Estimation Problems in Hybrid Systems

Recent developments in sensor and processor sophistication have created a need for effective estimation and control algorithms for hybrid, nonlinear systems. This book develops and illustrates a highly effective, computationally efficient, and flexible family of algorithms that can be used for the design of state estimators and feedback controllers for a variety of nonlinear plants. Several applications are studied, including tracking a maneuvering aircraft, automatic target recognition, and the decoding of signals transmitted across a wireless communications link.

The authors begin by setting out the necessary theoretical background, discussing infinite-dimensional algorithms and methods of nonlinear estimation. They develop a practical, finite-dimensional approximation to an optimal estimator and demonstrate its application to such problems as target tracking (including the jammed radar case), control of hybrid systems, warhead impact prediction, and an innovative signal demodulator. Throughout the book they illustrate theoretical results by simulation of real-world hybrid systems, drawn from a variety of engineering fields.

The book will be of great interest to graduate students and researchers in electrical and computer engineering. It will also be a useful reference for practicing engineers involved in the design of estimation, tracking, or wireless communications systems.

David Sworder received his Ph.D. from the University of California, Los Angeles and is Professor of Electrical and Computer Engineering at the University of California, San Diego. He has served as a consultant for many corporations, including Hughes Aircraft Company and Rockwell International. He is a Fellow of the IEEE.

John Boyd received his Ph.D. from the University of California, San Diego. He is a systems scientist at Cubic Defense Systems Inc., in San Diego, working on the design of novel communications and navigation systems.

Estimation Problems in Hybrid Systems

DAVID D. SWORDER
University of California, San Diego

JOHN E. BOYD
Cubic Defense Systems, Inc.

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 1999

Printed in the United States of America

Typeface Times Roman 11/14 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Sworder, David D.

Estimation problems in hybrid systems / David D. Sworder, John E. Boyd.
p. cm.

includes bibliographical references.
ISBN 0-521-62320-0
1. Feedback control systems. 2. Estimation theory. 3. Nonlinear theories. I. Boyd, John E., 1946– . II. Title.
TJ216. S96 1999
629.8′3 – dc21 99-12281
CIP

ISBN 0 521 62320 0 hardback

CONTENTS

List of Illustrations

Cambridge University Press
0521623200 - Estimation Problems in Hybrid Systems
David D. Sworder and John E. Boyd
Frontmatter
More information

Preface

Who Should Read This Book

This book is intended for engineers and designers who seek to develop effective estimation and control algorithms for nonlinear systems. The reader is assumed to have some background in random processes and estimation (see, for example, [Pap91]) along with familiarity with concepts of feedback control phrased within the context of linear state space models (see, for example, [DB95] or [Wol94]). This background should include knowledge of

- random variables and processes,
- probability density and distribution functions for random variables, both continuous and discrete,
- moments and cross moments, including correlation functions,
- second-order properties of stationary processes, including power spectral densities,
- conditional expectations with respect to an observation process,
- fundamental properties of feedback systems, including stability and controllability of a system model.

Both time-continuous and time-discrete processes will be encountered. Some familiarity with mean-square estimation is useful. For example, in the development of the Kalman filter [BW92, Chapter 7], linear state space models are integrated with Gaussian white noise. The Kalman filter will form a basis of comparison for many of the estimators that follow.

Our treatment is applications oriented, but the reader will find that nonlinear systems require more detailed analysis than is necessary in the study of linear systems. For example, a common approach to linear estimation uses a system model phrased as a set of ordinary differential equations with continuous white noise excitation [May79, Chapter 4]:

$$\dot{x}_t = Ax_t + Bu_t + Cw_t, \tag{0.1}$$

where x_t is the state, u_t is the actuating signal, and w_t is an exogenous, Gaussian white noise excitation. This model, or its time-discrete analogue, suffices to represent the dynamics of system evolution. The Kalman filter is based, in part, upon such a model.

Equation (0.1) is an adequate expression of system dynamics in many situations. But because of the pathological properties of white noise, it is sometimes preferable to replace the differential equation (0.1) with an integral equation:

$$x_t = x_0 + \int_0^t (Ax_s + Bu_s)\,ds + \int_0^t C\,dw_s. \tag{0.2}$$

It is easier to give consistent meaning to the integrals in Equation (0.2) than it is interpret the path properties of white noise [WH85, Chapter 3, Section 8]. Stochastic integral equations such as (0.2) are often written in a differential form using increments [Ell82]; for example, (0.2) would be written

$$dx_t = (Ax_t + Bu_t)\,dt + C\,dw_t, \tag{0.3}$$

where (0.3) is taken to be shorthand for (0.2). System analysis can then be carried out in terms of the increments, retaining only terms of order one or less in dt. This formal calculus of increments permits a coordinated treatment of processes, both continuous and piecewise continuous.

