ESTIMATION FOR POISSON PROCESSES WITH APPLICATIONS IN OPTICAL COMMUNICATION

by

John Robert Clark

B.S., Stanford University
(1966)

M.S., Polytechnic Institute of Brooklyn
(1968)

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 1971

Signature of Author ______________________
Department of Electrical Engineering, July 27, 1971

Certified by ______________________
Thesis Supervisor

Accepted by ______________________
Chairman, Departmental Committee on Graduate Students

ESTIMATION FOR POISSON PROCESSES WITH APPLICATIONS IN OPTICAL COMMUNICATION

by

John Robert Clark

Submitted to the Department of Electrical Engineering on July 27, 1971 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

ABSTRACT

A systematic theory of least-squares estimation for Poisson processes with random intensities is developed in this thesis. It is shown that the MMSE estimator of any L_2 process embedded in the intensity can be represented as an explicit functional of the observations of the counting process, via a stochastic integral. This representation easily yields stochastic differential equations for many different types of estimators, and is applicable to causal filtering, prediction, non-causal filtering, and parameter estimation. This representation, as well as many subsidiary results presented in the thesis, are applicable almost verbatim to a much wider class of observations known as regular point processes. This extension is noted but not elaborated upon.

The estimators obtained from the stochastic integral representation are in general nonlinear functionals of the observations, and depend on estimates of all higher moments, making implementation difficult. To facilitate implementation, various types of approximations are developed. Exact estimators are obtained which are constrained to be linear functionals of the data; these possess a structure which is independent of the observations. Approximate nonlinear estimators are obtained by expanding exact estimators in powers of the estimator error. Truncated series of this form yield estimators which are easily realized in the form of feedback loops, and should perform well if the error is indeed small.

Performance of estimators is obtained, both as bounds on mean-square error, and as a partial differential-difference equation for the first-order probability density function of the actual estimator error. The latter is valid for a wide class of approximate estimators and some exact ones, and can be solved under certain circumstances; e.g., for narrowband systems in stead-state operation.

The theory is applied to some representative problems in optical communication. A specific example -- angle modulation on a subcarrier -- is examined in detail. A "quasi-optimum" demodulator is derived, and its performance in all ranges of operation, including threshold and below, is obtained. It is shown that the system behaves much as a classical feedback loop realization as long as the transmitted signal is not too weak.

A new statistical characterization, intimately connected with least-squares estimation, is obtained for Poisson processes with random intensities. In the form of a joint probability density function of event occurrence times, this characterization is potentially useful for many detection and estimation problems. For example, it leads immediately to a likelihood ratio for deciding between two Poisson processes with arbitrary random intensities.

THESIS SUPERVISOR: Estil V. Hoversten
TITLE: Associate Professor of Electrical Engineering

ACKNOWLEDGMENT

I wish to express my sincere gratitude to Professor Estil V. Hoversten for his guidance and encouragement throughout the course of this study. Under his supervision, my research at M.I.T. has been a most valuable educational experience. I am also grateful to Dr. Robert M. Lerner, Professor Robert S. Kennedy, and Professor Harry L. Van Trees for their careful reading of my thesis.

I am thankful to Professor Donald L. Snyder of Washington University, St. Louis, for his generosity in sharing ideas and results, and for his continued interest. My colleague, Richard S. Orr, deserves special thanks, both for his friendship, and for the many stimulating discussions we have had.

My work at M.I.T. was made possible by the generous support of the National Science Foundation and Lincoln Laboratory, for which I am most appreciative.

I am ever grateful to my wife, Jan, for her constant love and encouragement, as well as for her expert typing.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY OF SYMBOLS

A = a matrix; a class of ω sets; a constant

A^{-1} = inverse of a square matrix

A^c = complement of a class ($= \Omega - A$)

A_d = detector area; detector aperture

A_t = a scalar time function

a_t = a process

α = a constant

α_t = a process; a martingale

a.s. = almost surely

B = a class of ω sets; a constant; σ-algebra of subsets in Ω ; band-limits

$\overline{B}$ = a completed σ-algebra

$B(\cdot)$ = σ-algebra generated by $(\cdot)$

B_{x_t} = σ-algebra generated by $\{x_\tau, \tau \in [t_0, t]\}$

B_1, B_2 = σ-subalgebras of a σ-algebra B

B_λ = σ-algebra generated by $\{\lambda_t, t \in T\}$

B_N = σ-algebra generated by $\{N_t, t \in T\}$

B_t = a scalar time function

b = a constant

b_t = a process

β = a constant; η/hf_0; inverse time constant

β_j = a constant

β_t = a Wiener process

$\mathrm{Cov}(\cdot)$	=	covariance of $(\cdot)$
C	=	a matrix; a constant; a class of ω sets
C_t	=	a scalar time function
$C(\cdot)$	=	a cost function
$\chi_A(\cdot)$	=	indicator of a set A
$\not{C}$	=	Euler's constant (0.57721566...)
c	=	subscript denoting "classical"; a constant
γ	=	a constant; a phase modulation index
D	=	dimensionality; diversity measure; a diagonal matrix
D_x	=	Jacobean operator
DSPP	=	doubly-stochastic Poisson process
d_j	=	element of a diagonal matrix
$dg(\cdot)$	=	diagonal matrix with on-diagonal elements $(\cdot)$
$\mathrm{diag}(\cdot)$	=	$dg(\cdot)$
"$\xrightarrow{d}$"	=	convergence in distribution
dN_t	=	differential element of $\{N_t,\ t \in T\}$
$d\mathcal{N}_t$	=	diag (dN_t)
δ	=	a constant; a fading intensity parameter; $\sqrt{1+\gamma^2\delta}$
$\delta(\cdot)$	=	Dirac impulse function
$\delta_{x,\,y}$	=	Kronecker delta = 1 if $x = y$; 0 otherwise
Δ	=	positive constant; forward difference operator
Δt	=	positive time increment
∇_x	=	gradient operator

E	=	expectation operator
E_x	=	expectation relative to measure induced by $\{x_t, t \in T\}$
$E\{\cdot\|\cdot\}$	=	Conditional expectation
$\mathcal{E}$	=	a constant; energy
$\{e_j\}$	=	a set of unit vectors
ϵ	=	"is contained in," a positive constant
$Ei(\cdot)$	=	Euler's integral
η	=	quantum efficiency; mean of an r.v.
η_t	=	a process
F_t	=	a matrix function of time
f_t	=	a process
$f_t(x_t)$	=	a memoryless function of t and x_t
f_o	=	optical carrier frequency, Hertz
f_k	=	a function of k variables
$f(t)$	=	a time function
$\mathcal{F}$	=	a set; a σ-algebra
$\{\mathcal{F}_t\}$	=	an increasing family of σ-subalgebras
Ψ	=	an angle
$\Psi_t(\cdot)$	=	differential generator of a process
φ	=	total phase error = $\gamma(x_t - \hat{x}_t)$
$\emptyset$	=	an angle; empty set
$\emptyset_t$	=	a process; a function of t and a process x_t
$\{\emptyset_i\}$	=	a set of eigenfunctions
G	=	an ω set; a matrix

G_n = a Borel function of n variables

G_t = a matrix function of time

g_t = a process; an integrand

$g_t(x_t)$ = a memoryless function of t and x_t

$g(t)$ = a time function

$\mathcal{G}$ = a set; a σ-algebra

H = a Hilbert space; an ω set

$H_2\{\cdot\}$ = Hilbert space of nonlinear functionals of $\{\cdot\}$

H_x = Hessian operator

H_t^f = "selector" matrix for lognormal fading

h = Planck's constant

$h(\cdot,\cdot)$ = a function of two variables; a filter

h_t = a process; an integrand

$h(t)$ = time function; impulse response

$h_t(x_t)$ = a memoryless function of t and x_t

$h_t^{(\cdot)}$ = "selector" vector for the process $(\cdot)$

$\mathcal{H}$ = an ω set

I = an integral; an interval

I_t = an indefinite integral

$I^n[\cdot]$ = an iterated stochastic integral

$Im(\cdot)$ = imaginary part of $(\cdot)$

$I_t(N,z)$ = mutual information between $\{N_t, t \in T\}, \{z_t, t \in T\}$

$I_n(\cdot)$ = modified Bessel function of first kind

i = $\sqrt{-1}$; index

1_m = m-dimensional vector of 1's = $\sum_{i=1}^{m} e_i$

J = a constant

J_t = an indefinite integral; a matrix function; or $2A_tB_t$

J_{ij} = ij-element of a matrix J_t

$J_n(\cdot)$ = a Bessel function of the first kind

j = index of summation or sequence

K = a constant

K_t = a matrix function; or $2A_tC_t$

$K_x(\cdot,\cdot)$ = covariance kernel of the process $\{x_t, t \in T\}$

k = index of summation or sequence; a parameter

k_n = a function of n variables

$\{\kappa_i\}$ = cumulants; semi-invariants

L = a constant

$L^+[\cdot]$ = Kolmogorov forward operator

L_t = a matrix function; or $2B_tC_t$; a kernel

$L_2[T]$ = a space of functions square-integrable on T

L_2 = space of r.v.'s having finite second moments

L_1 = space of absolutely integrable r.v.'s

$L_k^{\alpha}(\cdot)$ = associated Laguerre polynomial

$\mathcal{L}$ = σ-algebra of Lebesgue-measurable sets in T

ℓ = index of summation

$\ln(\cdot)$ = natural logarithm (possibly vector-valued)

l.i.m. = limit in the mean

λ = a constant

λ_0 = a constant

$\{\lambda_i\}$ = a set of eigenvalues

λ_t = intensity of a counting process

$\lambda_t(B_{N_t})$ = intensity function of a RPP

$\hat{\lambda}_t$ = MMSE estimate of λ_t

Λ = arbitrary set contained in a σ-algebra

Λ_t = diag(λ_t), if λ_t is a vector

MAP = maximum a posteriori probability

ML = maximum likelihood

MMSE = minimum mean-square error

M = Hilbert subspace; a constant

M_t = $A_t^2 + B_t^2 + C_t^2$

$M_z(iv)$ = characteristic function of z

m = index of summation; a modulation index; a general analog message

m.s. = mean-square

$\hat{m}$ = estimate of a general analog message

m_t = mean of a counting process

μ = a constant

μ_0 = intensity of additive background noise

μ_k = $[\mu_0]_k$, if μ_0 is a vector

$\mu(\cdot)$ = a general measure

$\mu(t)$ = an intensity function, non-random

N = a constant; upper limit of a sum

N_0 = 2-sided height of noise power spectrum

N_t = a counting process (RPP, DSPP, or PP)

$N(\eta, \sigma^2)$ = normal: mean η, variance σ^2

$\mathcal{N}_t$ = $-N_t$; also, = diag (N_t) if N_t is a vector

n = a summation index

n_t = a process: $N_t - \int \lambda_{t'} dt'$

n_π = $n_{t,r}$ = background noise (complex envelope)

ν = a constant; optical carrier frequency (Hertz)

ν_t = a process: $N_t - \int \hat{\lambda}_{t'} dt'$

$\|\cdot\|$ = a general norm

$\|\cdot\|_1$ = $\Sigma|\cdot|$

$\|\cdot\|_2$ = $(\Sigma|\cdot|^2)^{\frac{1}{2}}$

$O(h)$ = quantity with property: $\lim_{h \to 0} O(h)/h < \infty$

$o(h)$ = quantity with property: $\lim_{h \to 0} o(h)/h = 0$

P = probability measure; a parameter

P.P. = Poisson process

$Pr\{\cdot\}$ = probability measure of the event $\{\cdot\}$

P_N, P_λ = marginal probability measures

P_t = conditional error covariance matrix

P_{ij} = ij-element of P_t

$\bar{P}_t$ = error covariance matrix

$\bar{P}$ = total average noise power, completed measure

$P_x(X)$ = cumulative distribution function of r.v. x

$p_x(X)$ = probability density function of r.v. x

p.d.f. = probability density function

$p(\{t_i\};N_t)$ = joint p.d.f. of $\{t_i\}$ and N_t (t fixed)

$p(z)$ = p.d.f. of z

$\xrightarrow{p}$ = convergence in probability

p.lim = limit in probability

p = subscript denoting "projection"

π = (t, r) if a subscript; otherwise, = 3.14159...

Q = a constant

Q_t = a time function

q = a constant; a general measure

q.m. = quadratic mean

q_t = a process

$q_t(\hat{x}_t)$ = a memoryless function of t and $\hat{x}_t$

R = the real line, $(-\infty,\infty)$; a constant

$\bar{R}$ = $[-\infty,\infty]$

R^n = n-dimensional Euclidean space

$\mathcal{R}^n$ = σ-algebra of Borel sets in R^n

Re(•) = real part of (•)

$\{R_t^j\}$ = a set of matrices

R_t = a time function

RPP = regular point process

r = index of summation; a constant

r_t = a process; a received signal

$r_t(\hat{x}_t)$ = a memoryless function of t and $\hat{x}_t$

r(t) = a received signal

r.v. = random variable

r.p. = random process

$\rho(\cdot,\cdot)$ = a general metric

SNR = signal-to-noise ratio

s = a summation index; a constant

s_t = a signal; a process

$\mathscr{S}$ = effective SNR in message bandwidth

$\mathscr{S}_c$ = $\mathscr{S}$ for a classically designed demodulator

σ = dummy variable of integration

σ^2 = variance

${\sigma_\varphi}^2$ = total phase error variance

${\sigma_z}^2$ = variance of an r.v. z

Tr = trace

T = parameter set over which t ranges

T' = a subset of T

t = a running time variable

$t-$ = $\lim\limits_{\Delta \to 0+} t - \Delta$

t_0, t_1 = endpoints of a time interval

$\{t_i\}$ = set of event times of a counting process or RPP

$\{t_i^{(n)}\}$ = a sequence of partitions of T

τ = a time variable; a stopping time

U = a constant, depending on M^2, μ_0

θ = an angle

θ_t = a process

u = a dummy variable

$u_{-1}(t)$ = unit step function

u_t = a process

V = a constant (steady-state value of P_t)

V_c = V for a classically designed demodulator

v = a dummy variable

v_t = a process

v_p = a projection

W = bandlimits

W_t = an $L_2[T]$ function

w = a dummy variable; a constant

w_t = a standardized Wiener process; a process

w.p.1 = with probability one

X = a metric space; a range variable for x

$\tilde{X}$ = a range variable for $x - \hat{x}$

x = an ω function; a general random variable

x_t = a general random process; an Ito process

$\hat{x}_t$ = $E\{x_t | B_{N_t}\}$; an estimate of x or x_t

$\hat{x}_{t|\tau}$ = $E\{x_t | B_{N_\tau}\}$

x_t^* = a candidate for $\hat{x}_t$

$x_t(\omega)$ = a function of the variables t, ω

$x_t^{(0)}$ = $x_t - P$

$\tilde{x}_t$ = $x_t - \hat{x}_t$

x_k = $[x_t]_k$ if x_t is a vector

ξ = a dummy variable

ξ_t = a martingale; an L_2 process

Y = a range variable for y

y = an ω function; a general r.v.

y_t = a general r.p.; an increasing r.p.; $\exp i v' x_t$

Z = a range variable for z

z = an ω function; a r.v. to be estimated

z_t = a r.p. to be estimated

$\hat{z}_t$ = estimate of z or z_t

z_t^* = candidate for $\hat{z}_t$

$\hat{z}_t^{\ell}$ = estimate of z_t linear in $\{v_t, t \in T\}$

$z^{(n)}$ = Chi-squared r.v., n degrees of freedom

Ω = collection of points; universe

ω = point in Ω

ω_N, ω_λ = points in Ω_N, Ω_λ

$\omega_\theta, \omega_\phi, \omega_A$ = radian frequencies

ζ_t = a martingale

(Ω, B) = measurable space

(Ω, B, P) = probability space

$\{\cdot\}$ = a set

$\{\omega: \ldots\}$ = an ω set

$\{x_t, t \in T\}$ = a r.p., a r.v. of which is x_t

$[\ ], [\ ), (\ ], (\ )$ = closed, semi-open, open intervals

$(\cdot,\cdot)$ = inner product in H

_____ = end of theorem

▮ = end of proof

$\bar{x}$ = conjugate of x

$\Omega_1 \times \Omega_2$ = product of the sets Ω_1, Ω_2

$A \otimes B$ = product σ-algebra

x' = transpose of x

CHAPTER I

Introduction

In the past few decades, estimation theory has developed into a vigorous and important part of the theory of statistical inference. Along with detection theory, it forms the basis for much of the significant research in communications, and provides a common language and methodology for a wide diversity of disciplines.

One of the notable features of the theory is that it is highly systematized. This is especially valuable in areas such as communications, where the theory must be repeatedly applied under widely varying conditions in order to design and analyze practical systems. Until very recently, most of the problems in communications involved Gaussian statistics*, for which the theory of statistical inference is well-developed. When the Gaussian assumption was utterly unjustifiable, non-parametric (robust) techniques were usually applied, often to the detriment of performance when the robustness wasn't really required.

In the last few years there has been a rapidly growing interest in statistical inference for decidedly

*Often as only an approximation to reality.

non-Gaussian data, in such fields as biomedical instrumentation, epidemiology, neurophysiology, and, especially, optical communications. In these areas the raw observations which must be processed are usually the instants of occurrence of like events which occur at random points in time. Such records are realizations of what are descriptively called <u>point processes</u>. A point process disregards the microstructure of the individual events; it is concerned only with the points in time at which the events occur. In epidemiology, an event might be the diagnosis of a contagious sickness in a member of a population; in neurophysiology, an event might be the firing of a nerve cell. In optical communication, an event is the detection of a photon by a photodetector. Such diverse physical phenomena as these often have identical statistical descriptions. Poisson processes, renewal processes, and birth-and-death processes, for example, are well-known point processes which accurately model a myriad of natural phenomena.

The theory of statistical inference for point processes is poorly developed at this time -- the concise, systematic methodology that applies to Gaussian processes has no counterpart in point process theory, so that new problems often must be explored without the guidance of precedent.

It is this lack of a systematic theory, which motivated the study undertaken in this thesis. The inadequacy of available techniques first became evident, when the solution of a class of continuous estimation problems in optical communication was attempted, without success. As the theory was developed to treat these problems, it became obvious that the techniques could enjoy exceptionally wide applicability. Although the predominant emphasis of this thesis is on optical communications, the reader will find it relatively easy to apply the theory in other fields. This is aided by the organization of the report, which carefully separates theory from application.

A point process which is sufficiently general to model most optical communication problems is the <u>doubly-stochastic Poisson process</u>[23,24,64-67]. This process, which has a random intensity, is a Poisson process for each realization of the intensity process; it is an accepted statistical model for many photodetection devices[47,48,54-62]. The majority of the results in this thesis apply almost verbatim to a much wider class of point processes -- those which can be characterized by instantaneous intensity processes. Called <u>regular point processes</u>[65,68], they can accurately model many physical point phenomena. Explicit remarks about this generalization are interspersed throughout the thesis.

In optical communication, detection theory has received considerably greater attention than estimation theory, owing perhaps to the attractiveness of often-simpler finite hypothesis problems. Consequently, we limit the scope of this study by confining our attention almost wholly to the estimation problem. We further delimit the thesis by ignoring heterodyne systems, for which classical descriptions in terms of Gaussian statistics are appropriate[46,49,51,52].

For the remainder of this chapter, we shall give a statement of the mathematical problems being considered, discuss the physical motivation for studying them, review previous work, and summarize the organization and main results of the thesis.

A. Problem Statement.

A precise, mathematical formulation appears in Chapter III, Section B, so we shall be brief, here.

Our primary task in this thesis will be the estimation of random processes embedded in the intensity of a doubly-stochastic Poisson process. In general, we allow vector observations, and we consider such types of estimation as causal filtering, prediction, non-causal filtering, and parameter estimation. We also examine the hypothesis testing problem, and its relation to parameter estimation.

The criterion of optimality for most of the estimators in this thesis is minimum mean-square error (MMSE); i.e., the MMSE estimator z_t of z_t minimizes the quantity,

$$E\{(z_t - \hat{z}_t)(z_t - \hat{z}_t)'\}$$

where prime ($'$) indicates vector transposition.

An additional task will be the development of practical suboptimum estimators, with the allied goals of demonstrating their utility in applications and analyzing their performance.

In the next section we show how this problem is relevant to optical communications, and draw some connections with past work.

B. Motivation and Background.

A fairly general analog optical communication system is illustrated in Figure 1.1. A message source generates a continuous waveform, representing the intelligence that is ultimately to be communicated. This waveform enters a modulator which changes it to a form suitable for subsequent transmission over the channel. The output of the modulator, a nonlinear functional of the message, in general, is then used to vary some property of the optical source. The source, usually a laser, transmits a modulated optical field into the channel. The latter can commonly be the clear atmosphere, free space, clouds, water, an optical waveguide -- any medium which can effect the transfer of optical energy. The channel introduces random perturbations such as turbulent refraction and fading (scintillation), scattering, delay, spatio-temporal dispersion, and background light. The corrupted optical signal is intercepted at the channel output by a preprocessor, which performs spatial and possibly temporal filtering and amplification of the field. The preprocessed field impinges on a photodetection device, which converts the optical field into an electrical signal for subsequent extraction of the message by the demodulator.

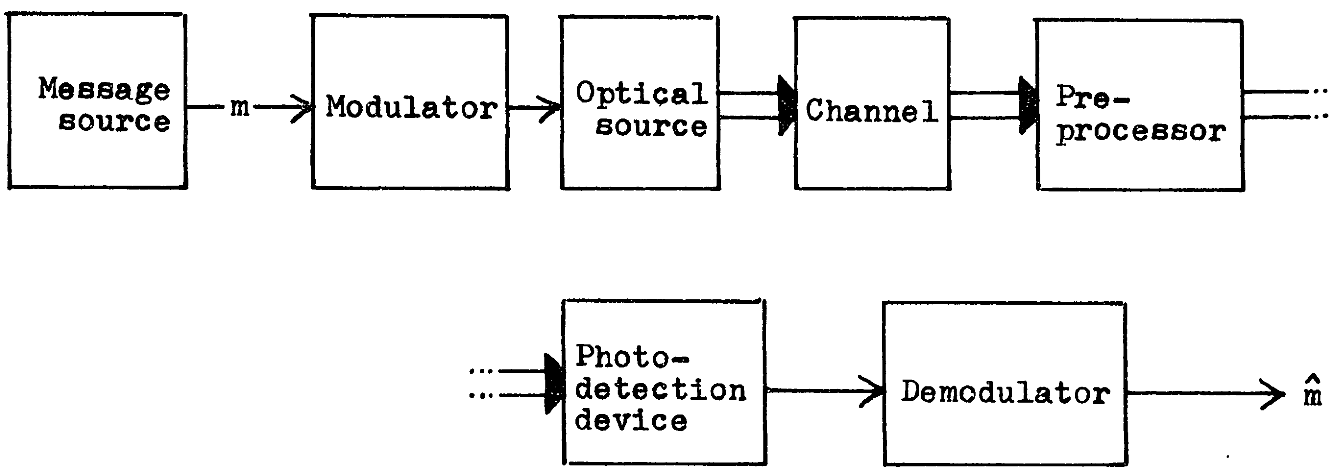

Message source
m
Modulator
Optical source
Channel
Pre-processor
Photo-detection device
Demodulator
m̂

Figure 1.1

Except for the fact that it uses optical frequencies, this system looks much like a classical communication system commonly employed at lower frequencies. There are many additional differences, however, which make the design of an optical communication system a drastically altered and more difficult problem. Most important, effects arising from the quantum nature of radiation cannot be neglected at optical frequencies, as they are at lower frequencies. Indeed, quantum effects may entirely swamp classical thermal effects at optical wavelengths. Thus, any methodology for design or analysis at these wavelengths must be firmly constructed on a foundation of quantum theory.

There are alternative ways to satisfy this requirement. In designing optimal receivers, for example, one can use a full quantum description of received fields, and optimize over an entire class of (possibly non-physical) measurements. This approach has received considerable attention recently[52,69-76,95,96]. On the other hand, one can use the considerably less general but more intuitive "structured" or "semi-classical" approach. This involves using a classical field description along with a specific photodetection device for which a quantum-mechanically-correct classical model exists. Optimization is then performed within these constraints using classical techniques. This approach has been very fruitful[15,25-28,37,49,96], and is employed exclusively in this thesis. Thus the reader need have no familiarity with quantum mechanics.

The specific photodetection device with which we are concerned is the ideal photodetector; classically, its output is a doubly-stochastic Poisson process with intensity proportional to the squared magnitude of the complex envelope of the incident field. To varying degrees of accuracy, this models such real devices as photomultipliers, photoconductors, and photodiodes[52,96]. Furthermore, it is accepted as an essentially correct quantum-mechanical description of such devices[47,48,54-62].

In our optical communication system, therefore, this establishes as a doubly-stochastic Poisson process the raw data from which the analog message must be extracted, and justifies the need to develop estimation techniques for point processes.

In quantum estimation, Helstrom[69,70] was the first to obtain Cramer-Rao bounds on mean-square error for parameter estimation. These were extended and generalized recently by Personick[75], who in his thesis developed and systematized quantum estimation theory for waveform channels. Except for these works, most of the research in quantum communication theory has dealt with the finite-hypothesis detection problem[71-74,95].

Stochastic point processes are discussed in many texts[1-5,36,78], often in terms of counting processes, which do nothing more than keep a running count of the number of past events. Beutler and Lenneman[82-85] have investigated the spectral properties of point processes in a series of papers. Macchi[79] has studied the representation problem for a wide class of point process, in terms of multicoincidences. However, these discussions do not touch on the statistical inference problem. Reiffen and Sherman[37] appear to be the first to have obtained a likelihood ratio for a point process hypothesis testing problem. Although the result was originally proved for only a narrow class of Poisson processes, with piecewise-constant intensities, it has since turned out to be a form having great generality. Bucknam[92] used Reiffen and Sherman's result to study the detection problem for the turbulent atmospheric optical channel. Bar-David[25] popularized a long-known statistical representation for Poisson processes, and used it to reproduce Reiffen and Sherman's result, and to solve some simple parameter estimation problems. This representation is gaining widespread use, and has been generalized by the author (See Chapter III, Section E of this thesis; also, Ref. 80), for doubly-stochastic Poisson processes, and more recently by Rubin[68], for regular point processes. Using the generalized representation, which is a joint probability density of event occurrence times and total number of events, Rubin has obtained a likelihood ratio valid for any regular point process. In an earlier work[81], he studied a variety of optimal and suboptimal detection schemes for renewal processes.

Karp and Clark[27] examined the first-order counting distribution of a doubly-stochastic Poisson process, and touched on the waveform communication problem. Clark[26] obtained unrealizable waveform estimators (smoothing filters), optimal in the MAP sense[33], for the turbulent atmospheric channel, and Clark and Hoversten[28] have studied approximate statistical models for photo-detectors. An interesting and potentially useful representation for the first-order counting statistics of a doubly-stochastic Poisson process with a lognormal intensity, has been proposed by Orr[31], and by Solimeno, Corti, and Nicoletti[32]. The latter have applied it to a detection problem in optical communication.

Snyder[23,24,44] was the first to obtain waveform estimators of broad applicability for doubly-stochastic Poisson processes. Using an analytical approach similar to that of his monograph[14], he has found stochastic differential equations for estimators of Markov processes. He was also the first to obtain general detection results for these processes. The approximate ("quasi-optimum") estimators developed by Snyder for scalar observations are extended to vector data in this thesis, and applied to some specific problems in optical communication.

Much of the recent fundamental work on estimation theory for Gaussian processes is relevant to this study. Of particular importance is the growing body of results built on the pioneering work of Kalman and Bucy[14,16,18,42,87-91]. The essence of these results is that estimators can be represented directly as solutions to stochastic differential equations valid on finite time intervals, rather than as the output of an optimum filter which satisfies some integral equation on a semi-infinite interval, as in the Wiener-Hopf theory. Stochastic differential equations play an important role in this thesis, as well as in the works just cited. As these works point out, manipulation of stochastic differential equations requires care, especially with regard to non-linear operations. A special calculus is required, based on the theory of square-integrable martingales[1-3,6-13]. In this calculus, quadratic and higher powers of differentials cannot be neglected in general, as they are in the deterministic calculus. This results in a new and unfamiliar set of operational rules, the neglect of which has resulted in some erroneous results in early work on nonlinear estimation[19].

C. Thesis Organization.

The first half of Chapter II develops a few of the essential ideas of the theory of point processes, especially doubly-stochastic Poisson processes. Peripheral properties which might be of interest are stated in an appendix. In another appendix, some of the important concepts of the mathematical theory of probability, as they apply to this thesis, are summarized very briefly. This appendix serves the purpose both of establishing notation, and of providing a quick reference for many of the mathematical definitions used in the report.

The stochastic calculus is used extensively in this thesis, so we devote the latter half of Chapter II to its exposition, in the context of point processes. Additional discussion is included in an appendix, for the benefit of the reader desiring additional background.

Chapter III contains the central mathematical results of the thesis. The important results are stated in a fairly general context, without explicit reference to optical communication.

Applications in optical communication are considered in Chapter IV. Models for turbulent optical channels and for optical detectors are discussed, as well as a mathematical model for real communication systems. A system is designed for the simultaneous estimation of multidimensional lognormal channel fading and a Gauss-Markov message which angle-modulates a subcarrier. Implementations are discussed, and performance is analyzed in detail for a particular example.

In Chapter V, conclusions are drawn and recommendations are made for further research.

D. Summary of Thesis Results.

For convenience, we note here the major contributions of this study.

In what follows, $\{N_t, t \in T\}$ is a doubly-stochastic Poisson process, scalar, for simplicity, with rate process $\{\lambda_t, t \in T\}$. Throughout this thesis, T will always be a finite interval.

Let $\{\lambda_t, t \in T\}$ be a function of a random variable z, where $E\{|z|^2\} < \infty$. Then the MMSE estimate $\hat{z}_t =$ $E\{z | N_\sigma, \sigma \in [t_0, t]\}$ can be represented by a stochastic integral as

$$\hat{z}_t = E(z) + \int_{t_0}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma) | N_\tau, \tau \in [t_0, \sigma]\} \hat{\lambda}_\sigma^{-1} d\nu_\sigma \tag{1.1}$$

where

$$\hat{\lambda}_\sigma = E\{\lambda_\sigma | N_\tau, \tau \in [t_0, \sigma]\}$$

$$\nu_t = N_t - \int_{t_0}^{t} \hat{\lambda}_\sigma d\sigma$$

(See Theorem 3.1). This is a fundamental result, from which many subsidiary results can be derived (See Chapter III, Section D). For example, if $\{x_t, t \in T\}$ is an Ito process* satisfying

*See II.C.1. for a definition.

$$dx_t = f_t dt + f_t d\zeta_t \tag{1.2}$$

with $\{\zeta_t,\ t \in T\}$ a zero-mean martingale, then it can be shown that the estimate $\{\hat{x}_t = E(x_t | N_\sigma,\ \sigma \in [t_0, t]), t \in T\}$ satisfies the stochastic differential equation,

$$d\hat{x}_t = E\{f_t | N_\sigma,\ \sigma \in (t_0, t]\} dt + E\{x_t(\lambda_t - \hat{\lambda}_t) | N_\sigma, \sigma \in (t_0, t]\} \hat{\lambda}_t^{-1} d\nu_t \tag{1.3}$$

In Section E. of Chapter III, this equation is used as the starting point of a series of approximations, leading ultimately to a pair of equations, the processor and variance equations, which can be implemented with relative ease. Theorem 3.8 provides a partial differential-difference equation, related to the Fokker-Planck equation, for the exact first-order probability density function of the estimator error. This theorem is valid for a wide class of estimators, including the approximate ones just mentioned, as well as some exact ones. For the cases in which the theorem is applicable, its validity and accuracy do not depend on the nature of any approximations used in the derivation of the particular estimators. In addition to the mean-square error of estimators, Theorem 3.8 is potentially useful for investigating anomalous behavior such as threshold and cycle skipping in angle-modulated systems. This potential is illustrated for some specific examples in Chapter IV.

Other results in Chapter III include equations for prediction, non-causal filtering (interpolation), optimum detection for diversity channels, optimal linear estimators (Theorem 3.6), mean-square error bounds (e.g., Theorem 3.7), and mutual information. The following result is also obtained in Chapter III.

Let $\{N_t, t \in T\}$ be an inhomogeneous Poisson counting process, with event times denoted by $t_1, t_2, \ldots, t_{N_t}$. It is well known[25] that $\{N_t, t \in T\}$ can be statistically represented by

$$p(\{t_i\}; N_t) = \exp\left[-\int_{t_0}^{t} \lambda_\sigma d\sigma\right] \prod_{i=1}^{N_t} \lambda_{t_i} \tag{1.4}$$

where λ_t is the intensity function of $\{N_t, t \in T\}$. $p(\{t_i\}; N_t)$ is the joint probability density function of the ordered set $\{t_1, \ldots, t_{N_t}\}$, and N_t. If $\{N_t, t \in T\}$ is instead a doubly-stochastic Poisson process with intensity process $\{\lambda_t, t \in T\}$, then (1.4) still holds, with λ_t replaced by its MMSE causal estimate $\hat{\lambda}_{t-}$,

$$\hat{\lambda}_{t-} = E\{\lambda_t | N_\sigma, \sigma \in [t_0, t)\} \tag{1.5}$$

(See Chapter III, Section E, or Reference 80).

In Chapter IV a signal model for angle modulation on a subcarrier is proposed. This model is used to design a demodulator for a turbulent diversity channel with correlated lognormal fading and additive noise. Using Theorem 3.8, an exact performance analysis, valid both above and below threshold, is carried out for this system in the absence of fading, assuming phase modulation (PM). It is shown that the system performs as a classical phase-lock loop over a wide range of parameter values. Bounds on the parameters, to guarantee classical behavior, are derived, and comparisons with classical PM systems are drawn.

It can be shown that the results we have quoted here, as well as others in the thesis, remain valid if $\{N_t, t \in T\}$ is a vector regular point process with conditionally independent components.

CHAPTER II

Preliminaries

A. Introduction.

In this chapter, we provide a brief review of some of the central concepts of point processes and the stochastic calculus, as they apply to this thesis. Broader discussions, emphasizing supplemental properties and peripheral ideas, can be found in the appendices. The reader desiring a deeper background in these subjects can consult References 1-5, 30, 43, 64, 66, 77, 78 (probability, random processes); References 5, 25, 41, 64-68, 78, 82-85 (point processes); and References 1-3, 6-13, 16-22, 77, 86, 90 (stochastic calculus).

The table at the beginning of this thesis lists most of the symbols used. Vectors are usually not explicitly distinguished from scalars; the transpose of a (column) vector x is denoted by x', and the inverse of a square matrix A by A^{-1}. When not explicitly stated, matrices appearing in products are assumed to be commensurate. Inner products specifically associated with a Hilbert space are denoted by the form, $(\cdot, \cdot)$, as opposed to vector inner products such as $x'y$.

Throughout this thesis T will be used to denote a Lebesgue-measurable parameter set; in most cases, this will mean that T is a finite interval $[t_0, t_1]$.

Theorems will be terminated with the symbol "________" in the right margin; proofs will be terminated by "■".

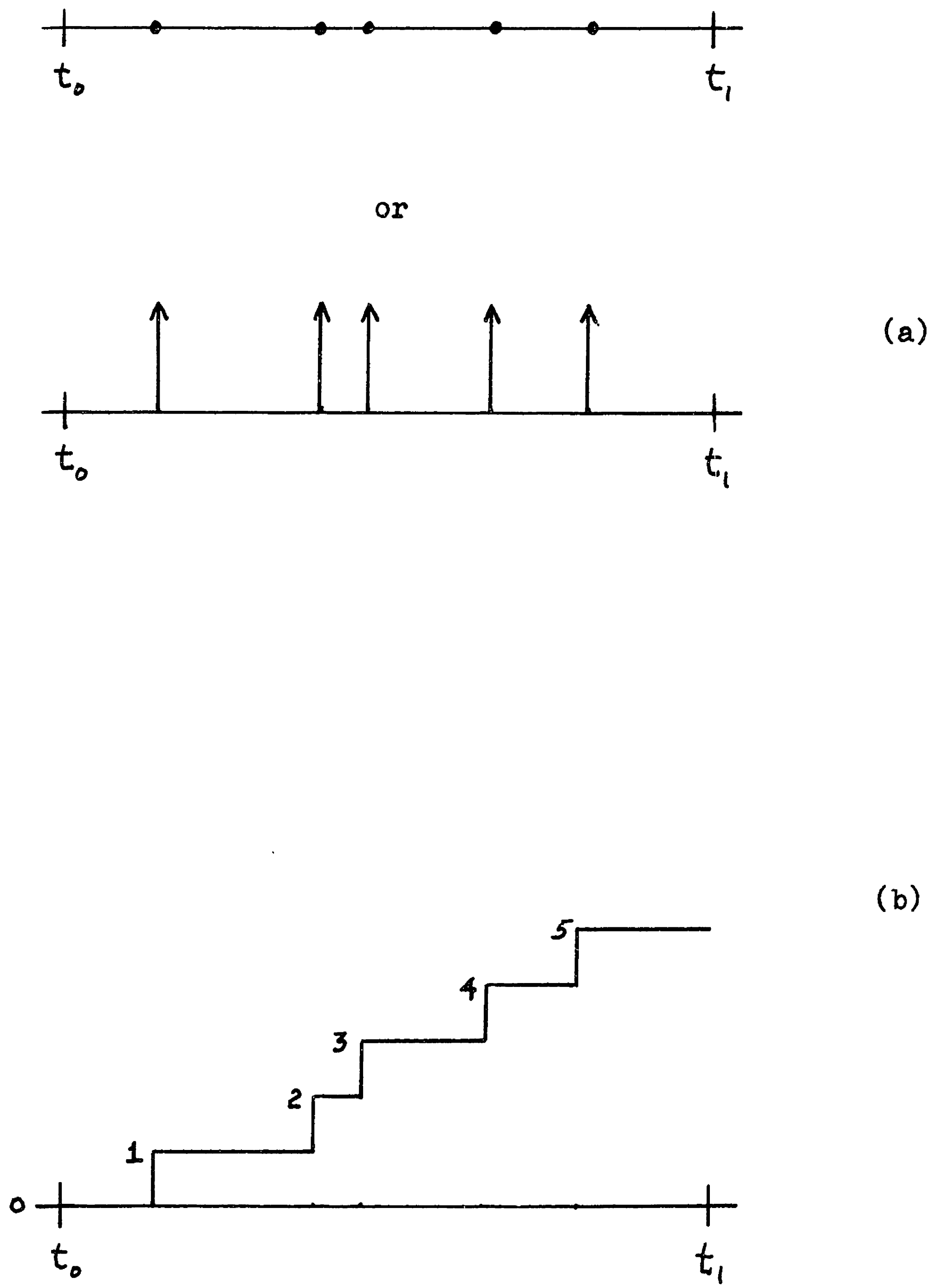
B. Point Processes and Counting Processes.

A point process is a random process which is completely characterized by the instants of occurrence of like events which occur at random points in time. A point process disregards the exact nature of the individual events; it is concerned only with the points in time at which the events occur.

A counting process is an integer-valued random process* which counts the number of events occurring in an interval, these events, or points, having been distributed by the stochastic mechanism of a point process.

A point process and its corresponding counting process are statistically equivalent. That is, knowledge of the event times of a point process is equivalent to knowledge of the jump times of the corresponding counting process, and vice-versa. The sample functions of these processes are notably different, however, as Figure 2.1 illustrates. This is immediately clear from the fact that a realization of a point process is a countable collection of ordered random variables, whereas a realization of a counting process is a piecewise constant function defined on a continuum. A realization of a point process can be viewed suggestively as

*To be distinguished from a jump process, which is not necessarily integer valued.



A point process (a), and its corresponding

counting process (b).

Figure 2.1

an impulse train, with impulses occurring at the event times. This establishes a point process as the formal derivative of a counting process.

In this thesis we shall be concerned almost exclusively with counting processes, which we denote in general by $\{N_t, t \in T\}$, where T is an interval. Counting processes can be defined on spaces of higher dimension than the line, such as the plane, but that generalization will not be used here.

To be useful mathematical entities, counting processes must be subject to certain analytical restrictions. The weakest condition usually imposed is that in sufficiently small intervals, at most one jump (event) can occur; i.e., $\{N_t, t \in T\}$ must possess an intensity. In mathematical terms, the following relations must hold w.p.1:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\{N_{t+\Delta t} - N_t = 1 \mid B_{N_t}\} = \lambda_t(B_{N_t}) \tag{2.1}$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[1 - \Pr\{N_{t+\Delta t} - N_t = 0 \mid B_{N_t}\}\right] = \lambda_t(B_{N_t}) \tag{2.2}$$

$$\lim_{\Delta t \to 0} \Pr\{N_{t+\Delta t} - N_t > 1 \mid B_{N_t}\} = 0 \tag{2.3}$$

$\lambda_t(B_{N_t})$ denotes the intensity, a non-negative functional of $\{N_\sigma, \sigma \leq t; \sigma, t \in T\}$.

Another condition usually imposed on $\{N_t, t \in T\}$ is that with probability one (w.p.1) only a finite number of jumps may occur in any finite interval; i.e., $Pr\{N_t = \infty, t \in T\} = 0$.

A point process whose counting process satisfies these conditions is sometimes called a regular point process[65,68] (RPP), and is a practical model for many physical phenomena. A vector RPP is defined to be a collection of scalar RPPs, not necessarily independent of one another. It is frequently convenient to restrict the class of vector RPP's slightly by defining vector RPP's with conditionally independent components. For this smaller class, we take the components of a "future" increment $N_{t+\Delta t} - N_t$ to be independent, given the "past" of the process, B_{N_t}. With this definition, it can be shown that w.p.1, jumps in the components of $\{N_t, t \in T\}$ do not coincide.

Poisson processes obviously belong to the class of regular point processes, and it has been shown[68,80] that doubly-stochastic Poisson processes are also members of this class. In the next two subsections, we briefly summarize some important facts regarding Poisson processes (PP) and doubly-stochastic Poisson processes (DSPP), as they apply to this thesis. Additional properties are stated in Appendix B.

1. Poisson Process.

A Poisson process $\{N_t, t \in T\}$ is an independent-increment counting process, with initial value zero, such that in any interval $T' \subset T$, the number of jumps or events is Poisson distributed:

$\Pr\{\#$ of jumps in $T' = n\} =$

$$\frac{1}{n!}\left[\int_{T'} \lambda_t dt\right]^n \exp\left[-\int_{T'} \lambda_t dt\right] \tag{2.4}$$

λ_t is some integrable, non-negative, non-random function called the rate or intensity. If λ_t is constant, $t \in T$, then $\{N_t, t \in T\}$ is called a stationary or homogeneous Poisson process. The incremental properties (2.1)-(2.3) hold, with $\lambda_t(B_{N_t}) = \lambda_t$ a deterministic function. We define a vector Poisson process as a collection of independent scalar PP's.

Let $\{N_t, t \in T\}$ be a scalar PP, with $T = [t_0, t_1]$. Let $\{t_i\} = \{t_1, t_2 \ldots, t_{N_t}\}$ be the set of occurrence times of the events that take place in $[t_0, t]$, where $t_0 < t_1 < t_2 < \ldots < t_{N_t} \leq t$. It can be shown[25] that the joint probability density function (p.d.f.) of N_t and the set $\{t_i\}$ is given by

$$p(\{t_i\}; N_t) = \exp\left[-\int_{t_0}^{t} \lambda_\sigma d\sigma\right] \prod_{i=1}^{N_t} \lambda_{t_i}, \quad t \in T, \tag{2.5}$$

where the product is defined to be unity if $N_t = 0$. This p.d.f. provides a complete statistical description of $\{N_t, t \in T\}$, and is beginning to see wide use in applications[25,26].

2. Doubly-Stochastic Poisson Process.

We now define a simple but important generalization of the Poisson process.

Let $\{\lambda_t, t \in T\}$ be a random process which is non-negative w.p.1. Then $\{N_t, t \in T\}$ is a DSPP if, given $\{\lambda_t, t \in T\}$, $\{N_t, t \in T\}$ is conditionally a Poisson process. Loosely speaking, $\{N_t, t \in T\}$ is a Poisson process with a stochastic intensity function; however, as is illustrated in Appendix B, the statistical properties of DSPP's are quite different in general from those of PP's, and care must be exercised not to confuse the two.

The DSPP $\{N_t, t \in T\}$ satisfies the following incremental relations w.p.1:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\{N_{t+\Delta t} - N_t = 1 \mid \lambda_\sigma, \sigma \in [t, t+\Delta t]\} = \lambda_t \qquad (2.6)$$

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[1 - \Pr\{N_{t+\Delta t} - N_t = 0 \mid \lambda_\sigma, \sigma \in [t, t+\Delta t]\}\right] = \lambda_t \qquad (2.7)$$

$$\lim_{\Delta t \to 0} \Pr\{N_{t+\Delta t} - N_t > 1 \mid \lambda_\sigma, \sigma \in [t, t+\Delta t]\} = 0 \qquad (2.8)$$

These equations describe incremental behavior relative to the intensity process $\{\lambda_t, t \in T\}$. It can be shown[68] that (2.1)-(2.3), which describe incremental behavior relative to

$\{N_t, t \in T\}$, also hold for DSPP's, with $\lambda_t(B_{N_t})$ given by

$$\lambda_t(B_{N_t}) = E\{\lambda_t | B_{N_t}\} . \tag{2.9}$$

A relation similar to (2.5) is available to DSPP's (See Reference 80 and Chapter III, Section E).

We define a vector DSPP as a collection of scalar DSPP's which, <u>given the intensities</u>, are conditionally independent PP's. Note that this does not rule out the possibility of statistical dependencies among the intensity processes; it merely removes the possibility of coincident jumps in different components of $\{N_t, t \in T\}$.

C. Stochastic Calculus.

In the theory of point processes, one often encounters quantities of the form

$$y(t) = \sum_k h(t, t_k) \tag{2.10}$$

with $\{t_i\}$ a realization of a point process, and $h(\cdot, \cdot)$ some function, possibly random. Indeed, we have already seen this form, somewhat disguised, in (2.5):

$$\prod_{i=1}^{N_t} \lambda_{t_i} = \exp \sum_{i=1}^{N_t} \ln \lambda_{t_i}$$

Eq. (2.10) can be viewed as a linear filter, by recalling that a point process can be represented formally as an impulse train; i.e.,

$$y(t) = \int h(t,\tau) \sum_k \delta(\tau - t_k) d\tau \tag{2.11}$$

The impulse train is itself the formal derivative of a counting process $\{N_t,\ t \in T\}$, so that (2.11) can be written formally as

$$y(t) = \int h(t,\tau) dN_\tau \tag{2.12}$$

Note that this is an integral with respect to the counting process, $\{N_t,\ t \in T\}$.

More generally, a point process might be the excitation for a non-linear, time-varying causal system. The state of such a system evolves according to the differential equation,

$$\frac{dx_t}{dt} = f_t + g_t z_t, \tag{2.13}$$

where x_t is the state vector, z_t is the excitation, and f_t and g_t are causal functionals of x_t and z_t, possibly possessing memory. If z_t were a point process, represented as an impulse train, we could take the same approach that led to (2.12), and write (2.13) as

$$dx_t = f_t dt + g_t dN_t \tag{2.14}$$

In integral form, this is

$$x_t - x_{t_0} = \int_{t_0}^{t} f_\sigma d\sigma + \int_{t_0}^{t} g_\sigma dN_\sigma \tag{2.15}$$

Again, as in (2.12), we are faced with an integral for which the integrator is a counting process. How does one manipulate such integrals, and what is the meaning of equations such as (2.15)? So far, we have merely utilized some rather suggestive symbolism. To answer the questions we have posed, we must make our definitions precise; this is the topic of the subsection which follows.

