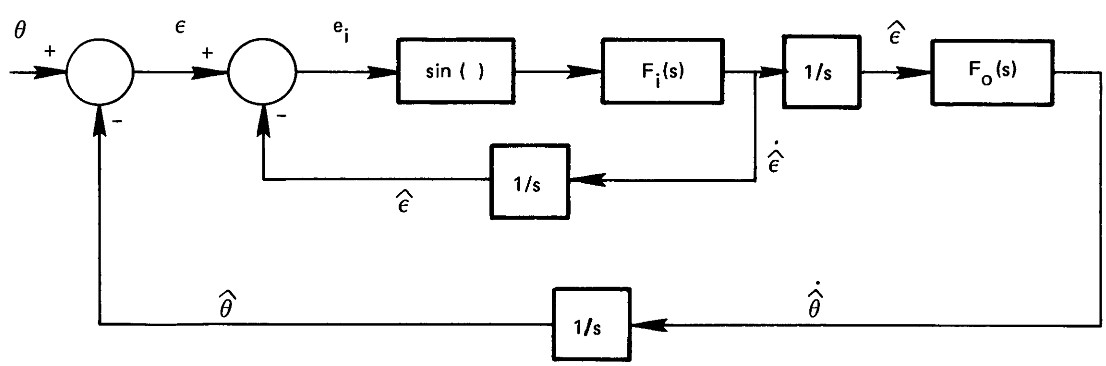


Fig. 2.2 Baseband Compound PLL

The basic idea behind this approach is to use a broadband filter $F_i(s)$ to quickly acquire and estimate ε. By integrating the output ($\dot{\hat{\varepsilon}}$) of this loop we obtain an estimate of ε that is linear over a wide range of phase errors, requiring only that the inner loop remain in lock. The outer loop then performs the desired filtering of $\hat{\varepsilon}'$, as in the classic PLL where $\hat{\varepsilon} = \sin \varepsilon$. Our construction here requires that the integrator after the inner loop perfectly match the inner VCO integrator. We will later describe an implementation that bypasses this unrealizable constraint.

The advantage of the compound loop approach is that the acquisition properties of a broadband loop can be combined with the noise attenuation properties of a narrow-band PLL.

2.2.2 Simple Implementation

An easier implementation of these ideas exists. Looking again at the baseband compound loop (Figure 2.2), we see that the two feedback integrators (for $\hat{\varepsilon}$ and $\hat{\theta}$) may be replaced by one acting on the sum $\dot{\hat{\varepsilon}}$ plus $\dot{\hat{\theta}}$. This single integrator and sine nonlinearity may be realized by a single PLL, as shown in Figures 2.3 and 2.4.

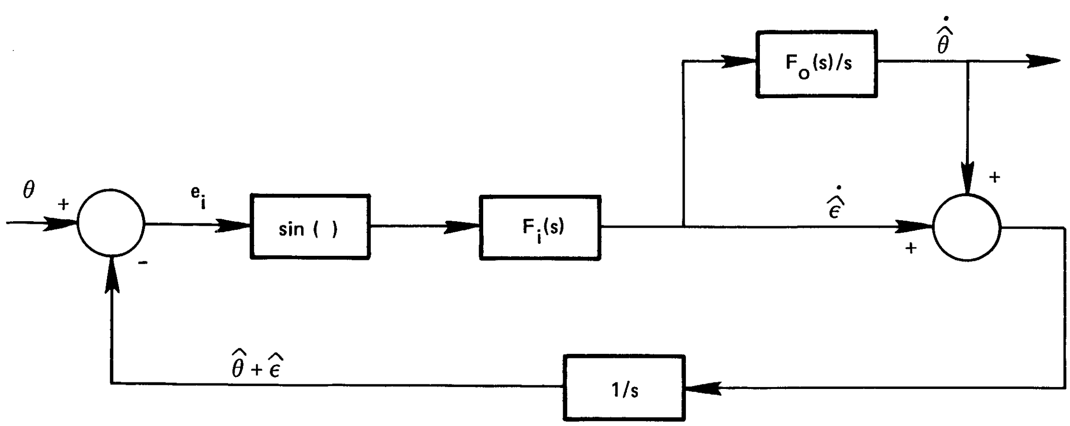


Fig. 2.3 Baseband Compound PLL

We see that the feedback path and estimate location are quite different from those of a classic PLL. We remark that one may be tempted to use $\hat{\varepsilon}$, as an estimate of the <u>error</u> in $\hat{\theta}$, to improve the phase estimate $\hat{\theta}$. Since $\hat{\varepsilon}$ comes from the wideband loop gain $F_i(s)$, however, $\hat{\varepsilon}$ will in fact be much noisier than $\hat{\theta}$ and should not be used to "improve" it. The signal $\hat{\varepsilon}$ should be used only to improve the acquisition and tracking of the loop, not the phase estimate.

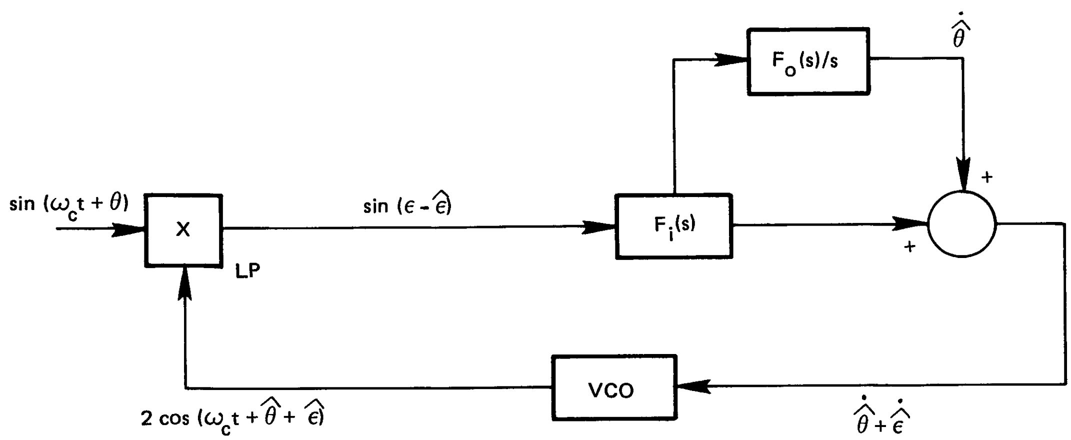


Fig. 2.4 <u>Compound PLL</u>

We prefer to use the augmented feedback realization of Figure 2.3 in analyzing the compound loop, since it may be compared to similar classical loops with different signal paths.

However, we will propose a third filter structure (Figure 2.9) for actually constructing our design. This final implementation replaces the inner-loop VCO with sine and cosine modulators and an integrator to provide an accurate $\hat{\varepsilon}$ as the output of the inner loop. The early implementations, especially that of Figure 2.4 may be practical for digital communication or finite-time estimation, where drift due to integrator offsets is small.

2.3 Brownian Motion Phase Example

2.3.1 Problem Statement

For a demonstration of our ideas, we consider the Brownian Motion phase process of section 1.4.3 and consider the design of a classic first-order PLL (as in section 1.4.3.1). When the carrier frequency is known and the noise strengths q and r are small, the PLL becomes the optimal filter, since the problem reduces to a linear, Gaussian one.

2.3.2 Acquisition Range for First-Order PLL

In all practical situations, however, ω_c is not known perfectly, and the PLL must first "acquire" the signal before the linear model becomes valid. We may examine the noise-free acquisition performance of a first-order loop by considering the error equations for

$$\theta = \Delta\omega t \tag{2.1}$$

where $\Delta\omega$ is the error in our knowledge of ω_c. The loop is shown in Fig. 2.5.

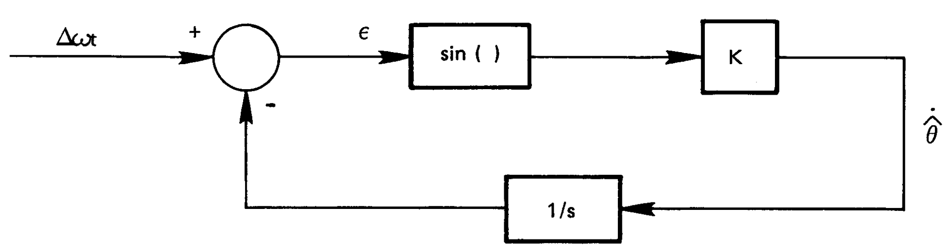


Fig. 2.5 Acquisition Model

The equation for $\dot{\varepsilon}$ becomes

$$\dot{\varepsilon} = \Delta\omega - K \sin(\varepsilon) \tag{2.2}$$

If $|\Delta\omega| < K$, then an equilibrium points exists where $\dot{\varepsilon} = 0$, and the loop will "lock". If $|\Delta\omega| > K$ however, the loop cannot acquire the input signal.

2.3.3 Second Order PLL

2.3.3.1 Design

PLL designers often use a second order loop because of its improved acquisition performance. The block diagram for such a loop is shown in Figure 2.6, where $F(s) = K(1 + a/s)$.

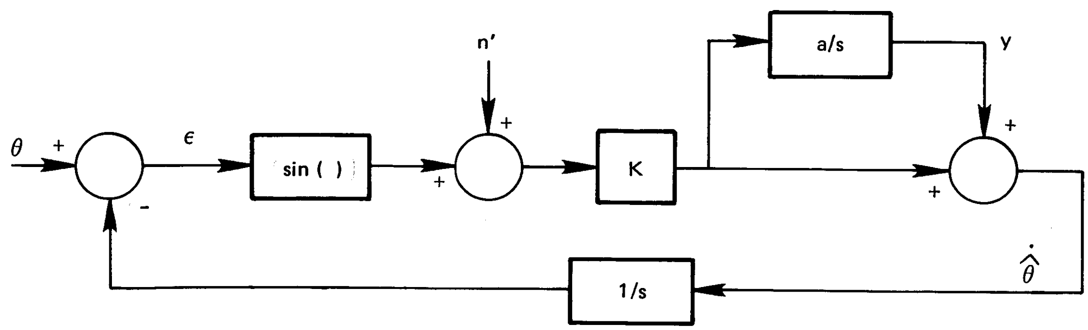


Fig. 2.6 Baseband Second Order PLL

The noise-free error equation for this loop is:

$$\ddot{\varepsilon} + K\left(\frac{d}{dt} + a\right) \sin \varepsilon = 0 \tag{2.3}$$

2.3.3.2 Acquisition Performance

This equation has not been solved, but Viterbi [41] shows that the (noise-free) acquisition range is infinite. Viterbi also develops an approximate formula for acquisition time (time until cycle-skipping stops) that is reasonable for low-noise environments:

$$t_L \approx \frac{1}{a} \quad \frac{\Delta\omega}{K} - \sin \Delta\theta)^2 \tag{2.4}$$

where

t_L = acquisition time (sec.)

$\Delta\theta$ = initial phase error (rad.)

$\Delta\omega$ = initial frequency error (rad/sec)

As $a \to 0$, the formula is accurate only for $|\Delta\omega|>K$; for $|\Delta\omega|<K$ in the first order loop, acquisition is immediate.

As Viterbi points out, although the frequency acquisition range is infinite, the acquisition time (proportional to $(\Delta\omega)^2$) may be prohibitively large. Furthermore, if the integration is imperfect (i.e. if $\frac{a}{s}$ is really $\frac{a}{s + \varepsilon'}$), the acquisition range is finite.

2.3.3.3 Noise Performance

We may examine the linear error equations of this second order system (with noise) to determine the penalty imposed by the added integrator (for a Brownian motion phase process). We let

$$x = \begin{pmatrix} \varepsilon \\ y \end{pmatrix} \tag{2.5}$$

then

$$\dot{x} = \begin{bmatrix} -K & -1 \\ aK & 0 \end{bmatrix} x + \begin{bmatrix} 1 & -K \\ 0 & aK \end{bmatrix} \begin{pmatrix} \dot{u} \\ \dot{n}' \end{pmatrix} \tag{2.6}$$

The steady-state phase error variance (see appendix B.1) becomes:

$$P_\theta = r(a+K) + q/2K \tag{2.7}$$

This equation may be minimized with respect to K by choosing

$$K° = \sqrt{\frac{q}{2r}} \tag{2.8}$$

which is the same gain as for the classic first order loop. We see also that the optimal "a" is 0 (which would produce a first order loop), and that P_θ increases linearly with "a" ("a" must be non-negative for a stable loop).

2.3.4 Compound PLL for Brownian Motion Phase Process

2.3.4.1 Design

We design the compound loop for this problem as follows. We choose a first order loop for the inner loop for convenience ($F_i(s) = K_1$). The outer loop is designed as an optimum linear filter for the phase process described and is thus also a first order loop. Recalling Figure 2.4, we may show the complete loop as in Figure 2.7 (where the single VCO input is not $\dot{\hat{\theta}}$ alone).

2.3.4.2 Acquisition Performance

We can associate this loop with the classic second order loop to determine its acquisition performance. The time until $\underline{e_i}$ stops skipping cycles is given by

$$t_L \approx \frac{1}{K_2} \left(\frac{\Delta\omega}{K_1} - \sin \Delta\theta\right)^2 \tag{2.9}$$

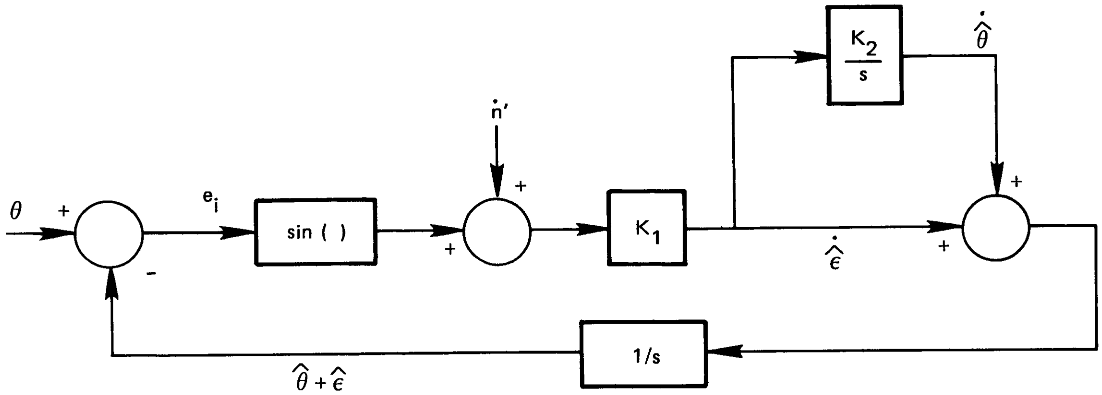


Figure 2.7 Compound PLL for Brownian Motion Phase Process

from equation 2.4. Since $K_1 >> K_2$ by design, as $K_2 \to 0$, this equation is valid only for $|\Delta\omega| > K_1$. For $|\Delta\omega| < K_1$, acquisition is instantaneous. As will be shown, the ability to relate acquisition performance to e_i (and not ε) is quite beneficial. We remark that t_L for the compound loop will refer to time until "linear operation," or frequency-lock. There will be a small additional delay while phase-lock is achieved by the inner loop, responding as a linear filter to the phase input. In regular PLL's, phase- and frequency-lock are nearly simultaneous for low-noise cases.

2.3.4.3 Noise Attenuation

We next examine the phase error variance of the compound loop. The linear equations are

$$\text{for } x = \begin{bmatrix} \varepsilon \\ \hat{\varepsilon} \end{bmatrix}, \quad \varepsilon = \theta - \hat{\theta}, \quad \dot{\theta} = \dot{u}$$

$$\dot{x} = \begin{bmatrix} 0 & -K_2 \\ K_1 & -K_1 \end{bmatrix} x + \begin{bmatrix} 1 & 0 \\ 0 & K_1 \end{bmatrix} \begin{bmatrix} \dot{u} \\ \dot{n}' \end{bmatrix} \tag{2.10}$$

From appendix B.2 we find the steady state variance to be

$$P_\theta = \frac{q}{2K_2} + rK_2 + \frac{q}{2K_1} \tag{2.11}$$

We see that P_θ is minimized for

$$K_2^\circ = \sqrt{\frac{q}{2r}} = K^\circ \tag{2.12}$$

$$K_1^\circ = \infty \tag{2.13}$$

Thus, K_1 is optimum when it is as large as possible. Here, the acquisition time is decreased by the same action that decreases the error variance. The classic loop has the opposite characteristics: the acquisition time can only be decreased by increasing the error variance.

2.3.4.4 Limitations

There are practical reasons, however, for not letting K_1 get too large. The noise threshold (η) of a PLL is that value of phase error variance (P_θ in the classic loop) at which the effects of the noise noticeably degrade performance over the predicted linear operation. Often, η is chosen as 0.25 rad^2. The important parameter in determining linear operation, however, is the variance of the signal fed to the sine non-linearity - ε in the classic loop but e_i in the compound case. We can find the linear predicted error variance of e_i by associating the compound loop variables with their classic counterparts in equation 2.7. We then have, in steady state:

$$P_{ei} = rK_1 + \frac{q}{2K_1} + rK_2 \tag{2.14}$$

We note that the K_1 that minimizes P_{ei} is

$$K_1' = \sqrt{\frac{q}{2r}} = K^{\circ} \tag{2.15}$$

Thus, if P_{θ}, with $K = K^{\circ}$ and "a" near zero in equation 2.7, is close to η, then P_{e_i} will be also, even with $K_1 = K_1'$, and no acquisition improvement is possible.

If the minimum phase error variance is much less than the threshold constraint ($P_{\theta} \ll \eta$), however, then we may choose a $K_1 \gg K^{\circ}$ that still results in $P_{e_i} < \eta$ (our criterion for linear performance). There exists a maximum K_1 that still results in $P_{e_i} \leq \eta$, where K_2 is chosen to minimize P_{θ} (i.e., $K_2 = K^{\circ}$). We can find this maximum K_1 from the larger root of:

$$K_1^2 + (K_2 - \frac{\eta}{r}) K_1 + \frac{q}{2r} = 0 \tag{2.16}$$

(which is a restatement of equation 2.14 with $P_{e_i} = \eta$).

As $P_{\theta_\ell} \to 0$, $P_{e_i} \to rK_1$, and $K_{1\,max}$ approaches:

$$K_{1_{max}} \to \frac{\eta}{r}$$

Thus, we see that η forms a ceiling for P_{e_i} and therefore for K_1. This means that the noise parameters of the problem (how far P_{θ_ℓ} is below η) impose a limit to the improvement that may be realized by using a compound loop.

2.3.5 Performance Comparison

2.3.5.1 Summary of Equations

We summarize the pertinent equations in Table 2.1 for $\Delta\theta = 0$ (in the acquisition equations).

Table 2.1 Second Order Filter Performance for Brownian Motion Phase Process

Classical Loop (2nd Order)	
Acquisition Time	$t_L \approx \frac{1}{a} \left(\frac{\Delta\omega}{K}\right)^2$
Phase error variance and Sine Noise Level	$P_\theta = r(a+K) + \frac{q}{2K}$
Compound Loop	
Acquisition Time	$t_L \approx \frac{1}{K_2} \left(\frac{\Delta\omega}{K_1}\right)^2$
Phase error variance	$P_\theta = rK^\circ + \frac{q}{2K^\circ} + \frac{q}{2K_1}$
Sine Noise Level	$P_{e_i} = rK_1 + \frac{q}{2K_1} + rK^\circ$
Minimum Error Variance	$P_{\theta_\ell} = \sqrt{2\,qr}$

2.3.5.2 Graph Explanation

In order to appreciate the improvement possible with the compound-loop technique, we plot the normalized acquisition time (without noise) versus the steady-state phase error variance (with noise) in Figure 2.8 for two values of $P_{\theta\ell}$.

For the classic loop, we first pick $K=K^{\circ}$ to minimize P_{θ} and then vary a/K. We see that the "improvement" in t_L is quite slow beyond $a/K = 1/2$, but that the penalty in phase error variance at this point is only 1 db. This value for a/K also represents a damping ratio of 0.7, and is used quite frequently in practice.

Next, using $a/K = 1/2$, we increase K above K° to see a much "faster" improvement in t_L versus P_{θ}.

For the compound loop, we choose $\eta = 0.1$ rad^2. as our linearity constraint, to be somewhat more conservative than the .25 mentioned earlier. We use $K_2 = K^{\circ}$ for minimum P_{θ}, and then select K_1 as large as possible such that $P_{e_i} \leq \eta$. This results in a very low t_L for a negligible increase in P_{θ} above $P_{\theta\ell}$. This is the furthest left point on the compound-loop curve. We also increased K_2 above K° and plotted t_L versus the resulting P_{θ}, but the improvement is slow, and $P_{ei} > \eta$ for this section of the curve. This technique is not recommended in general.

The normalization of $t_L/(\Delta\omega)^2$ by $(1/2r)^3$, if seemingly arbitrary, is done only to avoid the necessity of plotting different curves for different noise strengths. We regret the loss of physical "feel" that inevitably accompanies such normalization. We note that the asymptotes (for large P_{θ}) for the compound-loop curve and the classic-loop, constant a/K curve are the same for the two values of $P_{\theta\ell}$ shown. Thus, a rough estimate of the improvement possible for any $P_{\theta\ell}$ may be quickly obtained.

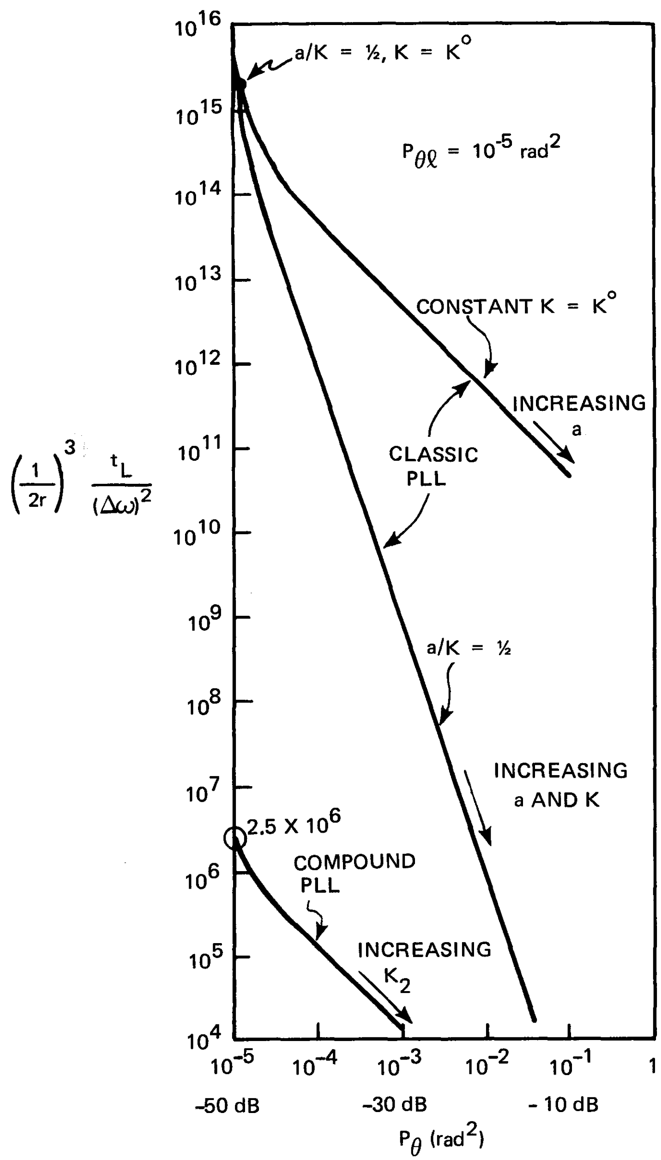
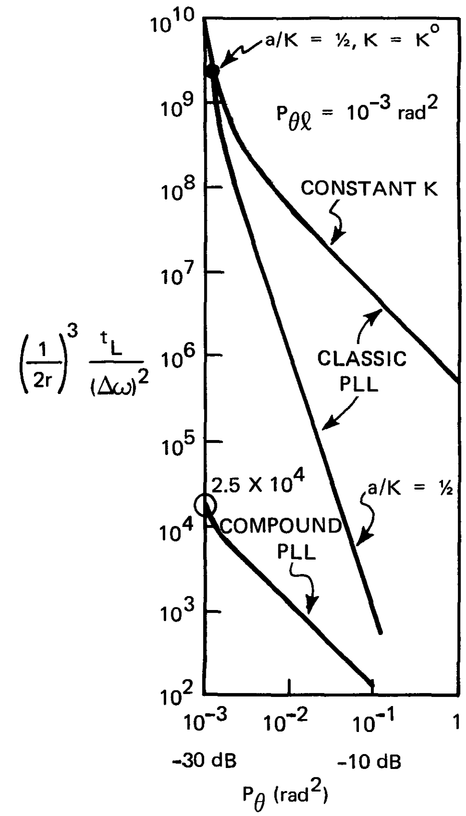


Figure 2.8 Normalized Acquisition Time vs. Phase Error Variance



2.3.5.3 Results

The results are very impressive for these low-noise cases ($P_{\theta\ell} \leq 10^{-3}$ rad^2.). Considering the minimum-P_θ points on the compound-loop curves, we see that for $P_{\theta\ell} = 10^{-5}$, the compound loop achieves a factor of 10^9 improvement in acquisition time over a classic loop of equal P_θ, and a 28db improvement in P_θ over a classic loop of equal t_L. At $P_{\theta\ell} = 10^{-3}$, the improvement margins are 10^5 and 15 db respectively.

One further improvement in the compound loop that is now shown on the graphs is the frequency range of "instantaneous acquisition" (no cycle-skipping). As $P_{\theta\ell} \to 0$, we may choose $K_1 \gg K_2$, and the tracking dynamics of the compound loop become those of a first order loop. Thus, for $|\Delta w| < K_1$, acquisition is essentially instantaneous. For the classic second order loop with $a \ll K$, or for the first order loop (with $a=0$), the range is $|\Delta w| < K$. Since $K \approx K_2$ (for similar noise filtering), and $K_1 \gg K_2$, the compound loop's quick acquisition range is much larger than that of the classic loop.

Thus, for a Brownian motion phase process and for low values of $P_{\theta\ell}$, the second-order compound PLL offers clear advantages in acquisition time and range over classic loops with similar output phase error variance.

2.4 General Technique

2.4.1 Generalization to Higher Order

The generalization of this technique to higher-order phase processes is straightforward (see Figure 2.4). We would, in general, advocate a first order "inner loop" ($F_i(s) = K_1$) for simplicity, and an outer loop designed as an optimal linear filter for the phase process. K_1 would then be made as large as possible such that the variance at the sine nonlinearity would be below some acceptable constraint level.

This would result in a loop of order "n+1" for an "nth" order problem, but this seems a small price to pay for the increased performance.

2.4.2 Alternate Implementation

We now construct another implementation of the compound loop that avoids the "perfect integrator" assumption. By using a normal integrator and sine and cosine modulators we can duplicate the function of the inner loop without using an extra integrator to obtain $\hat{\varepsilon}$. In this manner we have access to the $\hat{\varepsilon}$ that is fed back in the inner loop, removing the error caused by two different integrators (one realized by a VCO) producing two "$\hat{\varepsilon}$'s". This design is shown in Figure 2.9 and may be compared to Figure 2.1. We have used a first-order inner loop ($F_i(s) = K_1$) for simplicity.

The strengths of the noises shown are as follows:

$$E\{\dot{n}_I(t)\dot{n}_I(\tau)\} = E\{\dot{n}_Q(t)\dot{n}_Q(\tau)\} = 2r\,\delta(t-\tau) \tag{2.17}$$

$$E\{\dot{n}_3(t)\dot{n}_3(\tau)\} = 2r\,\delta(t-\tau) \tag{2.18}$$

The strength of $\dot{n}_3$ coincides with that used for the compound loop analysis in earlier sections of this chapter.

The signal "$\sin(\varepsilon-\hat{\varepsilon})$" follows from the identity

$$\sin(x-y) = (\sin x)\cos y - (\cos x)\sin y \tag{2.19}$$

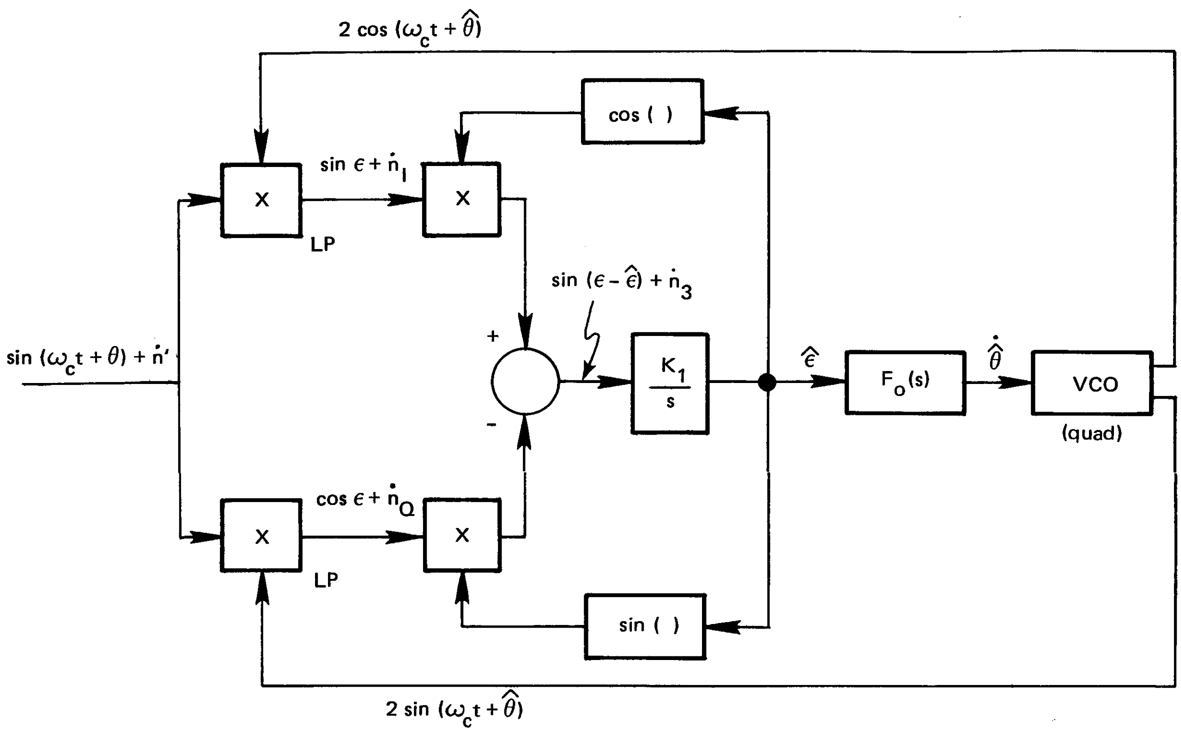


Figure 2.9 Alternate Compound Phase-Lock Loop

This allows us to avoid using an intermediate frequency, and also permits the subtraction of sinusoids to be carried out at baseband. We note that the sine and cosine modulators also operate at baseband.

2.4.3 VCO Replacement

The elimination of the inner-loop VCO is quite beneficial, and we are led to wonder if the outer-loop VCO may be similarly replaced. It does not seem advantageous at this time, for the following reasons. The VCO is used to transform the signal "$\dot{\hat{\theta}}$" into the signal "$2\cos(\omega_c t + \hat{\theta})$." We could add $\dot{\hat{\theta}}$ to ω_c, integrate the sum and then pass it through a cosine modulator, but this has two drawbacks. First, the integrator output is growing with time, and second, the cosine modulation must occur at ω_c.

The signal $2\cos(\omega_c t + \hat{\theta})$ also could be obtained from the identity

$$2\cos(\omega_c t + \hat{\theta}) = 2 \cos \omega_c t \cos \theta - 2\sin \omega_c t \sin \hat{\theta}$$

where $\sin \hat{\theta}$ and $\cos \hat{\theta}$ come from modulators and $\sin \omega_c t$ and $\cos \omega_c t$ from a VCO operating on ω_c. The principal drawback here is that the subtraction must be performed at carrier frequency which, like carrier-frequency modulation (eg. $\sin(\omega_c t + \theta)$), calls for more expensive components and more critical adjustments then similar operations at baseband.

We therefore conclude that it is impractical to replace the outer-loop VCO at this time. We may, however, replace the inner-loop VCO because we are operating on a baseband signal of finite range — the outer-loop phase error (ε).

2.4.4 More Inner Loops

The reader may wonder why, if one inner loop is so valuable, we don't add another, "inner-" inner loop to our designs. The reason is that it wouldn't help. Without noise, one inner loop could have "infinite" bandwidth, and no improvement in acquisition performance would be necessary (or possible). With noise, however, we are limited in the amount that we can open up the bandwidth of the inner loop. If we open the bandwidth up to our linearity constant, there will be no room for improvement by any inner-inner loop. Thus, one inner loop provides as much acquisition improvement as possible, with the least complexity.

2.5 Conclusion

2.5.1 Summary

We have shown that the acquisition performance of classic phase-lock loops may be greatly improved in low noise environments. The improvement may be achieved without penalizing the noise-attenuation properties of the PLL, and the amount of improvement increases as the minimum phase-error variance decreases.

2.5.2 Remarks

The name "compound PLL" was chosen for our original implementation concept of placing one PLL inside of another. This follows the terminology of Klapper and Frankle [24 ch8] who describe various combinations of FM detectors, FM feedback systems and PLL's inside of FM feedback loops (FMFB). These cascaded filters are distinguished from "multiple" loops which incorporate parallel internal filters. The section of compound loops does not, however, mention a PLL inside of a PLL.

Biswas and Banerjee [5] do consider such a design, but they augment the inner VCO by "injecting" the beat signal (at ω_{IF}). They mention, in passing, a "double phase-locked loop (DPLL)" that does not have this feature, but in their use the filters $F_i(s)$ and $F_o(s)$ are designed differently. In particular, they make no attempt to obtain a wider-bandwidth phase-error estimate $\hat{\varepsilon}$.