When the system is nonlinear, additional difficulties arise because nonlinear operations on white noise paths are difficult to interpret, whether in the form (0.1) or (0.2). Suppose the dynamic evolution of the system is represented by a vector stochastic differential equation:

$$dx_t = \mathbf{f}(x_t, u_t)\,dt + \mathbf{g}(x_t, u_t)\,d\eta_t, \tag{0.4}$$

where $\{\eta_t\}$ is a random process and represents the unpredictable disturbances that influence state evolution. Equation (0.4) is interpreted to say that, from state x_t at time t, the plant has a deterministic drift in the direction $\mathbf{f}(x_t, u_t)$. About this extrapolation, there is a random perturbation $(d\eta_t)$ with multiplier $\mathbf{g}(x_t, u_t)$.

In an introductory analysis, we might divide both sides of (0.4) by dt to arrive at a model that has a more conventional appearance:

$$\dot{x}_t = \mathbf{f}(x_t, u_t) + \mathbf{g}(x_t, u_t)\dot{\eta}_t.$$

If $\{\eta_t\}$ were Brownian motion, $\{\dot{\eta}_t\}$ would be *Gaussian white noise*. However, $\{\eta_t\}$ is not necessarily Brownian and may indeed have discontinuous sample paths. Equation (0.4) is better written

$$x_t = x_0 + \int_0^t \mathbf{f}(x_s, u_s)\,ds + \int_0^t \mathbf{g}(x_s, u_s)\,d\eta_s. \tag{0.5}$$

Because the integrands in (0.5) are random, care must be exercised in their explication.

Preliminaries

First, we present some basic notational conventions. An integer index set $\{1, \ldots, S\}$ will be designated $\mathbf{S}$. Boldface vectors labeled $\mathbf{e}$ are canonical unit vectors whose dimension is always clear from the context: $\mathbf{e}_i$ is the ith canonical unit vector in $\mathbb{R}^k$. We shall encounter $\mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}'_j$ and $\mathbf{E}_i = \mathbf{E}_{ii}$; $\mathbf{1}$, a vector of "ones;" and $\mathbf{I}$, the identity matrix. The dimension of these matrices will be determined by context. The statement "x is $\mathbf{N}(m, P)$" (or "$x \in \mathbf{N}(m, P)$") means that x is a Gaussian random variable with (mean, covariance) equal to (m, P), though sometimes $\mathbf{N}(m, P)$ will represent the probability density itself. The Hadamard product "$*$" is defined by $(x * y)_i = x_i y_i$. If λ is a vector, none of whose components is zero, we shall refer to the vector of inverses as λ^{-1}: $\lambda * \lambda^{-1} = \mathbf{1}$. An integral over the whole space is written $\int_\Omega$; for example, $\int_\Omega f(u)\, du$ indicates an integration over the full range of the variable labeled u.

In what follows, subscripts are used in a variety of ways. We wish to avoid iterated subscripts because such forms make the equations harder to read. Suppose we are dealing with the time interval $[0, \mathsf{T}]$, and at time t, ν_t is the value of the vector process $\{\nu_t; t \in [0, \mathsf{T}]\}$ (written $\{\nu_t\}$). Where no confusion will arise, a subscript may identify time, the component of the vector, or a particular set of components of the vector. For the process $\{\nu_t\}$, $\{\nu_1\}$ denotes the scalar process that is the first component of $\{\nu_t\}$, while $\{\nu_x\}$ is a subvector of processes in $\{\nu_t\}$ that is associated in some way with another process $\{x_t\}$.