1. Stochastic Equations and Integrals.

Consider the integral,

$$\int_{t_0}^{t} g_\sigma dN_\sigma , \tag{2.16}$$

where g_σ is a function of σ and a causal functional of $\{N_t,\ t \in T = [t_0,\ t_1]\}$. We wish this integral to possess certain properties. For example, if g_σ is non-random, the evaluation

$$\int_{t_0}^{t} g_\sigma \, dN_\sigma = \sum_{i=1}^{N_t} g_{t_i} \tag{2.17}$$

where the $\{t_i\}$ mark the jumps of $\{N_t,\ t \in T\}$, is desirable. As another example, if $\{N_t,\ t \in T\}$ is a Poisson process (an independent increment process), and $f_t(x_t)$ and $g_t(x_t)$ are memoryless functions of x_t, we want the process $\{x_t,\ t \in T\}$ specified by the stochastic differential equation,

$$dx_t = f_t(x_t)dt + g_t(x_t)dN_t \tag{2.18}$$

to be a Markov process.

Eq. (2.16) cannot be defined as the usual integral of deterministic calculus. This is obvious from the following example. Consider the integral $\int_{t_0}^{t} N_\sigma dN_\sigma$. The Riemann-Stieltjes definition of this integral

$$\int_{t_0}^{t} N_\sigma dN_\sigma = \lim_{n} \sum_{i=0}^{n-1} N_{\tau_i}[N_{\sigma_{i+1}} - N_{\sigma_i}] \tag{2.19}$$

where $\sigma_n = t$, τ_i is <u>any</u> point in the interval $[\sigma_i,\ \sigma_{i+1}]$, and the limit is taken in some appropriate probabilistic sense. This definition, when it applies, gives the same result regardless of where τ_i falls in the interval

$[\sigma_i, \sigma_{i+1}]$. This is not necessarily true for the integral $\int_{t_0}^{t} N_\sigma dN_\sigma$, however. Assume that a jump in the process $\{N_t, t \in T\}$ occurs in $[\sigma_i, \sigma_{i+1}]$. Then the value of N_{τ_i} depends on whether τ_i falls before or after the jump; hence (2.19) is an ambiguous definition. To remove this ambiguity, our only choice is to let τ_i be one of the endpoints, σ_i or σ_{i+1}. If we choose $\tau_i = \sigma_{i+1}$, it can be shown[1-3,16] that the solution of (2.18) is not a Markov process. Therefore, we must choose $\tau_i = \sigma_i$.

This choice yields a consistent definition possessing many desirable properties, including the ones we have mentioned here. However, the resulting integral does not obey the rules of ordinary calculus, and must be manipulated under a new set of rules. These will be developed in the next subsection. First, let us give a precise definition of the stochastic integral, (2.16).

Let $\{N_t, t \in T = [t_0, t_1]\}$ be an RPP, and let $\{g_t, t \in T\}$ be a random process with sample paths which are Lebesgue-measurable w.p.1. Assume that for every $t \in T$, g_t is a causal functional of $\{N_t, t \in T\}$. The stochastic integral $\int_{t_0}^{t} g_\sigma dN_\sigma$ is defined as

$$I_t = \int_{t_0}^{t} g_\sigma dN_\sigma = \underset{n}{\text{l.i.m.}} \sum_{i=0}^{n-1} g_{\sigma_i}\left[N_{\sigma_{i+1}} - N_{\sigma_i}\right], \tag{2.20}$$

where $\sigma_n = t$ and $\lim_{n\to\infty} \max_i(\sigma_{i+1} - \sigma_i) = 0$. Although I_t is defined as a limit in quadratic mean, it can be shown that it is unique w.p.1 (Doob[1], p. 442, Skorokhod, pp. 18, 29-31).

Because of the jump nature of $\{N_t, t \in T\}$, the l.i.m. is easily evaluated to give

$$I_t = \sum_{i=1}^{N_t} g_{t_i-} \tag{2.21}$$

where $t_i- \triangleq \lim_{\Delta \downarrow 0} t_i - \Delta$, and the $\{t_i\}$ are the event times of $\{N_t, t \in T\}^*$. The need for t_i- in (2.21) is evident from (2.20): if a jump occurs in $[\sigma_i, \sigma_{i+1}]$, it must not be reflected in g_{σ_i}.

If g_t is non-random, then $g_{t_i-} = g_{t_i}$, because w.p.1 the discontinuities of g_t, if any, cannot coincide with the jumps of $\{N_t, t \in T\}$. For this case, (2.17) holds, as desired. Moreover, if $\{N_t, t \in T\}$ is a Poisson process, the solution of

$$x_t - x_t = \int_{t_0}^{t} f_\sigma(x_\sigma)d\sigma + \int_{t_0}^{t} g_\sigma(x_\sigma)dN_\sigma \tag{2.22}$$

-- (2.18) in integral form -- is a Markov process (See Doob[1], p. 282).

That the definition (2.20) does not obey the rules of ordinary calculus is illustrated by the integral considered

*This evaluation is rigorously verified in Skorokhod[2], Ch. 3.

earlier, $\int_{t_0}^{t} N_\sigma dN_\sigma$. By ordinary rules, this evaluates as $(\frac{1}{2})N_t^{\,2}$; by (2.21) it works out to be

$$\sum_{i=1}^{N_t}(N_{t_i} - 1) = 0 + 1 + 2 + \ldots + (N_t - 1) = \tfrac{1}{2}N_t^{\,2} - \tfrac{1}{2}N_t$$

The definition (2.20) can be applied to a much wider class of integrals, such as $\int_{t_0}^{t} g_\sigma d\xi_\sigma$ where $\{\xi_t,\ t \in T\}$ is any L_2 process. If $\{\xi_t,\ t \in T\}$ is a L_2 martingale, the resulting integral possesses many remarkable and useful properties.

Let us briefly discuss stochastic integrals with martingale integrators, following Doob[1], Ch. IX, and Meyer[7,8]. Let $\{\xi_t,\ t \in T\}$ be an L_2 martingale. There exists[7] a process $\{y_t,\ t \in T\}$, w.p.1 continuous from the right and increasing, such that for $s < t$,

$$E\{(\xi_t - \xi_s)^2 \mid B_{\xi_s}\} = E\{y_t - y_s \mid B_{\xi_s}\} \tag{2.23}$$

The process $\{y_t,\ t \in T\}$, discussed further in Appendix C, will be used in the definition which follows. Let $\{g_t,\ t \in T\}$ be a process with sample paths which are Lebesgue-measurable w.p.1, such that for every $t \in T$, g_t is B_{ξ_t}-measurable. Assume also that $E\{\int_{t_0}^{t_1} g_\sigma^2\, dy_\sigma\} < \infty$. The stochastic integral $\int_{t_0}^{t} g_\sigma d\xi_\sigma$ is defined as

$$I_t = \int_{t_0}^{t} g_\sigma d\xi_\sigma = \underset{n}{l.i.m.} \sum_{i=0}^{n-1} g_{\sigma_i}[\xi_{\sigma_{i+1}} - \xi_{\sigma_i}], \tag{2.24}$$

where $\sigma_n = t$ and $\lim_{n\to\infty} \max_i (\sigma_{i+1} - \sigma_i) = 0$. As with (2.20), this definition is unique w.p.1. Eq. (2.24) is identical in form to (2.20), with $\{N_t,\ t \in T\}$ replaced by $\{\xi_t,\ t \in T\}$.

We now list some properties of the stochastic integral (2.24). We emphasize that they are not true in general for any other definition. Proofs can be found in the cited references. For s, $t \in T$:

1) I_t is, w.p.1, uniquely defined;

2) $E(I_t) = 0$; (2.25)

3) If $J_s = \int_{t_0}^{s} f_\sigma \, d\xi_\sigma$, and $\{f_t,\ t \in T\}$ satisfies the same conditions as $\{g_t,\ t \in T\}$, then

$$E(I_t J_s) = E\left\{\int_{t_0}^{t\wedge s} g_\sigma f_\sigma \, dy_\sigma\right\}, \text{ where } t\wedge s = \min(t,s). \quad (2.26)$$

4) I_t is an L_2 martingale.

These properties are extremely useful in statistical calculations. Properties 2) and 3), for example, greatly facilitate many moment calculations. Property 4) is a reproducing property, in the sence that both $\{\xi_t,\ t \in T\}$ and $\{I_t,\ t \in T\}$ are martingales; this makes immediate the definition of iterated or multiple integrals, which themselves possess the properties 1) - 4).

The definition (2.24) can be extended without difficulty to vector valued processes as long as $\{\xi_t,\ t \in T\}$ has independent components. In the sequel we shall use vector integrals without further comment. (See Appendix C for further discussion.)

The relevance of this broad definition (2.24) to our investigation derives from the fact that any integral of the form (2.20), with a counting process integrator, can be written as the sum of an ordinary integral of a stochastic process and a stochastic integral of the form (2.24). This decomposition follows from a theorem of Doob and Meyer[7,8], and is discussed further in Appendix C.

The result is obvious if $\{N_t,\ t \in T\}$ is a PP, because any independent increment process can be transformed to a martingale by subtracting the mean. Thus,

$$N_t - \int_{t_0}^{t} \lambda_\sigma \, d\sigma \qquad (2.27)$$

is a martingale, and can be used in the definition, (2.24).

Now that we have precisely defined stochastic integrals, we can proceed to give meaning to equations such as (2.14) and (2.15)*, which are generally called stochastic equations[1-3,16,77]. We are concerned in this thesis with vector equations of the form

$$x_t - x_t = \int_{t_0}^{t} f_\sigma \, d\sigma + \int_{t_0}^{t} g_\sigma \, d\xi_\sigma \qquad (2.28)$$

where $\{\xi_t,\ t \in T\}$ is a vector L_2 process, and $\{f_t,\ t \in T\}$ and $\{g_t,\ t \in T\}$ are vector and matrix processes, respectively, which are functions of $\{x_t,\ t \in T\}$, $\{\xi_t,\ t \in T\}$, and, possibly,

*The first integral on the r.h.s. of (2.15) -- an ordinary integral of a random process, probably familiar to most readers -- is discussed briefly in Appendix A, Section 8.

of other processes, as well. Eq. (2.28) is of interest because it provides a very general description of a large class of random processes. Note that it is a constructive relation; i.e., it gives a direct, sample function description of processes generated from the $\{\xi_t, t \in T\}$ process.

We frequently write equations such as (2.28) in the suggestive form of a stochastic differential equation,

$$dx_t = f_t dt + g_t d\xi_t. \tag{2.29}$$

We emphasize, though, that (2.29) is no more than shorthand for (2.28), and can be misleading if interpreted too literally. For example, operating on (2.28) and (2.29) with the conditional expectation $E(\cdot | B_{\xi_s})$, where $s \in [t_0, t]$, does not yield identical results, as can easily be verified.

Properties of (2.28) depend on the processes $\{\xi_t, t \in T\}$, $\{f_t, t \in T\}$, and $\{g_t, t \in T\}$. If $\{\xi_t, t \in T\}$ is an independent increment process, and f_t and g_t are memoryless functions of x_t obeying mild regularity conditions[1-3], then $\{x_t, t \in T\}$ is a Markov process, and is the unique solution of (2.28). This approach to the study of Markov processes has been nost fruitful, bringing forth many important results in recent years, both in theory[1-3,9-12-19,77], and in applications[13,14,16,18-23-33,42,77,86-91].

More generally, let $\{\zeta_t, t \in T\}$ be an L_2 martingale, and assume that f_t and g_t are finite-variance, causal functionals of $\{x_t, t \in T\}$ and $\{\zeta_t, t \in T\}$. We allow f_t and g_t to be causal functionals of other related processes, such as control functions, also. The process $\{x_t, t \in T\}$ defined by (2.28) is then called an Ito process, and (2.28), or (2.29), is termed an Ito equation[1-3]. This is a very broad definition, encompassing a wide variety of applications. In the next subsection, we discuss Ito processes in the context of this thesis.

2. Change of Variable.

In the previous paragraph we defined the Ito process. If $\{\zeta_t, t \in T\}$ in (2.28) is replaced by a counting process $\{N_t, t \in T\}$, the result is still an Ito process because, as was pointed out in the discussion following (2.26), $\{N_t, t \in T\}$ can always be transformed into a martingale (see Appendix C)*.

---

*After transforming $\{N_t, t \in T\}$ into a martingale, (2.30) takes the form,

$$dx_t = \tilde{f}_t dt + g_t d\zeta_t, \quad (\tilde{f}_t \neq f_t)$$

where $\{\zeta_t, t \in T\}$ is a martingale. It can be verified from the results of Appendix C and Lemma D.1, Appendix D, that $\{x_t, t \in T\}$ is an Ito process as long as $f_t dt + g_t d\zeta_t$ describes an Ito process.

In dealing with counting processes, we will frequently encounter Ito equations of this form,

$$dx_t = f_t dt + g_t dN_t. \tag{2.30}$$

The following result, known as the Ito differential rule for counting processes, provides us with the fundamental tool for manipulating equations such as (2.30). It allows us to write down immediately the Ito equation satisfied by any well-behaved function of x_t.

<u>Theorem 2.1.</u>

Let $\{x_t, t \in T\}$ be the vector Ito process defined by (2.30), and let $\{N_t, t \in T\}$ be a counting process of a vector RPP with conditionally independent components. Let ϕ_t be a scalar-valued, differentiable function of t and x_t for every t and x_t. Then the process $\{\phi_t, t \in T\}$ satisfies the Ito equation,

$$d\phi_t = \frac{\partial \phi_t}{\partial t} dt + (\nabla_{x_t} \phi_t)' f_t \, dt$$

$$+ \sum_{j=1}^{D} [\phi_t(x_t + g_t e_j) - \phi_t(x_t)] \, dN_t' e_j \tag{2.31}$$

where $\{N_t, t \in T\}$ is assumed to be D-dimensional, and $\{e_j\}$ is the set of unit vectors in D-dimensional Euclidean space (R^D).

Proof.

Eq. (2.31) can be obtained directly, with some effort, from Kunita and Watanabe[6], Th. 5.1. We present a short heuristic proof, modeled after Snyder's[23] proof of the scalar case.

Let $\Delta N_t = N_{t+\Delta t} - N_t$, and let $\{N_t^{(i)}, t \in T\}$ and $\{N_t^{(j)}, t \in T\}$ be any two components of the vector RPP $\{N_t, t \in T\}$, with corresponding intensities $\{\lambda_t^{(i)}(B_{N_t}), t \in T\}$, $\{\lambda_t^{(j)}(B_{N_t}), t \in T\}$.

Then,

$$\Pr\{\Delta N_t^{(i)} = 1, \Delta N_t^{(j)} = 1\}$$

$$= E\left[\Pr\{\Delta N_t^{(i)} = 1, \Delta N_t^{(j)} = 1 \mid B_{N_t}\}\right]$$

$$= E\left[\Pr\{\Delta N_t^{(i)} = 1 \mid B_{N_t}\} \Pr\{\Delta N_t^{(j)} = 1 \mid B_{N_t}\}\right]$$

where the last equality follows from the conditional independence assumption. Now,

$$E\left[\Pr\{\Delta N_t^{(i)} = 1 \mid B_{N_t}\} \Pr\{\Delta N_t^{(j)} = 1 \mid B_{N_t}\}\right]$$

$$= E\left[\left(\lambda_t^{(i)}(B_{N_t})\Delta t + o(\Delta t)\right)\left(\lambda_t^{(j)}(B_{N_t})\Delta t + o(\Delta t)\right)\right]$$

$$= o(\Delta t)$$

Thus, jumps in the components of $\{N_t, t \in T\}$ cannot coincide in time, and we see that, for small Δ_t, N_t assumes to $o(\Delta t)$ one of the values, 0, e_1, e_2, ..., or e_D. That is, for sufficiently small Δt, a change can occur in at most one component of N_t. Therefore, to $o(\Delta t)$,

$$\Delta x_t = x_{t+\Delta t} - x_t = \begin{cases} f_t \Delta t & , \Delta N_t = 0 \\ f_t \Delta t + g_t e_1 & , \Delta N_t = e_1 \\ \vdots & \quad \vdots \\ f_t \Delta t + g_t e_D & , \Delta N_t = e_D \end{cases}$$

It follows that

$$\begin{aligned} \Delta \phi_t &= \phi_{t+\Delta t}(x_{t+\Delta t}) - \phi_t(x_t) \\ &= \phi_{t+\Delta t}(x_t) - \phi_t(x_t) \\ &\quad + \left[\phi_t(x_t + f_t \Delta t) - \phi_t(x_t)\right] \delta_{0, \Delta N_t} \\ &\quad + \sum_{j=1}^{D} \left[\phi_t(x_t + f_t \Delta t + g_t e_j) - \phi_t(x_t)\right] \delta_{e_j, \Delta N_t} \\ &\quad + o(\Delta t) \end{aligned}$$

where

$$\delta_{x, y} = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$$

But

$$\phi_{t+\Delta t}(x_t) - \phi_t(x_t) \rightarrow \frac{\partial \phi_t}{\partial t} \Delta t + o(\Delta t) ;$$

$$\phi_t(x_t + f_t\Delta t) - \phi_t(x_t) \to (\nabla_{x_t}\phi_t)' f_t \Delta t + o(\Delta t);$$

and

$$\phi_t(x_t + f_t\Delta t + g_t e_j) - \phi_t(x_t) \to \phi_t(x_t + g_t e_j) - \phi_t(x_t) + O(\Delta t)^*$$

as $\Delta t \to 0$, so

$$\Delta\phi_t(x_t) = \frac{\partial \phi_t}{\partial t}\Delta t + (\nabla_{x_t}\phi_t)' f_t \Delta t\, \delta_{0,\Delta N_t}$$

$$+ \sum_{j=1}^{D} [\phi_t(x_t + g_t e_j) - \phi_t(x_t)]\, \delta_{e_j,\Delta N_t} + o(\Delta t)$$

Now, to $o(\Delta t)$,

$$\delta_{0,\Delta N_t} = 1 - \Delta N_t' 1_D \qquad \left(1_D = \sum_{j=1}^{D} e_j\right)$$

$$\delta_{e_j,\Delta N_t} = \Delta N_t' e_j$$

Substituting these in (2.32) and letting $\Delta t \to 0$, we get the desired result, (2.31). ∎

$^*O(\Delta t)$ denotes terms satisfying

$$\lim_{\Delta t \to 0} \frac{O(\Delta t)}{\Delta t} < \infty$$

while $o(\Delta t)$ denotes terms satisfying

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0 .$$

D. Summary.

In this chapter, we have provided a brief review of some of the central concepts of point processes and the stochastic calculus, as they apply to this thesis. Broader discussions can be found in the appendices, and in the references cited at the beginning of this chapter.

CHAPTER III

Estimation Theory for Doubly-Stochastic Poisson Processes

A. Introduction.

This chapter comprises the main body of theoretical results in this thesis.

After making a precise statement of the estimation problem and briefly discussing the limitations of elementary approaches, we prove a representation theorem, which allows any minimum mean-square error (MMSE) estimator to be expressed as an explicit functional of the observations via a stochastic integral. We apply this theorem to problems of prediction, interpolation, hypothesis testing, and causal filtering, with the last receiving the greatest emphasis.

For DSPP's we derive a new statistical characterization which relates directly to MMSE estimation, and provides an alternative tool for treating estimation and, especially, detection problems.

Based on causal estimators obtained from the representation theorem, a number of different approximation schemes are developed. Following a comparative evaluation of these schemes, we analyze the performance of estimators, both exact and approximate. In particular, we derive a partial differential-difference equation for the first-order p.d.f. of the estimator error. This equation is valid for a wide class of approximate estimators and some exact ones, and yields exact performance when it applies.

It is shown that virtually all the results of this chapter apply not only to DSPP's, but also to the broader class of RPP's. This generalization is discussed in detail following presentation of the representation theorem.

This chapter contains much mathematics and technical detail; combined with the fact that the results are stated in a general context, divorced from applications, this might tend to obscure the practical utility of the theory. Thus, the reader desiring a foretaste of some applications (in one particular field -- optical communications), might wish to skim Chapter 4 before reading this chapter.

B. Problem Formulation.

Let $\{N_t,\ t \in T = [t_o,\ t_1]\}$ be a vector DSPP, with rate process $\{\lambda_t,\ t \in T\}$ which is non-negative w.p.1 and satisfies $E\{\lambda_t\} < \infty$, $t \in T$. Assume that $\{\lambda_t,\ t \in T\}$ is a causal functional of some vector random process $\{z_t,\ t \in T\}$. Based on a realization of $\{N_t,\ t \in T_1 \subset T\}$ in some sub-interval T_1, we wish to find the minimum mean-square error (MMSE) estimate of z_t.

Depending on T_1, the estimate can be causal, non-causal, or predictive. The correspondence is given by:

$T_1 = [t_o,\ t] \Rightarrow$ causal;

$T_1 = [t_o,\ \tau],\ \tau > t \Rightarrow$ non-causal;

$T_1 = [t_o,\ s],\ s < t \Rightarrow$ predictive.

Each of these types of estimation will be considered in this chapter, with heavy emphasis on the first. If $\{z_t,\ t \in T\}$ is a constant random variable, we refer to the problem as parameter estimation.

Our objective is to obtain a MMSE point estimate of the random variable z_t, valid for continuously varying t. Interval estimates, as well as estimates based on other performance criteria, are discussed in Appendix E and References 26, 27, 33, 35, 75, and 93.

We have formulated the estimation problem in the context of DSPP's because of the relevance of these processes to optical communication. We shall show, however, that most of our results apply also to the wider class of regular point processes.

C. A Representation Theorem for Estimators.

1. Doubly-Stochastic Poisson Processes.

The MMSE estimate of random variable z relative to $\{N_\sigma, \sigma \in [t_0, t]\}$ is the conditional expectation, $E\{z|B_{N_t}\}$. This is proved in many texts on probability; Gikhman and Skorokhod[3], for example, give a simple proof in a general context. Straightforward calculation of the conditional mean is usually a prohibitively difficult task, except in a few simple cases (see Appendix E). However, using indirect methods a very useful representation of the conditional mean can be obtained in terms of a stochastic integral. This is the substance of the following theorem, which is the fundamental theoretical result of this thesis.

The proof of the theorem is quite technical, so we merely map out the important steps, relegating the details to Appendix D. The theorem, and the major arguments of its proof, were originally motivated by a similar result due to Frost[16], who examined the estimation problem for random signals in additive Gaussian noise.

Although the theorem is stated in the context of parameter estimation, there is nothing to prevent the parameter from being a random variable of some process. When this is the case, we shall subsequently show by use of a limiting procedure that the theorem leads directly to stochastic differential equations for process estimation.

Theorem 3.1.

Let $\{N_t,\ t \in T = [t_o, t_1]\}$ be a vector doubly-stochastic Poisson process, as defined in II.B.1, and let z be a vector random variable with components in L_2. Assume that the intensity $\{\lambda_t,\ t \in T\}$ satisfies the following conditions:

1) $E\{\lambda_t\} < \infty,\ t \in T;$

2) $\Pr\{\lambda_t < 0,\ t \in T\} = 0;$

3)* $B(z) \subset B_{\lambda_t},\ t \in T.$

Define the process $\{\nu_t,\ t \in T\}$ by the equation $d\nu_t = dN_t - \hat{\lambda}_t dt$, where $\hat{\lambda}_t = E\{\lambda_t | B_{N_t}\}$, and let $\hat{\Lambda}_t = \text{diag}(\hat{\lambda}_t)$. Then,

(i) $B_{N_t} = B_{\nu_t},\ t \in T;$

(ii) the conditional mean $\{\hat{z}_t = E(z|B_{N_t}),\ t \in T\}$ can be represented by the stochastic integral equation,

$$\hat{z}_t = \hat{z}_{t_o} + \int_{t_o}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{\nu_\sigma}\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma \qquad (3.1)$$

Proof.

Part (i) of the theorem is an immediate consequence of Lemma D.1.** By definition, we have

---

*This is simply a statement that $\{\lambda_t,\ t \in T\}$ depends on z, and possibly on other processes, as well.

**In this proof, lemma designations prefixed with the letter "D" refer to lemmas in Appendix D.

$$\nu_t = N_t - N_{t_0} - \int_{t_0}^{t} \hat{\lambda}_\sigma d\sigma$$

$$= N_t - \int_{t_0}^{t} \hat{\lambda}_\sigma d\sigma \tag{3.2}$$

where the latter equality follows from the fact that $N_{t_0} = 0$ w.p.1*. Now $\int_{t_0}^{t} \hat{\lambda}_\sigma d\sigma$ exists for all $t \in T$ because

$$\int_{t_0}^{t} E\{\hat{\lambda}_\sigma\} d\sigma = \int_{t_0}^{t} E\{\lambda_\sigma\} d\sigma$$

and $E\{\lambda_t\} < \infty$, $t \in T$, by hypothesis (see Doob[1], p. 62, or Appendix A, Section 8). Since $\hat{\lambda}_t = E(\lambda_t | B_{N_t})$ is B_{N_t}-measurable, the conditions of Lemma D.1 are satisfied, and we thus have $B_{N_t} = B_{\nu_t}$, $t \in T$.

The proof of (ii) is more involved. The argument centers on the result of Appendix A, Section 7, that conditional expectation can be defined by the orthogonality condition, (A.10). To prove that the r.h.s. of (3.1) is a version of the conditional expectation $E(z|B_{N_t})$, we show that the difference

$$z - \left[\hat{z}_{t_0} + \int_{t_0}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{\nu_\sigma}\} \hat{\lambda}_\sigma^{-1} d\nu_\sigma\right]$$

---

*This definition is motivated by the decomposition theorem of Appendix C. The martingale property of $\{\nu_t, t \in T\}$ will play an important role in the proof of (ii).

is orthogonal to every functional which is B_{N_t}-measurable. This is accomplished in a number of steps. First, we show that it is sufficient for the set of B_{N_t}-measurable "testing functions" to be a particular class of exponential functionals of $\{N_\sigma, \sigma \in [t_0, t]\}$ (Lemma D.2). Next we show, using Lemma D.1, that this class can in turn be replaced with a closely related class of functionals of $\{N_\sigma, \sigma \in [t_0, t]\}$ which are martingales. The martingale property simplifies subsequent calculations, in which we explicitly verify by means of the stochastic calculus that (3.1) satisfies the orthogonality condition

We shall prove (3.1) only for scalar observations $\{N_t, t \in T\}$. Other than unwieldy notation and additional algebraic complexity, the proof for vector observations contains nothing that is different.

<u>Step 1.</u>

Let

$$z_t^* = \hat{z}_{t_0} + \int_{t_0}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma) | B_{\nu_\sigma}\} \hat{\lambda}_\sigma^{-1} d\nu_\sigma \tag{3.3}$$

According to Lemma D.2, z_t^* is a version of $\hat{z}_t = E(z|B_{N_t})$ iff it satisfies the relation,

$$E\{(z - z_t^*) \exp i \int_{t_0}^{t} W_\sigma dN_\sigma\} = 0 , \tag{3.4}$$

for any non-random function W_σ which is square-integrable on T (i.e., $W_\sigma \in L_2[T]$).

Step 2.

Define

$$\eta_t = N_t - \int_{t_0}^{t} (iW_\sigma)^{-1}(e^{iW_\sigma}-1)\hat{\lambda}_\sigma \, d\sigma \tag{3.5}$$

We have already shown in the proof of (i) that $\hat{\lambda}_\sigma$ is integrable in σ; since $(iW_\sigma)^{-1}(e^{iW_\sigma}-1)$ is integrable, it follows that the product $(iW_\sigma)^{-1}(e^{iW_\sigma}-1)\ \hat{\lambda}_\sigma$ is integrable in σ. It is also B_{N_σ}-measurable, because W_σ is non-random and $\hat{\lambda}_\sigma$ is a conditional expectation with respect to B_{N_σ}. We conclude from Lemma D.1 that $B_{\eta_t} = B_{N_t}$, $t \in T$.

This allows us to use $\{\eta_t, t \in T\}$ in place of $\{N_t, t \in T\}$ in (3.4). The simplification which results from this rather arbitrary definition is due to the fact that

$\{\exp i\int_{t_0}^{t} W_\sigma d\eta_\sigma \, , \, B_{N_t} \, , \, t \in T\}$ is a martingale.

Consequently, it can be expressed very simply in terms of a particular stochastic integral, which fortuitously, has the same integrator process $\{\nu_t, t \in T\}$ as the integral in z_t^*. The properties of stochastic integrals of type (2.24) then make verification of the orthogonality condition a relatively easy task.

Step 3.

We must show that $E\{(z - z_t^*)\alpha_t\} = 0$,

where

$$\alpha_t = \exp i \int_{t_0}^{t} W_\sigma \, d\eta_\sigma \tag{3.6}$$

α_t is a function of the form exp r_t, where

$$dr_t = iW_t d\eta_t$$

$$= -(e^{iW_t}-1)\hat{\lambda}_t dt + iW_t dN_t$$

Let us apply the Ito rule for counting processes, Theorem 2.1, to the function exp(•). The first term on the r.h.s. of (2.31) is zero, so we are left with

$$d\alpha_t = -\alpha_t\hat{\lambda}_t(e^{iW_t}-1)dt + \alpha_t(e^{iW_t}-1)dN_t$$

$$= \alpha_t(e^{iW_t}-1)d\nu_t \ ;$$

the definition $d\nu_t = dN_t - \hat{\lambda}_t dt$ was used to obtain the second equality. In integral form, we have, using the obvious initial condition,

$$\alpha_t = 1 + \int_{t_0}^{t} \alpha_\sigma (e^{iW_\sigma}-1)d\nu_\sigma \qquad (3.7)$$

Now, α_t is B_{η_t}-measurable because it is defined in (3.6) to be a causal functional of $\{\eta_t,\ t\in T\}$. We have shown that $B_{\eta_t} = B_{N_t}$ and $B_{N_t} = B_{\nu_t}$, $t\in T$; therefore, α_t is B_{ν_t}-measurable.

This, combined with the fact that $\{\nu_t,\ B_{\nu_t},\ t\in T\}$ is a martingale (see Appendix C), implies that $\{\alpha_t,\ B_{\nu_t},\ t\in T\}$ is also a martingale.

Let $K_\sigma = E\{\mathfrak{z}(\lambda_\sigma-\hat{\lambda}_\sigma)|B_{\nu_\sigma}\}\hat{\lambda}_\sigma^{-1}$ and $dn_t = dN_t - \lambda_t dt$.

$$E\{(z - z_t^*)\alpha_t\} = E\{z\} + E\{z\int_{t_0}^t \kappa_\sigma (e^{iw_\sigma} - 1) d\nu_\sigma\} - E\{z_t^* \alpha_t\}$$

Substituting (3.3) and (3.7) and grouping terms, this becomes

$$E\{(z - \hat{z}_{t_0})\} + E\{z\int_{t_0}^t \alpha_\sigma (e^{iw_\sigma} - 1) dn_\sigma\}$$

$$+ E\{z\int_{t_0}^t \alpha_\sigma (\lambda_\sigma - \hat{\lambda}_\sigma)(e^{iw_\sigma} - 1) d\sigma\} - E\{\int_{t_0}^t K_\sigma d\nu_\sigma\}$$

$$- E\{\hat{z}_{t_0}\int_{t_0}^t \kappa_\sigma (e^{iw_\sigma} - 1) d\nu_\sigma\} - E\{\int_{t_0}^t \alpha_\sigma K_\sigma (e^{iw_\sigma} - 1)\hat{\lambda}_\sigma d\sigma\} \quad (3.8)$$

The last integral in (3.8) results from use of (2.23), (2.26) and Corollary C.1.1. of Appendix C.

The first term in (3.8) is zero because the <u>a priori</u> estimate $\hat{z}_{t_0}$ is simply the mean of z. The second term is zero since z is assumed to be B_{λ_t}-measurable and thus

$$E\{z\int_{t_0}^t \alpha_\sigma (e^{iw_\sigma} - 1) dn_\sigma\}$$

$$= E\{z\, E[\int_{t_0}^t \alpha_\sigma (e^{iw_\sigma} - 1) dn_\sigma | B_{\lambda_t}]\} = 0$$

The fourth and fifth terms are zero by virtue of (2.25) and the B_{ν_σ}-measurability of the integrands. We are left with

$$E\{(z - z_t^*)\alpha_t\} =$$

$$E\left\{\zeta\int_{t_0}^{t}\alpha_\sigma(\lambda_\sigma-\hat{\lambda}_\sigma)(e^{iW_\sigma}-1)d\sigma\right\}-E\left\{\int_{t_0}^{t}\alpha_\sigma K_\sigma\hat{\lambda}_\sigma(e^{iW_\sigma}-1)d\sigma\right\} \quad (3.9)$$

Now

$$E\left\{\zeta\int_{t_0}^{t}\alpha_\sigma(\lambda_\sigma-\hat{\lambda}_\sigma)(e^{iW_\sigma}-1)d\sigma\right\}$$

$$=\int_{t_0}^{t}E\left\{\alpha_\sigma(e^{iW_\sigma}-1)E\left[\zeta(\lambda_\sigma-\hat{\lambda}_\sigma)\,\middle|\,B_{\nu_\sigma}\right]\right\}d\sigma$$

$$=\int_{t_0}^{t}E\left\{\alpha_\sigma(e^{iW_\sigma}-1)K_\sigma\hat{\lambda}_\sigma\right\}d\sigma \quad ;$$

so the terms on the r.h.s. of (3.9) cancel, and $E\{(z - z_t^*)\,\alpha_t\} = 0$ for arbitrary $W_\sigma \in L_2[T]$. Therefore,

$$z_t^* = \hat{z}_t = \hat{z}_{t_0} + \int_{t_0}^{t} K_\sigma d\nu_\sigma \, .$$

This completes the proof. ∎

The proof of Theorem 3.1 (ii) appears to be very non-constructive, in that it looks as though $\hat{z}_t$, 3.1, was magically picked out of a hat. Indeed, our choice of $\hat{z}_t$ was really a well-motivated "educated guess." The conditional expectation $E(z|B_{\nu_t})$ is easily seen to be a martingale; thus, it is reasonable to assume that $\hat{z}_t$ can be represented as a stochastic integral, with the integrand a B_{N_t}-measurable functional[1-3,16]. With no further assumptions about the integrand, we could have arrived at Eq. (3.9) from which the form of K_σ is obvious.

If z_t^* is constrained to be the linear MMSE estimate† of z, the "projection theorem," Lemma D.2, becomes[16]

$$E\{(z - z_t^*)\int_{t_0}^{t} W_\sigma \, d\nu_\sigma\} = 0, \tag{3.10}$$

(the well-known linear orthogonality test[4]), and the linear estimate $\hat{z}_t^{\ell}$ is easily shown to be

$$\hat{z}_t^{\ell} = \hat{z}_{t_0}^{\ell} + \int_{t_0}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma)'\} E^{-1}\{\Lambda_\sigma\} d\nu_\sigma \tag{3.11}$$

Again,

$$d\nu_t = dN_t - \hat{\lambda}_t dt \tag{3.12}$$

The proof of (3.11) is a simple application of the properties of stochastic integrals. Note that (3.11) differs from the nonlinear case only in that the expectations in the integrand are total rather than conditional (see Frost[16] for a similar result in the Gaussian case).

By writing z_t in the form, for scalar observations,

$$\hat{z}_t = \hat{z}_{t_0} + \int_{t_0}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma) | B\nu_\sigma\} \hat{\lambda}_\sigma^{-\frac{1}{2}} dr_\sigma \tag{3.13}$$

with $dr_t = \hat{\lambda}_t^{-\frac{1}{2}} d\nu_t$, we can give an interesting alternative

† Linear functional of $\{\nu_\sigma, \sigma \in [t_0, t]\}$.

interpretation of $\hat{z}_t$. First we note that, by virtue of Eq. (2.25),

$$E\{r_t - r_s\} = E\{\int_s^t \hat{\lambda}_\sigma^{-\frac{1}{2}} d\nu_\sigma\} = 0 \quad , t > s . \tag{3.14}$$

Also,

$$E\{(r_t - r_s)^2 | B_{\nu_s}\} = E\{\int_s^t (\hat{\lambda}_\sigma^{-\frac{1}{2}})^2 \hat{\lambda}_\sigma d\sigma | B_{\nu_s}\}$$

$$= t - s, \; t > s. \tag{3.15}$$

i.e., $\{r_t, t \in T\}$ is a process with wide-sense stationary increments. It is obvious also that it is a martingale with orthogonal increments. (See Doob[1], p. 100). Thus, $\hat{\lambda}_t^{-\frac{1}{2}}$ "whitens" the process $\{\nu_t, t \in T\}$.

The process $\{r_t, t \in T\}$ has the attributes of an "innovations" process[16,98]; indeed, (3.13) is similar to a result obtained by Frost[16] for the problem of estimating random signals in Gaussian noise. Our whitened process $\{r_t, t \in T\}$ corresponds to his innovations process $\{\nu_t, t \in T\}$. It should be noted, however, that $\{r_t, t \in T\}$ might not contain all the information carried by $\{\nu_t, t \in T\}$; i.e., we have no assurance that $B_{\nu_t} = B_{r_t}$ (Lemma D.1 does not apply, here). Since $B_{\nu_t} = B_{N_t}$, it is perhaps more appropriate to call $\{\nu_t, t \in T\}$ the innovations process, even though it does not have stationary increments as in (3.15). Also, the formula defining $\{\nu_t, t \in T\}$,

$$d\nu_t = dN_t - \hat{\lambda}_t dt, \tag{3.16}$$

is identical in form to the relation defining Gaussian innovations.

2. Regular Point Processes.

The following theorem extends the results of the previous section to the broader class of regular point processes.

Theorem 3.2.

Let $\{N_t, t \in T\}$ be a vector RPP with conditionally independent components, and let z be a vector r.v. with components in L_2. Assume that the intensity $\{\lambda_t, t \in T\}$ depends causally on $\{N_t, t \in T\}$ and satisfies the three conditions listed in the statement of Theorem 3.1. Define $\{\nu_t, t \in T\}$, $\hat{\lambda}_t$ and $\hat{\Lambda}_t$ as in Theorem 3.1. Then,

(i) $B_{N_t} = B_{\nu_t}$, $t \in T$;

(ii) $\hat{z}_t = E(z | B_{N_t})$

$$= \hat{z}_{t_0} + \int_{t_0}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{\nu_\sigma}\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma \tag{3.17}$$

This theorem is simply a formal statement that (3.1) holds for regular point processes with conditionally independent components. Note that the intensity defined by (2.1)-(2.3) is given by $\lambda_t(B_{N_t}) = E\{\lambda_t | B_{N_t}\} = \hat{\lambda}_t$.

Proof.

Part (i) is immediate in view of Lemma D.1.

The proof of part (ii) is identical to the proof carried out for Th. 3.1 (ii). It is easily verified that the steps are valid for RPP's. Step 1 is valid by virtue of the generality of Lemma D.2. The definitions of Step 2 cause no difficulties. In Step 3, $\{\alpha_t, t \in T\}$ is again a martingale; Eq. (3.7) retains the same form, because the Ito rule, Th. 2.1, is the same whether $\{N_t, t \in T\}$ is a DSPP or a RPP. Following (3.7), the remaining manipulations of Step 3 are unchanged, and we can again show that $E\{(z - z_t^*)\alpha_t\} = 0$. ∎

In the next section we apply Theorems 3.1 and 3.2 to a variety of estimation problems. Since Th. 3.1 is a special case of Th. 3.2, we shall usually reference only the latter, unless we are specifically concerned with DSPP's.

D. Applications of the Representation Theorem.

In this section we apply the representation results of the preceding section to various problems in prediction, causal filtering, non-causal filtering, and parameter estimation.

We begin by proposing a particular model for the random process to be estimated. The structure of this model will allow us to derive continuously evolving process estimates from Theorems 3.1 or 3.2, even though those theorems are stated in the context of parameter estimation. Let $\{x_t, t \in T\}$ be a L_2 Ito process defined by the equation,

$$dx_t = f_t dt + g_t d\xi_t, \tag{3.18}$$

where $\{\xi_t, t \in T\}$ is a L_2 martingale, and $\{f_t, t \in T\}$ and $\{g_t, t \in T\}$ are L_2 processes causally dependent on $\{x_t, t \in T\}$ and $\{\xi_t, t \in T\}$. f_t and g_t can be causal functions of other processes, as well. This Ito process model is very general, embodying not only Markov processes specified by stochastic differential equations, but also processes which have a much more complex dependence on their past histories.

In what follows, we shall develop estimation equations for the Ito process $\{x_t, t \in T\}$, assuming that the intensity λ_t of the counting process $\{N_t, t \in T\}$ is a causal functional of $\{x_t, t \in T\}$; i.e., $B_{x_t} \subset B_{\lambda_t}$. Except where otherwise noted,

$\{N_t,\ t \in T\}$ can be taken to be either a vector DSPP, or more generally, a counting process of a vector RPP with conditionally independent components.

1. Prediction.

Let $\hat{x}_{t|\tau} = E\{x_t | B_{N_\tau}\}$. Suppose it is desired to find an expression for $\hat{x}_{t|\tau}$ for $t \geqslant \tau$, based upon observations of the vector process $\{N_\sigma,\ \sigma \in [t_o, \tau]\}$. Th. 3.2 yields the immediate result,

$$\hat{x}_{t|\tau} = \hat{x}_{t|t_o} + \int_{t_o}^{\tau} E\{x_t(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{\nu_\sigma}\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma \qquad (3.19)$$

This was obtained simply by substituting the r.v. x_t for z in the theorem.

2. Causal Filtering.

Set $\tau = t$ in (3.19) to get the equation for the filtered estimate,

$$\hat{x}_t = \hat{x}_{t|t} = \hat{x}_{t|t_o} + \int_{t_o}^{t} E\{x_t(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{\nu_\sigma}\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma \qquad (3.20)$$

The following theorem illuminates the structure of $\hat{x}_t$ by demonstrating that it is an Ito process.

Theorem 3.3.

The filtered estimate $\hat{x}_t = \hat{x}_{t|t}$ satisfies the stochastic differential equation,

$$d\hat{x}_t = \hat{f}_t dt + E\{x_t(\lambda_t - \hat{\lambda}_t)' | B_{\nu_t}\} \hat{\Lambda}_t^{-1} d\nu_t \qquad (3.21)$$

where $\hat{f}_t = E\{f_t | B_{\nu_t}\}$.

<u>Proof.</u>

Let $s > t$; then we can write, using Th. 3.2,

$$\hat{x}_{s|s} - \hat{x}_{t|t} = E\{x_s - x_t | B\nu_t\}$$

$$+ \int_t^s E\{x_\sigma(\lambda_\sigma - \hat{\lambda}_\sigma)' | B\nu_\sigma\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma$$

$$+ \int_t^s E\{(x_s - x_\sigma)(\lambda_\sigma - \hat{\lambda}_\sigma)' | B\nu_\sigma\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma \tag{3.22}$$

The last term is zero because λ_σ is independent of the increment $(x_s - x_\sigma)$; i.e., by causality,

$$E\{(x_s - x_\sigma)(\lambda_\sigma - \hat{\lambda}_\sigma)' | B\nu_\sigma\} =$$

$$E\{(x_s - x_\sigma) \underbrace{E[(\lambda_\sigma - \hat{\lambda}_\sigma)' | B\nu_\sigma \otimes B(x_\alpha, \alpha \in [\sigma, s])]}_{=0} | B\nu_\sigma\} = 0$$

The first term is, from (3.18),

$$E\{x_s - x_t | B\nu_t\} = E\{\int_t^s f_\sigma d\sigma + \int_t^s g_\sigma d\xi_\sigma | B\nu_t\}$$

$$= E\{\int_t^s f_\sigma d\sigma | B\nu_t\} + E\{E[\int_t^s g_\sigma d\xi_\sigma | B\nu_t \otimes B\xi_t] | B\nu_t\}$$

$$= E\{\int_t^s f_\sigma d\sigma | B\nu_t\} = E\{\int_t^s \hat{f}_\sigma d\sigma | B\nu_t\} \tag{3.23}$$

Now,

$$d\hat{x}_t = \lim_{s \downarrow t} (\hat{x}_{s|s} - \hat{x}_{t|t})$$

$$= \hat{f}_t dt + E\{x_t(\lambda_t - \hat{\lambda}_t)' | B\nu_t\} \hat{\Lambda}_t^{-1} d\nu_t, \text{a.s.}$$

The interchange of limit and expectation in the first term can be justified by the conditions imposed on $\{f_t, t \in T\}$ and the bounded convergence theorem[1,3]. ∎

We remark that (3.21) can be verified indirectly by appealing to Lemma D.2. We also note that (3.21) can be written in the equivalent form,

$$d\hat{x}_t = \hat{f}_t dt + E\{(x_t - \hat{x}_t)\lambda'_t \mid B_{\nu_t}\}\hat{\Lambda}_t^{-1} d\nu_t \tag{3.24}$$

by virtue of the fact that

$$E\{x_t(\lambda_t - \hat{\lambda}_t)' \mid B_{\nu_t}\} = E\{(x_t - \hat{x}_t)(\lambda_t - \hat{\lambda}_t)' \mid B_{\nu_t}\}$$

$$= E\{(x_t - \hat{x}_t)\lambda'_t \mid B_{\nu_t}\} \tag{3.25}$$

3. Canonical Filtering Equation.

Suppose it is desired to estimate causally the scalar process $y_t = \exp(iv'x_t)$ based upon observations of $\{N_\sigma, \sigma \in [t_o, t]\}$. $\{x_t, t \in T\}$ is the Ito process described by (3.18). It is of interest to estimate y_t because $\hat{y}_t$ is the conditional characteristic function of x_t. Theorem 3.4 gives a stochastic differential equation for the estimate, $\{\hat{y}_t, t \in T\}$.

Theorem 3.4.

The conditional characteristic function $\{\hat{y}_t, t \in T\}$ of $\{x_t, t \in T\}$ relative to $\{N_t, t \in T\}$ satisfies the equation

$$d\hat{y}_t = E\{\Psi_t(v)\exp(iv'x_t) \mid B_{\nu_t}\}dt \tag{3.26}$$

$$+ E\{(\lambda_t - \hat{\lambda}_t)'\exp(iv'x_t) \mid B_{\nu_t}\}\hat{\Lambda}_t^{-1} d\nu_t$$

where v is a non-random vector parameter and $\psi_t(v)$ is the characteristic form of the differential generator[2] of x_t,

$$\psi_t(v) = \mathop{p\,\lim}_{\Delta t \to 0}(\Delta t)^{-1} E\{\exp[iv'(x_{t+\Delta t} - x_t)] - 1 \mid x_t, B_{\nu_t}\} \tag{3.27}$$

Proof.