CHAPTER 3

THE REPRESENTATION THEOREM

3.1 Introduction

In this chapter we begin an investigation of the full nonlinear filtering problem discussed in Chapter 1. We are specifically interested in the threshold problem - the breakdown of filters based upon linearized analysis in regions of high noise. We will attempt to obtain workable approximations to the exact answer (the conditional density function) instead of exact solutions to the approximate problem. We will postpone the approximation stage of design from the problem to the solution.

What would be an optimal nonlinear filter? The conditional probability density function represents the complete solution to our problem. This density would allow one to compute both an estimate that minimized the expected value of any chosen cost function and the value of that minimum cost.

In the linear filtering problem, the conditional density is Gaussian, and we can determine the complete density by computing the conditional mean and covariance as functions of the measurement history, time, and the original density of the state. The Kalman filter does precisely that.

In nonlinear problems, however, the differential equation for the conditional mean depends on the conditional covariance, the covariance depends on the third moment, etc. [19]. The moment equations become an infinite chain and must be approximated. (The Kalman filter obeys the same equations, but the zero value of the third central moment of a

Gaussian density "breaks" the chain.)

But we want the conditional density, and the moments are only one way to express it. For problems "on the circle", where the state variable is a phase angle between $-\pi$ and π, the density is a periodic function, and the Fourier coefficients become a more useful set of statistics than the moments. Unfortunately, the differential equations for these variabiles are also infinitely coupled for the nonlinear measurement of interest in the PLL problem (see section 5.3.1).

There are expressions for the conditional density itself, however. The moment equations can be obtained from "Kushner's equation" [19], a partial differential equation (for the conditional density) that is similar to a Fokker Planck equation with a data-dependent forcing term. This equation is usually too complex to solve.

Kushner's equation can, in turn, be derived from another representation of the conditional density, an integral Bayes' rule type of formula. It is this expression that we will approximate.

3.2 Bucy's Representation Theorem

3.2.1 Motivation

We begin by deriving the basic formula, sometimes called "Bucy's Representation Theorem", which was first stated in 1965 by Bucy [7] and proven by Mortensen [32] (see Kailath [22] for a discussion of the development of the theorem). Our derivation will generally follow Wong [45], with a slightly different emphasis and notation.

The principal result of this chapter, the representation theorem, is not original, but we have two reasons for including it. First, this form of the conditional density is seldom noted by engineers, in part because of the difficulty in stating the theorem without recourse to measure-theoretic notions. We intend to offer a derivation of the theorem (and an explanation of the relevant mathematical concepts) that is straightforward and easy to understand.

Secondly, in our approximation method, we will use some mathematical operations that may seem strange to someone unfamiliar with the representation theorem, but otherwise interested in phase-lock loops. By deriving the theorem in this chapter and carefully defining the operations involved, we hope to make the justification for our approximation method (in the next chapter) more understandable.

We begin our derivation by introducing some notation and reformulating the problem. In general we follow Wong [45], with the most obvious difference being the interchange of x and z to conform to this author's conventions.

3.2.2 Notation

Let us consider a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ where Ω is a (non-empty) set of elements ω, $\mathcal{A}$ is a σ-algebra of subsets (A) of Ω, and $\mathcal{P}$ is a probability measure. We define a (real) random variable as a measurable mapping of $(\Omega,\mathcal{A})$ into $(R,\mathcal{R})$ where R is the real line and $\mathcal{R}$ is the Borel σ-algebra. If $\mathcal{P}_0$ is another probability measure on $(\Omega, \mathcal{A})$, we say that $\mathcal{P}_0$ is absolutely continuous, or differentiable, with respect to $\mathcal{P}$ ($\mathcal{P}_0 << \mathcal{P}$) if $\mathcal{P}(A) = 0$ implies that $\mathcal{P}_0(A) = 0$ for all A in $\mathcal{A}$. $\mathcal{P}$ and $\mathcal{P}_0$ are singular ($\mathcal{P} \perp \mathcal{P}_0$) if there exists an A such that $\mathcal{P}(A) = 0$ and

$P_0(\Omega - A) = 0$. We call P and P_0 equivalent ($P \equiv P_0$) if $P << P_0$ and $P_0 << P$.

If P is differentiable with respect to P_0, then the Radon-Nikodym theorem [45, p. 210] provides that there exists a unique A-measurement function Λ such that

$$P(A) = \int_A \Lambda(\omega) P_0(d\omega) \tag{3.1}$$

and we write

$$\Lambda = \frac{dP}{dP_0} \tag{3.2}$$

This Λ is called the Radon-Nikodym derivative of P with respect to P_0. The converse of the theorem (see Rudin [34], p. 23) allows us to define a measure P by specifying Λ and P_0.

If $\Omega = R$, $A = R$, and P is absolutely continuous with respect to the Lebesgue measure, then there exists a non-negative Borel function $p(x)$, $x \in R$, such that

$$P(A) = \int_A p(x)\,dx \qquad \text{for } A \in R \tag{3.3}$$

and $p(x)$ is called a probability density function.

We may write

$$\frac{dP}{dx} = p \tag{3.4}$$

This leads to an alternate notation for Λ. If P and P_0 are both absolutely continuous with respect to the Lebesgue measure, then

$$\Lambda = \frac{p}{p_0} \tag{3.5}$$

and Λ is called a likelihood ratio. This terminology derives from the use of density ratios in detection theory. Currently, however, the term "likelihood ratio" is used for many Radon-Nikodym derivatives that are unrelated to detection problems.

We denote the expectation of a random variable x by

$$Ex = \int_{\Omega} x(\omega) P(d\omega) \tag{3.6}$$

We call I_A and indicator function for A if

$$I_A(\omega) = 1 \quad \text{for} \quad \omega \in A$$
$$= 0 \quad \text{for} \quad \omega \notin A$$

3.2.3 Conditional Expectation

If $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$ ($\mathcal{B} \subset \mathcal{A}$), then we denote the conditional expectation of x with respect to $\mathcal{B}$ by

$$E^{\mathcal{B}}x \quad \text{or} \quad E(x|\mathcal{B})$$

and define it by the relations

a) $E^{\mathcal{B}}x$ is measurable with respect to $\mathcal{B}$ (3.7a)

b) $EI_A(E^{\mathcal{B}}x) = EI_A x \quad \forall A \in \mathcal{B}$ (3.7b)

Now we wish to show that the restriction of Λ to $\mathcal{B}$ is the conditional expectation (given $\mathcal{B}$) of Λ, i.e.,

$$\frac{dP^{\mathcal{B}}}{dP_0^{\mathcal{B}}} = E_0^{\mathcal{B}} \frac{dP}{dP_0} \tag{3.8}$$

or

$$\Lambda^{\mathcal{B}} = E_0^{\mathcal{B}} \Lambda \tag{3.9}$$

This follows from equation (3.7, b.) since

$$E_0 I_A x = E_0 I_A (E_0^{\mathcal{B}} x) \qquad A \varepsilon \mathcal{B} \tag{3.10}$$

then for all A in $\mathcal{B}$

$$P(A) = E_0 I_A \Lambda = E_0 I_A (E_0^{\mathcal{B}} \Lambda) \tag{3.11}$$

and by definition, since $P^{\mathcal{B}} << P_0^{\mathcal{B}}$

$$P^{\mathcal{B}}(A) = E_0 I_A \Lambda^{\mathcal{B}} \tag{3.12}$$

since

$$P^{\mathcal{B}}(A) = P(A) \qquad A \varepsilon \mathcal{B} \tag{3.13}$$

Equations 3.11 and 3.12 imply that

$$E_0 I_A \Lambda^{\mathcal{B}} = E_0 I_A (E_0^{\mathcal{B}} \Lambda) \qquad A \varepsilon \mathcal{B} \tag{3.14}$$

or simply that

$$\Lambda^{\mathcal{B}} = E_0^{\mathcal{B}} \Lambda \tag{3.15}$$

which is just equation 3.9.

We next want to demonstrate a very valuable result for conditional expectations and Radon-Nikodym derivatives:

$$E^{\mathcal{B}}x = \frac{E_0^{\mathcal{B}}\Lambda x}{E_0^{\mathcal{B}}\Lambda} \qquad (3.16)$$

For $A \in \mathcal{B}$, we have by definition (and equation 3.7 b.) that

$$EI_A x = E_0 I_A \Lambda x = E_0 I_A E_0^{\mathcal{B}} \Lambda x \qquad (3.17)$$

and also

$$EI_A x = EI_A \, E^{\mathcal{B}}x = E_0 I_A \Lambda^{\mathcal{B}}(E^{\mathcal{B}}x) \qquad (3.18)$$

So that equation 3.17 and 3.18 imply that

$$E_0 I_A E_0^{\mathcal{B}} \Lambda x = E_0 I_A \Lambda^{\mathcal{B}}(E^{\mathcal{B}}x) \qquad A \in \mathcal{B} \qquad (3.19)$$

or simply

$$E_0^{\mathcal{B}} \Lambda x = \Lambda^{\mathcal{B}} E^{\mathcal{B}} x \qquad (3.20)$$

which, with equation 3.9, is equivalent to equation 3.16.

Equations 3.9 and 3.16 will be most useful in what follows.

3.2.4 Stochastic Processes

We now introduce some notation for stochastic processes. We let x_t be a stochastic process, and sometimes write x_0^t for $\{x_s: 0 \le s \le t\}$. We also distinguish between the σ-algebras $\mathcal{A}_{xt}$ and $\mathcal{A}(x_t)$ by defining

$\mathcal{A}_{xt}$ = the smallest σ-algebra with respect to which x_0^t is measurable

$\mathcal{A}(x_t)$ = the smallest σ-algebra with respect to which x_t is measurable

In this chapter, we will be concerned with the time interval [0, 1], and we will denote $\mathcal{A}_{x1}$ by $\mathcal{A}_x$. If $\mathcal{A}$ and $\mathcal{B}$ are two σ-algebras of subsets of Ω, we refer to the smallest σ-algebra containing both $\mathcal{A}$ and $\mathcal{B}$ as $\underline{\mathcal{A} \vee \mathcal{B}}$.

3.2.5 Representation Theorem

3.2.5.1 Problem Statement

We now are ready to consider a general nonlinear filtering problem with additive Gaussian measurement noise. We consider two vector-valued (nx1) stochastic processes x_0^t and v_0^t on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$. We assume that, under $\mathcal{P}$, x_0^t and v_0^t are independent, x_0^t is a Markov process, the components of v_0^t are independent standard Brownian motions, and

$$\int_0^1 (x_t)^T x_t dt < \infty \qquad \text{(with probability 1)} \tag{3.21}$$

We also define the process

$$z_t = \int_0^t x_s ds + v_t \tag{3.22}$$

Here we are dealing with a non-Guassian x_t, but later we will reformulate our results to conform to our earlier notation, where we have a nonlinear transformation h of a Gaussian x_t. We may consider x_t the signal, v_t the measurement noise, and z_t the measurement. The assumptions are those usually satisfied in filtering problems. The integral constraint on x_t corresponds to a finite-energy requirement, and cannot be dispensed with, while the independence of the components of u_t can usually be relaxed.

We define the σ-algebra $\mathcal{A}_t$ by

$$\mathcal{A}_t = \mathcal{A}_{st} \vee \mathcal{A}_{zt}$$

and we let $\mathcal{A} = \mathcal{A}_1$ be the algebra of the probability space $(\Omega, \mathcal{A}, \mathcal{P})$. We will want the expectation of some function of x_t conditioned on all of the measurements up until time t. Thus we want to evaluate

$$E[f(x_t)|z_0^t] = E^{\mathcal{A}_{zt}} f(x_t)$$

As a first step, we construct a new measure $\mathcal{P}_0$ by defining the Radon-Nikodym derivative $d\mathcal{P}_0/d\mathcal{P}$. Under this $\mathcal{P}_0$, x_0^t and z_0^t will be independent, and z_0^t will be a standard Brownian motion. This is a consequence of Girsanov's theorem (see Wong [45] page 228), and is quite important in what follows.

3.2.5.2 $\underline{\mathcal{P}_0 \text{ Construction}}$

We define $\mathcal{P}_0$ on $(\Omega, \mathcal{A})$ by constructing the Radon Nikodym derivative:

$$\frac{d\mathcal{P}_0}{d\mathcal{P}} = \exp\{-\int_0^1 x_t^T\, dv_t - \frac{1}{2}\int_0^1 x_t^T x_t dt\} \tag{3.23}$$

where the first integral is an Ito integral whose existence is guaranteed by constraint 3.21 (see McKean [31]). We claim that $\mathcal{P}_0$ has the following properties

- $\mathcal{P}_0$ is a probability measure

- under $\mathcal{P}_0$, z_0^t has independent Brownian motions for components

- under $\mathcal{P}_0$, x_0^t and z_0^t are independent

- The restriction of $\mathcal{P}_0$ to $\mathcal{A}_x$ is the same as the restriction of $\mathcal{P}$ to $\mathcal{A}_x$.

The proof of these claims is detailed in Wong [45, proposition 5.1, p. 232]. The last point should be stressed: $\mathcal{P}_0^{\mathcal{A}_x} = \mathcal{P}^{\mathcal{A}_x}$. Thus the density for x_t (assuming one exists) is the same under $\mathcal{P}$ or $\mathcal{P}_0$. We will make use of this in our approximation method.

To complete Wong's proposition 5.1, we further claim that

- $\mathcal{P} << \mathcal{P}_0$

- $$\Lambda = \frac{d\mathcal{P}}{d\mathcal{P}_0} = \exp\left\{ \int_0^1 x_t dz_t - \frac{1}{2} \int_0^1 x_t^T x_t dt \right\} \tag{3.24}$$

In equation 3.24 the first integral is an Ito integral (under $\mathcal{P}_0$). We see that $\mathcal{P}$ and $\mathcal{P}_0$ are equivalent measures, and thus we can, using Λ or Λ^{-1}, obtain $\mathcal{P}$ from $\mathcal{P}_0$ and vice versa, or more importantly:

$$E[f(x_t)] = E_0[\Lambda f(x_t)]$$

We now define

$$\Lambda_t = E_0^{\mathcal{A}_t} \Lambda \tag{3.25}$$

and, as Wong shows

$$\Lambda_t = \exp\left\{ \int_0^t x_s dz_s - \frac{1}{2} \int_0^t x_s^T x_s ds \right\} \tag{3.26}$$

where the difference between Λ_t and Λ is only in the upper limit of the integrals. We see that equations 3.25 and 3.9 imply that

$$\Lambda^{A_t} = \frac{dP^{A_t}}{dP_0^{A_t}} = \Lambda_t \tag{3.27}$$

3.2.5.3 Conditional-Density Representation

We are now ready to provide an "answer" to our filtering problem. Using equations 3.27 and 3.16 we may write

$$E^{A_{zt}} f(x_t) = \frac{E_0^{A_{zt}}[\Lambda_t f(x_t)]}{E_0^{A_{zt}}\Lambda_t} \tag{3.28}$$

since $A_{zt} < A_t$.

We want to rewrite equation 3.28 in such a way that we may infer the conditional density from it. To do this, we note that, from equation 3.7 b.

$$E_0^{A_{zt}}(\Lambda_t f(x_t)) = E_0^{A_{zt}}[E_0^{A_{zt} \vee A(x_t)} \Lambda_t f(x_t)] \tag{3.29}$$

which, since $f(x_t)$ is measurable with respect to $A_{zt} \vee A(x_t)$, becomes

$$E_0^{A_{zt}}(\Lambda_t f(x_t)) = E_0^{A_{zt}}[f(x_t) E_0^{A_{zt} \vee A(x_t)} \Lambda_t] \tag{3.30}$$

For future reference, we identify

$$B = A_{zt} \vee A(x_t) \tag{3.31}$$

and

$$U_t(x_t,\ z_0^t) = E_0^{\mathcal{B}}\ \Lambda_t \tag{3.32}$$

where Λ_t is given by equation 3.26.

We now note that for any random variable y that is $\mathcal{A}(x_t)$-measurable

$$E_0^{\mathcal{A}_{zt}} y = \int_\Omega y\ d\mathcal{P}_0^{\mathcal{A}_{zt}} \tag{3.33}$$

But since $\mathcal{P}_0$ is restricted to $\mathcal{A}_{zt}$ (and using R^n for Euclidian n-space),

$$E_0^{\mathcal{A}_{zt}}\ y = \int_{R^n} y\ \mathcal{P}_0^{\mathcal{A}_{zt}}(dx_t) \tag{3.34}$$

which becomes

$$E_0^{\mathcal{A}_{zt}} y = \int_{R^n} y\ \mathcal{P}(dx_t) \tag{3.35}$$

since x_t is independent of z_0^t under $\mathcal{P}_0$, and $\mathcal{P}_0^{\mathcal{A}_{xt}} = \mathcal{P}^{\mathcal{A}_{xt}}$. Now if $\mathcal{P}$ is absolutely continuous with respect to the Lebesgue measure, x_t has a density denoted by

$$\frac{d\mathcal{P}}{dx} = p(x,t)$$

and

$$E_0^{\mathcal{A}_{zt}} y = \int_{R^n} y\ p(x,t)\,dx \tag{3.36}$$

Thus, equation 3.28 may be rewritten

$$E^{\mathcal{A}_{zt}} f(x_t) = \frac{\int_{R^n} f(x_t) U_t(x_t,z_0^t)\ p(x,t)\,dx}{\int_{R^n} U_t(x_t,z_0^t)\ p(x,t)\,dx} \tag{3.37}$$

The usual expression for the conditional expectation would be

$$\hat{E}^{z^t} f(x_t) = \int_{R^n} f(x)\, p(x,t|z_0^t)\, dx \tag{3.38}$$

Thus, by comparing equations 3.37 and 3.38, we may infer that the conditional density is

$$p(x,t|z_0^t) = \frac{U_t(x_t,\, z_0^t)\; p(x,t)}{\int_{R^n} U_t(x_t, z_0^t)\; p(x,t)\, dx} \tag{3.39}$$

where U_t is given by equation 3.32.

We note that since p(x,t) is the a priori density for x_t, equation 3.39 is a type of Bayes' rule, providing the ratio between the conditional and a priori densities for x_t. In this context, Λ_t (equation 3.26) is a type of "density" for the process z_0^t conditioned on the process x_0^t. The expectation in U_t is over x_0^t conditioned on x_t and with z_0^t fixed. Jazwinski [19] provides an intuitive argument along these lines.

3.2.6 Properties of the Conditional Density

3.2.6.1 Denominator

We want to examine some of the properties of this representation of the conditional density (equation 3.39). The denominator in the expression is

$$W_t = E_0^{\mathcal{A}z^t} \Lambda_t = \int_{R^n} E_0^{\mathcal{B}} \Lambda_t\, p(x,t)\, dx \tag{3.40}$$

First, we note that

$$d\Lambda_t = \Lambda_t x_t^T dz_t \quad , \qquad \Lambda_0 = 1 \tag{3.41}$$

or

$$\Lambda_t = 1 + \int_0^t \Lambda_s x_s^T dz_s \tag{3.42}$$

Thus

$$W_t = 1 + E_0^{\mathcal{A}_{zt}} \int_0^t \Lambda_s x_s^T dz_s \tag{3.43}$$

Now, we want to bring the expectation inside the Ito integral above. Mortensen [32] first developed a "Fubini" theorem for Ito integrals, and Marcus [30] proves a Fubini theorem for conditional expectations. A combination of these results allows us to write

$$W_t = 1 + \int_0^t E_0^{\mathcal{A}_{zt}} (\Lambda_s x_s^T) dz_s \tag{3.44}$$

Since $\Lambda_s x_s^T$ is $\mathcal{A}_s$-measurable, and since z_s^t is independent of $\mathcal{A}_s$ [see Wong [45], p. 300], we have

$$W_t = 1 + \int_0^t E_0^{\mathcal{A}_{zs}} (\Lambda_s x_s^T) dz_s \tag{3.45}$$

Now since $\mathcal{A}_{zs} \subset \mathcal{A}_s$, equation 3.16 implies that

$$E_0^{\mathcal{A}_{zs}} \Lambda_s x_s^T = (E_0^{\mathcal{A}_{zs}} \Lambda_s)(E^{\mathcal{A}_{zs}} x_s^T) \tag{3.46}$$

By defining

$$\hat{x}_s = E^{\mathcal{A}_{zs}} x_s \tag{3.47}$$

we see that

$$W_t = 1 + \int_0^t W_s \hat{x}_s^T dz_s \tag{3.48}$$

which is solved by

$$W_t = \exp \{ \int_0^t \hat{x}_s^T dz_s - \frac{1}{2} \int_0^t \hat{x}_s^T \hat{x}_s ds \} \tag{3.49}$$

Unfortunately, this representation of the denominator is not helpful in trying to estimate $\hat{x}_t$. The expression (3.49) finds its greatest use in detection theory, where it provides the hint of using the best estimate of an unknown signal (in the "unknown signal" problem) in precisely the same place in the likelihood-ratio as the known signal (in the "known signal" problem), as shown by Kailath [20].

It is an interesting result for filtering also, and we have included it for two reasons. The first is that the type of manipulation performed in order to bring the expectation inside the integral in equation 3.44 is the same as that needed to use our approximation method.

The second reason for including this derivation is that it demonstrates that the denominator produces an $\hat{x}_s$ term, which appears in Kushner's equation for the conditional density [19]. This makes Kushner's equation a partial integro-differential equation with a stochastic driving term. We will shortly discuss a less complex partial differential equation (first derived by Mortensen [32]), which describes the propagation of the numerator of the density.

3.2.6.2 Nonlinear-Measurement Formulation

Before that, however, we wish to return to the notation of Chapter 1, and formulate the equivalent of equation 3.39. We assume that

$$dx = Fx\, dt + G\, du \tag{3.50}$$

$$dz = h(x)dt + dn \tag{3.51}$$

We also require that

$$\int_0^t h^T(x_s) R^{-1} h(x_s) ds < \infty \tag{3.52}$$

with probability 1 for all finine t. Then the conditional density for x_t is (see Jazwinski [19]).

$$p(x,t|z_0^t) = \frac{(E_0^{\mathcal{B}} \Lambda_t)\ p(x,t)}{E_0^{\mathcal{A}zt}\ \Lambda_t} \tag{3.53}$$

where

$$\mathcal{B} = \mathcal{A}_{zt}\ v\ \mathcal{A}(x_t) \tag{3.54}$$

and

$$\Lambda_t = \exp\{\int_0^t h^T(x_s)R^{-1}dz_s - \frac{1}{2}\int_0^t h^T(x_s)R^{-1}h(x_s)ds\} \tag{3.55}$$

We note that the conditioning in $\mathcal{B}$ is on the <u>state</u> x_t and not the measurement $h(x_t)$.

In the expressions for the conditional density (equations 3.39 and 3.53), the only term that we need to compute from the data is U_t.

$$U_t(x_t,\ z_0^t) = E_0^{\mathcal{B}}\ \Lambda_t \tag{3.56}$$

Since we know the a priori density for x_t from the dynamics (3.50), and since the denominator W_t (in 3.57)

$$W_t = \int_{R^u} U_t(x, z_0^t)\, p(x, t)dx \tag{3.57}$$

can be computed from knowledge of U_t and $p(x,t)$, U_t represents the "new information" in the measurements. It is U_t which changes the shape of $p(x, t|z_0 t)$ away from that of $p(x,t)$. It is U_t, therefore, that we will approximate in the next chapter. Before then, we want to consider the possibility of exact solutions for the conditional density. To do this, we first examine differential forms for $p(x,t|z_0^t)$ and $U_t \cdot p(x,t)$.

3.2.6.3 Differential Density Forms

The integral expressions 3.39 and 3.53 appear to offer little hope of exact solution. Even in the linear, Gaussian case, when the Ito integral

$$\int_0^t x_s^T dz_s$$

is Gaussian under $P_0^{\mathcal{B}}$, the second guadratic term in the exponent of Λ_t,

$$-\frac{1}{2}\int_0^t x_s^T x_s ds$$

is unknown.

Since many control engineers prefer differential forms for filters, one is motivated to exanime differential forms for the conditional density in the hope that they will appear easier to solve. Kushner's equation is such a form [19]

$$dp = L(p)dt + (h_t-\hat{h}_t)^T R^{-1}(dz_t-\hat{h}_t dt)p \tag{3.58}$$

where

$$h_t = h(x_t)$$

$$P = p(x,t|z_0^t)$$

and

$$\hat{h}_t = \hat{E}^{zt} h_t$$

Here, $L(p)$ is the Fokker-Planck operator associated with the dynamics (eq. 3.50); that is, the a priori density for x_t satisfies

$$\frac{\partial p(x,t)}{\partial t} = L(p(x,t)) \tag{3.59}$$

Along with the general difficulties of solving a "forced" Fokker-Planck equation, Kushner's equation contains $\hat{h}_t$ terms which are <u>integrals</u> over the density $p(x,t|z_0^t)$. It is fortunate, but not widely recognized, that these $\hat{h}_t$ terms come from the normalizing denominator W_t, and that Kushner's equation can be simplified.

If we define

$$V(x,t|z_0^t) = U_t(x_t, z_0^t)\; p(x,t) \tag{3.60}$$

then

$$p(x,t \mid z_0^t) = \frac{V(x,t;\; z_0^t)}{W_t} \tag{3.61}$$

where

$$W_t = \int_{R^n} V(x,t;\; z_0^t)\,dx \tag{3.62}$$

Mortensen [32] showed that (see also Wong [45])

$$dV = L(V)\,dt + Vh_t^T R^{-1} dz_t \tag{3.63}$$

This equation (3.63) is somewhat easier to analyze than (3.58), but clearly no general solution is available. We note that equation 3.48 implies that

$$dW = W\hat{h}_t^T R^{-1} dz \tag{3.64}$$

Using equations 3.63 and 3.64, we may take the Ito derivative of the ratio V/W to obtain (see Jazwinski [19], p. 115)

$$\begin{aligned} d\left(\frac{V}{W}\right) = \frac{V}{W} h^T R^{-1} dz + \frac{1}{W} L(V)\,dt - \frac{V}{W} \hat{h}_t^T R^{-1} dz_t \\ - \frac{V}{W} h_t^T R^{-1} \hat{h}_t dt + \frac{V}{W} \hat{h}_t^T R^{-1} \hat{h}_t dt \end{aligned} \tag{3.65}$$

Since P = V/W and since W is not a function of x and may be taken inside of $L(\ )$, equation 3.65 reduces to

$$dp = L(p)\,dt + (h_t - \hat{h}_t)^T R^{-1} (dz_t - \hat{h}_t dt) \tag{3.66}$$

which is equation 3.58.

Thus Mortensen's equation is consistent with that of Kushner, while being somewhat easier to analyze. The only general nonlinear case where a solution is available, however, is when the process noise strength (Q) goes to zero. We consider this next.

3.3 No Process Noise

3.3.1 Linear-Measurement Problem

Bucy and Joseph [9, p.51] first pointed out that when the process driving noise went to zero and the state at time "s" became a measurable function of the state at time "t", the representation theorem provided an explicit formula for the conditional density. This result will be used to show that our approximation method converges as the process noise strength goes to zero. In this section we wish to demonstrate the phenomenon for two simple problems.

We recall that (from equation 3.32)

$$U_t(x_t, z_0^t) = E_0^{A_{zt} \vee A(x_t)} \Lambda_t \tag{3.67}$$

Now if $h(x_s)$ is a measurable function of x_t, then Λ_t is a measurable function of $A_{zt} \vee A(x_t)$, and

$$U_t(x_t, z_0^t) = \Lambda_t \tag{3.68}$$

To demonstrate this result, we consider the scalar linear system

$$\begin{aligned} dx &= 0 \\ dz &= x\, dt + dn \\ p(x_0) &\sim N(0, Q) \end{aligned} \tag{3.69}$$

Then from equation 3.68 and 3.55,

$$p(x,t|\, z_0^t) = \frac{\Lambda_t p(x,t)}{\int_{-\infty}^{\infty} \Lambda_t p(x,t)\, dx} \tag{3.70}$$

where

$$\Lambda_t = \exp \{\frac{1}{R} \int_0^t x\, dz_\tau - \frac{1}{2R} \int_0^t x^2 d\tau\} \tag{3.71}$$

so that

$$p(x,t|z_0^t) = \frac{1}{\sqrt{2\pi P_t}} e^{-(x-\hat{x}_t)^2/2P_t} \quad (3.72)$$

where

$$\hat{x}_t = \frac{P_t}{R} \int_0^t dz_\tau \quad (3.73)$$

$$P_t = \frac{QR}{R + tQ} \quad (3.74)$$

By taking derivatives of $\hat{x}_t$ and P_t we arrive at the more familiar relationships:

$$\dot{\hat{x}}_t = \frac{P_t}{R} (\dot{z}_t - \hat{x}_t) \quad (3.75)$$

and

$$\dot{P}_t = -\frac{P_t^2}{R} \quad (3.76)$$

Equations 3.75 and 3.76 are the Kalman Filter equations for this problem.

3.3.2 Phase-Measurement Problem

A second example, of more interest to us, concerns the system

$$d\theta = 0$$

$$dz = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} dt + \begin{pmatrix} dn_1 \\ dn_2 \end{pmatrix} \quad (3.77)$$

from Chapter I, with

$$p(\theta) = \frac{1}{2\pi} \qquad -\pi \leq \theta \leq \pi \quad (3.78)$$

We have, from 3.68 and 3.55, that

$$p(\theta,t|z_0^t) = \frac{\frac{1}{2\pi} \Lambda_t}{\frac{1}{2\pi} \int_{-\pi}^{\pi} \Lambda_t d\theta} \quad (3.79)$$

where

$$\Lambda_t = \exp\{\frac{1}{2r}\int_0^t (\sin\theta\, dz_1 + \cos\theta\, dz_2) - \frac{1}{4r}\int_0^t (\sin^2\theta + \cos^2\theta)\,d\tau\} \qquad (3.80)$$

or

$$\Lambda_t = \exp\{x_t \sin\theta_t + y_t \cos\theta_t - \frac{1}{4r}\, t\} \qquad (3.81)$$

where

$$x_t = \frac{1}{2r}\int_0^t dz_1, \qquad y_t = \frac{1}{2r}\int_0^t dz_2 \qquad (3.82)$$

Thus

$$p(\theta,t|z_0^t) = \frac{e^{x_t \sin\theta + y_t \cos\theta}}{2\pi I_0\left(\sqrt{x_t^2 + y_t^2}\right)} \qquad (3.83)$$

or

$$p(\theta,t\; z_0^t) = \frac{e^{\alpha_t \cos(\theta-\beta_t)}}{2\pi I_0(\alpha_t)} \qquad (3.84)$$

where

$$\alpha_t = \sqrt{x_t^2 + y_t^2}, \qquad \beta_t = \tan^{-1}\left(x_t/y_t\right) \qquad (3.85)$$

This result was first noted by Kailath [21], who derived it from the likelihood ratio for the detection problem. Mallinckrodt, et. al. [29]

later rederived it, although without reference to the Representation Theorem.

This density form is the same as that of the 1st-order PLL error density (equation 1.61), and seems somewhat "natural" for this problem. (See also J. T.-H. Lo's recent paper [28].) The densities, however, represent two very different results. The PLL error density comes from a steady-state analysis of the error in a chosen filter structure - the PLL. The density above (equation 3.84), however, is the data-dependent conditional density of the state, when there is no process noise, and may be used to obtain the optimal filter structure.

We want to demonstrate that this filter approaches the regular PLL design for this problem for large t. We note that

$$\frac{x_t}{t} = \frac{\sin\theta}{2r} + \frac{1}{2rt}\int_0^t dn_1$$

where the mean of the second term is zero and its variance is

$$\frac{1}{2rt} \to 0$$

as $t \to \infty$. Thus

$$\frac{x_t}{t} \to \frac{\sin\theta}{2r}$$

and similarly

$$\frac{y_t}{t} \to \frac{\cos\theta}{2r}$$

as $t \to \infty$, so that

$$\beta \to \theta \tag{3.86}$$

$$\alpha \to t/2r \tag{3.87}$$

We next need to obtain the differential equations for α_t and β_t. Taking the Ito derivatives of α and β as functions of x_t and y_t we have

$$d\alpha = \frac{1}{4r\alpha}\, dt + \frac{1}{2r} \begin{pmatrix} \sin\beta \\ \cos\beta \end{pmatrix}^T \begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \tag{3.88}$$

$$d\beta = \frac{1}{2r\alpha} \begin{pmatrix} \cos\beta \\ -\sin\beta \end{pmatrix}^T \begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \tag{3.89}$$

But letting $\beta = \tilde{\theta}$ in the "in-phase" and "quadrature" baseband signals (equation 1.21), we have

$$d\alpha = \frac{1}{4r\alpha}\, dt + \frac{1}{2r}\, dz_Q \tag{3.90}$$

$$d\beta = \frac{1}{2r\alpha}\, dz_I \tag{3.91}$$

We recognize that this is just a phase-lock loop with a data dependent gain $K(z_0^t)$,

$$K(z_0^t) = \frac{1}{2r\alpha_t} \tag{3.92}$$

and for large time t,

$$K(z_0^t) \to 1/t \tag{3.93}$$

It is interesting to examine a PLL design for this problem. Given that q=0, we have that

$$\dot{P} = - P^2/2r \quad \text{with } P(0) = \infty \tag{3.94}$$

or

$$P = 2r/t \tag{3.95}$$

so that

$$K = \frac{P}{2r} = \frac{1}{t} \tag{3.96}$$

Thus, the optimal filter approaches the classic PLL for large t, but outperforms the PLL for small t by using a data-dependent, rather than just time-programmed, gain. This is to be expected in nonlinear filters, since the nonlinear measurements provide some information about the value of a given measurement history that linear measurements do not.

