

Estimation Problems in Hybrid Systems

DAVID D. SWORDER
University of California, San Diego

JOHN E. BOYD
Cubic Defense Systems, Inc.



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

© Cambridge University Press 1999

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1999

Printed in the United States of America

Typeface Times Roman 11/14 pt. *System* L^AT_EX 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

<i>List of Illustrations</i>	<i>page</i> ix
<i>Preface</i>	xiii
1 Hybrid Estimation	1
1.1 Introduction	1
1.2 A Tracking Example	14
1.3 Infinite-Dimensional Algorithms	21
1.4 Multiple Model Algorithms	23
1.5 Modal Observations	34
2 The Polymorphic Estimator	41
2.1 Introduction	41
2.2 Modal Estimation	44
2.3 The Polymorphic Estimator	47
2.4 The Abridged PME: Time-Continuous Plant and Observation	50
3 Situation Assessment	56
3.1 Introduction to Situation Assessment	56
3.2 Decision Maker Dynamics	58
3.3 Order Bias in Human Decision Makers	60
3.4 Multilevel Situation Assessment	67
3.5 Decision-Making Phenotypes: An Example	72
3.6 Conclusions	82
4 Image-Enhanced Target Tracking	83
4.1 Tracking an Agile Target	83
4.2 Image Modeling and Interpretation	88

4.3	Tracking Maneuvering Targets	92
4.4	An Example: An Antiship Missile	97
4.5	Renewal Models for Maneuvering Targets	105
4.6	Performance Contrasts with Different Lifetime Modeling	107
5	Hybrid Plants with Base-State Discontinuities	117
5.1	Plant State Discontinuities	117
5.2	Plant State Rotation	119
5.3	A Maneuvering Aircraft with Variable Drag	122
5.4	Plant State Translation	126
5.5	A Maneuvering Aircraft with Sudden Translations	130
5.6	Variable Set Points	134
5.7	Estimating the Temperature of a Solar Panel	136
6	Mode-Dependent Observations	142
6.1	Problem Definition	142
6.2	The PME	143
6.3	Modal Estimation Using the PME	145
6.4	A Maneuvering Target Employing Countermeasures	148
7	Control of Hybrid Systems	155
7.1	Feedback Regulation of Hybrid Systems	155
7.2	The PME	161
7.3	An RPV Subject to Subsystem Failure	164
8	Target Recognition and Prediction	171
8.1	Problem Statement	171
8.2	Recognition and Tracking a Maneuvering Target	172
8.3	Automatic Target Recognition	175
8.4	Path Prediction	186
8.5	A Missile Test in Australia	188
9	Hybrid Estimation Using Measure Changes	197
9.1	Change of Measure	197
9.2	Gaussian Minimum Shift Keying	209
9.3	An Example	217
Appendix 1	PME Derivation Details	221
Appendix 2	COM Derivation Details	251

<i>Bibliography</i>	257
<i>Index</i>	263
<i>Glossary</i>	267

1

Hybrid Estimation

1.1 Introduction

Common problems in design require that an engineer devise a control or decision algorithm that converts measurements of system and environmental variables into signals that aid in system regulation. For example, a control node converts sensor outputs into an actuating signal that moves the system toward the desired operating point and keeps it there. At this foundational level, the engineer must formulate a mapping from the system observables into an action or report; for example, a feedback regulator converts the measured outputs of the system to be controlled (the *plant*) into an input that stabilizes the system.

Design is made difficult by disturbances internal to the system and by noise at its output. For example, there may be no sensors that measure those plant variables most useful for regulation, or, if measured, the variables may be masked by noise in the sensor-to-regulator link. Lacking omniscience, an engineer must process the available measurements to produce a good approximation to relevant but “hidden” variables. And this inference must be done on-line. The processing algorithm must not only be adapted to the incoming data stream, it must be of a form that can be implemented: An implementable estimation algorithm is an explicit mapping of the sensor output process (the *measurements*) into a (nearly) concurrent estimate of the required variables. In the applications studied here, the need for contemporaneous response limits consideration to finite-dimensional recursive algorithms; new observations are integrated into an estimate in an accretive manner.

Analytical design in estimation and control begins with a formal mathematical description of the system to be controlled (the *plant model*). The model delineates the response of the plant to endogenous actuating signals as well as representing the influence of exogenous disturbances common to the application. The system designer selects a control policy or a state estimation algorithm based in large part upon the behaviors predicted by the model. The practicality of analytic procedures is

linked closely to the realism of the plant model. However, realism must be tempered by the need to have a model that is simultaneously flexible and tractable.

One useful paradigm phrases the plant model in terms of a set of nonlinear stochastic differential equations. Let us start with a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and a time interval of interest, $[0, T]$. On this space there is a right-continuous filtration $\{\mathcal{F}_t; 0 \leq t \leq T\}$ and right-continuous, \mathcal{F}_t -adapted random processes, $\{\Phi_t\}$, $\{w_t\}$, and $\{n_t\}$. Subject to initial conditions χ_0 and g_0 , the plant model is written:

plant model

$$d\chi_t = \mathbf{f}(\chi_t, v_t, \Phi_t) dt + \mathbf{g}(\chi_t, v_t, \Phi_t) dw_t, \quad (1.1)$$

$$dg_t = \mathbf{r}(\chi_t, v_t, \Phi_t) dt + \mathbf{s}(\chi_t, v_t, \Phi_t) dn_t, \quad (1.2)$$

where $\{v_t\}$ is an s -dimensional actuating process (the *plant input*), $\{g_t\}$ is an r -dimensional observation process (the *plant output*), and $\{\chi_t\}$ is an n -dimensional internal process (the *plant state*). Equation (1.1) describes the temporal evolution of the internal variables within the plant, and (1.2) describes the sensor outputs available for estimation and/or control.

This plant model is more complicated than that encountered in introductory studies of feedback control. In applications, even when the actuating process is specified, the realizations of the state and output paths are unpredictable – there are many effects not well captured in a deterministic model. Chance influences in the plant and sensor are represented by the stochastic processes in (1.1) and (1.2). Various accretive effects are represented by $\{w_t\}$ and $\{n_t\}$; for example, $\{w_t\}$ could describe the high frequency modes ignored in a low-dimensional plant model, and $\{n_t\}$ could describe noise at the sensor output. The *environmental process*, $\{\Phi_t\}$, denotes external conditions of a more global sort that affect plant operation. The value of $\{\Phi_t\}$ might indicate the operational status of a subelement within the plant, external conditions that influence the plant dynamics (e.g., temperature), the level of loads placed upon the system by linked elements, etc. In contrast to $\{w_t\}$ and $\{n_t\}$, which tend to be aggregations of small increments, Φ_t may symbolize temporally distinct events. (Friedland called Φ_t the *metastate* when used in the context of adaptive control; see [Fri96, Chapter 10].) All of these disturbance processes are viewed by the designer as exogenous.