This notation becomes ambiguous when the process is time discrete: $\{\nu_t; t = kT, k \in \mathbf{N}\}$. The sequence $\{\nu_{kT}; k \in \mathbf{N}\}$ will be written $\{\nu[k]\}$. A component sequence from $\{\nu[k]\}$ would be $\{\nu_i[k]\}$. This notational convention becomes complicated when $\{\nu[k]\}$ is a sequence of functions of some spatial variable "z." When the spatial variable needs to be made explicit, the sequence is written $\{\nu[k](z)\}$ with components $\{\nu_i[k](z)\}$. If there is concern that the meaning of the subscript is hard to determine from context, the more explicit notation will be used (e.g., $(\nu_i[k] = \mathbf{e}'_i \nu[k])$.

Subscripts also appear as identifiers. To make explicit the variables involved in correlation and covariance matrices, they are sometimes identified with subscripts (e.g., $E[xy'] = R_{xy}$). The time dependence of the moment may be written as a direct argument (e.g., $E[x_t y'_t] = R_{xy}(t)$ or, alternatively, $E[x[k]y[k]'] = R_{xy}[k]$). Subvectors and submatrices will be denoted in different ways depending on context: If P_{xx} is a matrix related to a vector $\boldsymbol{x}$, $P_{xx}(r : s, t : v)$ is the submatrix formed from rows r through s of columns t through v; $P_{x_i x}$ is the ith row of the matrix (and is usually associated with the ith component of the vector x) and P_{xx_i} is the ith column. A matrix $S_{yy}(t)$ is the square root of the positive symmetric matrix $P_{yy}(t)$ if $S_{yy}(t)'S_{yy}(t) = P_{yy}(t)$. There are many square roots of a positive matrix and their differences are important in computation. We are concerned only with representation of matrices and thus any of the square roots will do for our purposes. Given the multiplicity of uses, when confusion regarding the interpretation of a subscript may exist, multiple subscripts will perforce be used.

Random Processes

In this book we will look at the properties of random processes defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ (see, [Ell82]) on a time interval [0, T] (alternatively $t = kT; k \in \mathbf{N}$). The set of events $\mathcal{F}$ is a σ-field. A *random variable* is a (real-valued) $\mathcal{F}$-measurable function. A random vector has components that are random variables, and a random process is a time-indexed set of random vectors [Pap91, Chapter 10].

In estimation and control the notion of conditional expectation is important. Our definition of conditional expectation differs from that found in introductory engineering texts. Suppose x and y are random variables (understood to be on $(\Omega, \mathcal{F}, \mathcal{P})$). Another σ-field, $\mathcal{Y}$, on Ω is said to be coarser than $\mathcal{F}$ if every element of $\mathcal{Y}$ is necessarily an element of $\mathcal{F}$: $\mathcal{Y}$ is *coarser* than $\mathcal{F}$ (i.e., $\mathcal{Y} \subset \mathcal{F}$); $\mathcal{F}$ is *finer* than $\mathcal{Y}$. The coarsest σ-field (necessarily within $\mathcal{F}$) with respect to which y is a random variable is said to be the σ-field generated by y and is possibly labeled descriptively (e.g., $\mathcal{Y}$). Clearly y is a $\mathcal{Y}$-random variable (a random variable on $(\Omega, \mathcal{Y}, \mathcal{P})$). The expectation of x given y (denoted $E[x \mid \mathcal{Y}]$) is the $\mathcal{Y}$-random variable with all of the orthodox properties of conditional expectation given in [Pap91, Chapter 7]. Idiomatically, we would say that $E[x \mid \mathcal{Y}]$ is a random variable expressible as a function of y.

We will deal with conditioning, not just on random variables, but on the sample paths of random processes. On $(\Omega, \mathcal{F}, \mathcal{P})$, there are several elemental random processes. All of them are piecewise *right continuous* (or continuous); that is, if $\{y_t\}$ is a random process, $y_t = y_{t+}$. Let $\{x_t\}$ and $\{y_t\}$ be random processes and consider $\{y_t\}$ on the interval $[0, s]$. There is a coarsest σ-field within $\mathcal{F}$ with respect to which events determined by $\{y_u; u \in [0, s]\}$ (the past and present of $\{y_t\}$) are measurable. This σ-field will be called $\mathcal{Y}_s$. The indexed family of σ-fields, $\{\mathcal{Y}_t\}$, is the *filtration* generated by $\{y_t\}$ (see, [Ell82, p. 332]). This filtration is right continuous because $\{y_t\}$ is right continuous and, moreover, is such that if $s \le t$ then $\mathcal{Y}_s \subset \mathcal{Y}_t$. There is also a left continuous filtration generated by $\{y_u; u \in [0, s)\}$ and labeled $\{\mathcal{Y}_{t-}\}$. A process $\{x_t\}$ is *$\mathcal{Y}_t$-adapted* if x_t is a $\mathcal{Y}_t$-random variable for every t. A process $\{x_t\}$ is *$\mathcal{Y}_t$-predictable* if x_t is $\mathcal{Y}_{t-}$-measurable for every t. Indeed, the predictable version[†] of a right continuous process is given by its left continuous modification. So, if $\{x_t\}$ is a right continuous random process and generates $\{\mathcal{X}_t\}$, then $\{x_{t-}\}$, the left continuous modification of $\{x_t\}$, generates $\{\mathcal{X}_{t-}\}$, the filtration of "past events."