We now want to examine the differential equation for the Fourier coefficients of the conditional density. We consider the expansion

$$p(\theta|z_0^t) = \frac{1}{2\pi}[1 + 2\sum_{n=1}^{\infty} (a_n \sin n\theta + b_n \cos n\theta)] \tag{3.97}$$

Using Kushner's equation (3.58) with $L(p) = 0$ and

$$E\left[\begin{pmatrix}\sin n\theta \\ \cos n\theta\end{pmatrix} \Big| z_0^t\right] = \begin{pmatrix} a_n \\ b_n \end{pmatrix} \tag{3.98}$$

we have

$$d\begin{pmatrix} a_n \\ b_n \end{pmatrix} = \frac{1}{2r} H_n \begin{pmatrix} dz_1 - a_1\, dt \\ dz_2 - b_1\, dt \end{pmatrix} \tag{3.99}$$

where

$$H_n = \begin{bmatrix} \dfrac{b_{n-1} - b_{n+1}}{2} - a_1 a_n & \dfrac{a_{n-1} + a_{n+1}}{2} - a_n b_1 \\ \dfrac{a_{n+1} - a_{n-1}}{2} - a_1 b_n & \dfrac{b_{n-1} + b_{n+1}}{2} - b_1 b_n \end{bmatrix} \quad (3.100)$$

For our density, moreover, we know that

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = \frac{I_n(\alpha)}{I_0(\alpha)} \begin{pmatrix} \sin n\beta \\ \cos n\beta \end{pmatrix} \quad (3.101)$$

as functions of α and β, satisfy equation 3.99. This will prove useful in computing sub-optimal filters when $q \neq 0$.

3.4 Summary

In this chapter we have laid the ground work for the rest of this thesis. We have derived an expression for the conditional density that we will approximate in the next chapter. We have used this expression to solve the phase-lock loop problem for a constant-phase system and have shown how the optimal filter differs from a phase-lock loop. Let us now proceed with the general approximation method.

CHAPTER 4

APPROXIMATION METHOD

4.1 General Approach

4.1.1 Introduction

In this chapter we will develop an approximation to the conditional density function (equation 3.53) which we will later use to develop sub-optimal filters for the phase-tracking problem. The approximation method is quite general, however, and may be used in other nonlinear filtering problems with additive Gaussian measurement noise. We will therefore retain the general problem format in this chapter.

We recall that (equation 3.53)

$$p(x,t|z_0^t) = \frac{U_t(x_t, z_0^t)\ p(x,t)}{W_t} \tag{4.1}$$

where

$$W_t = \int_{R^n} u_t(x,z_0^t)\ p(x,t)\,dx \tag{4.2}$$

$$U_t = E_0^{\mathcal{B}}\ e^{\zeta_t} \tag{4.3}$$

$$\mathcal{B} = A_{zt} \vee A(x_t) \tag{4.4}$$

$$\zeta_t = \int_0^t h_s^T R^{-1} dz_s - \frac{1}{2} \int_0^t h_s^T R^{-1} h_s\, ds \tag{4.5}$$

$$h_s = h(x_s) \tag{4.6}$$

We remark that $p(x,t)$ is the (known) a priori density for x_t, and that W_t is a normalization parameter which is independent of x_t. Intuitively, we see that if we develop an approximation $\tilde{U}_t$ to U_t and compute a normalization $\tilde{W}_t$ from $\tilde{U}_t$ and $p(x,t)$, then the approximate density

$$\tilde{p}(x,t|z_0^t) = \frac{\tilde{U}_t p(x,t)}{\tilde{W}_t} \tag{4.7}$$

should converge to the real density as $\tilde{U}_t \to U_t$. We will discuss convergence more thoroughly in the next section.

The approximations that we will propose for U_t are motivated by the fact that U_t (equation 4.3) is an expectation of an exponential of a random variable (ζ_t). Thus U_t is a "moment generating function" [23] for the conditional density

$$p_0(\zeta_t|z_0^t, x_t)$$

While this density for ζ_t is, in general, nearly impossible to find, any given <u>moment</u> of ζ_t is straight-forward (if tedious) to compute analytically as a function of the transition density for x_t and the measurements z_0^t.

Our technique will involve approximating the moment generating function with finite sums of moments, or exponentials of finite sums of cumulants, of ζ_t. (We stress that these are not the moments and cumulants of x_t conditioned on z_0^t, but rather the moments and cumulants of ζ_t conditioned on x_t and z_0^t under the P_0 measure.) We then approximate the conditional density for x_t by dividing our approximate numerator by the normalization from the integral of the approximate numerator.

4.1.2 <u>Convergence of Density Approximations</u>

Before describing our approximation method, we want to discuss density approximations, and convergence, in general. The numerical density approximation of Bucy and Senne [11] results in approximate density values at a finite number of points (x_i) in the state space for a given sample path z_0^t. Our method, however, like the Gaussian sum approach of [2] and [36], will develop a continuous <u>function</u> of x at time t for a given sample path z_0^t. Thus, like the actual condtional density, our approximations will be functions of x, t, and, through z_0^t, a sample point (ω) in the probability space.

We will approximate $U_t(x_t, z_0^t)$ in equation 4.3 by a series of functions $\tilde{U}_{tn}$ which will converge <u>pointwise</u> in x, t, and ω. We will then construct an approximate numerator

$$V_n = \tilde{U}_{tn}\, p(x, t) \tag{4.8}$$

and denominator

$$W_n = \int_{R^K} V_n(x,t,\omega)\,dx \tag{4.9}$$

which will define the approximate density

$$p_n(x,t,\omega) = \frac{V_n(x,t,\omega)}{W_n(t,\omega)} \tag{4.10}$$

For one of our approximation methods, we also will be able to demonstrate pointwise (in t and ω) convergence of the denominator W_n. This, along with the numerator convergence, may be shown to imply pointwise convergence of p_n in equation 4.10. (See Rudin [33], page 43, where the convergence of the ratio of two convergent sequences* follows from Theorem 3.3 c. and d.).

Unfortunately, we have been unable to prove the uniform (in x) convergence of our numerator or density approximations, in part because of the infinite domain (R^n) for x. Our approximate densities, however, like the exact density, are differentiable (in x and t) and therefore continuous, and, since we do not expect to encounter pathological cases, stronger convergence (e.g. L^2 in R^n for x) may in fact be provable. The well-behaved nature of the (Gaussian) a priori density p(x,t) also may help in demonstrating stronger results.

The pointwise convergence in ω implies (see Wong [45] p. 20 or Egoroff's theorem in Rudin [34] p. 72) convergence in probability ($\mathcal{P}$ or $\mathcal{P}_0$). We have, in general, been unable to demonstrate stronger convergence (e.g. quadratic mean) in ω for the numerator or density approximations, although it too may be provable. In section 4.3 we discuss bounding techniques for the numerator errors that may lead, in specific problems, to more useful results.

Before discussing the specific approximation technique for our conditional moment generating function U_t (equation 4.3), we want to examine the properties of "regular" moment generating functions. We depart slightly from our problem formulation, and introduce a new random variable (y), to avoid confusion with the rest of this chapter.

---

*The convergence only holds for the denominator not equal to zero. In our case, $w_n > 0$ for n sufficiently large, and the convergence will hold.

4.1.3 Moment Generating Functions

In this section we describe some of the properties of moment generating functions, moments, and cumulants which we will find useful. We assume that we have a random variable (y) with finite moments, that is

$$E[y^n] < \infty \tag{4.11}$$

for all $n < \infty$. Then the moment generating function $\phi(u)$ is a well-defined convex function of u [23] given by

$$\phi(u) = E[e^{uy}] \tag{4.12}$$

We also define the the log-moment generating function $\psi(u)$ by

$$\psi(u) = \ln \psi(u) \tag{4.13}$$

$\psi(u)$ is a convex function of u, since

$$\frac{\partial^2\psi}{\partial u^2} = \psi'' = \frac{\phi\phi'' - (\phi')^2}{\phi^2} \tag{4.14}$$

and

$$\{E[ye^{uy}]\}^2 \leq E[y^2e^{uy}]\ E[e^{uy}] \tag{4.15}$$

or

$$(\phi')^2 \leq \phi''\phi \tag{4.16}$$

by the Schwarz inequality. Thus $\psi'' > 0$ and ψ is convex.

From the definition of $\phi(u)$ (equation 4.12) we see that the n^{th} moment of y may be computed from

$$E[y^n] = \left. \frac{d^n \phi(u)}{du^n} \right|_{u=0} \tag{4.17}$$

Then $\phi(u)$ may be written as a Taylor (or Maclaurin) series [1, p. 880] in u about the point u=0, where the coefficient of $(u^n/n!)$ is the n th moment of y.

$$\phi(u) = \phi(0) + \phi' \Big|_{u=0} u + \phi'' \Big|_{u=0} \frac{u^2}{2} + \ldots \tag{4.18}$$

$$= \sum_{n=0}^{\infty} \frac{\mu_n u^n}{n!}$$

with

$$\mu_n = E[y^n] \tag{4.19}$$

$$\mu_0 = 1 \tag{4.20}$$

If we define the partial sum

$$\phi_N(u) = \sum_{n=0}^{N} \frac{\mu_n u^n}{n!} \tag{4.21}$$

then the remainder (R_N) is given by

$$R_N = \phi - \phi_n = \frac{u^{N+1}}{(N+1)!} E[y^{N+1} e^{vy}] \tag{4.22}$$

for some v with

$$0 \leq v \leq u \tag{4.23}$$

We may also consider a Taylor series for the log-moment generating function $\psi(u)$. We write

$$\psi(u) = \sum_{n=1}^{\infty} \frac{\lambda_n u^n}{n!} \tag{4.24}$$

where

$$\lambda_n = \frac{d^n}{du^n} \ln E[e^{uy}] \Big|_{u=0} \tag{4.25}$$

The λ_n's are called the cumulants, or the semi-invariants, of y [23 and 38], and ψ is sometimes called the "cumulant generating function." The cumulants have the important property of being invariant under a change of origin, except for the first cumulant , the mean. The n th cumulant can be expressed in terms of the first n moments by

$$\lambda_1 = \mu_1 \tag{4.26}$$

$$\lambda_2 = \mu_2 - \mu_1^2 = \sigma^2 \tag{4.27}$$

$$\lambda_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3 \tag{4.28}$$

$$\lambda_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4 \tag{4.29}$$

The formulae for the higher-order cumulants become quite tedious. Kendall and Stuart [23] tabulate them up to λ_{10}.

We see that λ_1 is the mean of y, λ_2 is the variance (σ^2), λ_3 is the 3rd central moment, and λ_4 is the 4th central moment minus $3\sigma^4$. For a Gaussian density, all of the cumulants after the second are zero. The cumulants for an arbitrary density thus represent a measure of the "non-Gaussianness" of the density. In this respect, λ_3 becomes an un-normalized coefficient of skewness, and λ_4 becomes an un-normalized coefficient of Kurtosis [23].

Since equation 4.24 is a Taylor series, we may form the partial sum ψ_N

$$\psi_N = \sum_{n=1}^{N} \frac{\lambda_n u^n}{n!} \tag{4.30}$$

and the remainder R_N'

$$R_N' = \psi - \psi_N = \frac{u^{N+1}}{(N+1)!} \left[\frac{d^{N+1}\psi(u)}{du^{N+1}} \right]_{u=v} \tag{4.31}$$

for some v with

$$0 \leq v \leq u \tag{4.32}$$

4.2 Approximation Method

4.2.1 General Design

We now propose a general method for approximating $U_t(x_t, z_0^t)$. We recall from equation 4.5 that

$$\zeta_t = \int_0^t h_s^T R^{-1} dz_s - \frac{1}{2} \int_0^t h_s^T R^{-1} h_s ds \tag{4.5}$$

Then ζ_t is a random variable, since it is A_t measurable. Furthermore, by the constraint (equation 3.52) needed to guarantee the existence (finiteness) of the Ito integral

$$\int_{0}^{t} h_s^T R^{-1} dz_s$$

we have that

$$E_0^{\mathcal{B}}[\zeta_t^n] < \infty \tag{4.33}$$

Therefore we may construct the conditional moment generating function

$$\phi^{\mathcal{B}}(u) = E_0^{\mathcal{B}} e^{u\zeta_t} \tag{4.34}$$

Then

$$U_t(x_t, z_0^t) = \phi^{\mathcal{B}}(1) \tag{4.35}$$

or

$$U_t(x_t, z_0^t) = e^{\psi^{\mathcal{B}}(1)} \tag{4.36}$$

for the conditional cumulant generating function $\psi^{\mathcal{B}}(u)$.

We can now approximate U_t by the partial sums of the moment and cumulant generating functions. In other words, we form

$$\phi_N^{\mathcal{B}}(u) = \sum_{n=0}^{N} \frac{\mu_n^{\mathcal{B}} u^n}{n!} \tag{4.37}$$

where

$$\mu_n^{\mathcal{B}} = E_0^{\mathcal{B}}[\zeta_t^n] \tag{4.38}$$

or

$$\psi_N^{\mathcal{B}}(u) = \sum_{n=0}^{N} \frac{\lambda_n^{\mathcal{B}} u^n}{n!} \tag{4.39}$$

where we form the conditional $\lambda_n^{\mathcal{B}}$'s from the relations 4.26 to 4.29.

The pointwise (in u) convergence of $\phi_N^{\mathcal{B}}$ (and $\psi_N^{\mathcal{B}}$) is guaranteed by equation 4.33, and we are only interested in the point u=1. Since we can approximate $U_t(x_t, z_0^t)$ and since we know p(x,t), we can form an approximation for $p(x,t|z_0^t)$ as in equation 4.7 that will converge (pointwise in x,t, and ω) to the correct density as $N \to \infty$.

We want to stress again that $\mu_n^{\mathcal{B}}$ is <u>not</u> a moment of the conditional density $p(x,t|z_0^t)$, but rather a (conditional) moment of the log of the Radon-Nikodym derivative

$$dP^{A}{}_t / dP_0^{A}{}_t$$

which will be useful in <u>approximating</u> the conditional density.

To demonstrate how the approximate densities can be computed, we examine the moments $\mu_n^{\mathcal{B}}$. While in general

$$E_0^{\mathcal{B}} e^{\zeta_t}$$

is impossible to find, each of the <u>moments</u> $\mu_n^{\mathcal{B}}$ is a straightforward, but tedious, functional (of x_t and z_0^t) to compute. For example, in

$$\mu_1^{\mathcal{B}} = E_0^{A_{zt} \vee A(x_t)} \left[\int_0^t h_s^T R^{-1} dz_s - \frac{1}{2}\int_0^t h_s^T R^{-1} h_s \, ds\right] \qquad (4.40)$$

the second term may be written

$$E_0^{A_{zt} \vee A(x_t)} \left[-\frac{1}{2}\int_0^t h_s^T R^{-1} h_s ds\right] = -\frac{1}{2}\int_0^t E_0^{A(x_t)} [h_s^T R^{-1} h_s] ds \quad (4.41)$$

because the integrand is only a function of x_s, and x_s is independent of

z_0^t under $\mathcal{P}_0$. (Strictly speaking, we also need a modified Fubini theorem for conditional expectations as discussed by Marcus [30], but this follows from our definition 3.7 b.) Also, since $\mathcal{P}^{\mathcal{A}_x} = \mathcal{P}_0^{\mathcal{A}_x}$ (as discussed in Section 3.2.5.2), we have that

$$E_0^{\mathcal{A}(x_t)}[f(x_s)] = E^{\mathcal{A}(x_t)}[f(x_s)] \tag{4.42}$$

for any function (f) of the state x_s.

The first term in equation 4.40 may be written as

$$E_0^{\mathcal{A}_{zt} \vee \mathcal{A}(x_t)}\left[\int_0^t h_s^T R^{-1} dz_s\right] = \int_0^t E^{\mathcal{A}(x_t)}[h_s]\, R^{-1} dz_s \tag{4.43}$$

by the same justification as in equations 3.44 and 4.41. Thus

$$\mu_1^{\mathcal{B}} = \int_0^t E^{\mathcal{A}(x_t)}[h(x_s)]\, R^{-1} dz_s - \frac{1}{2}\int_0^t E^{\mathcal{A}(x_t)}[h^T(x_s) R^{-1} h(x_s)]\, ds \tag{4.44}$$

We note that the expectations are over x_s conditioned on x_t for $\underline{0 \leq s \leq t}$. These expectations therefore are determined by the a priori transition density for x (specified by the dynamics, equation 3.50), independent of the measurements z_0^t.

The results of the expectations in equation 4.44 will be functions of s and x_t. $\mu_1^{\mathcal{B}}$ will then be a <u>known function</u> of x_t and z_0^t; therefore we can compute density approximations as

$$p_1(x,t|z_0^t) = \frac{e^{\mu_1^{\mathcal{B}}} p(x,t)}{\int e^{\mu_1^{\mathcal{B}}} p(x,t)\, dx} \tag{4.45}$$

or

$$p_1(x,t|z_0^t) = \frac{[1 + \mu_1^{\mathcal{B}}]\ p(x,t)}{\int [1 + \mu_1^{\mathcal{B}}]\ p(x,t)\,dx} \tag{4.46}$$

Higher order approximations will involve products of integrals which must be expressed as iterated integrals before the expectations can be taken. For instance, $\mu_2^{\mathcal{B}}$ will contain a term of the form

$$I = E_0^{\mathcal{B}}[\{\int_0^t h_s^T R^{-1} dz_s\}^2] \tag{4.47}$$

which can be written

$$I = E_0^{\mathcal{B}}[\int_0^t \int_0^t dz_s^T\ R^{-1} h_s h_r^T R^{-1} dz_r] \tag{4.48}$$

and finally

$$I = \int_0^t \int_0^t dz_s^T\ R^{-1} E^{A(x_t)} [h(x_s) h^T(x_r)]\ R^{-1} dz_r \tag{4.49}$$

so that the expectation may be evaluated as a function of s, r and x_t. Higher order terms become tedious to compute, but the expectations are always over the known density for x_s, $x_s x_r$, etc., conditioned on x_t.

Thus, we see that the moments for ζ_t conditioned on $\mathcal{B}$ can be computed even when the moment generating function cannot. The computation leads to analytical formulae for approximate conditional densities. We now investigate the relative merits of the different approximations that are possible.

4.2.2 Possible Approximations

In this section we want to investigate the relative merits of the possible expansions for U_t. The first expansion that one might consider is the partial sum of moments (equation 4.37) evaluated at u=1.

$$\phi_N^{\mathcal{B}}(u)\Big|_{u=1} = \sum_{n=0}^{N} \frac{\mu_n^{\mathcal{B}}}{n!} \tag{4.50}$$

This series converges for large N, but for small N there is no guarantee that

$$\phi_N^{\mathcal{B}}(1) > 0$$

even though

$$\phi^{\mathcal{B}}(1) > 0$$

by definition. Clearly our approximate density should be positive.

A less obvious drawback to the moment series is that the partial sum in equation 4.50 does not approach the known correct answer (for fixed N) as $Q \to 0$ (see section 3.3.1). An expansion in terms of functions of central moments does converge, however, since the central moments ($n \geq 2$) all go to zero as the variance does. We are thus led to consider the cumulant expansion of equation 4.39.

We propose approximations to U_t of the form

$$U_{t_N} = \exp\left\{\sum_{n=1}^{N} \frac{\lambda_n^{\mathcal{B}}}{n!}\right\} \tag{4.51}$$

where $\lambda_n^{\mathcal{B}}$ is the n'th cumulant. Since the cumulants are functions of the central moments (for $n \geq 2$) and since, for $Q = 0$, any function of x_s becomes a measurable function of x_t, we see that

$$\lambda_n^{\mathcal{B}} = 0 \qquad n \geq 2 \tag{4.52}$$

whenever $Q = 0$. We recall that $\lambda_1^{\mathcal{B}} = \mu_1^{\mathcal{B}}$, and for $Q=0$

$$\lambda_1^{\mathcal{B}} = E_0^{\mathcal{B}} \zeta_t = \zeta_t \tag{4.53}$$

Thus

$$U_{t_N} = U_t = e^{\zeta_t} \tag{4.54}$$

and the cumulant series of <u>any</u> N converges to the correct answer, as $Q \to 0$.

The cumulant approximation is positive by definition (equation 4.51) and converges to U_t as $N \to \infty$. Thus, for small Q, the cumulant series appears to be a better way to approximate U_t than the moment expansion. For large Q, the situation is less clear.

A third possible expansion may be obtained by noting that

$$E_0^{\mathcal{B}} e^{\zeta_t} = e^{E_0^{\mathcal{B}} \zeta_t} \left[E_0^{\mathcal{B}} \left(e^{\zeta_t - E_0^{\mathcal{B}} \zeta_t} \right) \right] \tag{4.55}$$

We can thus approximate U_t by

$$U_{t_N} = e^{E_0^{\mathcal{B}} \zeta_t} \left[\sum_{n=0}^{N} \frac{\nu_n^{\mathcal{B}}}{n!} \right] \tag{4.56}$$

where $\nu_n^{\mathcal{B}}$ is the n th <u>central</u> moment. This expansion, like the cumulant series, converges for any N as $Q \to 0$, as well as converging (for any Q) as $N \to \infty$. While it is not clear that U_{t_N} is positive for every N, we note that

$$U_{t_0} = e^{E_0^{\mathcal{B}} \zeta_t} > 0 \tag{4.57}$$

$$U_{t_1} = e^{E_0^{\mathcal{B}} \zeta_t} > 0 \tag{4.57}$$

$$U_{t_1} = e^{E_0^{\mathcal{B}} \zeta_t} > 0 \tag{4.58}$$

$$U_{t_2} = e^{E_0^{\mathcal{B}} \zeta_t} [1 + \frac{\nu_2^{\mathcal{B}}}{2}] > 0 \tag{4.59}$$

where equation 4.58 follows from $\nu_1^{\mathcal{B}} = 0$ and where equation 4.59 follows from the fact that the variance is guaranteed to be positive.

A final, less obvious way to approximate U_t is motivated by a Hermite polynomial expansion for the Radon-Nikodym derivative Λ_t (equation 3.55). McKean [31 p. 36] demonstrates that

$$\Lambda_t = e^{\zeta_t} = \sum_{n=0}^{\infty} H_n(A, B) \tag{4.60}$$

where A is the "intrinsic time" given by

$$A = \int_0^t h_s^T R^{-1} h_s \, ds \tag{4.61}$$

B is the Ito integral

$$B = \int_0^t h_s^T R^{-1} dz_s \tag{4.62}$$

and H_n is the n'th Hermite polynomial

$$H_n(r, s) = \frac{(-1)^n}{n!} \exp\left(\frac{s^2}{2r}\right) \frac{\partial^n}{\partial s^n} \exp\left(- \frac{s^2}{2r}\right) \tag{4.63}$$

Thus we may approximate U_t by

$$U_{t_N} = \sum_{n=0}^{N} E_0^{\mathcal{B}}[H_n(A,B)] \tag{4.64}$$

This expansion appears difficult to examine in detail, and we will not use it in the rest of this work.

We see, then, that there are several ways to approximate U_t. In general, we prefer the cumulant method (equation 4.51) because of its positivity and convergence properties. Because of the complexity of these approximations, it appears that the most useful terms will be those of low order, in particular, N=1. We therefore want to examine in detail the density $p(x_s|x_t)$, for $s \leq t$, which we will need for the first moment and cumulant.

4.2.3 Backward Transition Density

The expectation for the first moment and cumulant approximations is of the form

$$E^{\mathcal{A}(x_t)}[h(x_s)] = \int_{R^n} h(x_s)\, p(x_s|x_t)\,dx_s \tag{4.65}$$

where we see that we need the "backward" transition density

$$p(x_s|x_t) \quad \text{for } s \leq t$$

From Bayes' rule we have

$$p(x_s|x_t) = \frac{p(x_t|x_s)\ p(x,\ s)}{p(x,t)} \tag{4.66}$$

Thus, we can obtain the needed density from knowledge of the usual forward transition density and the a priori state density at times s and t.

If the initial density for x_t is

$$p(x,\ 0) = N(0,\ P_0) \tag{4.67}$$

and we have the dynamics of equation (1.2)

$$\dot{x} = Fx + G\dot{u} \tag{4.68}$$

then

$$p(x,\ t) = N(\bar{x}_t,\ P_t) \tag{4.69}$$

where

$$\dot{\bar{x}} = F\bar{x} \tag{4.70}$$

$$\dot{P} = FP + PF^T + GQG^T \tag{4.71}$$

If we let $\phi(t,s)$ be the transition matrix for F, that is

$$x_t = \phi(t,s)x_s + \int_s^t \phi(t,\sigma)Gu(\sigma)d\sigma \tag{4.72}$$

then

$$\bar{x}_t = \phi(t,s)\bar{x}_s \tag{4.73}$$

$$P_t = \phi(t,s)\ P_s\phi^T(t,s) + P_{t/s} \tag{4.74}$$

where

$$P_{t/s} = \int_s^t \phi(t,\sigma) GQG^T \phi^T(t,\sigma)\, d\sigma \tag{4.75}$$

This means that the forward transition density is

$$p(x_t | x_s) = N(\phi(t,s)x_s, P_{t/s}) \tag{4.76}$$

and using Bayes' rule (equation 4.66) we have that

$$p(x_s | x_t) = N(\bar{x}_{s/t}, P_{s/t}) \tag{4.77}$$

where

$$\bar{x}_{s/t} = P_{s/t}\, \phi^T(t,s)\, P_{t/s}^{-1} x_t \tag{4.78}$$

and

$$P_{s/t}^{-1} = \phi^T(t,s) P_{t/s}^{-1} \phi(t,s) + P_s^{-1} \tag{4.79}$$

Using the matrix manipulations familiar from the discrete Kalman filter (see e.g., Bryson and Ho [6] p. 357), we may write

$$\bar{x}_{s/t} = P_s \phi^T(t,s)\; [\phi(t,s)P_s\, \phi^T(t,s) + P_{t/s}]^{-1}\, x_t \tag{4.80}$$

and

$$P_{s/t} = P_s - P_s \phi^T(t,s)\; [\phi(t,s)P_s \phi^T(t,s) + P_{t/s}]^{-1}\, \phi(t,s)P_s \tag{4.81}$$

which, because of equation 4.74, become

$$\bar{x}_{s/t} = P_s \phi^T(t,s)\, P_t^{-1} x_t \tag{4.82}$$

and

$$P_{s/t} = P_s - P_s\phi^T(t,s)\; P_t^{-1}\phi(t,s)P_s \tag{4.83}$$

In addition, if F is not a function of time and P_t has a steady-state value

$$P_t = P_s = P \tag{4.84}$$

then the above equations may be written

$$\bar{x}_{s/t} = P\phi^T(t,s)P^{-1}x_t \tag{4.85}$$

$$P_{s/t} = P-P\phi^T(t,s)P^{-1}\phi(t,s)P \tag{4.86}$$

These equations simplify still further if x_t is a scalar. Then

$$\bar{x}_{s/t} = \phi(t,s)x_t \tag{4.87}$$

and

$$P^{-1}_{s/t} = \phi^2(t,s)P^{-1}_{t/s} + P^{-1} = \frac{\phi^2(t,s)P + P_{t/s}}{P_{t/s}\,P} \tag{4.88}$$

Using equation 4.74 we then have

$$P^{-1}_{s/t} = P^{-1}_{t/s}\,\frac{P}{P} \tag{4.89}$$

so that

$$P_{s/t} = P_{t/s} \tag{4.90}$$

4.3 Approximation Accuracy

4.3.1 General Considerations

One of the advantages of our approximate density approach to filtering is the availability of an approximate density with which to compute estimates, minimum costs, and parameter optimization studies. These computations, however, are only as good as the density approximation itself, and we are thus led to consider the overall accuracy of the moment and cumulant truncation methods.

In general, we know that the numerator approximations converge pointwise in x, t, and ω. But we would like to know, quantitatively, how good the numerator approximations are, and how much improvement each additional term is likely to bring, especially for the critical first few terms. It is difficult to say anything quantitative about this convergence in general, but we will discuss some bounding methods that may prove useful in individual problems.

4.3.2 Moment Approximations

4.3.2.1 Denominator Convergence

The moment approximation converges, pointwise in x, t, and ω, to the correct moment generating function U_t (equation 4.3). Thus, our approximate numerator V_N

$$V_N = \sum_{K=1}^{N} \frac{E_0^{\beta}\, \zeta_t^K}{K!}\, p(x,t) \tag{4.91}$$

converges to the correct numerator, V. We now demonstrate that the denominator W_N

$$W_N = \int_{R^N} V_N \, dx \tag{4.92}$$

also converges, pointwise in ω and t, to the correct denominator. From equation 3.37, we know that the exact denominator is

$$E_0^{\mathcal{A}_{zt}} e^{\zeta_t} = \int_{R^N} (E_0^{\mathcal{B}} e^{\zeta_t}) \; p(x,t)\,dx \tag{4.93}$$

but

$$\int_{R^n} V_N dx = \sum_{K=1}^{N} \frac{E_0^{\mathcal{A}_{zt}} \zeta^K}{K!} \tag{4.94}$$

Thus,

$$W_N = \sum_{K=1}^{N} \frac{E_0^{\mathcal{A}_{zt}} \zeta^K}{K!} \tag{4.95}$$

converges pointwise in ω and t just as the moment sum in the numerator converges pointwise in x, ω, and t. Both numerator and denominator represent partial sums in the moment expansion for the moment generating function. The only difference is that the numerator expectations are conditioned on the σ-algebra

$$\mathcal{B} = \mathcal{A}_{zt} \vee \mathcal{A}(x_t) \tag{4.96}$$

while the denominator is conditioned on $\mathcal{A}_{zt}$ alone. The pointwise convergence of the numerator <u>and</u> denominator therefore guarantees, as discussed in section 4.1.2, the pointwise (in x, t, and ω) convergence of the density approximations for the moment expansion method.

4.3.2.2 Moment-Approximation Bound

Next, we discuss a method for bounding the remainder in the Taylor series expansion for the moment generating function. Adapting the remainder in equation 4.22 to the conditional expectations used in our approximation, we have

$$R_N = \frac{1}{(N+1)!} E_0^{\mathcal{B}} [\zeta^{N+1} e^{v\zeta}] \tag{4.97}$$

for some v such that

$$0 \leq v \leq 1$$

Now for N even we have

$$\frac{\partial}{\partial v} R_N = \frac{1}{(N+1)!} E_0^{\mathcal{B}} [\zeta^{N+2} e^{v\zeta}] \geq 0 \tag{4.98}$$

so that R_N is an increasing function of v. Therefore

$$R_N \leq \frac{1}{(N+1)!} E_0^{\mathcal{B}} [\zeta^{N+1} e^{\zeta}] \tag{4.99}$$

which may be written (using equation 3.16)

$$R_N \leq \frac{1}{(N+1)!} E^{\mathcal{B}} [\zeta^{N+1}] E_0^{\mathcal{B}} [e^{\zeta}] \tag{4.100}$$

Now since $E_0^{\mathcal{B}} [e^{\zeta}]$ is the actual numerator, we may define a "normalized" remainder as

$$s_N = \frac{R_N}{E_0^{\mathcal{B}} e^{\zeta}} = \frac{E_0^{\mathcal{B}} e^{\zeta} - \phi_N^{\mathcal{B}}}{E_0^{\mathcal{B}} e^{\zeta}} \tag{4.101}$$

where $\phi_N^{\mathcal{B}}$ is the N'th moment approximation given by equation 4.50. Then

$$s_N \leq \frac{1}{(N+1)!} E^{\mathcal{B}} \zeta^{N+1} \tag{4.102}$$

so that

$$E\, s_N \leq \frac{1}{(N+1)!} E[\zeta^{N+1}] \tag{4.103}$$

for N even.

A more useful measure of the overall approximation accuracy may be found from the average value of s_N^2, where

$$s_N^2 = [\frac{1}{(N+1)!}]^2 [E^{\mathcal{B}} \zeta^{N+1}]^2 \tag{4.104}$$

Jensen's inequality provides that

$$[E^{\mathcal{B}} \zeta^{N+1}]^2 \leq E^{\mathcal{B}} \zeta^{2N+2} \tag{4.105}$$

so that

$$E\, s_N^2 \leq [\frac{1}{(N+1)!}]^2 E[\zeta^{2N+2}] \tag{4.106}$$

for N even. The usefulness of these bounds (equations 4.103 and 4.106) of course, is determined by the ability to evaluate or bound the expectation of ζ for a particular problem. If the expectation can be evaluated or bounded by some function of t (independent of x and ω), the bound will represent uniform convergence in x and convergence in mean (for equation 4.103) or mean square (equation 4.106) for ω, in place of the convergence "in probability" mentioned in section 4.1.2.

4.3.3 Cumulant Bound

The cumulant bound (R_N' in equation 4.31) corresponding to equation 4.97 is, in general, somewhat more difficult to evaluate because of the difficult formulae relating the higher-order cumulants and moments. We concentrate, therefore, on the error in the first cumulant approximation

$$E_0^{\mathcal{B}} e^{\zeta} \simeq e^{E_0^{\mathcal{B}} \zeta} \tag{4.107}$$

We consider the normalized error

$$\Delta = \frac{E_0^{\mathcal{B}} e^{\zeta} - e^{E_0^{\mathcal{B}} \zeta}}{e^{E_0^{\mathcal{B}} \zeta}} \tag{4.108}$$

If we define ξ by

$$\xi = \zeta - E_0^{\mathcal{B}} \zeta \tag{4.109}$$

then

$$\Delta = E_0^{\mathcal{B}} e^{\xi} - 1 \tag{4.110}$$

Now it is possible to write

$$\Delta^2 = (E_0^{\mathcal{B}} e^{\xi})^2 - 2E_0^{\mathcal{B}} e^{\xi} + 1 \tag{4.111}$$

$$\leq E_0^{\mathcal{B}} e^{2\xi} - 2E_0^{\mathcal{B}} e^{\xi} + 1 \tag{4.112}$$

by Jensen's inequality. Then we have, finally,

$$E_0\ \Delta^2 \leq E_0[e^{2\xi} - 2e^{\xi} + 1] \tag{4.113}$$

This bound, like those in equations 4.103 and 4.106, depends on the ability to compute the expectation on the right. For the 1st order PLL problem which we discuss in the next chapter, we will demonstrate how $E_0\ \Delta^2$ may be bounded in the case where ζ is bounded.