In both estimation and the control, the output signal, $\{g_t\}$, is processed to create causal estimates of important system variables. A *filter* provides estimates of the current values of both the plant state vector and the environmental process. A *predictor* estimates future values of the same variables. Often, the environmental process has a character fundamentally different from the plant state. The value of Φ_t

may be a symbolic variable (e.g., $\Phi_t \in \{\text{normal operation, degraded operation}\}$). In this event, the average value of Φ_t has no meaning. Rather, the probability distribution of Φ_t is required to properly assess the status of the plant. Denote the filtration generated by $\{g_t\}$ by $\{\mathcal{G}_t\}$. If mean square error is used as a performance index, the estimation problem can be posed as follows:

Find an explicit processing algorithm to generate (or approximate) the mean plant state $\hat{\chi}_t = E[\chi_t | \mathcal{G}_t]$ and the \mathcal{G}_t -probability distribution of Φ_t .

There are applications in which even this will not suffice and more comprehensive statistical properties of the plant processes are required.

Unfortunately, even when formal descriptions of the exogenous processes are integrated with (1.1) and (1.2), an elementary solution to this estimation problem does not currently exist. There is, however, one special case in which astounding success has been achieved. So much so that the solution thus derived is used in circumstances far removed from those in which it was developed. Specifically, suppose that the system has “smooth” nonlinearities, that the plant noise, $\{w_t\}$, is a Brownian motion, and that the environmental process, $\{\Phi_t\}$, is constant with known value Φ_c . Associated with Φ_c there is a nominal operating condition, both in the state and in the actuating signal labeled (χ_n, ν_n) . Frequently (χ_n, ν_n) is a condition of plant stasis: $f(\chi_n, \nu_n, \Phi_c) = 0$. The operating condition (or *regime*) is known by different names: in the process control industry, (χ_n, ν_n) is referred to as the set point or the operating point; in aircraft flight control, (χ_n, ν_n) is referred to as the trim condition; in other applications, (χ_n, ν_n) is simply the reference point. We will use these terms interchangeably and note in this context that Φ_t simply points to the operating mode or regime with its value having no intrinsic meaning.

For a particular regime, there is a local description of the plant phrased in terms of a set of perturbation variables. These are defined as the (usually small) deviations in state and excitation from the set point: $x_t = \chi_t - \chi_n$; $u_t = \nu_t - \nu_n$. Using orthodox methods and neglecting higher order terms, the perturbation processes are commonly represented by a linear stochastic differential equation with initial condition taken to be Gaussian: x_0 is $\mathbf{N}(\hat{x}_0, P_{xx}(0))$, and

$$dx_t = (Ax_t + Bu_t) dt + C dw_t, \quad (1.3)$$

where $\{w_t\}$ is a Brownian motion with intensity $W(d\langle w, w \rangle_t = W dt)$. Call $\{x_t\}$ the *base-state process* to distinguish it from the plant state process, $\{\chi_t\}$; call $\{u_t\}$ the *regulation signal* to distinguish it from the plant input, $\{\nu_t\}$. Equation (1.3) relates the base-state to the inputs $\{u_t\}$ (endogenous) and $\{w_t\}$ (exogenous). The base-state excitation is a Brownian motion with intensity $CWC' = R_x$. Of course, if the plant is linear over a large region of the state space, (1.3) is valid without consideration of the set point. In such applications, it is understood that χ_n and ν_n are both zero.

The set point is known ($\mathcal{P}[\Phi_t \equiv \Phi_c] = 1$) and need not be estimated, but the plant state is frequently not known and must be inferred from sensor outputs. Suppose a sensor provides a noisy but linear plant state measurement,

plant state measurement: time-continuous

$$dy_t = H\chi_t dt + dn_t, \quad (1.4)$$

where $\{n_t\}$ is a Brownian motion independent of $\{w_t\}$, with intensity $R_x > 0$ ($d\langle n, n \rangle_t = R_x dt$), and $y_0 = 0$. By subtracting the contribution of the set point from the output, (1.4) can be written as a noisy, linear measurement of the base-state: $dy_t - H\chi_n dt = Hx_t dt + dn_t$. The innovation increment $dv_t = dy_t - d\hat{y}_t$ can be written $H\tilde{x}_t dt + dn_t$, where $\tilde{x}_t = x_t - \hat{x}_t$. When there is only one sensor, $g_t \equiv y_t$. To differentiate this case from others that follow, denote the filtration generated by $\{y_t\}$ by $\{\mathcal{Y}_t\}$ ($= \mathcal{G}_t$ in this case), where a circumflex may be used to denote \mathcal{Y}_t -expectation if no confusion will result. Equations (1.3) and (1.4) will be called a *linear-Gauss-Markov* (LGM) model even when x_0 is not Gaussian. Although the observation is unconventional, the regime offset is known and is accommodated in a direct fashion. The base-state estimator is known for the LGM problem: the Kalman filter. The Kalman filter generates $\{\hat{x}_t\}$ using a simple recursive algorithm. The plant state estimator is $\hat{\chi}_t = \chi_n + \hat{x}_t$.

In the systems we will study, $\{\Phi_t\}$ is not nearly so obliging. Instead of a single operating point, $\{\Phi_t\}$ may move about in its range space in response to the macroevents that influence the plant. The temporal structure of the regime process has a fundamental impact on system analysis. If, for example, $\{\Phi_t\}$ has sample paths that are well described by a diffusion process, then $\{\Phi_t\}$ can be integrated into (1.1) as an additional plant state. This is an attractive option when the time constants of $\{\Phi_t\}$ are comparable with those of the plant, though this inclusion compounds the plant nonlinearity.

In other applications, $\{\Phi_t\}$ has a distinguishing feature that precludes orthodox state augmentation. Suppose the plant has S possible operating regimes, and at any particular time, Φ_t takes on a value selected from a set of size S : $\Phi_t \in \{\Phi_i; i \in \mathbf{S}\}$. The plant now has S possible reference points (or set points, etc.), and these are identified with the S possible values of $\{\Phi_t\}$; that is, there are S vector pairs, $\{(\chi_i, v_i); i \in \mathbf{S}\}$, which designate the S relevant stasis conditions for the plant. For example, the k th nominal operating point for the plant is (χ_k, v_k) , and if $\Phi_t = \Phi_k$ the plant input and state should be near (χ_k, v_k) .