Using the same sort of argument that led to (3.21), we find

$$d\hat{y}_t = \lim_{s \downarrow t} E\{y_s - y_t \mid B_{\nu_t}\} + E\{y_t(\lambda_t - \hat{\lambda}_t)' \mid B_{\nu_t}\}\hat{\Lambda}_t^{-1} d\nu_t$$

The first term is

$$\lim_{s \downarrow t} E\{e^{iv'x_s} - e^{iv'x_t} \mid B_{\nu_t}\} =$$

$$= \lim_{s \downarrow t} E\{e^{iv'x_t}[e^{iv'(x_s - x_t)} - 1] \mid B_{\nu_t}\}$$

$$= \lim_{s \downarrow t} E\{e^{iv'x_t} E[e^{iv'(x_s - x_t)} - 1 \mid x_t, B_{\nu_t}] \mid B_{\nu_t}\}$$

$$= E\{e^{iv'x_t}\psi_t(v) \mid B_{\nu_t}\} dt$$

The interchange of limit and expectation can be justified by the bounded convergence theorem and the assumptions[1,3] on $\{x_t, t \in T\}$. ∎

This "canonical filtering theorem" was obtained by Snyder[23] by different means. Note that $\hat{y}_t$ is a sufficient statistic for $\hat{x}_t$ relative to $\{N_\sigma, \sigma \in [t_0, t]\}$. Indeed, (3.21) can be obtained from (3.26) by differentiation, as can higher conditional moments. As for the form of $\psi_t(v)$, if $\{\xi_t, t\epsilon T\}$ is, for example, a standardized* Wiener process, then

$$\psi_t(v) = iv'f_t - \tfrac{1}{2}v'g_tg_t'v. \tag{3.28}$$

Other forms can be found in Skorokhod[2].

Eq. (3.26) contains a number of subsidiary results of interest. If we average over the data $\{N_\sigma, \sigma \epsilon [t_0, t]\}$, (3.26) reduces to

$$\frac{\partial}{\partial t} E\{y_t\} = E\{\psi_t(v)e^{iv'x_t}\} \tag{3.29}$$

The last term in (3.26) disappears as a consequence of (2.25). Eq. (3.29) describes the time evolution of the (unconditional) characteristic function of $\{x_t, t\epsilon T\}$. Substituting (3.28), the differential generator of a Wiener process, into (3.29) and inverse Fourier transforming results in the forward Fokker-Planck equation for the p.d.f. $p(x_t)$ of x_t,

$$\frac{\partial}{\partial t} p(x_t) = L^+[p(x_t)] \tag{3.30}$$

$L^+[\cdot]$ is the forward Kolomogorov differential operator[2,3] for the Markov process $\{x_t, t \epsilon T\}$:

$$L^+[\cdot] = -\sum_i \frac{\partial}{\partial x_i}[f_i(\cdot)] + \frac{1}{2}\sum_i\sum_j [g_tg_t']_{ij} \frac{\partial^2(\cdot)}{\partial x_i \partial x_j} \tag{3.31}$$

---

*A standardized Wiener process has an identity (I) variance matrix; e.g., $E\{\xi_t \xi_s'\} = I t\wedge s$.

We have assumed in (3.31) that g_t is a non-random time function. Substituting (3.28) into (3.26) and inverse transforming leads to the important stochastic differential equation for the conditional p.d.f. of x_t relative to $\{N_\sigma, \sigma \in [t_o, t]\}$,

$$dp(x_t|B_{N_t}) = L^+[p(x_t|B_{N_t})]dt$$
$$+ p(x_t|B_{N_t})(\lambda_t - \hat{\lambda}_t)' \hat{\Lambda}_t^{-1} d\nu_t$$

Snyder[23] has pointed out the striking similarity between this equation, which is now data-dependent, and the corresponding equation for conditionally Gaussian observations. Eq. (3.21) for $d\hat{x}_t$ can be obtained directly from the equation above by multiplying by x_t and integrating, with appropriate boundary conditions[14,23].

4. Non-causal Filtering.

Let us estimate x_t based on observations $\{N_\sigma, \sigma \in [t_o, \tau]\}$, $\tau \geq t$. From Theorem 3.2 we have

$$\hat{x}_{t|\tau} = \hat{x}_{t|t} + \int_t^\tau E\{x_t(\lambda_\sigma - \hat{\lambda}_\sigma)'|B_{\nu_\sigma}\}\hat{\Lambda}_\sigma^{-1} d\nu_\sigma \qquad (3.32)$$

This is easily seen to be the sum of two orthogonal estimates, because the stochastic integral is a martingale in its upper limit (see Ch. II), and has initial value zero. The first term is the causal filtered estimate, and the second is an estimate "backward in time" of x_t given data up to time τ.

We use the term "backward" because in (3.32) x_t is evaluated at the lower limit of the integral, whereas in (3.20) x_t is evaluated at the upper limit of the integral. Clearly, with $\tilde{x}_{a|b} = x_a - \hat{x}_{a|b}$,

$$\tilde{x}_{t|t} = \tilde{x}_{t|\tau} + \int_t^\tau E\{x_t(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{\nu_\sigma}\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma \tag{3.33}$$

The two terms on the right are orthogonal, because the error is orthogonal to measurable functionals of the data (see Appendix A, Section 7); thus,

$$E\{\tilde{x}_{t|t}\tilde{x}'_{t|t}\} = E\{\tilde{x}_{t|\tau}\tilde{x}'_{t|\tau}\} + E\{II'\} \tag{3.34}$$

where

$$I \triangleq \int_t^\tau E\{x_t(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{\nu_\sigma}\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma$$

That is, for $\tau > t$, $E\{II'\} > 0$; hence,

$$E\{\tilde{x}_{t|t}\tilde{x}'_{t|t}\} > E\{\tilde{x}_{t|\tau}\tilde{x}'_{t|\tau}\} \tag{3.35}$$

5. Parameter Estimation -- Detection.

Suppose the intensity process of $\{N_t, t \in T\}$ is $\{\lambda_t = \alpha \mu_t + \mu_0, t \in T\}$ where α is a zero-one random variable, μ_t is a process non-negative w.p.1, and μ_0 is a positive constant. It is desired to estimate the parameter α.

Applying Theorem 3.2, we have

$$\hat{\alpha}_t = \hat{\alpha}_{t_0} + \int_{t_0}^{t} E\{\alpha(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{\nu_\sigma}\} \hat{\Lambda}_\sigma^{-1} d\nu_\sigma \tag{3.36}$$

or equivalently,

$$d\hat{\alpha}_t = E\{\alpha(\lambda_t - \hat{\lambda}_t)' | B_{\nu_t}\} \hat{\Lambda}_t^{-1} d\nu_t \tag{3.37}$$

Now,

$$\hat{\lambda}_t = E\{\lambda_t | B_{N_t}\} = E\{\alpha\mu_t + \mu_0 | B_{N_t}\}$$

$$= \hat{\alpha}_t \hat{\mu}_t + \mu_0 \tag{3.38}$$

where $\hat{\mu}_t$ is the "pseudo-estimate,"

$$\hat{\mu}_t = E\{\mu_t | B_{N_t}, \alpha = 1\}, \tag{3.39}$$

$$\hat{\alpha}_t = E\{\alpha | B_{N_t}\} = Pr\{\alpha = 1 | B_{N_t}\}$$

and $\hat{\Lambda}_t = \mathrm{diag}(\hat{\alpha}_t \hat{\mu}_t + \mu_0)$

Using these results, Eq. (3.37) becomes

$$d\hat{\alpha}_t = E\{\alpha(\alpha\mu_t' - \hat{\alpha}_t \hat{\mu}_t') | B_{\nu_t}\} \hat{\Lambda}_t^{-1} d\nu_t \tag{3.40}$$

This can be simplified by using $d\nu_t = dN_t - \hat{\lambda}_t dt$ and noting that $\widehat{\alpha^2}_t = \hat{\alpha}_t$; Eq. (3.40) reduces to

$$d\hat{\alpha}_t = \hat{\alpha}_t(1-\hat{\alpha}_t)\hat{\mu}_t'\hat{\Lambda}_t^{-1}d\nu_t$$

$$= -\hat{\alpha}_t(1-\hat{\alpha}_t)\|\hat{\mu}_t\|_1 dt + \hat{\alpha}_t(1-\hat{\alpha}_t)\hat{\mu}_t'\hat{\Lambda}_t^{-1}dN_t \qquad (3.41)$$

where $\|\cdot\|_1$ denotes the operation of summing the magnitudes of the components of $(\cdot)$.

Eq. (3.41) is a stochastic differential equation in $\hat{\alpha}_t$, which can be solved explicitly by making the change of variable $\ell_t = \ell n(\hat{\alpha}_t/1 - \hat{\alpha}t)$. To determine the equation for ℓ_t, we apply the Ito rule for Poisson processes to ℓ_t, using (3.41). The result of this is

$$d\ell_t = -\|\hat{\mu}_t\|_1 dt + \{\ell n[\Lambda_0^{-1}(\mu_0+\hat{\mu}_t)]\}' dN_t \qquad (3.42)$$

where Λ_0 is a diagonal matrix with non-zero entries the components of μ_0, and $\ell n(\cdot)$ is the vector natural logarithm; i.e., $\ell n(x)$ is a vector with components the natural logs of the components of x. Note that our change of variable has effected a separation of variables, so that ℓ_t is given explicitly by

$$\ell_t = \ell_{t_0} - \int_{t_0}^{t}\|\hat{\mu}_\sigma\|_1 d\sigma + \int_{t_0}^{t}\{\ell n[\Lambda_0^{-1}(\mu_0+\hat{\mu}_\sigma)]\}' dN_\sigma \qquad (3.43)$$

Then, of course,

$$\hat{\alpha}_t = \frac{e^{\ell_t}}{1+e^{\ell_t}} \cdot \tag{3.44}$$

Although (3.44) is the answer to our problem, the quantity ℓ_t is of much greater inherent interest, as ℓ_t is the <u>log-likelihood ratio</u> for deciding between the hypotheses

H_1: intensity process of $\{N_t, t \in T\}$ is

$$\{\mu_t + \mu_0 , t \in T\} ;$$

H_0: intensity process of $\{N_t, t \in T\}$ is μ_0, $t \in T$; based on the observations $\{N_\sigma, \sigma \in [t_0, t]\}$. ℓ_{t_0} is the <u>a priori</u> log-likelihood ratio $\ln[\Pr(\alpha = 1)/\Pr(\alpha = 0)]$. It should be emphasized that $\hat{\mu}_t = E\{\mu_t | B_{N_t}, \alpha = 1\}$ has an interpretation as a MMSE causal estimate only when $\alpha = 1$; thus the name, "pseudo-estimate."

Eq. (3.43) is the extension of the well-known Reiffen-Sherman detector[37] to stochastic intensity functions and vector observations. In scalar form,

$$\ell_t = \ell_{t_0} - \int_{t_0}^{t} \hat{\mu}_\sigma d\sigma + \int_{t_0}^{t} \ln \frac{\mu_0 + \hat{\mu}_\sigma}{\mu_0} dN_\sigma \tag{3.45}$$

it is identical to the result obtained by Snyder[23] under the somewhat more restrictive assumption that λ_t be a memoryless function of a Markov process. Eq. (3.45) was also obtained by Evans[15] using different methods.

Eq. (3.43) is the fundamental detection equation for binary hypothesis testing. The first integral in (3.43) is an energy bias term: the second integral is nothing more than a weighted counting operation: the individual components of the vector observations are weighted continuously in time according to the estimated intensities of the components. We emphasize that no special restrictions were placed on the process $\{\mu_t, t \in T\}$; it need not be a function of a Markov process, nor need it have uncorrelated components. We required only that $\{N_t, t \in T\}$ have conditionally independent components.

Eq. (3.43) is even more general than it appears, in that log-likelihood ratios for other detection problems can be obtained from it by using the chain rule for likelihood ratios[13,23]. For example, the following two problems are easily handled:

(1) $H_1 : \lambda_t = \mu_t^{(1)} + \mu_o$

$$H_0 : \lambda_t = \mu_t^{(0)} + \mu_o \qquad (3.46)$$

(2) $H_1 : \lambda_t = s(t) + \mu_t + \mu_o$

$$H_0 : \lambda_t = \mu_t + \mu_o \quad [s(t) \text{ nonrandom}] \qquad (3.47)$$

E. Another Approach to Detection.

We saw in the preceding section that detection results could be obtained from the solution to a certain estimation problem. The problem we considered served to illustrate the use of Th. 3.2 for parameter estimation; using the result of the following theorem, we can solve, with both elegance and economy, a wider range of hypothesis testing problems. Indeed, the theorem to be proved below has potential value in a diversity of problems, including process characterization and estimation, as well as detection.

Theorem 3.5.

Let $\{N_t, t \in T\}$ be a doubly-stochastic Poisson process with rate process $\{\lambda_t, t \in T\}$ satisfying $E(\lambda_t) < \infty$, $t \in T$, and event times $\{t_1, t_2, \ldots, t_{N_t}\} \in [t_0, t]$. Then the joint p.d.f. of the $\{t_i\}$ and N_t is given by

$$p(\{t_i\}; N_t) = \exp\left[-\int_{t_0}^{t} \hat{\lambda}_\sigma d\sigma\right] \prod_{i=1}^{N_t} \hat{\lambda}_{t_i-} \tag{3.48}$$

$$= \exp\left\{-\int_{t_0}^{t} \hat{\lambda}_\sigma d\sigma + \int_{t_0}^{t} \ln \hat{\lambda}_\sigma dN_\sigma\right\} \tag{3.49}$$

where $\hat{\lambda}_t = E\{\lambda_t \mid B_{N_t}\}$, $t \in T$,

and $t_i - = \lim\limits_{h \downarrow 0}(t_i - h)$

This theorem, first stated by the author in Reference 80, gives a complete statistical description of $\{N_t, t \in T\}$ in terms of the causal MMSE estimate $\hat{\lambda}_t$. Except for replacing λ_t with $\hat{\lambda}_t$, $p(\{t_i\}; N_t)$ is identical in form to (2.5). Of course, the notation in (3.48) and (3.49) can be rather misleading, as $\hat{\lambda}_t$ depends on $\{t_i\}$ and N_t in a complex way, whereas λ_t in (2.5) is not a function of $\{t_i\}$ or N_t. Nevertheless, the similarity of these representations is striking.

Proof.

For notational simplicity, let $p_{t|\lambda} = p(\{t_i\}; N_t | B_{\lambda_t})$. Also, define $x_t = -\int_{t_0}^{t} \lambda_\sigma d\sigma + \int_{t_0}^{t} \ln \lambda_\sigma dN_\sigma$; then,

$$p_{t|\lambda} = \exp x_t. \tag{3.50}$$

Clearly,

$$dx_t = -\lambda_t dt + \ln\lambda_t dN_t$$

Now apply the Ito rule for Poisson processes, Th. 2.1, to $p_{t|\lambda}(x_t) = \exp x_t$:

$$dp_{t|\lambda} = -p_{t|\lambda}\lambda_t dt + p_{t|\lambda}(\lambda_t - 1) dN_t \tag{3.51}$$

Equivalently, using the obvious initial condition, $p_{t_0|\lambda} = 1$,

$$p_{t|\lambda} = 1 - \int_{t_0}^{t} p_{\sigma|\lambda}\lambda_\sigma d\sigma + \int_{t_0}^{t} p_{\sigma|\lambda}(\lambda_\sigma - 1) dN_\sigma \tag{3.52}$$

Taking the expectation of (3.52) over all possible sample paths of $\{\lambda_\sigma, \sigma \in [t_0, t]\}$, holding $\{N_\sigma, \sigma \in [t_0, t]\}$ fixed, we get

$$p_t = p(\{t_i\}; N_t) = E_\lambda\{p_{t|\lambda}\}$$

$$= 1 - \int_{t_o}^{t} E_\lambda\{p_{\sigma|\lambda}\lambda_\sigma\}d\sigma + \int_{t_o}^{t} E_\lambda\{p_{\sigma|\lambda}(\lambda_\sigma - 1)\}dN_\sigma \tag{3.53}$$

The interchange of expectation and integration is valid for the stochastic integral above, because $\{N_\sigma, \sigma \in [t_o, t]\}$ is fixed and the integral is well-defined (i.e., exists) in the Lebesgue-Stieltjes sense. The interchange in the time-integral in (3.53) is valid because $E_\lambda\{p_{\sigma|\lambda}\lambda_\sigma\} < \infty$ (See Appendix A, Section 8, or Doob[1], Ch. II).

Let us evaluate $E_\lambda\{p_{t|\lambda}\lambda_t\}$. Assume that $\{N_\sigma, \sigma \in [t_o, t]\}$ and $\{\lambda_\sigma, \sigma \in [t_o, t]\}$ are defined on the probability spaces (Ω_N, B_N, P_N) and $(\Omega_\lambda, B_\lambda, P_\lambda)$, where ω_N and ω_λ denote points in Ω_N and Ω_λ. Define the product space $(\Omega_N \times \Omega_\lambda, B_N \otimes B_\lambda, P)$, where P is the joint probability measure of $\{N_\sigma, \sigma \in [t_o, t]\}$ and $\{\lambda_\sigma, \sigma \in [t_o, t]\}$. Let $\mathcal{N} = \{\omega_N : t_i \le \tau_i ; i = 1, \ldots, N_t ; N_t = n\}$; then, using Bayes' Theorem (clearly applicable here, because $\{N_t, t \in T\}$ is w.p.1 characterized by a finite number of r.v.'s[1]),

$$E_\lambda\{\lambda_t P(\mathcal{N}|B_{\lambda_t})\} = \int_{\Omega_\lambda} \lambda_t P(\mathcal{N}|B_{\lambda_t}) P_\lambda(d\omega_\lambda)$$

$$= \int_{\Omega_\lambda} \lambda_t P(\mathcal{N}, d\omega_\lambda) = \int_{\Omega_\lambda} \lambda_t P(d\omega_\lambda|\mathcal{N}) P_N(\mathcal{N})$$

$$= \hat{\lambda}_t P_N(\mathcal{N}) \tag{3.54}$$

Thus,

$$E_\lambda\{\lambda_t p_{t|\lambda}\} = \hat{\lambda}_t p_t$$

and (3.53) becomes

$$p_t = 1 - \int_{t_0}^{t} p_\sigma \hat{\lambda}_\sigma d\sigma + \int_{t_0}^{t} p_\sigma (\hat{\lambda}_\sigma - 1) dN_\sigma$$

Comparing this with (3.52) and (3.50), we see that the solution is

$$p_t = \exp\left\{ -\int_{t_0}^{t} \hat{\lambda}_\sigma d\sigma + \int_{t_0}^{t} \ln \hat{\lambda}_\sigma \, dN_\sigma \right\}$$

This completes the proof. ∎

The reader can easily verify that the detection result of the previous section can be obtained from (3.49) by inspection.

We remark that the same method by which we proved Th. 3.5, particularly the steps following (3.53), can be used to average the equation[5]

$$\frac{\partial}{\partial t} p_{N_t}(n|\lambda) = -\lambda_t p_{N_t}(n|\lambda) + \lambda_t p_{N_t}(n-1|\lambda) \tag{3.55}$$

over the sample paths of $\{\lambda_\sigma, \sigma \in [t_o, t]\}$ to get

$$\frac{\partial}{\partial t} p_{N_t}(n) = -\hat{\lambda}_t^0 p_{N_t}(n) + \hat{\lambda}_t^1 p_{N_t}(n-1) \tag{3.56}$$

where $p_{N_t}(n|\lambda) = \Pr\{N_t = n | B_{\lambda_t}\}$, and $\hat{\lambda}_t^i = E\{\lambda_t | N_t = i\}$.

Here we observe that $\hat{\lambda}_t$ is an expectation conditioned only on N_t, the total number of events in $[t_0, t]$. The equation above governs the evolution in time of the counting distribution of any doubly-stochastic Poisson process, and is potentially useful for hypothesis testing problems in which the only available data is N_t, rather than $\{N_\sigma, \sigma \in [t_0, t]\}$.

For a particular type of doubly-stochastic Poisson process, the equation for $p_{N_t}(n)$ takes the form of a simple recursive relation for which calculations are especially easy. Let $\lambda_t = z\mu(t)$, with z a non-negative random variable and $\mu(t)$ a non-negative deterministic function. Then the equation for $p_{N_t}(n)$, together with Eq. (E.20) (Appendix E) yields

$$p_{N_t}(n + 1) = \frac{1}{n+1}\left\{n\, p_{N_t}(n) - \frac{m(t)}{\mu(t)}\frac{\partial}{\partial t} p_{N_t}(n)\right\} \tag{3.57}$$

where $m(t) = \int_{t_0}^{t}\mu(\sigma)d\sigma$ and $p_{N_t}(0) = E\{e^{-zm(t)}\}$.

In the statement of Theorem 3.5 we required that $\{N_t, t \in T\}$ be a doubly-stochastic Poisson process. Rubin subsequently proved in a recent paper[68] that (3.49) and (3.56) also hold if $\{N_t, t \in T\}$ is a regular point process. This generalization is discussed briefly in Appendix B.

F. Approximations.

All of the causal estimates for which differential equations have been derived, although realizable in principle because of their independence of future observations, are difficult to implement exactly, because of their dependence on estimates of the rate process $\{\lambda_t, t \in T\}$. Furthermore, exact estimates depend, in general, on estimates of moments of all orders. These difficulties occur in much the same way in the nonlinear estimation of random signals in additive Gaussian noise. Indeed, as we shall demonstrate, many approximation techniques developed for the latter problem are applicable to our estimators.

In this section we shall examine three types of approximations to the exact estimator equations. The first, the Taylor expansion method, is widely used in nonlinear estimation. The second is related to Volterra expansions and appears to be only of academic interest. The third type of approximation requires that estimators be linear functionals of the data. We conclude this section with a brief critique of these approximations.

1. Taylor Expansions.

The sub-optimum estimator equations to be derived below are based on finite series approximations of the exact stochastic equations. Their accuracy hinges on the degree to which the estimation error $\tilde{x}_t = x_t - \hat{x}_t$ is "small."

Eq. (3.21), reproduced below, describes the evolution of the MMSE estimate $\hat{x}_t$ of the Ito process x_t:

$$d\hat{x}_t = \hat{f}_t dt + E\{(x_t - \hat{x}_t)\lambda_t' \mid B_{\nu_t}\}\hat{\Lambda}_t^{-1} d\nu_t \qquad (3.58)$$

To use the Taylor expansion approach, we must restrict the processes $\{N_t, t \in T\}$ and $\{x_t, t \in T\}$ somewhat. We require $\{N_t, t \in T\}$ to be a DSPP; we also assume that $\{x_t, t \in T\}$ satisfies

$$dx_t = f_t(x_t)dt + g_t dw_t, \qquad (3.59)$$

where

1) $\{w_t, t \in T\}$ is a standardized vector Wiener process;
2) $f_t(x_t)$ is a memoryless function of t and x_t;
3) g_t is a non-random time function.

That is, $\{x_t, t \in T\}$ is a Markov diffusion process*. Conditions 1) and 3) are imposed primarily for convenience. Condition 2), however, is essential, as is the following additional condition:

4) $\lambda_t(x_t)$ is a memoryless function of t and x_t.

The necessity of conditions 2) and 4) is due to the fact that we shall be expanding f_t and λ_t in Taylor series in x_t.

*For this to be strictly true, f_t and g_t must obey mild regularity conditions[2].

To use these expansions, it will be necessary to have a stochastic differential equation for $P_t \triangleq E\{\tilde{x}_t\tilde{x}_t' | B_{N_t}\}$, the conditional error covariance matrix. We shall verify that P_t satisfies,

$$
\begin{aligned}
dP_t = & \{ \widehat{f_t\tilde{x}_t'} + \widehat{\tilde{x}_t f_t'} + g_t g_t' \} dt \\
& - x_t \tilde{\lambda}_t' \hat{\Lambda}_t^{-1} d\eta_t \hat{\Lambda}_t^{-1} \tilde{\lambda}_t x_t' \\
& + \widehat{\tilde{x}_t\tilde{x}_t'\tilde{\lambda}_t'} \hat{\Lambda}_t^{-1} d\nu_t
\end{aligned}
\tag{3.60}
$$

where $\tilde{\lambda}_t = \lambda_t - \hat{\lambda}_t$ and $d\eta_t = \text{diag}(dN_t)$.

For simplicity we prove only the scalar case. Using the properties of conditional expectation, we can write

$$
\begin{aligned}
\Delta P_t = P_{t+\Delta t} - P_t = & \widehat{x^2}_{t+\Delta t|t+\Delta t} - \widehat{x^2}_{t|t} \\
& - \left(\hat{x}^2_{t+\Delta t|t+\Delta t} - \hat{x}^2_{t|t} \right)
\end{aligned}
\tag{3.61}
$$

Thus we have the formal relation,

$$
dP_t = d\widehat{(x_t^2)} - d(\hat{x}_t^2), \tag{3.62}
$$

obtained from (3.61) by letting $\Delta t \to 0$. Eq. (3.62) makes sense as long as we interpret it as a limiting form of (3.61). Let us first obtain a stochastic differential equation for x_t^2. Since $\{x_t, t \in T\}$ is a diffusion, we can use the Ito rule for Wiener processes (Appendix C, Section 2) to write down an Ito equation for $\{x_t^2, t \in T\}$. The Ito rule immediately gives

$$d(x_t^2) = (2x_t f_t + g_t^2)\, dt + 2x_t g_t dw_t \tag{3.63}$$

Because (3.63) is an Ito equation, we can apply Theorem 3.3 to obtain a formula for the causal estimate $\widehat{x^2}_t$:

$$\widehat{d(x_t^2)} = (\widehat{2x_t f_t} + g_t^2)\, dt + \widehat{x_t^2 \tilde{\lambda}_t}\, \hat{\lambda}_t^{-1} d\nu_t \tag{3.64}$$

This takes care of the first term in (3.62). To evaluate the second term, we note that $\{\hat{x}_t, t \in T\}$ satisfies the Ito equation (3.21) (or, (3.58)); since Eq. (3.21) is driven by a counting process, we can use the differential rule for counting processes (Theorem 2.1) to obtain an Ito equation for $\{\hat{x}_t^2, t \in T\}$. We get, after some manipulation,

$$\begin{aligned} d(\hat{x}_t^2) = 2\hat{x}_t \hat{f}_t dt + (\widehat{x_t \tilde{\lambda}_t}\, \hat{\lambda}_t^{-1})^2 dN_t \\ + 2\hat{x}_t \widehat{x_t \tilde{\lambda}_t}\, \hat{\lambda}_t^{-1} d\nu_t \end{aligned} \tag{3.65}$$

Now, forming the difference in (3.62),

$$\begin{aligned} dP_t = (\widehat{2x_t f_t} - 2\hat{x}_t \hat{f}_t + g_t^2)\, dt - (\widehat{x_t \tilde{\lambda}_t}\hat{\lambda}_t^{-1})^2 dN_t \\ + \widehat{x_t^2 \tilde{\lambda}_t}\, \hat{\lambda}_t^{-1} d\nu_t - 2\hat{x}_t\, \widehat{x_t \tilde{\lambda}_t}\, \hat{\lambda}_t^{-1} d\nu_t \end{aligned} \tag{3.66}$$

Since $\hat{x}_t$ is B_{N_t}-measurable, we have $\widehat{\hat{x}_t{}^2\tilde{\lambda}_t} = 0$ (see Appendix A, Section 7). Thus we can add the term $\widehat{\hat{x}_t{}^2\tilde{\lambda}_t}\hat{\lambda}_t^{-1}d\nu_t$ to the r.h.s. of (3.66) without altering the equation. Collecting terms, (3.66) then becomes,

$$dP_t = (2\widehat{\tilde{x}_t f_t} + g_t{}^2)dt - (\widehat{x_t\tilde{\lambda}_t}\hat{\lambda}_t^{-1})^2 dN_t$$
$$+ \widehat{\tilde{x}_t{}^2\tilde{\lambda}_t}\hat{\lambda}_t^{-1}d\nu_t \tag{3.67}$$

This is the scalar version of (3.60). The proof of the vector form differs only in algebraic complexity. Now assume that f_t and λ_t have Taylor series expansions about the estimate $\hat{x}_t$ as

$$f_t = f_t(\hat{x}_t) + D_{\hat{x}_t}[f_t(\hat{x}_t)]'(x_t - \hat{x}_t)$$
$$+ \tfrac{1}{2}\{H_{\hat{x}_t}[f_t(\hat{x}_t)] : (x_t - \hat{x}_t)(x_t - \hat{x}_t)'\} + \cdots \tag{3.68}$$

and

$$\lambda_t = \lambda_t(\hat{x}_t) + D_{\hat{x}_t}[\lambda_t(\hat{x}_t)]'(x_t - \hat{x}_t)$$
$$+ \tfrac{1}{2}\{H_{\hat{x}_t}[\lambda_t(\hat{x}_t)] : (x_t - \hat{x}_t)(x_t - \hat{x}_t)'\} + \cdots$$

$$\tag{3.69}$$

where

$$D_x[f(x)] = \left[\frac{\partial f_j}{\partial x_i}\right] \tag{3.70}$$

is the Jacobian matrix of f,

$$H_x\left[f_k(x)\right] = \left[\frac{\partial^2 f_k}{\partial x_i \partial x_j}\right] \tag{3.71}$$

is the Hessian matrix of f, and, with A a matrix,

$$\left\{H_x\left[f(x)\right] : A\right\} = \begin{bmatrix} Tr\{H_x[f_1(x)]A\} \\ Tr\{H_x[f_2(x)]A\} \\ \vdots \end{bmatrix} \tag{3.72}$$

Using these expansions, the r.h.s. of (3.58) can be expressed as an infinite series of terms involving estimates of x_t and powers of $\tilde{x}_t$. Assuming convergence of the series, the resulting equation is still exact. The same substitutions can be made in (3.60), the conditional covariance equation. Approximate estimation equations are then obtained by truncating the series. The point of truncation determines the type of approximate equations obtained. Truncation serves the purpose of reducing the dimension of the system of equations to be solved.

For a <u>first order</u> approximation, we ignore in the expansions of λ_t and f_t terms which are quadratic or higher degree in $\tilde{x}_t$. By thus linearizing λ_t and f_t, we obtain the following approximate equations:

$$d\hat{x}_t = \{f_t(\hat{x}_t) - P_t \nabla_{\hat{x}_t}[\|\lambda_t(\hat{x}_t)\|_1]\}dt + P_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)]dN_t \tag{3.73}$$

$$dP_t = \{ D_{\hat{x}_t}[f_t(\hat{x}_t)]' P_t + P_t D_{\hat{x}_t}[f_t(\hat{x}_t)] + g_t g_t' \} dt$$

$$- P_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)] d\eta_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)]' P_t \qquad (3.74)$$

The last term on the r.h.s. of (3.74) is the value of $-d\hat{x}_t d\hat{x}_t'$ to o (dt). $d\eta_t$ is the matrix $dN_t dN_t'$; to o(dt), $d\eta_t = \mathrm{diag}(dN_t)$. $\nabla_{\hat{x}_t}$ is the gradient operator, and $\ln(\cdot)$ is the vector natural logarithm.

For a <u>second order</u> approximation, we retain quadratic as well as linear and constant terms in the expansions for λ_t and f_t. This results in approximate equations which are more complex than the first order equations, but hopefully more accurate:

$$d\hat{x}_t = \{ f_t(\hat{x}_t) + \tfrac{1}{2}\{ H_{\hat{x}_t}[f_t(\hat{x}_t)] : P_t \} - P_t \nabla_{\hat{x}_t}[\| \lambda_t(\hat{x}_t) \|_1] \} dt$$

$$+ P_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)] dN_t \qquad (3.75)$$

$$dP_t = \{ D_{\hat{x}_t}[f_t(\hat{x}_t)]' P_t + P_t D_{\hat{x}_t}[f_t(\hat{x}_t)] + g_t g_t'$$

$$+ \tfrac{1}{2} P_t \,\mathrm{Tr}\left(H_{\hat{x}_t}[\| \lambda_t(\hat{x}_t) \|_1] P_t \right) \} dt$$

$$- \tfrac{1}{2} P_t \{ H_{\hat{x}_t}[\lambda_t(\hat{x}_t)] : P_t \}' \Lambda_t^{-1}(\hat{x}_t) dN_t$$

$$- P_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)] d\eta_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)]' P_t \qquad (3.76)$$

In (3.76) we used the approximation, obtained from the Taylor series,

$$\widehat{\tilde{x}_t \tilde{x}_t' (\lambda_t - \hat{\lambda}_t)'} \cong -\frac{1}{2} P_t \{ H_{\hat{x}_t}[\lambda_t(\hat{x}_t)] : P_t \}' \tag{3.77}$$

When (3.77) is multiplied on the right by the Taylor approximation for $\hat{\Lambda}_t^{-1}[dN_t - \hat{\lambda}_t dt]$, the result is the fourth and fifth terms on the r.h.s. of (3.76).

These second order equations can be simplified somewhat by assuming that terms of the form $E\{\tilde{x}_t \tilde{x}_t' \tilde{x}_t \tilde{x}_t' | B_{N_t}\}$ can be factored into sums of products of conditional covariances, as though $\tilde{x}_t$ were Gaussian. It is clear that this assumption can only be justified a priori, on the grounds that it provides considerable simplification. As with the entire Taylor expansion approach, the moment factoring step is ad hoc and can only be evaluated based on performance in specific applications. The equations for the <u>Gaussian second order</u> filter are:

$$d\hat{x}_t = \{ f_t(\hat{x}_t) + \frac{1}{2}\{ H_{\hat{x}_t}[f_t(\hat{x}_t)] : P_t \} - P_t \nabla_{\hat{x}_t}[\| \lambda_t(\hat{x}_t) \|_1] \} dt$$

$$+ P_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)] \, dN_t \tag{3.78}$$

$$dP_t = \{ D_{\hat{x}_t}[f_t(\hat{x}_t)]' P_t + P_t D_{\hat{x}_t}[f_t(\hat{x}_t)] + g_t g_t'$$

$$- P_t H_{\hat{x}_t}[\|\lambda_t(\hat{x}_t)\|_1] P_t \} dt$$

$$+ P_t H_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)' dN_t] P_t \tag{3.79}$$

Note that (3.78) and (3.75) are the same. Eqs. (3.78) and (3.79) were originally derived by Snyder[23] for scalar observations. In the derivation of (3.79) we have used the approximation,

$$\widehat{\tilde{x}_t \tilde{x}_t' (\lambda_t - \hat{\lambda}_t)'} \hat{\Lambda}_t^{-1} [dN_t - \hat{\lambda}_t dt]$$

$$\cong \widehat{\tilde{x}_t \tilde{x}_t' (\lambda_t - \hat{\lambda}_t)'} \Lambda_t^{-1}(\hat{x}_t) [dN_t - \lambda_t(\hat{x}_t) dt] \tag{3.80}$$

$$\cong P_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)] d\eta_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)]' P_t$$

$$+ P_t H_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)' dN_t] P_t - P_t H_{\hat{x}_t}[\|\lambda_t(\hat{x}_t)\|_1] P_t dt \tag{3.81}$$

Eq. (3.80) was obtained by substituting the Taylor approximations of λ_t and Λ_t^{-1}. Higher order terms which would subsequently be neglected were not included. Eq. (3.81) resulted from the use of the moment factoring theorem.

The approximate filters we have determined have been based on truncated moment expansions. Another reasonable expansion upon which to base approximations is a truncated quasi-moment series[41-43]. It can be shown[86] that a quasi-moment expansion truncated after the second quasi-moment leads to the Gaussian second-order filter just derived. Moreover, Fisher's recent work[42] suggests that the retention of higher-than-second order quasi-moments results in an alarming increase in system complexity and computational effort.

Hence, we shall not pursue quasi-moment expansions any further at this time.

We emphasize that the approximations developed in this section are _ad hoc_ in nature. They depend strongly on the degree to which the error $x_t - \hat{x}_t$ is "small," and on the form of f_t and λ_t. The efficacy of the approximate filters can ultimately be gauged only in individual applications.

2. Iterated Integrals.

For the Gaussian estimation problem Frost[16] has proposed that the estimate $\hat{z}_t$ of a random variable z be expressed as an infinite series or orthogonal multiple Wiener integrals, using the "polynomial chaos" introduced by Wiener[39] and later refined by Ito[40]. This approach suggests potentially valuable approximations as well as interesting interpretations of many estimators, so we extend it here to the counting process estimation problem. It is worth mentioning now that the usefulness of the iterated integral approach depends strongly on the availability of the intensity estimate $\hat{\lambda}_t$, because the "innovations" process we shall use here is defined in terms of $\hat{\lambda}_t$.

Let $\{\xi_t, t \in T\}$ be a scalar martingale and let $\{k_n(\sigma_1, \sigma_2, \ldots, \sigma_n); t_o \leq \sigma_1 \leq \ldots \leq \sigma_n \leq t_1\}$ be $B_{\xi_{\sigma_n}}$ - measurable, L_2 functions. Define for n = 0, 1, 2, ... the multiple integral $I^n[k_n]$ as

$$I^n[k_n] = \int_{t_o}^{t} \int_{t_o}^{\sigma_n} \cdots \int_{t_o}^{\sigma_2} k_n(\sigma_1, \ldots, \sigma_n)\, d\xi_{\sigma_1} \cdots d\xi_{\sigma_n} \tag{3.82}$$

$$I^0[k_n] = \text{const.}$$

This is just a special case of the stochastic integral defined in Eq. (2.24), with

$$g_\sigma = \int_{t_o}^{\sigma} \int_{t_o}^{\sigma_{n-1}} \cdots \int_{t_o}^{\sigma_2} k_n(\sigma_1, \ldots, \sigma_{n-1}, \sigma)\, d\xi_{\sigma_1} \cdots d\xi_{\sigma_{n-1}}$$

For our purposes, it is convenient to let $\{\xi_t, t \in T\}$ be the "innovations" process $\{r_t, t \in T\}$, where $dr_t = \hat{\lambda}_t^{-\frac{1}{2}}[dN_t - \hat{\lambda}_t dt]$, and $\{N_t, t \in T\}$ is a RPP, It is easily shown that

$$E\{I^n[k_n]\} = 0, \; n = 1, 2, \ldots, \tag{3.83}$$

and that the $\{I^n[k_n]\}$ have the orthogonality property

$$E\left\{I^n[k_n]I^m[f_m]\right\}$$

$$= \begin{cases} 0, & n \neq m \\ E\left[\int_{t_0}^{t}\int_{t_0}^{\sigma_n}\cdots\int_{t_0}^{\sigma_2} k_n(\sigma_1,\dots,\sigma_n) f_n(\sigma_1,\dots,\sigma_n)\,d\sigma_1\cdots d\sigma_n\right], & n = m \end{cases} \tag{3.84}$$

where f_m is a function of the same type as k_n. Eq. (3.84) is a consequence of (2.26) and the fact that $E\{(dr_t)^2\} \sim dt$; i.e.,

$$E\left\{\int_{t_0}^{t} f_\sigma\, dr_\sigma \int_{t_0}^{t} g_\sigma\, dr_\sigma\right\} = E\left\{\int_{t_0}^{t} f_\sigma g_\sigma\, d\sigma\right\} \tag{3.85}$$

For the argument which follows we assume without proof that any L_2 scalar functional g of $\{r_t, t \in T\}$ can be expressed as

$$g = \sum_{n=0}^{\infty} I^n[k_n] \tag{3.86}$$

by a suitable choice of the k_n. Eq. (3.86) is proved rigorously by Ito[40] for the case of $\{r_t, t \in T\}$ a Wiener process; the author is not aware of a general proof, however. One can argue plausibly that (3.86) is probably true. Except for sample function continuity, the second-order properties of $\{r_t, t \in T\}$ are the same as those of a Wiener process; in particular, the rules for manipulating stochastic integrals are the same. It is not unreasonable to assume, then, that (3.86) holds (in a m.s. sense), because of the close connection between $\{r_t, t \in T\}$ and Wiener processes.

Based on (3.86), the following result can be proven. With the hypotheses of Theorem 3.2, modified appropriately for scalar observations, a version of $\{\hat{z}_t = E(z|B_{N_t}), t \in T\}$ is given by

$$\hat{z}_t = \hat{z}_{t_0} + \sum_{n=1}^{\infty} I^n[k_n] \tag{3.87}$$

where the non-random function k_n is defined as

$$k_n(\sigma_1, \dots, \sigma_n) = \frac{E\{z(\lambda_{\sigma_1} - \hat{\lambda}_{\sigma_1}) \cdots (\lambda_{\sigma_n} - \hat{\lambda}_{\sigma_n})\}}{E\{\hat{\lambda}_{\sigma_1}^{1/2} \cdots \hat{\lambda}_{\sigma_n}^{1/2}\}} \tag{3.88}$$

We have restricted ourselves to scalar observations for notational convenience only; there is no difficulty in extending this result to vector RPP's. We omit the proof of (3.87), as it is essentially identical to Frost's proof of a similar result for the Gaussian estimation problem[16].

This series, Eq. (3.87), has an interesting interpretation. Write out the first few terms,

$$\hat{z}_t = \hat{z}_{t_0} + \int_{t_0}^{t} \frac{E\{z(\lambda_\sigma - \hat{\lambda}_\sigma)\}}{E\{\hat{\lambda}_\sigma^{1/2}\}} d\nu_\sigma + \int_{t_0}^{t}\int_{t_0}^{\sigma_2} \frac{E\{z(\lambda_{\sigma_1} - \hat{\lambda}_{\sigma_1})(\lambda_{\sigma_2} - \hat{\lambda}_{\sigma_2})\}}{E\{\hat{\lambda}_{\sigma_1}^{1/2}\hat{\lambda}_{\sigma_2}^{1/2}\}} d\nu_{\sigma_1} d\nu_{\sigma_2} + \cdots$$

and observe that z_t is the sum of an a priori estimate, an estimate linear in $\{r_t, t \in T\}$, an estimate quadratic in

$\{r_t, t \in T\}$, and so forth. Furthermore, each term in the series is orthogonal to all the other terms. We emphasize that each term in (3.87) is a functional of $\{r_t, t \in T\}$ and not $\{N_t, t \in T\}$; thus, the usefulness of (3.87) depends on the ease with which $\hat{\lambda}_t$, and thus r_t, can be obtained. This is the primary limitation of (3.87). In applications one would always truncate the series at some point, so there is the additional question of the relative importance of truncation errors.

3. Linear Estimates.

Clearly a more desirable series version of z_t would involve terms which are linear, quadratic, cubic, etc., functionals of the actual observations $\{N_\sigma, \sigma \in [t_o, t]\}$. Obtaining the general term in the series appears to be an extremely tedious exercise in stochastic calculus; therefore we confine our attention to the linear term, which is probably of the greatest utility for approximations.

The following theorem specifies the linear MMSE estimate of the r.v. z relative to the DSPP $\{N_\sigma, \sigma \in [t_o, t]\}$.

Theorem 3.6.

Let z be a vector random variable with components in L_2, and let λ_t be a function of z. Define the matrices

$$\Lambda_t = \operatorname{diag}(\lambda_t) \text{ and } K_\lambda(t,\tau) = E\{[\lambda_t - E(\lambda_t)][\lambda_\tau - E(\lambda_\tau)]'\}.$$

Then the MMSE estimate of z based on linear functionals of $\{N_\sigma, \sigma \in [t_o, t]\}$ is given by

$$\hat{z}_t = \hat{z}_{t_o} - \int_{t_o}^{t} L_\sigma E(\lambda_\sigma) d\sigma + \int_{t_o}^{t} L_\sigma dN_\sigma \tag{3.89}$$

where the non-random matrix L_σ satisfies the linear (Fredholm) integral equation of the second kind,

$$L_\sigma E(\Lambda_\sigma) + \int_{t_o}^{t} L_\tau K_\lambda(\sigma,\tau) d\tau = E\{z[\lambda_\sigma - E(\lambda_\sigma)]'\} \tag{3.90}$$

Proof.

We prove this theorem using the linear orthogonality test. The error $(z - \hat{z}_t)$ is orthogonal to any linear functional of the data $\{N_\sigma, \sigma \in [t_o, t]\}$ iff $\hat{z}_t$ is the linear MMSE estimate of z. $\hat{z}_t$ must satisfy

$$E\{(z - \hat{z}_t)' \int_{t_o}^{t} W_\sigma dN_\sigma\} = 0 \tag{3.91}$$

for an arbitrary non-random matrix W_σ with components in $L_2[T]$. Adding on both sides of (3.91) the non-random term $-\int_{t_o}^{t} W_\sigma E(\lambda_\sigma) d\sigma$, does not affect the validity of the test. Thus, (3.91) is equivalent to

$$E\{(z - \hat{z}_t)' \int_{t_o}^{t} W_\sigma d\nu_\sigma\} = 0 \tag{3.92}$$

where

$$d\nu_\sigma = dN_\sigma - E(\lambda_\sigma) d\sigma. \tag{3.93}$$

Applying the test, we have

$$E\{(z - E(z) - \int_{t_0}^{t} L_\sigma dv_\sigma)' \int_{t_0}^{t} W_\sigma dv_\sigma\}$$

$$= E\{z'\int_{t_0}^{t} W_\sigma dv_\sigma\} - E(z)'E\{\int_{t_0}^{t} W_\sigma dv_\sigma\} - E\{(\int_{t_0}^{t} L_\sigma dv_\sigma)'(\int_{t_0}^{t} W_\sigma dv_\sigma)\}$$

Although $\{v_\sigma, \sigma \in [t_0, t]\}$ is not a martingale, it is easily seen that $\int_{t_0}^{t} W_\sigma dv_\sigma$ has mean zero, eliminating the second term above. The other terms can be evaluated by defining $dn_t = dN_t - \lambda_t dt$, which is a martingale given B_{λ_t}. Using the properties of the trace operator to keep W_σ on the left, we have

$$E\{z'\int_{t_0}^{t} W_\sigma dv_\sigma\} = E\{z'\int_{t_0}^{t} W_\sigma [dn_\sigma + (\lambda_\sigma - E(\lambda_\sigma))d\sigma]\}$$

$$= E\{z'E[\int_{t_0}^{t} W_\sigma dn_\sigma | B_{\lambda_t}]\} + \int_{t_0}^{t} Tr\{W_\sigma' E[z(\lambda_\sigma - E(\lambda_\sigma))']\} d\sigma$$

$$= \int_{t_0}^{t} Tr\{W_\sigma' E[z(\lambda_\sigma - E(\lambda_\sigma))']\} d\sigma$$

Also,

$$E\{(\int_{t_0}^{t} L_\sigma dv_\sigma)'(\int_{t_0}^{t} W_\sigma dv_\sigma)\}$$

$$= E\{(\int_{t_0}^{t} L_\sigma [dn_\sigma + (\lambda_\sigma - E(\lambda_\sigma))d\sigma])'(\int_{t_0}^{t} W_\sigma [dn_\sigma + (\lambda_\sigma - E(\lambda_\sigma))d\sigma])\}$$

$$= E\{(\int_{t_0}^{t} L_\sigma dn_\sigma)'(\int_{t_0}^{t} W_\sigma dn_\sigma)\} + \int_{t_0}^{t}\int_{t_0}^{t} Tr\{W_\sigma' L_\tau K_\lambda(\sigma, \tau)\} d\sigma d\tau$$

By conditioning on B_{λ_t}, the cross-terms are easily seen to have zero mean. The first term on the r.h.s., above, is evaluated as

$$E\left\{\left(\int_{t_0}^{t} L_\sigma dn_\sigma\right)'\left(\int_{t_0}^{t} W_\sigma dn_\sigma\right)\right\} = \int_{t_0}^{t} Tr\left\{W_\sigma' L_\sigma E(\Lambda_\sigma)\right\} d\sigma$$

(see Appendix C, Eq. (C.11)).