4.3.4 Statistical Bounds

Because the general bounds of sections 4.3.2 and 4.3.3 are difficult to evaluate, we now consider a somewhat different approach. We recall that the numerator U_t is a (conditional) moment generating function for a random variable (ζ_t) with unknown density but with computable moments and cumulants. It seems reasonable, therefore, to bound the error in a finite cumulant approximation by that error achieved for a "worst-case" density for ζ. Here a "worst-case" density is one in which the higher-order cumulants have the greatest effect.

Recalling that the cumulants of order greater than two are a measure of the "non-Gaussianness" of the density (section 4.1.3), we consider the least "Gaussian" density - the uniform density. Our argument here is not rigorous. One cannot find a density with the "highest" cumulants, and clearly the uniform density, being symmetric, will have only even cumulants. Nonetheless, we feel that by picking a uniform density of the same variance as that of the real density we may obtain a conservative estimate of the approximation error. The uniform density may not be the worst "conceivable" case, but it should be the worst "practical" case. We let

$$p(y) = \frac{1}{2a} \quad \text{for } |y| \leq a \tag{4.114}$$

for a random variable y. Then

$$E[e^y] = \frac{1}{2a}[e^a - e^{-a}] \tag{4.115}$$

$$E[y] = 0 \tag{4.116}$$

and

$$E[y^2] = \frac{a^2}{3} = \sigma^2 \tag{4.117}$$

We now approximate the moment generating function (equation 4.115) by the second cumulant approximation

$$E[e^y] \simeq e^{\sigma^2/2} \tag{4.118}$$

Then we can define a per cent error in this approximation as

$$PE = \frac{E[e^y] - e^{\sigma^2/2}}{E[e^y]} \cdot 100 \tag{4.119}$$

This equation may be written as a function of σ by using equations 4.115 and 4.117

$$PE(\sigma) = \left[1 - \frac{2\sqrt{3}\ \sigma\ e^{\sigma^2/2}}{e^{\sqrt{3}\ \sigma} - e^{-\sqrt{3}\ \sigma}}\right]100 \tag{4.120}$$

which is evaluated in Table 4.1 for a few values of σ.

Table 4.1 Percent Error in Second Cumulant Approximation for Uniform Density

σ	PE (%)
0.1	.0005
0.2	.008
0.3	.04
0.4	.12
0.5	.30
1.0	4.3
1.414	15.8
2.0	60.4
3.0	418

The error is very small for low σ, but increases rapidly beyond $\sigma = 1$. We stress that this is the error in a second-order approximation (cumulant) for the case when the random variable is uniformly distributed. This is expected to be an upper bound on the approximation error for our variable ζ.

This analysis illustrates the type of bound one might get by considering the statistical nature of the problem. A tighter bound would be more useful, but having an error indication as a function of the (observed) variance is, by itself, an interesting result.

4.4 Conclusion

We have developed a general approximation method for the conditional density function based on Bucy's representation theorem. The recommended cumulant version of our method produces a positive approximation density which converges to the correct density as the process noise strength goes to zero and as the number of terms in the approximation becomes infinite. This method, unlike numerical approximation schemes, produces a set of statistics which are functionals of the current state and past measurements. These statistics are then used to specify completely an actual density function which, in place of the exact conditional density, can provide "optimal" estimates and "minimum" costs.*

In the next chapter we apply this method to the design of suboptimal filters for the first-order phase-lock loop problem. We show that the first cumulant approximation produces a filter (first proposed by Mallinckrodt, et al. [29]) which outperforms the PLL (the extended Kalman filter) for this problem.

---

*"Optimal" and "minimal" with respect to the <u>approximate</u> density.

CHAPTER 5

THE FIRST-ORDER PHASE-LOCK LOOP PROBLEM

5.1 Introduction

5.1.1 Chapter Organization

In this chapter we consider the Brownian motion phase problem first discussed in section 1.4.3. We begin by examining the results of recent researchers in this area, first analyzing (unrealizable) optimal filters and then realizable sub-optimal designs. For some of the sub-optimal filters, we bring out relationships not noted before, including "hidden" filter equivalences and high signal-to-noise ratio (SNR) convergence properties. We then show that the first cumulant approximation (of the last chapter) leads to a filter, first suggested by Mallinckrodt, et. al. [29], that outperforms the extended Kalman filter for this problem - the PLL. We investigate the accuracy of this filter and higher-order approximations and note the difficulty of obtaining estimates and the generally slow convergence to the optimum. Finally, we examine a modification of the first cumulant filter that offers increased performance at a slight increase in complexity.

5.1.2 Problem Statement

We consider the Brownian motion phase process (θ_t) first encountered in section 1.4.3. We let

$$d\theta = du \tag{5.1}$$

where

$$E[du^2] = q\ dt \tag{5.2}$$

We assume that the baseband measurements of equation 1.18

$$d\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} dt + \begin{pmatrix} dn_1 \\ dn_2 \end{pmatrix} \tag{5.3}$$

are available, and we sometimes use the notation

$$dz = \begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \text{ and } dn = \begin{pmatrix} dn_1 \\ dn_2 \end{pmatrix}$$

where the noise strength for n is given by

$$E[dn\ dn^T] = \begin{pmatrix} 2r & 0 \\ 0 & 2r \end{pmatrix} dt \tag{5.4}$$

We will also use the quadrature measurements (from equation 1.21)

$$\begin{pmatrix} dz_I \\ dz_Q \end{pmatrix} = \begin{pmatrix} \sin(\theta-\hat{\theta}) \\ \cos(\theta-\hat{\theta}) \end{pmatrix} dt + \begin{pmatrix} dn_I \\ dn_Q \end{pmatrix} \tag{5.5}$$

for some estimate $\hat{\theta}$. The noise strength for the quadrature baseband noise is given by

$$E\left[\begin{pmatrix} dn_I \\ dn_Q \end{pmatrix}\begin{pmatrix} dn_I \\ dn_Q \end{pmatrix}^T\right] = \begin{pmatrix} 2r & 0 \\ 0 & 2r \end{pmatrix} dt \tag{5.6}$$

Our problem is to estimate θ_t given z_0^t; more precisely, we wish to find the conditional density

$$p(\theta,t|z_0^t)$$

5.2 The Phase-Lock Loop

The phase-lock loop is commonly used to estimate θ_t for this problem. The loop resulting from the linearized analysis is a first-order PLL as discussed in section 1.4.3.1. This loop is the extended Kalman filter for this problem (see section 1.4.2) and, as such, represents a useful performance benchmark. We will consider three representations for the first-order PLL.

The first representation is the most common, where the PLL generates a phase estimate through the equation (see equation 1.54 and Figure 1.3)

$$d\hat{\theta} = K\ dz_I \tag{5.7}$$

where

$$K = P_{\theta\ell}/2r \tag{5.8}$$

and

$$P_{\theta\ell} = \sqrt{2rq} \tag{5.9}$$

This $P_{\theta\ell}$ is the linear-predicted phase-error variance and represents a useful noise parameter for the problem.

The second loop representation comes from the extended Kalman-filter interpretation of the PLL. Equation 5.7 may be written

$$d\hat{\theta} = K\begin{pmatrix}\cos\hat{\theta}\\ -\sin\hat{\theta}\end{pmatrix}\left[dz - \begin{pmatrix}\sin\hat{\theta}\\ \cos\hat{\theta}\end{pmatrix}dt\right] \tag{5.10}$$

A third representation for the first-order PLL may be found by considering the "rectangular coordinates"

$$x = \begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}\sin\hat{\theta}\\ \cos\hat{\theta}\end{pmatrix} \tag{5.11}$$

By taking the Ito derivatives of x_1 and x_2 as functions of $\hat{\theta}$ (in equation 5.10) we obtain

$$d\begin{pmatrix}x_1\\ x_2\end{pmatrix} = -\frac{q}{2}\begin{pmatrix}x_1\\ x_2\end{pmatrix}dt + K\begin{bmatrix}x_2^2 & -x_1x_2\\ -x_1x_2 & x_1^2\end{bmatrix}\left[dz - \begin{pmatrix}x_1\\ x_2\end{pmatrix}dt\right] \tag{5.12}$$

These three representations all describe the same filter, and will be useful in what follows.

It is worthwhile to examine the actual performance of this filter. We recall that the steady-state error density of the PLL is given by (see equation 1.61)

$$p(\varepsilon) = \frac{e^{\frac{1}{P}\cos\varepsilon}}{2\pi\, I_0(1/P)} \tag{5.13}$$

for

$$\varepsilon = \theta - \hat{\theta} \tag{5.14}$$

and that the gain K that minimizes P ($P_{\theta\ell}$ = minimum P) also minimizes the actual error variance (equation 1.64) and the expected value of the cosine cost function (sections 1.4.3.2 and 1.4.3.3). In particular, for the density above (equation 5.13), the (minimum) expected cost is

$$E[1-\cos\varepsilon] = 1 - \frac{I_1(1/P_{\theta\ell})}{I_0(1/P_{\theta\ell})} \tag{5.15}$$

This is a well-behaved function of $P_{\theta\ell}$, going from 0 at $P_{\theta\ell} = 0$ to 1 at $P_{\theta\ell} = \infty$. (The upper bound on this cost function is 1 — the value it attains for a uniform error density.)

If the phase measurement were linear with the same noise strength, then the conditional density would be Gaussian, and the modulo-2π error density would be a "folded-normal" (see [43] or [4])

$$p_\ell(\varepsilon) = \frac{1}{2\pi}\left[1 + 2\sum_{n=1}^{\infty} e^{-n^2 P_{\theta\ell}/2} \cos n\varepsilon\right] \tag{5.16}$$

The expected (cosine) cost would become

$$E_\ell[1-\cos\varepsilon] = 1-e^{-P_{\theta\ell}/2} \tag{5.17}$$

We plot these two cost functions (equation 5.15 and 5.17) versus $P_{\theta\ell}$ in figure 5.1. In most communications applications, $P_{\theta\ell}$ is usually much less than 1, and the actual PLL performance is only slightly worse than the linear prediction. For very high $P_{\theta\ell}$, however, the PLL outperforms its linear prediction for the cosine cost function.

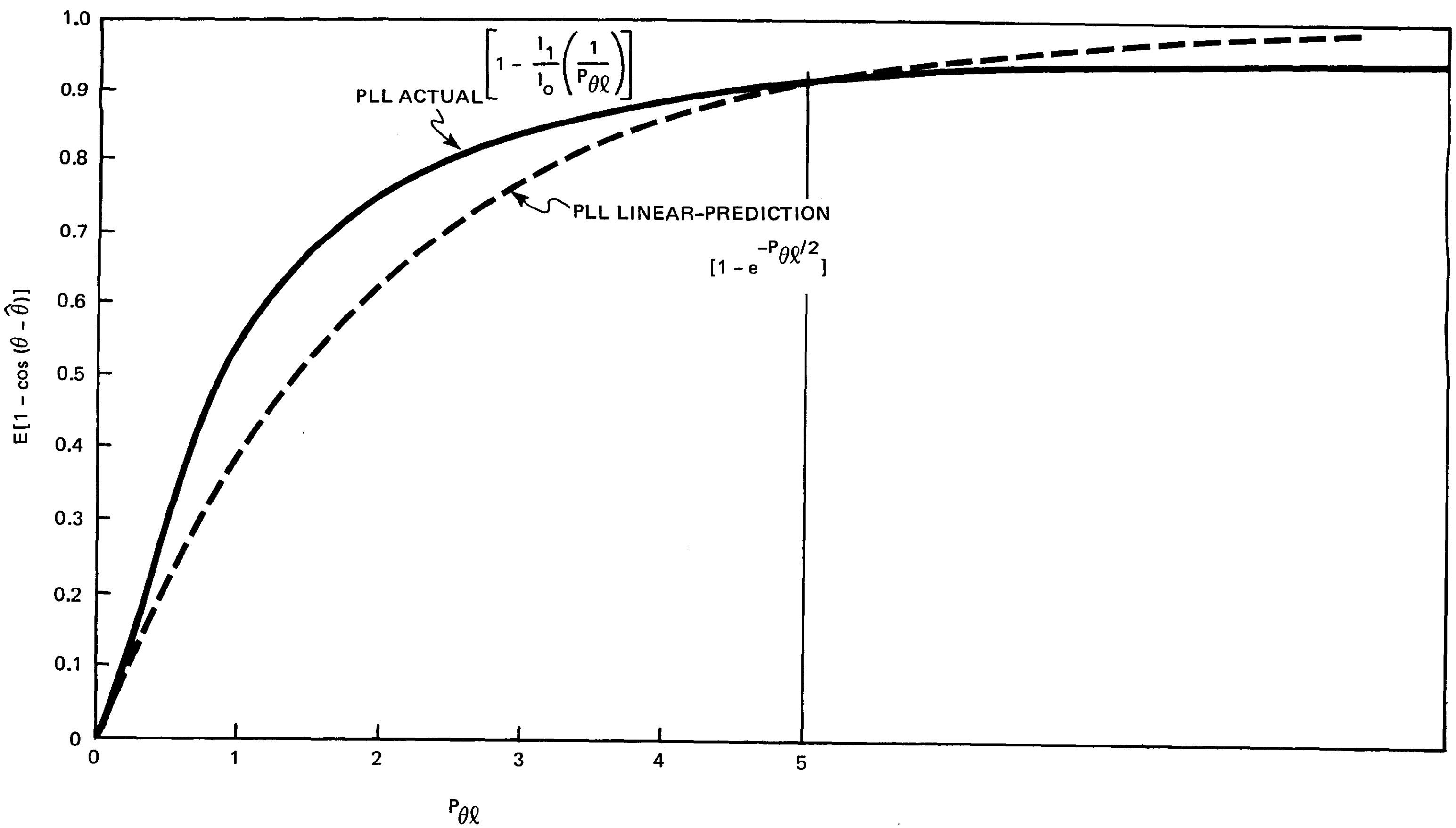


Figure 5.1 PLL Actual and Linear-Predicted Performance vs. $P_{\theta\ell}$

This is a somewhat surprising result. For the more usual error-squared cost function, the actual PLL performance is always worse than the linear predicted variance ($P_{\theta\ell}$), as shown by Galdos [14] (using the method of Snyder and Rhodes [35]). For the cosine cost function, however, there is a point at which the PLL, with its nonlinear measurement, performs better than the optimal linear filter with a linear phase measurement. Thus the linear-predicted performance is not a lower bound on the optimal filter performance for all value of $P_{\theta\ell}$, although it remains a practical lower bound for most reasonable values of $P_{\theta\ell}$.

5.3 Optimal-Filter Descriptions

5.3.1 Stratonovich

Stratonovich [37] was the first to describe the optimal filer, rather than the optimal PLL. He examined the (conditional) cosine cost function

$$E[1-\cos(\theta_t-\hat{\theta}_t)\,|\,z_0^t] \tag{5.18}$$

and noted that the optimal estimate ($\hat{\theta}_t$) that minimized this function was given by

$$\hat{\theta} = \tan^{-1}\left(a_1/b_1\right) \tag{5.19}$$

where a_1 and b_1 are the conditional estimates of the sine and cosine, that is

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = E\left[\begin{pmatrix} \sin\theta_t \\ \cos\theta_t \end{pmatrix} \middle| \; z_0^t\right] \tag{5.20}$$

The minimum cost is therefore

$$1 - \sqrt{a_1^2 + b_1^2}$$

He then derived the stochastic differential equations for a_1 and b_1 as part of the general set of Fourier coefficients for the conditional density. If we define

$$p(\theta, t \mid z_0^t) = \frac{1}{2\pi}\left[1 + 2\sum_{n=1}^{\infty} a_n \sin n\theta + b_n \cos n\theta\right] \tag{5.21}$$

where

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = E\left[\begin{pmatrix} \sin n\theta_t \\ \cos n\theta_t \end{pmatrix} \middle| \; z_0^t \right] \tag{5.22}$$

then

$$d\begin{pmatrix} a_n \\ b_n \end{pmatrix} = \frac{-n^2 q}{2}\begin{pmatrix} a_n \\ b_n \end{pmatrix} dt + \frac{1}{2r} \; H_n \left[dz - \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} dt \right] \tag{5.23}$$

where

$$H_n = \begin{bmatrix} \dfrac{b_{n-1} - b_{n+1}}{2} - a_1 a_n & \dfrac{a_{n-1} + a_{n+1}}{2} - a_n b_1 \\ \\ \dfrac{a_{n+1} - a_{n-1}}{2} - a_1 b_n & \dfrac{b_{n-1} + b_{n+1}}{2} - b_1 b_n \end{bmatrix} \tag{5.24}$$

with

$$a_0 = 0$$

$$b_0 = 1$$

We recognize the matrix H_n from the no-process-noise analysis of section 3.3.2. These equations represent an infinite-dimensional optimal filter for the Brownian motion phase problem and cannot be truncated because of the coupling of each nth set of coefficients to the (n-1)th and (n+1)th equations. We are therefore led to consider approximate methods for computing the conditional density.

5.3.2 Mallinckrodt, Bucy and Cheng

Mallinckrodt, Bucy and Cheng [29] performed the first large-scale effort to analyze the Brownian motion phase problem from the viewpoint of estimation theory. They approximated the (smooth) conditional density by approximately 100 point masses in the state (phase) space. This approximate density was propagated numerically through the dynamics (by the Chapman-Kolmogorov equation) and measurements (by Bayes' rule) in a rather complex (and slow) computer program, which was used in digital simulations.* These Monte Carlo simulations showed what the minimum variance was, but otherwise provided little insight into the problem. Later work by the authors and their students [8, 10, 12, 16] has centered on improving the computational speed of the optimal filters for the first- and second-order PLL problems.

---

*For propagating the complete density, see also Levieux [25-27].

5.3.3 Gaussian Sum Approximations

In [39], Tam and Moore apply the Gaussian sum approach of Sorenson and Alspach [2, 36] to the design of phase estimators for this problem. They concentrated on simplifying the ad hoc reinitialization required in this method and developed a class of estimators that performed optimally as the number of Gaussian densities in the sum became large. When only one density was used, their filter became the extended Kalman filter, and therefore the PLL, for this problem. For two densities in the sum, their filter performance was close to that of the Fourier coefficient filter of Willsky [44].

This approach, like the point-mass method of Mallinckrodt, et al., produces a density which numerically approximates the conditional density, resulting in a complex filter which performs well. These methods, however, do not offer simple parametric approximations to the conditional density and do not readily provide insight into the structure of the optimal filter.

5.4 Sub-Optimal Filters

5.4.1 Mallinckrodt, Bucy and Cheng

In addition to their numerical studies (as discussed in section 5.3.2 above), Mallinckrodt, et al. [29], proposed a sub-optimal filter of the form

$$d\begin{pmatrix} x \\ y \end{pmatrix} = -\frac{q}{2}\begin{pmatrix} x \\ y \end{pmatrix} dt + \frac{1}{2r}\begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \qquad (5.25)$$

$$\hat{\theta} = \tan^{-1}(x/y) \tag{5.26}$$

for the case when $q \neq 0$. The authors claimed that this filter, which they called a "static phase filter," performed well, although they did not include any simulation results. The structure of the filter was motivated by that of the no-process-noise ($q=0$) optimal filter discussed in section 3.3.2.

Mallinckrodt, et al., also described a general static phase filter, where (changing their notation)

$$d\begin{pmatrix} x \\ y \end{pmatrix} = -f\begin{pmatrix} x \\ y \end{pmatrix} dt + g\begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \tag{5.27}$$

$$f \to 0 \quad \text{as } q \to 0$$

and

$$\hat{\theta} = \tan^{-1}(x/y) \tag{5.28}$$

for arbitrary positive gains f and q. Since any constant multiplying x <u>and</u> y will not affect the phase estimate $\hat{\theta}$, we may assume, without loss of generality, that $g = 1/2r$.

They also demonstrated that any static phase filter may be regarded as a phase-lock loop with a data-dependent gain. To show this, we define

$$\alpha = \sqrt{x^2 + y^2} \tag{5.29}$$

$$\beta = \tan^{-1}(x/y) \tag{5.30}$$

Then the Ito derivatives of α and β as functions of x and y (in equation

5.27) produce (see Appendix C).

$$d\alpha = (\frac{1}{4r\alpha} - f\alpha)dt + \frac{1}{2r} \; dz_Q \tag{5.31}$$

$$d\beta = \frac{1}{2r\alpha} dz_I \tag{5.32}$$

where dz_I and dz_Q are given by equation 5.5 for $\hat{\theta}=\beta$.

Thus, all static phase filters may be regarded as phase-lock loops with data-dependent gains

$$K = \frac{1}{2r\alpha} \tag{5.33}$$

This concept, first discussed in [29], is interesting. It means that bandpass filters around the carrier frequency (ω_c), which may be implemented with low-pass filters on the baseband signals (such as the static phase filter in equation 5.27), may be considered "special" phase-lock loops if some care is taken in viewing the estimate. For instance, when looking for a sinsoid synchronized with θ, we should take (using equation 5.28)

$$\sin \hat{\theta} = x/\sqrt{x^2 + y^2} \tag{5.34}$$

and <u>not</u>

$$\sin \hat{\theta} = x \tag{5.35}$$

as is usually done.

It is also interesting that the data-dependence of the gain (α) is in the quadrature (cosine) measurement "channel" $\dot{z}_Q$. The quadrature channel is often used in phase-lock loops to provide "lock" and "loss-of-lock" indications, but the static phase filters show how to use $\dot{z}_Q$ to provide information about the quality of the measurements. This is an intuitively pleasing idea, since $\dot{z}_Q$ results from the correlation between the carrier-frequency measurement $\dot{z}'$ (in equation 1.16) and the filter estimate of $\dot{z}'$, $\sin(\omega_c t + \hat{\theta})$. We next examine a static phase filter which was derived in a much different way.

5.4.2 Linear Minimum-Variance Filters

In [15] Gustafson and Speyer demonstrated that the first-order phase-lock loop problem could be viewed as a state-dependent noise problem in the "measurement space". To see this, we define

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} \tag{5.36}$$

where

$$d\theta = du$$

as in equation 5.1 and dz is given by equation 5.3. Then we may take the Ito derivative of x_1 and x_2 to obtain

$$d\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -\frac{q}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dt + \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} du \tag{5.37}$$

with

$$dz = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dt + dn \tag{5.38}$$

For state-dependent noise systems with linear measurements, as above, one can derive the optimal linear filter (the linear filter with the minimum error variance) as was done in [15]. In steady-state this filter becomes

$$d\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\end{pmatrix} = -\frac{q}{2}\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\end{pmatrix}dt + K\left[dz - \begin{pmatrix}\hat{x}_1\\ \hat{x}_2\end{pmatrix}dt\right] \tag{5.39}$$

where

$$K = \left[\frac{\sqrt{rq(rq+1)}}{2r} - \frac{q}{2}\right] \tag{5.40}$$

is obtained from the solution to a Riccati equation. We also pick

$$\hat{\theta} = \tan^{-1}(x_1/x_2) \tag{5.41}$$

The authors called this filter (at carrier frequency, rather than at baseband as shown) a "linear, minimum-variance unbiased quadrature filter," which they abbreviated LQF.

The LQF may be rewritten as

$$d\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\end{pmatrix} = -f\begin{pmatrix}\hat{x}_1\\ \hat{x}_2\end{pmatrix}dt + K\begin{pmatrix}dz_1\\ dz_2\end{pmatrix} \tag{5.42}$$

where

$$f = \frac{\sqrt{rq(rq+1)}}{2r} \tag{5.43}$$

Then an equivalent filter (one with the same phase estimate) may be obtained by defining

$$\begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = \frac{1}{2rK} \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} \tag{5.44}$$

where

$$d \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} = - f \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \end{pmatrix} dt + \frac{1}{2r} \begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \tag{5.45}$$

$$\hat{\theta} = \tan^{-1}(\hat{y}_1/\hat{y}_2) \tag{5.46}$$

This is a static phase filter with a gain f (chosen to minimize the measurement space error variance) which goes to zero as q does (for constant r). Thus, this filter should approach the optimal as $q \to 0$, and the simulation results in [15] seem to support this.

The analysis in [15], however, indicates that the LQF has an actual phase error variance that is approximately 6% <u>above</u> $P_{\theta\ell}$ for very small $P_{\theta\ell}$. The actual PLL performance (equation 1.61) is also slightly worse than $P_{\theta\ell}$, although apparently less than 6%. Thus the LQF (and possibly other static phase filters) does not outperform the phase-lock loop for all values of $P_{\theta\ell}$, although it does converge to the optimal filter as $q \to 0$.

5.4.3 Assumed-Density Filter

In [44], Willsky presented results for a sub-optimal, nonlinear filter that closely matched the performance of the LQF for high $P_{\theta\ell}$ and approached that of the (optimal) PLL for low $P_{\theta\ell}$. We derive a baseband version of this filter and show that it does converge to the classic PLL, for low $P_{\theta\ell}$, in form as well as performance.

We recall first the Fourier coefficient equations of Stratonvich (equation 5.23), where the a_n's and b_n's form an infinite set of coupled differential equations. We can truncate these equations by assuming that the conditional density is a folded-normal density (see equation 5.16), so that the Fourier coefficients become

$$\begin{pmatrix} a_n \\ b_n \end{pmatrix} = e^{-n^2\gamma/2} \begin{pmatrix} \sin n\beta \\ \cos n\beta \end{pmatrix} \tag{5.47}$$

where β is the mean, and γ the variance, of the normal density which generates the folded-normal.

We then use a_1 and b_1 to solve for γ and β, so that all the higher-order coefficients may be written as functions of a_1 and b_1. Specifically, we write

$$\begin{pmatrix} \sin \beta \\ \cos \beta \end{pmatrix} = \frac{1}{\sqrt{a_1^2 + b_1^2}} \begin{pmatrix} a_1 \\ b_1 \end{pmatrix} \tag{5.48}$$

and

$$e^{-\gamma} = a_1^2 + b_1^2 \tag{5.49}$$

Using these identities (equations 5.48 and 5.49), the differential equations for a_1 and b_1 may be written as

$$d\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = -\frac{q}{2}\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} dt + \frac{1}{2r}\begin{pmatrix} K_1 & K_2 \\ K_2 & K_3 \end{pmatrix}\left(dz - \begin{pmatrix} a_1 \\ a_1 \end{pmatrix} dt\right) \tag{5.50}$$

where

$$K_1 = \frac{1}{2}\,(1-e^{-2\gamma}) - a_1^2(1-e^{-\gamma}) \tag{5.51 a}$$

$$K_2 = -\,a_1 b_1 (1-e^{-\gamma}) \tag{5.51 b}$$

$$K_3 = \frac{1}{2}(1-e^{-2\gamma}) - b_1^2(1-e^{-\gamma}) \tag{5.51 c}$$

This Fourier coefficient filter (FCF) with

$$\hat{\theta} = \tan^{-1}(a_1/b_1) \tag{5.52}$$

worked well in simulations, as reported in [44] and later in this chapter, marginally outperforming the LQF design of [15]. Also, since the conditional density becomes Gaussian as $P_{\theta\ell}\to 0$, and since the normal and folded-normal densities converge as the variance approaches zero, the FCF performance approaches that of the optimal PLL as $P_{\theta\ell}\to 0$.

We now want to demonstrate that, for small $P_{\theta\ell}$, the FCF approaches the PLL in <u>form</u> as well as performance. We begin by showing that $\gamma=P_{\theta\ell}$ for small $P_{\theta\ell}$. First we obtain the Ito derivative (appendix <u>C</u>) of $e^{-\gamma}$, as a function of a_1 and b_1 in equations 5.49 and 5.50, as

$$d(e^{-\gamma}) = -q\, e^{-\gamma}dt + \frac{1}{2r}\, e^{-\gamma}(1-e^{-\gamma})^2 dz_Q + \frac{1}{4r}\, (1-e^{-\gamma})^4\, dt \qquad (5.53)$$

where (see equation 5.5)

$$dz_Q = e^{\gamma}(a_1,\, b_1)\,dz = \cos(\theta-\hat{\theta})\,dt + dn_Q \qquad (5.54)$$

Equation 5.53 is stable for $e^{-\gamma} < 1$, which is the expected range of $e^{-\gamma}$ considering the cost function interpretation of the true Fourier coefficients discussed in section 5.3.1. Next we obtain the Ito derivative of γ, as a function of $e^{-\gamma}$ (above) as

$$d\gamma = q\, dt - \frac{1}{2r}\,(1-e^{-\gamma})^2\, dz_Q - \frac{e^{\gamma}}{4r}(1-e^{-\gamma})^5 dt \qquad (5.55)$$

For small $P_{\theta\ell}$, we expect the phase estimate to closely approximate the actual phase. Therefore, to order ε (for $\varepsilon = \theta-\hat{\theta}$)

$$\sin\varepsilon = \varepsilon$$

$$\cos\varepsilon = 1$$

Thus, the quadrature measurement (dz_Q in equation 5.54) contains no phase information. Since, for low noise, the variance of the dn_Q term also becomes negligible (with zero mean), we may assume that

$$dz_Q \simeq 1\ dt$$

and that the γ from equation 5.55 (a random variable) is "quite close" to a γ_d from the deterministic equation (cf. equation 5.55)

$$\dot{\gamma}_d = q - \frac{(1 - e^{-\gamma_d})^2}{2r} - \frac{e^{\gamma_d}}{4r} (1 - e^{-\gamma_d})^5 \tag{5.56}$$

The meaning of "quite close" (above) is uncertain. We have been unable to show the mean-square (or any other) convergence of γ to γ_d because of the nonlinear differential equations involved (5.55 and 5.56). Nonetheless, we believe this discussion, while not rigorous, provides useful insight into the low-noise operation of the FCF.

Now, since we expect γ_d to be small for small $P_{\theta\ell}$, we let

$$e^{\pm\gamma_d} = 1 \pm \gamma_d \tag{5.57}$$

in equation 5.56 to obtain

$$\dot{\gamma}_d = q - \frac{\gamma_d^2}{2r} - \frac{(1+\gamma_d)}{4r} \gamma_d^5 \tag{5.58}$$

or, to order γ_d^2

$$\dot{\gamma}_d = q - \frac{\gamma_d}{2r} \tag{5.59}$$

which is precisely the equation satisfied by $P_{\theta\ell}$ (see equations 1.51, 1.52, and 5.9). Thus, for small $P_{\theta\ell}$

$$\gamma_d = P_{\theta\ell} \tag{5.60}$$

and γ is reasonably close to γ_d.

Next, we examine the filter structure for small $P_{\theta\ell}$. We let

$$e^{-2\gamma_d} = 1-2\gamma_d \tag{5.61}$$

and, using equation 5.57 and 5.61 (with $\gamma=\gamma_d$), the FCF gains (equations 5.51 a-c) become

$$K_1 = \gamma(1-a_1^2) \tag{5.62 a}$$

$$K_2 = -a_1 b_1 \gamma \tag{5.62 b}$$

$$K_3 = \gamma(1-b_1^2) \tag{5.62 c}$$

where

$$a_1^2 + b_1^2 = e^{-\gamma} \approx 1-\gamma \tag{5.63}$$

The gains above become, to order γ,

$$K_1 = \gamma b_1^2 \tag{5.64 a}$$

$$K_2 = -a_1 b_1 \gamma \tag{5.64 b}$$

$$K_3 = \gamma a_1^2 \tag{5.64 c}$$

Thus, the FCF becomes (using equations 5.50, 5.60, 5.64 a-c and $\gamma=\gamma_d$)

$$d\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = -\frac{q}{2}\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} dt + \frac{P_{\theta\ell}}{2r}\begin{pmatrix} b_1^2 & -a_1 b_1 \\ -a_1 b_1 & a_1^2 \end{pmatrix}\left[dz - \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}\right] dt \tag{5.65}$$

By making the obvious associations

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{5.66}$$

we see that the PLL (equation 5.12 with K from equation 5.8) and the FCF (equation 5.65) are identical for small $P_{\theta\ell}$.

The convergence of the FCF to the PLL as $P_{\theta\ell}\to 0$ is satisfying, since the FCF performance is good for large $P_{\theta\ell}$ as well. The only drawback to the FCF is its complexity when compared to the simple static phase filters. In addition to the complexity of the gains themselves, the filter states mulitiplying the measurements require Wong-Zakai correction terms in implementation or digital simulation of the FCF (see appendix C). Since the filter gains (equations 5.49 and 5.51) are composed of terms like a_1^4 and b_1^4, the correction terms will be of the order of a_1^7 and b_1^7, and the FCF becomes significantly more difficult to implement than the static phase filters.