For simplicity, array the nominal states (respectively nominal actuating signals) as an $n \times S$ matrix χ (respectively an $s \times S$ matrix v): $\chi = [\chi_i]$ (respectively

$\mathbf{v} = [v_i]$). During operation, the system will operate in one regime for a time ($\Phi_t = \Phi_i$ for $t \in [a, b)$) and then suddenly shift ($\Phi_b = \Phi_j$) to another in response to an external event or change in the surrounding environment. In most applications, the discontinuous sample paths of $\{\Phi_t\}$ are an approximation to the continuous though abrupt modal transitions that actually occur. Nevertheless, the representation of $\{\Phi_t\}$ with a process of piecewise constant paths is a useful abstraction when the interval over which the modal transition takes place is short as compared to the important time constants of the plant.

Since the environmental process has a finite state space, $\{\Phi_t\}$ can be represented using a more illuminating notation. Let ϕ_t be a pointer to the current regime: The state space of ϕ_t consists of the S canonical unit vectors in \mathbb{R}^S ($\phi_t \in \{\mathbf{e}_1, \dots, \mathbf{e}_S\}$). The component in ϕ_t with value one marks the current mode of operation: If $\Phi_t = \Phi_k$ then $\phi_t = \mathbf{e}_k$. The $\{\phi_t\}$ process is called the *modal-state process* to differentiate it from the base-state process. The base-state variables are deviations from the current set point: $x_t = \chi_t - \chi\phi_t$; $u_t = v_t - \mathbf{v}\phi_t$. The comprehensive state of the system is the composition of the base- and modal-states: The *zygostate* is the pair (x_t, ϕ_t) . Since ϕ_t is an indicator vector, the expectation of the modal-state is actually the conditional probability vector $\hat{\phi}_t = [\mathcal{P}\{\phi_t = \mathbf{e}_i | \mathcal{G}_t\}]$.

Control in a multiregime environment presents some subtle challenges. When the regime is known and constant (e.g., $\phi_t \equiv \mathbf{e}_i$), the actuating signal has a natural decomposition ($v_t = u_t + \mathbf{v}\phi_t$) into a feedforward component associated with the set point ($\mathbf{v}\mathbf{e}_i = v_i$) and a feedback component (u_t) that maintains the plant state near the set point ($\chi_t \approx \chi_i$). When the modal-state is neither known nor measured, this implementation is not possible because proper feedforward control cannot be generated. In applications, a variety of replacements for $\{\mathbf{v}\phi_t\}$ have been proposed. We will not explore issues of feedforward control in any depth here. We will simply employ $\{\mathbf{v}\hat{\phi}_t\}$ as the “feedforward” component of the actuating signal: Ideal set point actuation will be replaced with its expectation. Note, however, that a failure to generate the proper feedforward actuating signal has an influence that must be included in the base-state dynamics.

A comprehensive plant model requires a representation of evolution, both intramodal and intermodal. Consider the former first. During an extended (known) modal sojourn, proper control will place and maintain the plant state vector near the correct set point. The natural plant model in this circumstance would be that local model, selected from a family of regime-specific, linear models, associated with the present mode of operation. The modal-state is a pointer, and the intrasojourn model can be written:

$$dx_t = \sum_i ((A_i x_t + B_i (u_t + \mathbf{v}(\hat{\phi}_t - \mathbf{e}_i))) dt + C_i dw_t) \phi_i, \quad (1.5)$$

where $\{A_i, B_i, C_i; i \in \mathbf{S}\}$ are determined from (1.1) in precisely the way (1.3) was in the unimodal (or unimorphic[†]) case.

Suppose the plant is in the i th mode ($\phi_t = \mathbf{e}_i$) and the modal estimate is a good one ($\hat{\phi}_t \approx \mathbf{e}_i$). The base-state dynamic equation is the i th selection from the family of models: $(A_i x_t + B_i u_t) dt + C_i dw_t$. The exogenous excitation is a Brownian motion with intensity $R_\chi(i) = C_i W C_i'$. There is an atypical term in (1.5) that is connected with failure to implement the proper feedforward excitation ($-B_i v \tilde{\phi}_t \phi_i dt$). When the estimate of ϕ_t is good, this last term is negligible, and the intramodal dynamics are LGM.

The intramodal representation is but a part of the model of plant evolution. When the regime changes, many things can happen to the plant state. There will be no attempt to be exhaustive in this list, but we will encounter situations in which the plant state translates, rotates, and/or is scaled. More specifically, suppose $\{\Phi_t\}$ makes the transition $\mathbf{e}_i \mapsto \mathbf{e}_l$ at time t . Then $\{\chi_t\}$ may experience:

Translation: $\Delta \chi_t = \rho(i, l); i \neq l$.

Rotation and/or scaling: $\Delta \chi_t = M(i, l) \chi_{t-}; i \neq l$,

where $\Delta \chi_t = \chi_t - \chi_{t-}$.

When the mode changes, the plant state may be transformed in a way that creates a path discontinuity. This abrupt change in plant state is an approximation in most cases. But, if the interval over which a change takes place is small, a discontinuous path model may provide a far simpler representation of the state variation than would a continuous path model created from an intricate diffusion process. To fill out the list of transformation matrices, let $\rho(i, i) = 0$, $M(i, i) = 0$; $i \in \mathbf{S}$. The indicator vector of the discontinuity event $\mathbf{e}_i \mapsto \mathbf{e}_l$ at time t can be written as $\phi_i \mathbf{e}_l' \Delta \phi_t$. The plant state discontinuity can be written explicitly as

$$\Delta \chi_t = \sum_{i,l} (M(i, l) \chi_{t-} + \rho(i, l)) \phi_i \mathbf{e}_l' \Delta \phi_t.$$

Discontinuities in $\{\chi_t\}$ are reflected directly in $\{x_t\}$, but there is an additional source of base-state discontinuity. When the mode changes $\mathbf{e}_i \mapsto \mathbf{e}_l$, the base-state reference level changes from χ_i to χ_l . Even if the plant state were continuous, the base-state would experience a discontinuity:

$$\Delta x_t = -\chi \Delta \phi_t.$$

These intermodal transition conditions can be combined to yield the base-state

[†] We say a system may have one or several modes or, equivalently, forms. Hence a single-mode plant is called unimodal (or unimorphic) to distinguish it from a polymodal (polymorphic) system.

