A NEW GENERALIZED CONSIDER COVARIANCE ANALYSIS AND OTHER TOOLS FOR CHALLENGING ESTIMATION SCENARIOS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Joanna C. Hinks

May 2012

A NEW GENERALIZED CONSIDER COVARIANCE ANALYSIS AND OTHER TOOLS FOR CHALLENGING ESTIMATION SCENARIOS

Joanna C. Hinks, Ph.D.

Cornell University 2012

Several strategies are presented for dealing with various situations where traditional estimation techniques may fail. These methods complement existing solutions by improving estimation robustness or by providing analysis of the estimator performance. Applications include spacecraft attitude and orbit determination problems. The first contribution determines attitude and angular velocity for a spinning spacecraft using only time-spaced unit vector measurements. Several algorithms are developed that are suitable for initialization of an extended Kalman filter so as to prevent filter divergence due to high nonlinearity. A second focus is on orbit determination for multiple satellites encountering highly uncertain environmental perturbations to their orbits and signals. A filter that incorporates estimation of upper atmospheric and ionospheric parameters along with the satellite orbits is shown to be observable. Consider covariance analysis demonstrates the improvement in the orbit solution that results from this additional state estimation. Lastly, the technique of Consider covariance analysis is extended to analyze square-root information filters and smoothers with a wide variety of modeling errors. The new analysis is the most general Consider analysis for square-root information filters, and the only generalized Consider analysis for Rauch-Tung-Striebel square-root information smoothers. It can study filters and smoothers with incorrect noise, incorrect initialization, unmodeled biases or dynamics, erroneous system matrices, and other error classes.

BIOGRAPHICAL SKETCH

Joanna Hinks was born in Toronto, Canada, and lived in a number of locations during her childhood, including Florida, Ohio, and Wisconsin. She graduated as valedictorian of her high school class in 2002. In 2006 she received her B.S. degree in Mechanical Engineering from Cedarville University, with a cumulative undergraduate GPA of 4.0. Joanna received an M.S. degree in Mechanical Engineering from Cornell University in 2009, and completed a Ph.D. in Mechanical Engineering in 2012, also at Cornell University.

Joanna's research interests lie in the areas of estimation and filtering, dynamic modeling, inverse problems, and GNSS algorithms and technology. Specific applications include spacecraft attitude and orbit determination problems and modeling of the satellite environment.

O God, my help in courses past
My hope for jobs to come
My shelter from the stormy blast
And my eternal home.

Like Abraham I hear Your call
But know not where You guide.
I doubt and stray, and often fall
Yet You remain beside.

Let not my path be effortless
Lest I forget my need.
Let me not trust my own success
Nor worldly praises heed.

Christ, lead me always where You choose.
Each day my trust renew.
No joy or trial would I refuse
But hide my life in You.

ACKNOWLEDGMENTS

I am grateful to the National Science Foundation for three years of financial support in the form of an NSF Graduate Research Fellowship.

The orbit determination work presented in Chapter 3 was supported in part by Boeing Defense, Space & Security through Contract #8CC0789. Peter M. Fyfe is the Technical Monitor.

My advisor, Dr. Mark Psiaki, provided guidance and support (and when necessary, reproof) throughout every step of my graduate career, and made it more of an adventure than an ordeal. I have benefited greatly from both his factual knowledge and his example as a researcher and a teacher. My research has also been furthered by many valuable discussions with my other committee members, my officemates, and the past and current members of the Cornell GNSS Lab.

I would have been in dire straits many times without the assistance of Marcia Sawyer, who never failed to answer my questions about paperwork patiently. I think I had to ask the difference between "academic program" and "degree program" at least five times.

To all my teachers over the past 23 years, thank you for teaching me how to learn. And to my classmates and fellow-sufferers, thank you for all the experiences we shared - I learned more from you than from any textbook.

I owe my happiness and sanity to all those who helped to make Ithaca a home for the past six years. In particular, I want to thank the members of the Graduate Christian Fellowship and New Life Presbyterian Church for being a family when I couldn't visit my own very often.