Putting these results together, we have

$$\int_{t_0}^{t} Tr\left\{W_\sigma'\left[E\{\mathfrak{z}(\lambda_\sigma - E(\lambda_\sigma))'\} - L_\sigma E(\Lambda_\sigma) - \int_{t_0}^{t} L_\tau K_\lambda(\sigma,\tau)d\tau\right]\right\} d\sigma = 0$$

for arbitrary W_σ. For this to be satisfied, L_σ must be a solution of (3.90), proving the theorem. ∎

We remark that (3.89) is notably different from (3.11), the linear estimate relative to $\{\nu_t,\ t \in T,\ d\nu_t = dN_t - \hat{\lambda}_t dt\}$. Given only the actual observations $\{N_\sigma,\ \sigma \in [t_0,\ t]\}$, we must solve the integral equation (3.90) to obtain the estimate. On the other hand, given the equivalent martingale $\{\nu_\sigma,\ \sigma \in [t_0,\ t]\}$ we are led to the simpler form, (3.11). To take advantage of this simplicity, however, we must somehow generate the process $\{\nu_\sigma,\ \sigma \in [t_0,\ t]\}$. Theorem 3.6 avoids that problem, at the expense of introducing a new integral equation.

With only slight changes, the proof above will verify that Theorem 3.6 also holds for regular point processes with conditionally independent components.

4. Summary of Approximation Methods.

Of the three different approximation techniques we have discussed, no single approach stands out as being the most desirable under all circumstances. We summarize below some of the advantages and disadvantages of each method.

1) Taylor expansions

Advantages: simple to implement; well-suited to digital computation; mean-square error is readily available (P_t); $\hat{\lambda}_t$ is not required.

Disadvantages: approximations are ad hoc; filter divergence is possible; filter structure is not fixed, in general; process $\{x_t, t \in T\}$ to be estimated must be a diffusion; $\lambda_t(x_t)$ must be a memoryless function; g_t must be non-random; $\{N_t, t \in T\}$ must be a DSPP.

2. Iterated Integrals

Advantages: intuitively satisfying; given $\hat{\lambda}_t$, filter is completely specified; any L_2 process can be estimated; $\{N_t, t \in T\}$ can be any RPP.

Disadvantages: requires knowledge of $\hat{\lambda}_t$; truncation error is unknown; implementation is complex, in general.

3. Linear Estimators

Advantages: implementation completely determined by solution to a non-random integral equation; filter structure depends only on second-order statistics of $\{\lambda_t, t \in T\}$; filter is linear in $\{N_t, t \in T\}$; any suboptimality is due only to linearity restriction; $\hat{\lambda}_t$ is not required; $\{N_t, t \in T\}$ can be any RPP.

Disadvantages: inadequate for highly non-linear problems; requires solution of an integral equation.

Some of the weaknesses of the Taylor expansion approach are easily remedied. With appropriate modifications, we can allow $\{x_t, t \in T\}$, the process to be estimated, to be a more general Markov process. g_t can be a memoryless function of x_t, and $\{N_t, t \in T\}$ can be a RPP.

Although it is intuitively pleasing, the iterated integral method is strongly compromised by the fact that $\hat{\lambda}_t$ is required. This is a fundamental difficulty, the resolution of which determines the usefulness of the method. The only serious disadvantage of the linear estimators is their inherent linearity. The numerous weaknesses of the Taylor approximations are tolerable only when a nonlinear estimator would be expected to perform substantially better than a linear estimator. As many communication schemes, for example, are highly nonlinear, it is not surprising that the Taylor expansion approach is widely used.

In Chapter IV we consider some estimation problems which can be solved quite satisfactorily by means of the Taylor series method. We also examine a problem for which a linear estimator is better suited than an approximate nonlinear estimator.

G. Mean-Square Error.

In this section we discuss various means of bounding the mean-square error of estimators. In addition to well-known bounds of the Cramer-Rao type, we also consider differential equations for the error, integral forms derived from the representation theorem (Th. 3.2), and a form related to mutual information.

1. Cramer-Rao-Type Bounds.

Let z_t^* be any estimate of the vector random variable z, based on observations of the DSPP $\{N_\sigma, \sigma \in [t_o, t]\}$. Then, subject to certain regularity conditions,

$$E\{(z - z_t^*)\ (z - z_t^*)'\} \geq J_t^{-1}, \tag{3.94}$$

where J_t is the Fisher information matrix

$$E\{\nabla_z[\ln p(z|B_{N_t})]\nabla_z[\ln p(z|B_{N_t})]'\}.$$

∇_z is the gradient operator with respect to z. Snyder has shown[23] that the general element J_{ij} of J_t is given by

$$J_{ij} = -E\left\{\frac{\partial^2 \ln p(z)}{\partial z_i\, \partial z_j}\right\} + E\left\{\int_{t_o}^{t} \frac{\partial^2 \lambda_\sigma(z)}{\partial z_i\, \partial z_j} d\sigma\right\}$$

$$- E\left\{\int_{t_o}^{t} \lambda_\sigma(z) \frac{\partial^2 \ln \lambda_\sigma(z)}{\partial z_i\, \partial z_j} d\sigma\right\}$$

This can also be written,

$$J_{ij} = - E\left\{ \frac{\partial^2 \ln p(z)}{\partial z_i \, \partial z_j} \right\} \tag{3.95}$$

$$+ E\left\{ \int_{t_0}^{t} \lambda_\sigma(z) \frac{\partial \ln \lambda_\sigma(z)}{\partial z_i} \frac{\partial \ln \lambda_\sigma(z)}{\partial z_j} d\sigma \right\}$$

As long as z is a random variable, it can be seen that an efficient estimate does not exist, because the posterior p.d.f. of z is not Gaussian (see Van Trees[33], pp. 72-73). If z is non-random, it can be shown that efficient estimates exist (see Appendix E for an example).

Except for certain parameter estimation problems, bounds such as (3.94) are of limited use, though they are often relatively easy to derive. Unfortunately, the determination of general bounds for process estimation is not so simple. We can bound the error of interval estimates† by extending (3.94) to countably infinite dimensional random vectors, via an "information kernel" and a limiting argument. This is a relatively straightforward generalization[33], but the end result is not especially enlightening for our purposes.

Recently, however, Snyder[44] has gone a step further and obtained an information kernel type bound valid for causal as well as non-causal estimates. This appears to be of potential value for certain forms of $\{\lambda_t, t \in T\}$; as it is not yet

†See Appendix E for an example of an interval estimate.

generally available, we reproduce the derivation here.

Let $\{z_\sigma, \sigma \in [t_0, t]\}$ be a zero-mean, scalar, Gaussian r.p. which we wish to estimate based on observations of $\{N_\sigma, \sigma \in [t_0, t]\}$. Consider the approximation

$$z_\sigma^K = \sum_{k=1}^{K} z_k \psi_k(\sigma), \quad \sigma \in [t_0, t] \tag{3.96}$$

to $\{z_\sigma, \sigma \in [t_0, t]\}$ suggested by the Karhunen-Loeve representation $z_\sigma = \text{l.i.m. } z_\sigma^K$, $\sigma \in [t_0, t]$, where the $\{z_k\}$ are independent Gaussian r.v.'s and the $\{\psi_k\}$ are the eigenfunctions of the covariance kernel K_z of $\{z_\sigma, \sigma \in [t_0, t]\}$. z_k has mean zero and variance μ_k, where μ_k is the eigenvalue corresponding to the eigenfunction ψ_k.

Now let $\{y_\sigma, \sigma \in [t_0, t]\}$ denote the scalar observation record†† available for causally estimating the $\{z_k\}$ at time t. Let $\ell_t(\{z_k\}) = \ln p(\{z_k\} | B_{y_t})$ where $p(\{z_k\} | B_{y_t})$ is the posterior joint p.d.f. of the $\{z_k\}$, and define $R_d(t, \tau) = E\{d_t d_\tau\}$, where $\{d_\sigma, \sigma \in [t_0, t]\}$ is a "derivative signal" of the form $\partial s_\sigma / \partial z_\sigma$, $\sigma \in [t_0, t]$, with s_σ a known memoryless function of z_σ. If $J_{ki}^K = E\{\partial^2 \ell_t / \partial z_k \partial z_i\}$ has the form

---

††We do not yet restrict $\{y_\sigma, \sigma \in [t_0, t]\}$ to be the record of a counting process.

$$J_{ki}^{K} = \mu_i^{-1}\delta_{ki} + \alpha \int_{t_o}^{t} \varphi_k(v)\varphi_i(v) R_d^K(v,v)dv \qquad (3.97)$$

with α some constant and $R_d^K(v, v) = E\{d_v^2(z_v^K)\}$, then the m.s. error in causally estimating $\{z_\sigma, \sigma \in [t_o, t]\}$ with z_t^* satisfies

$$E\{(z_t - z_t^*)^2\} \geq \gamma(t, t), \qquad (3.98)$$

with equality iff s_t is linear in z_t and the $\{z_k\}$ can be estimated efficiently. $\gamma(t, \tau)$ is the solution of the integral equation

$$\gamma(t,\tau) + \alpha \int_{t_o}^{t} \gamma(t,v) R_d(v,v) K_z(v,\tau) dv = K_z(t,\tau), \qquad (3.99)$$
$$\tau \in [t_o, t]$$

This result is derived as follows. Let

$$\tilde{z}_t^K = \sum_{k=1}^{K} \tilde{z}_k(t)\varphi_k(t),$$

where $\tilde{z}_k(t) = z_k - z_k^*(t)$. Then

$$E\{(\tilde{z}_t^K)^2\} = \sum_{j=1}^{K}\sum_{k=1}^{K} E\{\tilde{z}_k(t)\tilde{z}_j(t)\}\varphi_j(t)\varphi_k(t)$$

$$\geq \sum_{j=1}^{K}\sum_{k=1}^{K} \gamma_{jk}^{K}\varphi_j(t)\varphi_k(t) \triangleq \gamma^K(t,t),$$

where $[\gamma_{jk}^{K}]$ is the inverse of the information matrix $[J_{jk}^{K}]$. The inequality is the Cramer-Rao inequality.

Define

$$\gamma^K(t,v) = \sum_{j=1}^{K}\sum_{k=1}^{K} \gamma_{jk}^K \varphi_j(t)\varphi_k(v), \quad v \in [t_0, t].$$

Multiply this by $\alpha R_d^K(v,v)\mu_i \varphi_i(v)$, integrate over $[t_0, t]$ and use (3.97) to obtain

$$\alpha \int_{t_0}^{t} \gamma^K(t,v) R_d^K(v,v)\mu_i \varphi_i(v) dv = \sum_{j=1}^{K}\sum_{k=1}^{K} \gamma_{jk}^K \varphi_j(t)[\mu_i J_{ki}^K - \delta_{ki}].$$

Since $\sum_{k=1}^{K} \gamma_{jk}^K J_{ki}^K = \delta_{ji}$, this becomes

$$\alpha \int_{t_0}^{t} \gamma^K(t,v) R_d^K(v,v)\mu_i \varphi_i(v) dv + \sum_{j=1}^{K} \gamma_{ji}^K \varphi_j(t) = \mu_i \varphi_i(t).$$

Now multiply both sides by $\varphi_i(u)$, $u \in [t_0, t]$, and sum on i from 1 to K. Then use the definition of $\gamma^K(t,u)$ and Mercer's theorem, and let $K \to \infty$ to obtain the desired result.

For $y_\sigma = N_\sigma$, $\sigma \in [t_0, t]$, a comparison of (3.95) and (3.97) makes it clear that J_{ki}^K has the form required for (3.98) (with $\alpha = 4\beta$), if $\lambda_t = \beta s_t^2$. For $y_\sigma = s_\sigma + n_\sigma$, $\sigma \in [t_0, t]$, with $\{n_\sigma, \sigma \in [t_0, t]\}$ a white Gaussian process of spectral height N_0, J_{ki}^K again has the form required for (3.98) (with $\alpha = 1/N_0$). Thus the limiting performance for the Poisson intensity $\lambda_t = \beta s_t^2$ is the same as that for observing s_t in additive white Gaussian noise with spectral density $N_0 = 1/4\beta$.

2. Differential and Integral Forms.

We have seen in previous sections that exact as well as various approximate stochastic differential equations can be derived for the conditional covariance matrix of the error in MMSE estimates. It is obvious that the error covariance itself can be obtained by averaging the conditional covariance over the observations and solving the resulting non-random differential equation. In most cases the latter operation will have to be performed numerically.

For reference, we enumerate some of the covariance equations. Let $\bar{P}_t = E\{P_t\}$, and assume that $\{x_t,\ t \in T\}$ is a diffusion. To obtain the following formulas, we have used $d\nu_t = dN_t - \hat{\lambda}_t dt$ and the zero-expectation property of stochastic integrals (2.25). In the approximate formulas, we have also used the appropriate Taylor approximation for $\hat{\Lambda}_t$.

From (3.60) (exact):

$$\dot{\bar{P}}_t = E\Big[f_t \tilde{x}_t' + \tilde{x}_t f_t' + g_t g_t'$$

$$- E\{x_t(\lambda_t - \hat{\lambda}_t)' | B_{\nu_t}\} \hat{\Lambda}_t^{-1} E\{(\lambda_t - \hat{\lambda}_t) x_t' | B_{\nu_t}\}\Big] \tag{3.100}$$

From (3.74) (first order approximation):

$$\dot{\bar{P}}_t = E\Big[D_{\hat{x}_t}[f_t(\hat{x}_t)]' P_t + P_t D_{\hat{x}_t}[f_t(\hat{x}_t)] + g_t g_t'$$

$$- P_t D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)] \Lambda_t(\hat{x}_t) D_{\hat{x}_t}[\ln \lambda_t(\hat{x}_t)]' P_t\Big] \tag{3.101}$$

From (3.76) (second order approximation):

$$\dot{\overline{P}}_t = E\Big[D_{\hat{x}_t}[f_t(\hat{x}_t)]'P_t + P_t D_{\hat{x}_t}[f_t(\hat{x}_t)] + g_t g_t'$$

$$- P_t D_{\hat{x}_t}[\ln\lambda_t(\hat{x}_t)]\Lambda_t(\hat{x}_t)D_{\hat{x}_t}[\ln\lambda_t(\hat{x}_t)]'P_t\Big] \qquad (3.102)$$

Note that this is identical to (3.101).

From (3.79) (Gaussian second order approximation):

$$\dot{\overline{P}}_t = E\Big[D_{\hat{x}_t}[f_t(\hat{x}_t)]'P_t + P_t D_{\hat{x}_t}[f_t(\hat{x}_t)] + g_t g_t'$$

$$- P_t H_{\hat{x}_t}[\|\lambda_t(\hat{x}_t)\|_1]P_t + P_t H_{\hat{x}_t}[\ln\lambda_t(\hat{x}_t)'\lambda_t(\hat{x}_t)]P_t\Big] \quad (3.103)$$

In general, these differential equations resist efforts to study analytically the behavior of the m.s. error. The following theorem provides an alternative representation for $\overline{P}_t$, which is more general and in many ways more revealing.

Theorem 3.7.

Let z be a vector random variable with components in L_2, and define the matrices $\Lambda_t = \text{diag}(\lambda_t)$ and $K_\lambda(t, \tau) = E\{[\lambda_t - E(\lambda_t)][\lambda_\tau - E(\lambda_\tau)]'\}$. Then any estimate z_t^* based on the DSPP $\{N_\sigma, \sigma \in [t_0, t]\}$ is bounded in mean-square error as

$$E\{(z - z_t^*)(z - z_t^*)'\} \geq \text{Cov}(z)$$

$$- \int_{t_0}^{t} E\Big\{E[z(\lambda_\sigma - \hat{\lambda}_\sigma)'|B_{N_\sigma}]\hat{\Lambda}_\sigma^{-1}E[(\lambda_\sigma - \hat{\lambda}_\sigma)z'|B_{N_\sigma}]\Big\}d\sigma \quad (3.104)$$

Furthermore, (3.104) is an equality if z_t^* is the MMSE estimate of z.

Corollary 3.7.1.

Any estimate z_t^* based on $\{N_\sigma, \sigma \in [t_o, t]\}$ is bounded in m.s. error as

$$E\{(z - z_t^*)(z - z_t^*)'\} \geqslant \mathrm{Cov}(z) - \int_{t_o}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma)'\hat{\Lambda}_\sigma^{-1}(\lambda_\sigma - \hat{\lambda}_\sigma)z'\}\,d\sigma \qquad (3.105)$$

Corollary 3.7.2.

Any estimate z_t^* based on linear functionals of $\{N_\sigma, \sigma \in [t_o, t]\}$ is bounded in m.s. error as

$$E\{(z - z_t^*)(z - z_t^*)'\} \geqslant \mathrm{Cov}(z) - \int_{t_o}^{t} L_\sigma E\{[\lambda_\sigma - E(\lambda_\sigma)]z'\}\,d\sigma \qquad (3.106)$$

with equality when z_t^* is the MMSE linear estimate of z. L_σ satisfies the integral equation,

$$L_\sigma E(\Lambda_\sigma) + \int_{t_o}^{t} L_\tau K_\lambda(\sigma,\tau)\,d\tau = E\{z[\lambda_\sigma - E(\lambda_\sigma)]'\}, \qquad (3.107)$$

as in Theorem 3.6.

Proof.

The proof of the theorem and corollaries is simple. Note that $E\{(z - z_t^*)(z - z_t^*)'\}$ can be written as

$$E\{([z_t^* - E(z)] - [z - E(z)])([z_t^* - E(z)] - [z - E(z)])'\}$$

$$= \mathrm{Cov}(z) - E\{[z_t^* - E(z)]\,[z_t^* - E(z)]'\} \qquad (3.108)$$

The last term above can be evaluated from the result of Theorem 3.2 and (2.26), as $z_t^* - E(z)$ is simply $\int_{t_0}^{t} E\{z(\lambda_\sigma - \hat{\lambda}_\sigma)' | B_{N_\sigma}\}\hat{\Lambda}_\sigma^{-1} d\nu_\sigma$ when $z_t^* = \hat{z}_t = E(z|B_{N_t})$.

The first corollary is a direct consequence of Jensen's inequality for conditional expectations[1].

The second corollary follows by noting (see Th. 3.6) that $z_t^* - E(z) = \int_{t_0}^{t} L_\sigma dv_\sigma$ when z_t^* is the linear MMSE estimate, and by using $dv_\sigma = dn_\sigma + [\lambda_\sigma - E(\lambda_\sigma)]\,d\sigma$ and (2.26). ∎

This theorem, which can be shown to hold for RPP's with conditionally independent components, expresses the error covariance as the initial (zero data) error, Cov(z), minus a term, growing in time, which reflects the information imparted by the observations. This evolving "correction term" is clearly bounded above by Cov(z). This form for the error clearly has some conceptual advantages over both Cramer-Rao type bounds and the differential equations (3.100) - (3.103). This gain is not without a price, however, as the quantities in the integrals, although total expectations, contain estimates as well as the variable z. Nonetheless, the integral forms will probably respond best to bounding efforts, owing to their simplicity.

Since Theorem 3.7 is valid for any L_2 random variable z, we can let z be a r.v. of a diffusion process $\{x_t, t \in T\}$.

Then (3.104) can be viewed as a solution of (3.100), just as (3.21) can be thought of as a solution of (3.1).

Beyond Corollary 3.7.1, the results of the theorem immediately suggest a number of approximations that can be made. The most obvious is applying the Taylor series approximations of the previous section. It would also be worthwhile to bound the difference of the linear error and the nonlinear error, in order to evaluate the efficacy of the linear estimate as an approximation.

3. Mutual Information.

Using Theorem 3.5 we can obtain an interesting relation for the m.s. error of the MMSE estimate $\{\hat{\lambda}_\sigma, \sigma \in [t_0, t]\}$ in terms of the mutual information $I_t(N, z)$ between the scalar DSPP $\{N_\sigma, \sigma \in [t_0, t]\}$ and some process $\{z_\sigma, \sigma \in [t_0, t]\}$ which satisfies $B_{z_t} = B_{\lambda_t}$, $t \in T$. The mutual information[38] is defined as

$$I_t(N, z) \triangleq E\left\{\ln \frac{p(\{t_i\}; N_t \mid B_{\lambda t})}{p(\{t_i\}; N_t)}\right\} \tag{3.109}$$

where the numerator and denominator are given by Eqs. (2.5) and (3.49). Substituting these into (3.109) yields

$$I_t(N, z) = E\left\{-\int_{t_0}^{t} (\lambda_\sigma - \hat{\lambda}_\sigma) d\sigma + \int_{t_0}^{t} \ln \frac{\lambda_\sigma}{\hat{\lambda}_\sigma} dN_\sigma\right\} \tag{3.110}$$

The first integral clearly has zero mean, so we are left with

$$I_t(N, z) = E\left\{ \int_{t_0}^{t} \ln \frac{\lambda_\sigma}{\hat{\lambda}_\sigma} dN_\sigma \right\}$$

$$= E\left\{ \int_{t_0}^{t} \ln \lambda_\sigma dN_\sigma - \int_{t_0}^{t} \ln \hat{\lambda}_\sigma dN_\sigma \right\}$$

Define $d\nu_t = dN_t - \hat{\lambda}_t dt$ and $dn_t = dN_t - \lambda_t dt$;
then

$$I_t(N, z) = E\left\{ \int_{t_0}^{t} \ln \lambda_\sigma dn_\sigma + \int_{t_0}^{t} \lambda_\sigma \ln \lambda_\sigma d\sigma \right.$$

$$\left. - \int_{t_0}^{t} \ln \hat{\lambda}_\sigma d\nu_\sigma - \int_{t_0}^{t} \hat{\lambda}_\sigma \ln \hat{\lambda}_\sigma d\sigma \right\} \qquad (3.111)$$

The two stochastic integrals have zero expectation by virtue of the measurability of their integrands and Eq. (2.25); thus,

$$I_t(N, z) = E\left\{ \int_{t_0}^{t} \lambda_\sigma \ln \lambda_\sigma d\sigma - \int_{t_0}^{t} \hat{\lambda}_\sigma \ln \hat{\lambda}_\sigma d\sigma \right\} \qquad (3.112)$$

Noting that $\ln x \leq x - 1$, $x > 0$, we can approximate $I_t(N, z)$ as

$$I_t(N, z) \cong E\left\{ \int_{t_0}^{t} \widehat{\lambda^2_\sigma} d\sigma - \int_{t_0}^{t} \hat{\lambda}^2_\sigma d\sigma \right\}$$

or, what is the same,

$$I_t(N, z) \cong E\left\{ \int_{t_0}^{t} (\lambda_\sigma - \hat{\lambda}_\sigma)^2 d\sigma \right\}. \qquad (3.113)$$

This is the desired relation. It can be shown to hold for $\{N_t, t \in T\}$ a RPP, also. (3.113) can be sharpened somewhat by specifying more precisely the form of $\{\lambda_t, t \in T\}$. Using the binary detection example, Evans[15] has obtained such an expression by a formal manipulation of (3.45).

H. The Distribution of Estimator Errors

In our discussions of estimator performance, we have dealt solely with the mean-square error measure, which is only a second moment description of the perhaps complex statistical behavior of the error process, $\tilde{x}_t = x_t - \hat{x}_t$. For some fixed $t \in T$, the probability distribution of the r.v. $\tilde{x}_t$ would be at least as revealing as the m.s. error. The following theorem provides a partial differential equation for the time evolution of the first-order error p.d.f., assuming that the processes $\{x_t,\ t \in T\}$ and $\{\hat{x}_t,\ t \in T\}$ possess a particular structure (motivated by the Taylor approximations of III.F.1).

Theorem 3.8.

Let $\{N_t,\ t \in T\}$ be a vector DSPP with vector rate process $\{\lambda_t,\ t \in T\}$ which is non-negative w.p.1 and satisfies $E(\lambda_t) < \infty$, $t \in T$. Let $\{w_t,\ t \in T\}$ be a standardized Wiener process. Assume that $\{x_t,\ t \in T\}$ and $\{\hat{x}_t,\ t \in T\}$ satisfy the equations,

$$dx_t = f_t(x_t)dt + g_t dw_t$$

$$d\hat{x}_t = q_t(\hat{x}_t)dt + r_t(\hat{x}_t)dN_t$$

where f_t and λ_t are memoryless L_2 functions of x_t, and q_t and

r_t are memoryless L_2 functions of $\hat{x}_t$; g_t is a function of time alone.

Then the p.d.f. of the r.v. $\tilde{x}_t = x_t - \hat{x}_t$ satisfies the partial differential-difference equation,

$$\frac{\partial}{\partial t} p_{\tilde{x}_t}(\tilde{X}) = -\nabla_{\tilde{X}} \cdot \left[E\{f_t - q_t \mid \tilde{x}_t = \tilde{X}\} p_{\tilde{x}_t}(\tilde{X})\right] + \frac{1}{2} \operatorname{Tr}\left[g_t g_t' \nabla_{\tilde{X}\tilde{X}} p_{\tilde{x}_t}(\tilde{X})\right] + \sum_{j=1}^{D} E\left\{\lambda_t'(x_t) e_j \left[p_{\tilde{x}_t}(\tilde{X} + r_t(\hat{x}_t) e_j) - p_{\tilde{x}_t}(\tilde{X})\right] \Big| \tilde{x}_t = \tilde{X}\right\} \tag{3.114}$$

where the $\{e_j\}$ are unit vectors in R^D.

We remark that (3.114) is nothing more than a Fokker-Planck equation if $q_t = r_t = 0$, as $\{x_t, t \in T\}$ is a Markov diffusion process. It is an easy generalization to allow g_t to be a function of x_t, but that is seldom necessary for communications problems, where g_t is usually non-random. It is also a straightforward extension to allow $\{x_t, t \in T\}$ to be a general Markov process, rather than a diffusion.

To prove this theorem, we need the following Lemma, which is usually associated with Markov processes[16,23], but is applicable to a much wider class of processes[64].

<u>Lemma</u>

Let $\{z_t, t \in T\}$ be a vector stochastic process which satisfies the following conditions:

1) $\frac{1}{\Delta t}\left|E\{e^{iv'(z_{t+\Delta t}-z_t)}-1|z_t\}\right| \le g_t(v, z_t)$ a.s.,

where $E\{|g_t(v, z_t)|\} < \infty$;

2) $\Psi_t(v|z_t) \triangleq \lim_{\Delta t \to 0} \frac{1}{\Delta t} E\{e^{iv'(z_{t+\Delta t}-z_t)}-1|z_t\}$

exists as a limit in probability.

Assume that the characteristic function $M_{z_t}(iv) = E(e^{iv'z_t})$ is differentiable in t, $t \in T$. Then $M_{z_t}(iv)$ satisfies

$$\frac{\partial}{\partial t} M_{z_t}(iv) = E\{\Psi_t(v|z_t)e^{iv'z_t}\},$$

with $M_{z_{t_0}}(iv) = E(e^{iv'z_{t_0}})$

Proof.

Clearly,

$$M_{z_{t+\Delta t}}(iv) = E\{e^{iv'z_{t+\Delta t}}\}$$

$$= E\{e^{iv'z_t} e^{iv'(z_{t+\Delta t}-z_t)}\}$$

$$= E\{e^{iv'z_t} E[e^{iv'(z_{t+\Delta t}-z_t)}|z_t]\}$$

Forming the first difference, we have

$$\frac{M_{z_{t+\Delta t}}(iv) - M_{z_t}(iv)}{\Delta t} = E\left\{e^{iv'z_t}\frac{1}{\Delta t}E[e^{iv'(z_{t+\Delta t}-z_t)}-1|z_t]\right\}$$

Taking the limit as $\Delta t \to 0$ we obtain the desired result; the interchange of limit and expectation is valid by virtue of the Lebesgue bounded convergence theorem[1,3] and condition (1) of the lemma. ∎

We now proceed to prove the theorem.

Form z_t of the lemma by adjoining the vectors x_t and $\hat{x}_t$:

$$z_t' = [x_t', \hat{x}_t']$$

Partition the vector v of real numbers into v_1 and v_2 such that

$$v'z_t = v_1'x_t + v_2'\hat{x}_t$$

Let us now compute $\psi_t(v|z_t) = \psi_t(v_1, v_2|x_t, \hat{x}_t)$. This is accomplished by applying the appropriate Ito rules (see Ch.II and Appendix C) to the function $e^{iv(\cdot)}$. Accordingly,

$$e^{iv_1'(x_{t+\Delta t}-x_t)} = 1 + \int_t^{t+\Delta t}[iv_1'f_\sigma - \tfrac{1}{2}v_1'g_\sigma g_\sigma' v_1]e^{iv_1'(x_\sigma - x_t)}d\sigma$$

$$+ iv_1'\int_t^{t+\Delta t} e^{iv_1'(x_\sigma - x_t)}g_\sigma dw_\sigma ;$$

$$e^{iv_2'(\hat{x}_{t+\Delta t}-\hat{x}_t)} = 1 + \int_t^{t+\Delta t} iv_2'q_\sigma e^{iv_2'(\hat{x}_\sigma - \hat{x}_t)}d\sigma$$

$$+ \sum_{j=1}^{D}\int_t^{t+\Delta t}(e^{iv_2'r_\sigma e_j}-1)e^{iv_2'(\hat{x}_\sigma - \hat{x}_t)}dN_\sigma' e_j$$

Defining $dn_t = dN_t - \lambda_t dt$, the latter can be written as

$$e^{iv_2'(\hat{x}_{t+\Delta t}-\hat{x}_t)} = 1+\int_t^{t+\Delta t}[iv_2' q_\sigma + \sum_{j=1}^{D}(e^{iv_2' r_\sigma e_j}-1)\lambda_\sigma' e_j]e^{iv_2'(\hat{x}_\sigma-\hat{x}_t)}d\sigma$$

$$+\sum_{j=1}^{D}\int_t^{t+\Delta t}(e^{iv_2' r_\sigma e_j}-1)e^{iv_2'(\hat{x}_\sigma-\hat{x}_t)}dn_\sigma' e_j \; .$$

Combining these results, we get

$$e^{iv'(z_{t+\Delta t}-z_t)}-1 = \int_t^{t+\Delta t}[iv_1' f_\sigma - \tfrac{1}{2}v_1' g_\sigma g_\sigma' v_1]e^{iv_1'(x_\sigma-x_t)}d\sigma$$

$$+\int_t^{t+\Delta t}[iv_2' q_\sigma + \sum_{j=1}^{D}\lambda_\sigma' e_j(e^{iv_2' r_\sigma e_j}-1)]e^{iv_2'(\hat{x}_\sigma-\hat{x}_t)}d\sigma + iv_1'\int_t^{t+\Delta t}e^{iv_1'(x_\sigma-x_t)}g_\sigma dw_\sigma$$

$$+\sum_{j=1}^{D}\int_t^{t+\Delta t}(e^{iv_2' r_\sigma e_j}-1)e^{iv_2'(\hat{x}_\sigma-\hat{x}_t)}dn_\sigma' e_j \quad + o(\Delta t) \tag{3.115}$$

Now we must evaluate $E\{e^{iv'(z_{t+\Delta t}-z_t)}-1 \mid x_t, \hat{x}_t\}$. There is no contribution from the third and fourth terms in (3.115) because

$$E\{\int_t^{t+\Delta t} e^{iv_1'(x_\sigma-x_t)}g_\sigma dw_\sigma \mid x_t, \hat{x}_t\} = 0$$

and

$$E\left\{\int_t^{t+\Delta t}(e^{iv_2' r_\sigma e_j}-1)e^{iv_2'(\hat{x}_\sigma-\hat{x}_t)}dn_\sigma' e_j \;\middle|\; x_t,\hat{x}_t\right\}$$

$$= E\left\{E\left[\int_t^{t+\Delta t}(e^{iv_2' r_\sigma e_j}-1)e^{iv_2'(\hat{x}_\sigma-\hat{x}_t)}dn_\sigma' e_j \;\middle|\; B_{x_{t+\Delta t}} \otimes B_{\hat{x}_{t+\Delta t}}\right]\middle|\; x_t,\hat{x}_t\right\}$$

$$= 0$$

Thus we are left with

$$E\{e^{iv'(z_{t+\Delta t}-z_t)}-1 \mid x_t,\hat{x}_t\}$$

$$= \int_t^{t+\Delta t} E\{[iv_1' f_\sigma - \tfrac{1}{2}v_1' g_\sigma g_\sigma' v_1]e^{iv_1'(x_\sigma - x_t)} \mid x_t,\hat{x}_t\}d\sigma$$

$$+ \int_t^{t+\Delta t} E\{[iv_2' q_\sigma + \sum_{j=1}^{D}\lambda_\sigma' e_j(e^{iv_2' r_\sigma e_j}-1)]e^{iv_2'(\hat{x}_\sigma-\hat{x}_t)} \mid x_t,\hat{x}_t\}d\sigma$$

$$+ o(\Delta t), \qquad (3.116)$$

where the exchange of integration and expectation can be justified by the boundedness of the integrals in (3.116). We verify this boundedness, and thus condition(1) of the lemma, as follows.

Use the triangle inequality where appropriate, to get

$$\frac{1}{\Delta t}\left|E\{e^{iv'(z_{t+\Delta t}-z_t)}-1|z_t\}\right|$$

$$\leq \frac{1}{\Delta t}E\left\{\int_t^{t+\Delta t}[|v_1'f_\sigma|+\tfrac{1}{2}v_1'g_\sigma g_\sigma'v_1+|v_2'q_\sigma|+2\|\lambda_\sigma\|_1]d\sigma \Big| x_t,\hat{x}_t\right\} + \frac{|o(\Delta t)|}{\Delta t}$$

$$\leq E\left\{\sup_{\substack{\sigma\in[t,t+\Delta t]\\ t,t+\Delta t\in T}}[|v_1'f_\sigma|+\tfrac{1}{2}v_1'g_\sigma g_\sigma'v_1+|v_2'q_\sigma|+2\|\lambda_\sigma\|_1]\Big| x_t,\hat{x}_t\right\} + \frac{|o(\Delta t)|}{\Delta t}$$

Taking the expectation of this, we have

$$E\left\{\sup_{\substack{\sigma\in[t,t+\Delta t]\\ t,t+\Delta t\in T}}[|v_1'f_\sigma|+\tfrac{1}{2}v_1'g_\sigma g_\sigma'v_1+|v_2'q_\sigma|+2\|\lambda_\sigma\|_1]\right\} + \frac{|o(\Delta t)|}{\Delta t} < \infty,$$

as a consequence of the conditions that f, g, q, and λ are assumed to satisfy.

Dividing (3.116) by Δt, and using the bounded convergence theorem[1,3] to take the limit inside the expectation as $\Delta t \to 0$, we get

$$\psi_t(v_1,v_2|x_t,\hat{x}_t) = iv_1'f_t - \tfrac{1}{2}v_1'g_\sigma g_\sigma'v_1 + iv_2'q_t + \sum_{j=1}^{D}\lambda_t'e_j(e^{iv_2'v_\sigma e_j}-1) \tag{3.117}$$

Using the Markov inequality,

$$\Pr\{|\cdot| \geq \epsilon\} \leq \frac{E\{|\cdot|\}}{\epsilon}$$

it is easily verified that (3.117) exists as a limit in probability.

Now according to the lemma,

$$\frac{\partial}{\partial t} M_{x_t,\hat{x}_t}(iv_1, iv_2) = E\{\psi_t(v_1,v_2|x_t,\hat{x}_t)e^{iv_1'x_t + iv_2'\hat{x}_t}\}$$

It is clear from this that the characteristic function of $\tilde{x}_t = x_t - \hat{x}_t$ satisfies

$$\frac{\partial}{\partial t} M_{\tilde{x}_t}(iu) = E\{\psi_t(u, -u\,|\,x_t, \hat{x}_t)e^{iu'\tilde{x}_t}\}$$

$$= E\{[iu'(f_t - g_t) - \tfrac{1}{2}u'g_t g_t' u$$

$$+ \sum_{j=1}^{D} \lambda_t' e_j (e^{-iu' r_t e_j} - 1)] e^{iu'\tilde{x}_t}\}$$

$$= E\{[iu' E(f_t - g_t | \tilde{x}_t) - \tfrac{1}{2}u' g_t g_t' u$$

$$+ \sum_{j=1}^{D} E(\lambda_t' e_j (e^{-iu' r_t e_j} - 1) | \tilde{x}_t)] e^{iu'\tilde{x}_t}\}$$

Inverting this expression we get (3.114). This completes the proof. ∎

The usefulness of this theorem depends for the most part on the ease with which the conditional expectations in (3.114) can be evaluated. If $f_t - q_t$ is proportional to $\tilde{x}_t$, as is often the case, the first expectation presents no problems. The second, however, almost always requires some approximation, unless we are in the degenerate situation of λ_t and r_t being non-random. In general we must solve for the joint p.d.f. of x_t and $\hat{x}_t$ in order to evaluate the conditional expectations. This joint p.d.f., which satisfies

$$\frac{\partial}{\partial t} p_{x_t,\hat{x}_t}(X,\hat{X}) = -\nabla_X \cdot [f_t \, p_{x_t,\hat{x}_t}(X,\hat{X})]$$

$$-\nabla_{\hat{X}} \cdot [q_t \, p_{x_t,\hat{x}_t}(X,\hat{X})] + \tfrac{1}{2} \mathrm{Tr} [g_t g_t' \nabla_{XX} p_{x_t,\hat{x}_t}(X,\hat{X})]$$

$$+ \sum_{j=1}^{D} \lambda_t' e_j [p_{x_t,\hat{x}_t}(X, \hat{X} - r_t e_j) - p_{x_t,\hat{x}_t}(X,\hat{X})], \tag{3.118}$$

is usually difficult to treat analytically, even in simple cases.

If the conditional expectations in (3.114) are specified, it will still seldom be possible to solve directly for $p_{\tilde{x}_t}(\tilde{X})$, owing to the lack of a systematic theory for mixed (differential-difference) equations of this type[63]. However, even when $p_{\tilde{x}_t}$ cannot be obtained directly, the characteristic function $M_{\tilde{x}_t}(iu)$ can sometimes be evaluated.

It should be noted that the structure this theorem presumes for the process $\{\hat{x}_t, t \in T\}$ is quite restrictive, in that q_t and r_t must be memoryless functions of $\hat{x}_t$. This rules out most exact estimators, but allows for approximate estimators which can be updated in time based only on the most recent estimate and causal observations of $\{N_t, t \in T\}$. Fortunately, most of our approximate estimators fall into this class. Indeed, no-memory updating is a feature usually striven for in practical suboptimum systems which are to be implemented. The theorem can be extended to handle exact estimators, but the resulting partial differential equation will be of little use as it will not possess solutions which evolve uniquely from the initial conditions[64].

I. Summary.

In this chapter we have presented the theoretical basis for solving the estimation problem for counting process observations.

First, we proved a theorem which provides a stochastic integral representation for any estimator of a L_2 process. It was then demonstrated that stochastic differential equations for estimates of Ito processes could be derived from the integral representation. This led to a canonical filtering theorem for causal estimation, as well as results for prediction, non-causal filtering, and parameter estimation.

A new statistical characterization of DSPP's, applicable to hypothesis testing problems as well as some estimation problems, was obtained.

Using Taylor series expansions, approximate estimator equations were developed. It was seen that these equations often yield estimators which perform well if the error is small, and are relatively easy to implement. Other approximations were presented, based on estimates relative to linear functionals of the observations, and on estimates relative to linear, quadratic, etc., functionals of the process $\{r_t, t \in T\}$ (Eq. (3.13)).

Finally, the performance of estimators was examined via a number of bounds and equalities. In particular, a partial differential-difference equation for the p.d.f. of the estimator error was obtained. This is an exact equation which applies to many approximate estimators, such as those developed in II.F.1, as well as some exact estimators.

CHAPTER IV

Applications in Optical Communication

A. Introduction.

In the previous chapter, methods were developed for solving the estimation problem for Poisson processes. The problem as formulated in III.B. was largely mathematical; little reference was made to practical problems for which the Poisson model is appropriate. In this chapter we give detailed consideration to some selected problems in optical communication, chosen to be meaningful in themselves as well as illustrative of estimation techniques. We shall be concerned with analog communication through the clear, turbulent atmosphere; we leave to others problems involving scatter channels such as clouds or water, imaging, optical radar, etc. We also ignore obvious applications in biomedical measurements[23], gravity wave research[45], and astronomy.

Specifically, we consider the following problem. Given a continuously modulated optical signal, corrupted by multiplicative atmosphere fading as well as additive background noise, what are "good" receivers for demodulating this signal? As stated, this question is too general. We narrow its scope by imposing certain reasonable restrictions on the modulation and the receiver; briefly, we rule out spatial modulation at the transmitter, and heterodyning* at the receiver. Further, we specify the energy-receptive elements at the receiver to be ideal quantum photodetectors.

---

*In most cases heterodyne receivers can be treated via the techniques of classical Gaussian detection and estimation theory[46].

B. Channel Model.

Random variations in the refractive index of the clear atmosphere, resulting from turbulent mixing of thermal layers and wind, cause optical signals to fade in space and time (but not in frequency)[49]. If we assume that a transmitted signal at the input of the channel is a linearly polarized plane wave, that a noise with statistically independent polarization components is added at the output, and that depolarization effects due to turbulence are negligible[50], then the scalar channel model described by the expression*

$$y_{\pi} = z_{\pi} s_t + n_{\pi} \tag{4.1}$$

is appropriate[49]. Here, y_{π}, z_{π}, s_t, and n_{π} are complex envelopes.

The fading factor z_{π}, which accounts for such miscellaneous "constants" as antenna patterns, is assumed to be stationary in time and space. The vast majority of

*Throughout this chapter, except where confusion would result, we follow the convention of denoting time (t) dependence by the subscript t, and time-and-space (t, r) dependence by the subscript π. r defines the position coordinate in the receiving aperture.

experimental evidence suggests that z_{π} is a complex lognormal process[51,52]; moreover, theoretical arguments support this, at least for short path lengths and weak fading. It has been shown that the p.d.f. of a real lognormal variate can in some instances be approximated quite closely by a non-central chi-squared p.d.f.[34,53]. The latter is the distribution of the square of the Euclidean norm of a non-zero-mean Gaussian random vector. This fact, and the analytical difficulties presented by the lognormal process, suggest that Gaussian as well as lognormal fading be considered.

The complex noise n_{π} accounts for background light, but not receiver shot noise, and is taken to be a zero-mean Gaussian process, with independent real and imaginary parts, stationary in time and space. It is further assumed that n_{π} is essentially "white" in space, and that[49]

$$E\{ n_{\pi_1} n_{\pi_2} \} = 0, \qquad \forall \pi_1, \pi_2 \tag{4.2}$$

This completes the general formulation of our channel model; as we proceed, additional simplifying assumptions will be made to facilitate the solution of particular problems.

Before pursuing the subject of detector models, we mention briefly some of the methods of representing the scalar field at the receiving aperture[46,51]. Two of the most useful are the spatial sampling expansion and the orthogonal series expansion. The latter is known as the plane-wave decomposition, or wave-number sampling, when complex spatial sinusoids are chosen for the basis. For finite apertures the sampling expansion, which can also be viewed as an orthogonal expansion, is only approximate, because the received field cannot be wave-number-limited. Finite sample size also limits accuracy. However, the spatial field samples are relatively easy to characterize statistically[46,51,96], so this physically satisfying representation finds great utility. Indeed, it appears to be the only tractable representation for lognormal fields.

The orthogonal series expansion can be made exact for finite apertures, but it sacrifices some mathematical convenience. By choosing the basis to satisfy the Karhunen-Loéve integral equation, the coefficients are guaranteed to be uncorrelated, but not necessarily independent (unless the field is Gaussian). When the field is non-Gaussian, statistical characterization of the expansion coefficients is a difficult problem[46,51]. The plane-wave decomposition is exact for finite apertures, but the coefficients are uncorrelated only in the limit of very large apertures.

In this chapter, we postulate that the receiver gathers energy by distributing a number of ideal photodetectors over the receiving aperture. For sufficiently small detectors, (i.e., detectors over which the field is spatially coherent) the spatial sampling expansion is the logical field representation for this arrangement.

C. Detector Model.

The output of an idealized quantum photodetector is usually taken to be an inhomogeneous Poisson process, conditioned on the received field incident upon the detector surface[47,48,54-62]. That is, the detector output is a doubly-stochastic Poisson process with intensity process.

$$\lambda_t = \frac{\eta}{h f_0} \int_{A_d} |a_r|^2 dr \, , \quad t \in T \tag{4.3}$$

where a_r is the complex envelope of a linearly polarized classical field at the detector surface A_d, η is the quantum efficiency, h is Planck's constant, and f_0 is the frequency of the radiation. η is assumed constant over the frequency range of interest, and the field is normalized such that the characteristic impedance of the medium can be taken to be unity. If the incident field has an arbitrary polarization, then the integrand in (4.3) is the sum of the squared magnitudes of the orthogonal polarization components.

This detector model easily admits generalization to account for various manifestations of non-ideal behavior[52,96]. Finite detector bandwidth can be accommodated by assuming that the output is accessible only through a realizable filter ("filtered Poisson process"). This filter can be randomized in various ways. Chaotic thermal noise and detector dark current can be accounted for by independent additive Gaussian and shot noises at the output.

For the applications considered in this chapter we will not concern ourselves with these generalizations; nor shall we consider the actual behavior of real devices, for which detailed discussions can be found in the literature[52,96].

Output statistics for the ideal quantum photodetector can be calculated using classical (i.e., non-quantum) tools[47,48]. Theorem 3.5 gives a complete formal statistical description of the photodetector counting process, in terms of the MMSE causal estimate of the intensity (4.3). A useful marginal statistic is the probability p(k) of k photocounts in a given time interval, say $[t_0, t_1] = T$. Conditioned on the incident field, the counts at time t_1 obey a Poisson law,

$$p(k|a_{\pi};\ t \in T,\ r \in A_d) = \frac{m_{t_1}}{k!} e^{-m_{t_1}} \tag{4.4}$$

$$m_{t_1} = \alpha \iint_{T\,A_d} |a_{\pi}|^2 dr\,dt \tag{4.5}$$

where $\alpha = \eta/hf_0$ and a_{π} is the complex envelope of the field. It is often useful to know p(k), the unconditional counting distribution, for a random field.