discontinuity model:

$$\Delta x_t = \sum_{i,l} (M(i,l)x_{t-} + (\chi_i - \chi_l) + M(i,l)\chi_i + \rho(i,l))\phi_i \mathbf{e}'_l \Delta \phi_t. \quad (1.6)$$

Now combine the intermodal discontinuity with the intramodal dynamics to yield:

base-state model

$$\begin{aligned} dx_t = & \sum_i ((A_i x_t + B_i(u_t - \mathbf{v}\phi_t)) dt + C_i dw_t)\phi_i + \sum_{i,l} (M(i,l)x_t \\ & + (\chi_i - \chi_l) + M(i,l)\chi_i + \rho(i,l))\phi_i \mathbf{e}'_l \Delta \phi_t. \end{aligned} \quad (1.7)$$

Equation (1.7) is the fundamental model of time-continuous base-state evolution. Its appearance is formidable. Be assured that while the various discontinuity and set point conditions will appear in what follows, in no application will all occur simultaneously! In many cases, (1.7) takes on a strikingly simpler form. It is advantageous to set apart some special instances of (1.7) because they are easier to interpret.

LJS: The most often studied specialization of (1.7) is called a linear jump system (LJS). In an LJS there is no regime-specific set point reference ($\chi = 0$, $\mathbf{v} = 0$), nor are there plant state discontinuities at modal transition [Mar90]. The LJS model is simply

$$dx_t = \sum_i ((A_i x_t + B_i u_t) dt + C_i dw_t)\phi_i. \quad (1.8)$$

Often the intensity of the Brownian excitation is constant across regimes and there is no feedback control:

$$dx_t = \sum_i A_i x_t \phi_i dt + C dw_t. \quad (1.9)$$

We will find this simpler model to be useful in certain tracking applications.

JTS: In some applications, the plant state discontinuity has a particular structure. There is neither rotation nor scaling. The plant state discontinuity is a translation in the form of a difference between mode-specific levels: $\rho(i,l) = \rho_l - \rho_i$. Array these levels as rows of an $s \times n$ matrix $\rho = [\rho_i]$. The base-state dynamic equation of a jump translating system

(JTS) can be written

$$dx_t = \sum_i ((A_i x_t + B_i(u_t - v\phi_t)) dt + C_i dw_t)\phi_i + (\rho' - \chi) d\phi_t. \quad (1.10)$$

If the plant state is a continuous process and there is no control, the JTS-model becomes even simpler:

$$dx_t = \sum_i A_i x_t \phi_i dt - \chi d\phi_t + C dw_t, \quad (1.11)$$

where the model is shown with constant Brownian intensity.

In interpreting the results derived on the basis of (1.7), we should recognize the approximations inherent in the model. If we ignore the drift identified with the feedforward implementation, the intrasojourn base-state dynamics are LGM. This model is easily justified in a region about the set point where higher order deviation variables are negligible. Exactly this kind of linearization procedure is accepted practice in applications involving unimodal plants, and during quiescent periods, Equation (1.5) – the intermodal restriction of (1.7) – is reasonable. If the set point changes, the magnitude of the base-state vector will increase abruptly. The state of a well-regulated plant will move expeditiously toward the new set point. In (1.7) the evolution model uses the dynamics of the successor regime. There are systems for which this concatenation of local models would be inappropriate (e.g., an unstable system moves away from the new set point). We will not pursue this issue further and will accept (1.7) as an adequate for our purposes.

The comprehensive plant state (base, mode) is a combination of continuous and discrete elements. The base-state moves within \mathbb{R}^n , and though $\phi_t \in \mathbb{R}^s$, the modal-state has a finite range space. The modal process is usually thought to be exogenous: The path of $\{\Phi_t\}$ is indifferent to $\{x_t\}$. Because it modulates the base-state motion, $\{\Phi_t\}$ is not, however, independent of $\{x_t\}$. With this heterogeneous state space structure, such plants are called *hybrid*. Heterogeneity of various kinds is becoming more common in applications, and the adjective “hybrid” is applied quite broadly. Nevertheless, because it is so descriptive, we will use hybrid to refer to plants and systems with this state space decomposition.

To complete the plant model, the temporal evolution of the modal-state must be quantified. In much of what follows, $\{\phi_t\}$ will be represented by an \mathcal{F}_t -Markov process satisfying the stochastic differential equation:

modal-state model

$$d\phi_t = Q' \phi_t dt + dm_t \quad (1.12)$$

with initial condition ϕ_0 . The $S \times S$ matrix Q is called the modal transition rate matrix: If $i \neq j$, $\mathcal{P}(\phi_{t+dt} = \mathbf{e}_j | \phi_t = \mathbf{e}_i) = Q_{ij} dt$ with $Q_{ii} = -\sum_{l \neq i} Q_{il}$. The off-diagonal elements of the Q -matrix are nonnegative. The diagonal elements are such as to make the row sums of Q equal zero. It is known that the mean sojourn time in state $\phi_t = \mathbf{e}_i$ is $-1/Q_{ii}$, and if $\phi_t = \mathbf{e}_i$, the probability that the next modal transition will be $\mathbf{e}_i \mapsto \mathbf{e}_j$ is $-Q_{ij}/Q_{ii}$. Consequently, Q can be particularized from observations of the modal process. The second term in (1.12) is a purely discontinuous \mathcal{F}_t -martingale increment: $E[dm_t | \mathcal{F}_t] = 0$.

Equation (1.12) can be integrated into (1.7). Note that $\phi_i \mathbf{e}'_l d\phi_t = (Q_{il} dt + dm_l)\phi_i$. So

$$\begin{aligned} dx_t = & \sum_i ((A_i x_t + B_i(u_t - \mathbf{v}\phi_t)) dt + C_i dw_t)\phi_i + \sum_{i,l} (M(i,l)x_t \\ & + (\chi_i - \chi_l) + M(i,l)\chi_i + \rho(i,l))(Q_{il} dt + dm_l)\phi_i. \end{aligned} \quad (1.13)$$

Though not a particularly appealing relation, (1.13) can be made easier to interpret if we collect some of the terms that have a common influence. Let

$$\begin{aligned} \mathbf{A}_i &= A_i + \sum_l Q_{il} M(i,l), \\ \Theta(i,l) &= \chi_i - \chi_l + M(i,l)\chi_i + \rho(i,l), \\ \rho_i &= \sum_l \Theta(i,l) Q_{il}. \end{aligned} \quad (1.14)$$

In these terms, the base-state model can be written

$$\begin{aligned} dx_t = & \sum_i ((\mathbf{A}_i x_t + B_i(u_t - \mathbf{v}\phi_t)) dt + C_i dw_t)\phi_i \\ & + \sum_{i,l} (M(i,l)x_t + \Theta(i,l))\phi_i dm_l + \rho' \phi_t dt. \end{aligned} \quad (1.15)$$

The equation of base-state dynamics has the general appearance of an LGM model but it differs in important particulars. The state matrix, \mathbf{A}_i , of $\{x_t\}$ is composed of the intramodal component (A_i) plus a component determined by both the direction of the linear, intermodal discontinuity and its likelihood ($\sum_l Q_{il} M(i,l)$). The control matrix, B_i , is that of the intramodal model. The translational discontinuity in the plant state is reflected in $\rho' \phi_t dt$. There is a collection of terms in the drift of $\{x_t\}$ not found in the classical models of control and estimation. The model is highly nonlinear with the modal-state a multiplier throughout.