There may be different filtrations on $(\Omega, \mathcal{F}, \mathcal{P})$ relevant to the application, and if we wish to distinguish the filtration of interest we will write the probability space and filtration as $(\Omega, \mathcal{F}, \mathcal{P}; \mathcal{F}_t)$. All of the elemental processes in this book are $\mathcal{F}_t$-adapted.

If $\{x_t\}$ is a state process and $\{y_t\}$ is an observation process, an engineer may seek the expectation of x_t conditioned on the past of $\{y_t\}$: Idiomatically this estimate is said to be a function of the observations up to time t or to be *causal*. Write this conditional expectation $E[x_t \mid \mathcal{Y}_t]$. The conditional mean $E[x_t \mid \mathcal{Y}_t]$ is a $\mathcal{Y}_t$-random variable, which we will write as $\hat{x}_t$ if the conditioning filtration is apparent from context. The random process $\{\hat{x}_t\}$ is a $\mathcal{Y}_t$-adapted random process (also called a $\mathcal{Y}_t$-random process). If confusion might arise as to the conditioning filtration, the conditional mean process would be written $\{\hat{x}_t; \mathcal{Y}_t\}$. Since $\mathcal{Y}_t$ is coarser than $\mathcal{F}_t$ the estimation error is an $\mathcal{F}_t$-random variable. For example, x_t is $\mathcal{F}_t$-adapted and $\hat{x}_t$ is $\mathcal{Y}_t$-adapted. Hence, the error $\tilde{x}_t = x_t - \tilde{x}_t$ must be an $\mathcal{F}_t$-random variable but likely not a $\mathcal{Y}_t$-random variable.

The structure of random processes on $(\Omega, \mathcal{F}, \mathcal{P}; \mathcal{F}_t)$ can be quite complex. Fortunately we will not have to face processes of a general type. Instead, only two circumscribed classes of elemental $\mathcal{F}_t$-random processes will appear in the applications that follow. The first is composed of $\mathcal{F}_t$-Brownian motions: A (vector) random process $\{w_t\}$ is a *Brownian motion* if $w_0 = 0$, and when $s \le t$, $w_t - w_s$ is $\mathbf{N}(0, W(t-s))$ and independent of $\mathcal{F}_s$ (see, [Ell82, Definition 12.27]). We will refer to W as the *intensity* of the Brownian motion. It is easily seen that $E[w_t w_t'] = Wt$. Brownian motion is an $\mathcal{F}_t$-martingale process: If $s \le t$, $E[w_t \mid \mathcal{F}_s] = w_s$.

It is useful to develop a formal calculus of increments. Increments are defined in the forward direction (e.g., $dw_t = w_{t+dt} - w_t$ with $dt > 0$). Associated with $\{w_t\}$ is the $\mathcal{F}_t$-predictable quadratic variation process, $\langle w, w; \mathcal{F}_t\rangle_t$. The *$\mathcal{F}_t$-predictable quadratic variation process* is the integral of its increments, where $d\langle w, w; \mathcal{F}_t\rangle_t = E[dw_t dw_t' \mid \mathcal{F}_t]$. Since $dw_t dw_t' = W\,dt$ [WH85, Proposition 3.4], it follows that $\langle w, w; \mathcal{G}_t\rangle_t = Wt$ for any filtration $\{\mathcal{G}_t\}$ $(d\langle w, w; \mathcal{G}_t\rangle_t = W dt)$.