5.5 First-Cumulant Filter

5.5.1 General Design

Using the approximation method of Chapter 4, we now develop a sub-optimal filter for the first-order PLL problem. We begin by considering the log of the Radon-Nikodym derivative (Λ_t) given by equation (4.5)

$$\zeta_t = \int_0^t h^T R^{-1} dz_s - \frac{1}{2}\int_0^t h^T R^{-1} h\, ds$$

For the baseband measurements (equation 5.3) this becomes

$$\zeta_t = \frac{1}{2r}\int_0^t \begin{pmatrix}\sin\theta_s\\ \cos\theta_s\end{pmatrix}^T \begin{pmatrix}dz_1\\ dz_2\end{pmatrix} - \frac{1}{4r}\int_0^t (\sin^2\theta_s + \cos^2\theta_s)\, ds \qquad (5.67)$$

or

$$\zeta_t = \frac{1}{2r} \int_0^t \begin{pmatrix} \sin \theta_s \\ \cos \theta_s \end{pmatrix}^T dz_s - \frac{t}{4r} \tag{5.68}$$

Now since any term in ζ that is not a function of θ can be factored out of the numerator and denominator of the conditional density (equation 3.53), we may write the conditional density for <u>all</u> baseband measurement problems as

$$p(\theta, t | z_0^t) = \frac{\left(E_0^{\mathcal{B}} e^{\zeta_t'}\right) p(\theta,t)}{E_0^{\mathcal{A}_{zt}} e^{\zeta_t'}} \tag{5.69}$$

where

$$\zeta_t' = \frac{1}{2r} \int_0^t (\sin \theta_s \, dz_{1_s} + \cos \theta_s \, dz_{2_s}) \tag{5.70}$$

For the first cumulant (and moment) filters (as in section 4.2.2) we need

$$E_0^{\mathcal{B}} \zeta_t' = \frac{1}{2r} \int_0^t \hat{E}\left[\begin{pmatrix} \sin \theta_s \\ \cos \theta_s \end{pmatrix} \middle| \theta_t\right]^T dz \tag{5.71}$$

The forward transition density for θ_t is

$$p(\theta_t | \theta_s) = N(\theta_s, q(t-s)) \tag{5.72}$$

This density may be "folded" to obtain the modulo-2π phase density

$$p(\theta_t | \theta_s) = \frac{1}{2\pi} [1 + 2 \sum_{n=1}^{\infty} e^{-\frac{n^2 q}{2}(t-s)} \cos n(\theta_t - \theta_s)] \tag{5.73}$$

If we let the a priori density be uniform

$$p(\theta_t) = p(\theta_s) = \frac{1}{2\pi} \qquad -\pi < \theta_{\substack{s\\t}} \leq \pi \tag{5.74}$$

then from Bayes' rule

$$p(\theta_s|\theta_t) = \frac{p(\theta_t|\theta_s)p(\theta_s)}{p(\theta_t)} = p(\theta_t|\theta_s) \tag{5.75}$$

Therefore, we can evaluate the expectation in equation 5.71 as

$$E\left[\begin{pmatrix}\sin\theta_s\\ \cos\theta_s\end{pmatrix}\middle|\theta_t\right] = e^{\frac{-q(t-s)}{2}}\begin{pmatrix}\sin\theta_t\\ \cos\theta_t\end{pmatrix} \tag{5.76}$$

so that

$$E_0^{\mathcal{B}}\,\zeta_t' = x_1 \sin\theta_t + y_1 \cos\theta_t \tag{5.77}$$

where

$$\begin{pmatrix}x\\ y\end{pmatrix} = \frac{1}{2r}\int_0^t e^{-\frac{q}{2}(t-s)}\begin{pmatrix}dz_{1_s}\\ dz_{2_s}\end{pmatrix} \tag{5.78}$$

or

$$d\begin{pmatrix}x\\ y\end{pmatrix} = -\frac{q}{2}\begin{pmatrix}x\\ y\end{pmatrix}dt + \frac{1}{2r}\begin{pmatrix}dz_1\\ dz_2\end{pmatrix} \tag{5.79}$$

We form an approximate density by using the first cumulant approximation

$$E_0^{\mathcal{B}}\, e^{\zeta_t'} \simeq e^{E_0^{\mathcal{B}}\,\zeta_t'} \tag{5.80}$$

the a priori density (equation 5.74), and the normalizing denominator

$$\int_{-\pi}^{\pi} e^{E_0^{\beta} \zeta_t'} \frac{1}{2\pi} \, d\theta_t$$

to obtain

$$p(\theta,t|z_0^t) \approx \frac{e^{x \sin \theta_t + y \cos \theta_t}}{2\pi \, I_0 \sqrt{x^2 + y^2}} \tag{5.81}$$

We obtain a phase estimate from this density

$$E\left[\begin{pmatrix} \sin \theta_t \\ \cos \theta_t \end{pmatrix} \middle| z_0^t \right] = \frac{I_1(\alpha)}{I_0(\alpha)} \begin{pmatrix} \sin \beta \\ \cos \beta \end{pmatrix} \tag{5.82}$$

where

$$\alpha = \sqrt{x^2 + y^2} \tag{5.83}$$

$$\beta = \tan^{-1}(x/y) \tag{5.84}$$

so that

$$\hat{\theta} = \tan^{-1} \frac{\left[I_1(\alpha)/I_0(\alpha)\right] \sin \beta}{\left[I_1(\alpha)/I_0(\alpha)\right] \cos \beta} = \beta \tag{5.85}$$

$$= \tan^{-1}(x/y) \tag{5.86}$$

This estimate is the same as that obtained by the static phase filter in equations 5.25 and 5.26, and the filters are therefore identical. This filter, as a static phase filter and because of the cumulant interpretation of the approximate density, converges to the known optimal filter for the case where q=0.

5.5.2 <u>Sub-Optimal Filter Comparison</u>

We now compare the performance of the first cumulant (static phase) filter to that of other suboptimal designs. The Fourier coefficient (FCF), state-dependent noise (LQF), first cumulant (approximate density filter or APDF), and phase-lock loop (PLL) filter were all simulated. We used a fourth-order Runge-Kutta integration routine, with a time step of 1/100 th of the PLL time constant (1/K), to minimize the effect of the discretization. We ran the filters for four runs of 500 time constants each and discarded the data from the first twenty-five time constants (in each run) to avoid "start-up" transients. This resulted in 1900 "effective" degrees of freedom, as in [15, 29, 39], for a 3% predicted standard deviation in the computed error variances (see [15] and [39]). The same pseudo-noise sequences were used for all of the filters.

We compared the filters for $P_{\theta\ell} = 1$ rad.^{2} (q=1, r=1/2), where the PLL degradation (over its linear-predicted performance) is near maximum.* The computed averages for the error-squared and cosine cost functions are listed in Table 5.1, along with the percent improvement in each filter with respect to the PLL. As an indication of the optimal filter performance, the results of Tam and Moore [39] for their 6-density filter and Mallinckrodt, et. al. [29], for their point-mass filter are included,

*See Van Trees [40] and section 5.2.

along with the per cent improvement relative to their PLL simulations. The PLL linear-predicted performance, exact performance (using the density of equation 5.13), and simulated performance are included for comparison.

	$E[\varepsilon^2]$ (rad.^{2})		$E[1 - \cos\varepsilon]$	
	Actual	% Improvement	Actual	% Improvement
PLL: Lin Pred.	1.0	—	.393	—
Exact	1.604	—	.554	—
Simulated	1.648	(Ref.)	.567	(Ref.)
FCF	1.437	12.8%	.506	10.8%
LQF	1.456	11.7%	.511	9.9%
APDF	1.498	9.1%	.525	7.4%
Results from [39]				
PLL Simulated	1.586	(Ref.)	.547	(Ref.)
Optimal	1.374	13.4%	.490	10.8%
Results From [29]				
PLL Simulated	1.614	(Ref.)	N/A	
Optimal	1.395	13.6%	N/A	

Table 5.1 Sub-Optimal Filter Performance Comparison

These results indicate that the Fourier coefficient filter performs marginally better than the LQF, and that both are much better (considering the possible improvement) than the PLL. The APDF performs slightly worse than the LQF and FCF, but also much better than the PLL. The APDF performance is encouraging; it means that the first cumulant filter outperforms the extended Kalman filter for this problem.

At this point we are led to consider two questions: Can the static phase filter performance be analyzed, and can it be improved upon? The rest of this chapter trys to answer both of these questions. We conclude this section on the first cumulant filter with a demonstration of one of the bounding techniques discussed in chapter four.

5.5.3 Accuracy of First-Cumulant Approximation

In this section we use the cumulant remainder to bound the cumulant approximation as mentioned in section 4.33. Our goal is to demonstrate how it is possible to use some of the bounding methods even when the expectations required cannot be evaluated.

We recall that

$$E_0\Delta^2 \leq E_0[e^{2\xi} - 2e^{\xi} + 1] \tag{5.87}$$

where

$$\Delta = \frac{E_0^{\mathcal{B}} e^{\zeta} - e^{E_0^{\mathcal{B}}\zeta}}{e^{E_0^{\mathcal{B}}\zeta}} \tag{5.88}$$

and

$$\xi = \zeta - E_0^{\mathcal{B}}\zeta \tag{5.89}$$

For this problem, we may replace ζ by ζ' and ξ becomes

$$\xi = \frac{1}{2r}\int_0^t \left[\begin{pmatrix}\sin\theta_s\\ \cos\theta_s\end{pmatrix} - e^{-\frac{q}{2}(t-s)}\begin{pmatrix}\sin\theta_t\\ \cos\theta_t\end{pmatrix}\right]^T dz \tag{5.90}$$

We see that

$$E_0^{A_{\theta t}} \xi = 0 \tag{5.91}$$

and

$$E_0^{A_{\theta t}} \xi^2 = \frac{1}{2r} \int_0^t \left\{ \left[1 + e^{-q(t-s)}\right] - 2e^{-\frac{q}{2}(t-s)} \cos(\theta_t - \theta_s) \right\} ds \tag{5.92}$$

because under P_0

$$E[dz\ dz^T] = \begin{pmatrix} 2r & 0 \\ 0 & 2r \end{pmatrix} dt \tag{5.93}$$

Thus, we may write

$$E_0^{A_{\theta t}} \xi^2 \leq \sigma_m^2 \tag{5.94}$$

where

$$\sigma_m^2 = \frac{t}{2r} + \frac{1}{2rq}[1 - e^{-qt}] + \frac{2}{rq}[1 - e^{-qt/2}] \tag{5.95}$$

where we have bounded the cosine in equation 5.92 and taken the integral. We may now write

$$E_0^{A_{\theta t}} \xi^2 = \sigma^2 \leq \sigma_m^2 \tag{5.96}$$

and then

$$E_0 \Delta^2 \leq E_0 E_0^{A_{\theta t}} [e^{2\xi} - 2e^{\xi} + 1] \tag{5.97}$$

Now since ξ conditioned on $A_{\theta t}$ is a Gaussian random variable with zero mean (equation 5.91) and variance σ^2 (equation 5.92 and 5.96), the right-hand side in equation 5.97 equals

$$E_0[e^{2\sigma^2} - 2e^{\sigma^2/2} + 1]$$

which is a monotonically increasing function of σ^2. Thus

$$E_0\, \Delta^2 \leq 1 + e^{2\sigma_m^2} - 2e^{-\sigma_m^2/2} \tag{5.98}$$

where σ_m^2 is given by equation 5.95.

For very small t

$$\sigma_m^2 \simeq \frac{2t}{r}$$

and

$$1 + e^{2\sigma_m^2} - 2e^{\sigma_m^2/2} \simeq \sigma_m^2 \tag{5.99}$$

This bound is hard to evaluate qualitatively. Clearly, for very small t, the cumulant approximation will be a very good one, and we may place high confidence in the filter. For larger t, the bound becomes larger, but the filter performance is still very good, as shown by the simulations.

Thus, despite the lack of a tight bound, we know that the first cumulant filter performs well, and we are led to examine other ways of predicting and analyzing that performance.

5.6 Static Phase Filter Performance

5.6.1 Approximate-Density Interpretation

This section investigates several aspects of static phase filter performance. Unlike the density accuracy that the last section considered, the filter performance (error variance or expected cosine cost function) is the subject here. While we are concerned here with static phase filters in general, we begin by recalling that the first cumulant filter creates the approximate density

$$p(\theta,t|z_0^t) = \frac{e^{x \sin \theta_t + y \cos \theta_t}}{2\pi I_0(\sqrt{x^2 + y^2})} \tag{5.100}$$

Using this density, the conditional cosine cost function becomes

$$E[1 - \cos(\theta - \hat{\theta})|z_0^t] = 1 - \frac{I_1(\sqrt{x^2 + y^2})}{I_0(\sqrt{x^2 + y^2})} \tag{5.101}$$

Thus, the filter states x and y may be used to provide an estimate of how well the filter is tracking. The accuracy of this estimate, of course, depends on the accuracy of the cumulant approximation, which, as discussed in the last section, is hard to determine. We now consider a similar performance indication for a general static phase filter that does not have this limitation.

5.6.2 Exact Performance

This section presents a technique for obtaining the actual value of the cosine cost function for a general static phase filter. We consider the nonlinear, two-state error equations

$$d\varepsilon = \frac{1}{2r\alpha}(\sin \varepsilon \, dt + dn_I) + du \tag{5.102}$$

$$d\alpha = [-f\alpha + \frac{1}{4r\alpha}]dt + \frac{1}{2r}[\cos \varepsilon \, dt + dn_Q] \tag{5.103}$$

for

$$\varepsilon = \theta - \hat{\theta}$$

which we obtain from equations 5.1, 5.31, and 5.32.

Now by averaging over the noise dn_Q

$$d\overline{\alpha} = [-f\overline{\alpha} + \frac{1}{4r}\overline{(\frac{1}{\alpha})}]dt + \frac{1}{2r}\overline{\cos \varepsilon}\, dt \tag{5.104}$$

where

$$\overline{(\ )} = E(\ \ )$$

and assuming that $\overline{\alpha}$ goes to a steady-state*, we have that

$$1-\overline{\cos \varepsilon} = 1-2rf\overline{\alpha} + \frac{1}{2}\overline{(\frac{1}{\alpha})} \tag{5.105}$$

This means that by computing an average value for α and $1/\alpha$ in actual filter operation, it is possible to obtain

$$1 - \overline{\cos \varepsilon}$$

through equation 5.105 without using the actual phase θ. The next section demonstrates a useful application of this result.

5.6.3 Filter Behavior When Signal Lost

We now examine the behavior of a static phase filter when there is no signal present, that is, when

$$\begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} = \begin{pmatrix} dn_1 \\ dn_2 \end{pmatrix} \tag{5.106}$$

This situation might arise if the actual currier frequency were much dif-

---

*We have been unable to completely justify this assumption, but "evidence" that the density of α approaches a steady-state is given in appendix D.

ferent from the heterodyning frequency used to obtain the baseband measurements. In this case

$$d\begin{pmatrix} x \\ y \end{pmatrix} = -f\begin{pmatrix} x \\ y \end{pmatrix} dt + \frac{1}{2r}\begin{pmatrix} dn_1 \\ dn_2 \end{pmatrix} \tag{5.107}$$

which describes a Gaussian system with zero mean and a steady-state variance of

$$E\left[\begin{pmatrix} x \\ y \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}^T\right] = \begin{pmatrix} v & 0 \\ 0 & v \end{pmatrix} \tag{5.108}$$

where

$$v = \frac{1}{4rf} \tag{5.109}$$

Since x and y are zero mean, independent Gaussian random variables, α is Rayleigh distributed, with

$$p(\alpha) = \frac{\alpha}{v} e^{-\alpha^2/2v} \qquad \alpha \geq 0 \tag{5.110}$$

so that

$$E[\alpha] = \sqrt{\frac{\pi}{2}}\ \sqrt{v} \tag{5.111}$$

and

$$E\left[\frac{1}{\alpha}\right] = \sqrt{\frac{\pi}{2}}\ \frac{1}{\sqrt{v}} \tag{5.112}$$

with v given by equation 5.109 for any f.

We note that if one uses the performance estimating technique of equation 5.105, then

$$1 - \overline{\cos \varepsilon} = 1 + \sqrt{\frac{\pi}{2}}\ \frac{1}{2}\left[\frac{1}{\sqrt{v}} - 4rf\sqrt{v}\right] = 1 \tag{5.113}$$

Thus the static phase filter can detect a signal loss using the α and $1/\alpha$ averages suggested by equation 5.105.

5.6.4 Low-Noise Filter Performance

The two-state error equations 5.102 and 5.103 may be used to obtain low-$P_{\theta\ell}$ convergence results, as was done in [15] for the LQF. For very small $P_{\theta\ell}$, the variance of dn_Q becomes negligible, and

$$\cos \varepsilon \simeq 1 \tag{5.114}$$

so that equation 5.103 may be written

$$\dot{\alpha} = -f\alpha + \frac{1}{4r\alpha} + \frac{1}{2r} \tag{5.115}$$

which becomes, in steady-state

$$f\alpha - \frac{1}{4r\alpha} = \frac{1}{2r} \tag{5.116}$$

The error equation (5.102) becomes (where $\sin \varepsilon \simeq \varepsilon$)

$$\dot{\varepsilon} = -\frac{1}{2r\alpha}(\varepsilon + \dot{n}_I) + \dot{u} \tag{5.117}$$

For any constant α, the error variance is minimized (as was done for the PLL by the optimal Kalman filter gain K, in equation 1.53) by choosing

$$\frac{1}{2r\alpha^0} = K = \frac{P_{\theta\ell}}{2r} \tag{5.118}$$

or

$$\alpha^0 = \frac{1}{P_{\theta\ell}} \tag{5.119}$$

This means the optimal f (from equation 5.116) is

$$f^0 = \frac{P_{\theta\ell}}{2r} + \frac{P_{\theta\ell}^2}{4r} \tag{5.120}$$

which becomes

$$f^0 = \sqrt{\frac{q}{2r}} + \frac{q}{2} \tag{5.121}$$

As was mentioned in section 5.4.2, in [15] Gustafson and Speyer found that the LQF had a 6% degradation in error variance over the PLL as $P_{\theta\ell} \to 0$. The static phase filter with gain f^0, however, will exactly approach the PLL as $P_{\theta\ell} \to 0$.* Unfortunately, this filter (with gain f^0) does not perform as well as the LQF static phase filter for high $P_{\theta\ell}$.

We simulated the LQF, APDF, and static phase filter with gain f^0 on one of the four 500 time-constant noise sequences used to evaluate the sub-optimal filters in section 5.5.2. The results are listed in Table 5.2. The absolute accuracy of these numbers is not as great as that of

---

*We note that the LQF was designed to have the lowest error variance (of any static phase filter) in the measurement space, but not, apparently, in the phase space.

table 5.1, but the relative ratings should be reliable.

	$E[\varepsilon^2]$	$E[1 - \cos \varepsilon]$
PLL	1.9728	.6462
APDF	1.5869	.5483
LQF	1.5277	.5297
f^0 Filer	1.5950	.5454

Table 5.2 Static Phase Filter Comparison

These results indicate that the f^0 filter performs worse than the LQF but about the same as the APDF for $P_{\theta\ell} = 1$ rad.^{2} Apparently none of the gains (f, as a function of r and q) provides optimal performance for all $P_{\theta\ell}$, and we are led to consider modifications of the static phase filter to improve performance.

Two such modifications are considered in this chapter. The first uses the higher-order density approximations of the last chapter to improve the accuracy of the density shape. The second uses a modified, nonlinear damping gain $f(\alpha)$ in place of the constant f of the static phase filters.

5.7 Higher-Order Density Approximations

5.7.1 First-Moment Information

Now we investigate the advantages and disadvantages of adding more terms in our density approximation. We consider densities using the first, second and third moments of ζ_t', combined in a variety of ways, to create cumulant, central moment, and moment approximate densities (see section 4.2.2).

We recall that the first moment of ζ_t' is

$$E_0^{\beta}\, \zeta_t' = x_1 \sin \theta_t + y_1 \cos \theta_t \tag{5.122}$$

for

$$\zeta_t' = \frac{1}{2r} \int_0^t (\sin \theta_s \, dz_{1_s} + \cos \theta_s \, dz_{2_s}) \tag{5.123}$$

and

$$d \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = -\frac{q}{2} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} dt + \frac{1}{2r} \begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \tag{5.124}$$

This was used (see section 5.5.1) to create the first cumulant approximate density

$$p(\theta,t|z_0^t) = \frac{e^{\alpha_1 \cos(\theta-\beta_1)}}{2\pi I_0(\alpha_1)} \tag{5.125}$$

for

$$\alpha_1 = \sqrt{x_1^2 + y_1^2} \tag{5.126}$$

$$\beta_1 = \tan^{-1} x_1/y_1 \tag{5.127}$$

We note that the Fourier series of this density is

$$p(\theta,t|z_0^t) = \frac{1}{2\pi} \left[1 + 2 \sum_{n=1}^{\infty} \frac{I_n(\alpha_1)}{I_0(\alpha_1)} \cos n(\theta-\beta_1) \right] \tag{5.128}$$

(see the Bessel function appendix A), and the optimal phase estimate*

---

*We use the term "optimal phase estimate" to denote the Stratonovich estimate, with respect to the approximate conditional density, as discussed in section 5.3.1.

$$\hat{\theta} = \tan^{-1} \frac{I_1(\alpha_1)/I_0(\alpha_1)\sin\beta_1}{I_1(\alpha_1)/I_0(\alpha_1)\ \cos\beta_1} = \beta_1 \tag{5.129}$$

The first moment may also be used in a moment approximation, where

$$E_0^{\mathcal{B}}\, e^{\zeta_t'} \simeq 1 + E_0^{\mathcal{B}}\, \zeta_t' \tag{5.130}$$

which leads to an approximate density (not necessarily positive) of the form

$$p(\theta, t\,|\,z_0^t) = \frac{1}{2\pi}\,[1 + 2\alpha_1 \cos(\theta-\beta_1)] \tag{5.131}$$

which is a finite Fourier series with optimal phase estimate

$$\hat{\theta} = \tan^{-1}(x/y) = \beta_1 \tag{5.132}$$

Thus the first moment and cumulant filters lead to different density functions but the <u>same</u> phase estimate, and therefore the same performance.

5.7.2 Second-Moment Filters

We now investigate the second moment of ζ_t'. We have that

$$E_0^{\mathcal{B}}(\zeta_t')^2 = \frac{1}{4r^2}\int_0^t\int_0^t E_0^{\mathcal{B}}(\sin\theta_\tau dz_{1_\tau} + \cos\theta_\tau d_{z_{2\tau}})(\sin\theta_s dz_{1s}+\cos\theta_s dz_{2s}) \tag{5.133}$$

Using the technique of section 4.2.1 and trigonometric sum and difference formulae, we may rewrite this moment as

$$E_0^{\mathcal{B}}(\zeta_t')^2 = \frac{1}{8r^2}\int_0^t\int_0^t \Big\{ E^{A(\theta_t)}[\cos(\theta_\tau-\theta_s)-\cos(\theta_\tau+\theta_s)]dz_{1\tau}dz_{1s} \tag{5.134}$$

$$+ E^{A(\theta_t)}[\sin(\theta_\tau+\theta_s) + \sin(\theta_s-\theta_\tau)]dz_{1s}dz_{2\tau}$$

$$+ E^{A(\theta_t)} [\sin (\theta_\tau+\theta_s) + \sin (\theta_\tau-\theta_s)] dz_{1\tau}\, dz_{2s}$$

$$+ E^{A(\theta_t)} [\cos (\theta_\tau+\theta_s) + \cos (\theta_\tau-\theta_s)] dz_{2\tau}\, dz_{2s}$$

We recall (equation 5.75) that θ_s given θ_t is a "backward" Brownian motion with mean θ_t and variance $q(t-s)$. Thus ϕ_- and ϕ_+, conditioned on θ_t, are Brownian motions with

$$\phi_- = \theta_\tau - \theta_s \tag{5.135}$$

$$\phi_+ = \theta_\tau + \theta_s \tag{5.136}$$

and

$$\phi_- \sim N(0,\ q[2 \max (\tau,s) - (\tau+s)]) \tag{5.137}$$

$$\phi_+ \sim N(2\theta_t,\ q[4t - (\tau+s) - 2 \max (\tau, s)] \tag{5.138}$$

Therefore

$$E\left[\begin{pmatrix}\sin \phi_- \\ \cos \phi_-\end{pmatrix} \Big| \theta_t\right] = e^{-\frac{q}{2}[2 \max (\tau, s) - (\tau+s)]} \begin{pmatrix}0 \\ 1\end{pmatrix} \tag{5.139}$$

and

$$E\left[\begin{pmatrix}\sin \phi_+ \\ \cos \phi_+\end{pmatrix} \Big| \theta_t\right] = e^{-\frac{q}{2}[4t - (\tau+s) - 2 \max (\tau, s)]} \begin{pmatrix}\sin 2\theta_t \\ \cos 2\theta_\tau\end{pmatrix} \tag{5.140}$$

where we have used the folded-normal density as in section 5.5.1 to evaluate the expectations.

This means that

$$E_0^{\mathcal{B}}(\zeta_t')^2 = \frac{1}{8r^2} \int_0^t \int_0^t \left\{ e^{-\frac{q}{2}[2 \max(\tau, s) - (\tau+s)]} (dz_{1z}dz_{1s}+dz_{2\tau}dz_{2s}) \right. \tag{5.141}$$

$$+ \cos 2\theta_t \left[e^{-\frac{q}{2}[4t-(\tau+s)-2 \max(\tau,s)]} (dz_{2\tau}dz_{2s}-dz_{1\tau}dz_{1s}) \right]$$

$$\left. + \sin 2\theta_t \left[e^{-\frac{q}{2}[4t-(\tau+s)-2 \max(\tau,s)]} (dz_{1\tau}dz_{2s}-dz_{2\tau}dz_{1s}) \right] \right\}$$

which may be written

$$E_0^{\mathcal{B}}(\zeta_t')^2 = c + x_2 \sin 2\theta_t + y_2 \cos 2\theta_t \tag{5.142}$$

where

$$dc = \frac{1}{2r}(x_1\, dz_1 + y_1\, dz_2) \tag{5.143}$$

$$dx_2 = -2qx_2 dt + \frac{1}{2r}(x_1 dz_2 + y_1 dz_1) \tag{5.144}$$

$$dy_2 = -2qy_2 dt + \frac{1}{2r}(y_1 dz_2 - x_1 dz_1) \tag{5.145}$$

and x_1 and y_1 are given by equation 5.124. We note that Wong-Zakai correction terms are required for simulating or implementing these equations. However, because of the independence of n_1 and n_2, $\dot{x}_2$ needs no correction term and the terms for $\dot{y}_2$ cancel, so that only $\dot{c}$ needs a correction (of $-1/2r$). Thus the filter states to be implemented become

$$\dot{c} = \frac{1}{2r}(x_1\dot{z}_1 + y_1\dot{z}_2) - \frac{1}{2r} \tag{5.146}$$

$$\dot{x}_2 = -2qx_2 + \frac{1}{2r}(x_2\dot{z}_2 + y_1\dot{z}_1) \tag{5.147}$$

$$\dot{y}_2 = -2qy_2 + \frac{1}{2r}(y_1\dot{z}_2 - x_1\dot{z}_1) \tag{5.148}$$

This new information may be combined with

$$(E_0^{\mathcal{B}}\ \zeta_t')^2 = \frac{x_1^2 + y_1^2}{2} + x_1 y_1 \sin 2\theta_t + \frac{y_1^2 - x_1^2}{2} \cos 2\theta_t \tag{5.149}$$

to evaluate the conditional variance of ζ_t' as

$$E_0^{\mathcal{B}}\ (\zeta_t' - E_0^{\mathcal{B}}\zeta_t')^2 = V_1 + V_2 \sin 2\theta_t + V_3 \cos 2\theta_t \tag{5.150}$$

where

$$V_1 = c - \frac{x_1^2 + y_1^2}{2} \tag{5.151}$$

$$V_2 = x_2 - x_1 y_1 \tag{5.152}$$

$$V_3 = y_2 - \frac{(y_1^2 - x_1^2)}{2} \tag{5.153}$$

Then the numerator of the density in the second cumulant approximation may be written as

$$U_t' = e^{\alpha_1 \cos(\theta - \beta_1) + \frac{\alpha_2}{2} \cos 2(\theta - \beta_2)} \tag{5.154}$$

where α_1 and β_1 are given by equations 5.126 and 5.127 and

$$\alpha_2 = \sqrt{V_2^2 + V_3^2} \tag{5.155}$$

$$\beta_2 = \frac{1}{2} \tan^{-1} V_2/V_3 \tag{5.156}$$

(Since V_1 does not multiply a function of θ_t, it may be factored out of the numerator.) Unfortunately, this function (equation 5.154) of θ is very difficult to analyze. It is an exponential Fourier series* and we

*See Lo [28] for a discussion of these series.

need the <u>regular</u> Fourier coefficients

$$E\left\{\begin{pmatrix}\sin\theta\\ \cos\theta\end{pmatrix}\Big| z_0^t\right\}$$

for our optimal estimate. Since no straightforward way to evaluate the integral

$$\int_{-\pi}^{\pi}\begin{pmatrix}\sin\theta\\ \cos\theta\end{pmatrix} U_t' \, d\theta$$

exists, we attempted instead to search for the approximate mode of U_t' as our phase estimate.

We define

$$F(\theta) = \alpha_1 \cos(\theta-\beta_1) + \frac{\alpha_2}{2} \cos 2(\theta-\beta_2) \tag{5.157}$$

and take

$$\frac{\partial F}{\partial\theta}\Big|_{\theta=\hat\theta} = -\alpha_1 \sin(\hat\theta-\beta_1) - \alpha_2 \sin 2(\hat\theta-\beta_2) = 0 \tag{5.158}$$

Now by assuming that $\hat\theta \simeq \beta_1$, we let

$$\sin(\hat\theta-\beta_1) \simeq \hat\theta-\beta_1 \tag{5.159}$$

$$\sin 2(\hat\theta-\beta_1) \simeq 2(\hat\theta-\beta_1) \tag{5.160}$$

$$\cos 2(\hat\theta-\beta_1) \simeq 1 \tag{5.161}$$

so that equation 5.158 may be solved for

$$\hat\theta = \beta_1 - \frac{\alpha_2 \sin 2(\beta_1-\beta_2)}{\alpha_1 + 2\alpha_2 \cos 2(\beta_1-\beta_2)} \tag{5.162}$$

A filter using this estimate (called A2C) was simulated (see section 5.7.4) with poor results. The added density information in α_2 and β_2 did not result in a better phase estimate after our approximations (equations 5.159 - 5.161) were made.

There are other ways to use the second moment information, however. The moment approximation for the numerator would be

$$E_0^{\mathcal{B}} e^{\zeta_t'} \simeq 1 + E_0^{\mathcal{B}}\zeta_t' + \frac{1}{2} E_0^{\mathcal{B}}(\zeta_t')^2 \tag{5.163}$$

where the right-hand side equals

$$1 + \frac{c}{2} + x_1 \sin\theta + y_1 \cos\theta + \frac{x_2}{2}\sin 2\theta + \frac{y_2}{2}\cos 2\theta$$

While the normalization will make this density

$$p(\theta,t|z_0^t) = \frac{1}{2\pi}[1 + a_1 \sin\theta + b_1 \cos\theta + a_2 \sin 2\theta + b_2 \cos 2\theta] \tag{5.164}$$

for

$$a_1 = x_1/(1+c/2) \tag{5.165}$$

$$b_1 = y_1/(1+c/2) \tag{5.166}$$

we see that the phase estimate, which depends on a_1/b_1, will be unchanged from the first moment filter. Thus, the second moment changes the density shape but not the phase estimate for the moment approximations.

A third way to use the second moment is in the central moment approximation, where

$$E_0^{\mathcal{B}}\, e^{\zeta'_t} \simeq e^{E_0^{\mathcal{B}}\zeta'_t}\,[1 + \frac{1}{2} E_0^{\mathcal{B}}[\zeta'_t - E_0^{\mathcal{B}}\zeta'_t]^2] \tag{5.167}$$

Here, the right-hand side becomes

$$e^{\alpha_1 \cos\,(\theta-\beta_1)}\,[1 + \frac{V_1}{2} + \frac{\alpha_2}{2}\;\cos 2(\theta-\beta_2)]$$

where V_1, α_2 and β_2 are the same as those in the cumulant approximation.

This results in an approximate density function of the form

$$p(\theta,t|z_0^t) = \frac{e^{\alpha_1 \cos\,(\theta-\beta_1)}\,[A + \frac{\alpha_2}{2} \cos 2(\theta-\beta_2)]}{2\pi[I_0(\alpha_1)A + \frac{\alpha_2}{2} I_2(\alpha_1)\;\cos 2(\beta_1-\beta_2)]} \tag{5.168}$$

for

$$A = 1 + \frac{V_1}{2} = 1 + \frac{c}{2} - \frac{x_1^2 + y_1^2}{4} \tag{5.169}$$

The phase estimate for this density is equal to

$$\hat{\theta} = \tan^{-1} a_1/b_1 \tag{5.170}$$

where

$$a_1 = AI_1(\alpha_1)\sin\beta_1 + \frac{\alpha_2}{4}(I_3(\alpha_1)\;\sin\Delta_1 + I_1(\alpha_1)\;\sin\Delta_2) \tag{5.171}$$

$$b_1 = AI_1(\alpha_1)\cos\beta_1 + \frac{\alpha_2}{4}(I_3(\alpha_1)\;\cos\Delta_1 + I_1(\alpha_1)\;\cos\Delta_2) \tag{5.172}$$

$$\Delta_1 = 3\beta_1 - 2\beta_2 \tag{5.173}$$

and

$$\Delta_2 = 2\beta_2 - \beta_1 \tag{5.174}$$

This filter, after a start-up transient, performed identically to the first cumulant filter, as shown in section 5.7.4 (where this filter is called 2CM). The explanation is that c (in the second moment) grew rapidly with time, so that the filter (through A) put increasing weight on the first cumulant terms (those due to α_1 and β_1) in equations 5.171 and 5.172.

5.7.3 Third-Moment Designs

We also examined the effects of the third moment $E_0^{\mathcal{B}}(\zeta_t')^3$ on the moment and cumulant approximate densities. The third moment (see appendix E) is of the form

$$E_0^{\mathcal{B}}(\zeta_t')^3 = 6\,[x_s \sin\theta + x_c \cos\theta + x_3 \sin 3\theta + y_3 \cos 3\theta] \qquad (5.175)$$

Thus, in the third moment approximate density, the third moment has components (x_s and x_c) in the first Fourier coefficients, resulting in a phase estimate of

$$\hat{\theta} = \tan^{-1} \frac{x_1 + x_s}{y_1 + x_c} \qquad (5.176)$$

In simulating this filter, it was observed that the $\dot{x}_s$ and $\dot{x}_c$ equations could be simplified by consideration of only the dominant gains (as discussed in appendix E). The simulation results reported for the third moment filter are in fact for the simplified third moment filter. As discussed in section 5.7.4, this filter slightly outperformed the first moment filter.

We next consider the effect of the third central moment on the first term in the cumulant expansion. The third cumulant approximate numerator would be of the form

$$U_t = \exp\{\alpha_1' \cos(\theta-\beta_1') + \frac{\alpha_2}{2} \cos 2(\theta-\beta_2) + \frac{\alpha_3}{6} \cos 3(\theta-\beta_3)\} \qquad (5.177)$$

where α' and β' have components from the first cumulant (α_1 and β_1) and the third central moment. By assuming that the single angle term ($\alpha_1' \cos(\theta-\beta_1')$) is dominant, we may form the approximate density

$$p(\theta,t|z_0^t) = \frac{e^{\alpha_1' \cos(\theta-\beta_1')}}{2\pi I_0(\alpha_1')} \qquad (5.178)$$

with phase estimate

$$\hat{\theta} = \beta_1' \qquad (5.179)$$

This filter (called A3C in section 5.7.4) was also simulated with generally poor results.