Finally, to all my family members and especially Mom, my first teacher, and Dad, the first Dr. Hinks, thank you again for everything. I don't say it often

enough, and I know I haven't been easy. Thank you for being there for me in both the good times and the bad. I would never have gotten to this point without all your prayers, love, support, encouragement, rebuke, teaching, discipline, advice, sympathy, wisdom, humor,...and most importantly, the examples you set in daily life.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgments . . . v
Table of Contents . . . vii
List of Tables . . . ix
List of Figures . . . x

1 Introduction **1**

2 Solution Strategies for an Extension of Wahba's Problem to a Spinning Spacecraft **7**
2.1 Introduction . . . 7
2.2 Problem Formulation and Background . . . 12
2.2.1 Problem Formulation . . . 12
2.2.2 Background from Dynamics . . . 14
2.2.3 Background from Attitude Determination . . . 18
2.3 Solutions to the Restricted Problem . . . 20
2.3.1 Aliasing . . . 20
2.3.2 Solution Paradigm . . . 22
2.3.3 Equation Sets . . . 25
2.3.4 Initialization Procedures . . . 36
2.4 Results . . . 39
2.4.1 Baseline Cases . . . 42
2.4.2 Measurement Error Effects . . . 44
2.4.3 Further Discussion of Results . . . 47
2.4.4 A Practical Implementation Strategy . . . 48
2.5 Conclusions . . . 49
2.6 Appendix: Equivalence of Wahba Performance Metric and Derived Cost Function . . . 50

3 Estimation of Atmospheric and Ionospheric Effects in Multi-Satellite Orbit Determination Using Crosslinks **52**
3.1 Introduction . . . 53
3.2 Atmospheric and Ionospheric Density Parameterizations . . . 57
3.3 System Dynamics and Measurements . . . 62
3.3.1 State Vector and Dynamics . . . 63
3.3.2 Measurement Models . . . 65
3.3.3 Linearization . . . 67
3.4 Observability and Consider Analysis Theory . . . 69
3.4.1 Linearized Observability Analysis . . . 69
3.4.2 Consider Covariance Analysis . . . 73
3.5 Results . . . 76

3.5.1 "Truth" Model . . . 77
3.5.2 Observability Results . . . 79
3.5.3 Consider Analysis Results . . . 82
3.6 Conclusions . . . 84

4 A Multipurpose Square-Root Consider Covariance Analysis for Linear Filters 86
4.1 Introduction . . . 86
4.2 Introductory Example . . . 93
4.3 Consider System Formulation . . . 95
4.4 Consider Analysis Algorithms . . . 102
4.4.1 Initialization and First Steps . . . 102
4.4.2 Main Algorithm . . . 108
4.4.3 Consider Covariance Calculations . . . 114
4.5 Examples . . . 116
4.5.1 Example: Incorrectly Modeled Noise . . . 117
4.5.2 Example: Incorrect System Matrices . . . 124
4.6 Conclusions . . . 132
4.7 Appendix: QR and LQ Factorization Calculations . . . 132

5 A Multipurpose Square-Root Consider Covariance Analysis for Linear Smoothers 134
5.1 Introduction . . . 134
5.2 Background: Consider System Model & Forward-Pass Filter Analysis . . . 139
5.2.1 Consider Form . . . 140
5.2.2 Needed Outputs from Forward-Pass Consider Filter Analysis . . . 143
5.3 Consider Smoother Analysis Algorithms . . . 146
5.3.1 Initial Definitions and Setup . . . 147
5.3.2 Backwards Propagation . . . 149
5.3.3 Smoothed Estimation Error Covariances . . . 154
5.4 Examples . . . 157
5.4.1 Filter/Smoother Assumed Model Description . . . 158
5.4.2 Example: Incorrect Initialization . . . 159
5.4.3 Example: Correlated Noise . . . 165
5.4.4 Example: Unestimated Disturbances . . . 169
5.5 Conclusions . . . 174

6 Summary & Conclusions 175

References 180

LIST OF TABLES

2.1 Characteristics of "truth" model scenarios. 40

5.1 Consider system model summary. 140

LIST OF FIGURES

2.1 Attitude variables; major-axis spinner. 42
2.2 Attitude variables; minor-axis spinner. 42
2.3 Angular-momentum variables; major-axis spinner. 43
2.4 Angular-momentum variables; minor-axis spinner. 43
2.5 Attitude error standard deviations. 46
2.6 Angular veloicty error standard deviations. 46

3.1 Least-squares fit of ionospheric parameters. 62
3.2 Estimation error standard deviations for satellite position. 80
3.3 Atmospheric density normalized standard deviations. 82
3.4 Ionospheric density normalized standard deviations. 82
3.5 Consider standard deviations for satellite position 83

4.1 Filter error standard deviations; incorrect noise. 123
4.2 RMS filter errors; incorrect system matrices. 130

5.1 RMS smoother errors; incorrect initialization. 164
5.2 Filter error standard deviations; correlated noise. 168
5.3 Smoother error standard deviations; correlated noise. 168
5.4 Filter error standard deviations; unestimated disturbances. . . . 173
5.5 Smoother error standard deviations; unestimated disturbances. . 173