The increment in $\{x_t\}$ also contains exogenous forcing terms. One is a wideband noise term ($C_i dw_t$) also found in LGM models. The other is neither linear nor Gaussian. The plant state discontinuity term, $\sum_{i,l} (M(i,l)x_t + \Theta(i,l))\phi_i dm_l$, is

an increment of a purely discontinuous martingale. The coefficient, $(M(i, l)x_t + \Theta(i, l))\phi_i$, contains base- and modal-state products.

The specialized dynamics of an LJS are not changed when the modal process is Markovian because the modal dynamics do not enter the base-state equation. The base-state evolution of the JTS can be written:

$$dx_t = \sum_i ((A_i x_t + B_i(u_t - v\phi_t) + (\rho' - \chi)Q'e_i) dt + C_i dw_t)\phi_i + (\rho' - \chi) dm_t. \quad (1.16)$$

Equation (1.16) contains the same types of excitation found in the more comprehensive model, (1.15), but the simpler structure of (1.16) will be reflected in the estimation algorithm; compare $(\rho' - \chi) dm_t$ with $\sum_{i,l} (M(i, l)x_t + \Theta(i, l))\phi_i dm_t$.

In this book, we will present algorithms for generating (or approximating) $\{\hat{x}_t\}$ and $\{\hat{\phi}_t\}$. The accuracy of the estimates depends upon the quality and kind of sensors available in the application. A model for one kind of sensor is displayed in (1.4). The measurement is time continuous, linear in plant state, and the noise is additive and Gaussian. We will refer to (1.4) as the model of the *plant state sensor* even though $\{y_t\}$ may be generated by a collection of individual devices arrayed in a suite. For example, there may be radars aboard a set of geographically diverse platforms (shipboard, land-based, and air-based) with all tracking the same target. It is this aggregate that is called the plant state sensor. The noise in the observation is determined by both the sensor and the geometry (e.g., range), after linearization if necessary.

When the measurement frequency is too slow to justify using (1.4), the plant state sensor outputs are more accurately viewed as a time-discrete sequence. Suppose observations occur with intersample period T . A linear, time-discrete measurement of the plant state at time $t = kT$ is a direct analogue of (1.4):

plant state measurement: time-discrete

$$y[k] = H\chi[k] + n[k], \quad (1.17)$$

where $\{n[k]\}$ is a Gaussian white noise process with covariance R_x ($R_x > 0$), independent of the exogenous processes in (1.13). As is the case when the measurements are time continuous, if $\{\phi_t\}$ is known, $\{y[k]\}$ can be recast as a measurement of the base-state uncontaminated by the mode: $y[k] - H\chi\phi[k] = Hx[k] + n[k]$, and the *measurement residual* is defined to be the difference between what the output is and what it is predicted to be:

$$r[k] = y[k] - E[y[k]|\mathcal{G}[k-1]] = H\tilde{x}[k]^- + n[k].$$

In this case, $r[k]$ is equal to the innovations increment $\Delta v[k]$. With imperfect knowledge of $\{\phi[k]\}$, the measurement residual (and the innovations increment) is

$$r[k] = H\tilde{x}[k]^- + H\chi\tilde{\phi}[k]^- + n[k].$$

There is thus a mixing of base-state and modal-state errors in which the base-state error, $\tilde{x}[k]^-$, is conflated with a *base-state equivalent error*, $\chi\tilde{\phi}[k]^-$. Of course, during long sojourns in a regime, the modal-state is probably identified rather well ($\tilde{\phi}[k] \approx 0$), and the observation reverts to its orthodox form.

For LJS with known modal path, the Kalman filter generates the conditional mean of the base-state for either time-continuous or time-discrete measurements. Look at the time-continuous case, and denote the filtration generated by $g_t = \text{vec}(y_t, \phi_t)$ by \mathcal{G}_t^ϕ . The ϕ superscript is used to distinguish this filtration (perfect modal knowledge) from those that follow (noisy modal measurements or none). The Kalman filter generates two base-state moments: the conditional mean, $\hat{x}_t = E[x_t | \mathcal{G}_t^\phi]$, and the conditional error covariance, $P_{xx}(t) = E[\tilde{x}_t\tilde{x}_t' | \mathcal{G}_t^\phi]$. If x_0 is $\mathbf{N}(\hat{x}_0, P_{xx}(0))$, the estimate and the error are Gaussian. The Kalman filter generates the \mathcal{G}_t^ϕ -conditional distribution of the base-state (x_t is $\mathbf{N}(\hat{x}_t, P_{xx}(t))$) from which other statistical properties of the estimate can be derived. One form of the Kalman filter is [BW92, Figure 7.1]

Kalman filter: time-continuous state, time-continuous measurement

$$d\hat{x}_t = \sum_i A_i \hat{x}_t \phi_i dt + \gamma_x dv_t \quad (1.18)$$

subject to

$$\frac{d}{dt} P_{xx} = \sum_i (A_i P_{xx} + P_{xx} A_i' + R_\chi(i)) \phi_i - \gamma_x R_x \gamma_x' \quad (1.19)$$

The factor $\gamma_x = P_{xx} H' R_x^{-1}$ is the Kalman gain and

$$dv_t = dy_t - H(\chi\phi + \hat{x}_t) dt$$

is a \mathcal{G}_t^ϕ -innovation increment.