We remark that the third central moment could have been used in a central moment approximation, but it would not have affected the A term in equation 5.168, and hence the eventual convergence to the first cumulant filter. We now describe the actual simulation results for these higher-order density approximations.

5.7.4 Filter Comparison

We have shown how to generate approximate densities for the first-order PLL problem. The moment approximation method leads to a Fourier series representation for the conditional density. The cumulant approximation leads to a less-workable exponential Fourier series, and the central-moment approach creates a hybrid density shape. The moment and

and central moment densities may be solved for the "optimal" phase estimate, while we have found it necessary to approximate the higher-order cumulant phase estimate.

We remark that, in this problem, the optimal estimate and minimum cost are determined by the first Fourier coefficients a_1 and b_1 (see section 5.3.1). In our moment approximation method, the nth moment affects the Fourier coefficients up to order n, so that it is possible to use the first n moments in a partial-sum approximation for the first Fourier coefficients. (Actually, only the odd moments directly affect the first Fourier coefficient; the even moments affect the "constant" in the Fourier expansion, however, and enter the first coefficients through the normalization process.) Thus, it may be possible to obtain a finite-sum approximation to the optimal filter performance by examining the higher order moments in more detail.

We concentrate, instead, on examining the improvement in filter performance obtained by adding the first few moments and cumulants. The first-cumulant (APDF, also the first-moment filter), approximate-second-cumulant (A2C), second-central moment (2CM), third moment (3M), and approximate-third-cumulant (A3C) filters were simulated on one of the four 500-time-constant noise sequences used in section 5.5.2. The results are shown in Table 5.3. (We remark, as in section 5.6.4, that the absolute accuracy of the numbers in Table 5.3 is not as great as that of Table 5.1, but the relative ratings should be reliable.

	$E[\varepsilon^2]$ (rad.^{2})	$E[1-\cos\varepsilon]$
PLL	1.9728	0.6462
APDF	1.5869	0.5483
A2C	1.7923	0.6034
2CM	1.5866	0.5482
3M	1.5845	0.5478
A3C	2.4315	0.7637

Table 5.3 Higher-Order Approximate Density Filter Comparison

We see that the approximate second-and third-cumulant filters perform worse than the first-cumulant filter, the second central moment filter performs almost the same as the first cumulant, and the third-moment filter slightly outperforms APDF. To verify the third moment filter performance, we tested it on all four noise sequences, resulting in a mean-squared error of 1.495 and a mean cosine error of 0.524 to compare to the first cumulant filter's 1.498 and 0.525, respectively, from Table 5.1. These results, like those for the 2CM filter, are very close to those of the APDF, but the 3M filter, unlike 2CM, did not produce the same estimate (for each sample path) as the APDF after the start-up transients died down.

We conclude that the errors in obtaining phase estimates from the higher-order cumulant approximations outweigh the increased accuracy of the density approximation. The convergence of the second-central-moment-filter apparently means that this expansion is inappropriate for the conditional density for this problem, and that the first cumulant alone is just as good. We have been unable to determine if this convergence is inherent in the central-moment approximation method, or is only true for the first-order PLL problem.

We also note the moment approximation's slow convergence to the optimal filter (as the number of moments increases). Since the first moment improves performance about 9% over the PLL (in error variance), and the possible improvement is only about 13.5%, it is not surprising that the additional moments improve performance very slowly. In other nonlinear problems where there is more room for improvement, the benefits of more terms in the approximation should be more apparent.

5.8 Bessel-Function Filter

We see that the higher-order cumulant approximations lead to density functions from which it is difficult to obtain phase estimates, and that the moment approximations, while generating simple phase estimates, converge very slowly to the correct conditional density. We also recall that the LQF of [15] is a static phase filter, with a different gain f, which outperforms the first cumulant static phase filter. We know that the exponential cosine density form is well suited to this problem and now consider how we obtain better density approximations by modifying the gain f.

We consider a type of static phase filter with nonlinear damping $f(\alpha)$, that is, we let

$$d\begin{pmatrix} x \\ y \end{pmatrix} = -f(\alpha)\begin{pmatrix} x \\ y \end{pmatrix} dt + \frac{1}{2r}\begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \tag{5.180}$$

$$\alpha = \sqrt{x^2 + y^2} \tag{5.181}$$

$$\beta = \tan^{-1}\left(\frac{x}{y}\right) \tag{5.182}$$

and form the approximate density

$$p(\theta,t|z_0^t) = \frac{e^{\alpha\cos(\theta-\beta)}}{2\pi I_0(\alpha)} \tag{5.183}$$

We now demonstrate that by careful choice of $f(\alpha)$, we can make the differential equations for the first Fourier coefficients ($\hat{a}_1$ and $\hat{b}_1$) of our approximate density match those of the true density. For the true density, we recall, from equations 5.21 to 5.24, that

$$d\begin{pmatrix} a_n \\ b_n \end{pmatrix} = -\frac{n^2 q}{2}\begin{pmatrix} a_n \\ b_n \end{pmatrix}dt + \frac{H_n}{2r}\left[dz - \begin{pmatrix} a_1 \\ b_1 \end{pmatrix}dt\right] \tag{5.184}$$

where

$$H_n = \begin{bmatrix} \frac{b_{n-1} - b_{n+1}}{2} - a_1 a_n & \frac{a_{n-1} + a_{n+1}}{2} - a_n b_1 \\ \frac{a_{n+1} - a_{n-1}}{2} - a_1 b_n & \frac{b_{n-1} + b_{n+1}}{2} - b_1 b_n \end{bmatrix} \tag{5.185}$$

For the approximate density, we have

$$\begin{pmatrix} \hat{a}_n \\ \hat{b}_n \end{pmatrix} = g_n(\alpha)\begin{pmatrix} \sin n\beta \\ \cos n\beta \end{pmatrix} \tag{5.186}$$

where we introduce the notation

$$g_n(\alpha) = \frac{I_n(\alpha)}{I_0(\alpha)} \tag{5.187}$$

We want to obtain the Ito derivatives of the approximate-density Fourier coefficients ($\hat{a}_n$ and $\hat{b}_n$) as functions of the statistics x and y (in equation 5.180). From the no-process-noise analysis of section 3.3.2, where a_n and b_n were the same functions of x and y as $\hat{a}_n$ and $\hat{b}_n$ here, we may infer that

$$d\begin{pmatrix}\hat{a}_n\\ \hat{b}_n\end{pmatrix} = \begin{bmatrix}\partial\hat{a}_n/\partial x & \partial\hat{a}_n/\partial y\\ \partial\hat{b}_n/\partial x & \partial\hat{b}_n/\partial y\end{bmatrix}\begin{pmatrix}dx\\ dy\end{pmatrix} + (\text{ITO})\ dt \tag{5.188}$$

where

$$\begin{bmatrix}\partial\ \hat{a}_n/\partial x & \partial\ \hat{a}_n/\partial y\\ \partial\ \hat{b}_n/\partial x & \partial\ \hat{b}_n/\partial y\end{bmatrix} = \hat{H}_n \tag{5.189}$$

and

$$(\text{ITO}) = -\frac{1}{2r}\hat{H}_n\begin{pmatrix}\hat{a}_1\\ \hat{b}_1\end{pmatrix} \tag{5.190}$$

where $\hat{H}_n$ is the matrix H_n evaluated at the approximate-density coefficients.

Then for our approximate density,

$$d\begin{pmatrix}\hat{a}_n\\ \hat{b}_n\end{pmatrix} = -f(\alpha)\hat{H}_n\begin{pmatrix}x\\ y\end{pmatrix}dt + \frac{1}{2r}\hat{H}_n\left[dz-\begin{pmatrix}\hat{a}_1\\ \hat{b}_1\end{pmatrix}dt\right] \tag{5.191}$$

We also note that, after much simplification,

$$\hat{H}_n\begin{pmatrix} x \\ y \end{pmatrix} = \left[\frac{g_{n-1} + g_{n+1}}{2} - g_1 g_n\right] \frac{\alpha}{g_n} \begin{pmatrix} \hat{a}_n \\ \hat{b}_n \end{pmatrix} \tag{5.192}$$

Thus, the differential equations for the Fourier coefficients of the approximate density become

$$d\begin{pmatrix} \hat{a}_n \\ \hat{b}_n \end{pmatrix} = -\frac{\alpha}{g_n} f(\alpha) \left(\frac{g_{n-1} + g_{n+1}}{2} - g_1 g_n\right)\begin{pmatrix} \hat{a}_n \\ \hat{b}_n \end{pmatrix} dt + \frac{1}{2r} \hat{H}_n [dz - \begin{pmatrix} \hat{a}_1 \\ \hat{b}_1 \end{pmatrix} dt] \tag{5.193}$$

and we see that this would equal equation 5.184 if

$$\frac{\alpha}{g_n(\alpha)} f(\alpha) \left(\frac{g_{n-1}(\alpha) + g_{n+1}(\alpha)}{2} - g_1(\alpha) g_n(\alpha)\right) = n^2 \frac{q}{2} \tag{5.194}$$

This equation will not hold (for all n) for any $f(\alpha)$, but by choosing

$$f(\alpha) = \frac{q}{2} \frac{g_1(\alpha)}{\alpha} \frac{1}{\left(\frac{1+g_2(\alpha)}{2} - g_1(\alpha) g_n(\alpha)\right)} \tag{5.195}$$

we may obtain

$$d\begin{pmatrix} \hat{a}_1 \\ \hat{b}_1 \end{pmatrix} = -\frac{q}{2} \begin{pmatrix} \hat{a}_1 \\ \hat{b}_1 \end{pmatrix} dt + \frac{1}{2r} \hat{H}_1 [dz - \begin{pmatrix} \hat{a}_1 \\ \hat{b}_1 \end{pmatrix} dt] \tag{5.196}$$

as desired. Unfortunately, this also implies that

$$d\begin{pmatrix} \hat{a}_n \\ \hat{b}_n \end{pmatrix} = -\frac{q}{2} \frac{g_1}{g_n} \frac{\left(\frac{g_{n-1} + g_{n+1}}{2} - g_1 g_n\right)}{\left(\frac{1 + g_2}{2} - g_1^2\right)} \begin{pmatrix} \hat{a}_n \\ \hat{b}_n \end{pmatrix} dt + \frac{1}{2r} \hat{H}_n [dz - \begin{pmatrix} \hat{a}_1 \\ \hat{b}_1 \end{pmatrix} dt] \tag{5.197}$$

Thus, the first Fourier coefficients propagate in the same manner as the optimal pair, with correct damping and coupling to the higher-order terms. Unfortunately, the higher-order terms are improperly damped, so that even the first Fourier coefficients are incorrect. Nonetheless, this density should closely approximate the conditional density, especially as $P_{\theta\ell} \to \infty$ and the density becomes uniform, where the higher-order coefficients become negligible.

The damping function $f(\alpha)$ (equation 5.195), divided by q, is plotted in figures 5.2 and 5.3. We see that, for very small α,

$$f(\alpha) \to q/2$$

while for very large α, $f(\alpha)$ becomes linear in α. This function should be easy to approximate, while the lack of filter states multiplying dz in equation 5.180 means that no Wong-Zakai correction terms are required. Thus, this filter (with nonlinear gain $f(\alpha)$) is only slightly more complex than the static phase designs.

This filter (which we call the "Bessel filter") was simulated on the same noise sequences as the filters listed in Table 5.1, and the results are shown in Table 5.4. We see that the Bessel filter performs very well, with the average errors falling in the small region between the LQF and the FCF.

Considering the approximate density that the Bessel filter defines (see equation 5.183), we also computed (during the simulations) the average value of

$$1 - \frac{I_1(\alpha)}{I_0(\alpha)}$$

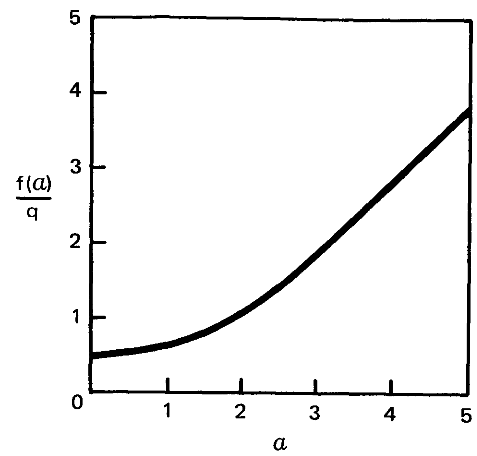


Figure 5.2 Nonlinear Gain for Small α

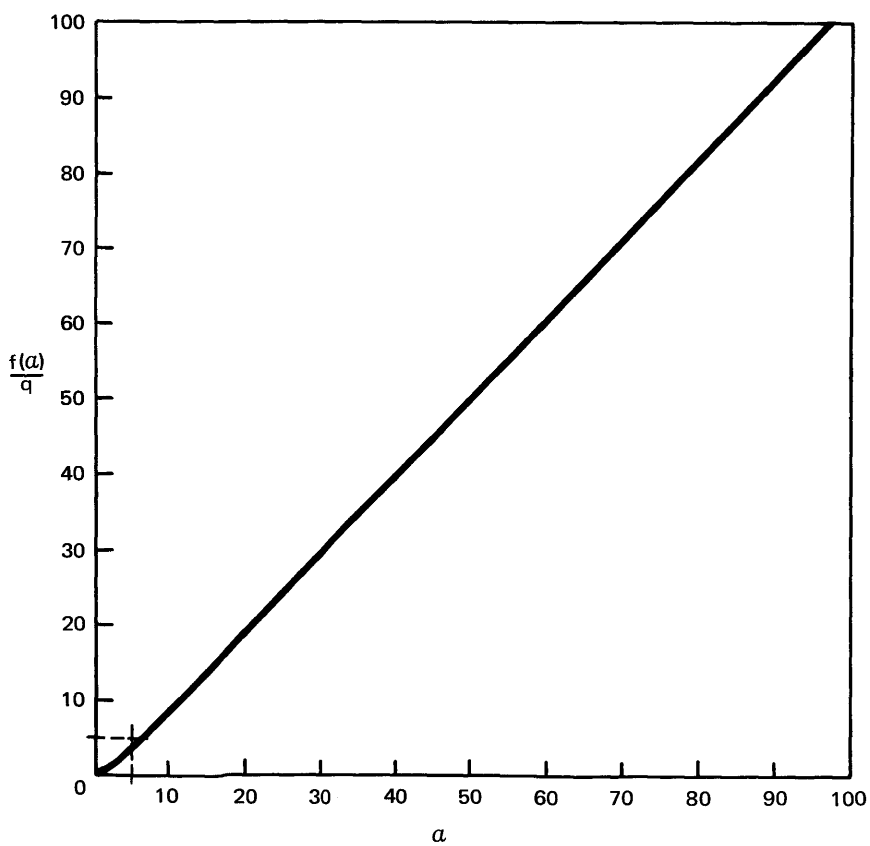


Figure 5.3 Nonlinear Gain for Large α

	$E[\varepsilon^2]$ (rad.^{2})		E[1-cos ε]	
	Actual	% Improvement	Actual	% Improvement
PLL (Simulated)	1.648	(Ref.)	0.567	(Ref.)
APDF	1.498	9.1%	0.525	7.4%
LQF	1.456	11.7%	0.511	9.9%
BESSEL	1.444	12.4%	0.508	10.4%
FCF	1.437	12.8%	0.506	10.8%

Table 5.4 BESSEL FILTER PERFORMANCE

as an indication of how well the filter could predict its "mean cosine error." For comparison, the analagous quantity

$$1 - \sqrt{a_1^2 + b_1^2}$$

was computed for the FCF (see section 5.4.3). These results, along with the actual performance on the same simulations, are shown in Table 5.5. We see that the Bessel filter is accurate in its "self-analysis," being slightly less optimistic than the FCF, and very close to its actual performance.

	E[1 - cos ε]	
	Predicted	Actual
Bessel	0.501	0.508
FCF	0.492	0.506

Table 5.5 Predicted and Actual Performance for the Bessel Filter and FCF

5.9 Conclusions

This chapter has investigated the first-order phase-lock loop problem in detail. We first examined the PLL, the optimal, infinite dimensional filter, and certain practical sub-optimal designs. Then the approximation method of chapter 4 was used to generate several approximate densities for this problem, with encouraging results. The first-cumulant approximation produced an exponential cosine density which seemed to fit the problem well. The higher-order cumulants produced exponential Fourier series, however, which could not be used to generate better phase estimates. The moment equations, on the other hand, produced regular Fourier series which easily provided phase estimates. Unfortunately, the addition of the extra moments resulted in a very slow improvement in filter performance.

Nonetheless, two facts about the first-cumulant density stand out. First, the first-cumulant filter outperforms the classic PLL - the extended Kalman filter for this problem - for high $P_{\theta\ell}$. As a test of our general approximation method, it is considered significant that the first term outperforms the extended Kalman filter. Secondly, a very similar filter to the first-cumulant filter, the linear minimum-variance filter of [15], performs better with only a slighly different gain. This indicates that the first cumulant filter does not make optimal use (i.e., pick the best parameters) of its own filter structure. Indeed, we found it easier to significantly improve performance by modifying the first-cumulant filter (designing the Bessel filter) than by taking more terms in the approximation series.

In the first-order PLL problem we considered, however ($P_{\theta\ell}$=1 rad.^{2}), the first term in the moment and cumulant approximations improved upon PLL performance (error variance) by about 9%, where the possible improvement is only 13.5%. In other nonlinear problems were there is more room for improvement, the higher-order approximations may be more useful. For the PLL problem, we found the approximation method to be more useful in providing <u>clues</u> about practical, sub-optimal filter structures than in completely specifying those structures.

CHAPTER 6

SECOND-ORDER PHASE-LOCK LOOP PROBLEMS

6.1 Introduction

Encouraged by the good performance of the approximate density filters in the first order phase-lock loop problem (Chapter 5), we now design first-cumulant approximate density filters for three phase-estimation problems usually "solved" by second-order phase-lock loops [40]. These problems consider more complicated, but more realistic, phase models than the first-order problem (Brownian motion phase), while retaining the baseband measurements of chapters one and five. The first problem (section 6.2) adds carrier frequency uncertainty to the Brownian motion phase model of the last chapter. The second problem (section 6.3) models the phase as the integral of Brownian motion, and the third problem (section 6.4) is a general FM (frequency modulation) problem often encountered in communications.

We note that for all three of these problems, no exact (nonlinear) analysis for the PLL operation in noise exists to compare with Viterbi's result (section 1.4.3.1) for the first-order PLL. Only in the linear case (high signal-to-noise ratio) have closed-form solutions (Kalman filters) been obtained. For the special case when there is no process noise, however, frequency estimation results have been obtained as solutions to parameter estimation problems [41]. By their nature, our approximate densities are closer to the infinite-dimensional parameter estimation result than to the linearized PLL.

6.2 Brownian Motion Phase with Unknown Carrier Frequency

6.2.1 Problem Statement

In this section we consider the problem of a Brownian motion phase process transmitted at a carrier frequency with a (stationary) Gaussian density. This problem is an extension of the first-order problem we considered in the last chapter to a more realistic filtering situation. In practice, the carrier frequency is not known exactly (or more precisely, the local oscillator - VCO - frequency cannot be matched perfectly to the carrier), and a frequency difference exists between the real carrier and the carrier estimate we heterodyne by to obtain the baseband measurements (section 1.4.1).

If we model the carrier frequency as a Gaussian random variable with mean $\bar{\omega}_c$ and variance σ^2 and then heterodyne by $\bar{\omega}_c$, the difference frequency becomes a zero mean random variable with variance σ^2. We may model this signal with the state equation*

$$d\begin{pmatrix}\theta\\ \omega\end{pmatrix} = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}\theta\\ \omega\end{pmatrix}dt + \begin{pmatrix}1\\ 0\end{pmatrix}du \tag{6.1}$$

The measurements are given by

$$d\begin{pmatrix}z_1\\ z_2\end{pmatrix} = \begin{pmatrix}\sin\theta_t\\ \cos\theta_t\end{pmatrix}dt + \begin{pmatrix}dn_1\\ dn_2\end{pmatrix} \tag{6.2}$$

*Throughout this chapter, "ω" will refer to the frequency state and not a point in the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ as in Chapter 4.

where u, n_1 and n_2 are as in section 5.1.2, and ω is a Gaussian difference frequency with density

$$p(\omega) = N(0, \sigma^2) \tag{6.3}$$

We remark that the associated linear system (with linear phase measurement) is observable but not controllable. The uncontrollability is because of the lack of an input to the frequency derivative. This means that the input noise does not get to the frequency (ω), and perfect frequency estimation is possible with an infinite observation interval.

As discussed in section 5.5.1, the important term in the Radon-Nikodym derivative for this baseband measurement problem is

$$\zeta'_t = \frac{1}{2r} \int_0^t \left(\sin \theta_s \, dz_{1_s} + \cos \theta_s \, dz_{2_s}\right) \tag{6.4}$$

The actual conditional density may then be written as

$$p(\theta, \omega, t | z_0^t) = \frac{(E_0^{\mathcal{B}} e^{\zeta'_t}) \, e^{-\omega^2/2\sigma^2}}{W^t} \tag{6.5}$$

where

$$\mathcal{B} = \mathcal{A}_{zt} \vee \mathcal{A}(\theta_t, \omega) \tag{6.6}$$

and

$$W_t = \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} E_0^{\mathcal{B}} e^{\zeta'_t} e^{-\omega^2/2\sigma^2} \, d\omega d\theta \tag{6.7}$$

We note that it is possible to modify a first-order PLL to improve its acquisition performance, as was discussed in section 2.3.3, by adding an integrator and forming a second-order loop. This method works very poorly in high-noise situations, however, and we do not pursue it here. Instead we will discuss a "bank" of PLL's at different frequencies in section 6.2.5.1.

6.2.2 First-Cumulant Filter

From the system model (equation 6.1) we may write

$$\theta_t = \theta_s + \omega(t-s) + (u_t - u_s) \tag{6.8}$$

$$= \theta_0 + \omega(t) + u_t \tag{6.9}$$

where u_t and ω are independent Gaussian random variables. Thus, the density for θ_s conditioned on θ_t and ω (for $s \leq t$) is Gaussian with

$$p(\theta_s | \theta_t, \omega) = N(\theta_t - \omega(t-s), q(t-s)) \tag{6.10}$$

The expectations needed for the first cumulant approximation (section 4.2.2) become (from the folded-normal identities of equation 5.47)

$$E\left[\begin{pmatrix}\sin \theta_s \\ \cos \theta_s\end{pmatrix} \middle| \theta_t, \omega\right] = e^{-\frac{q}{2}(t-s)} \begin{bmatrix}\sin (\theta_t - \omega(t-s)) \\ \cos (\theta_t - \omega(t-s))\end{bmatrix} \tag{6.11}$$

Now using the approximation

$$E_0^{\mathcal{B}} e^{\zeta_t'} \simeq e^{E_0^{\mathcal{B}} \zeta_t'} \tag{6.12}$$

we form the approximate density

$$P_1(\theta,\ \omega,\ t|z_0^t) = \frac{e^{x(\omega)\ \sin\ \theta_t + y(\omega)\ \cos\ \theta_t}}{W_1}\ e^{-\omega^2/2\sigma^2} \tag{6.13}$$

where

$$W_1 = \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} e^{x(\omega)\ \sin\ \theta + y(\omega)\ \cos\ \theta}\ e^{-\omega^2/2\sigma^2}\ d\omega d\theta \tag{6.14}$$

$$x(\omega) = \frac{1}{2r} \int_0^t e^{-\frac{q}{2}(t-s)}\ [\cos\ \omega(t-s)dz_{1_s} + \sin\ \omega(t-s)dz_{2_s} \tag{6.15}$$

$$y(\omega) = \frac{1}{2r} \int_0^t e^{-\frac{q}{2}(t-s)}\ [\cos\ \omega(t-s)dz_{2_s} - \sin\ \omega(t-s)dz_{1_s}] \tag{6.16}$$

where we have used equation 6.11 and standard trigonometric identities to obtain $x(\omega)$ and $y(\omega)$ from $E_0^{\mathcal{B}}\ \zeta_t'$. We may also write equations 6.15 and 6.16 as

$$d\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -q/2 & \omega \\ -\omega & -q/2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} dt + \frac{1}{2r} \begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \tag{6.17}$$

for any chosen (<u>fixed</u>) ω. We comment that these equations will generate the <u>exact</u> conditional density (at any chosen ω) when q=0.

Care must be taken in interpreting this density as a "function" of ω. The value of the density will change as ω varies, and this is the type of functional relationship guaranteed by the conditional expectation in equations 6.5 and 6.12. Unfortunately, however, the density requires an infinite number of statistics to specify its value for all ω. That is, for any finite number (n) of frequency values

$\omega_i (i=1,2, \ldots, n)$, the value of

$$p_1(\theta, \omega, t|z_0^t)\Big|_{\omega=\omega_i}$$

may be obtained by keeping the 2n filter states $x(\omega_i)$ and $y(\omega_i)$, where each x and y are evaluated for all time with fixed $\omega=\omega_i$. In order to obtain a continuous function of ω, therefore, we need an infinite number of evaluation points ω_i.

For the case when q=0, equation 6.13 specifies the exact value of the conditional density at the chosen ω. This result is usually used in "m-hypothesis-test" problems, where ω is assumed, a priori, to be one of m values ω_i, and a filter state pair $x(\omega_i)$ and $y(\omega_i)$ is constructed at each possible ω_i. By its nature, the density (equation 6.13) is much more difficult to examine when ω is a continuous random variable.

6.2.3 Implementation of the First-Cumulant Filter

We now discuss various implementations of the approximate density filter for any value of q (i.e., not necessarily q=0). The filter states (equation 6.17) may be obtained in three simple ways for any chosen ω_i. The most straightforward implementation is suggested by the realization that equation 6.17 is a static phase filter at a frequency ω_i instead of at "baseband" (see chapter 5). Thus we consider implementing the regular static phase filter

$$d\begin{pmatrix} x \\ y \end{pmatrix} = -\frac{q}{2}\begin{pmatrix} x \\ y \end{pmatrix} dt + \frac{1}{2r}\, dz^i \qquad (6.18)$$

where dz^i results from heterodyning the carrier-frequency measurement dz' (see equation 1.16) by $\bar{\omega}_c + \omega_i$ instead of $\bar{\omega}_c$ alone. This results in a bank of n static phase filters (one at each ω_i) which requires n heterodyning operations. Alternatively, equation 6.17 can be implemented directly at each ω_i, which requires only one heterodyning operation.

The third implementation is suggested by the impulse response of the filter states in equations 6.15 and 6.16. This method also has the advantage of requiring only one heterodyning operation. We define

$$\begin{bmatrix} x_s(\omega_i) \\ x_c(\omega_i \end{bmatrix} = \frac{1}{2r} \int_0^t e^{-\frac{q}{2}(t-s)} \begin{pmatrix} \sin \omega(t-s) \\ \cos \omega(t-s) \end{pmatrix} dz_{1_s} \tag{6.19a}$$

$$\begin{bmatrix} y_s(\omega_i) \\ y_c(\omega_i) \end{bmatrix} = \frac{1}{2r} \int_0^t e^{-\frac{q}{2}(t-s)} \begin{pmatrix} \sin \omega(t-s) \\ \cos \omega(t-s) \end{pmatrix} dz_{2_s} \tag{6.19b}$$

so that

$$x(\omega_i) = x_c(\omega_i) + y_s(\omega_i) \tag{6.20a}$$

$$y(\omega_i) = y_c(\omega_i) - x_s(\omega_i) \tag{6.20b}$$

Then we note that these states may be obtained from the Laplace transforms (see, e.g., Hildebrand [17])

$$L\{e^{-\frac{q}{2}t} \sin \omega_i t\} = \frac{\omega_i}{(s + \frac{q}{2})^2 + \omega_i^2} \tag{6.21}$$

$$L\{e^{-\frac{q}{2}t} \cos \omega_i t\} = \frac{s + \frac{q}{2}}{(s + \frac{q}{2})^2 + \omega_i^2} \tag{6.22}$$

We show the complete filter forms in figure 6.1.

While this type of filter would be needed at each ω_i for which we wanted to evaluate the approximate density, certain savings in implementation are possible. The filter forms are all identical, only the parameters change, so that the filter computations may be efficiently time-shared in digital implementations. Also, for analog or digital implementation, the values of $x(-\omega_i)$ and $y(-\omega_i)$ are available from different combinations of $x_s(\omega_i)$, $x_c(\omega_i)$, $y_s(\omega_i)$ and $y_c(\omega_i)$. Using equations 6.19 and 6.20 we have

$$x(-\omega_i) = x_c(\omega_i) - y_s(\omega_i) \tag{6.23a}$$

$$y(-\omega_i) = y_c(\omega_i) + x_s(\omega_i) \tag{6.23b}$$

This means that, for a symmetric bank of filters about $\omega_i = 0$, n four-state (equations 6.19 and 6.20) filters will provide (2n-1) density evaluation points (assuming one filter is set at $\omega_i=0$).

6.2.4 Frequency Estimation

Next we examine some techniques for obtaining phase and frequency estimates from the approximate density filters. The marginal (approximate) density for the frequency is obtained from

$$p_1(\omega, t|z_0^t) = \int_{-\pi}^{\pi} p_1(\theta,\omega,t|z_0^t)\,d\theta \tag{6.24}$$

Using equation 6.13 and the Bessel function appendix (A) we find that

$$p_1(\omega,t|z_0^t) = \frac{I_0\left(\sqrt{x^2(\omega) + y^2(\omega)}\right)e^{-\omega^2/2\sigma^2}}{W_1'} \tag{6.25}$$

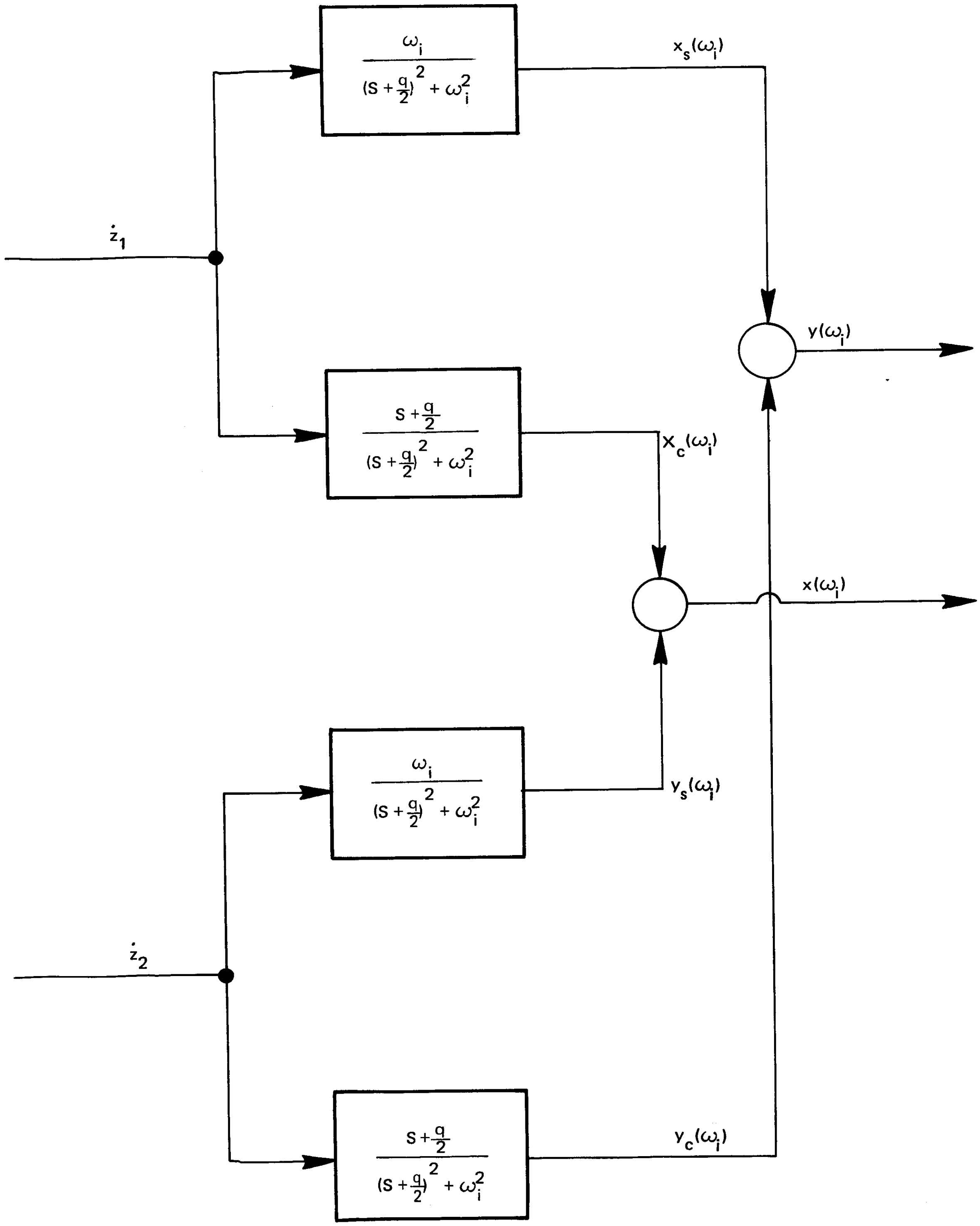


Figure 6.1 Laplace Form of Filter

where

$$W_1' = \int_{-\infty}^{\infty} I_0\left(\sqrt{x^2(\omega) + y^2(\omega)}\right) e^{-\omega^2/2\sigma^2} d\omega \tag{6.26}$$

This density is, of course, awkward to analyze. For the case when $\sigma^2 \to \infty$, however, the density reduces to

$$p_1(\omega,t|z_0^t) = \frac{I_0\left(\sqrt{x^2(\omega) + y^2(\omega)}\right)}{\int_{-\infty}^{\infty} I_0\left(\sqrt{x^2(\omega) + y^2(\omega)}\right) d\omega} \tag{6.27}$$

which is a monotonic function of the magnitude

$$F(\omega) = x^2(\omega) + y^2(\omega) \tag{6.28}$$

Thus, the maximum a posteriori (MAP) estimate of the frequency may be found by maximizing $F(\omega)$. For the case with $q=0$, this estimator (the optimum) matches that from the parameter estimation analysis of Viterbi [41 p. 272]. Regrettably, this is still an infinite-dimensional filter for continuous ω.