The second class of elemental processes contains $\mathcal{F}_t$-Markov processes on the canonical unit vectors in $\mathbb{R}^S$. Such a process, $\{\phi_t\}$, is characterized by its initial probability distribution, $(\hat{\phi}_0)$, and its transition rates. Let the $S \times S$-matrix Q have as its elements $Q_{ij} = \mathcal{P}(\phi_{t+dt} = \mathbf{e}_j \mid \phi_t = \mathbf{e}_i)/dt$ if $i \ne j$, with $Q_{ii} = -\sum_{j \ne i} Q_{ij}$. The *generator* of the Markov process $\{\phi_t\}$ is Q', a matrix with nonnegative elements

† If x_t and y_t are random variables on the same probability space and $\mathcal{P}(x_t = y_t) = 1$ for all t, then the variables are said to be versions or modifications of each other.

off the diagonal and column sums equal to zero. Because the state space of $\{\phi_t\}$ is the canonical unit vectors, $\mathcal{P}(\phi_t = \mathbf{e}_i) = E[\mathbf{e}_i{}'\phi_t]$. Let us add the discontinuities to the forward increment. Define $d\phi_t$ as $\phi_{t+dt} - \phi_{t-}$. If $\{\phi_t\}$ has a discontinuity at time t, this will be denoted $\Delta\phi_t = \phi_t - \phi_{t-}$. It is easily shown that $E[d\phi_t \,|\, \mathcal{F}_t] = Q'\phi_t dt$.

Define $dm_t = d\phi_t - Q'\phi_t\, dt$. Then $\{m_t\}$ has discontinuities where $\{\phi_t\}$ does:

$$\Delta m_t = m_t - m_{t-} = \phi_t - \phi_{t-}.$$

Clearly, $E[dm_t \,|\, \mathcal{F}_t] = 0$: $\{m_t\}$ is an $\mathcal{F}_t$-martingale (see, [EAM95, Section 7.2]). The $\mathcal{F}_t$-predictable quadratic variation of $\{m_t\}$ is defined as with $\{w_t\}$, but $\{m_t\}$ has a fundamentally different character. First, if $\{\phi_t\}$ makes no transitions in the interval $[t, t + dt]$, then $dm_t dm_t' \approx 0$: $\{m_t\}$ is called a purely discontinuous $\mathcal{F}_t$-martingale because, excluding jumps, its quadratic variation is zero. Alternatively, if $\{\phi_t\}$ makes the transition $\mathbf{e}_i \mapsto \mathbf{e}_j$ in $[t, t + dt]$, then $\Delta m_t \Delta m_t' = (\mathbf{e}_j - \mathbf{e}_i)(\mathbf{e}_j - \mathbf{e}_i)' = \mathbf{E}_i + \mathbf{E}_j - \mathbf{E}_{i,j} - \mathbf{E}_{j,i}$. If $\phi_t = \mathbf{e}_i$,

$$d\langle m, m; \mathcal{F}_t\rangle_t = \sum_j (\mathbf{E}_i + \mathbf{E}_j - \mathbf{E}_{ij} - \mathbf{E}_{ji})\mathcal{P}(\phi_{t+dt} = \mathbf{e}_j \,|\, \phi_t = \mathbf{e}_i).$$

The general expression for $d\langle m, m; \mathcal{F}_t\rangle_t = E[\Delta m_t \Delta m_t' \,|\, \mathcal{F}_{t-}]$ is given as a function of Q in the Appendix 1.

Sometimes martingales with continuous paths (e.g., $\{w_t\}$) appear in combination with martingales with discontinuous paths (e.g., $\{m_t\}$) to form a *composite* martingale $\{\eta_t\}$. In fact, any $\mathcal{F}_t$-martingale can be separated into its continuous and discontinuous parts: $\eta_t = \eta_t^c + \eta_t^d$ where $\{\eta_t^c\}$ is a continuous process and $\{\eta_t^d\}$ is purely discontinuous. The two are mutually orthogonal: $d\langle \eta_t^c, \eta_t^d; \mathcal{F}_t\rangle_t = 0$ [Ell82, Chapter 9]. For example, $d\langle w, m; \mathcal{F}_t\rangle_t = E[dw_t dm_t' \mid \mathcal{F}_t] = 0$. Additionally, two purely discontinuous processes without common jump times are orthogonal: If $\{\Delta\phi_t \Delta\psi_t'\}$ is essentially the zero process, $d\langle \phi, \psi; \mathcal{F}_t\rangle_t = 0$.