The Kalman filter is familiar to engineers, and comprehensive studies of its properties are available. There are features of the Kalman filter that warrant comment. The base-state estimate prescribes concurrent extrapolation ($\sum_i A_i \hat{x}_t \phi_i dt$) and correction ($\gamma_x dv_t$). The direction of extrapolation is determined by the current A -matrix (selected by ϕ_t). Correction is achieved by weighting the innovations

increment with γ_x . The Kalman gain, $\gamma_x = P_{xx} H' R_x^{-1}$, increases with observation quality (H) and decreases with sensor noise intensity (R_x). The gain also increases as the estimation uncertainty increases. (The base-state error covariance, P_{xx} , is sometimes called the *uncertainty matrix*.) An increase in P_{xx} makes the Kalman filter more data-driven whereas a decrease in P_{xx} makes the Kalman filter more model-driven.

Modal dependence enters the Kalman filter in a direct manner. The $\{\hat{x}_t\}$ and $\{P_{xx}\}$ dynamics change in concert with $\{\phi_t\}$. Although $\{\phi_t\}$ is a random process, $\{P_{xx}\}$ is random only because the coefficients in (1.19) are random: If the modal process were known a priori, P_{xx} could be precomputed. In any case, the error covariance is independent of the base-state observation $\{y_t\}$.

Equation (1.18) has an intuitively appealing form. Note that

$$E[dx_t | \mathcal{G}_t^\phi] = \sum_i A_i \hat{x}_t \phi_i dt$$

and that the innovation process is a \mathcal{G}_t^ϕ -martingale. The equation of evolution of the base-state estimate is

$$d\hat{x}_t = E[dx_t | \mathcal{G}_t^\phi] + d\mu_t,$$

where μ_t is a \mathcal{G}_t^ϕ -martingale [Kri84]. The increment in the mean is the mean of the increment plus a correction that is a martingale increment. For the system under study, all \mathcal{G}_t^ϕ -martingales are integrals with respect to the innovations process: All \mathcal{G}_t^ϕ -martingale increments are \mathcal{G}_t^ϕ -predictable multiples of the innovation increment. The last term above must be of the form of a gain multiplying dv_t [Ell82].

In applications in which the measurement is time discrete, the Kalman filter can be deduced formally from (1.18) and (1.19). Begin at time kT with the filter in state $(\hat{x}[k], P_{xx}[k])$. For the discrete-time case, it is convenient to distinguish the *pre-update* version of the base-state estimate from the *post-update* estimate. Denote the extrapolated state vector estimate at time $(k+1)T$ by $\hat{x}[k+1]^- = \hat{x}_{(k+1)T^-}$, and similarly denote the covariance by $P_{xx}[k+1]^- = P_{xx}((k+1)T^-)$. Integration of the measurement at time $(k+1)T$ gives rise to a correction to the pre-update estimate: $\Delta\hat{x}[k+1] = \hat{x}[k+1] - \hat{x}[k+1]^-$ and similarly $\Delta P_{xx}[k+1] = P_{xx}[k+1] - P_{xx}[k+1]^-$. The filter residual $r[k+1] = y[k+1] - H(\chi\phi[k+1] - \hat{x}[k+1]^-)$ is the innovations increment. The residual process is a white Gaussian process with covariance $R_{yy}[k]$ (with inverse $D_{yy}[k] = S_{yy}[k]' S_{yy}[k]$):

$$R_{yy} = E[r[k]r[k]'] | \mathcal{G}^\phi[k-1] = H P_{xx}[k]^- H' + R_x = D_{yy}[k]^{-1}.$$

The discrete form of the $\mathcal{G}^\phi[k]$ -filter is given in [BW92, Figure 5.9].

Kalman filter: time-continuous state; time-discrete measurements

Between observations:

$$\frac{d}{dt}\hat{x}_t = \sum_i A_i \hat{x}_t \phi_i, \quad (1.20)$$

$$\frac{d}{dt}P_{xx} = \sum_i (A_i P_{xx} + P_{xx} A_i' + R_x(i)) \phi_i. \quad (1.21)$$

At an observation:

$$\Delta \hat{x}[k+1] = \gamma_x r[k+1], \quad (1.22)$$

$$\Delta P_{xx}[k+1] = -\gamma_x R_{yy}[k+1] \gamma_x', \quad (1.23)$$

where $\gamma_x = P_{xx}[k+1]^{-1} H' D_{yy}[k+1]$. Equations (1.20)–(1.23) follow from (1.18) and (1.19) by making the replacements

- (i) in extrapolation, $R_x^{-1} \rightarrow 0$,
- (ii) in correction, $E[dv_t dv_t' | \mathcal{G}_t^\phi]/dt \rightarrow R_{yy}[k]$.

Like the continuous Kalman filter, its discrete sibling is a predictor–corrector, but the prediction and correction are not concurrent. Correction takes place at the observation times with a difference in the gain: For time-discrete measurements, $\gamma_x = P_{xx} H' R_{yy}^{-1}$; for time-continuous measurements, $\gamma_x = P_{xx} H' R_{yy}^{-1}/dt$, where $R_{yy}^{-1}/dt (= E[dv_t dv_t' | \mathcal{G}_t^\phi]/dt)$ is the intensity of the residual process. As in the time-continuous case, the observation gain increases with improved sensor quality. The time-discrete residual process is a white, Gaussian process: $r_k \in \mathbf{N}(0, R_{yy}[k])$. If the residual is scaled by $S_{yy}[k]$, a unit Gaussian white sequence is obtained: $S_{yy}[k] r[k] \in \mathbf{N}(0, I)$.

The Kalman filter is a complete solution to the estimation problem as posed, but most applications do not fall neatly within the modeling paradigm. Nonlinearities and discontinuities neglected in the model cause the performance of the Kalman filter to degrade. The influence of mismodeling is seen frequently in simulation exercises where the size of the estimation error can be contrasted with the computed error covariance. In an actual system, the true error is not known. But the residuals can be measured, and if $\{S_{yy}[k]r[k]\}$ is not a unit white noise process, the model of the plant and sensor may need to be refined. When $\{r[k]r[k]'\}$ consistently exceeds $R_{yy}[k]$, the filter is said to exhibit *excess error*; if P_{xx} is small, the filter residual may exceed the standard deviation of the noise in a single measurement.

1.2 A Tracking Example

To illustrate some of the issues that arise in hybrid estimation within the context of a concrete example, consider a tracking problem in which we wish to determine the position and velocity of an evasive aircraft moving in the X - Y plane. (The altitude is essentially constant.) Targets with limited thrust control maneuver by *jinking*: The turn rate tends to be nearly constant over intervals with sudden changes at unpredictable times. Suppose the aircraft is detected at a range of 36 km ($t = 0$) traveling at a speed of 300 m/s. The aircraft coasts (nearly constant velocity flight) for three seconds ($t \in [0, 3)$), makes a 7 g turn to the right for six seconds ($t \in [3, 9)$), coasts for two seconds ($t \in [9, 11)$), makes a 7 g turn to the left for five seconds ($t \in [11, 16)$), and then returns to coast. Increased drag during a turn causes the aircraft to slow to 60% of the speed that it had entering the turn with a 40% increase in speed when a turn transitions to coast. During an interval of constant turn rate (including coast), the speed is fairly constant.