With only slight modifications, the results of Reference 27 apply if we assume that $\{a_{\pi};\ t \in T,\ r \in A_d\}$ is a Gaussian field. We specialize some of these results here for the case of small detectors. We assume that a_{π} is the sum of a zero-mean white Gaussian noise $\{n_t, t \in T\}$, bandlimited to $\pm$ B Hertz, and a non-random function s_t occupying the same band. We have suppressed r-dependence because, by assumption, a_{π} as a function of r varies negligibly over the detector surface. Taking the first $2B(t_1 - t_0) + 1$ eigenvalues of the covariance of $\{n_t, t \in T\}$ to be the same (N_0), and the rest to be zero, it can be shown that the counting distribution is given by

$$p(k) = \frac{(\beta N_0)^k}{(1+\beta N_0)^{k+2B(t_1-t_0)+1}} \exp - \frac{\beta \int_{t_0}^{t_1} |s_t|^2 dt}{1+\beta N_0} L_k^{2B(t_1-t_0)} \left[- \frac{\int_{t_0}^{t_1} |s_t|^2 dt}{N_0(1+\beta N_0)} \right] \quad (4.6)$$

where β is a constant directly proportional to detector quantum efficiency and surface area, and inversely proportional to photon energy. For a wide range of signal intensities $|s_t|^2$, this can be approximated[27] by a Poisson distribution with intensity

$$\beta |s_t|^2 + \beta \overline{P} \quad (4.7)$$

when $\beta N_0 \ll 1$ and $2B(t_1 - t_0) \gg 1$. $\overline{P}$ is the total average noise power $2BN_0$. This Poisson approximation (4.7) is often valid under other circumstances, when $\{n_t, t \in T\}$ is not necessarily white[28]. This approximation is a traditional one in optical communications, as few results would have been forthcoming without it. In effect, it models the photodetector as a device which translates a known signal in Gaussian noise into a Poisson signal in Poisson noise. In principle, our MMSE results of the previous chapter have no need of this approximation, as they are designed for doubly-stochastic Poisson processes; however, significant simplifications are often possible when the Poisson approximation is used. For the specific examples considered later in this chapter, we use the Poisson approximation.

D. Communication Model.

In deriving recursive MMSE estimator equations for random processes in Chapter III, we assumed that the intensity process $\{\lambda_t, t \in T\}$ was a causal functional of the Ito process $\{x_t, t \in T\}$ to be estimated. However, we did not specify how $\{x_t, t \in T\}$ originated. In this section we present a recipe for $\{x_t, t \in T\}$ which embodies all the salient features of our channel and detector models. This prescription is similar in many ways to the very general state variable model postulated by Snyder[14] for the Gaussian estimation problem.

Let x_t be the vector obtained by adjoining the vectors a_t, u_t, w_t, and b_t; i.e.,

$$x_t' = [\, a_t' \;\; u_t' \;\; w_t' \;\; b_t' \,] \tag{4.8}$$

The order is not significant, but we shall use that shown, for consistency. We now define the vectors comprising x_t.

Let $\{a_t, t \in T\}$ be the Ito process (see Ch. II) defined by the equation,

$$da_t = f_t^{(a)} dt + g_t^{(a)} d\alpha_t, \; a_t \text{ real} \tag{4.9}$$

where $\{\alpha_t, t \in T\}$ is a martingale. We take $\{a_t, t \in T\}$, or elements thereof, to be the message for which an estimate is ultimately desired.

$\{a_t,\ t\in T\}$ and $\{v_t,\ t\in T\}$ are the input and output of a linear filter described by the equations,

$$\dot{u}_t = F_t^{(u)} u_t + G_t^{(u)} a_t$$
$$v_t = A_t^{(v)} u_t + B_t^{(v)} a_t \tag{4.10}$$

A nonlinear memoryless transformation maps the filter output $\{v_t,\ t\in T\}$ into the "signal" $\{s_t,\ t\in T\}$, which in turn is subjected to additional linear filtering via the equations,

$$\dot{w}_t = F_t^{(w)} w_t + G_t^{(w)} s_t$$
$$z_t = A_t^{(z)} w_t + B_t^{(z)} s_t \tag{4.11}$$

The first filter and the nonlinear transformation are invariably associated with a "modulator," which generates a signal suitable for transmission over the channel. Depending on circumstances, the second linear filter, with output $\{z_t,\ t\in T\}$, can be associated either with the modulator or with the channel. This message-modulator model is a most versatile one, covering a wide range of linear and nonlinear modulation schemes[14].

Let $\{b_t,\ t\in T\}$ be the Ito process defined by the equation

$$db_t = f_t^{(b)} dt + g_t^{(b)} d\beta_t,\ b_t \text{ real} \tag{4.12}$$

where $\{\beta_t,\ t \in T\}$ is a martingale. We take $\{b_t,\ t \in T\}$ to be a source of channel disturbance which interacts with $\{z_t, t \in T\}$ in a possibly nonlinear way.

The channel output $\{y_t,\ t \in T\}$ is described by the equation

$$dy_t = h_t(x_t)dt + g_t^{(y)} d\eta_t,\ y_t \text{ real}, \tag{4.13}$$

where $\{\eta_t,\ t \in T\}$ is an independent increment process. $\dot{\eta}_t$ formally represents additive white noise in the channel. h_t is a memoryless nonlinear transformation of x_t.

For the optical channel, we are concerned only with the complex envelope of the channel output. Therefore it is appropriate to let $\{y_t,\ t \in T\}$, and often $\{b_t,\ t \in T\}$ and $\{z_t,\ t \in T\}$, be composed of the real and imaginary parts of complex envelopes, thus doubling their dimension. Then, for D "small" photodetectors distributed over the receiving aperture, the rate process for the j^{th} detector is specified by

$$\lambda_{t_j} = \beta_j\, |y_{t_j}|^2 \tag{4.14}$$

where $|y_{t_j}|^2$ denotes $[\text{Real}\ (y_{t_j})]^2 + [\text{Imaginary}\ (y_{t_j})]^2$, and β_j is a positive constant accounting for quantum efficiency, photon energy, and detector surface area. If the Poisson approximation discussed in the previous section is employed, λ_{t_j} can be written

$$\lambda_{t_j} = \beta_j \left| h_{t_j}(x_t) \right|^2 + \beta_j \lambda_{o_j} \tag{4.15}$$

λ_{o_j} represents an independent additive Poisson noise at the output of the j^{th} photodetector, and easily accounts for detector dark current as well as channel noise.

With $\{x_t,\ t \in T\}$ defined by (4.8), it can be seen that $\{x_t,\ t \in T\}$ satisfies

$$dx_t = f_t dt + g_t d\xi_t \tag{4.16}$$

with

$$f_t = \left[\begin{array}{c} f_t^{(a)} \\ \hdashline F_t^{(u)} u_t + G_t^{(u)} a_t \\ \hdashline F_t^{(w)} w_t + G_t^{(w)} s_t \\ \hdashline f_t^{(b)} \end{array} \right] \tag{4.17}$$

$$g_t = \left[\begin{array}{cccc} g_t^{(a)} & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & g_t^{(b)} \\ \hdashline \end{array} \right] \tag{4.18}$$

$$\xi_t = \begin{bmatrix} \alpha_t \\ \hline 0 \\ \hline 0 \\ \hline \beta_t \end{bmatrix} \qquad (4.19)$$

As with the modulator model, this channel model is most versatile, containing many commonly occurring channels[14].

In estimating x_t we are actually jointly estimating all the variables in the communication model. To extract that desired estimates we simply multiply x_t by an appropriately chosen matrix of zeroes and ones. This is a valid operation in the context of MMSE estimation.

This completes the definition of our communication model. It accounts for diversity reception by ideal small photodetectors, and includes a variety of channels and modulators. Although the model was developed here in the context of Ito processes, it retains much of its generality when all the messages and disturbances are constrained to be Markov diffusion processes. For example, modulator and/or channel memory can still be accounted for by the linear filters (4.10) and (4.11).

Photodetectors of arbitrary size can be accommodated in a straightforward manner. By definition, such a detector intercepts more than one spatial mode of the incident field; λ_{t_j} is the sum of the squared magnitudes of each spatial mode, the modes being determined by an orthogonal spatial decomposition of the field over the detector surface. A stochastic equation such as (4.13) describes each spatial mode. Thus, photodetectors of any size can be handled by taking the spatial modes of the field envelope as the fundamental channel output descriptors. The number of such modes is the spatial diversity measure of the channel.

E. Applications.

In this section we consider some specific applications for Poisson estimation theory. The diversity of examples emphasizes the essential unity of the approach, and illustrates the tractability of techniques for many heretofore difficult problems which have not yielded to earlier, piecemeal attacks.

Our set of examples is by no means exhaustive, or even entirely representative of real problems in optical communication. We attempt, as far as possible, however, to present some problems which are meaningful, and yet stripped of the complexities which could easily obscure the central issues.

1. Channel Measurement: Lognormally Fading Signal.

Let the intensity of the scalar observations $\{N_t,\ t \in T\}$ be

$$\lambda_t = \beta s_t e^{2x_t^{(o)}}, \quad t \in T \tag{4.20}$$

where β is a positive constant, s_t is a known positive function, and $\{x_t^{(o)},\ t \in T\}$ is a Gaussian process given by

$$x_t^{(o)} = x_t - P_o \tag{4.21}$$

with $\{x_t,\ t \in T\}$ satisfying

$$dx_t = -kx_t dt + \sqrt{2P_o k}\, dw_t \tag{4.22}$$

$\{w_t,\ t \in T\}$ is a standardized scalar Wiener process, so $\{x_t,\ t \in T\}$ is a Gaussian first-order Markov process with covariance $P_o e^{-k|t-\tau|}$. We assume that $E\{x_{to}\} = 0$, which implies that $E\{x_t\} = 0$, $t \in T$. Thus $\{\lambda_t,\ t \in T\}$ is a lognormal process with mean βs_t, $t \in T$, and covariance $(t, \tau \in T)$

$$K_\lambda(t,\tau) = \beta^2 s_t s_\tau \left[\exp\left(4P_o e^{-k|t-\tau|}\right) - 1\right] \qquad (4.23)$$

It is desired to obtain the MMSE estimate of $x_t^{(o)}$; in terms of $\hat{x}_t$, this estimate is given by $\hat{x}_t^{(o)} = \hat{x}_t - P_o$. In the past, a "fast fading" lognormal intensity has resisted most systematic attempts at estimation. We shall see, however, that it is a tractable intensity for our approach.

Whether we use the first order, second order, or Gaussian second order approximation, the filter (processor) equation is the same:

$$d\hat{x}_t = -k\hat{x}_t dt - 2P_t \beta e^{-2P_o} s_t e^{2\hat{x}_t} dt + 2P_t dN_t \qquad (4.24)$$

The variance equation, however, depends on the approximation. We will use the Gaussian second order equation, because it turns out to be explicitly independent of the observations:

$$dP_t = -2kP_t dt + 2P_o k dt - 4\beta e^{-2P_o} s_t e^{2\hat{x}_t} P_t^2 dt \qquad (4.25)$$

Recalling (3.79), this is seen to be a natural consequence of the lognormal form of $\{\lambda_t,\ t \in T\}$. Eq. (4.25), which is a Ricatti equation, can be solved in advance in terms of $\hat{x}_t$, and the solution can be substituted directly in (4.24), resulting in an observation-invariant filter. This important result also holds when $\{x_t,\ t \in T\}$ and $\{N_t,\ t \in T\}$ are multidimensional. Figure 4.1 indicates how (4.24) might be implemented.

It is clear from (4.24) that the filter for x_t is nonlinear in the observations. As an alternative approach, we might wish to find a linear estimator for x_t. According to Theorem 3.6, an optimum linear MMSE estimator will have the form of a linear filter,

$$\hat{x}_t = \int_{t_0}^{t} L_\sigma dN_\sigma - \int_{t_0}^{t} L_\sigma E(\lambda_\sigma) d\sigma \tag{4.26}$$

where L_σ satisfies the linear integral equation,

$$L_\sigma E(\lambda_\sigma) + \int_{t_0}^{t} K_\lambda(\sigma, \tau) L_\tau d\tau = E\{x_t(\lambda_\sigma - E(\lambda_\sigma))\}$$

To specify this filter, we must have $E(x_t\lambda_\sigma)$. It is easily seen that

$$E(x_t\lambda_\sigma) = \beta e^{-2P_0} s_\sigma \frac{\partial}{\partial u} M_{x_t, x_\sigma}(u, v)\Big|_{\substack{u=0 \\ v=\lambda}} \tag{4.27}$$

where $M_{x_t, x_\sigma}(u, v)$ is the joint moment generating function of x_t and x_σ. Since these r.v.'s are jointly Gaussian, we have

$$M_{x_t, x_\sigma}(u, v) = \exp\left\{\frac{1}{2} P_0 \left[(u^2+v^2) + 2uve^{-k|t-\sigma|}\right]\right\} \tag{4.28}$$

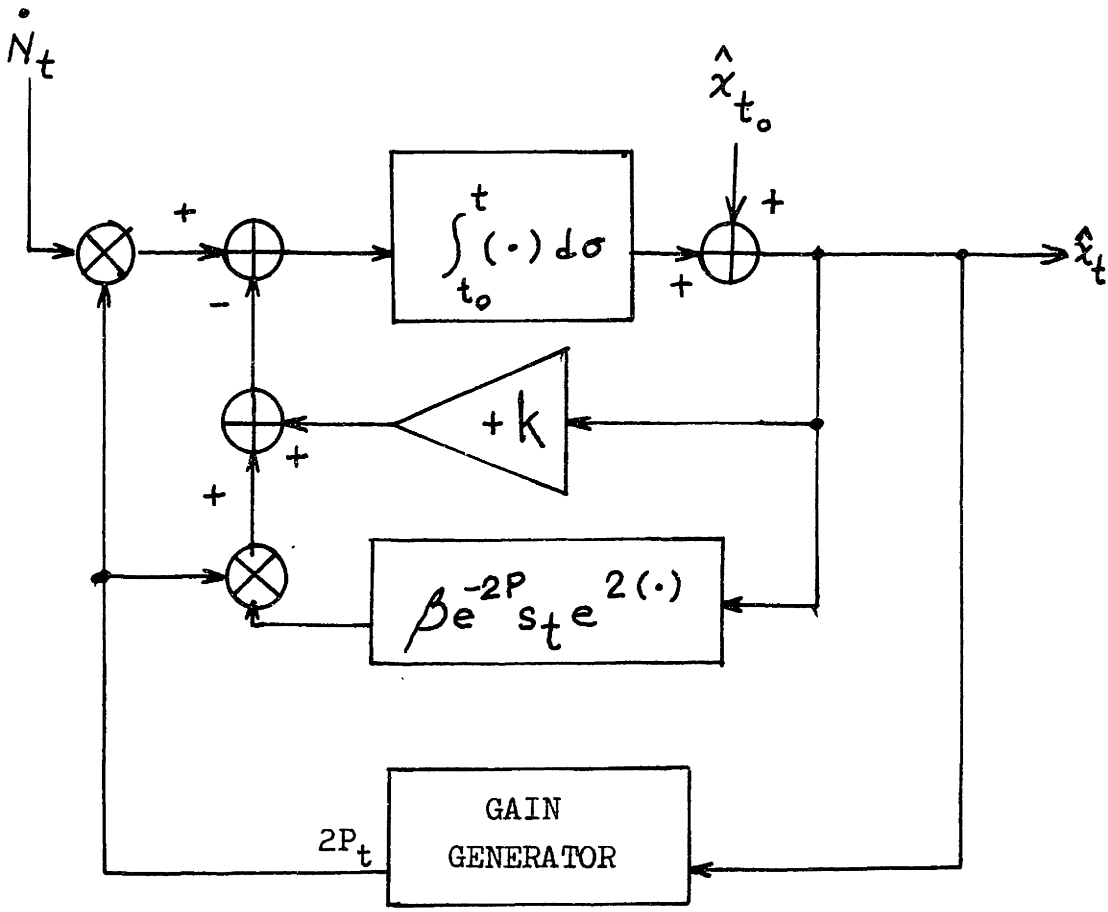

$\dot{N}_t$
+
-
$\int_{t_0}^{t}(\cdot)\,d\sigma$
$\hat{x}_{t_0}$
+
+
$\hat{x}_t$
$+k$
+
+
$\beta e^{-2P}s_t e^{2(\cdot)}$
$2P_t$
GAIN
GENERATOR

Figure 4.1

Thus,

$$E(x_t \lambda_\sigma) = 2\beta P_o s_\sigma e^{-k|t-\sigma|} \tag{4.29}$$

and the equation for L_σ is ($\sigma \in [t_o, t]$)

$$L_\sigma + \beta \int_{t_o}^{t} L_\tau s_\tau [\exp(4P_o e^{-k|\sigma-\tau|}) - 1] d\tau = 2P_o e^{-k|t-\sigma|} \tag{4.30}$$

The m.s. error of the optimum linear estimate is, from Corollary 3.7.2 of Theorem 3.7,

$$E\{(x_t - \hat{x}_t)^2\} = P_o - 2P_o \beta e^{-kt} \int_{t_o}^{t} L_\sigma s_\sigma e^{k\sigma} d\sigma \tag{4.31}$$

To facilitate comparison with a Cramer-Rao bound, let $k \to 0$; this implies that $\{x_t, t \in T\}$ becomes slowly varying in time, and we can take it to be a constant r.v. over $[t_o, t]$. Then L_σ is independent of σ and is given by

$$L = \frac{2P_o}{1 + \beta(e^{4P_o} - 1)\int_{t_o}^{t} s_\tau d\tau} \tag{4.32}$$

$\hat{x}_t$ takes the form,

$$\hat{x}_t = \frac{2P_o\left(N_t - \beta\int_{t_o}^{t} s_\tau d\tau\right)}{1 + \beta(e^{4P_o} - 1)\int_{t_o}^{t} s_\tau d\tau}, \tag{4.33}$$

expected for slow fading, for which N_t is a sufficient statistic. The m.s. error of (4.33) is

$$E\{(x_t - \hat{x}_t)^2\} = P_o - \frac{4P_o^2\beta\int_{t_o}^{t} s_\tau d\tau}{1+\beta(e^{4P_o}-1)\int_{t_o}^{t} s_\tau d\tau} \quad (4.34)$$

$$\rightarrow P_o - \frac{4P_o^2}{e^{4P_o}-1} \qquad \text{for large } t$$

From III.G.1, it is seen that the Cramer-Rao bound for this problem is expressible as

$$E\{(x_t - \hat{x}_t)^2\} > P_o - \frac{4P_o^2\beta\int_{t_o}^{t} s_\tau d\tau}{1+4P_o\beta\int_{t_o}^{t} s_\tau d\tau} \quad (4.35)$$

$$\rightarrow 0 \text{ for large } t$$

for <u>any</u> estimate $\hat{x}_t$ of x. An efficient estimate does not exist for the lognormal intensity, as long as $P_o > 0$. Note, however, that the performance of the linear estimator is quite close to the bound for small fading variance P_o, indicating that the linear approximation is reasonable for weak fading. Indeed, for this situation there is no point in implementing a non-linear filter; the <u>optimum</u> nonlinear estimator is bounded above and below in performance by (4.34) and (4.35), whereas the <u>quasi-optimum</u> estimator, (4.24), might actually perform <u>worse</u> than the linear estimator, by virtue of the ad hoc approximations leading to the Gaussian second order equations.

2. Channel Measurement: Lognormally Fading Signal, Poisson Noise.

As a generalization of the preceding example, we estimate $x_t^{(o)}$ when the intensity process is

$$\lambda_t = \beta s_t e^{2x_t^{(o)}} + \beta\mu_o \;, \quad t \in T, \tag{4.36}$$

where μ_o is a non-negative constant. This intensity is often a good model for a fading optical signal in wideband background noise, as was pointed out in IV.C. We make the same assumptions as in the noiseless example; in addition, we assume that $\mu_o \ll s_t$, an assumption sometimes valid for the clear atmospheric channel. $\{\lambda_t, t \in T\}$ is a lognormal process with mean $\beta s_t + \beta\mu_o$ and covariance given by (4.23). Again, $x_t^{(o)} = x_t - P_o$.

As in the previous example, the first order, second order, and Gaussian second order approximations yield identical filter equations,

$$d\hat{x}_t = -k\hat{x}_t dt - 2P_t\beta e^{-2P_o} s_t e^{2\hat{x}_t} dt$$

$$+ (e^{-2P_o} s_t e^{2\hat{x}_t} + \mu_o)^{-1} 2e^{-2P_o} s_t e^{2\hat{x}_t} P_t dN_t \tag{4.37}$$

which, in light of the small μ_o assumption, can be written as

$$d\hat{x}_t = -k\hat{x}_t dt - 2P_t\beta e^{-2P_o} s_t e^{2\hat{x}_t} dt$$

$$+ 2(1 - \mu_o e^{2P_o} s_t^{-1} e^{-2\hat{x}_t}) P_t dN_t \tag{4.38}$$

The Gaussian second order variance equation is again the most attractive:

$$dP_t = 2kP_t dt + 2P_o k dt - 4\beta\, e^{-2P_o} s_t e^{2\hat{x}_t} P_t^2 dt$$

$$+ P_t^2 \frac{\partial^2}{\partial \hat{x}_t^2} \ln \lambda_t(\hat{x}_t) dN_t \qquad (4.39)$$

We can use the small μ_o assumption to simplify (4.39) somewhat:

$$dP_t = -2kP_t dt + 2P_o k dt - 4\beta\, e^{-2P_o} s_t e^{2\hat{x}_t} P_t^2 dt$$

$$+ 4\mu_o e^{2P_o} s_t^{-1} e^{-2\hat{x}_t} P_t^2 dN_t \qquad (4.40)$$

In contrast to the noiseless case, we see that this equation is not independent of the observations; however, for small μ_o the last term in (4.40) becomes negligible, and (4.40) is then the same as (4.25). Thus, for sufficiently small noise and fading, x_t can be estimated with substantially the same accuracy as in the noiseless case. Even if (4.25) seriously underestimates the estimator error, it might be worthwhile to assume an observation-invariant filter, and optimize P_t empirically when the processor is implemented. This approach would help to alleviate any possible problems of filter divergence due to data-induced variations in P_t.

With only slight modifications of the noise-free example, we can derive a linear MMSE estimator for x_t, with no restrictions on the magnitude of μ_0. In the equation for L_σ, the only quantity which changes to account for μ_0 is $E(\lambda_\sigma)$; the result is ($\sigma \in [t_0, t]$)

$$L_\sigma(s_\sigma + \mu_0) + \beta s_\sigma \int_{t_0}^{t} L_\tau s_\tau \left(e^{4P_0 e^{-k|\sigma-\tau|}} - 1\right) d\tau$$

$$= 2P_0 s_\sigma e^{-k|t - \sigma|} \qquad (4.41)$$

As in (4.31), the m.s. error is

$$E\{(x_t - \hat{x}_t)^2 = P_0 - 2P_0 \beta e^{-kt} \int_{t_0}^{t} L_\sigma s_\sigma e^{k\sigma} d\sigma \qquad (4.42)$$

It can be seen that $L_\sigma, \sigma \in [t_0, t]$ is a continuous, monotone decreasing function of μ_0; for small μ_0, the solution of (4.41) differs only slightly from that of (4.30). However, for $k \to 0$ and $\mu_0 > 0$, L_σ is not independent of σ, as it was in the noise-free case, unless s_t is a constant. For $k = 0$ and $s_t = s$, $t \in T$, the solution of (4.41) is

$$L = \frac{2P_0 s}{\mu_0 + s + \beta s^2 (e^{4P_0} - 1)(t - t_0)} \qquad (4.43)$$

and $\hat{x}_t$ is given by

$$\hat{x}_t = \frac{2P_0 s (N_t - \beta s (t - t_0))}{\mu_0 + s + \beta s^2 (e^{4P_0} - 1)(t - t_0)} \qquad (4.44)$$

Observe that for very large μ_0, $\hat{x}_t \cong 0$, the a priori mean of x_t. The m.s. error of this estimate is

$$E\{(x_t - \hat{x}_t)^2\} = P_o - \frac{4P_o^2 s\beta(t-t_o)}{1 + \frac{\mu_o}{s} + \beta s(e^{4P_o}-1)(t-t_o)} \qquad (4.45)$$

$$\rightarrow P_o - \frac{4P_o^2}{e^{4P_o}-1} \qquad \text{for large } t.$$

The C-R bound can be evaluated by using (3.94), (3.95), and the expansion,

$$\frac{1}{1+a} = \sum_{n=0}^{\infty} (-a)^n$$

The result is

$$E\{(x_t - \hat{x}_t)^2\} > P_o - \frac{4P_o^2 s\beta(t-t_o)}{\varepsilon^{-1} + 4P_o s\beta(t-t_o)} \qquad (4.46)$$

$$\rightarrow 0 \text{ for large } t$$

where

$$\varepsilon = \sum_{n=0}^{\infty} \left(-\frac{\mu_o}{s}\right)^n e^{2P_o n(n-1)} \qquad (4.47)$$

Observe that, for small P_o and $\mu_o \ll s$, (4.45) and (4.46) are quite close, indicating that if the additive noise is small, the linear estimator is again reasonable for weak fading. Note also that, for large t, the estimate in the presence of noise, (4.44), has the same asymptotic m.s. error as the estimate (4.33) for the noiseless case.

3. Signal Model for Communications.

We now shift our attention to the simultaneous estimation of intelligence and lognormal channel fading. In the most general setting, we take the vector rate process of $\{N_t,\ t \in T\}$ to be

$$\lambda_t = \beta e^{H_t^{f'} x_t} |s_t|^2 + \beta\mu_0 \tag{4.48}$$

where $e^{(\cdot)}$ is the vector exponential function, the elements of which are the exponentials of the elements of $(\cdot)$, and the complex signal envelope s_t (scalar) is

$$s_t = A_t + B_t e^{i\theta t} + C_t e^{i\phi t}. \tag{4.49}$$

β and μ_0 play the same roles as in the previous examples; x_t is a vector whose components represent the message and fading processes. H_t^f is a matrix which picks out the fading components; some or possibly all of the variables in s_t depend on the message components in x_t.

In practice s_t will seldom be as complicated as we have indicated; some of the parameters in (4.49) will not be present. In particular, s_t will usually take one of the forms,

$$s_t = A_t \tag{4.50}$$

$$s_t = A_t + B_t e^{i\theta_t} \tag{4.51}$$

$$s_t = B_t e^{i\theta_t} + C_t e^{i\theta_t} \tag{4.52}$$

The first represents intensity modulation, with A_t related to x_t by $A_t = h_t^{A\prime} x_t$; h_t^A is a vector which picks out the message component in x_t. The second form, (4.51), represents phase or frequency modulation on a sideband, possibly with pre-emphasis; it also allows for a transmitted reference or pilot tone signal (i.e., A_t). The last form is related to heterodyning; θ_t and ϕ_t appear in λ_t only as the difference $\theta_t - \phi_t$. If $B_t e^{i\theta_t}$ is the transmitted signal and $C_t e^{i\phi_t}$ the local oscillator, then the intermediate frequency (I.F.) signal is proportional to $|s_t|^2 = M_t^2 + L_t^2 \cos(\theta_t - \phi_t)$. θ_t and ϕ_t, in (4.51) and (4.52), are related to x_t by $\theta_t = \omega_\theta t + h_t^{\theta\prime} x_t$ and $\phi_t = \omega_\phi t + h_t^{\phi\prime} x_t$; the coefficients h_t^θ, h_t^ϕ again extract appropriate components of x_t.

In general, the real signal leading to (4.49) is

$$E_t = \mathrm{Re}\left\{s_t e^{i2\pi f_0 t}\right\}$$

$$= A_t \cos 2\pi f_0 t + B_t \cos(2\pi f_0 t + \theta_t) + C_t \cos(2\pi f_0 t + \phi_t)$$

The essential quantity $|s_t|^2$ is given by

$$|s_t|^2 = M_t^2 + J_t^2 \cos \theta_t + K_t^2 \cos \phi_t + L_t^2 \cos(\theta_t - \phi_t) \tag{4.53}$$

with

$$
\begin{aligned}
M_t^{\,2} &= A_t^{\,2} + B_t^{\,2} + C_t^{\,2} \\
J_t^{\,2} &= 2A_tB_t \\
K_t^{\,2} &= 2A_tC_t \\
L_t^{\,2} &= 2B_tC_t
\end{aligned}
\qquad (4.54)
$$

Our task now is to find the approximate estimator equation, in terms of (4.48) and (4.53). In general, this is a tedious and unenlightening job, even if we take A_t, B_t, and C_t to be non-random. However, it is instructive to determine the estimator equations for certain simple cases. The extension to more complicated examples is straightforward.

4. Angle Modulation on a Subcarrier.

We now consider a particular example in detail. Let s_t be given by

$$s_t = A_t + B_t e^{i\theta_t} \qquad (4.55)$$

where A_t and B_t are real, non-random functions of time, and $\theta_t = \omega_\theta t + h_t^{\theta\prime} x_t$. It is assumed that there are no initial frequency or phase offsets; i.e., at $t = t_0$ the receiver is assumed to be perfectly synchronized in phase and frequency with the transmitter. Initial phase lock is not an essential requirement, as the effects of the initial condition become negligible for $t \gg t_o$. λ_t is given by

$$\lambda_t = \beta e^{H_t^{f\prime} x_t}(M_t^2 + J_t^2 \cos\theta_t) + \beta\mu_0 \tag{4.56}$$

and

$$M_t^2 = A_t^2 + B_t^2$$

$$J_t^2 = 2A_tB_t \tag{4.57}$$

For convenience, we shall again use the Gaussian second-order equations developed in III.F.1. First, the various derivatives appearing in the estimator and variance equations must be evaluated. Omitting obvious steps, and abbreviating $\hat{\theta}_t = \omega_\theta t + h_t^{\theta\prime}\hat{x}_t$,

$$\nabla_{\hat{x}_t}(\|\lambda_t\|_1) = -\beta \| e^{H_t^{f\prime}\hat{x}_t}\|_1 J_t^2 h_t^\theta \sin\hat{\theta}_t$$

$$+ \beta H_t^f e^{H_t^{f\prime}\hat{x}_t}[M_t^2 + J_t^2\cos\hat{\theta}_t] \tag{4.58}$$

$$D_{\hat{x}_t}[\ln\lambda_t] = \left[H_t^f - \frac{J_t^2 h_t^\theta 1_D' \sin\hat{\theta}_t}{M_t^2 + J_t^2\cos\hat{\theta}_t}\right]$$

$$\cdot \operatorname{diag}\left(e^{H_t^{f\prime}\hat{x}_t}[M_t^2+J_t^2\cos\hat{\theta}_t]\right)\operatorname{diag}^{-1}\left(e^{H_t^{f\prime}\hat{x}_t}[M_t^2+J_t^2\cos\hat{\theta}_t]+\mu_0\right) \tag{4.59}$$

$$H_{\hat{x}_t}(\|\lambda_t\|_1) = -\beta \|e^{H_t^{f\prime}\hat{x}_t}\|_1 J_t^2 h_t^\theta h_t^{\theta\prime} \cos\hat{\theta}_t$$

$$-2\beta H_t^f e^{H_t^{f\prime}\hat{x}_t} h_t^{\theta\prime} J_t^2 \sin\hat{\theta}_t$$

$$+\beta H_t^f \operatorname{diag}(e^{H_t^{f\prime}\hat{x}_t}) H_t^{f\prime}[M_t^2 + J_t^2\cos\hat{\theta}_t] \tag{4.60}$$

Observe that $\nabla_{\hat{x}_t}(\|\lambda_t\|_1)$ and $H_{\hat{x}_t}(\|\lambda_t\|_1)$ are independent of the noise vector, μ_0. The expression for $H_{\hat{x}_t}[dN_t'\ln\lambda_t]$ is quite complicated; however, it will be needed soon, so we write it out. Denote the <u>columns</u> of H_t^f by h_t^1, h_t^2, h_t^3, ..., etc. Let R_t be a 3-dimensional array with 2-dimensional sections R_t^j given by

$$R_t^j = M_t^2 h_t^j h_t^{j\prime} + J_t^2(h_t^j h_t^{j\prime} - h_t^\theta h_t^{\theta\prime})\cos\hat{\theta}_t$$

$$-J_t^2(h_t^\theta h_t^{j\prime} + h_t^j h_t^{\theta\prime})\sin\hat{\theta}_t$$

Define the matrix-valued function $R_t \odot D$ as

$$R_t \odot D \triangleq \sum_j R_t^j d_j ,$$

where D is a diagonal matrix. Then

$$H_{\hat{x}_t}[dN_t' \ln\lambda_t]$$

$$= R_t \odot \left[\operatorname{diag}(\mu_0)\operatorname{diag}\left(e^{H_t^{f'}\hat{x}_t}\right)\operatorname{diag}^{-2}\left(e^{H_t^{f'}\hat{x}_t}[M_t^2+J_t^2\cos\hat{\theta}_t]+\mu_0\right)\operatorname{diag}(dN_t)\right]$$

$$-[J_t^4 + M_t^2 J_t^2 \cos\hat{\theta}_t]\, h_t^{\theta} h_t^{\theta'} \tag{4.61}$$

$$\bullet\ \operatorname{Tr}\left[\operatorname{diag}^2\left(e^{H_t^{f'}\hat{x}_t}\right)\operatorname{diag}^{-2}\left(e^{H_t^{f'}\hat{x}_t}[M_t^2+J_t^2\cos\hat{\theta}_t]+\mu_0\right)\operatorname{diag}(dN_t)\right]$$

Equations (4.58)-(4.61) still yield excessively complicated estimator equations, so we shall attempt to find further approximations. As the additive noise μ_0 is often quite small in practice, and appears to have a non-singular effect on the estimator equations, we might choose to ignore its presence in (4.59) and (4.61). Eqs. (4.59) and (4.61) then become

$$D_{\hat{x}_t}[\ln\lambda_t] \cong H_t^f - \frac{J_t^2 \sin\hat{\theta}_t}{M_t^2+J_t^2\cos\hat{\theta}_t}\, h_t^{\theta} 1_D' \tag{4.62}$$

$$H_{\hat{x}_t}[dN_t'\ln\lambda_t] \cong -\frac{[J_t^4+M_t^2J_t^2\cos\hat{\theta}_t]\, h_t^{\theta}h_t^{\theta'}}{[M_t^2+J_t^2\cos\hat{\theta}_t]^2}\, \|dN_t\|_1 \tag{4.63}$$

By ignoring the noise in the observations, we can arrive at further simplifications for $H_{\hat{x}_t}[dN_t'\ln\lambda_t]$. Let

$$dN_t = dn_t + \beta e^{H_t^{f'}x_t}[M_t^2+J_t^2\cos\theta_t]\,dt\ ; \tag{4.64}$$

then

$$\|dN_t\|_1 = \|dn_t\|_1 + \beta\|e^{H_t^{f'}}x_t\|_1[M_t^2+J_t^2\cos\theta_t]dt$$

$$= \beta\|e^{H_t^{f'}}\hat{x}_t\|_1[M_t^2+J_t^2\cos\hat{\theta}_t]dt + \|dn_t\|_1$$

$$+ \{\beta\|e^{H_t^{f'}}x_t\|_1[M_t^2+J_t^2\cos\theta_t]$$

$$-\beta\|e^{H_t^{f'}}\hat{x}_t\|_1[M_t^2+J_t^2\cos\hat{\theta}_t]\}dt \tag{4.65}$$

where the last equality results from adding and subtracting identical terms. Using the above in (4.63), we have

$$H_{\hat{x}_t}[dN_t' \ln\lambda_t] \cong$$

$$-\frac{\beta\|e^{H_t^{f'}}\hat{x}_t\|_1[J_t^4+M_t^2J_t^2\cos\hat{\theta}_t]h_t^\theta h_t^{\theta'}}{M_t^2+J_t^2\cos\hat{\theta}_t}dt$$

$$-\frac{[J_t^4+M_t^2J_t^2\cos\hat{\theta}_t]h_t^\theta h_t^{\theta'}}{[M_t^2+J_t^2\cos\hat{\theta}_t]^2}\{\|dn_t\|_1 +$$

$$+\beta\|e^{H_t^{f'}}x_t\|_1[M_t^2+J_t^2\cos\theta_t]dt - \beta\|e^{H_t^{f'}}\hat{x}_t\|_1[M_t^2+J_t^2\cos\hat{\theta}_t]dt\} \tag{4.66}$$

Following Snyder[23], we drop the last term on the r.h.s. of (4.66). We do so because $\{n_t, t \in T\}$ is a zero-mean process, and the difference term in the braces is of the order of $x_t - \hat{x}_t$. Thus, since the covariance P_t appears quadratically with $H_{\hat{x}_t}[dN_t' \ln \lambda_t]$ in the variance equation (3.79), this difference term results in a quantity of the order of $(x_t - \hat{x}_t)$ to the fifth power. We have dropped such quantities in the derivation of the Gaussian second-order variance equation (which contains terms only up to fourth order in the error).

This approximation has the useful effect of decoupling the variance equation from the data $\{N_t, t \in T\}$; i.e.,

$$H_{\hat{x}_t}[dN_t' \ln \lambda_t] \cong -\beta \| e^{H_t^{f\prime} \hat{x}_t} \|_1 \frac{J_t^2 [J_t^2 + M_t^2 \cos \hat{\theta}_t] h_t^{\theta} h_t^{\theta\prime}}{M_t^2 + J_t^2 \cos \hat{\theta}_t} dt \tag{4.67}$$

Equations (4.58), (4.60), (4.62), and (4.67) now specify our approximate estimator. If we assume, as is usually the case, that $\sin(\omega_\theta t + h_t^{\theta\prime} \hat{x}_t)$ and $\cos(\omega_\theta t + h_t^{\theta\prime} \hat{x}_t)$ are rapidly varying relative to other quantities in the estimator equations, then for ω_θ large enough we can reasonably neglect products of slowly and rapidly varying functions, which would have only slight effect on $\hat{x}_t$ and P_t. Then, letting "$\sim$" denote "slowly varying part,"

$$\nabla_{\hat{x}_t}(\|\lambda_t\|_1) \sim \beta M_t^2 H_t^f e^{H_t^{f\prime}\hat{x}_t} \tag{4.68}$$

$$H_{\hat{x}_t}(\|\lambda_t\|_1) \sim \beta M_t^2 H_t^f \, diag\left(e^{H_t^{f\prime}\hat{x}_t}\right) H_t^{f\prime} \tag{4.69}$$

These simplifications are obvious; it is less obvious what we do with $H_{\hat{x}_t}[dN_t' \ln \lambda_t]$. ($D_{\hat{x}_t}[\ln \lambda_t]$, (4.62), cannot be simplified further.)

Only the slowly varying ("DC") component of $H_{\hat{x}_t}[dN_t' \ln \lambda_t]$ is significant; this component is given by

$$\begin{aligned} DC &= \frac{1}{2\pi}\int_{-\pi}^{\pi} H_{\hat{x}_t}[dN_t' \ln \lambda_t]\, d\rho, \quad \rho = \hat{\theta}_t, \\ &= -\beta \|e^{H_t^{f\prime}\hat{x}_t}\|_1 \left(M_t^2 - \sqrt{M_t^4 - J_t^4}\right) h_t^\theta h_t^{\theta\prime} dt \\ &= -2\beta \|e^{H_t^{f\prime}\hat{x}_t}\|_1 B_t^2 h_t^\theta h_t^{\theta\prime} dt, \end{aligned} \tag{4.70}$$

where, recall, B_t is a parameter in s_t. Hence,

$$H_{\hat{x}_t}[dN_t' \ln \lambda_t] \sim -2\beta B_t^2 \|e^{H_t^{f\prime}\hat{x}_t}\|_1 h_t^\theta h_t^{\theta\prime} dt \tag{4.71}$$

Assume for convenience that $\{x_t, t \in T\}$ is a Gaussian diffusion; i.e.,

$$dx_t = F_t x_t dt + G_t dw_t, \tag{4.72}$$

where $\{w_t, t \in T\}$ is a standardized Wiener process, and F_t and G_t are non-random matrix functions. Eq.(4.72) is just a special case of the general diffusion equation,

$$dx_t = f_t(x_t)dt + g_t(x_t)dw_t,$$

in which f_t and g_t are memoryless nonlinear functions of t and x_t.

Using (4.72) and all the approximations we have made, we find that $\hat{x}_t$ and P_t satisfy the following equations (illustrated in Fig. 4.2):

$$d\hat{x}_t = F_t\hat{x}_t dt - \beta M_t^2 P_t H_t^f e^{H_t^{f\prime}\hat{x}_t} dt$$

$$+P_t\left\{H_t^f - \frac{J_t^2 \sin\hat{\theta}_t \, h_t^\theta 1_D^\prime}{M_t^2 + J_t^2\cos\hat{\theta}_t}\right\} dN_t \qquad (4.73)$$

$$dP_t = \left[F_t^\prime P_t + P_t F_t + G_t G_t^\prime\right]dt$$

$$- \beta M_t^2 P_t H_t^f \operatorname{diag}\left(e^{H_t^{f\prime}\hat{x}_t}\right) H_t^{f\prime} P_t dt$$

$$- 2\beta B_t^2 \left\|e^{H_t^{f\prime}\hat{x}_t}\right\|_1 P_t h_t^\theta h_t^{\theta\prime} P_t dt \qquad (4.74)$$

The series of approximations we have outlined have been carried through without ignoring the noise μ_0, but the results are very complicated and tedious to derive; hence, we

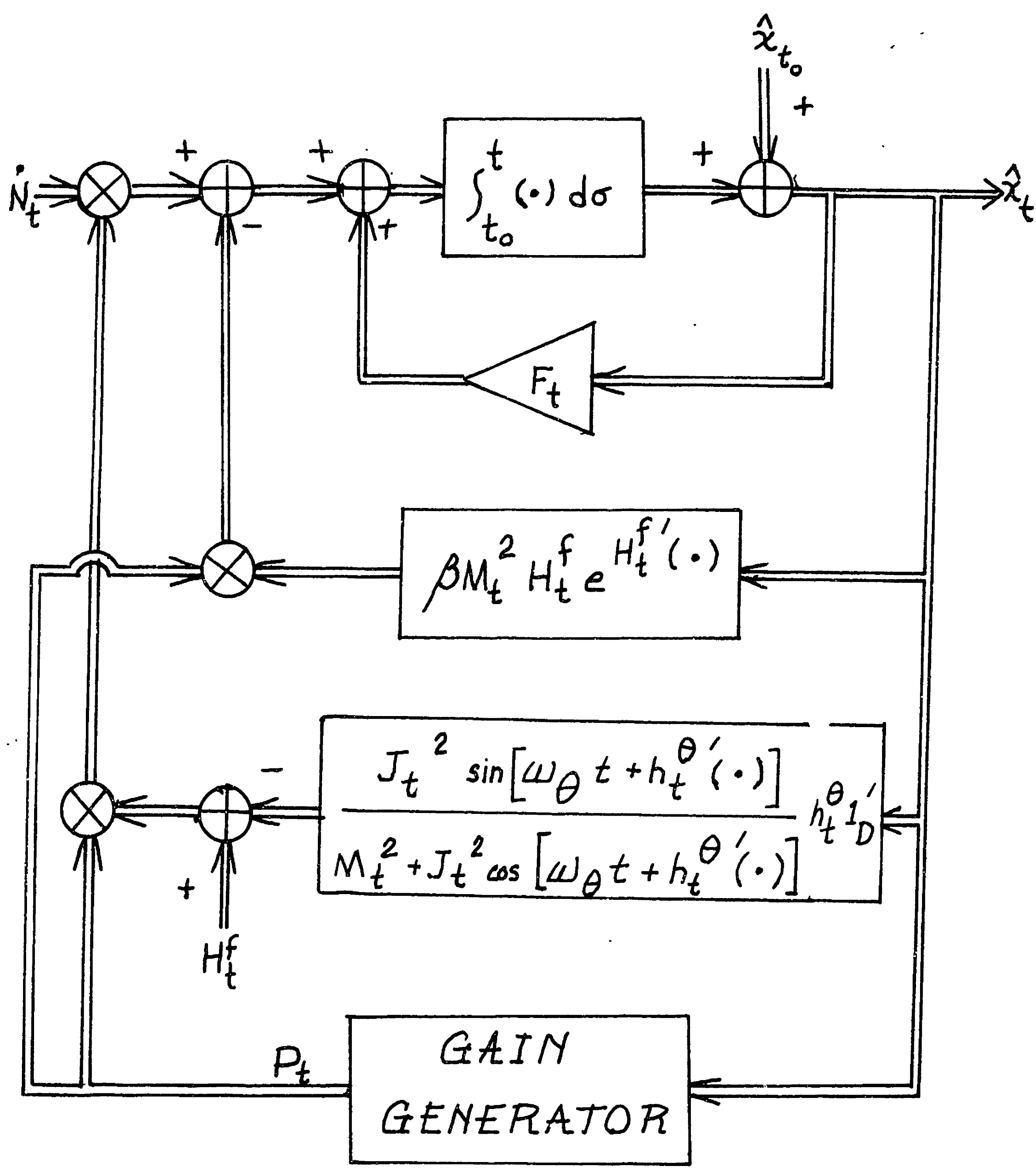

$\dot{N}_t$
$\hat{x}_{t_0}$
$\int_{t_0}^{t} (\cdot)\, d\sigma$
$\hat{x}_t$
F_t
$\beta M_t^2 H_t^f e^{H_t^{f\prime}(\cdot)}$
$\frac{J_t^2 \sin[\omega_\theta t + h_t^{\theta\prime}(\cdot)]}{M_t^2 + J_t^2 \cos[\omega_\theta t + h_t^{\theta\prime}(\cdot)]} h_t^{\theta} 1_D'$
H_t^f
P_t
GAIN GENERATOR

Figure 4.2

do not include them here. In the absence of fading, however, additive noise can be accounted for almost trivially. Nothing is lost in this case by taking D = 1, and μ_0 then simply augments the quantity M_t^2. Thus (4.73) and (4.74) apply, provided we replace M_t^2 by $M_t^2 + \mu_0$, and $2B_t^2$ (in the last term of (4.74)) by

$$(M_t^2 + \mu_0)\left(1 - \sqrt{1 - \left(\frac{J_t^2}{M_t^2+\mu_0}\right)^2}\right) \tag{4.75}$$

The latter replacement arises from the calculation in (4.70) of the slowly varying component of $H_{\hat{x}_t}[dN_t' \ln \lambda_t]$. It can be seen that, in the absence of fading, the effects of μ_0 are indistinguishable from the effects due to M_t^2. This phenomenon will be examined more closely in a subsequent example.

We remark that the approximations leading to (4.73) and (4.74) can be applied with equal facility (in the absence of noise) when $s_t = B_t e^{i\theta t} + C_t e^{i\phi t}$, because $|s_t|^2 = M_t^2 + L_t^2 \cos(\theta_t - \phi_t)$. The only additional requirement is that sin and cos of $\theta_t - \phi_t$ vary rapidly relative to other quantities.

If $s_t = A_t$, then $|s_t|^2$ can have the form $\cos^2(\omega_A t + h_t^{A'} x_t)$; this is just a special case of $s_t = A_t + B_t e^{i\theta t}$, with $A_t = B_t = \frac{1}{2}$, $\theta_t = 2(\omega_A t + h_t^{A'} x_t)$, and has been considered by Snyder[23].

In what follows, we shall confine our attention to (4.73) and (4.74), as $s_t = A_t + B_t e^{i(\omega_\theta t + h_t^{\theta\prime} x_t)}$ provides a rich fund of meaningful examples. It can be seen that if either A_t or B_t (but not both) is zero, the estimator equations yield information only about the fading process and not the message, as expected. Similarly, if H_t^f is zero, then fading is absent and we estimate only the message. Finally, if h_t^θ is zero, the effect is the same as if A_t or B_t were zero.

a. Phase Modulation: One-Dimensional Message; Diversity.