6.2.5 Phase Estimation

6.2.5.1 Phase Estimation from Joint Filter

We now examine phase estimation for the approximate density filter of equation 6.13 with $q\neq 0$ and $\sigma^2\neq\infty$. We approximate the smooth a priori density for ω with n point masses, each at a different frequency, and each weighted (relatively) by $e^{-\omega_i^2/2\sigma^2}$. The phase density then becomes

$$p(\theta,t|z_0^t) = \int_{-\infty}^{\infty} p(\theta,\omega,t|z_0^t)\,d\omega \tag{6.29}$$

$$= \frac{1}{n}\sum_{i=1}^{n} p(\theta,\omega_i,t|z_0^t) \tag{6.30}$$

where

$$p(\theta,\omega_i,t|z_0^t) = \frac{e^{x(\omega_i)\ \sin\ \theta + y(\omega_i)\ \cos\ \theta}\ e^{-\omega_i^2/2\sigma^2}}{K\ 2\pi\ I_0(\sqrt{x^2(\omega_i) + y^2(\omega_i)})} \qquad (6.31)$$

$$K = \sum_{i=1}^{n} e^{-\omega_i^2/2\sigma^2} \qquad (6.32)$$

Each $x(\omega_i)$ and $y(\omega_i)$ is obtained by a filter of the form of equation 6.20.

The phase estimate for this density is obtained from

$$\hat{\theta} = \tan^{-1}\ a_1/b_1 \qquad (6.33)$$

where

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \frac{1}{n} \sum_{i=1}^{n} \int_{-\pi}^{\pi} \begin{pmatrix} \sin\ \theta \\ \cos\ \theta \end{pmatrix} p(\theta,\ \omega_i,\ t|z_0^t)\, d\theta \qquad (6.34)$$

which becomes, using equation 6.31

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \frac{1}{2\pi Kn} \sum_{i=1}^{n} \frac{I_1\left(\sqrt{x^2(\omega_i) + y^2(\omega_i)}\right)}{I_0\left(\sqrt{x^2(\omega_i) + y^2(\omega_i)}\right)}\ e^{-\omega_i^2/2\sigma^2} \begin{bmatrix} x(\omega_i) \\ y(\omega_i) \end{bmatrix} \qquad (6.35)$$

This may be simplified slightly by ignoring the constant $2\pi nK$ term which is cancelled in the ratio a_1/b_1 in equation 6.33.

The sin θ and cos θ estimates (a_1 and b_1) are composed of weighted sums (equation 6.35) of the sin θ and cos θ estimates of the individual filters (at each ω_i). The weightings contain both a priori ($\exp -\omega_i^2/2\sigma^2$) and a posteriori ($I_1/I_0$) information, but it is the a posteriori "adaptive" weighting that makes the filter interesting. The weighting term I_1/I_0 is a function of the same statistic ($F(\omega)$ in equation 6.28) which is used to compute the most probable frequency. Thus, this filter uses the estimate of

$$\sqrt{x^2(\omega_i) + y^2(\omega_i)}$$

from the quadrature channel of each static phase filter (see section 5.4.1, where this statistic was called "α") to judge how "reasonable" the phase estimate from that filter (at ω_i) is. This filter bank should outperform a classic PLL, which has no way of sensing a frequency offset in high noise.

One may wonder why we can't compare this bank of filters to an analogous bank of PLL's, also spread over the range of possible frequencies. That is, why not use (2n-1) PLL's, arranged at $\pm\,\omega_i$ (and $\omega_1 = 0$) for $i = 2, \ldots, n$, and weight each phase estimate by

$$\frac{e^{-\omega_i^2/\sigma^2}}{1 + 2\sum_{i=2}^{n} e^{-\omega_i^2/\sigma^2}}$$

to obtain an overage phase estimate? (The filter bank would be symmetric about ω=0 because the difference frequency had a zero-mean Gaussian a priori density.) The reason for not using such a scheme is that a classic PLL at ω_i, if closer to the real frequency, would have a "balancing loop" at $-\omega_i$ which was further from the real frequency. Since the classic loop has no way of updating its error covariance prediction, both loops (at $+\omega_i$ and $-\omega_i$) would have the same weighting (above), and the net performance would not change noticeably. The bank of static phase filters can update the relative weightings; a bank of PLL's cannot.

6.2.5.2 Phase Estimation Only

There is an alternate method for obtaining a phase estimate in this problem when we are not interested in simultaneously estimating frequency. Using the conditional density formula (eq. 6.5) and integrating over the frequency, we obtain the marginal phase density

$$p(\theta,\ t|z_0^t) = \frac{E_0^C e^{\zeta_t'}}{W_t'} \tag{6.36}$$

where

$$C = A_{2t} \vee A(\theta_t) \tag{6.37}$$

$$W_t' = \int_{-\pi}^{\pi} E_0^C\, e^{\zeta_t'}\, d\theta \tag{6.38}$$

Our goal is to create a first-cumulant approximation for the σ-algebra C

$$E_0^C\, e^{\zeta_t'} \simeq e^{E_0^C \zeta_t'} \tag{6.39}$$

We recall from equation 6.8 that

$$\theta_s = \theta_t - \omega(t-s) - (u_t-u_s) \tag{6.40}$$

where ω is a Gaussian random variable, independent of u_t. This means that θ_s given θ_t is a Gaussian random variable with mean θ_t and variance $\sigma^2(t-s)^2 + q(t-s)$. We therefore have that

$$E\left[\begin{pmatrix}\sin\theta_s\\ \cos\theta_s\end{pmatrix}\Big|\theta_t\right] = e^{-\frac{\sigma^2}{2}(t-s)^2 - \frac{q}{2}(t-s)} \begin{pmatrix}\sin\theta_t\\ \cos\theta_t\end{pmatrix} \tag{6.41}$$

so that our approximate density becomes

$$p_1(\theta,t|z_0^t) = \frac{1}{W_1'} e^{\bar{x}\sin\theta + \bar{y}\cos\theta} \tag{6.42}$$

where

$$\begin{pmatrix}\bar{x}\\ \bar{y}\end{pmatrix} = \frac{1}{2r}\int_0^t e^{-\frac{\sigma^2}{2}(t-s)^2 - \frac{q}{2}(t-s)} \begin{pmatrix}dz_{1_s}\\ dz_{2_s}\end{pmatrix} \tag{6.43}$$

$$W_1' = 2\pi I_0(\sqrt{\bar{x}^2 + \bar{y}^2}) \tag{6.44}$$

$$\hat{\theta} = \tan^{-1} \bar{x}/\bar{y} \tag{6.45}$$

This filter is (of course) independent of ω and the infinite-dimensional problems of the filter of equation 6.13. Unfortunately, the above filter is still infinite-dimensional, since $\bar{x}$ and $\bar{y}$ cannot be computed as the outputs of a finite-dimensional linear system (for all t). For any fixed t, we can compute $\bar{x}(t)$ and $\bar{y}(t)$, but we cannot compute these states as functions of time from any finite set of derivatives and the measurements. The reason for this is that $e^{-\Delta^2}$ for $\Delta = t-s$ is not separable. That is, we cannot write

$$\begin{pmatrix} \overline{x} \\ \overline{y} \end{pmatrix} = f(t) \int_0^t g(\tau)\, dz_\tau \tag{6.46}$$

for any f and g. (We could do this for the first-order PLL problem, where the static phase filter was the same as this filter with $\sigma^2 = 0$.) We can approximate $\overline{x}$ and $\overline{y}$ by a sum of separable functions by finding f_j and g_j such that

$$e^{-\frac{\sigma^2}{2}(t-s)^2} \simeq \sum_{j=1}^{n} f_j(t) g_j(s) \tag{6.47}$$

While we do not pursue this course here, it remains an interesting avenue of investigation.

We note that, for q=0 (where this is the optimal phase filter), the impulse response of the above filter (equation 6.43) resembles that of a dispersive delay line. Such a device has been used, in a much different way, to estimate the frequency in this problem (see Viterbi [41, p.283]). The operation of delay lines is usually limited to very high frequencies with short delay times, however, and such devices may not be useful for all values of σ^2.

Finally, we want to compare this phase estimator (equation 6.43 with $\sigma^2 \neq 0$, $q \neq 0$) to the static phase filter of chapter 5 to examine how the frequency uncertainty has affected the approximate density filter. We first note that the frequency uncertainty filter (equation 6.43) converges to the static phase filter as $\sigma^2 \to 0$, a reasonable property. We also see that the only difference between the filters (when $\sigma^2 \neq 0$) is the faster decay with time in the impulse response of the frequency uncertainty filter. While the static phase filter discounts old data by a factor of $\exp(-q\Delta/2)$ where $\Delta = t-s$, the frequency-uncertainty filter uses the function $\exp(-q\Delta/2 - \sigma^2\Delta^2/2)$. The filter of equation 6.43 realizes that the frequency uncertainty makes old data less reliable than the Brownian motion alone does, and the filter compensates by throwing away the old information faster than before. The reasonable nature of this result, which comes from the approximation method itself rather than as a consequence of a design assumption, is further evidence of the value of the approximation technique.

6.3 Brownian Motion Frequency

6.3.1 Problem Statement

We now examine a simple frequency modulation problem considered by Bucy, Hecht and Senne [8]. While the problem itself is not very realistic, it is interesting because it generates a very familiar second-order PLL as a solution (in the low-noise region). The phase is modelled as the integral of a Brownian motion frequency, that is

$$d\begin{pmatrix}\theta\\ \omega\end{pmatrix} = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}\theta\\ \omega\end{pmatrix}dt + \begin{pmatrix}0\\ 1\end{pmatrix}du \tag{6.48}$$

where

$$E[du^2] = q\,dt \tag{6.49}$$

and we assume that we have the usual baseband measurement (dz) of equation 6.2.

One reason for considering this phase model is that it represents a "slower", more realistic phase than the Brownian motion process considered in chapter 5 and the last section (6.2). This more realistic phase is obtained, however, by modelling the <u>frequency</u> as a Brownian motion, with an ever increasing variance, which is a poor assumption in general. (Brownian motion phase is not as disturbing since the modulo-2π nature of our phase message translates a Gaussian phase with infinite variance into a uniformly distributed random variable on the circle.) We note that a first-order Markov model for the frequency, which we consider in section 6.5, is better suited to most problems.

6.3.2 Classic PLL

Our main reason for considering the model of equation (6.48), however, is that it generates a second-order phase-lock loop with a 0.707 damping ratio (a value often used). To see this, we consider a Kalman filter designed for the dynamics of equation (6.48) with a linear pseudo-measurement (see section 1.4.1)

$$dz_p = \theta dt + dn_I$$

where dn_I is the in-phase noise of equation 5.5. The linear-predicted PLL error variance then propagates as

$$\dot{P}_\ell = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} P_\ell + P_\ell \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix} - P_\ell \begin{bmatrix} 1/2r & 0 \\ 0 & 0 \end{bmatrix} P_\ell \tag{6.50}$$

which is solved in steady-state by

$$P_\ell = \begin{bmatrix} 2[2r^3q]^{1/4} & (2rq)^{1/2} \\ (2rq)^{1/2} & [2^3rq^3]^{1/4} \end{bmatrix} \tag{6.51}$$

The extended Kalman filter (PLL) may then written as

$$d\begin{pmatrix} \hat{\theta} \\ \hat{\omega} \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{pmatrix} \hat{\theta} \\ \hat{\omega} \end{pmatrix} dt + \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} dz_I \tag{6.52}$$

where

$$\begin{bmatrix} K_1 \\ K_2 \end{bmatrix} = \begin{bmatrix} (2q/r)^{1/4} \\ (q/2r)^{1/2} \end{bmatrix} \tag{6.53}$$

and

$$dz_I = \sin\,(\theta\ \hat{\theta}) + dn_I \tag{6.54}$$

as in equations 1.21 and 5.5

For the very-high signal-to-noise ratio case, we may diagram the linear system (with the "small sine" approximation as shown in figure 6.2.

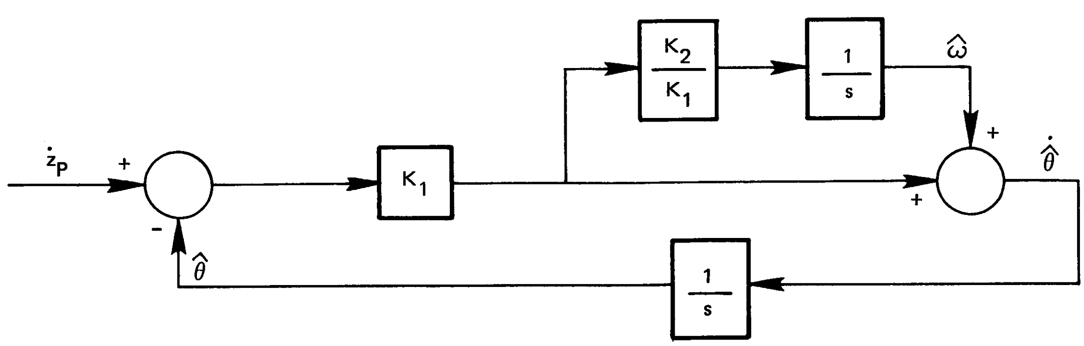


Figure 6.2 Linear 2nd-Order PLL

This filter has a transfer function of

$$\frac{\hat{\theta}(s)}{\dot{z}_p(s)} = \frac{K_1(s + K_2/K_1)}{s^2 + K_1 s + K_2} \tag{6.55}$$

with a damping ratio* (ζ) of

$$\zeta = \frac{1}{2}\left[\frac{K_1^2}{K_2}\right]^{1/2} = \frac{\sqrt{2}}{2} \tag{6.56}$$

As discussed in Van Trees [40, p. 48 and 67], this is a very common design for PLL's.

6.3.3 Approximate-Density Filter

We now derive a first-cumulant approximate-density filter for this problem. We will use the approximation

$$p_1(\theta,\omega,t|z_0^t) = \frac{e^{E_0^{\mathcal{B}} \zeta_t'}}{W_1} p(\theta,\omega,t) \tag{6.57}$$

where

$$\zeta_t' = \frac{1}{2r} \int_0^t \begin{pmatrix} \sin \theta_s \\ \cos \theta_s \end{pmatrix}^T dz_s \tag{6.58}$$

$$\mathcal{B} = \mathcal{A}_{zt} \vee \mathcal{A}(\theta_t, \omega_t) \tag{6.59}$$

We note that ω_t is not constant for all time in this model as it was in section 6.2 .

---

*This ζ is, of course, unrelated to any of our Radon-Nikodym derivative terms ζ_t and ζ_t'.

We first examine the a priori density for θ and ω. This density is Gaussian with zero mean and covariance P, where

$$\dot{P} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} P + P \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix} \tag{6.60}$$

The solution to this equation is

$$P_{11}(t) = \frac{qt^3}{3} + t^2 P_{22}(0) + 2t P_{12}(0) + P_{11}(0) \tag{6.61a}$$

$$P_{12}(t) = \frac{qt^2}{2} + P_{22}(0)t + P_{12}(0) \tag{6.61b}$$

$$P_{22}(t) = qt + P_{22}(0) \tag{6.61c}$$

We also note that the transition matrix for the dynamics (equation 6.48) is

$$\phi(t,s) = \begin{bmatrix} 1 & (t-s) \\ 0 & 1 \end{bmatrix} \tag{6.62}$$

The first step in obtaining our filter is to find the backward transition density

$$p(\theta_s, \omega_s | \theta_t, \omega_t) \text{ for } s \leq t$$

needed in the expectation of ζ_t'. This density is Gaussian with mean and covariance (see section 4.2.3 and equations 4.78 and 4.79)

$$m_{s/t} = P_{s/t} \, \phi^T(t,s) \, P_{t/s}^{-1} \begin{pmatrix} \theta_t \\ \omega_t \end{pmatrix} \tag{6.63}$$

$$P_{s/t}^{-1} = \phi^T(t,s)\ P_{t/s}^{-1}\ \phi(t,s) + P_s^{-1} \tag{6.64}$$

where the forward transition covariance $P_{t/s}$ may be found from equations 6.61 as

$$P_{t/s} = q(t-s) \begin{bmatrix} (t-s)^2/3 & (t-s)/2 \\ (t-s)/2 & 1 \end{bmatrix} \tag{6.65}$$

Since we are interested in a filter for steady-state phase and frequency processes, and for ease of computation, we let the initial covariance P(0) become infinite (its "steady-state" value), which makes

$$P_s^{-1} = 0 \tag{6.66}$$

in equation 6.64 without otherwise affecting the transition densities. (However, this will affect the a priori density term $p(\theta,\omega,t)$ in our approximate density, which we will consider shortly.) Letting the covariance become infinite is convenient, but not necessary in our method. It is equivalent to assuming that we have no a priori state information, and therefore our filter will be an approximate <u>maximum-likelihood</u> estimator.

We may now solve equation 6.64 for $P_{s/t}$, finding (after much simplification)

$$P_{s/t} = q(t-s) \begin{bmatrix} (t-s)^2/3 & -(t-s)/2 \\ -(t-s)/2 & 1 \end{bmatrix} \tag{6.67}$$

Then equation 6.63 may be solved for the means

$$\begin{pmatrix} \bar{\theta}_{s/t} \\ \bar{\omega}_{s/t} \end{pmatrix} = \begin{bmatrix} 1 & -(t-s) \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \theta_t \\ \omega_t \end{pmatrix} \tag{6.68}$$

We now use these results in a folded-normal density (equation 5.47) to obtain

$$E\left[\begin{pmatrix}\sin\theta_s\\ \cos\theta_s\end{pmatrix}\middle|(\theta_t,\ \omega_t)\right] = e^{-\frac{q}{6}(t-s)^3}\begin{bmatrix}\sin\ (\theta_t-\omega(t-s))\\ \cos\ (\theta_t-\omega(t-s))\end{bmatrix} \tag{6.69}$$

so that we may write

$$E_0^{\mathcal{B}}\ \zeta_t' = x(\omega)\ \sin\theta_t + y(\omega)\ \cos\theta_t \tag{6.70}$$

where

$$x(\omega) = \frac{1}{2r}\int_0^t e^{-\frac{q}{6}(t-s)^3}(\cos\omega\ (t-s)dz_{1_s} + \sin\omega(t-s)\ dz_{2_s}) \tag{6.71a}$$

$$y(\omega) = \frac{1}{2r}\int_0^t e^{-\frac{q}{6}(t-s)^3}(\cos\omega(t-s)dz_{2_s} - \sin\omega(t-s)\ dz_{1_s}) \tag{6.71b}$$

Now since we assume $P_t^{-1} = 0$ for all t, we have that our approximate conditional density is

$$p_1(\theta,\omega,t|z_0^t) = \frac{e^{x(\omega)\ \sin\theta + y(\omega)\ \cos\theta}}{W_1} \tag{6.72}$$

where

$$W_1 = \int_{-\pi}^{\pi}\int_{-\infty}^{\infty} e^{x(\omega)\ \sin\theta + y(\omega)\ \cos\theta}\ d\omega\ d\theta \tag{6.73}$$

Phase and frequency estimates (approximate maximum-likelihood estimates) may be obtained from this filter in the same way as in sections 6.2.4 and

6.2.5. This filter is doubly complex, however, since it is simultaneously infinite-dimensional in ω and not separable in t and s. We note that no phase-only estimate is possible (in the manner of section 6.2.5.2), because of the infinite variance of the frequency which we assumed ($P_{22}(0) = \infty$) in order to obtain $P_{s/t}$ (in equation 6.67).

Finally, we note the overall similarity of this filter to that of the frequency-uncertainty problem of section 6.2. In this filter (equation 6.71), the log of the weighting function

$$e^{-q\Delta^3/6} \quad \text{for } \Delta = t-s$$

decreases as the cube of the time difference. We recall that the frequency-uncertainty filter (with an unknown but constant frequency) had a Δ^2 decay, while the original static phase filter of chapter 5 decreased with Δ only. Here, the approximation method recognizes that a changing frequency makes old data even less reliable than the uncertain frequency of section 6.2.

6.4 FM Problem

6.4.1 Problem Statement

We complete this chapter by analyzing an approximate density filter for a standard frequency-modulation (FM) problem (see, e.g., Van Trees [40 p. 94]). This problem has recently been examined by Tam and Moore [39], who obtained Gaussian-sum approximate filters for this system as well as for the first-order PLL problem as discussed in section 5.3.3. The phase is modelled as the integral of a first-order Markov frequency process. That is

$$d\begin{pmatrix}\theta\\ \omega\end{pmatrix}=\begin{pmatrix}0 & 1\\ 0 & -f\end{pmatrix}\begin{pmatrix}\theta\\ \omega\end{pmatrix}dt + \begin{pmatrix}0\\ 1\end{pmatrix}du \tag{6.74}$$

where

$$E[du^2] = qdt \tag{6.75}$$

and we use the baseband measurements as before (equation 6.2). This problem is reducible to that of the last section (with f=0) but yields much different results (for $f\neq 0$). The main advantage of this model is that the frequency variance has a finite steady-state value, unlike that of section 6.3.

6.4.2 Classic PLL Design

We first note that a classic phase-lock loop can be designed for this problem as in section 6.3.2. The results are summarized below for completeness (see, e.g., Van Trees [40, p.94] for details). The filter equations are

$$d\begin{pmatrix}\hat{\theta}\\ \hat{\omega}\end{pmatrix}=\begin{pmatrix}0 & 1\\ 0 & -f\end{pmatrix}\begin{pmatrix}\hat{\theta}\\ \hat{\omega}\end{pmatrix}dt + \begin{pmatrix}K_1\\ K_2\end{pmatrix}dz_I \tag{6.76}$$

where

$$\begin{bmatrix}K_1\\ K_2\end{bmatrix}=\frac{1}{2r}\begin{bmatrix}P_{\ell_{11}}\\ P_{\ell_{12}}\end{bmatrix} \tag{6.77}$$

and the linear-predicted error variance becomes (in steady-state)

$$P_{\ell_{11}} = [4r^2f^2 + (32\ r^3q)^{1/2}]^{1/2} - 2rf \tag{6.78a}$$

$$P_{\ell_{12}} = P^2_{\ell_{11}}/4r \tag{6.78b}$$

$$P_{\ell_{22}} = \frac{fP^2_{\ell_{11}}}{4r} + \frac{1}{8r^2} P^3_{\ell_{11}} \tag{6.78c}$$

The filter structure of this loop is the same as that of figure 6.2 except that $1/(s+f)$ replaces $(1/s)$ in the $\hat{\omega}$ signal path and, of course, the gains are different.

6.4.3 Approximate-Density Filter

We now design a first-cumulant approximate density filter for this system (equation 6.74). We first consider the covariance equation for the dynamics

$$\dot{P} = \begin{bmatrix} 0 & 1 \\ 0 & -f \end{bmatrix} P + P \begin{bmatrix} 0 & 0 \\ 1 & -f \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix} \tag{6.79}$$

We are interested in two solutions to this equation. The first is the solution for $P_{t/s}$, that is, a solution for P_t given that $P_s = 0$ (for $t \geq s$). This is found to be

$$P_{t/s} = \begin{bmatrix} \frac{q}{f^2}\Delta - \frac{q}{2f^3}[3-4e^{-f\Delta} + e^{-2f\Delta}] & \frac{q}{2f^2}[1 - e^{-f\Delta}]^2 \\ \frac{q}{2f^2}[1 - e^{-f\Delta}]^2 & \frac{q}{2f}[1 - e^{-f\Delta}] \end{bmatrix} \tag{6.80}$$

for

$$\Delta = t-s \geq 0 \tag{6.81}$$

The second covariance we want is that for the steady-state system, that is, P_t for large t. If the initial conditions for P_{22} and P_{12} are equal to their steady-state values ($q/2f$ and $q/2f^2$ respectively), then the variance P_t may be written

$$P_t = \begin{bmatrix} \infty & q/2f^2 \\ q/2f^2 & q/2f \end{bmatrix} \tag{6.82}$$

where we let $P_{11}(0)$ go to ∞ to represent the a priori uniform phase density (module 2π). (Strictly speaking, we will want to use the identity

$$P_s^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & 2f/q \end{bmatrix} \tag{6.83}$$

for time s, which will be true if we compute P_s^{-1} for $P_{11}(0) < \infty$, and then take the limit of P_s^{-1} as $P_{11}(0) \to \infty$.)

We note that the transition matrix for the dynamics of equation 6.74 is

$$\phi(t,s) = \begin{bmatrix} 1 & \frac{1}{f}(1-e^{-f\Delta}) \\ 0 & e^{-f\Delta} \end{bmatrix} \tag{6.84}$$

We now may obtain $P_{s/t}$ (as in section 6.3.2) from equation 4.79

$$P_{s/t}^{-1} = \phi^T(t,s)\; P_{t/s}^{-1}\; \phi(t,s) + P_s^{-1} \tag{6.85}$$

Using equations 6.80, 6.83, and 6.84, and after much simplification, we find that

$$P_{s/t} = \begin{bmatrix} \frac{q}{f^2}\Delta - \frac{q}{2f^3}(3-4e^{-f\Delta} + e^{-2f\Delta}) & -\frac{q}{2f^2}(1-e^{-f\Delta})^2 \\ -\frac{q}{2f^2}(1-e^{-f\Delta})^2 & \frac{q}{2f}(1-e^{-f\Delta}) \end{bmatrix} \tag{6.86}$$

Similarly, we use equation 4.78 for the mean

$$\begin{pmatrix} \overline{\theta}_{s/t} \\ \overline{\omega}_{s/t} \end{pmatrix} = P_{s/t}\; \phi^T(t,s)\; P_{t/s}^{-1} \begin{pmatrix} \theta_t \\ \omega_t \end{pmatrix} \tag{6.87}$$

and equations 6.80, 6.84, and 6.86 to obtain

$$\begin{pmatrix} \overline{\theta}_{s/t} \\ \overline{\omega}_{s/t} \end{pmatrix} = \begin{bmatrix} 1 & -\frac{1}{f}(1-e^{-f\Delta}) \\ 0 & e^{-f\Delta} \end{bmatrix} \begin{pmatrix} \theta_t \\ \omega_t \end{pmatrix} \tag{6.88}$$

Thus, since (θ_s, ω_s) conditioned on (θ_t, ω_t) is a Guassian random variable with mean given by equation 6.88 and variance given by 6.86, we may take the (by now) familiar expectations to obtain

$$E\left[\begin{pmatrix} \sin\theta_s \\ \cos\theta_s \end{pmatrix} \Big| \theta_t, \omega_t\right] = e^{-\frac{(P_{s/t})}{2}} \begin{bmatrix} \sin\left(\theta_t - \frac{1-e^{-f\Delta}}{f}\omega_t\right) \\ \cos\left(\theta_t - \frac{1-e^{-f\Delta}}{f}\omega_t\right) \end{bmatrix} \tag{6.89}$$

This means that the first-cumulant approximate density (using the steady-state a priori density including the uniform phase) becomes

$$p_1(\theta,\omega,t|z_0^t) = \frac{e^{x(\omega)\sin\theta + y(\omega)\cos\theta}\, e^{-(w^2/2)q/2f}}{W_1} \tag{6.90}$$

where

$$W_1 = \int_{-\infty}^{\infty}\int_{-\pi}^{\pi} e^{x(\omega)\sin\theta + y(\omega)\cos\theta}\, e^{-(w^2/2)q/2f}\, d\theta\, d\omega \tag{6.91}$$

$$x(\omega) = \frac{1}{2r}\int_0^t e^{-(P_{s/t})_{11}/2}\left(\cos[\omega\,\frac{1-e^{-f\Delta}}{f}]dz_1 + \sin\omega\,[\frac{1-e^{-f\Delta}}{f}]dz_2\right) \tag{6.92a}$$

$$y(\omega) = \frac{1}{2r}\int_0^t e^{-(P_{s/t})_{11}/2}\left(\cos\omega\,[\frac{1-e^{-f\Delta}}{f}]dz_2 - \sin\omega\,[\frac{1-e^{-f\Delta}}{f}]dz_1\right) \tag{6.92b}$$

$$(P_{s/t})_{11} = \frac{q}{f^2}\Delta - \frac{q}{2f^3}(3-4e^{-f\Delta} + e^{-2f\Delta}) \tag{6.93}$$

and

$$\Delta = t-s \tag{6.94}$$

The statistics $x(\omega)$ and $y(\omega)$ contain integrals of complex functions of s and ω, and therefore will be difficult to approximate. Nonetheless, some interesting facts about the filter may be obtained by a simple examination. We let

$$x(\omega) = x_s + x_\ell \tag{6.95a}$$

$$y(\omega) = y_s + y_\ell \tag{6.95b}$$

where

$$\begin{pmatrix} x_s \\ y_s \end{pmatrix} = \int_\tau^t \begin{pmatrix} f_x(s,t,\omega) \\ f_y(s,t,\omega) \end{pmatrix} dz \tag{6.96}$$

and

$$\begin{pmatrix} x_\ell \\ y_\ell \end{pmatrix} = \int_0^\tau \begin{bmatrix} f_x(s,t,\omega) \\ f_y(s,t,\omega) \end{bmatrix} dz \tag{6.97}$$

for

$$f_x(s,t,\omega) = e^{-(P_{s/t})_{11}/2} \begin{pmatrix} \cos \omega \left[\frac{1-e^{-f\Delta}}{f} \right] \\ \sin \omega \left[\frac{1-e^{-e\Delta}}{f} \right] \end{pmatrix}^T \tag{6.98}$$

and

$$f_y(s,t,\omega) = e^{-(P_{s/t})_{11}/2} \begin{pmatrix} -\sin \omega \left[\frac{1-e^{-f\Delta}}{f} \right] \\ \cos \omega \left[\frac{1-e^{-f\Delta}}{f} \right] \end{pmatrix}^T \tag{6.99}$$

We want to divide the integrals from 0 to t into two regions, one where Δ is "small" (τ to t) and one where Δ is "large" (0 to τ). Then for small Δ we have that, to order Δ^3 *

$$(P_{s/t})_{11} = \frac{q}{3} \Delta^3 \tag{6.100}$$

---

*Interestingly $(P_{s/t})_{11} = 0$ to orders Δ and Δ^2.

and to order Δ

$$\frac{1-e^{-f\Delta}}{f} = \Delta \tag{6.101}$$

Then we may approximate

$$\begin{pmatrix} x_s \\ y_s \end{pmatrix} \simeq \int_{\tau}^{t} e^{-\frac{q}{6}\Delta^3} \begin{bmatrix} \cos \omega \Delta & \sin \omega \Delta \\ -\sin \omega \Delta & \cos \omega \Delta \end{bmatrix} dz \tag{6.102}$$

which is very similar to the Brownian-motion-frequency filter derived in section 6.3. For large Δ however,

$$(P_{s/t})_{11} \simeq \frac{q}{f^2} \Delta \tag{6.103}$$

$$\frac{1-e^{-f\Delta}}{f} = \frac{1}{f} \tag{6.104}$$

and

$$\begin{pmatrix} x_\ell \\ y_\ell \end{pmatrix} \simeq \begin{pmatrix} \cos \frac{\omega}{f} \ \bar{x} + \sin \frac{\omega}{f} \bar{y} \\ \cos \frac{\omega}{f} \ \bar{y} - \sin \frac{\omega}{f} \bar{x} \end{pmatrix} \tag{6.105}$$

where

$$\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} = \int_0^{t} e^{-\frac{q}{f^2}\Delta} \begin{pmatrix} dz_1 \\ dz_2 \end{pmatrix} \tag{6.106}$$

Thus, the recent information (small Δ) is treated like that of the Brownian motion frequency filter with a Δ^3 weighting, while the old data is treated, for phase estimation, like that of the static phase filters with Δ weighting. It is interesting that the old data is not used for frequency estimation, however. To see this, we examine the statistic $F(\omega)$ (see equation 6.28) for the old data states x_ℓ and y_ℓ. We have that

$$F_\ell(\omega) = x_\ell^2(\omega) + y_\ell^2(\omega) = \bar{x}^2 + \bar{y}^2 \qquad (6.107)$$

which is independent of ω!

This filter, unlike that of section 6.3, recognizes that the older data does not contain useful frequency information. (The filter of section 6.2 is designed for a constant frequency, and the old data is useful in that case.) Finally, we note that if we let $q \to 0$ in the filter of equation 6.92, we obtain the (optimal) static phase filter first encountered in section 3.3.2. This is because we defined the initial frequency covariance to be equal to its steady-state value

$$P_{22_0} = q/2f \qquad (6.108)$$

so that setting q to 0 removes all frequency uncertainty, and the problem reduces to the no-process-noise phase estimation problem discussed earlier.

6.5 Summary

In this chapter we have applied the first cumulant approximation method to the design of sub-optimal filters for three second-order PLL problems. We have shown that this method produces complex but intuitively pleasing filters which converge to the optimum as the process-noise strength goes to zero. The implementation of these filters awaits further investigation and will require certain approximations. If frequency estimation is desired, a bank of filters spanning the range of possible frequencies will be needed. In addition, for the filters with inseparable functions of s and t, the impulse response of each filter will have to be approximated by a finite number of separable functions in the manner of equation 6.47.

Despite the implementation difficulties, we feel that these filters represent practical approaches to phase and frequency estimation in high-noise environments. Clearly, in low-noise applications, there is little reason <u>not</u> to use the classic PLL, which approximates the optimal linear filter after acquisition. (If acquisition in low-noise areas is a problem, we suggest considering the compound PLL structure of chapter 2.) For high-noise applications, however, the infinite-dimensional bank of filters (e.g. that discussed in section 6.2) is closer to the optimal (q=0) parameter estimation result than the compact but poorly-performing PLL.