A rudimentary motion model for the aircraft between changes in turn mode is

$$d \begin{bmatrix} X \\ Y \\ V_x \\ V_y \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\Phi \\ 0 & 0 & \Phi & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ V_x \\ V_y \end{bmatrix} dt + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} d \begin{bmatrix} w_x \\ w_y \end{bmatrix}. \quad (1.24)$$

In this tracking problem, there is no plant state reference point; that is, the plant dynamics are linear over \mathbb{R}^4 ($\chi = \mathbf{0}$), and the plant state is the base-state. Moreover, there is no endogenous actuating signal; that is, the tracker has no control over target motion ($u_t \equiv 0$). The base-state consists of $\{X, Y\}$, the position coordinates, and $\{V_x, V_y\}$, the associated velocities. The target is subject to two types of acceleration: (i) a wide band, omnidirectional acceleration described by the Brownian motion $\{w_x, w_y\}$ with intensity W and (ii) a maneuver acceleration represented by the turn rate process $\{\Phi_t\}$. The speed is slowly varying when the turn rate is constant, and so the omnidirectional acceleration is small: Let $C_i = \mathbf{e}_2 \otimes \mathbf{I}_2$ for all i and $W = \mathbf{I}_2$. The intensity of the acceleration is about 0.1 g.

The jinking behavior can be captured by partitioning the range of possible turn rates into three levels:

$$\Phi_t \in \{a_1 = 0.2r/s, \phi_t = \mathbf{e}_1; a_2 = 0r/s, \phi_t = \mathbf{e}_2; a_3 = -0.2r/s, \phi_t = \mathbf{e}_3\}.$$

The turn rate is given by $\Phi_t = a' \phi_t$. A change in motion mode causes a change in speed, but no rotation:

- At the beginning or end of a turn, the position process is continuous: $[X_{t+}, Y_{t+}] = [X_{t-}, Y_{t-}]$.
- At the beginning of a turn ($\Phi \mapsto a_1$ or $\Phi \mapsto a_3$) the target slows by 40%: $[V_x(t+), V_y(t+)] = 0.6[V_x(t-), V_y(t-)]$.

- At the end of a turn ($\Phi \mapsto \mathbf{e}_2$), the target speed increases by 40%, but not enough to attain the pre-turn velocity: $[V_x(t+), V_y(t+)] = 1.4 [V_x(t-), V_y(t-)]$

In this example, a turn-to-turn transition is not allowed. The intraregime model of the aircraft can be written as the stochastic differential equation:

$$dx_t = \sum_i A_i x_t \phi_i dt + dw_t, \quad (1.25)$$

where

$$A_i = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -a_i \\ 0 & 0 & a_i & 0 \end{bmatrix}.$$

At the origin of the coordinate system, $(0, 0)$, there is a sensor. A radar measures the position of the target every second with Gaussian errors of 40 m in range and 1.75 m in bearing (approximately 63 m at 35 km). This measurement is not linear in the coordinate system selected for the motion model: $y[k] = \mathbf{r}(x[k]) + n[k]$ instead of $y[k] = Hx[k] + n[k]$. The measurement relation can be linearized, not about the set point, but about the computed state estimate, \hat{x}_t , itself. A replacement for the measurement residual is $r[k] = y[k] - \mathbf{r}(\hat{x}[k]^-)$. The covariance update is computed using the χ -gradient of \mathbf{r} evaluated at $\hat{x}[k]$ in place of H in (1.22) and (1.23). This ancillary linearization is commonly done when the sensor nonlinearities are smooth and the estimation errors reasonably small, and it leads to an instance of the extended Kalman filter (EKF) [GA93, Table 5.4]. Similar output linearization will be performed in what follows wherever required without further comment.

The most rudimentary approach to the tracking problem would be to ignore the turn process and design an EKF based upon the specification of radar quality given above. Suppose $\hat{x}_0 = x_0$ and the initial covariance is taken to be diagonal with standard deviation in position (100 m) and velocity (20 m/s): The tracker is initialized at the true state of the aircraft and the initial uncertainty is larger than the single-measurement sensor error. The Brownian disturbances on the path are small: Set $W = 1$. Figure 1.1 shows a sample path of the nominal EKF as a *feather plot* referenced to a target path generated with $W = 0$. (A feather plot connects the estimates of location after a measurement to the true location. A point is shown every 0.1 s for clarity. The speed changes are not visible on the target path.) With the advantageous initialization, EKF($W=1$) begins well. The target model ignores turns and none occur at first.

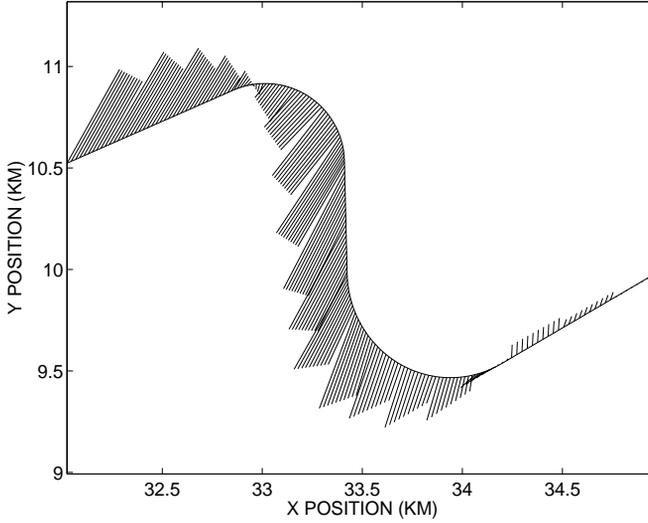


Figure 1.1. The path of a target with estimates from EKF(W=1).

When the target turns and slows, the performance of EKF(W=1) degrades. As the target turns to the right, EKF(W=1) fails to follow and extrapolates between radar measurements in the direction of the initial velocity. The position error is corrected in part when a radar measurement is received, but the gain is too small to bring \hat{x}_t back to x_t . The velocity correction is also far too small. With no direct velocity measurement, EKF(W=1) misinterprets $\{y[k]\}$. This creates tracking errors far in excess of the raw radar noise (about 60 m). It is not until the reverse turn has begun that EKF(W=1) identifies the velocity, but this is an artifact of the path. The error going into the final coast is quite large and the velocity estimate is abysmal.