In this example we derive a demodulator for a phase-modulated signal suffering lognormal fading. D-order receiver diversity is assumed. The message, as well as the fading in each channel, is taken to be one-dimensional, and the message is assumed to be independent of the fading. We partition x_t as

$$x_t = \begin{bmatrix} \left.\begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{matrix}\right\} \text{fading} \\ x_{D+1} \leftarrow \text{message} \end{bmatrix} ;$$

also,

$$F_t = \begin{bmatrix} F_1 & & 0 \\ & \ddots & \\ 0 & & F_{D+1} \end{bmatrix} ; \quad G_t = \begin{bmatrix} G_1 & & 0 \\ & \ddots & \\ 0 & & G_{D+1} \end{bmatrix} \tag{4.76}$$

$$dx_t = F_t x_t dt + G_t d\omega_t$$

$$\omega_t = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_D \\ \omega_{D+1} \end{bmatrix} \tag{4.77}$$

Consequently,

$$H_t^f = \delta \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ 0 & & \ddots & \\ & & & 1 \\ \hline 0 & 0 & \cdots & 0 \end{bmatrix} \begin{matrix} \uparrow \\ D+1 \\ \downarrow \end{matrix} \tag{4.78}$$

$$\longleftarrow D \longrightarrow$$

$$h_t^\theta = \gamma \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \begin{matrix} \uparrow \\ D+1 \\ \downarrow \end{matrix} \tag{4.79}$$

The numbers δ and γ are proportionality factors; δ is a measure of fading intensity, and γ is a phase modulation index.

We note that $[P_t]_{j,\, D+1} = [P_t]_{D+1,\, j} = 0$, $j = 1, \ldots D$, because of the independence of the message and the fading. Omitting the intermediate algebra, we arrive at the following equations for the estimates:

$$d\hat{x}_j = F_j \hat{x}_t dt - \beta M_t^2 \delta \sum_{k=1}^{D} P_{jk} e^{\delta \hat{x}_k} dt + \delta \sum_{k=1}^{D} P_{jk} dN_k \,, \qquad (j = 1, \ldots, D) \tag{4.80}$$

$$d\hat{x}_{D+1} = F_{D+1}\hat{x}_{D+1}dt - \frac{J_t^2 \sin(\omega_0 t+\gamma\hat{x}_{D+1})}{M_t^2+J_t^2\cos(\omega_0 t+\gamma\hat{x}_{D+1})}\gamma P_{D+1,D+1}\,\|dN_t\|_1 \tag{4.81}$$

Similarly,

$$dP_{ij} = [(F_i+F_j)P_{ij}+G_iG_j]dt - \beta M_t^2\delta^2\sum_{k=1}^{D} P_{ik}P_{jk}e^{\delta\hat{x}_k}dt \tag{4.82}$$

$$(i,\ j = 1,\ \ldots,\ D)$$

$$dP_{D+1,D+1} = 2F_{D+1}P_{D+1,D+1}dt + G^2_{D+1},dt - 2\beta B_t^2\sum_{k=1}^{D} e^{\delta\hat{x}_k}\gamma^2 P^2_{D+1,D+1}\,dt \tag{4.83}$$

The latter are differential equations of the Ricatti type; because they have no explicit dependence on the data they can be solved in advance in terms of $\hat{x}_t$, and substituted into the processor equations, (4.80) and (4.81).

These equations, (4.80)-(4.83), completely specify the processing which must be performed on the observations $\{N_t,\ t\in T\}$ to get the desired estimates. Figure 4.3 illustrates this processing. It can be seen from the equations or the diagram, that the system has what might be described as an "estimator-demodulator" structure; that is the demodulator, which yields $\hat{x}_{D+1}$, incorporates the fading estimates in exactly the same way that it would incorporate the fading itself were it known perfectly. This

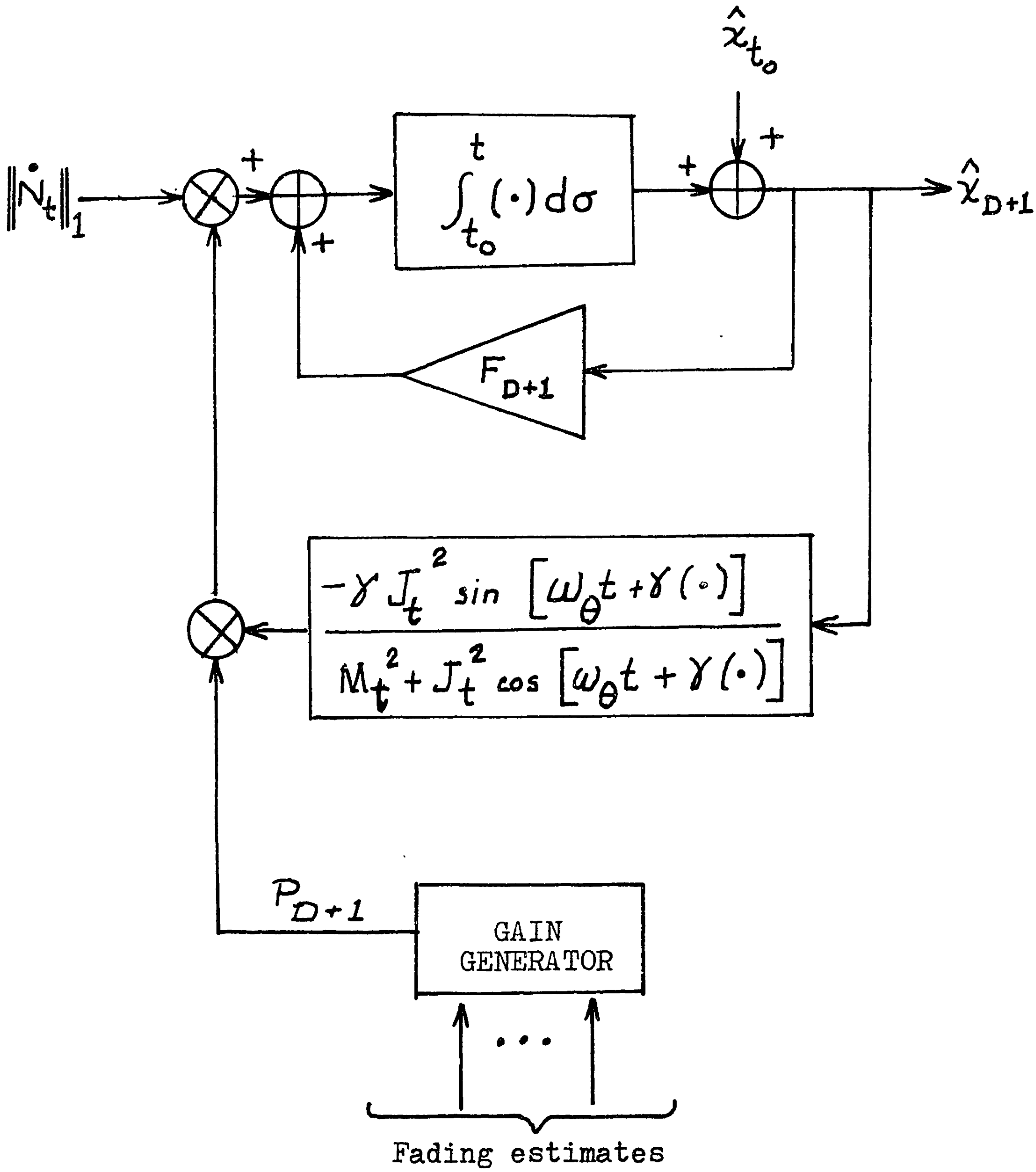

$\|\dot{N}_t\|_1$
+
+
$\int_{t_o}^{t} (\cdot)\, d\sigma$
$\hat{x}_{t_o}$
+
+
$\hat{x}_{D+1}$
F_{D+1}
$\dfrac{-\gamma J_t^2 \sin[\omega_\theta t + \gamma(\cdot)]}{M_t^2 + J_t^2 \cos[\omega_\theta t + \gamma(\cdot)]}$
P_{D+1}
GAIN GENERATOR
Fading estimates

Figure 4.3

is reminiscent of the well-known estimator-correlator structure exhibited by detectors for Gaussian random channels[13].

A number of other interesting observations consistent with intuition can be made simply by inspecting the estimator equations, (4.80)-(4.83).

1) The message estimate is influenced by the fading estimates only through $P_{D+1,D+1}$, which in turn is dependent on the fading estimates via the sum $\sum_{k=1}^{D} e^{\delta \hat{x}_k}$. This sum, upon which $P_{D+1,D+1}$ is monotonically dependent, is a measure of the total fading intensity, and acts as a gain factor in (4.81). The diversity advantage is in clear evidence, particularly when D is large and the $\{x_k\}_{k=1}^{D}$ exhibit some lack of correlation.

2) Given $P_{D+1,D+1}$, the message estimate depends on the unweighted sum of the data $\|dN_t\|_1$. On the other hand, unless the fading variables $\{x_k\}_{k=1}^{D}$ are completely correlated all the fading estimates depend upon weighted sums of the observations from the various diversity channels. Thus, given $P_{D+1,D+1}$, the demodulator can be implemented using one large photodetector, rather than D small detectors. Diversity, with its additional complexity, is needed only to obtain $P_{D+1,D+1}$. Since $P_{D+1,D+1}$ is a function of a total intensity measure, it might be possible to replace $\sum e^{\delta \hat{x}_k}$

with $\widehat{\sum e^{\delta x_k}}$ to get a reasonable "suboptimum" processor which does not use diversity. This observation is in agreement with similar ones made elsewhere relative to MAP demodulators[26].

3) The coupling among the fading estimates and among their error covariances is a function of the correlations among the $\{x_k\}_{k=1}^{D}$. If the fading variables are mutually independent, for example, then all the cross-covariances vanish, resulting in considerable simplification of the processing. For correlated fading, the coupling between the message estimate and the fading estimates is one-way; that is, $\hat{x}_{D+1}$ depends on the fading estimates, but the $\{\hat{x}_k\}_{k=1}^{D}$ are in no way influenced by $\hat{x}_{D+1}$, and can be evaluated separately. This is not unreasonable, since $\hat{x}_{D+1}$ is a phase estimate, and the $\{\hat{x}_k\}_{k=1}^{D}$ are logarithmic intensity estimates; ω_θ is assumed to be large, so relative phase should have negligible effect on estimates involving the amplitude of a sinusoid.

Up to now we have, for convenience, ignored additive background noise. To illustrate some of the effects of noise, we write out the equations for $\hat{x}_t$, omitting the messy preliminaries. The equations for P_t, although straightforward to derive, are also omitted, since they are far too complicated to be written compactly.

$$d\hat{x}_j = F_j \hat{x}_j dt - \beta M_t^2 \delta \sum_{k=1}^{D} P_{jk} e^{\delta \hat{x}_k} dt \quad +$$

$$+ \quad \delta\left[M_t^2 + J_t^2 \cos(\omega_\theta t + \gamma \hat{x}_{D+1})\right]$$

$$\cdot \sum_{k=1}^{D} P_{jk}\left[M_t^2 + J_t^2 \cos(\omega_\theta t + \gamma \hat{x}_{D+1}) + e^{-\delta \hat{x}_k}\mu_k\right]^{-1} dN_k \quad (4.84)$$

$$d\hat{x}_{D+1} = F_{D+1}\hat{x}_{D+1} - \gamma P_{D+1,D+1} J_t^2 \sin(\omega_\theta t + \gamma \hat{x}_{D+1})$$

$$\cdot \sum_{k=1}^{D} \left[M_t^2 + J_t^2 \cos(\omega_\theta t + \gamma \hat{x}_{D+1}) + e^{-\delta \hat{x}_k}\mu_k\right]^{-1} dN_k \qquad (4.85)$$

$\mu_k = [\mu_o]_k$ = kth component of μ_o

The main effect of the noise is to introduce direct coupling, in both directions, between the message equation and the fading equations, and to alter the simple dependence of $P_{D+1,D+1}$ on a total fading intensity estimate. It can be shown that the estimates have a continuous dependence on the noise; i.e., small changes in the noise induce correspondingly small changes in the estimates. Thus there is some justification, and considerable savings in complexity, for ignoring the noise μ_o in the receiver structure altogether when it is in fact known to be weak.

It is worthwhile to elaborate a bit further on the effects of noise, when it is accounted for in the processing. First, it is clear that large noise in a particular channel tends to deemphasize the data arriving from that channel.

This effect acts in concert with deep fading, so that a combination of the two strongly inhibits the data. Conversely, "favorable" fading ($\hat{x}_k$ large) tends to cancel the effects of large noise. Indeed, in both the processor and variance equations, the noise appears only in a "noise-to-fading" ratio, $\mu_k / e^{\delta \hat{x}_k}$, so that the absolute strength of the noise is, in a sense, irrelevant.

Another effect of noise is to introduce, in the equation for $\hat{x}_{D+1}$, a variable weighting of the observations from the various diversity channels, as well as an explicit dependence on the $\{\hat{x}_k\}_{k=1}^{D}$. Unless fading is absent altogether, this additional complexity is not diminished by assuming that all the μ_k are the same. This weighting is intuitively plausible, as we would always wish to deemphasize data from <u>relatively</u> noisy channels.

b. Performance Analysis of a PM System.

For this example, P_t provides an approximate performance measure for the estimating system. With regard to the message estimate $\hat{x}_{D+1}$, it is of interest to know other factors relative to performance, such as cycle skipping and threshold* behavior. This behavior can often be deduced from

*Threshold denotes the onset of relatively severe performance degradation resulting from excessive noise and/or weak signal. It can be shown that threshold behavior is always possible in nonlinear systems[101].

the p.d.f. of the estimator error $\tilde{x}_{D+1} = x_{D+1} - \hat{x}_{D+1}$. To this end we now attempt to apply Theorem 3.8 to our example.

To render the problem tractable, we assume that D = 1 and $H_t^f = 0$ i.e., no diversity and no fading. We further assume that A_t and B_t are constant in time, and that the one-dimensional Gauss-Markov message satisfies

$$dx_t = -k x_t dt + \sqrt{2k}\, dw_t \tag{4.86}$$

with $\{w_t, t \in T\}$ a standard Wiener process. The variance of x_t has been normalized to unity ($P_0 = 1$). For the moment we place no restrictions on the magnitude of the additive noise μ_0.

The estimator equations are special cases of (4.81) and (4.83), with M^2 replaced by $M^2 + \mu_0$ and $2B^2$ by (4.75):

$$d\hat{x}_t = -k\hat{x}_t dt - \frac{\gamma P_t J^2 \sin(\omega_0 t + \gamma \hat{x}_t)}{M^2 + \mu_0 + J^2 \cos(\omega_0 t + \gamma \hat{x}_t)} dN_t \tag{4.87}$$

$$dP_t = -2kP_t dt + 2k dt - \beta \gamma^2 (M^2 + \mu_0)$$

$$\cdot \left(1 - \sqrt{1 - \left(\frac{J^2}{M^2+\mu_0}\right)^2}\,\right) P_t^2 dt \tag{4.88}$$

Before embarking on a study of anomalous behavior, let us first comment briefly about performance above threshold.

In this region, we can assume that the conditions under which the estimator equations were derived (i.e., small error) are satisfied; then P_t closely describes performance. If steady-state operation is assumed, then $P_t = V$ = constant, $dP_t/dt = 0$, and we can solve for V as

$$V = \frac{2}{1+\sqrt{1+\gamma^2\delta}} \tag{4.89}$$

where

$$\delta = \frac{2\beta U}{k} \tag{4.90}$$

$$U = (M^2+\mu_0)\left(1-\sqrt{1-\left(\frac{J^2}{M^2+\mu_0}\right)^2}\right) \tag{4.91}$$

Thus the variance of the total phase error $\varphi = \gamma(x_t - \hat{x}_t)$ is

$$\sigma_\varphi^2 = \frac{2\gamma^2}{1+\sqrt{1+\gamma^2\delta}} \tag{4.92}$$

It is suggestive to call δ an effective "signal-to-noise ratio in the message bandwidth," by comparison with Snyder's[14] phase error results for the classical, additive, white Gaussian noise channel. For that channel, the observations are

$$r_t = C\sin(\omega_0 t + \gamma x_t) + n_t$$

with n_t a white Gaussian process of spectral height N_0 watts/Hz. The steady-state, above-threshold error variance is given by (4.92), with δ replaced by C^2/kN_0, the SNR in the message bandwidth. Thus, when both systems are operating above threshold, our phase demodulator performs as well as Snyder's in terms of error variance if

$$\frac{C^2}{N_0} = 2\beta U \tag{4.93}$$

Let us illustrate this with an example. Suppose $\mu_0 = 0$ and $A^2 = B^2$; then $U = 2B^2$, and B^2 is the average transmitted signal energy. Let C^2, the transmitted energy in the white noise channel, be equal to B^2. Then the above-threshold performance of the two demodulators is the same if the noise in the Gaussian channel is

$$N_0 = \frac{1}{4\beta} = \frac{hf_0}{4\eta} \tag{4.94}$$

It can be verified from (4.91) that the quantity U has a maximum value of $2B^2$, achieved when $A^2 = B^2$ and $\mu_0 = 0$. For any other values of A, B, and μ_0, the value of N_0 required to equalize the above-threshold performance of the two demodulators will exceed $hf_0/4\eta$. In other words, for a fixed transmitted energy $\mathcal{E}$, performance of our Poisson demodulator is lower-bounded by

$$2r^2\Big/\left(1+\sqrt{1+\frac{4r^2\eta\mathcal{E}}{khf_0}}\right) \tag{4.95}$$

This bound is achieved only when $A^2 = B^2 = \mathcal{E}$ and $\mu_0 = 0$.

To get an accurate indication of performance near and below threshold, σ_φ^2 as given by (4.92) is no longer adequate; we must solve for $p_{\tilde{x}_t}(\tilde{X})$, the p.d.f. of the error $x_t - \hat{x}_t$, and evaluate

$$\sigma_\varphi^2 = \gamma^2 \int_{-\infty}^{\infty} \tilde{X}^2 p_{\tilde{x}_t}(\tilde{X}) d\tilde{X} \tag{4.96}$$

Defining

$$Q_t = (M^2 + \mu_0)/\gamma P_t J^2 ; \qquad R_t = (\gamma P_t)^{-1} ;$$

the equation for $p_{\tilde{x}_t}(\tilde{X})$ takes the form (see Theorem 3.8):

$$\frac{\partial}{\partial t} p_{\tilde{x}_t}(\tilde{X}) = \frac{\partial}{\partial \tilde{X}}\left[k\tilde{X} p_{\tilde{x}_t}(\tilde{X})\right] + k \frac{\partial^2}{\partial \tilde{X}^2} p_{\tilde{x}_t}(\tilde{X})$$

$$+ E\Big\{\left[\beta(M^2 + \mu_0) + \beta J^2 \cos(\omega_0 t + \gamma x_t)\right]$$

$$\cdot \left[p_{\tilde{x}_t}\left(\tilde{X} - \frac{\sin(\omega_0 t + \gamma \hat{x}_t)}{Q_t + R_t \cos(\omega_0 t + \gamma \hat{x}_t)}\right) - p_{\tilde{x}_t}(\tilde{X})\right]\Big|_{\tilde{x}_t = \tilde{X}}\Big\} \tag{4.97}$$

The solution of this equation exactly describes the performance of the demodulator over all ranges of operation. The "design value" of error variance, given by (4.89), is used in (4.97) when it is desired to evaluate steady-state system performance in the region where (4.89) is no longer valid; i.e., at or below threshold. The presence of the last term in (4.97) makes the equation exceedingly difficult to solve. However, by assuming steady-state operation and replacing P_t with V (4.89), we can obtain an approximate solution.

Consider the equation for the characteristic function of the total phase error φ, obtained by transforming (4.97):

$$0 = -iuk \frac{d}{d(iu)} M_\varphi(iu) + k\gamma^2 (iu)^2 M_\varphi(iu)$$

$$+ E\Big\{E\Big[\big(\beta(m^2+\mu_0) + \beta\sigma^2 \cos(\omega_0 t + \gamma x_t)\big)$$

$$\cdot \left(\exp \frac{iu\gamma \sin(\omega_0 t + \gamma \hat{x}_t)}{Q + R\cos(\omega_0 t + \gamma \hat{x}_t)} - 1\right) \Big| \tilde{x}_t \Big] e^{iu\gamma \tilde{x}_t} \Big\} \quad (4.98)$$

The l.h.s. of this equation is zero by virtue of the steady-state assumption. To be consistent with this assumption, the r.h.s. cannot be explicitly dependent on time. Thus we must isolate and retain only the slowly varying components. This can be accomplished by using the expansion,

$$\exp \frac{iu\gamma \sin(\omega_0 t + \gamma \hat{x}_t)}{Q + R\cos(\omega_0 t + \gamma \hat{x}_t)} - 1 = \sum_{n=1}^{\infty} \frac{(iu)^n}{n!} \left[\frac{\gamma \sin(\omega_0 t + \gamma \hat{x}_t)}{Q + R\cos(\omega_0 t + \gamma \hat{x}_t)}\right]^n \quad (4.99)$$

We multiply this by $\beta(M^2 + \mu_0) + J^2\beta\cos(\omega_0 t + \gamma x_t)$ and retain those terms which are constant or depend only on $\varphi = \gamma(x_t - \hat{x}_t)$. In general there are infinitely many such terms. For convenience we use only the first two terms in the series above. This is reasonable if

$$\left| \frac{\gamma \sin \theta}{Q + R \cos \theta} \right| \ll 1, \text{ all } \theta \tag{4.100}$$

The trigonometric expression in (4.100) has a maximum magnitude given by

$$\frac{\gamma / Q}{\sqrt{1 - R^2/Q^2}} \tag{4.101}$$

Thus it is sufficient to require

$$\frac{\gamma / Q}{\sqrt{1 - R^2/Q^2}} \ll 1 \tag{4.102}$$

i.e.,

$$\sigma_\varphi^2 \frac{\dfrac{J^2}{M^2 + \mu_0}}{\sqrt{1 - \left(\dfrac{J^2}{M^2 + \mu_0}\right)^2}} \ll 1 \tag{4.103}$$

We now have

$$\exp \frac{iu\gamma \sin(\omega_0 t + \gamma \hat{x}_t)}{Q + R\cos(\omega_0 t + \gamma \hat{x}_t)} - 1 = \tag{4.104}$$

$$\cong iu \frac{\gamma \sin(\omega_\theta t + \gamma \hat{x}_t)}{Q + R\cos(\omega_\theta t + \gamma \hat{x}_t)} + \frac{1}{2}(iu)^2 \frac{\gamma^2 \sin^2(\omega_\theta t + \gamma \hat{x}_t)}{(Q + R\cos(\omega_\theta t + \gamma \hat{x}_t))^2}$$

The first term on the r.h.s. is an odd function of $\omega_\theta t + \gamma \hat{x}_t$ so we can expand it in a Fourier sine series on $(-\pi, \pi)$, the first term of which can be shown to be

$$iu \frac{2\gamma^2 VU}{J^2} \sin(\omega_\theta t + \gamma \hat{x}_t)$$

Similarly, the second term on the r.h.s. of (4.104) is an even function of $\omega_\theta t + \gamma \hat{x}_t$, and so admits a cosine expansion; the first term of this series is the constant,

$$\frac{(iu)^2}{2} \frac{\gamma^4 V^2 U}{(M^2 + \mu_0)\sqrt{1 - \frac{R^2}{Q^2}}}$$

and the second is

$$\frac{(iu)^2}{2} \left[\frac{2\gamma^4 V^2 U}{J^2} - \frac{2\gamma^4 V^2 U}{J^2 \sqrt{1 - R^2/Q^2}} \right] \cos(\omega_\theta t + \gamma \hat{x}_t)$$

We can ignore higher order harmonics because no slowly-varying quantities result from them when we multiply by $\beta(M^2 + \mu_0) + \beta J^2 \cos(\omega_\theta t + \gamma x_t)$. Thus, ignoring higher harmonics,

$$\exp \frac{iu \gamma \sin(\omega_\theta t + \gamma \hat{x}_t)}{Q + R\cos(\omega_\theta t + \gamma \hat{x}_t)} - 1 \quad \cong$$

$$= \tfrac{1}{2}(iu)^2 \frac{\gamma^4 V^2 U}{(M^2+\mu_0)\sqrt{1-R^2/Q^2}}$$

$$+ iu \frac{2\gamma^2 V U}{J^2} \sin(\omega_\theta t + \gamma \hat{x}_t)$$

$$+ \tfrac{1}{2}(iu)^2 \left[\frac{2\gamma^4 V^2 U}{J^2} - \frac{2\gamma^4 V^2 U}{J^2\sqrt{1-R^2/Q^2}}\right] \cos(\omega_\theta t + \gamma \hat{x}_t) \qquad (4.105)$$

Carrying out the multiplication, we find the slowly varying component to be

$$\tfrac{1}{2}(iu)^2 \frac{\beta\gamma^4 V^2 U}{\sqrt{1-R^2/Q^2}} - iu\beta\gamma^2 VU \sin\varphi$$

$$+ \tfrac{1}{2}(iu)^2 \left[\beta\gamma^4 V^2 U - \frac{\beta\gamma^4 V^2 U}{\sqrt{1-R^2/Q^2}}\right] \cos\varphi \qquad (4.106)$$

It is consistent with our previous assumptions to take $R^2/Q^2 \ll 1$; this reduces the expression above to the simpler form,

$$\tfrac{1}{2}(iu)^2 \beta\gamma^4 V^2 U - iu\beta\gamma^2 VU \sin\varphi \qquad (4.107)$$

Substituting this in the conditional expectation in (4.98), we obtain the following differential equation for the characteristic function of the phase error :

$$0 = -iu \frac{d}{d(iu)} M_\varphi(iu) + k\gamma^2 (iu)^2 M_\varphi(iu) -$$

$$-iu\beta\gamma^2 VUE\{\sin\varphi\, e^{iu\varphi}\} + \tfrac{1}{2}(iu)^2\beta\gamma^4 V^2 U M_\varphi(iu) \qquad (4.108)$$

The corresponding equation for the p.d.f. of the steady-state phase error is easily found from this to be

$$0 = \frac{d}{d\varphi}\left[\left(\varphi + \frac{\gamma^2\mathcal{S}}{1+\delta}\sin\varphi\right)p(\varphi)\right] + \frac{2\gamma^2\delta}{1+\delta}\,\frac{d^2}{d\varphi^2}p(\varphi) \qquad (4.109)$$

where we have combined some terms and defined $\delta = \sqrt{1+\gamma^2\mathcal{S}}$.

This is identical to the phase-error equation obtained by Snyder[14] for the classical, additive, white Gaussian noise channel. Again, as in (4.92), $\mathcal{S}$ is an effective signal-to-noise ratio in the message bandwidth.

Integrating (4.109), we obtain

$$p(\varphi) = \frac{1}{C}\exp\left\{-\frac{1+\delta}{4\gamma^2\delta}\varphi^2 + \frac{\mathcal{S}}{2\delta}\cos\varphi\right\} \qquad (4.110)$$

where

$$C = 2\gamma\left(\frac{\pi\delta}{1+\delta}\right)^{1/2}\sum_{n=-\infty}^{\infty} I_n\left(\frac{\mathcal{S}}{2\delta}\right)\exp\left[-\frac{\delta\gamma^2}{1+\delta}n^2\right] \qquad (4.111)$$

and I_n is the modified Bessel function of the first kind and the n^{th} order. This p.d.f., illustrated for various values of $\mathcal{S}$ and γ in Figure 4.4, gives the steady-state probability that the phase error φ will be found in any

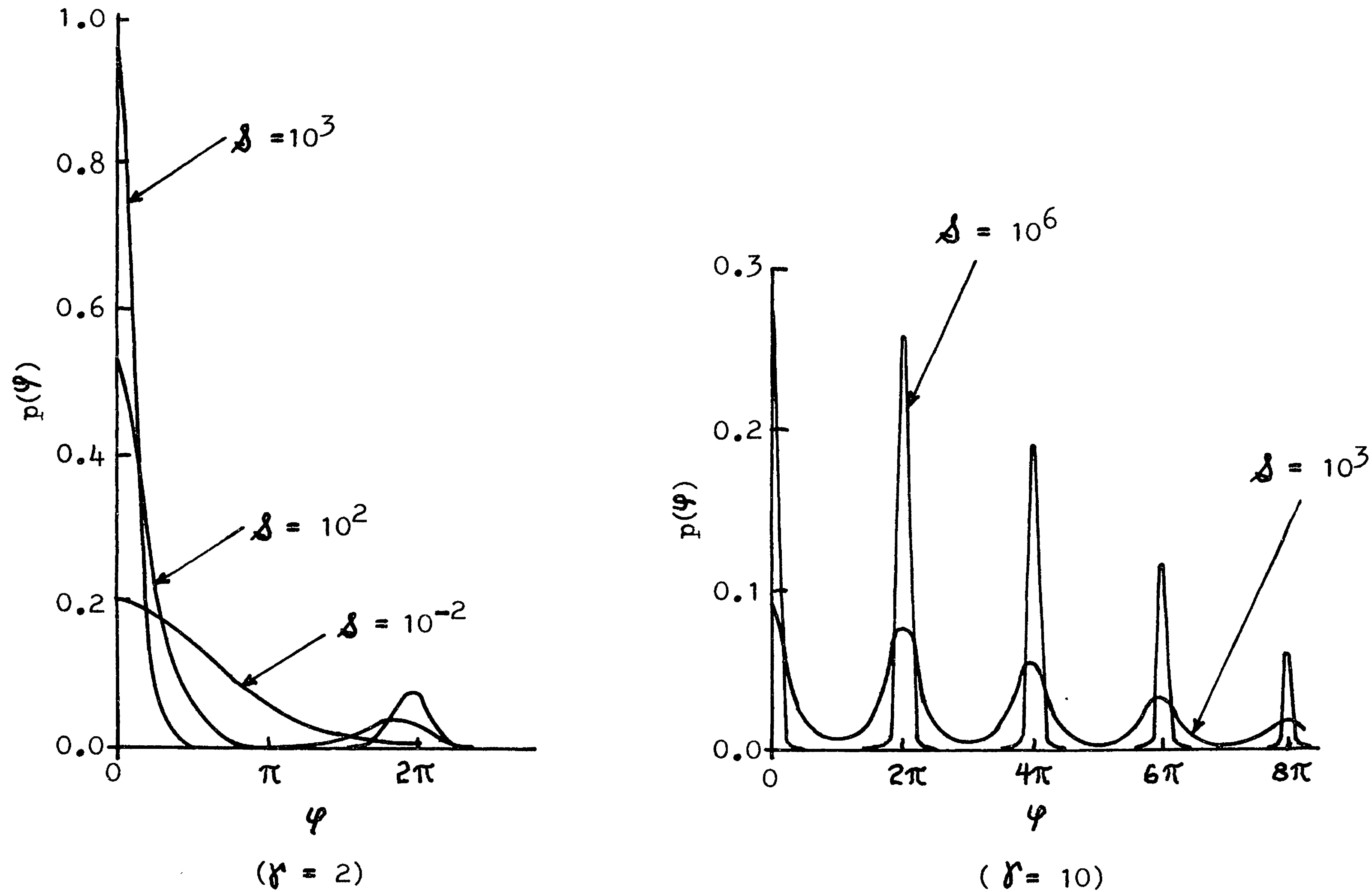


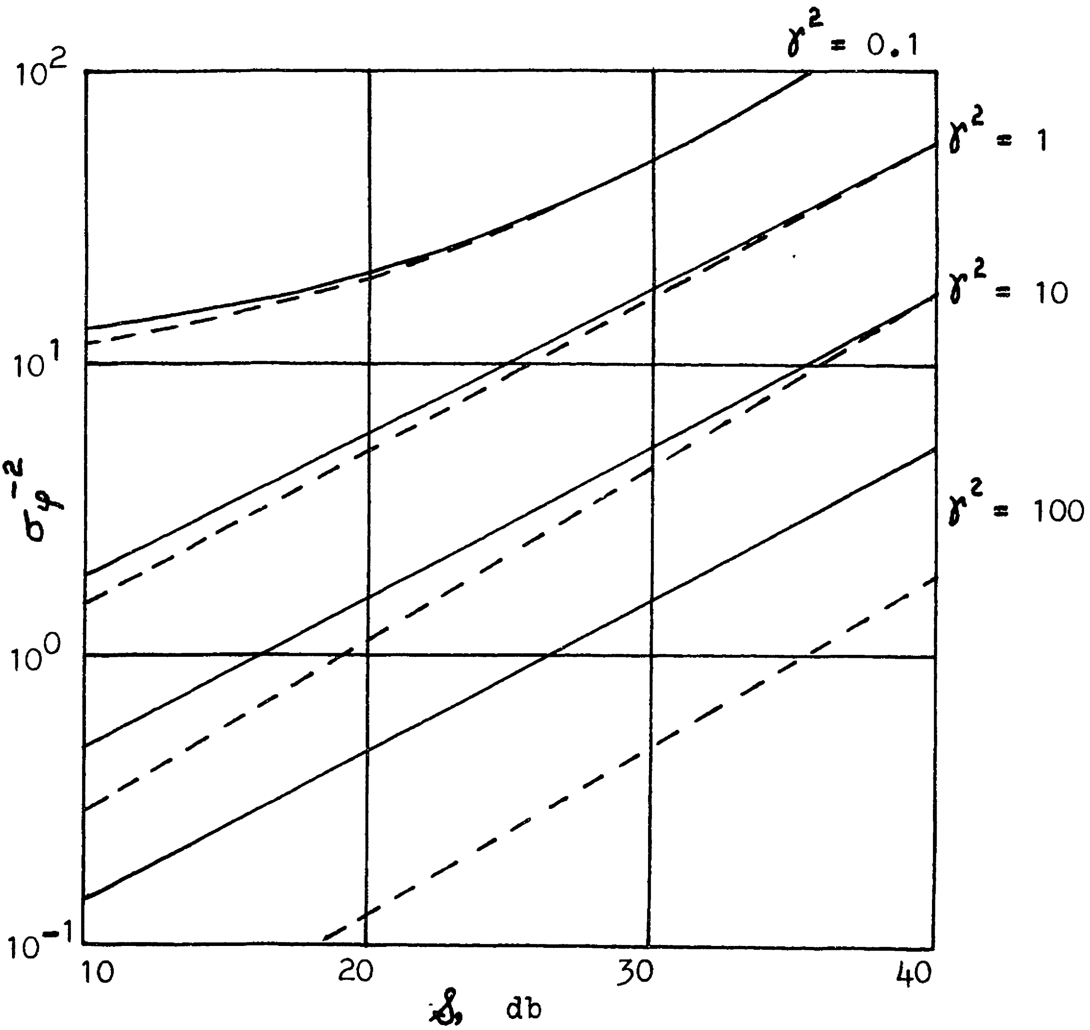
Figure 4.4 (from Ref. 14)

given interval. By its multimodal character, it indicates that φ is likely to be near an integral multiple of 2π; in particular, for small γ and large δ, the error φ is likely to be near 0. Also, the envelope of $p(\varphi)$ is normal with variance $2\gamma^2\delta/(1+\delta)$; thus for small enough γ, $p(\varphi)$ has only one lobe; i.e., no cycle skipping or threshold behavior occurs. Note that φ is measured with respect to zero phase error in a relative sense. Since phase can only be measured modulo-2π, the zero phase error with respect to which φ is measured is the zero associated with the most recent occurrence of phase lock. For example, suppose that the demodulator skips one cycle ($\pm 2\pi$), and regains lock without skipping back a cycle. Then at some later time, after steady-state has been achieved, the relative phase error in (4.110) is referred to the lock obtained after the cycle-skip.

We have computed the error variance using

$$\sigma_\varphi^2 = \int_{-\infty}^{\infty} \varphi^2 p(\varphi) d\varphi$$

This is plotted for representative parameter values in Figure 4.5, along with the above-threshold form of σ_φ^2, (4.92). It is assumed that (4.103) is satisfied. As predicted, there is negligible threshold behavior for small γ. It becomes very pronounced, however, as γ increases; for $\gamma = 10$, (4.92) is a wholly inadequate measure of performance, off by more than 10 db over the entire range of δ, 10-40 db.



Performance of Phase Demodulator

——————— = above-threshold form

– – – – – – – = exact form

Figure 4.5

The error variance can also be obtained analytically by differentiating (4.108) twice with respect to $(i\omega)$, and setting $i\omega$ equal to zero. This procedure yields the equation,

$$0 = -2k\sigma_\varphi^2 + 2k\gamma^2 + \beta\gamma^4 V^2 U - 2\beta\gamma^2 VU\, E\{\varphi \sin\varphi\} \qquad (4.112)$$

valid above and below threshold. Above threshold, (4.112) gives a phase error variance very close to (4.92), as we would expect.

Recall that the primary assumptions required to obtain (4.108), and thus (4.109)-(4.112), were

1) steady-state operation;

2) the inequality (4.103).

If 1) does not hold, we can make essentially no progress with the error equation, (4.97). If 1) holds, but not necessarily 2), then we must retain higher order terms in the Taylor expansion of the complex exponential, (4.99). We can still isolate the slowly varying parts of each of the terms but, because of the presence of higher-than-quadratic powers of $(i\omega)$, the end result will be an n^{th}-order differential equation for the phase-error p.d.f., $n > 2$. This equation will be of the form,

$$0 = \frac{d}{d\varphi}\left[\left(\varphi + \frac{\gamma^2 \delta}{1+\delta}\sin\varphi\right)p(\varphi)\right] +$$

$$+\sum_{n=2}^{N}[a_n+b_n\cos\varphi+c_n\sin\varphi]\frac{d^n}{d\varphi^n}p(\varphi), \quad (N>2) \tag{4.113}$$

and might have to be solved numerically.

Condition 2) of the previous paragraph is essentially a restriction on $\beta J^2/k$, the average number of signal photoelectrons per phase correlation-time. As a numerical example, suppose that $J^2/(M^2+\mu_0) = \frac{1}{4}$ and $\gamma^2 = 1$. Then (4.103) can be evaluated approximately as

$$(1/2)\Big/\left(1+\sqrt{1+\tfrac{1}{4}\left(\tfrac{\beta J^2}{k}\right)}\right) \ll 1$$

If the value of 0.1 for the l.h.s. of the inequality above is considered adequate, then it can be seen that $\beta J^2/k$ must be at least 60 for the inequality to be satisfied. This is a satisfying result, because we intuitively expect the demodulator to perform classically when the phase process changes very slowly between photoelectron arrivals. In general, we observe that if γ is sufficiently small and $J^2/(M^2+\mu_0) \ll 1$, then (4.110) is applicable without regard for the size of $\beta J^2/k$.

If $M^2 = J^2$ and $\mu_0 = 0$, then (4.103) cannot be satisfied; as a consequence, (4.110) does not apply. The more general form (4.113) is also ruled out, because

$$\left[\frac{\gamma\sin(\omega_0 t+\gamma\hat{x}_t)}{Q+R\cos(\omega_0 t+\gamma\hat{x}_t)}\right]^n = \left[\gamma^2 V\tan\tfrac{1}{2}(\omega_0 t+\gamma\hat{x}_t)\right]^n$$

does not possess a well defined Fourier series expansion unless n = 1. However, the function

$$\exp iu\,\gamma^2 V \tan \tfrac{1}{2}\hat{\theta}_t$$

admits a Fourier exponential series. Thus we can proceed as before to isolate slowly varying components and find the differential equation for the total phase error. We omit this result here, except to remark that it is not a "singular" result; i.e., it yields a phase error variance which is consistent above threshold with (4.92), and below threshold with the error variance obtained from (4.113).

c. Comparative Analysis of an Ad Hoc PM System.

The estimator equation for our PM demodulator is given by (4.87). Noting that it is quite similar in appearance to a classical PM demodulator for additive Gaussian channels[14], one might logically ask the following question: How would performance suffer if one, knowing nothing of optimum Poisson demodulators, used a classically-designed system in the optical receiver? The estimator and variance equations for such a system are obtained by taking

$$\lambda_t = \beta\,[J^2\cos(\omega_0 t + \gamma x_t) + M^2 + \mu_0]$$

to be the "received signal," analogous to the classical received signal,

$$r_t = C\sin(\omega_0 t + \gamma x_t) + n_t. \tag{4.114}$$

The estimator equations for the message (4.86) and the observations (4.114) are[14]

$$d\hat{x}_t = -k\hat{x}_t\,dt + \frac{\gamma C}{N_0} P_t \cos(\omega_0 t + \gamma \hat{x}_t)\,dr_t \tag{4.115}$$

$$dP_t = -2kP_t\,dt + 2k\,dt - \frac{\gamma^2 C^2}{2N_0} P_t^2\,dt \tag{4.116}$$

To obtain the corresponding equations for our ad hoc system, we compare λ_t and r_t, and deduce the replacements;

$$C \rightarrow \beta J^2$$

$$N_0 \rightarrow \beta(M^2 + \mu_0)$$

$$dr_t \rightarrow dN_t$$

$$\cos(\omega_0 t + \gamma\hat{x}_t) \rightarrow -\sin(\omega_0 t + \gamma\hat{x}_t)$$

The minus sign in the last replacement arises from the fact that the cosine in (4.115) resulted from differentiating the sine in (4.114). The ad hoc estimator equations can now be written,

$$d\hat{x}_t = -k\hat{x}_t\,dt - \frac{\gamma P_t J^2}{M^2 + \mu_0} \sin(\omega_0 t + \gamma\hat{x}_t)\,dN_t \tag{4.117}$$

$$dP_t = -2kP_t\,dt + 2k\,dt - \frac{\beta\gamma^2 J^4 P_t^2}{2(M^2 + \mu_0)}\,dt \tag{4.118}$$

The steady-state solution $P_t = V_c$ (c = "classically-designed") of (4.118) is

$$V_c = \frac{2}{1 + \sqrt{1 + \gamma^2 \delta_c}} \tag{4.119}$$

where

$$\delta_c = \frac{\beta J^4}{k(M^2 + \mu_0)} \tag{4.120}$$

is the signal-to-noise ratio in the message bandwidth. δ_c has a physically meaningful form[46]; i.e., it is the ratio of signal intensity squared and k times a noise intensity. Again, we emphasize that (4.119) is a good indicator of performance only above threshold.

To obtain performance below threshold, we needn't carry out a detailed analysis of the demodulator specified by (4.117) and (4.118), because it is just a special case of the Poisson demodulator we have been studying, (4.87) and (4.88). Eqs. (4.117) and (4.118) follow from (4.87) and (4.88) by assuming $J^2/(M^2 + \mu_0) \ll 1$. Thus we conclude that the classically designed PM demodulator performs as well as the demodulator (4.87) as long as $J^2/(M^2 + \mu_0) \ll 1$. If the stronger condition (4.103) is satisfied, then performance will be closely described by (4.110). If $J^2/(M^2 + \mu_0) \ll 1$, but (4.103) does not necessarily hold, then (4.113) will have to be solved to obtain performance.

If $\mu_0 = 0$ and $M^2 = J^2$, the classically designed receiver differs notably from that designed specifically for Poisson data; viz.,

Classically designed:

$$d\hat{x}_t = -k\hat{x}_t dt - \gamma P_t \sin(\omega_0 t + \gamma \hat{x}_t) dN_t \tag{4.121}$$

$$dP_t = -2kP_t dt + 2k dt - \frac{1}{2}\beta\gamma^2 J^2 P_t^2 dt \tag{4.122}$$

Poisson designed:

$$d\hat{x}_t = -k\hat{x}_t dt - \gamma P_t \tan\frac{1}{2}(\omega_0 t + \gamma\hat{x}_t) dN_t \tag{4.123}$$

$$dP_t = -2kP_t dt + 2k dt - \beta\gamma^2 J^2 P_t^2 dt \tag{4.124}$$

Using (4.119) and (4.89) we see that the steady-state solutions of (4.122) and (4.124) are given respectively by

$$V_c = \frac{2}{1+\sqrt{1+\frac{\gamma^2\beta J^2}{k}}} \tag{4.125}$$

$$V = \frac{2}{1+\sqrt{1+\frac{2\gamma^2\beta J^2}{k}}} \tag{4.126}$$

Thus, for equal performance above threshold, the system designed specifically for Poisson data requires 3 db less signal power than the classically designed system. We have seen that the systems perform identically above and below threshold when $J^2/(M^2 + \mu_o) \ll 1$, so we conclude that 3 db represents the maximum difference between them.

Although 3 db apart above threshold, it can be shown by use of Th. 3.8 that (4.121) and (4.123) perform identically well below threshold. Hence, in all ranges of operation, the classically designed demodulator (4.117) performs within 3 db of the demodulator (4.87) designed specifically for the data $\{N_t, t \in T\}$.

The conclusion to be drawn from this analysis is that, if there are "sufficiently many" photoelectrons relative to the phase coherence time (i.e., if (4.103) is satisfied), then a classically designed phase demodulator will perform as well as the demodulator (4.87), and its behavior can be analyzed classically (i.e., Theorem 3.8 is not required).

We can also draw some conclusions about signal design, since(4.95) lower bounds performance. It is clear from the discussion preceding (4.95) that, for a fixed background μ_o, it is desirable to maximize the transmitted signal energy and to require unity modulation index; i.e., $m = J^2/M^2 = 1$. When such optimal signaling is used, it cannot be established in general whether or not the demodulator is behaving classically. This will depend, according to (4.103), on the phase deviation parameter γ, and on the relative magnitudes of the signal and background noise energies.

In the last two sections we have studied the performance of two classes of PM demodulators: those designed specifically for the data $\{N_t, t \in T\}$, and those adapted from the classical Gaussian channel PM problem. We found that behavior depended on a number of different conditions. To tie all this together and facilitate future reference, we have indexed the equations governing the various systems in Table 4.1.

PM on a Subcarrier

Demodulator	Effective SNR	Phase Error Variance Above Thresh.	Phase Error Variance Everywhere	Error p.d.f./cf
(4.87)-(4.88); (4.103) satisfied	(4.90)-(4.91)	(4.92) Fig. 4.5	(4.112) Fig. 4.5	(4.110)-(4.108)
(4.87)-(4.88); (4.103) not satisfied	(4.90)-(4.91)	(4.92) Fig. 4.5	-	(4.113)
(4.117)-(4.118) (4.103) satisfied	(4.120) or (4.90)-(4.91)	(4.92)or(4.119) Fig. 4.5	(4.112) Fig. 4.5	(4.110)-(4.108)
(4.117)-(4.118) (4.103) not satisfied	(4.120)	(4.119)	-	(4.113)
Same as above, plus $\frac{J^2}{M^2 + \mu_0} \ll 1$	(4.120) or (4.90)-(4.91)	(4.92)or(4.119) Fig. 4.5	-	(4.113)

Table 4.1

F. Summary.

In this chapter, we have shown how to apply the estimation theory of Ch. III to practical problems in optical communication. Two classes of problems were examined: channel measurement, and the demodulation of phase-modulated subcarriers.

For the first problem, a lognormally-fading multiplicative channel was assumed. In estimating the fading parameter approximate nonlinear equations were easily obtained, with the convenient feature .t the variance equation was uncoupled from the data. This simplification did not occur when additive channel noise was accounted for in the estimator. In the limit of very slow fading and small a priori variance, it was found that a linear estimator performed very close to the Cramer-Rao bound. Asymptotically ($t \rightarrow \infty$), additive channel noise had no effect on linear estimator performance.

As a communication example, we considered the problem of estimating a Gaussian message which modulated the phase of a subcarrier signal. Using the approximations of III.F.1, non-linear estimators were obtained which allowed for both multidimensional lognormal fading and multidimensional messages. The effects of fading and additive noise were discussed for a particular example, and the probability density function of the estimator error was evaluated for the case of additive noise but no fading. The phase demodulator for this case was found to behave much as a classical phase-lock loop for a wide range of parameter values. For comparison, a practical ad hoc demodulator was proposed, and was shown to perform within 3 db of the systematically-designed "quasi-optimum" demodulator over its range of operation.