The composite martingale $\{\eta_t\}$ has associated with it another quadratic process. The *optional quadratic variation*, $[\eta, \eta]_t$, is determined from its increments: $d[\eta, \eta]_t = d\eta_t d\eta_t'$. It can also be found by adding the outer product of the jumps in $\{\eta_t\}$ to the predictable quadratic variation:

$$d[\eta, \eta]_t = d\langle \eta_t^c, \eta_t^c; \mathcal{F}_t\rangle_t + \Delta\eta_t \Delta\eta_t'.$$

The *optional cross quadratic variation* of two martingales is similarly defined [Kri84, Chapter 4].

Stochastic Differential Equations

Equation (0.5) relates the actuating signal to the system state. This is an integral equation with differential embodiment given in (0.4). For this model to be useful, each of the terms on the right side of (0.5) must be given clear meaning. The first integral, $\int_{[0,t]} \mathbf{f}(x_s, u_s)ds$, is of a conventional sort if the sample functions of $\{\mathbf{f}(x_t, u_t)\}$ are well behaved. The second integral, $\int_{[0,t]} \mathbf{g}(x_s, u_s)d\eta_s$, is more problematic. The integrand $\{\mathbf{g}(x_t, u_t)\}$ is an $\mathcal{F}_t$-random process and in what follows $\{\eta_t\}$ is an $\mathcal{F}_t$-martingale. It is advantageous to define the integral using the predictable version of the integrand; that is, $\int_{[0,t]} \mathbf{g}(x_s, u_s)d\eta_s$ is better written $\int_{[0,t]} \mathbf{g}(x_{s-}, u_{s-})d\eta_s$ [Ell82, Theorem 11.44]. For consistency, the stochastic differential equation could be written

$$dx_t = \mathbf{f}(x_{t-}, u_{t-})\, dt + \mathbf{g}(x_{t-}, u_{t-})\, d\eta_t$$

since the increment in $\{x_t\}$ depends upon the antecedent values of the arguments rather than their current values. For simplicity, we will not distinguish the predictable versions of the random processes in the differential equations even though the left continuous version of the integrands will appear in the integrals.

The output of the system is represented with a stochastic differential equation too:

$$dg_t = \mathbf{r}(x_t, u_t)\, dt + \mathbf{s}(x_t, u_t)\, dn_t, \tag{0.6}$$

where $\{g_t\}$ is the output process or the observation process as appropriate. The components of the $\mathcal{F}_t$-martingale $\{n_t\}$ that appear in (0.6) would be called *observation noise* or the equivalent. It is through $\{g_t\}$ that the value of $\{x_t\}$ can be determined. Let the filtration generated by $\{g_t\}$ be labeled $\{\mathcal{G}_t\}$. This output filtration is a subfiltration of $\mathcal{F}_t$. For any $\mathcal{F}_t$-random process $\{\zeta_t\}$, denote the $\mathcal{G}_t$-conditional expectation with a circumflex and the $\mathcal{F}_t$-conditional error with a tilde: For example, $\hat{\zeta}_t = E[\zeta_t \,|\, \mathcal{G}_t]$; $\tilde{\zeta}_t = \zeta_t - \hat{\zeta}_t$.

An important process related to the $\mathcal{G}_t$-mean is the *innovation process*. The innovation process, labeled $\{\nu_t\}$, is generated from its increments:

$$d\nu_t = dg_t - E[dg_t \,|\, \mathcal{G}_t].$$

This terminology is "motivated by the observation that, formally, $\nu_{t+h} - \nu_t$ represents the 'new' information about (the system state) obtained from observations between t and $t + h$" [Ell82, Definition 18.6].

Sometimes the observations are not time continuous but instead have a natural sampling interval. An example of this is a radar tracking an aircraft. The aircraft path is continuous (i.e., modeled as in (0.4)), but the observations occur every T seconds beginning at $t = 0$. In this case, we would replace (0.6) with

$$g[k] = \mathbf{r}(x[k], u[k]) + \mathbf{s}(x[k], u[k])n[k], \tag{0.7}$$

where $g[k]$ is the output (or observation) at time $t = kT$, and similarly for $x[k]$ and $u[k]$. In this model, $\{n[k]\}$ is not typically an $\mathcal{F}_t$-martingale but may be a sequence of martingale increments ($E[n[k] \mid \mathcal{F}_{kT}] = 0$). The output sequence, $\{g[k]\}$, generates a filtration, $\{\mathcal{G}_t\}$, which is defined for all $t \in [0, \top]$, not just $t \in \{0, T, 2T, ...\}$. However, since new information appears at the output at distinct times, it is true that $\{\hat{x}_t\}$ tends to have discontinuities at sample times.