This facet of our approximation method will be, we suspect, typical of most of its applications. The approximate filters generated by the methods of chapter 4 are closer to multiple-hypothesis testing detectors than to filters generated for the linearized problem. This is not surprising, since the same likelihood ratio used in the hypothesis tests is used in the representation theorem on which our approximations are based. We therefore believe that the ability to interpret the approximate density filters as modifications of the linearized filters (e.g. the static phase filters as "special" PLL's as discussed in section 5.4.1) will be the exception rather than the rule. We do not consider this to be a drawback to our method at all, since if the noise is low, the linearized filters will probably work quite well, while if the noise is high, there is no reason to expect the linearized filters to work at all. It is in the high-noise region, where the linearized filters break down, that alternate filters are needed, and it is here that we expect our method to help, both by directly providing useful filter structures and by giving a great deal of insight into the characteristics of the particular nonlinear filtering problem being considered.

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

We have considered a class of nonlinear filtering problems characterized by a Gaussian state, nonlinear measurement, and additive white Gaussian measurement noise. Much of this thesis has been devoted to first- and second-order phase-lock loop problems, both because they are examples of the general class and because they are interesting in their own right. A particularly interesting feature of these phase-lock loop problems is that PLL's are extended Kalman filters, and represent practical performance benchmarks for any other nonlinear filter. We have been concerned with two aspects of these problems where the usual linearized analysis (and the PLL) breaks down: acquisition in low noise and state estimation in high noise.

Our major contributions have been:

1) For high signal-to-noise ratio applications, we have developed a technique, the compound PLL, for greatly improving PLL acquisition performance without significantly affecting the (optimal) loop noise attenuation.

2) For the general problem with high noise, we have developed a method for approximating the conditional density function that requires no a priori assumptions about the shape or moments of the density. This technique results in an approximate density, constructed from a set of statistics which are functionals of the measurements, which may be used

to obtain state and performance estimates.

3) We have applied this general method to the design of phase estimators without making the low-noise assumptions inherent in PLL's. For the high-noise case simulated, the first term in our approximation resulted in a filter which out-performed the PLL.

4) In the course of applying the approximation method to the first-order PLL problem, we have examined several sub-optimal filters and compared them to the phase-lock loop. We have noted "hidden" filter equivalences and discussed low-noise convergence for some of the designs. We have also developed a modification of the approximate density filter in 3 (above) that offers improved performance at a slight increase in complexity.

The common thread through our analyses has been the avoidance of the linearizing assumptions usually used in nonlinear problems. This has allowed us to "use" the nonlinear properties of one phase-lock loop to improve the acquisition performance of another loop for the compound PLL. It has also allowed us to generate approximate-density filters for the <u>full nonlinear problem</u> that are free from Gaussian prejudices. This has been both a blessing and a curse: some of the approximate densities for the first-order PLL problem seemed quite reasonable for the problem, but unfortunately we were unable to obtain good phase estimates from the unfamiliar density forms. We expect more of these problems as other non-Gaussian densities arise in nonlinear estimation.

This leads to a discussion of future research topics which have been suggested by this thesis:

1) The compound loop analysis is relatively complete, we feel, for phase-lock loops. Simple acquisition aids, however, will always be in demand for linearized filters that need to be brought into their linear performance range. The acquisition problem has hardly been solved in such a general context.

2) The approximation technique is still fertile ground for development also. It should be possible to strengthen our convergence results for the approximations. It also seems possible to obtain tighter performance bounds on the use of approximate-density filters.

3) We hope it will prove possible to simplify the first-cumulant approximation so that problems of higher dimension may be handled more easily. It seems reasonable that the Gaussian state and first-cumulant method may be combined to obtain differential equations for the filter states directly. We now use two distinct steps to obtain these states: a backward conditional expectation followed by a forward integration.

4) As discussed above, the approximate densities are useless without a methodology for obtaining state and performance estimates. In nonlinear filtering in general, we believe there has been too much emphasis on mean and convariance estimation, due to our Gaussian upbringing, and not enough emphasis on other density statistics that may better fit the problem. We hope this thesis kindles such an interest.

5) Finally, we hope the approximation method, whether simplified or not, will be tried on other problems where the noise is high enough and the nonlinearity strong enough to make extended Kalman filters ill-

suited to the task. For these problems, where the payoff is high, we hope to see approximate-density filters perform well. Also, in these problems, the benefits of the higher-order approximations may be more apparent.

APPENDIX A

Bessel Function Relationships

At several places in this thesis (e.g., sections 1.4.3.1, 3.3.2, and 5.5.1) we use density numerators of the form

$$\hat{u} = e^{\alpha \cos (\theta-\beta)} \tag{A.1}$$

or

$$\hat{u} = e^{x \sin \theta + y \cos \theta} \tag{A.2}$$

for

$$\beta = \tan^{-1} x/y \tag{A.3}$$

$$\alpha = \sqrt{x^2 + y^2} \tag{A.4}$$

Bessel functions arise naturally when integrating these numerators.

We define the modified Bessel function $I_n(\alpha)$ for any integer n (including zero) as (see Abramowitz and Stegun [1])

$$I_n(\alpha) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(n\,\theta)\, e^{\alpha \cos \theta}\, d\theta \tag{A.5}$$

This is a positive function, for any positive α, with

$$I_n(0) = 0 \qquad n \neq 0 \tag{A.6}$$

$$I_0(0) = 1 \tag{A.7}$$

$$I_n(\alpha) = I_{-n}(\alpha) \tag{A.8}$$

Several interesting relationships follow from definition A.5. In particular, we have the recurrence relations

$$I_{n-1}(\alpha) - I_{n+1}(\alpha) = \frac{2n}{\alpha} I_n(\alpha) \tag{A.9}$$

$$\frac{dI_n(\alpha)}{d\alpha} = I_{n-1}(\alpha) - \frac{n}{\alpha} I_n(\alpha) \tag{A.10}$$

$$\frac{dI_n(\alpha)}{d\alpha} = I_{n+1}(\alpha) + \frac{n}{\alpha} I_n(\alpha) \tag{A.11}$$

and

$$\frac{dI_n(\alpha)}{d\alpha} = \frac{1}{2} (I_{n-1}(\alpha) + I_{n+1}(\alpha)) \tag{A.12}$$

Also, since

$$I_1(\alpha) = I_{-1}(\alpha) \tag{A.13}$$

from equation A.8, we have (using equation A.11)

$$\frac{dI_0(\alpha)}{d\alpha} = I_1(\alpha) \tag{A.14}$$

For convenience, we also will define the Bessel function ratio

$$g_n(\alpha) = \frac{I_n(\alpha)}{I_0(\alpha)} \tag{A.15}$$

which is a well-defined function of positive α with

$$g_n(0) = 0 \qquad n \neq 0 \tag{A.16}$$

$$g_0(0) = 1 \tag{A.17}$$

and

$$\lim_{\alpha\to\infty} g_n(\alpha) = 1 \qquad n\neq 0 \tag{A.18}$$

We may derive recurrence relations for g_n from equations A.9 and A.12 as

$$g_{n-1}(\alpha) - g_{n+1}(\alpha) = \frac{2n}{\alpha} g_n(\alpha) \tag{A.19}$$

and

$$\frac{dg_n(\alpha)}{d\alpha} = \frac{g_{n-1}(\alpha) + g_{n+1}(\alpha)}{2} - g_1(\alpha) g_n(\alpha) \tag{A.20}$$

Finally, we remark that, in addition to [1], Watson [42] and Erdélyi, et al. [13], are excellent references on Bessel functions.

APPENDIX B

Linear Error Equations

B.1 Classic PLL

In section 2.3.3.3., we define the state

$$x = \begin{bmatrix} \varepsilon \\ y \end{bmatrix} \tag{B.1}$$

where ε is the phase error and y is the output of the a/s integration in Figure 2.6. This state propagates according to the differential equation

$$\dot{x} = \begin{bmatrix} -K & -1 \\ aK & 0 \end{bmatrix} x + \begin{bmatrix} 1 & -K \\ 0 & aK \end{bmatrix} \begin{pmatrix} \dot{u} \\ \dot{n}' \end{pmatrix} \tag{B.2}$$

The convariance of this system therefore propagates as

$$\dot{P}_{11} = -2KP_{11} - 2P_{12} + q + 2r\ K^2 \tag{B.3}$$

$$\dot{P}_{12} = aKP_{11} - KP_{12} - P_{22} - 2raK^2 \tag{B.4}$$

$$\dot{P}_{22} = 2aKP_{12} + 2ra^2K^2 \tag{B.5}$$

This may be solved for the steady-state covariances

$$P_{12} = -\ arK \tag{B.6}$$

and

$$P_{11} = -P_{12}/K + q/2K + rK \tag{B.7}$$

---

*P_{ij} is the ij th element of the 2x2 covariance matrix for the vector x.

so that

$$P_{11} = r(a + K) + \frac{q}{2K} \tag{B.8}$$

where P_{11} is the phase error variance.

B.2 Compound PLL

To obtain the error equations for

$$\varepsilon = \theta - \hat{\theta} \tag{B.9}$$

in the compound loop (section 2.3.4.3), we find it useful to consider $\hat{\varepsilon}$ as a second state variable, although the variance of $\hat{\varepsilon}$ is of no interest. We consider the error equations

$$\dot{\varepsilon} = -K_2 \hat{\varepsilon} + \dot{u} \tag{B.10}$$

$$\dot{\hat{\varepsilon}} = K_1(\varepsilon - \hat{\varepsilon} + \dot{n}') \tag{B.11}$$

If we let

$$x = \begin{bmatrix} \varepsilon \\ \hat{\varepsilon} \end{bmatrix} \tag{B.12}$$

then

$$\dot{x} = \begin{bmatrix} 0 & -K_2 \\ K_1 & -K_1 \end{bmatrix} x + \begin{bmatrix} 1 & 0 \\ 0 & K_1 \end{bmatrix} \begin{pmatrix} \dot{u} \\ \dot{n}' \end{pmatrix} \tag{B.13}$$

The convariance of this system propagates as

$$\dot{P}_{11} = -2K_2 P_{12} + q \tag{B.14}$$

$$\dot{P}_{12} = -K_2 P_{22} + K_1 P_{11} - K_1 P_{12} \tag{B.15}$$

$$\dot{P}_{22} = 2K_1 P_{12} - 2K_1 P_{22} + 2rK_1^2 \tag{B.16}$$

The steady-state covariance is therefore

$$P_{12} = \frac{q}{2K_2} \tag{B.17}$$

$$P_{22} = P_{12} + rK_1 \tag{B.18}$$

$$= \frac{q}{2K_2} + rK_1$$

$$P_{11} = \frac{K_2}{K_1} P_{22} + P_{12} \tag{B.19}$$

$$= \frac{q}{2K_2} + \frac{q}{2K_1} + rK_2$$

APPENDIX C

Stochastic Calculus

Several processes and filters in this thesis obey equations of the form

$$x_t = \int_0^t f(x_\tau)\,d\tau + \int_0^t g(x_\tau)\,du_\tau \tag{C.1}$$

which we write

$$dx = f(x)\,dt + g(x)\,du \tag{C.2}$$

where

$$E[du\ du^T] = Q\,dt \tag{C.3}$$

These are not ordinary integral and differential equations, but Ito (stochastic) equations. The presence of a function (g) of x multiplying the Brownian motion separates these equations from those of the usual calculus, and special care must be taken in dealing with them. This appendix will not attempt to justify any of the rules we will present, but only list them for reference. The reader is referred to Wong [45], Jazwinski [19], or McKean [31] for details.

The first "problem" with stochastic equations is that the chain rule does not hold. That is, for a twice-differentiable scalar function (ψ) of the vector x, the derivative of ψ is <u>not</u> just

$$d\psi = \left(\frac{\partial\psi}{\partial x}\right)^T dx \tag{C.4}$$

but instead

$$d\psi = \left(\frac{\partial\psi}{\partial x}\right)^T dx + \frac{1}{2}\,\mathrm{tr}\,\{g(x)Qg^T(x)\psi_{xx}\}dt \tag{C.5}$$

where "tr" denotes the trace of the matrix argument and

$$(\psi_{xx})_{ij} = \frac{\partial^2\psi}{\partial x_i \partial x_j} \tag{C.6}$$

The second difficulty in analyzing equations of the form of (C.2) is that implementation or simulation of stochastic differential equations requires "Wong-Zakai" correction terms in order to compensate correctly for the difference between ordinary "white noise" (Wiener) integrals (where g is not a function of x) and the stochastic integrals in our analysis (see Wong [45] p. 161, or [46] or [47]). This means that the Ito equation for the <u>scalar</u> x

$$dx = f(x)dt + g(x)du \tag{C.7}$$

with

$$E[du^2] = (1)dt \tag{C.8}$$

is simulated (e.g., a Runge-Kutta integration routine) or implemented by the equation

$$\dot{x} = f(x) - \frac{1}{2}\left[g(x)\,\frac{\partial g(x)}{\partial x}\right] + g(x)\dot{u} \tag{C.9}$$

where $\dot{u}$ is implemented by Gaussian "white noise" (noise with a flat spectral density of unit height over the frequency range of interest). For digital simulations with an integration step size of Δ, $\dot{u}$ is approximated by a piece-wise constant (for interval lengths Δ) function of time. The value of this function over each interval is obtained from a sequence of independent, zero-mean, Gaussian random variables with variances of $1/\Delta$.

The vector correction terms are more awkward. For a vector x described by the Ito equation C.2 with Q (equation C.3) equal to the identity matrix, the ith component of x is implemented by

$$\dot{x}_i = \left[f_i(x) - \frac{1}{2} \sum_{jK} g_{Kj}(x) \frac{\partial g_{kj}(x)}{\partial x_K} \right] + \sum_j g_{ij}(x) \dot{u}_j \qquad (C.10)$$

where each $\dot{u}_j$ is obtained as above.

APPENDIX D

Steady-State Density for α

We want to justify our assumption that $\bar{\alpha}$ goes to a steady-state value in equation 5.104. First, we show that $\bar{\alpha}^2$ goes to a steady-state. We then argue that all of the even moments of α go to steady-state, although no general formula is available. (This is the missing link in our proof.) Finally, we point out that if all of the even moments of a positive random variable go to steady-state, then the density of the random variable (and therefore all functions and moments) goes to a steady-state value.

In equation 5.29 α is defined as

$$\alpha = \sqrt{x^2 + y^2} \tag{D.1}$$

where x and y may be written as

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2r} \int_0^t e^{-f(t-\tau)} \, dz_\tau \tag{D.2}$$

for

$$dz = \begin{pmatrix} \sin\theta \\ \cos\theta \end{pmatrix} dt + dn \tag{D.3}$$

We cannot obtain $E[\alpha]$ directly because of the square root, but we may write

$$E[\alpha^2] = \int_0^t \int_0^t \frac{1}{4r^2} e^{-f[2t-\tau-s]} \, E[dz_{1_\tau} dz_{1_s} + dz_{2_\tau} dz_{2_s}] \tag{D.4}$$

Now we may write

$$dz_{1_s} dz_{1_\tau} = \sin\theta_s \sin\theta_\tau \, ds d\tau + \sin\theta_s \, ds dn_{1_\tau} \tag{D.5}$$

$$+ \sin\theta_\tau \, d\tau dn_{1_s} + dn_{1_s} dn_{1_\tau}$$

$$dz_{2_s} dz_{2_\tau} = \cos\theta_s \cos\theta_\tau \, ds d\tau + \cos\theta_s \, ds \, dn_{2_\tau} \tag{D.6}$$

$$+ \cos\theta_\tau \, d\tau \, dn_{2_s} + dn_{2_s} dn_{2_s} dn_{2_\tau}$$

Since n_t is a Brownian motion process, we have that

$$E[\alpha^2] = \int_0^t \int_0^t \frac{1}{4r^2} e^{-f[2t-\tau-s]} \{E[\cos(\theta_\tau - \theta_s)] \tag{D.7}$$

$$+ 4r \, \delta(\tau - s)\} d\tau ds$$

Also, since θ_t is a Brownian motion, we may use the folded-normal density relations (equation 5.76) to obtain

$$E[\cos(\theta_\tau - \theta_s)] = e^{-\frac{q}{2}|\tau - s|} \tag{D.8}$$

Thus

$$E[\alpha^2] = \int_0^t \int_0^t \frac{1}{4r^2} e^{-f[2t-\tau-s]} e^{-\frac{q}{2}|\tau-s|} d\tau ds \tag{D.9}$$

$$+ \frac{1}{2rf}(1 - e^{-2ft})$$

which may be written (since τ and s are interchangeable above)

$$E[\alpha^2] = \frac{1}{2r^2} \int_0^t \left[\int_0^\tau e^{-f[2t-\tau-s] - \frac{q}{2}(\tau-s)} ds \right] d\tau \qquad \text{(D.10)}$$

$$+ \frac{1}{2rf}(1-e^{-2ft})$$

Equation D.10 is solved by

$$E[\alpha^2] = \frac{1}{2r^2} \left[\frac{1-e^{-2ft}}{2f(f+q/2)} - \frac{e^{-t(f+q/2)} - e^{-2ft}}{f^2-(q/2)^2} \right] \qquad \text{(D.11)}$$

$$+ \frac{1}{2rf}(1-e^{-2ft})$$

for

$$f \neq q/2 \qquad \text{(D.12)}$$

and by

$$E[\alpha^2] = \frac{1}{2r^2} \left[\frac{1-e^{-qt}}{q^2} - \frac{t}{q} e^{-qt} \right] + \frac{1-e^{-qt}}{rq} \qquad \text{(D.13)}$$

for

$$f = q/2 \qquad \text{(D.14)}$$

In either case, (equation D.11 or D.13)

$$\lim_{t\to\infty} E[\alpha^2] = \frac{1}{2r^2(2f)(f+\frac{q}{2})} + \frac{1}{2rf} \qquad \text{(D.15)}$$

which becomes, for $f=q/2$

$$\lim_{t\to\infty} E[\alpha^2] = \frac{1}{2r^2q^2} + \frac{1}{rq} \qquad \text{(D.16)}$$

We cannot evaluate the density for α

$P_{\alpha}(\alpha, t)$

or the odd moments, $E[\alpha^{2n-1}]$, because of the square root in equation D.1. But we know that the second moment goes to a steady-state value (equation D.15), and we suspect that all even moments, $E[\alpha^{2n}]$, also go to steady-state, although no general formula is available. If this were proven, it would be sufficient to guarantee that $P_{\alpha}(\alpha,t)$ went to steady-state. To see this, we define a new random variable

$$y = \alpha^2 \tag{D.17}$$

which is a well-defined one-to-one function of α since α is non-negative. We assume that all of the moments of y (all of the even moments of α)

$$E[y^n] = E[\alpha^{2n}] \tag{D.18}$$

go to steady-state values. Then clearly $p_y(y,t)$, the density for y, goes to a steady state, since it may be formed from the inverse Fourier transform of the characteristic function of y, which is specified by all of the moments. This means that

$$P_{\alpha}(\alpha,t) = 2\alpha P_y(\alpha^2, t) \tag{D.19}$$

must go to a steady state also, since the change of variables formula is independent of t.

APPENDIX E

Third-Moment Approximate-Density Equations

This appendix evaluates the third moment

$$E_0^{\mathcal{B}}(\zeta_t')^3 = \frac{1}{8r^3} E_0^{\mathcal{B}} \int_0^t \int_0^t \int_0^t \begin{pmatrix} \sin\theta_{t_1} \\ \cos\theta_{t_1} \end{pmatrix} dz_{t_1} \begin{pmatrix} \sin\theta_{t_2} \\ \cos\theta_{t_2} \end{pmatrix}^T dz_{t_2} \begin{pmatrix} \sin\theta_{t_3} \\ \cos\theta_{t_3} \end{pmatrix}^T dz_{t_3} \tag{E.1}$$

for the first-order PLL problem approximate-density filters (see section 5.7.3). After carrying out the multiplications in equation E.1, we obtain integrands of the form (cosine also)

$$\sin\,(\theta_{t_1} \pm \theta_{t_2} \pm \theta_{t_3})\; dz_{i_{t_1}}\; dz_{j_{t_2}}\; dz_{K_{t_3}}$$

Since θ_{t_1} is a Brownian motion (conditioned on θ_t for $t_1<t$) with

$$p(\theta_{t_1}|\theta_t) = N(\theta_t,\, q(t-t_1)) \tag{E.2}$$

we have that

$$p(\theta_{t_1} + \theta_{t_2} - \theta_{t_3}|\theta_t) = N(\theta_t,\, q[t-t_1-t_2-t_3-2\max(t_1, t_2) + 2\max(t_1, t_3) + 2\max(t_2, t_3)] \tag{E.3}$$

and

$$p(\theta_{t_1} + \theta_{t_2} + \theta_{t_3}|\theta_t) = N(3\theta_t,\, q[9t-t_1-t_2-t_3-2\max(t_1, t_2) - 2\max(t_2, t_3) - 2\max(t_1, t_3)]) \tag{E.4}$$

Using equations E.3 and E.4 and the folded normal identities (for $p(\psi) = N(m, \gamma)$)

$$E\begin{bmatrix}\sin\psi \\ \\ \cos\psi\end{bmatrix} = e^{-\gamma/2}\begin{bmatrix}\sin m \\ \\ \cos m\end{bmatrix}, \tag{E.5}$$

we find (after much simplification) that

$$E_0^{\beta}(\zeta_t')^3 = 6[x_s \sin\theta_t + x_c \cos\theta_t + x_3 \sin 3\theta_t + y_3 \cos 3\theta_t] \tag{E.6}$$

where

$$dx_s = -\frac{q}{2} x_s\, dt + \frac{1}{8r}[(x_{12} + x_{21})dz_2 + (x_{11} - x_{22})dz_1] + \frac{1}{4r}(y_{11} + y_{22})dz_1 \tag{E.7}$$

$$dx_c = -\frac{q}{2} x_c dt + \frac{1}{8r}[(x_{12} + x_{21})dz_1 + (x_{22} - x_{11})dz_2] + \frac{1}{4r}(y_{11} + y_{22})dz_2 \tag{E.8}$$

$$dx_3 = -\frac{9}{2} qx_3 dt + \frac{1}{8r}[(x_{22} - x_{11})dz_1 + (x_{12} + x_{21})dz_2] \tag{E.9}$$

$$dy_3 = -\frac{9}{2} qy_3 dt + \frac{1}{8r}[(x_{22} - x_{11})dz_2 - (x_{12} + x_{21})dz_1] \tag{E.10}$$

with the "intermediate states"

$$dx_{11} = -2qx_{11}dt + \frac{1}{2r} x_1 dz_1 \tag{E.11}$$

$$dx_{12} = -2qx_{12}dt + \frac{1}{2r} x_1 dz_2 \tag{E.12}$$

$$dx_{21} = -2qx_{21}dt + \frac{1}{2r}\, y_1 dz_1 \qquad (E.13)$$

$$dx_{22} = -2qx_{22}dt + \frac{1}{2r}\, y_1 dz_2 \qquad (E.14)$$

$$dy_{11} = \frac{1}{2r}\, x_1 dz_1 \qquad (E.15)$$

$$dy_{22} = \frac{1}{2r}\, y_1 dz_2 \qquad (E.16)$$

and the familiar

$$dx_1 = -\frac{q}{2}\, x_1 dt + \frac{1}{2r}\, dz_1 \qquad (E.17)$$

$$dy_1 = -\frac{q}{2}\, y_1 dt + \frac{1}{2r}\, dz_2 \qquad (E.18)$$

The computer simulations of these states require Wong-Zakai correction terms as discussed in appendix C. Using the above differentials, we actually simulated

$$\dot{x}_s = \frac{dx_s}{dt} - \frac{1}{4r}\, x_1 \qquad (E.19)$$

$$\dot{x}_c = \frac{dx_c}{dt} - \frac{1}{4r}\, y_1 \qquad (E.20)$$

$$\dot{x}_3 = \frac{dx_3}{dt} \qquad (E.21)$$

$$\dot{y}_3 = \frac{dy_3}{dt} \qquad (E.22)$$

$$\dot{x}_{11} = \frac{dx_{11}}{dt} - \frac{1}{4r} \qquad (E.23)$$

$$\dot{x}_{12} = \frac{dx_{12}}{dt} \qquad (E.24)$$

$$\dot{x}_{21} = \frac{dx_{21}}{dt} \tag{E.25}$$

$$\dot{x}_{22} = \frac{dx_{22}}{dt} - \frac{1}{4r} \tag{E.26}$$

$$\dot{y}_{11} = \frac{dy_{11}}{dt} - \frac{1}{4r} \tag{E.27}$$

$$\dot{y}_{22} \quad \frac{dy_{22}}{dt} - \frac{1}{4r} \tag{E.28}$$

and

$$\dot{x}_{1} = \frac{dx_{1}}{dt} \tag{E.29}$$

$$\dot{y}_{1} = \frac{dy_{1}}{dt} \tag{E.30}$$

In our simulations, the y_{11} and y_{22} states grew much faster than any of the x_{ij} states, so that the equations for x_s and x_c could be simplified to (cf. E.7 and E.8)

$$dx_s = -\frac{q}{2}\, x_s dt + \frac{1}{4r}\,(y_{11} + y_{22})\, dz_1 \tag{E.31}$$

$$dx_c = -\frac{q}{2}\, x_s dt + \frac{1}{4r}\,(y_{11} + y_{22})\, dz_2 \tag{E.32}$$

The simulated states became

$$\dot{x}_s = \frac{dx_s}{dt} - \frac{1}{8r}\, x_1 \tag{E.33}$$

$$\dot{x}_c = \frac{dx_c}{dt} - \frac{1}{8r}\, y_1 \tag{E.34}$$

REFERENCES

1. Abramowitz, M., and Stegun, I.A., eds., Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, Nat. Bureau of Standards, Dec. 1972.

2. Alspach, D.L., and Sorenson, H.W., "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," IEEE Trans. on Auto. Control, Vol. AC-17, No. 4, Aug. 1972, pp. 439-448.

3. de Bellescize, H., "La Reception Synchrone," Onde Elect., Vol. 11, June, 1932, pp. 230-240.

4. Bellman, R., A Brief Introduction to Theta Functions, Holt, Rinehart and Winston, New York, 1961.

5. Biswas, B.N. and Banerjee, P., "Range Extension of a Phase-Locked Loop," IEEE Trans. on Com., Vol. COM-21, No. 4, April 1973, pp. 293-296.

6. Bryson, A.E., Jr., and Ho., Y.-C., Applied Optimal Control, Ginn and Co., Waltham, Mass., 1969.

7. Bucy, R.S., "Nonlinear Filtering Theory," Correspondence, IEEE Trans. on Auto. Control, Vol. AC-10, No. 2, April 1965, p. 198.

8. Bucy, R.S., Hecht, C., and Senne, K.D., "An Engineer's Guide to Building Nonlinear Filters," Final Report, F.J. Seiler Research Lab., May 1972.

9. Bucy, R.S., and Joseph, P.D., Filtering for Stochastic Processes with Applications to Guidance, John Wiley and Sons, New York, 1968.

10. Bucy, R.S., and Mallinckrodt, A.J., "An Optimal Phase Demodulator, Stochastics, Vol. 1, Feb. 1973, pp.3-23.

11. Bucy, R.S. and Senne, K.D., "Digital Synthesis of Nonlinear Filters," Automatica, Vol. 7, 1971, pp. 287-298.

12. Bucy, R.S. and Youssef, H., "Fourier Realization of the Optimal Phase Demodulator," Proc. Fourth Symp. on Non-Linear Est., Sept. 1973, pp. 34-38.

13. Erdélyi, A., et al., eds., Higher Transcendental Functions, Volume II, McGraw Hill, New York, 1953.

14. Galdos, J.I., "Information and Distortion in Filtering Theory," Ph.D., M.I.T., Jan. 1975.

15. Gustafson, D.E., and Speyer, J.L., "Linear Minimum Variance Filters Applied to Carrier Tracking," IEEE Trans. on Auto. Control, Vol. AC-21, No. 1, Feb. 1976, pp. 65-73.

16. Hecht, C., "Digital Realization of Non-Linear Filters," Proc. Second Symp. on Non-Linear Est., Sept. 1971.

17. Hildebrand, F.B., Advanced Calculus for Applications, Prentice-Hall, Englewood Cliffs, N.J., 1962.

18. Ho, H.N., "Transient Analysis of Phase-Locked Loops in the Presence of Noise," Sc.D., George Washington U., Feb. 1975.

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

20. Kailath, T., "A Generalized Likelihood-Ratio Formula for Random Signals in Gaussian Noise," IEEE Trans. on Info. Theory, Vol. IT-15, No. 3, May 1969, pp. 350-361.

21. Kailath, T., "A Note on Least-Squares Estimates from Likelihood Ratios," Information and Control, Vol. 13, 1968, pp. 534-540.

22. Kailath, T., "A View of Three Decades of Linear Filtering Theory," IEEE Trans. on Info. Theory, Vol. IT-20, No. 2, March 1974, pp. 146-181.

23. Kendall, M.G., and Stuart, A., The Advanced Theory of Statistics, Volume 1, Distribution Theory, Charles Griffin, and Co., London, 1958.

24. Klapper, J. and Frankle, J.T., Phase-Locked and Frequency-Feedback Systems, Academic Press, New York, 1972.

25. Levieux, F., Filtrage Non-lineaire et Analyse Fonctionnelle, Rapport de Recherche No. 57, IRIA, Feb. 1974.

26. Levieux, F., "Functional Analysis Approach of the Partial Differential Equation Arising from Non-Linear Filtering Theory," Third Symp. on Non-Linear Est., Sept. 1972, pp. 230-243.

27. Levieux, F., "Treatment of Noise in Nonlinear Systems: Conception of New Algorithms and Applications," from New Directions in Signal Processing in Communications and Control, ed. by J.K. Skwirzynski, NATO Adv. Studies Inst. Series Noordhoff-Leyden, 1975.

28. Lo, J. T.-H., "Exponential Fourier Densities and Optimal Estimation and Detection on the Circle," to be published in IEEE Trans. on Info. Theory.

29. Mallinckrodt, A.J., Bucy, R.S., and Cheng, S.Y., "Final Project Report for a Design Study for an Optimal Non-Linear Receiver/ Demodulator," for NASA Goddard Space Flight Center, Contract No. NAS 5-10789, Aug. 1970.

30. Marcus, S.I., "Estimation and Analysis of Nonlinear Stochastic Systems," Ph.D., MIT May, 1975, ESL Report R-601.

31. McKean, H.P., Jr., Stochastic Integrals, Academic Press, New York, 1969.

32. Mortensen, R.E., "Optimal Control of Continuous-Time Stochastic Systems," U. Cal. Berkely, Elec. Res. Lab. Report ERL-66-1, Aug. 1966.

33. Rudin, W., Principles of Mathematical Analysis Second Edition, McGraw-Hill, New York, 1964.

34. Rudin, W., Real and Complex Analysis, McGraw-Hill, New York, 1966.

35. Synder, D.L., and Rhodes, I.B., "Filtering and Control Performance Bounds with Implications on Asymptotic Separation," Automatica, Vol. 8, 1972, pp. 747-753.

36. Sorenson, H.W., and Alspach, D.L., "Recursive Bayesian Estimation Using Gaussian Sums," Automatica, Vol. 7, 1971, pp. 465-479.

37. Stratonovich, R.L., Conditional Markov Processes and Their Application to the Theory of Optimal Control, Trans. by R.N. and N.B. McDonough, American Elsevier, New York, 1968.

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

39. Tam, P.K.S., and Moore, J.B., "A Gaussian Sum Approach to Phase and Frequency Estimation," U. of Newcastle, Australia, Report EE 7503, Jan. 1975.

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

41. Viterbi, A.J., Principles of Coherent Communication, McGraw-Hill, New York, 1966.

42. Watson, G.N., A Treatise on the Theory of Bessel Functions, Cambridge, U. Press, Cambridge, 1966.

43. Willsky, A.S., "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication - Part I: Analysis," _IEEE Trans. on Info. Theory_, Vol. IT-20, No. 5, Sept. 1974, pp. 577-583.

44. Willsky, A.S., "Fourier Series and Estimation on the Circle with Applications to Synchronous Communication - Part II: Implementation," _IEEE Trans. on Info. Theory_, Vol. IT-20, No. 5, Sept. 1974, pp. 584-590.

45. Wong, E., _Stochastic Processes in Information and Dynamical Systems_, McGraw-Hill, New York, 1971.

46. Wong, E., and Zakai, M., "On the Convergence of Ordinary Integrals to Stochastic Integrals," _Ann. Math. Stat._, Vol. 36, pp. 1560-1564.

47. Wong, E., and Zakai, M., "On the Relation Between Ordinary and Stochastic Differential Equations," _Int. J. Engrg. Sci._, Vol. 3, pp. 213-229.

BIOGRAPHY

John Eterno was born in Chicago, Illinois on February 9, 1950. He attended Emerson and Mary, Queen of Heaven grammar schools and Immaculate Conception High School in Elmhurst, Illinois, graduating in June, 1967. He then attended Case Institute of Technology of Case Western Reserve University in Cleveland, Ohio, where he majored in Fluid, Thermal and Aerospace Sciences, receiving the BS degree, with honors, in Engineering in June, 1971.

During the summer of 1971, he was a "summer intern" at the NASA Houston Space Center, now called the Johnson Space Center, where he worked on terminal area guidance techniques for the space-shuttle orbiter. In September, 1971, he enrolled in graduate school at MIT in the Department of Aeronautics and Astronautics, receiving the S.M. degree in February, 1974. While at MIT, he was a research assistant at the C.S. Draper Laboratory (formerly the Instrumentation Laboratory) where he worked in gravimetry, seismology, control, and signal processing.

Mr. Eterno is married to the former Becky Dickson of Bellefontaine, Ohio, who is presently working on her Ph.D. in English Literature at Boston University.