CHAPTER V

Conclusions

We have investigated in this thesis the problem of statistical estimation for doubly-stochastic Poisson processes. Statistical and analytical properties of these and other point processes were discussed, and a stochastic calculus appropriate for their manipulation was developed.

With these tools we were able to obtain a representation of any MMSE estimator as an explicit functional of the observations, via a particular stochastic integral. This representation yielded estimators for many problems of interest, including prediction, filtering, smoothing, and parameter estimation.

A statistical characterization, in the form of a joint p.d.f. of event times, was derived for doubly-stochastic Poisson processes. This p.d.f. is identical in form to that of an inhomogeneous Poisson process, with the intensities replaced by their least-squares causal estimates. This characterization was found to provide a convenient approach to many detection and estimation problems. In particular, it yielded directly a likelihood ratio for a general binary problem in terms of the MMSE estimate of the intensity process. It was seen that this likelihood preserved the form of the well-known Reiffen-Sherman detector[37] originally derived for Poisson processes.

Various approximations for estimators were developed. A truncated expansion in powers of the estimator error was seen to yield practical systems which were both reasonable for most non-linear modulation schemes, and relatively simple to implement. The "Gaussian second-order" approximation was judged to be particularly convenient, and was used for all the examples in Chapter IV. For many problems, a linear estimator is adequate, and often performs nearly optimally. A general linear estimator, whose form is observation-independent, was developed. This estimator was applied to a lognormal fading problem in Chapter IV, and found to be close to optimum for weak fading.

Bounds on the mean-square error of an estimator as well as various differential and integral forms, were obtained. A much stronger result -- a partial differential-difference equation for the p.d.f. of the actual estimator error -- was derived in Theorem 3.8. Valid for most approximate estimators and many exact ones, this equation can be solved for many problems of interest; e.g., narrowband systems in steady-state operation. The validity of the theorem, when it applies, does not depend on the approximations used to derive the estimators.

Applications in optical communications were discussed in Chapter IV. Models for turbulent atmospheric channels, ideal photodetectors, and real communication systems were proposed. Several realistic examples were investigated in detail, including the estimation of lognormal channel fading, both with and without background noise; the simultaneous estimation of correlated multidimensional lognormal fading and Gauss-Markov message; and the demodulation of a phase-modulated subcarrier in the presence of background noise. For the latter example a system implementation was determined, and performance was obtained for all ranges of operation, both above and below threshold. It was found that the demodulator operated much as a classical phase-lock loop over a wide range of parameter values. For weak signals, this was no longer the case; operating near the quantum limit, the system could no longer be described classically.

Much research remains to be done in the future. Based on the results of this thesis, as well as on previous results, the following recommendations are made:

1. The representation theorem, 3.2, although valid for a wide class of problems, was applied primarily to problems of causal filtering. Further results in prediction and smoothing would do much to broaden the theory of estimation for point processes.
2. The Taylor expansion approximations we have used so extensively are <u>ad hoc</u> in nature, and can only be evaluated in particular applications. A general appraisal of these approximations, as well as the determination of different and

perhaps better approximations, would be most useful.

3. We have studied only a small selection of examples in optical communication. The theory would benefit by being tested in other applications. For example, a detailed analysis of an FM subcarrier system would be of interest.

4. Any techniques which could be brought to bear on the analytical or numerical solution of differential-difference equations, would add to the value of Theorem 3.8. It would also be useful to generalize this theorem to a wider class of processes and estimators.

5. We have made little use in this thesis of Theorem 3.5, the joint p.d.f. of the event times of a doubly-stochastic Poisson process. This result appears to have wide applicability in the theory of point processes, and should be investigated further.

Appendix A -- Some Definitions from Probability and Stochastic Processes.

In this appendix we review some well-known definitions and concepts which find frequent use in the body of the thesis. Our discussion covers considerably less than the barest rudiments of a minimal probability theory; thus the reader who feels uneasy with this skeletal summary can profit by consulting the several excellent texts from which it was constructed: Doob[1], Skorokhod[2], Gikhman and Skorokhod[3], Neveu[94], and Wong[77]. Wong's recent book is especially clear on the points which are relevant to this thesis.

Throughout this appendix, Ω denotes an arbitrary set of points ω.

1. Algebras of Point Sets

A class $\mathcal{F}$ of subsets of Ω is called an algebra or field of sets of Ω if (i) $A \in \mathcal{F}$ and $B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$; (ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$. That is, an algebra is a class of sets closed under all finite Boolean operations. Denoting the empty set by $\emptyset$, we see that every algebra contains both $\emptyset$ and Ω. Indeed, the set $B_0 = \{\emptyset, \Omega\}$ is itself an algebra, as can easily be verified.

An algebra of sets is called a σ-algebra if it is closed under all countable Boolean operations. The sets in a σ-algebra are said to be <u>σ-measurable</u>. Intersections of arbitrarily many σ-algebras of subsets of Ω is again a σ-algebra of subsets of Ω. Therefore, given $G \in \Omega$, there is a smallest σ-algebra $B(G)$ such that $G \subset B(G)$. This is so because we can take $B(G)$ to be the intersection of all σ-algebras containing G, and there is at least one such intersection, viz., all subsets of Ω. The minimal σ-algebra $B(G)$ is said to be the σ-algebra <u>generated</u> by G.

Let us illustrate these ideas with a simple example. Let $\Omega = [0,1)$: i.e., $0 \leq \omega < 1$. The set

$$A = \{\Omega,\ \emptyset,\ [a,b),\ 0 \leq a < b \leq 1, \text{ finite unions of semi-open intervals}\} \tag{A.1}$$

is easily seen to be an algebra, but it is not a σ-algebra because it does not contain all open intervals (a,b) -- which can be expressed as countable unions of semi-open intervals. However, the set

$$B = \{\Omega,\ \emptyset, \text{ all open, semi-open, closed intervals in } \Omega\} \tag{A.2}$$

is clearly a σ-algebra, and is, in fact, the σ-algebra generated by A.

Let R^n be n-dimensional Euclidean space, and denote by $\mathcal{R}^n$ the smallest σ-algebra of subsets of R^n containing all rectangles. $\mathcal{R}^n$ is called an n-dimensional <u>Borel</u> σ-algebra, and sets in $\mathcal{R}^n$ are called <u>Borel sets</u>.

A countably additive non-negative set function defined on an algebra and satisfying $\mu(\emptyset) = 0$ is called a <u>measure</u>. If $\mu(\Omega) = 1$, $\mu(\cdot)$ is called a <u>probability measure</u>. Countable additivity, sometimes called complete additivity, means

$$\mu\left(\bigcup^{\infty} A_n\right) = \sum^{\infty} \mu(A_n) \tag{A.3}$$

where $\{A_n\}$ is a collection of disjoint sets. Measures can be defined unambiguously on σ-algebras as well as algebras. Let $\mathfrak{F}$ be an algebra, with a measure μ defined on it. It can be shown[1,77,94] that there is a unique set function q, defined on $B(\mathfrak{F})$, such that

$q(\Lambda) = \mu(\Lambda)$, for $\Lambda \in \mathfrak{F}$

This important result allows us to calculate probabilities on σ-algebras, rather than on arbitrary sets in Ω. The closure property under countable Boolean operations renders σ-algebras much more tractable analytically than arbitrary ω sets.

Let B be a σ-algebra of subsets of Ω. The pair (Ω, B) is called a <u>measurable space</u>. If a specific measure q is specified, the triplet (Ω, B, q) is called a space with measure q. (Ω, B, P) is a <u>probability space</u> if P is a probability measure. In a probability space (Ω, B, P), we take Ω to be a sample space of points (simple events) ω, B the σ-algebra of measurable events in Ω, and $P(\Lambda)$ the probability of an event $\Lambda \in B$.

Sets of measure zero in Ω are called <u>null sets</u>, and needn't be events. (Ω, B, P) is said to be <u>complete</u>* if every null set is an event (necessarily, P(null)=0.) Completion is always possible; if a space (Ω, B, P) is not complete, we can extend P uniquely to $\bar{P}$, where $\bar{P}$ is defined on the σ-algebra $\bar{B}$ of all events, including null events, in Ω[1,77,94]. Our completed space is then (Ω, $\bar{B}$, $\bar{P}$). Since completion is always possible, we shall take all probability spaces to be complete in this thesis. We remark that the σ-algebra B of the example, (A.2), does not yield a complete space (Ω, B, P) if we logically take P to be the lengths of intervals, because B does not contain all null sets; i.e., single points in $[0, 1)$. To complete the space, we augment B by including degenerate intervals of length zero; a point x in $[0, 1)$ is then expressible as $[x, x]$. Sets in the augmented σ-algebra $\bar{B}$ are called Lebesgue-measurable sets of $[0, 1)$, and $\bar{P}$ defined on $\bar{B}$ is just Lebesgue measure.

Let (Ω_1, B_1) and (Ω_2, B_2) be two measurable spaces, and let $x\colon \Omega_1 \to \Omega_2$ be a function or map from Ω_1 to Ω_2. x is said to be a <u>measurable function</u> if for every $B \subset B_2$, the inverse image $x^{-1}(B) = \{\omega : x(\omega) \in B\}$ is in B_1.

---

*This notion of completeness has nothing in common with the completeness of metric spaces.

It is convenient to classify measurable functions according to the types of spaces on which they are defined. If x: $(R^n, \mathcal{R}^n) \to (R^m, \mathcal{R}^m)$, then x is called a <u>Borel function</u>. The class of Borel functions is the smallest class which contains all continuous functions and is closed under point-wise limits. If x is a measurable mapping from an arbitrary measurable space to the real line, x: $(\Omega, B) \to (R, \mathcal{R})$, then x is said to be <u>measurable with respect to B</u>, or <u>B-measurable</u>. In other words, x is B-measurable if for every real c, $\{\omega : x(\omega) \in (-\infty, c]\} \in B$. It is sufficient for c to be chosen from a set dense in the real line[1]. Note that if x is B-measurable, and $B_1 \supset B$, then x is obviously B_1-measurable, also.

Denote by B(x) the σ-algebra generated by sets of the form $\{\omega : x(\omega) \in (-\infty, c], c \text{ real}\}$, where x is the measurable map, x: $(\Omega, B) \to (R, \mathcal{R})$. B (x) is said to be the <u>σ-algebra generated by the variable x</u>. This is not a new definition; it is simply a shorthand for B(G), $G = \{\omega : x(\omega) \in (-\infty, c], c \text{ real}\}$. To extend this shorthand, let $\{x_t, t \in T\}$ be an indexed family of B-measurable functions x_t: $(\Omega, B) \to (R, \mathcal{R})$, $t \in T$. Then the σ-algebra generated by sets of the form $\{\omega : x_t(\omega) \in (-\infty, c], c \text{ real}, t \in T\}$ is denoted by $B(x_t, t \in T)$, and is called the <u>σ-algebra generated by $\{x_t, t \in T\}$</u>. $B(x_t, t \in T)$ is contained in B, and it is the smallest σ-algebra with respect to which the

x_t's are all measurable. To simplify writing, we abbreviate B_{x_t} for $B(x_\sigma, \sigma \in [t_o, t])$ in this thesis. Then, if $s < t$, $B_{x_s} \subset B_{x_t} \subset B$.

We have defined B-measurable functions as real-valued scalar variables. The extension to vector functions $x: (\Omega, B) \rightarrow (R^n, \mathcal{R}^n)$ is immediate. We simply replace the interval $(-\infty, c]$ with an arbitrary n-dimensional Borel set C. Then x is B-measurable if for every C, $\{\omega: x(\omega) \in C\} \in B$. With this generalization, the definition of complex-valued functions is obvious.

In this thesis, functions will be real-valued vector variables, in general. Deviations from this are noted where confusion would be likely.

2. Random Variable.

A real-valued random variable (r.v.) defined on the space (Ω, B, P) is a finite B-measurable ω function. Two r.v.'s are considered equal (a.s., or w.p.1) if they differ only on null sets; i.e., sets of P-measure zero. A complex r.v. is defined to be a pair of B-measurable real r.v.'s, and a random vector is defined to be a countable collection of real or complex r.v.'s.

3. Distribution Function.

Consider the set of r.v.'s $\{x_1, \ldots, x_n\}$ defined on the space (Ω, B, P). Since the r.v.'s are all B-measurable (by definition), the ω set $\Lambda_n = \{\omega: x_i(\omega) \leq X_i;$

$i = 1, \ldots, n\}$ is included in B, and $P(\Lambda_n)$ is called the (multidimensional) <u>cumulative distribution function</u> (c.d.f.) of the set $\{x_1, \ldots, x_n\}$. This function, denoted $P_{x_1, \ldots, x_n}(X_1, \ldots, X_n)$, is right continuous, monotone nondecreasing, and defines a Lebesgue-Stieltjes probability measure on R^n. The derivative

$$p_{x_1,\ldots,x_n}(X_1,\ldots,X_n) = \frac{\partial^n}{\partial X_1 \cdots \partial X_n} P_{x_1,\ldots,x_n}(X_1,\ldots,X_n)$$

when it exists, is called the <u>probability density function</u> (p.d.f.) of $\{x_1, \ldots, x_n\}$.

4. Random Process.

A <u>random process</u> (r.p.) is any family of r.v.'s $\{x_t, t \in T\}$ defined on (Ω, B, P). A function of $t \in T$ obtained by fixing ω and letting t vary is called a <u>sample path</u>, or <u>sample function</u>, or <u>realization</u>, of the process.

5. Stochastic Equivalence.

Two r.p.'s $\{x_t, t \in T\}$ and $\{y_t, t \in T\}$ defined on the same probability space (Ω, B, P) are said to be <u>stochastically equivalent</u> if for every $t \in T$, $P(x_t \neq y_t) = 0$. In symbols, $x_t = y_t$ w.p.1, $t \in T$. This definition is motivated by the fact that, in practice, experiment enables us to distinguish only between hypotheses involving finite-order statistics of r.p. If $\{x_t, t \in T\}$ and $\{y_t, t \in T\}$ are stochastically equivalent, their n^{th}-order statistics coincide for every finite n. Note, however, that processes which are stochastically equivalent do not necessarily have the same sample paths.

6. Expectation.

Let x be a r.v. defined on (Ω, B, P). If x is integrable; i.e., if

$$\int_{\Omega} |x| dP < \infty,$$

then $\int_{\Omega} x dP$ is called the expectation of x and is denoted E(x).

7. Conditional Expectation.

Let x be a r.v. defined on (Ω, B, P), and let B_1 be a σ-subalgebra of B ($B_1 \subset B$). Any ω function φ which is integrable and B_1-measurable, and satisfies the equation,

$$\int_{\Lambda} \varphi dP = \int_{\Lambda} x dP, \quad \text{all } \Lambda \in B_1 \tag{A.4}$$

is called the conditional expectation of x relative to B_1, and is denoted $E(x \mid B_1)$.

Since $E(x \mid B_1)$ is a finite B-measurable ω function, it is itself a random variable. If B_1 is indexed on some parameter set T; i.e., if $B_1 = (B_1)_t$, then $\{E[x \mid (B_1)_t], t \in T\}$ is a random process on (Ω, B, P). If B_1 is the σ-algebra generated by a family of r.v.'s $\{y_t, t \in T\}$, we sometimes write $E\{x \mid y_t, t \in T\}$ instead of $E(x \mid B_1)$. We use this notation primarily when we wish to indicate a particular realization of $E(x \mid B_1)$. When $B_1 = B(y_t, t \in T)$, $E(x \mid B_1)$ can always be expressed as a Borel function of the r.v.'s in the family $\{y_t, t \in T\}$ (See Doob[1], Supplement, Th. 1.5). Thus it is meaningful to interpret $E\{x|y_t, t \in T\}$ as the evaluation of the conditional expectation $E(x|B_1)$ for a particular realization of $\{y_t, t \in T\}$.

The form $E\{x|y_t, t \in T\}$ is often used indiscriminately in place of $E(x|B_1)$. In many applications this is a harmless artifice; however, it is meaningless if B_1 is an arbitrary σ-algebra, rather than one known to be generated by a particular family of r.v.'s. Furthermore, it is in any case incompatible with the definition (A.4).

The implicit definition (A.4) is valid for any σ-algebra B_1, whether or not a conditional probability distribution exists. For the purposes of this thesis, (A.4) is more convenient for manipulations than some of the explicit definitions which are stated in terms of conditional probabilities (See Doob[1], Ch. I). Moreover, all of the essential properties of conditional expectation can be obtained directly from (A.4)[1,77,94]. Some of these are discussed below.

Let (Ω, B, P) be a probability space, and let B_1 and B_2 be σ-algebras satisfying $B_1 \subset B_2 \subset B$. Then if x is an integrable random variable,

$$E\{E[x|B_2]|B_1\} = E\{x|B_1\} \text{ w.p.1} \qquad (A.5)$$

If z is a B_1-measurable r.v. such that xz is integrable, then

$$E\{zx|B_1\} = zE\{x \mid B_1\} \text{ w.p.1} \tag{A.6}$$

Let $B_o = \{\Omega, \emptyset\}$; then

$$E\{x|B_o\} = E(x) \quad \text{w.p.1.}; \tag{A.7}$$

thus,

$$E\{E[x|B_1]\} = E(x) \quad \text{w.p.1.} \tag{A.8}$$

Using (A.6) and (A.8), one can easily verify the following simple property, which motivates an important estimation result in this thesis:

$$E\{[x - E(x|B_1)]z\} = 0 \tag{A.9}$$

That is, if xz is integrable, the difference $x - E(x|B_1)$ is orthogonal to any B_1-measurable r.v. Thus, if x^* is a B_1-measurable r.v. such that $E\{|x^*z|\} < \infty$, a necessary condition for x^* to be a version of $E(x|B_1)$ is

$$E\{[x - x^*]z\} = 0, \tag{A.10}$$

for all B_1-measurable r.v.'s z. It is also sufficient, because it can be shown (see Neveu[94], p. 121) that (A.10) is equivalent to the definition of conditional expectation. We can show sufficiency directly as follows. We have

$$E\{(x - x^*)z\} = E\{E[(x - x^*)z|B_1]\}$$

$$= E\{[E(x|B_1) - x^*]z\} = 0, \tag{A.11}$$

for arbitrary B_1-measurable z -- in particular, for $z = E(x|B_1) - x^*$. Thus, $E\{[E(x|B_1) - x^*]^2\} = 0$, which implies that $x^* = E(x|B_1)$ w.p.1.

8. Measure and Integration.

In this section we give brief consideration to the time behavior of a process $\{x_t, t \in T\}$. We ignore questions of separability and differentiability; these topics, particularly the former, are quite technical, and not of primary concern in this thesis. The interested reader is referred to Doob[1], Gikhman and Skorokhod[3], Wong[77], or Neveu[94].

Let $\mathcal{L}$ be the σ-algebra generated by Lebesgue-measurable sets in T, where $T \subset R$. This is the minimal σ-algebra containing all intervals and completed with respect to measure assigning lengths to intervals. Assume that $\{x_t, t \in T\}$ is defined on the probability space (Ω, B, P). Then $\{x_t, t \in T\}$ is said to be a <u>measurable process</u> if $x_t(\omega)$ is a (t,ω) function measurable with respect to $\mathcal{L} \otimes B$, the product σ-algebra of $\mathcal{L}$ and B. That is, $\{x_t, t \in T\}$ is a measurable process if for every real c, $\{(t,\omega): x_t(\omega) \in (-\infty, c]\} \in \mathcal{L} \otimes B$. Denoting Lebesgue measure by μ, the process $\{x_t, t \in T\}$ is thus defined on the product space $(T \times \Omega, \mathcal{L} \otimes B, \mu \times P)$ as a function of two variables: $x: (T \times \Omega, \mathcal{L} \otimes B, \mu \times P) \to (R, \mathcal{R})$.

According to Fubini's theorem, the following formula is meaningful as long as x is integrable on the space $(T \times \Omega, \mathcal{L} \otimes B, \mu \times P)$:

$$E\{\int_T x_t \, dt\} = \int_T E\{x_t\} \, dt \qquad (A.12)$$

That is, if $\int_T E\{|x_t|\}\, dt < \infty$, then almost all sample functions of $\{x_t,\ t \in T\}$ are Lebesgue integrable on T, and $\int_T x_t\, dt$ defines a random variable (See Doob[1], Ch. II, or Neveu[94], pp. 76, 91). This result justifies all time-integrals of processes used in this thesis.

Time integrals of processes can be defined in a mean-square sense, also. Such integrals are not used here; we refer the interested reader to Jazwinski[86] for a detailed discussion of the m.s. calculus.

Stochastic integrals, which are different from integrals such as $\int_T x_t\, dt$, play a major role in this thesis, and are discussed in Ch. II and Appendix C.

9. Continuity.

Let $\{x_t,\ t \in T\}$ be a random process defined on (Ω, B, P) with values in R^n, and let $|\cdot|$ denote Euclidean distance in R^n. $\{x_t,\ t \in T\}$ is said to be continuous in probability, or stochastically continuous, at a point $s \in T$ if for every $\epsilon > 0$,

$$\lim_{t \to s} P\,(|x_t - x_s| > \epsilon\,) = 0 \tag{A.13}$$

It is said to be continuous in probability on T if (A.13) is satisfied at every point $s \in T$.

$\{x_t,\ t \in T\}$ is continuous w.p.1 at $s \in T$ if for almost all $\omega \in \Omega$, the set of sample paths which are discontinuous at s is a null set. If $\{x_t,\ t \in T\}$ is continuous w.p.1 at every point in T, it is said to be continuous w.p.1 on T.

$\{x_t, t \in T\}$ is said to have <u>a.s. continuous sample paths</u> if the set of its sample paths which are discontinuous on T is a null set.

10. Convergence.

Let $\{\xi_n\}$ be a sequence of random variables defined on (Ω, B, P) and let S denote the event that the sequence $\{\xi_n\}$ converges to a finite limit. If P(S) = 1; that is, if the sequence diverges on, at most, an ω set of P-measure zero, the sequence $\{\xi_n\}$ is said to converge <u>almost surely</u> (a.s.) or <u>with probability 1</u> (w.p.1).

If a random variable ξ exists such that $P\{|\xi_n - \xi| > \epsilon\} \to 0$ as $n \to \infty$ for any $\epsilon > 0$, we say that the sequence $\{\xi_n\}$ converges <u>in probability</u>; in symbols, $\xi = \text{p lim } \xi_n$, or $\xi_n \xrightarrow{p} \xi$. Note that convergence w.p.1 implies, but is not implied by, convergence in probability. However, it can be shown that a sequence of r.v.'s that converges in probability contains a subsequence that converges a.s.[3] A necessary and sufficient condition for convergence in probability which does not require knowledge of the limit of the sequence is the <u>Cauchy criterion</u>: $\{\xi_n\}$ converges in probability iff for any $\epsilon > 0$ and $\delta > 0$ there exist m and n such that $P\{|\xi_n - \xi_m| > \epsilon\} < \delta$.

If a random variable ξ exists such that $E\{|\xi_n - \xi|^\nu\} \to 0$ as $n \to \infty$, the sequence $\{\xi_n\}$ is said to converge <u>in ν^{th}-mean</u>. An important special case is $\nu = 2$, which we call <u>convergence in mean-square</u> (m.s.) or <u>in quadratic mean</u>

(q.m.); in symbols, ξ = l.i.m. ξ_n, or $\xi_n \xrightarrow{m.s.} \xi$, or $\xi_n \xrightarrow{q.m.} \xi$. It is obvious that a Cauchy criterion can also be stated for quadratic mean convergence.

Let $\{P_{\xi_n}\}$ be the sequence of distribution functions associated with the sequence of r.v.'s, $\{\xi_n\}$. If a random variable ξ with distribution P_ξ exists such that $P_{\xi_n} \to P_\xi$ at all points of continuity as $n \to \infty$, we say that the sequence $\{\xi_n\}$ converges <u>in distribution</u>, and we write $\xi_n \xrightarrow{d} \xi$.

<u>11. Martingale.</u>

Let $\{x_t, t \in T\}$ be a random process defined on (Ω, B, P), such that $E\{|x_t|\} < \infty$, $t \in T$, and let $\{\mathcal{F}_t, t \in T\}$ be an increasing family of σ-algebras such that x_t is $\mathcal{F}_t$-measurable for each $t \in T$. Then $\{x_t, \mathcal{F}_t, t \in T\}$ is called a <u>martingale</u> if, for $t > s$,

$$E\{x_t \mid \mathcal{F}_s\} = x_s \quad \text{w.p.1} \tag{2.7}$$

$\{x_t, \mathcal{F}_t, t \in T\}$ is said to be a <u>supermartingale</u> (<u>submartingale</u>) if the equality in (2.7) is replaced with $\leq$ ($\geq$).

Martingales possess remarkable properties, many of which are discussed in Doob[1], and some recent research papers[6-8]. One important result, which facilitates many proofs, is the <u>martingale convergence theorem</u> (Feller[36], vol. II, p. 236). Let $\{\xi_n\}$ be a martingale sequence defined on (Ω, B, P). If for all n, $E\{|\xi_n|\}$ is uniformly bounded, then there exists a r.v. ζ such that $\zeta_n \to \zeta$ w.p.1. This is particularly useful for proving the convergence of a sequence of conditional expectations of the form $E(x|B_n)$, where $B_n \subset B_{n+1}$, because $\{E(x|B_n), B_n, n = 1, 2, 3, \ldots\}$ is a martingale, as can easily be verified.

Appendix B -- Point Processes and Counting Processes

In Chapter II we introduced some definitions from the theory of point processes. It is the purpose of this appendix to summarize a few of the more useful properties of such processes, and to briefly touch upon recent work in the area which is related to this thesis.

1. Regular Point Processes.

Early in Ch. II we indicated that the class of regular point processes (RPP's) -- that is, point processes with incremental behavior described by (2.1)-(2.3) -- is a very large class, encompassing most point processes in common use. We found that counting processes can be conveniently manipulated, according to the rules of a particular stochastic calculus much like the Ito calculus associated with Wiener processes. In Ch. III it was shown that a wide class of MMSE estimation problems, for which the observations are RPP's, can be attacked systematically using a common set of tools.

Recently, Rubin[68] has studied the problems of statistical characterization and hypothesis testing for scalar RPP's. Of particular importance, he has shown that the class of RPP's is closed under "compounding." What Rubin calls a compound point process could be called a doubly-stochastic point process, in the terminology of this thesis. Such a process possesses an intensity $\{\lambda_t, t \in T\}$ which is causally dependent not only on $\{N_t, t \in T\}$, but on other processes, as well; i.e., λ_t is not B_{N_t}-measurable. Denote these additional processes by $\{S_t, t \in T\}$. Then, <u>for each realization</u> of $\{S_t, t \in T\}$, the compound or doubly-stochastic point process $\{N_t, t \in T\}$ is a RPP satisfying incremental relations of the form (2.1)-(2.3). This definition is an obvious generalization of the DSPP.

Rubin's result states that, <u>without conditioning</u> on a particular realization of $\{S_t, t \in T\}$, a compound point process is still a RPP, satisfying (2.1)-(2.3) with intensity $\lambda_t(B_{N_t}) = \hat{\lambda}_t = E\{\lambda_t | B_{N_t}\}$, the causal MMSE estimate of λ_t. In other words, if a point process $\{N_t, t \in T\}$ has an intensity $\{\lambda_t, t \in T\}$ such that λ_t is B_{N_t}-measurable, $t \in T$, then λ_t can be used directly in the incremental relations for a RPP, (2.1)-(2.3). If, on the other hand, λ_t is not B_{N_t}-measurable, then the causal MMSE estimate $\hat{\lambda}_t$, which <u>is</u> B_{N_t}-measurable, must be used in (2.1)-(2.3). In either case, the point process in question belongs to the class of regular point processes.

This is an important result, because it frees us from having to treat compound point processes as a special, distinct class.

Rubin has also shown that class of RPP's is closed under finite independent superposition; i.e., the sum of a finite number of independent RPP's is again an RPP.

RPP's admit a complete statistical characterization in terms of a joint p.d.f. of event times. Let $\{N_t, t \in T = [t_0, t_1]\}$ be (the counting process of) an RPP, with intensity function $\lambda_t(B_{N_t})$, as in (2.1)-(2.3). Then the joint p.d.f. of the r.v. N_t, and the event times $\{t_i\}$, can be written as

$$p(\{t_i\};\ N_t) = \exp\left[-\int_{t_0}^{t} \lambda_\sigma(B_{N_\sigma})d\sigma + \int_{t_0}^{t} \ln \lambda_\sigma(B_{N_\sigma})dN_\sigma\right] \qquad (t \in T) \tag{B.1}$$

This useful result is a generalization of the joint p.d.f. obtained in III.E. for DSPP's. As indicated in III.E., detection results follow immediately from the p.d.f., $p(\{t_i\};\ N_t)$. In particular, the log-likelihood ratio for distinguishing between two RPP's is

$$\ln \ell_t = -\int_{t_0}^{t} [\lambda_\sigma^{(2)}(B_{N_\sigma}) - \lambda_\sigma^{(1)}(B_{N_\sigma})]d\sigma + \int_{t_0}^{t} \ln \frac{\lambda_\sigma^{(2)}(B_{N_\sigma})}{\lambda_\sigma^{(1)}(B_{N_\sigma})} dN_\sigma \tag{B.2}$$

Rubin has obtained a number of other, more specialized results, which we shall not discuss here.

2. Poisson Processes.

The Poisson process (PP), defined in II. 1., is probably the simplest non-trivial example of a regular point process. We summarize here some of its properties, most of which are well known (See Parzen[5], for example).

We defined a Poisson process $\{N_t, t \in T = [t_0, t_1]\}$ to be an independent-increment counting process with Poisson-distributed random variables. These defining properties can be deduced if we instead define PP axiomatically. Let $\{N_t, t \in T\}$ be a counting process, and require that it satisfy the following axioms[5]:

1) $N_{t_0} = 0$ w.p.1;
2) $\{N_t, t \in T\}$ has independent increments;
3) For any $t > t_0$, $0 < Pr\{N_t > 0\} < 1$;
4) For any $t > t_0$ and any $\Delta t > 0$,

$$\lim_{\Delta t \to 0} \frac{Pr\{N_{t+\Delta t} - N_t \geq 2\}}{Pr\{N_{t+\Delta t} - N_t \geq 1\}} = 0;$$

It can be shown that, with these axioms, $\{N_t, t \in T\}$ is a PP.

The characteristic function $M_{N_t}(iv)$ of the random variable N_t is given by

$$M_{N_t}(iv) = E\{e^{iv N_t}\} = \exp\{m_t(e^{iv} - 1)\} \tag{B.3}$$

where

$$m_t = \int_{t_0}^{t} \lambda_\sigma \, d\sigma \tag{B.4}$$

is the mean of N_t, and λ_t is the intensity function. By differentiating (B.3) it is easily shown that m_t is also the variance of N_t.

We now state a useful expectation formula, which is evidently due to Bar-David[25].

Theorem B.1.

Let $\{N_t,\ t \in T\}$ be a scalar PP, with rate function λ_t and event times $\{t_i\}$, and let f(t) be an arbitrary integrable function defined on T. Then,

$$E\left\{\prod_{i=1}^{N_t} f(t_i)\right\} = \exp \int_{t_0}^{t} \lambda_\sigma [f(\sigma)-1]\, d\sigma \tag{B.5}$$

where the product is defined to be unity if $N_t = 0$.

Proof.

By using the stochastic calculus, we can prove this theorem very simply.

Using (2.21), we can write

$$\phi_t \triangleq \prod_{i=1}^{N_t} f(t_i) = \exp \int_{t_0}^{t} \ln f(\sigma)\, dN_\sigma$$

$$\triangleq e^{x_t}$$

In stochastic differential form,

$$dx_t = \ln f(t)\, dN_t$$

Now let us apply the Ito differential rule, Th. 2.1, to the function $\phi_t(x_t) = \exp x_t$. The first two terms in (2.31) are zero, so

$$d\phi_t = \phi_t [f(t) - 1]\, dN_t$$

Since $\{N_t, t \in T\}$ is an independent increment process, we can write

$$N_t = n_t + \int_{t_0}^{t} \lambda_\sigma d\sigma, \; t \in T,$$

where $\{n_t, B_{N_t}, t \in T\}$ is a martingale* (See Doob[1], Ch. VII). Thus,

$$d\phi_t = \phi_t[f(t) - 1]dn_t + \phi_t\lambda_t[f(t) - 1]dt;$$

or, in integral form,

$$\phi_t = \int_{t_0}^{t} \phi_\sigma[f(\sigma)-1]dn_\sigma + \int_{t_0}^{t} \phi_\sigma\lambda_\sigma[f(\sigma)-1]d\sigma$$

Using (2.25), we have

$$E\{\phi_t\} = \int_{t_0}^{t} E\{\phi_\sigma\}\lambda_\sigma[f(\sigma)-1]d\sigma$$

This is a non-random integral equation for $E\{\phi_t\}$; it is easily verified by substitution that the solution is

$$E\{\phi_t\} = \exp \int_{t_0}^{t} \lambda_\sigma[f(\sigma)-1]d\sigma$$

This proves the theorem. ∎

A simple example of a use for this theorem is the derivation of the characteristic function of a r.v. of a filtered Poisson process[5]. Let $y_t = \sum_{i=1}^{N_t} h(t_i)$, where $\{N_t, t \in T\}$ is a PP, and $h(t)$ is some integrable function, called the filter impulse response, defined on T. $\{y_t, t \in T\}$ is then a filtered PP. Using (B.5), we immediately have

$$E\{e^{iv y_t}\} = E\{\prod_{i=1}^{N_t} \exp iv h(t_i)$$

*This decomposition is a special case of Th. C.1, in Appendix C.

$$= \exp \int_{t_0}^{t} \lambda_\sigma [e^{i v h(\sigma)} - 1] d\sigma \tag{B.6}$$

3. Doubly-Stochastic Poisson Processes.

Since DSPP's play a central role in optical communications, we give them special attention, apart from RPP's.

The DSPP is a particular case of Rubin's[68] "compound" or doubly-stochastic point process. In the general case, the intensity λ_t depends both on the past of $\{N_t, t \in T\}$ and on the past of other processes $\{S_t, t \in T\}$, the latter possibly representing signals, channel disturbances, etc. In the case of a DSPP, the intensity λ_t depends only on the past of $\{S_t, t \in T\}$, and not on $\{N_t, t \in T\}$. Thus, for each realization of $\{S_t, t \in T\}$, the counting process is a Poisson process.

As was pointed out in Ch. II, DSPP's possess statistical properties which are quite different in general from those of a PP. For example, the counting statistics of a PP can always be expressed in the form (2.4); for a DSPP, however, the situation is much more complicated. The conditional counting distribution, given $\{\lambda_t, t \in T\}$, of a DSPP obeys (2.4); the unconditioned distribution is obtained formally by averaging (2.4) over all possible realizations of $\{\lambda_t, t \in T\}$. Carrying out such an average is in general a very difficult problem for which few answers exist. In III.E. a differential-difference equation for the counting distribution, in terms of a certain MMSE estimate, was obtained. Some explicit results have also been obtained by Karp and Clark[27] for the important case when λ_t is the squared magnitude of a complex Gaussian process. These results were found by studying the statistics of $m_t = \int_{t_o}^{t} \lambda_\sigma d\sigma$ for a fixed t. m_t and N_t are related statistically by the fact, obvious from (B.3), that the characteristic function $M_{N_t}(iv)$ of N_t is the moment generating function $M_{m_t}(s)$ of m_t evaluated at $s = e^{iv} - 1$.

The Gaussian process on which λ_t depends is expanded in a Karhunen-Loève series; by virtue of the independence of the coefficients and the orthogonality of the basis functions, m_t can then be expressed as an infinite sum of independent r.v.'s, each of which is the squared magnitude of a complex Gaussian r.v. This type of infinite sum has been studied extensively, and much is known about its properties. It is then a straightforward matter to deduce the relevant properties of N_t. The reader is referred to Ref. 27 for details.

Appendix C -- Stochastic Calculus

In this appendix we supplement some of the material on the stochastic calculus presented in Chapter II. Doob[1] and Skorokhod[2] are the primary references for stochastic integrals; the decomposition theorem for supermartingales is due to Meyer[7, 8].

1. Stochastic Integrals and Martingales.

In Chapter II we defined stochastic integrals with respect to martingale integrators (see (2.24)), and indicated some useful properties which followed from the definition. The right continuous increasing process $\{y_t, t \in T\}$ which satisfies (2.23) is a part of that definition, and must be known in order to make use of certain of the properties, e.g., (2.26). We now discuss the role of the process $\{y_t, t \in T\}$ relative to counting processes and stochastic integrals. We show that any counting process $\{N_t, t \in T\}$ can be decomposed into the sum of a martingale $\{\zeta_t, t \in T\}$ and a right continuous increasing process. The latter process, which can be expressed as a simple functional of the intensity of $\{N_t, t \in T\}$, is precisely the process $\{y_t, t \in T\}$ required for the definition of stochastic integrals relative to $\{\zeta_t, t \in T\}$. This result is summarized by the following theorem and corollary.

Theorem C.1.

Let $\{N_t,\ t \in T = [t_0,\ t_1]\}$ be the counting process of a scalar RPP, with intensity function $\lambda_t(B_{N_t})$ which is integrable; i.e., $E[\lambda_t] < \infty$, $t \in T$. Then $\{N_t,\ t \in T\}$ admits the unique decomposition

$$N_t = \zeta_t + \int_{t_0}^{t} \lambda_\sigma(B_{N_\sigma})\,d\sigma,\ t \in T, \tag{C.1}$$

where $\{\zeta_t,\ t \in T\}$ is an L_2 martingale which satisfies

$$E\{(\zeta_t - \zeta_s)^2 \,|\, B_{\zeta_s}\} = E\{\int_s^t \lambda_\sigma(B_{N_\sigma})\,d\sigma \,|\, B_{\zeta_s}\} \tag{C.2}$$

for $t > s$.

Corollary C.1.1.

Let $\{N_t,\ t \in T\}$ be the counting process of a doubly-stochastic point process; i.e., a point process possessing an integrable intensity $\{\lambda_t,\ t \in T\}$ which is causally dependent not only on $\{N_t,\ t \in T\}$ but on other processes, as well. Then $\{N_t,\ t \in T\}$ can be decomposed as

$$N_t = \nu_t + \int_{t_0}^{t} \hat{\lambda}_\sigma(B_{N_\sigma})\,d\sigma,\ t \in T, \tag{C.3}$$

where $\{\nu_t,\ t \in T\}$ is an L_2 martingale which satisfies

$$E\{(\nu_t - \nu_s)^2 \,|\, B_{\nu_s}\} = E\{\int_s^t \hat{\lambda}_\sigma(B_{N_\sigma})\,d\sigma \,|\, B_{\nu_s}\} \tag{C.4}$$

for $t > s$, and $\hat{\lambda}_\sigma(B_{N_\sigma}) = E[\lambda_\sigma \,|\, B_{N_\sigma}]$.

We remark that the corollary would appear to be more general than the theorem, although this is actually not the case, as Rubin[68] has shown recently that any doubly-stochastic point process is also a RPP (See Appendix B for further discussion).

A comparison of (C.2) with (2.23) reveals that the process $\{y_t,\ t \in T\}$ required in the definition of stochastic integral, (2.24), is given by

$$y_t = \int_{t_0}^{t} \lambda_\sigma (B_{N_\sigma}) d\sigma,\ t \in T \tag{C.5}$$

Proof.

This theorem is a direct consequence of a conjecture of Doob, first proved by Meyer[7,8], that a continuous-parameter supermartingale can be expressed as the difference of a martingale and a right-continuous, increasing process.

A process $\{x_t,\ t \in T\}$ is a supermartingale if $E\{|x_t|\} < \infty$ and $E\{x_t | B_{x_s},\ s < t\} \leq x_s$; s, $t \in T$. Define the negative counting process $\{\eta_t,\ t \in T\}$ by $\eta_t = -N_t$, $t \in T$; then $\{\eta_t,\ t \in T\}$ is clearly a supermartingale because, given the past B_{η_s}, the average value of a future random variable η_t cannot be greater than η_s. In symbols,

$$E\{\eta_t | B_{\eta_s},\ s < t\} \leq \eta_s$$

(One can refer to the incremental properties, (2.1)-(2.3), if this does not seem obvious enough.)

To apply Meyer's decomposition to $\{\eta_t,\ t \in T\}$, we must verify that his hypotheses are satisfied; i.e., that $\{\eta_t,\ t \in T\}$, belongs locally to his class (D). The negative of any right continuous increasing process belongs to (D) locally (p. 196)*. $\{N_t, t \in T\}$ can be defined to be right continuous, or modified on sets of probability zero to make it so (Neveu[94], p. 140); thus, $\{\eta_t, t \in T\}$ belongs locally to class (D).

Applying the decomposition, we have

$$\eta_t = \rho_t - y_t$$

where $\{\rho_t, t \in T\}$ is a martingale and $\{y_t, t \in T\}$ is a right continuous increasing process. We identify the process $\{y_t, t \in T\}$ as follows (p. 199). Define

$$y_t^h = \int_{t_0}^{t} \frac{\eta_\sigma - E\{\eta_{\sigma+h} \mid B_{\eta_\sigma}\}}{h} \, d\sigma , \quad h > 0 ;$$

then

$$y_t = \lim_{h \to 0} y_t^h \tag{C.6}$$

The process $\{y_t, t \in T\}$ so determined is the unique increasing process to be used in the decomposition of $\{\eta_t, t \in T\}$. Now, from (2.1),

$$E\{\eta_{\sigma+h} \mid B_{\eta_\sigma}\} = \eta_\sigma - \lambda_\sigma(B_{N_\sigma})\, h + o(h)$$

Thus,

$$y_t^h = \int_{t_0}^{t} \lambda_\sigma(B_{N_\sigma})\, d\sigma + \frac{o(h)}{h} ,$$

---

* Unless indicated otherwise, page references in this proof refer to Meyer[7].

and

$$y_t = \lim_{h \to 0} y_t^h = \int_{t_0}^{t} \lambda_\sigma(B_{N_\sigma})d\sigma$$

The decomposition $\eta_t = \rho_t - y_t$ implies $\rho_t = \eta_t + y_t$; i.e., $\eta_t = (\eta_t + y_t) - y_t$. The latter is equivalent to $N_t = (N_t - y_t) + y_t$, which is the desired decomposition for $\{N_t, t \in T\}$ with the martingale $\{\zeta_t, t \in T\}$ given by $\zeta_t = -\rho_t = N_t - y_t$, $t \in T$.

It is easy to verify that $\{\zeta_t, t \in T\}$ is an L_2 martingale. For $s < t$

$$\zeta_t = \zeta_s - \int_s^t \lambda_\sigma(B_{N_\sigma})d\sigma + N_t - N_s$$

Since $B_{\zeta_s} \subset B_{N_s}$

$$E\{\zeta_t | B_{\zeta_s}\} = E\{E[\zeta_t | B_{N_s}] | B_{\zeta_s}\}$$

Now,

$$E\{\zeta_t | B_{N_s}\} = \zeta_s - E\{\int_s^t \lambda_\sigma(B_{N_\sigma})d\sigma | B_{N_s}\}$$

$$+ E\{N_t - N_s | B_{N_s}\},$$

and it can be seen from the incremental properties (2.1)-(2.3) that

$$E\{N_t - N_s | B_{N_s}\} = E\{\int_s^t \lambda_\sigma(B_{N_\sigma})d\sigma | B_{N_s}\}$$

Therefore, $E\{\zeta_t | B_{N_s}\} = \zeta_s$, and $E\{\zeta_t | B_{\zeta_s}\} = \zeta_s$. This establishes that $\{\zeta_t, t \in T\}$ is a martingale.

To show that $\{\zeta_t, t \in T\}$ is L_2, and to verify (C.2), we use the Ito rule for counting processes to write, for $t > s$,

$$\zeta_t^2 - \zeta_s^2 = \int_s^t 2\zeta_\sigma \lambda_\sigma (B_{N_\sigma}) d\sigma + \int_s^t (2\zeta_\sigma + 1) dN_\sigma$$

$$= \int_s^t \lambda_\sigma (B_{N_\sigma}) d\sigma + \int_s^t (2\zeta_\sigma + 1) d\zeta_\sigma \qquad \text{(C.7)}$$

Since $\{\zeta_t, t \in T\}$ is a martingale,

$$E\{(\zeta_t - \zeta_s)^2 | B_{\zeta_s}\} = E\{\zeta_t^2 - \zeta_s^2 | B_{\zeta_s}\}$$

The last term on the r.h.s. of (C.7) is a stochastic integral of the type, (2.24); thus,

$$E\{(\zeta_t - \zeta_s)^2 | B_{\zeta_s}\} = E\{\int_s^t \lambda_\sigma (B_{N_\sigma}) d\sigma | B_{\zeta_s}\}$$

This proves (C.2). Now let $s = t_0$; the quantity $\zeta_{t_0} = 0$ a.s. because $N_{t_0} = 0$ a.s. Hence,

$$E\{\zeta_t^2\} = \int_{t_0}^t E\{\lambda_\sigma (B_{N_\sigma})\} d\sigma,$$

and this exists, by hypothesis. Therefore $\{\zeta_t, t \in T\}$ is square-integrable. This completes the proof of the theorem.

∎

The proof of the corollary is immediate in view of Rubin's result[68] that a doubly-stochastic ("compound") point process with intensity process $\{\lambda_t, t \in T\}$ is equivalent to an RPP with intensity function $\hat{\lambda}_t(B_{N_t})$ (See Appendix B). ∎

The extension of these results to vector counting processes is immediate. The Theorem is essentially unchanged, except that we must verify that each component of $\{\varsigma_t, t \in T\}$ is L_2. (C.2) applies to each component of $\{\varsigma_t, t \in T\}$, provided jumps in different components of $\{N_t, t \in T\}$ don't coincide. If this is the case, (C.2) can be written

$$E\{(\varsigma_t-\varsigma_s)(\varsigma_t-\varsigma_s)' \mid B_{\varsigma_s}\} = E\{\int_s^t diag[\lambda_\sigma(B_{N_\sigma})]d\sigma \mid B_{\varsigma_s}\}$$

$$= E\{\int_s^t \Lambda_\sigma(B_{N_\sigma})d\sigma \mid B_{\varsigma_s}\} \tag{C.8}$$

Thus we can define stochastic integrals on $\{\varsigma_t, t \in T\}$ on a component-by-component basis, and if

$$I_t = \int_{t_0}^t g_\sigma d\varsigma_\sigma \quad ; \quad J_s = \int_{t_0}^s f_\sigma d\varsigma_\sigma$$

where g_σ and f_σ are commensurate matrices and ς_σ a vector, then,

1) $E\{I_t\} = 0$ (C.9)

2) $E\{I_t J_s'\} = E\{\int_{t_0}^{t\wedge s} g_\sigma dy_\sigma f_\sigma'\}$

$$= E\{\int_{t_0}^{t\wedge s} g_\sigma \Lambda_\sigma f_\sigma' d\sigma\} \tag{C.10}$$

3) $E\{I_t' J_s\} = E\{\int_{t_0}^{t\wedge s} Tr[g_\sigma \Lambda_\sigma f_\sigma'] d\sigma\}$ (C.11)

4) $\{I_t, B_{\varsigma_t}, t \in T\}$ is a martingale

These are obvious generalizations of the properties of the scalar integral, (2.24).