In some circumstances, the system state is time discrete:

$$x[k+1] = \mathbf{f}(x[k], u[k]) + \mathbf{g}(x[k], u[k])\eta[k+1]. \tag{0.8}$$

If the state and measurement grid are the same, a time-discrete system with time-discrete measurements has a structure like that given above with natural changes in terminology.

Some Useful Results from Martingale Theory

This section lists some useful results from martingale theory. The statements do not include certain qualifications to be found in the references [Ell82].

Definition 1 *The process $\{X_t\}$ is **corlol** (for continuous on the right, limits on the left) if there is a modification of $\{X_t\}$ such that*

$$X_t(\omega) = \lim_{s \to t^+} X_s(\omega)$$

and

$$X_{t^-}(\omega) = \lim_{s \to t^-} X_s(\omega).$$

Definition 2 *Given any process $\{X_t\}$ adapted to $\mathcal{F}_t$, if there exists a process $\{A_t\}$ such that $A_0 = 0$, $\{A_t\}$ is $\mathcal{F}_t$-predictable, $\{A_t\}$ has corlol sample paths of locally finite variation, and $\{X_t - A_t\}$ is an $\mathcal{F}_t$-martingale, then $\{A_t\}$ is called the predictable compensator of $\{X_t\}$ relative to $\mathcal{F}_t$*

Theorem 1 (Doob–Meyer Decomposition Theorem) *If the random process $\{X_t\}$ has a predictable compensator, then it is unique in the sense that any two predictable compensators are equal to each other for all t.*

This statement of the Doob–Meyer Decomposition Theorem is Proposition 3.2 in [WH85]. See [DM82] for a proof.

Theorem 2 (Martingale Representation Theorem) *Suppose the filtration $\{\mathcal{F}_t\}$ is generated by the local semimartingale $X_t = B_t + W_t$, where $\{B_t\}$ is of bounded variation and $\{W_t\}$ is a Brownian motion or a point process. Then any $\mathcal{F}_t$-local semimartingale $\{Z_t\}$ can be written as a stochastic integral against $\{W_t\}$. That is, there exists an $\mathcal{F}_t$-predictable function $\{\gamma_t\}$*

such that

$$Z_t = Z_0 + \int_0^t \gamma_s \, dW_s.$$

The Brownian motion version is due to a generalization by Fujisaki et al. [FKK72] of the works of Itô [Itô51], Kunita and Watanabe [KW67], and Clark [Cla70]. The extension to point processes is due to Brèmaud [Brè72].

Theorem 3 *Let $\{S_t\}$ be a process, not necessarily adapted to $\mathcal{G}_t$, a filtration generated by the continuous or purely discontinuous martingale $\{\nu_t\}$. Let $\hat{S}_t = E[S_t|\mathcal{G}_t]$, and define a process $\{B_t\}$ by $dB_t = E(dS_t|\mathcal{G}_t)$. Then $\{\hat{S}_t - B_t\}$ is a $\mathcal{G}_t$-martingale, and there exists a $\mathcal{G}_t$-predictable function γ_t such that*

$$d\hat{S}_t = E(dS_t|\mathcal{F}_t) + \gamma_t \, d\nu_t.$$

Proof ([WH85]): Since $\mathcal{G}_t$ is increasing,

$$\begin{aligned} E(d\hat{S}_t \,|\, \mathcal{G}_t) &= E\{[E(S_{t+dt} \,|\, \mathcal{G}_{t+dt}) - E(S_t \,|\, \mathcal{G}_t)]\} \\ &= E(S_{t+dt} \,|\, \mathcal{G}_t) - E(S_t) \,|\, \mathcal{G}_t) \\ &= dB_t. \end{aligned}$$

Therefore, $\hat{S}_t - B_t = M_t$ is a $\mathcal{G}_t$-martingale. By Theorem 2, M_t can be represented as a stochastic integral against $\{\nu_t\}$.

Theorem 3 provides a cornerstone for system estimation theory. It implies that under modest conditions, the estimator of S_t is the solution to a stochastic differential equation driven by the innovation process.