The EKF, in contrast to less structured estimators (e.g., the $\alpha - \beta$ tracker), not only generates an estimate of the base-state, but it also provides an assessment of its own performance. The upper left submatrix $P_{xx}(1 : 2, 1 : 2)$ gives the error covariance in position. A one- σ region of target location is found by centering an ellipse determined by $P_{xx}(1 : 2, 1 : 2)$ about \hat{x}_t . In some adaptive estimators, the radar pulse shape (and the signal-to-noise ratio (SNR) of the sensor) and the tracking window are dependent on the size, shape, and location of this error ellipse.

Figure 1.2 displays the target path along with the one- σ error ellipses (shown every 0.2 s for clarity) centered at the location estimates. The ellipses are near circles in this case because of the symmetry in the measurement. On the first coast, when the dynamic hypotheses of the EKF match the motion, tracking uncertainty

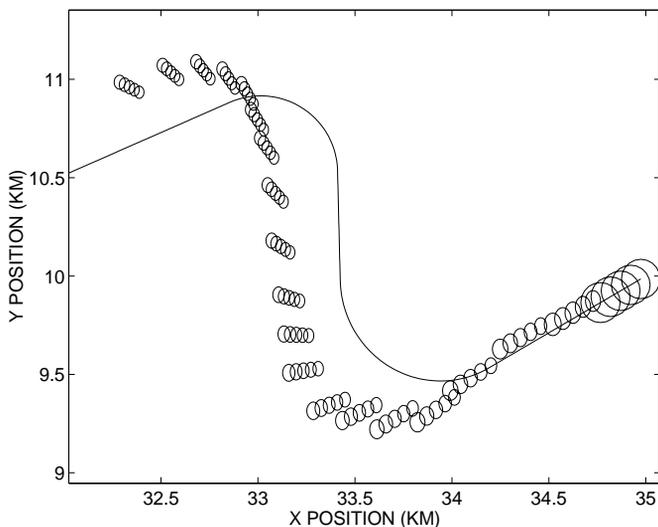


Figure 1.2. The path of a target with error ellipses generated by EKF($W=1$).

is reduced (the ellipses shrink) with each radar measurement. The true path lies within or next to the envelope of the one- σ error circles. When the target turns to the right and slows, EKF($W=1$) fails to react. Although EKF($W=1$) tacks away from the true path, the error ellipses evidence no sensitivity to the growth in the size of the measurement residuals. The residuals may exceed 10σ , a near impossibility if the errors were truly Gaussian. After completing the first turn, the target path lies several standard deviations away from $\{\hat{x}_t\}$ except when the target turns back into the estimate: The estimates are bad but the filter fails to acknowledge just how bad they are. A Gaussian density has thin tails, and the persistent presence of excess error as shown in Figure 1.1 is highly unlikely. The EKF's sanguine attitude would lead to loss-of-lock if the radar energy were focused in a three- σ window about $\{\hat{x}_t\}$.

With the approximations we have made in the design of this EKF, it is not surprising that it may need to be adjusted or tuned for this application. The nominal EKF is too sluggish to follow an agile target. In principle, any of the coefficients in EKF($W=1$) could be changed to make it more responsive. However, the aircraft dynamics and the observation equation are constrained by the physics of the path (e.g., the $\{A_i, i \in \mathbf{S}\}$) or the geometry of the sensors (e.g., H). Tuning in the EKF usually concentrates on the intensities of the exogenous disturbances: $W = E[dw_t dw_t'] / dt$ (the plant noise) or $R_x = E[n_k n_k']$ (the sensor noise). In fact, the focus is more commonly on the former because there are stronger empirical restrictions on the latter.

When W is increased to account for various modeling inaccuracies, *pseudonoise* is said to be added to the plant. For example, the motion model given in (1.24) is a low dimension representation of a very complicated object. An engineer could argue that the neglect of dynamic modes in the model causes the computed value of P_{xx} to be smaller than the true error covariance. For example, when the turns are ignored, the primary plant excitation is disregarded. If W is increased, the computed $\{P_{xx}\}$ is made larger. This increases the filter gain and the responsiveness of the EKF as well. While pseudonoise augmentation has proved useful in applications, the higher gains do magnify the noise. Additionally, additive white noise does not preserve the path geometry associated with the modes that are ignored, and this mismodeling may lead to performance that is far from optimal.

Let us try to improve the response of EKF($W=1$) by pseudonoise augmentation. To rationalize the level of pseudonoise, recognize that the target accelerations also include the turns. Set $W = 100$. The standard deviation of ΔV_x over one second is 10 m/s, which is equivalent to a 1 g constant acceleration. Over six seconds or so this would be roughly the *white-noise equivalent* of the turn process involving a 7 g turn over six seconds and intervening coasts. Of course, this equivalence is crude: The Brownian motion is continuous whereas the turn rate process is not; the Brownian motion acts throughout the tracking interval, whereas the turn rate changes at isolated times; the Brownian motion is omnidirectional, whereas the turn places specific geometric constraints on the target path.

Figure 1.3 shows the feather plot of a sample of an EKF with this pseudonoise augmentation. The effect of pseudonoise is beneficial for the most part. After the first

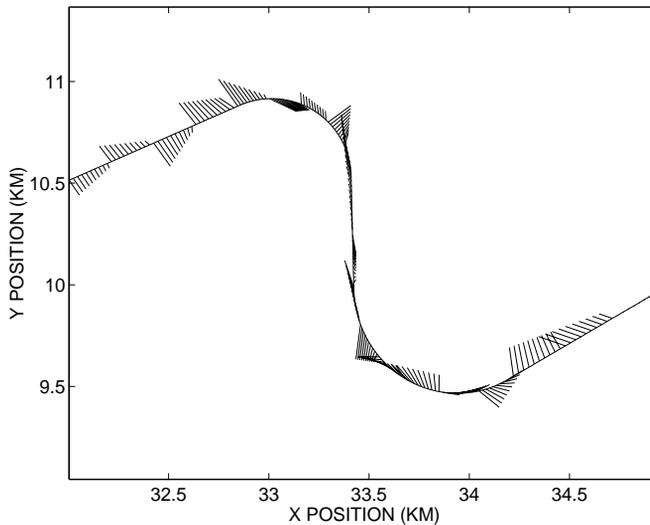


Figure 1.3. The feather plot for EKF($W=100$).