2. Ito Rule for Wiener Processes.

In Chapter III the Ito differential rule for Wiener processes was required. For ease of rererence, we state that rule here. Proofs can be found in Duncan[20] or Ito[11,12]; the result can also be obtained with some effort directly from Kunita and Watanabe[6]. A discussion of the rule, with examples, is in Jazwinski[86]. Define the vector Ito process* $\{x_t, t \in T\}$ by the differential formula,

$$dx_t = f_t dt + g_t d w_t, \tag{C.12}$$

where $\{w_t, t \in T\}$ is a standardized vector Wiener process; i.e., a vector of independent Wiener processes each of which has variance parameter unity. Let ϕ_t be a scalar-valued function, continuously differentiable in t and having continuous second mixed partial derivatives with respect to the elements of x_t. Then $\{\phi_t, t \in T\}$ satisfies the Ito equation,

$$d\phi_t = \frac{\partial \phi_t}{\partial t} dt + (\nabla_{x_t} \phi_t)' dx_t + \frac{1}{2} \operatorname{Tr}[g_t' (\nabla_{x_t x_t} \phi_t) g_t] dt \tag{C.13}$$

where $\nabla_{x_t x_t} \phi_t$ denotes the matrix of second partials of ϕ_t. By substituting (C.12), this can also be written,

*See Ch. II for the definition of Ito process.

$$d\phi_t = \left[\frac{\partial \phi_t}{\partial t} + (\nabla_{x_t} \phi_t)' f_t + \frac{1}{2} \mathrm{Tr}\left[g_t' (\nabla_{x_t x_t} \phi_t) g_t\right]\right] dt + (\nabla_{x_t} \phi_t)' g_t \, dw_t \tag{C14}$$

Note that (C13) differs from the common differential of deterministic calculus by the term, $\frac{1}{2}\mathrm{Tr}[g_t'(\nabla_{x_t x_t}\phi_t)g_t]\,dt$ This term arises, loosely speaking, from the fact that $(dw_t)^2$, for a Wiener process, behaves as dt and is not negligible.

Appendix D -- Details of the Proof of Theorem 3.1

To facilitate reading, some technical details were omitted from the proof of Theorem 3.1, the representation theorem. In this appendix we prove two lemmas which are central to the proof of the theorem. The first lemma is a rather obvious but useful result relating the σ-algebras generated by certain processes. The second lemma is a projection theorem for conditional expectations.

It is often convenient to be able to replace a particular conditional expectation of a random variable with an equivalent expectation, conditioned on a different σ-algebra. Such a replacement is valid only if the conditioning σ-algebras are equivalent. The lemma which follows establishes the σ-equivalence of certain frequently-occurring processes.

Lemma D.1.

Let $\{y_t, t \in T\}$ be a jump process with finite jumps, defined on a probability space (Ω, B, P), and let $\{r_t, t \in T\}$ be a process which is sample-function integrable in time and B_{y_t}-measurable. Assume that fixed discontinuities of $\{r_t, t \in T\}$, if any, are finite. Define the process $\{x_t, t \in T\}$ by the formula,

$$x_t = y_t + \int_{t_0}^{t} r_\sigma \, d\sigma .$$

Then $B_{x_t} = B_{y_t}$, all $t \in T$; i.e., $\{x_t, t \in T\}$ and $\{y_t, t \in T\}$ "contain the same information."

<u>Proof.</u>

Since $y_t + \int_{t_0}^{t} r_\sigma d\sigma$ is a causal functional of $\{y_t, t \in T\}$, it is B_{y_t}-measurable for all $t \in T$; i.e., $B_{x_t} \subset B_{y_t}$, $t \in T$. It remains to show that $B_{y_t} \subset B_{x_t}$. Because the discontinuities of $\{y_t, t \in T\}$ are finite, the discontinuities of $\{r_t, t \in T\}$ are finite also (See Appendix A, Section 1, or Doob[1], Supplement). Thus the integral $\int_{t_0}^{t} r_\sigma d\sigma$ is continuous in t, and the discontinuities of $\{x_t, t \in T\}$ coincide in time and magnitude with the discontinuities of $\{y_t, t \in T\}$. Since almost all realizations of $\{y_t, t \in T\}$ are completely described by the sizes and locations of the jumps, it is clear that $B_{y_t} \subset B_{x_t}$; i.e., given $\{r_t, t \in T\}$, $\{x_t, t \in T\}$ and $\{y_t, t \in T\}$ are causally recoverable from each other. This proves the theorem. ∎

We remark that the simplicity of the proof derives completely from the jump nature of the process $\{y_t, t \in T\}$. If $\{y_t, t \in T\}$ has a continuously varying component, the proof is much more difficult; indeed, the only proofs presently known for this case place a uniform boundedness condition on $\{r_t, t \in T\}$. This is an exceptionally strong restriction; however, it has been conjectured that it is probably unnecessary[98].

Before stating Lemma D.2, we establish some definitions. We shall need the notion of a Hilbert space. Such a space is a complete, linear, inner product space. More formally, a Hilbert space* H has the following properties. Let x, y, and z belong to H and let a and b be numbers, real or complex. Then,

1) H is a linear space; i.e.,

   a) There is an addition operation (+):

      $x + y = y + x$

   b) There is a unique zero (0)

      $x + 0 = x$

   c) For every x there exists an element -x such that

      $x + (-x) = 0$

   d) $a(x + y) = ax + ay$

      $(a + b)x = ax + bx$

      $abx = a(bx)$

      $1x = x$

2) H is an inner product space; i.e., there exists an operation $(\cdot, \cdot)$, called the inner product, such that

   a) $(x, x) \geqslant 0$

   b) $(x, x) = 0 \Rightarrow x = 0$

   c) $(x, y) = \overline{(y, x)}$, (overbar denotes complex conjugate)

   d) $(ax + by, z) = a(x, z) + b(y, z)$

*Most texts on mathematical analysis contain tutorial discussions of Hilbert spaces; thus, we shall be brief.

e) $(x + y, \ x + y)^{\frac{1}{2}} \leq (x, x)^{\frac{1}{2}} + (y, y)^{\frac{1}{2}}$

3) H is complete in the norm

$\|x\| = (x, x)^{\frac{1}{2}}$.

The notion of "projection" is an important concept, closed related to Hilbert space. Let M be a Hilbert subspace of the Hilbert space H. If $u \in H$ and $v_p \in M$, then v_p is called the <u>projection</u> of u onto M if

$$\|u - v_p\| = \inf_{v \in M} \|u - v\|, \tag{D.1}$$

where $\|\cdot\|$ is the norm corresponding to the inner product in H. Given that v_p must belong to the subspace M, it is natural to interpret v_p as the best approximation (in the sense of the norm $\|\cdot\|$) to the element u, where $u \in H$. That is, v_p is the element of M which is "closest" to u.

It can be shown[97] that v_p is unique in the sense that the norm of the difference of any two projections is zero. It can also be shown that v_p is the unique projection of u onto M iff

$$((u - v_p), v) = 0, \text{ every } v \in M \tag{D.2}$$

For our own purposes, we define the following specific Hilbert space. We call this space the Hilbert space of nonlinear functionals of a process. Let $\{y_t, t \in T\}$ be a L_2 random process. Let $H_2\{y_t, t \in T\}$ be the Hilbert space consisting of all random variables which are either: (1) finite linear combinations of square-integrable Borel functions* of y_{t_0}, y_{t_1}, ..., y_{t_n}, $n = 1, 2, 3$ ...; or (2) limits of such combinations in the norm corresponding to the inner product $(u, v) = E\{u' \bar{v}\}$. That is, $H_2\{y_t, t \in T\}$ is the space of all nonlinear functionals of $\{y_t, t \in T\}$ which yield r.v.'s with finite second moments.

We have defined this space for the reason that, if z is a L_2 random variable, $\hat{z}_t = E\{z | B_{y_t}\}$ is the projection of z onto $H_2\{y_\sigma, \sigma \in [t_0, t]\}$. This interpretation of $\hat{z}_t$ is a direct consequence of (D.2) and the property of conditional expectation (see Appendix A, Section 7),

$$E\{[z - E(z|B_{y_t})] x_t\} = 0, \tag{D.3}$$

valid for every x_t which is B_{y_t}-measurable. Since the norm in $H_2\{y_\sigma, \sigma \in [t_0, t]\}$ corresponds to the inner product $(u, v) = E\{u' \bar{v}\}$, the projection interpretation of $\hat{z}_t$ leads us back to the established fact that $\hat{z}_t = E(z|B_{y_t})$ minimizes the mean-square error $E\{(z - \hat{z}_t)'(z - \hat{z}_t)\}$. Thus, finding $\hat{z}_t$ amounts to finding the projection of z onto $H_2\{y_\sigma, \sigma \in [t_0, t]\}$. This is a result well-known in probability theory (see Neveu[94], pp. 122-123).

It is clear from the definition (D.1) that it is sufficient for v to range over any dense subset $\tilde{M}$ of M. This is obviously true since, by the definition of density, any element of M is arbitrarily close (in the sense of the norm $\| \cdot \|$) to some element of $\tilde{M}$.

*See Appendix A, Section 1, for a definition.

Thus,

$$\|u - v_p\| = \inf_{v \in M} \|u - v\| = \inf_{v \in \tilde{M}} \|u - v\| \tag{D.4}$$

and v_p is the unique projection of u onto M iff

$$((u - v_p), v) = 0, \text{ every } v \in \tilde{M} \tag{D.5}$$

According to (D.5), if z and z_t^* are L_2 random variables, then z_t^* is a version of $\hat{z}_t = E(z|B_{y_t})$ iff z_t^* is B_{y_t}-measurable and satisfies

$$E\{(z - z_t^*)\, x_t\} = 0 \tag{D.6}$$

for every L_2 random variable x_t which belongs to a dense subset of $H_2\{y_\sigma, \sigma \in [t_0, t]\}$.

In the lemma which follows, we exhibit such a subset. Although the lemma does not represent a new result, it is apparently little-known in estimation theory and communications engineering. Frost[16] was evidently the first to use the lemma in a communications context; our proof is a simplification of his proof, the substance of which can be traced to Hida[99].

<u>Lemma D.2.</u>

Let z and z_t^* be L_2 r.v.'s, and let $\{y_t, t \in T = [t_0, t_1]\}$ be a L_2 vector random process. Then z_t^* is a version of $\hat{z}_t = E(z|B_{y_t})$ iff z_t^* is B_{y_t}-measurable and satisfies the relation,

$$E\{(z - z_t^*) \exp i \int_{t_0}^{t} W_\sigma' \, dy_\sigma\} = 0 \tag{D.7}$$

any non-random vector function W_t with components which are square-integrable (i.e., $\in L_2[T]$). The integral above is defined by the r.h.s. of (2.24).

This "projection theorem" is nothing more than a nonlinear orthogonality principle, much akin to the well-known linear orthogonality principle[4]. We remark that the set of exponential functionals indicated in (D.7) is not the only dense subset of $H_2\{y_\sigma, \sigma \in [t_0, t]\}$; however, it turns out to be a convenient one for calculations.

Proof.

Since $x_t = \exp i\int_{t_0}^{t} W_\sigma' dy_\sigma$ is a causal functional of $\{y_t, t \in T\}$, x_t is B_{y_t}-measurable. The necessity of (D.7) is an immediate consequence of this fact. To prove sufficiency we need only show that the set of functionals,

$$\tilde{H}_2 = \{\exp i\int_{t_0}^{t} W_\sigma' dy_\sigma \,;\; W_\sigma \in L_2[T]\}$$

is dense in $H_2\{y_\sigma, \sigma \in [t_0, t]\}$. Now, it can be shown (Bachman and Narici[97], p. 173) that $\tilde{H}_2$ is dense in $H_2\{y_\sigma, \sigma \in [t_0, t]\}$ iff the only element of $H_2\{y_\sigma, \sigma \in [t_0, t]\}$ which is orthogonal to every element of $\tilde{H}_2$ is the zero element of $H_2\{y_\sigma, \sigma \in [t_0, t]\}$.

Let $h \in H_2\{y_\sigma, \sigma \in [t_0, t]\}$. Write the integral in (D.7) as the limit of a sum, and appeal to the bounded convergence theorem[1,3] to exchange limit and expectation. It can then be seen that it suffices to show the following: if h satisfies

$$E\{h \exp i \sum_{j=0}^{n-1} W_j' \Delta y_j\} = 0, \qquad n = 1,2,3,\ldots, \qquad (D.8)$$

for any real vectors $\{W_j\}$ having square-summable components, then h is the zero element of $H_2\{y_\sigma, \sigma \in [t_0,t]\}$. Suppose (D.8) is true. Then by conditioning on the variables y_0, $\Delta y_1, \ldots, \Delta y_{n-1}$ and noting that $\exp i \sum_j W_j' \Delta y_j$ is a function of these variables, we have

$$E\{E[h | B(y_0, \Delta y_1, \ldots, \Delta y_{n-1})] \exp i \sum_{j=0}^{n-1} W_j' \Delta y_j\} = 0 \qquad (D.9)$$

The exponentials $\{\exp i \sum_{j=0}^{n-1} W_j' \Delta y_j ; n = 1, 2, \ldots\}$ span the space of L_2 functions of the variables $y_0, \Delta y_1, \ldots, \Delta y_{n-1}$ (Hida, p. 76); thus we have

$$E\{h | B(y_0, \Delta y_1, \ldots, \Delta y_{n-1})\} = 0 \quad \text{w.p.1} \qquad (D.10)$$

It is clear that we can choose successive refinements of the partition of $[t_0,t]$ to construct a sequence of σ-algebras which increases monotonically to B_{y_t}. The conditional expectation above is easily seen to be a L_2 martingale; the martingale convergence theorem (Appendix A, Section 11) then guarantees that

$$\lim_{n} E\{h \mid B(y_o, \Delta y_1, \ldots, \Delta y_{n-1})\} = E\{h \mid B_{y_t}\} \quad \text{w.p.1.} \qquad (D.11)$$

Thus $E\{h \mid B_{y_t}\} = 0$ w.p.1, and, as h is B_{y_t}-measurable, we have $h = 0$ w.p.1.

Thus the set $\tilde{H}_2$ is dense in $H_2\{y_\sigma, \sigma \in [t_o, t]\}$, and the lemma is proved. ∎

Appendix E -- Some Elementary Estimation Examples

At the beginning of Chapter III we briefly indicated some of the elementary approaches which could be used to solve certain simple estimation problems. As the approaches are generally well-known, it was felt to be inappropriate to elaborate on them in the body of the thesis. However, some of the results which can be obtained for DSPP's are both interesting in themselves and relevant to optical communications. In this appendix we work a few selected examples using elementary techniques. We consider three kinds of estimates: maximum a posteriori probability (MAP), maximum likelihood (ML), and minimum mean-square error (MMSE).

Let $\{N_t,\ t \in T\}$ be a DSPP, with intensity $\{\lambda_t,\ t \in T\}$. Assume that λ_t is a known function $\lambda_t(z)$ of a single random variable z. We wish to estimate z based upon the observations $\{N_\sigma,\ \sigma \in [t_0, t]\}$. To evaluate any of the estimates we are considering, MAP, ML, or MMSE, we need the statistics for $\{N_\sigma,\ \sigma \in [t_0,\ t]\}$ conditioned on knowledge of z. Eq. (2.5) meets this need by providing a conditional joint p.d.f. of the event times $\{t_i\}$ and the total number of events N_t:

$$p(\{t_i\}; N_t \mid z) = \exp\left[-\int_{t_0}^{t} \lambda_\sigma(z)\,d\sigma\right] \prod_{i=1}^{N_t} \lambda_{t_i}(z) \tag{E.1}$$

Let p(z) be the p.d.f. of z. Then necessary conditions for $\hat{z}_{ML}$ and $\hat{z}_{MAP}$ to be the ML and MAP estimates of z are[33]:

$$\frac{\partial}{\partial z} p(\{t_i\}; N_t | z)\Big|_{z=\hat{z}_{ML}} = 0 \tag{E.2}$$

$$\left[\frac{\partial}{\partial z} p(\{t_i\}; N_t | z) + \frac{\partial}{\partial z} p(z)\right]_{z=\hat{z}_{MAP}} = 0 \tag{E.3}$$

The MMSE estimate can be written in the form

$$\hat{z}_{MMSE} = \frac{\int z\, p(\{t_i\}; N_t | z)\, p(z)\, dz}{\int p(\{t_i\}; N_t | z)\, p(z)\, dz} \tag{E.4}$$

The ML equation, of course, applies only to estimating a non-random, but unknown parameter z; or a random parameter z which is uniformly distributed over a sufficiently large interval that $\frac{\partial}{\partial z} p(z) = 0$.

For our first example let $\lambda_t(z) = z\,\mu(t)$, where z is a non-negative r.v. and $\mu(t)$ is a given non-negative function. In optical communications this models a received signal which is known except for a constant factor, introduced perhaps by a slowly-varying random channel gain.

For convenience, let $z=x^2$, where x N(0, σ_x^2); that is, normal with mean zero and variance σ_x^2. Then z is a chi-square r.v. with one degree of freedom, and

$$\frac{\partial}{\partial z} \ln p(z) = -\frac{1}{2}\left(\frac{1}{z} + \frac{1}{\sigma_x^2}\right).$$

With these assumptions it is easily shown that

$$\hat{z}_{MAP} = \frac{N_t - \frac{1}{2}}{\int_{t_0}^{t} \mu(\sigma)d\sigma + \frac{1}{2\sigma_x^2}} \tag{E.5}$$

$$\hat{z}_{ML} = \frac{N_t}{\int_{t_0}^{t} \mu(\sigma)d\sigma} \tag{E.6}$$

$$\hat{z}_{MMSE} = \frac{N_t + \frac{1}{2}}{\int_{t_0}^{t} \mu(\sigma)d\sigma + \frac{1}{2\sigma_x^2}} \tag{E.7}$$

If $z^{(n)} = x_1^2 + x_2^2 + \ldots + x_n^2$, where the $\{x_i\}$ are independent and $N(0, \sigma_x^2)$, then $z^{(n)}$ is chi-square with n degrees of freedom, and the MAP and MMSE estimates of $z^{(n)}$ are

$$\hat{z}^{(n)}_{MAP} = \frac{N_t + \frac{n}{2} - 1}{\int_{t_0}^{t} \mu(\sigma)d\sigma + \frac{1}{2\sigma_x^2}} \tag{E.8}$$

$$\hat{z}^{(n)}_{MMSE} = \frac{N_t + \frac{n}{2}}{\int_{t_0}^{t} \mu(\sigma)d\sigma + \frac{1}{2\sigma_x^2}} \tag{E.9}$$

$\hat{z}_{ML}$ remains unchanged, of course. In terms of mean-square error, the (unbiased)* estimate $\hat{z}_{MMSE}$ is optimum. It has m.s. error,

*An estimate $\hat{z}$ of r.v. z is said to be <u>biased</u> if $E(\hat{z}) \neq E(z)$; if z is non-random, the condition is $E(\hat{z}) \neq z$.

$$E\left\{\left(z^{(n)} - \hat{z}^{(n)}_{MMSE}\right)^2\right\} = \frac{n\sigma_x^2}{\int_{t_0}^{t} \mu(\sigma)d\sigma + \frac{1}{2\sigma_x^2}} \tag{E.10}$$

On the other hand, the equally plausible estimate of (E.8) has m.s. error,

$$E\left\{\left(z^{(n)} - \hat{z}^{(n)}_{MAP}\right)^2\right\} = \frac{n\sigma_x^2}{\int_{t_0}^{t} \mu(\sigma)d\sigma + \frac{1}{2\sigma_x^2}} + \frac{1}{\left[\int_{t_0}^{t} \mu(\sigma)d\sigma + \frac{1}{2\sigma_x^2}\right]^2} \tag{E.11}$$

and is biased. The ML estimate is unbiased, and its variance is easily found to be

$$\mathrm{var}(\hat{z}_{ML}) = \frac{z}{\int_{t_0}^{t} \mu(\sigma)d\sigma} \tag{E.12}$$

For non-random parameters the Cramer-Rao bound (See III.G.1.) reduces to

$$\mathrm{var}(\hat{z}) \geq \left[\int_{t_0}^{t} \lambda_\sigma(z)\left[\frac{\partial}{\partial z}\ln\lambda_\sigma(z)\right]^2 d\sigma\right]^{-1} \tag{E.13}$$

Since this works out to be the same as (E.12) for $\lambda_\sigma(z) = z\mu(\sigma)$, we conclude that $\hat{z}_{ML}$, Eq. (E.6), is an efficient estimator for the <u>non-random</u> parameter z.

As a second example, let $\lambda_\sigma(z) = e^{2z}\mu(\sigma)$, with $z \sim N(\eta, \sigma_z^2)$. In optical communications, this represents a known signal suffering slow, lognormal fading. Using (E.3), we can show that $\hat{z}_{MAP}$ satisfies

$$2\sigma_z^2 e^{2\hat{z}_{MAP}} \int_{t_0}^{t} \mu(\sigma)d\sigma + \hat{z}_{MAP} = 2\sigma_z^2 N_t + \eta \qquad \text{(E.14)}$$

-- a transcendental equation that cannot be solved analytically. It can be shown[33] that ML (but not MAP) estimation commutes with non-linear operations; i.e., the ML estimate of a non-linear function of a r.v. is equal to the non-linear function of the ML estimate of the r.v. Thus we can obtain $\hat{z}_{ML}$ for this example directly from (E.6) as

$$\hat{z}_{ML} = \ln\left[\frac{N_t}{\int_{t_0}^{t}\mu(\sigma)d\sigma}\right]^{1/2} \qquad \text{(E.15)}$$

We ignore for the moment the MMSE estimate of z, as the integrals in (E.4) cannot be evaluated in closed form.

Using the Cramer-Rao bound of III.G.1, we can show that any estimate $\hat{z}$ of z, where $\lambda_\sigma(z) = e^{2z}\mu(\sigma)$ and $z \sim N(\eta, \sigma_z^2)$, is bounded in m.s. error by

$$E\{(z-\hat{z}_t)^2\} \geq \left[\frac{1}{\sigma_z^2} + 4e^{2(\eta+\sigma_z^2)}\int_{t_0}^{t}\mu(\sigma)d\sigma\right]^{-1} \qquad \text{(E.16)}$$

For both of these simple examples, it is clear that the estimates can be updated continuously in time, and that the m.s. error is non-increasing in time. We also note that the estimates depend only on the random variable N_t and not on the past history of the process $\{N_t, t \in T\}$. It can be seen from (E.2), (E.3), and (E.4) that this property holds true as long as $\lambda_t(z)$ can be expressed in the multiplicative form $f(z)\mu(t)$. If $\lambda_t(z)$ does not have this form, then the simple techniques outlined here seldom yield practical results; indeed, (E.2)-(E.4) are usually mathematically intractable.

Among the elementary techniques, the relative efficacy of the MMSE approach depends largely on the ease with which the integrals in (E.4) can be evaluated. We have mentioned that the integrals cannot be carried out explicitly for $\lambda_\sigma(z) = e^{2z}\mu(\sigma)$, for example. Except for the simplest intensities, this is generally the case. Though MAP and ML estimates are usually less than optimum in the sense of m.s. error, they are often tractable where MMSE estimates are not.

If we let $\lambda_t(z) = z\mu(t)$ with z any non-negative r.v., we can obtain a useful alternative expression for $\hat{z}_{MMSE}$. For this rate function (E.4) becomes

$$\hat{z}_{MMSE} = \frac{\int_0^\infty z^{N_t+1} \exp\left[-z\int_{t_0}^t \mu(\sigma)d\sigma\right] p(z)\,dz}{\int_0^\infty z^{N_t} \exp\left[-z\int_{t_0}^t \mu(\sigma)d\sigma\right] p(z)\,dz} \tag{E.17}$$

The moment generating function of z, if it exists, is given by

$$M_z(u) = \int_0^\infty e^{uz}\, p(z)\,dz \; ; \tag{E.18}$$

therefore

$$\hat{z}_{MMSE} = \left. \frac{\dfrac{\partial^{N_t+1}}{\partial u^{N_t+1}} M_z(u)}{\dfrac{\partial^{N_t}}{\partial u^{N_t}} M_z(u)} \right|_{u = -\int_{t_0}^{t} \mu(\sigma)d\sigma} \tag{E.19}$$

It is easily shown that this can be written as[27]

$$\hat{z}_{MMSE} = \frac{(N_t+1)\; p_{N_t}(N_t+1)}{p_{N_t}(N_t) \cdot \int_{t_0}^{t} \mu(\sigma)\,d\sigma} \tag{E.20}$$

where $p_{N_t}(k) = \Pr\{N_t = k\}$ for the doubly-stochastic Poisson process with intensity $z\mu(\sigma)$. For large N_t, $\hat{z}_{MMSE}$ tends to $N_t / \int_{t_0}^{t} \mu(\sigma)d\sigma$, the ML result of Eq. (E.6). We point out that (E.20) could have been derived in a more direct manner, as N_t is a sufficient statistic and the conditional event-time statistics are irrelevant. The conditional distribution of N_t, Eq. (2.4), would take the place of $p(\{t_i\}; N_t | z)$ in (E.4).

To illustrate the use of (E.20), let z be chi-square with two degrees of freedom. Then $p_{N_t}(k)$ is the geometric (Bose-Einstein) distribution[27],

$$p_{N_t}(k) = \frac{\left[2\sigma_x^2 \int_{t_0}^{t} \mu(\sigma)d\sigma\right]^k}{\left[1 + 2\sigma_x^2 \int_{t_0}^{t} \mu(\sigma)d\sigma\right]^{k+1}} \tag{E.21}$$

and $\hat{z}_{MMSE}$ is

$$\hat{z}_{MMSE} = \frac{N_t + 1}{\int_{t_0}^{t} \mu(\sigma)d\sigma + \frac{1}{2\sigma_x^2}} \tag{E.22}$$

This agrees with Eq. (E.9).

This alternative form, Eq. (E.20), valid for $\lambda_t(z) = z\mu(t)$, is sometimes convenient when the integrals in (E.4) are difficult to evaluate and $p_{N_t}(k)$ is known, either exactly or approximately. Such is the case when z is lognormal, $z = e^{2x}$, $x \sim N(\eta, \sigma_x^2)$; $p_{N_t}(k)$ can be evaluated exactly (though not necessarily in closed form) using the theory of transformations in a Hilbert space[31,32]. If this is inadequate, closed form approximations to $p_{N_t}(k)$ exist[34].

Consider now the more difficult problem of estimating, by elementary techniques, the process $\{z_t, t \in T\}$, where $\lambda_t = \lambda_t(z_t)$ is a memoryless function of z_t. To apply any of the estimator equations, (E.2)-(E.4), we must find a representation for $\{z_t, t \in T\}$ in terms of a countable set of r.v.'s. The Karhunen-Loéve series is such a representation. Let

$$z_t = \sum_i \zeta_i \phi_i(t), \quad t \in T, \tag{E.23}$$

where

$$z_i = \int_{t_0}^{t_1} \zeta_\sigma \phi_i(\sigma)\, d\sigma \tag{E.24}$$

and the $\{\phi_i\}$ are eigenfunctions of the integral equation,

$$\lambda_i \phi_i(u) = \int_{t_0}^{t_1} K_\zeta(u,\sigma)\phi_i(\sigma)\, d\sigma \tag{E.25}$$

$K_2(u, \sigma) = E\{[z_u - E(z_u)][z_\sigma - E(z_\sigma)]\}$ is the covariance function of $\{z_t, t \in T\}$. Now we can estimate $\{z_t, t \in T\}$ by estimating each of the $\{z_i\}$ and applying the prescription,*

$$\hat{z}_t = \sum_i \hat{\zeta}_i \phi_i(t), \quad t \in T. \tag{E.26}$$

An important consequence of this approach is that it always yields unrealizable (non-causal) interval estimates[33].

---

*More precisely, Eq. (E.23) is at first truncated after a finite number of terms, say N. Then, after calculating the set of N estimates, $\{\hat{z}_i\}_{i=1}^{N}$, and applying (E.26) as a finite sum, we take the limit (in the mean-square sense) as $N \to \infty$ to get $\{\hat{z}_t, t \in T\}$.

That is, $\hat{z}_t$ for any $t \in T$ depends on the values of the observation record over the entire interval T. Although this limits somewhat the practical usefulness of this type of estimator, the mathematical forms obtained sometimes suggest suboptimal estimators which are useful.

This technique is intractable for MMSE estimates; however, it is relatively easy to use in the MAP case, if $\{z_t, t \in T\}$ is assumed to be a zero-mean Gaussian process. We shall simply quote the results here; details and calculations can be found in Clark[26].

It can be shown that for $t \in T$, the MAP estimator $\hat{z}_t$ satisfies the equation,

$$\hat{z}_t = \int_{t_0}^{t_1} K_z(t,\sigma) \frac{\partial \lambda_\sigma(\hat{z}_\sigma)}{\partial \hat{z}_\sigma} \lambda_\sigma^{-1}(\hat{z}_\sigma)\left[dN_\sigma - \lambda_\sigma(\hat{z}_\sigma)d\sigma\right] \qquad \text{(E.27)}$$

This result bears a marked resemblance both to the well-known MAP equation for the Gaussian-signal-in-Gaussian-noise problem (Ref. 33, Ch. 5), and to our representation theorem of Chapter III. If λ_t depends not only on z_t but also on an "undesired" process, such as a random channel disturbance, a result similar to (E.27) can be derived[26]. When the undesired process enters λ_t as a constant multiplicative factor x,

$$\lambda_t(z_t, x) = x\,\mu_t(z_t) + \mu_0, \qquad \text{(E.28)}$$

with μ_0 a non-negative constant, the MAP equation contains terms which can be identified as MMSE estimates of x! These estimates appear as a result of averaging (E.27), the MAP equation for a given value of x, over all x.

The appearance of MMSE estimates in a MAP equation is somewhat less surprising in view of the p.d.f. obtained in III.E. for DSPP's:

$$p(\{t_i\};\ N_{t_1}) = \exp\left\{-\int_{t_0}^{t_1} \hat{\lambda}_{\sigma,MMSE}\, d\sigma + \int_{t_0}^{t_1} \ln \hat{\lambda}_{\sigma,MMSE}\, dN_\sigma\right\} \qquad \text{(E.29)}$$

Substituting (E.28) and holding $\{z_t,\ t \in T\}$ fixed, this gives

$$p(\{t_i\};\ N_t | z_t,\ t \in T) =$$

$$= \exp\left\{-\int_{t_0}^{t_1} [\hat{x}_{\sigma,MMSE}\, \mu_\sigma(z_\sigma) + \mu_0]\, d\sigma + \int_{t_0}^{t_1} \ln[\hat{x}_{\sigma,MMSE}\, \mu_\sigma(z_\sigma) + \mu_0]\, dN_\sigma\right. \qquad \text{(E.30)}$$

This equation can now be used instead of (E.1) as a starting point for the derivation of MAP estimates of $\{z_t,\ t \in T\}$.

It should be noted that the MMSE estimates appearing in (E.30) differ somewhat from those which arise by averaging (E.27) over x. It can be seen that the latter estimates depend on all the data $\{N_t,\ t \in T\}$, regardless of the value of the dummy integration variable σ, whereas those in (E.30) are "causal" in the sense that they depend on σ. Kailath[13] has pointed out that a similar discrepancy occurs in "Gauss-in-Gauss" problems, and does not represent an unresolvable inconsistency. The appearance of causal estimates in (E.30) is a direct consequence of the derivation of (E.29). On the other hand, one does not expect to get causal estimates of x by averaging (E.27) over x, because, in evaluating that average, the entire record $\{N_t,\ t \in T\}$ is held fixed.

REFERENCES

1. J. L. Doob, "Stochastic processes," New York: Wiley, 1953.

2. A. V. Skorokhod, "Studies in the theory of random processes," Reading, Mass: Addison-Wesley, 1965.

3. I. I. Gikhman and A. V. Skorokhod, "Introduction to the theory of random proce s," Philadelphia: W. B. Saunders, 1969.

4. A. Papoulis, "Probability, random variables, and stochastic processes," New York: McGraw-Hill, 1965.

5. E. Parzen, "Stochastic processes," San Francisco: Holden-Day, 1962.

6. H. Kunita and S. Watanabe, "On square integrable martingales," Nagoya Math. J., vol. 30, pp. 209-245, August 1967.

7. P. Meyer, "A decomposition theorem for supermartingales," Ill. J. of Math., vol. 6, pp. 193-205, 1962.

8. P. Meyer, "Decomposition of supermartingales: the uniqueness theorem," Ill. J. of Math., vol. 7, pp. 1-17, 1963.

9. S. N. Bernstein, "Principles de la theorie des equations differentielles stochastiques," Trudy Fiz.-Mat. Inst. Akad, Nauk., vol. 5, pp. 95-124, 1934.

10. I. I. Gikhman, "On the theory of differential equations of random processes," Ukr. Mat. Zh., vol. 2, pp. 317-339, 1951.

11. K. Ito, "Stochastic integral," Proc. Imp. Acad. Tokyo, vol. 20, pp. 519-524, 1944.

12. K. Ito, "On stochastic differential equations," Mem. Amer. Math. Soc., no. 4, pp. 1-51, 1951.

13. T. Kailath, "A general likelihood ratio formula for signals in Gaussian noise," IEEE Trans. on Info. Th., vol. IT-15,pp. 350-361, May 1969.

14. D. L. Snyder, "The state-variable approach to analog communication theory," Cambridge: MIT Press, 1968.

15. J. Evans, "Likelihood ratios and mutual information for nonhomogeneous Poisson processes having stochastic intensity functions," MIT-RLE, QPR No. 98, July 15, 1970.

16. P. A. Frost, "Nonlinear estimation in continuous time systems," Ph.D. Dissertation, Dept. of Elec. Engrg., Stanford Univ., Stanford, Calif., May 1968.

17. E. Wong and M. Zakai, "On the convergence of ordinary integrals to stochastic integrals," Ann. Math. Stat., vol. 36, pp. 1560-1564, 1965.

18. J. M. C. Clark, "The representation of nonlinear stochastic systems with applications to filtering," Ph.D. Dissertation, Elec. Engrg. Dept., Imperial College, London, April 1966.

19. R. L. Stratonovich, "Conditional Markov process theory," Prob. Appl., vol. 5, pp. 156-178, 1960.

20. T. E. Duncan, "Probability densities for diffusion processes with applications to nonlinear filtering theory and detection theory," Ph.D. Dissertation, Dept. of Elec. Engrg., Stanford Univ., Stanford, Calif., May 1967.

21. T. E. Duncan, "Evaluation of likelihood functions," Info. and Control, vol. 13. pp. 62-74, 1968.

22. T. E. Duncan, "Likelihood functions for stochastic signals in white noise," Info. and Control, vol. 16, pp. 303-310, 1970.

23. D. L. Snyder, "Filtering and detection for doubly-stochastic Poisson processes," to appear, IEEE Trans. on Info. Th., 1972.

24. D. L. Snyder, "Approximate filtering for a Poisson process with a stochastic intensity function," Washington Univ., Biomedical Computer Lab., Mono. No. 140, July 1970.

25. I. Bar-David, "Communication under the Poisson regime," IEEE Trans. on Info. Th., vol. IT-15, pp. 31-37, January 1969.

26. J. R. Clark, "Waveform estimators for the turbulent atmospheric optical channel," MIT-RLE, QPR No. 97, April 15, 1970.

27. S. Karp and J. R. Clark, "Photon counting: a problem in classical noise theory," IEEE Trans. on Info. Th., vol. IT-16, pp. 672-680, November 1970.

28. J. R. Clark and E. V. Hoversten, "Poisson process as a statistical model for photodetectors excited by Gaussian light," MIT-RLE, QPR No. 98, July 15, 1970.

29. I. S. Gradshteyn and I. M. Ryzhik, "Table of integrals series and products," New York: Academic Press, 1965.

30. M. G. Kendall and A. Stuart, "The advanced theory of statistics," vol. 1, London: Griffin, 1963.

31. R. S. Orr, "Photocount statistics: a representation theorem," MIT-RLE, QPR No. 100, January 15, 1971.

32. S. Solimeno, E. Corti, and B. Nicoletti, "Optical communication in turbulent media," JOSA, vol. 60, pp. 1245-1251, September 1970.

33. H. L. Van Trees, "Detection, estimation, and modulation theory," vol. I, New York: Wiley, 1968.

34. J. R. Clark and S. Karp, "Approximations for lognormally fading optical signals," Proc. IEEE, December 1970.

35. C. W. Helstrom, "Estimation of modulation frequency of a light beam," in "Optical Space Communication," R. S. Kennedy and S. Karp (Eds.), NASA SP-217, 1969.

36. W. Feller, "An introduction to probability theory and its applications," Vol. II, New York: Wiley, 1966.

37. B. Reiffen and H. Sherman, "An optimum demodulator for Poisson processes: photon source detectors," Proc. IEEE, vol. 51, pp. 1316-1320, October 1963.

38. R. G. Gallager, "Information theory and reliable communication," New York: Wiley, 1968.

39. N. Wiener, "The homogeneous chaos," Amer. J. Math., vol. 60, pp. 897-936, 1938.

40. K. Ito, "Multiple Wiener integral," J. Math. Society of Japan, vol. 3, pp. 157-169, 1951.

41. P. I. Kuznetsov, R. L. Stratonovich, and V. I. Tikhonov, "Nonlinear transformations of stochastic processes," New York: Pergamon Press, 1965.

42. J. R. Fisher, "Optical nonlinear filtering," Advances in Control Systems, vol. 5, pp. 198-301, 1967.

43. H. Cramer, "Mathematical methods of statistics," Princeton, N.J.: Princeton Univ. Press, 1961.

44. D. L. Snyder and I. B. Rhodes, "Bounds on the accuracy in causal filtering for nonlinear observations with some implications on asymptotic separation in stochastic control," to be presented at the 5th Congress of the International Federation on Automatic Control (IFAC), Paris, 1972.

45. Gaylord Moss, private communication.

46. E. V. Hoversten, R. O. Harger, and S. J. Halme, "Communication theory for the turbulent atmosphere," Proc. IEEE, vol. 58, pp. 1626-1650, October 1970.

47. L. Mandel and E. Wolf, "Letter to the Editor," J. Phys. A. (Proc. Phys. Soc.) ser. 2, vol. 1, pp. 625-627, 1968.

48. J. P. Gordon, "Comments on the quantal theory of communication," Appendix F, p. 125, in Reference 52.

49. R. S. Kennedy and E. V. Hoversten, "On the atmosphere as an optical communication channel," IEEE Trans. on Info. Th., vol. IT-14, pp. 716-725, September 1968.

50. A. Saleh, "An investigation of laser wave depolarization due to atmospheric transmission," IEEE J. Quantum Electronics, vol. QE-3, pp. 540-543, 1967.

51. S. J. Halme, "Efficient optical communication in a turbulent atmosphere," Ph.D. Thesis, Dept. Elec. Engrg., M.I.T., Cambridge, Mass., October 1969.

52. R. S. Kennedy and S. Karp (Eds.), "Optical space communication," NASA SP-217, 1969.

53. M. Ohta and T. Koizumi, "Intensity fluctuations of stationary random noise containing an arbitrary signal wave," _Proc. IEEE_, vol. 57, p. 1231, 1969.

54. R. J. Glauber, "Optical coherence and photon statistics," _Quantum optics and electronics_, C. de Witt, et al, Eds., New York: Gordon and Breach, 1965.

55. G. Bedard, "Photon counting statistics of Gaussian light," _Phys. Rev._, vol. 151, pp. 1038-1039, Nov. 25, 1966.

56. E. Jakeman and E. R. Pike, "The intensity fluctuation distribution of Gaussian light," _J. Phys. A._, (Proc. Phys. Soc.), ser. 2, vol. 1, pp. 128-138, 1968.

57. L. Mandel, "Flucuations of light beams," _Progress in Optics_, _II_, E. Wolf, Ed., New York: Wiley, 1963.

58. R. J. Glauber, "Photon counting and field correlations," _Physics of quantum electronics_, P. L. Kelley, et al, Eds., New York: McGraw-Hill, 1965.

59. C. W. Helstrom, "The distribution of photoelectric counts from partially polarized Gaussian light," _Proc. Phys. Soc._, vol. 83, pp. 777-782, 1964.

60. J. Perina, "Determination of the statistical properties of light from photoelectric measurements," _Czech. J. of Physics_, vol. B17, p. 1086, 1967.

61. J. R. Klauder and E. C. G. Sudarshan, "Fundamentals of quantum optics," New York: W. A. Benjamin, 1968.

62. P. L. Kelley and W. H. Kleiner, "Theory of electromagnetic field measurement and photoelectron counting," _Phys. Rev._, vol. 136, #2A, pp. 316-334, Oct. 19, 1964.

63. R. E. Bellman and K. L. Cooke, "Differential-difference equations," New York: Academic Press, 1963.

64. M. S. Bartlett, "Stochastic processes," Cambridge: Cambridge University Press, 1966.

65. D. R. Cox and P. A. W. Lewis, "Multivariate point processes," IBM Tech. Rpt. RC 3029, IBM Thomas J. Watson Res. Ctr., Yorktown Hgts., N.Y., Sept. 8, 1970.

66. D. R. Cox and P. A. W. Lewis, "The statistical analysis of a series of events," London: Methuen, 1966.

67. D. R. Cox, "Some statistical methods connected with series of events," J. Roy. Stat. Soc. B, vol. 17, pp. 129-164, 1955.

68. I. Rubin, "Regular point processes and their detection," submitted for publication to IEEE Trans. on Info. Th.

69. C. W. Helstrom, "Minimum mean-square error of estimates in quantum statistics," Phys. Let., vol. 25A, pp. 101-102, July 31, 1967.

70. C. W. Helstrom, "The minimum variance of estimates in quantum signal detection," IEEE Trans. on Info. Th., vol. IT-14, pp. 234-242, March 1968.

71. C. W. Helstrom, "Detection and quantum mechanics," Inform. Contr., vol. 10, pp. 254-291, March 1967.

72. C. W. Helstrom, "Detection and quantum mechanics (II)," Inform. Contr., vol. 13, p. 156, August 1968.

73. C. W. Helstrom, "The detection and resolution of optical signals," IEEE Trans. on Info. Th., vol. IT-10, pp. 275-287, October 1964.

74. J. W.-S. Liu, "Reliability of quantum mechanical communication systems," Tech. Rep. No. 468, Research Lab. of Electronics, M.I.T., Cambridge, Mass., Dec. 31, 1968.

75. S. D. Personick, "Efficient analog communication over quantum channels," Tech. Rpt. No. 477, Research Lab. of Electronics, M.I.T., Cambridge, Mass., May 15, 1970.

76. H. P. H. Yuen, "Communication theory of quantum systems," Ph.D. Thesis, Dept. of Elec. Engrg., M.I.T., Cambridge, Mass., June 1970.

77. E. Wong, "Stochastic processes in information and dynamical systems," New York: McGraw-Hill, 1971.

78. R. L. Stratonovich, "Topics in the theory of random noise," vol. I, New York: Gordon and Breach, 1963.

79. O. Macchi, "Stochastic point processes and multicoincidences," IEEE Trans. on Info. Th., vol. IT-17 pp. 2-7, January 1971.

80. J. R. Clark, "Estimation for Poisson processes," MIT-RLE, QPR No. 100, pp. 146-150, January 15, 1971.

81. I. Rubin, "Detection of point processes and applications to photon and radar detection," Tech. Rpt. No. 32, Information Sciences and Systems Lab., Dept. of Elect. Engrg., Princeton Univ., Princeton, N.J., September 1970.

82. F. J. Beutler and O. A. Z. Lenneman, "The theory of stationary point processes," Acta Math., vol. 116, pp. 159-197, 1966.

83. F. J. Beutler and O. A. Z. Lenneman, "Random sampling of random processes: stationary point processes," Inform. Contr., vol. 9, pp. 325-346, 1966.

84. F. J. Beutler and O. A. Z. Lenneman, "The spectral analysis of impulse processes," Inform. Contr., vol. 12, pp. 236-258, 1968.

85. F. J. Beutler and O. A. Z. Lenneman, "On the statistics of random pulse processes," Inform. Contr., vol. 18, pp. 326-341, 1971.

86. A. H. Jazwinski, "Stochastic processes and filtering theory," New York: Academic Press, 1970.

87. H. J. Kushner, "Dynamical equations for optimal nonlinear filtering," J. Diff. Eqs., vol. 3, pp. 171-190, 1967.

88. Yu. Sosulin and R. L. Stratonovich, "Optimum detection of a diffusion process in white noise," Radio Engrg. Electron. Phys., vol. 10, pp. 704-714, 1965.

89. R. L. Stratonovich, "Conditional Markov processes," Th. Prob. and Its Appl., vol. 5, pp. 156-178, 1960.

90. W. M. Wonham, "Some applications of stochastic differential equations to optimal nonlinear filtering," J. SIAM Contr., vol. 2, pp. 347-369, 1965.

91. D. C. Beste and C. L. Weber, "On the specification and performance of optimal non-anticipative analog receivers," presented at the 1970 International Symp. Info. Th., Noordwijk, The Netherlands, June 15-19, 1970.

92. J. N. Bucknam, "Direct detection of optical signals in the presence of atmospheric turbulence," S. M. Thesis, Dept. of Elect. Engrg., M.I.T., Cambridge, Mass., June 1969.

93. O. Macchi and B. Picinbono, "Estimation and detection of weak optical signals," submitted for publication to IEEE Trans. on Infor. Th.

94. J. Neveu, "Mathematical foundations of the calculus of probability," San Francisco: Holden-Day, 1965.

95. C. W. Helstrom, J. P. Gordon, and J. W.-S. Liu, "Quantum communication theory," Proc. IEEE, vol. 58, October 1970.

96. Special issue on optical communications, Proc. IEEE, October 1970.

97. G. Bachman and L. Narici, "Functional analysis," New York: Academic Press, 1966.

98. P. A. Frost and T. Kailath, "An innovations approach to least-squares estimation - part III: nonlinear estimation in white Gaussian noise," IEEE Trans. on Auto. Contr., June 1971.

99. T. Hida, "Stationary stochastic processes," Princeton, N.J.: Princeton University Press, 1970.

100. S. T. Hu, "Elements of real analysis," San Francisco: Holden-Day, 1967.

101. J. M. Wozencraft and I. M. Jacobs, "Principles of communication engineering," New York: Wiley, 1965.

BIOGRAPHICAL NOTE

John Robert Clark was born in Minneapolis, Minnesota, on April 7, 1944. He received the B.S. degree With Distinction from Stanford University, Stanford, California, in 1966, and the M.S. degree from Polytechnic Institute of Brooklyn, New York, in 1968, both in electrical engineering.

From 1966 to 1968 he was a Member of the Technical Staff of Bell Telephone Laboratories, Holmdel, New Jersey, where he participated in the Graduate Study Program, and was involved in the design of troposcatter communication systems. In 1969 he was a summer employee at NASA Electronics Research Center, Cambridge, Massachusetts, where he was engaged in the study of optical communication systems, and was the recipient of a NASA Superior Achievement Award.

His current interests are in the areas of applied probability theory, communication theory, and optical communications. He has published papers on photon counting statistics in IEEE Transactions on Information Theory, November 1970, and Proceedings of the IEEE, December 1970.

Mr. Clark is a member of Tau Beta Pi, Sigma Xi, and the American Association for the Advancement of Science.