CHAPTER 1

INTRODUCTION

A common task in many areas of engineering is that of estimation: combining information from mathematical models and noisy measurements to obtain the best possible estimate of some quantity of interest. For dynamic systems in state-space form, estimation is most frequently accomplished by a Kalman filter (KF) [1] or some closely-related method, such as an unscented Kalman filter, square-root information filter, or particle filter. These methods propagate past estimates using a dynamics model and combine the result with new information from measurements with a given measurement model. Weighting of old and new information is accomplished by using the modeled statistics associated with the dynamic process noise and the measurement noise. In addition to computing estimates, KF-based techniques compute the uncertainty of those estimates in the form of estimation error covariance.

In order to produce optimal estimates, KFs require certain assumptions to hold. First, the system of interest must be represented by linear state-space models. If nonlinearities exist, estimation can often be accomplished by an extended Kalman filter (EKF), which performs a linearization around some reference trajectory. EKFs are prone to divergence, however, if the state estimates are not close enough to the true states for the linearizations to hold. Both linear Kalman filters and EKFs also require that the available measurements provide sufficient information to uniquely determine a state estimate; this is related to the concept of *observability*. Finally, the mathematical models for the system of interest must be accurate. If the assumed dynamics or measurement models do not capture the true system behavior in some respect, the resulting estimates and estima-

tion error covariances will usually be unreliable. The remainder of this dissertation presents some tools for dealing with estimation scenarios where these basic Kalman filter assumptions have been violated.

Chapter 2 addresses a highly-nonlinear problem in the area of satellite attitude determination. Specifically, it examines the case of a rigid, axially-symmetric spacecraft undergoing torque-free motion. The only available measurements are unit-vector measurements at a sequence of distinct times. This problem is formulated as an extension of the classic Wahba's problem [2]. Although an EKF can work in this scenario, it requires a very accurate initialization for both attitude and angular velocity in order to avoid divergence. Most previous work in this area ignores the issue of initialization, assumes multiple measurements are available at each time to fix the instantaneous attitude, or assumes that some form of angular rate information is available. One contribution of Chapter 2 is a partially analytical, partially numerical solution paradigm for this problem. Its algorithms exploit the known form of the spacecraft's attitude dynamics to write a set of coupled nonlinear equations. A series of algebraic manipulations reduces the solution space to a lower dimension with fewer equations and unknowns, and the remaining set of equations can be solved by numerical techniques. In addition to this overarching paradigm, the chapter provides several specific algebraic strategies by which the equations can be reduced to a simpler form. These strategies are demonstrated by Monte Carlo simulations, and practical implementation issues are discussed.

In Chapter 3, two issues are addressed: system observability and high model uncertainty. This chapter's system of interest is a set of satellites occupying a single low-Earth, polar orbital plane. Ranging measurements are available from

the satellites to ground stations, and between adjacent satellites in the plane. The orbital dynamics are perturbed by atmospheric drag, and the crosslink ranging signals between satellites are delayed by the ionosphere. Both of these effects are poorly understood, and the associated physical processes contain high levels of uncertainty. Chapter 3 contributes empirical dynamic models for the upper atmosphere and ionosphere. These models have properties that are more suitable for the estimation scenario than most physics-based models, yet they provide enough degrees of freedom to capture the physical effects on the satellite orbits and signals. Chapter 3 also examines the system observability when the dynamically-varying parameters related to the atmosphere and ionosphere are estimated simultaneously with the satellite orbits. As the assumed dynamic models for the atmosphere and ionosphere are so uncertain, one might question whether estimating these extra parameters is really beneficial. A Consider covariance analysis thus studies the effects on orbit determination performance that result from estimating the environmental parameters rather than assuming standard atmospheric and ionospheric models. The observability and Consider covariance analyses illustrate some of the challenges that arise in complicated systems with poorly-understood mathematical models. Chapter 3's Consider analysis is adapted from the square-root information formulation of Ref. [3], but there are limits to what this formulation can accomplish. Therefore, this chapter provides the motivation for the development of the more general Consider covariance analysis tools in the following chapters.

Chapters 4 and 5 together develop a powerful tool for analysis of square-root information filters and smoothers with model errors. Consider covariance analysis is a well-established technique that has applications in both design and analysis. It computes the true estimation error covariance of a filter or smoother

with incorrect models of the system dynamics or measurements, given a true system model. Types of errors that are typically investigated include incorrect noise statistics or incorrect *a priori* covariance, incorrect system matrices, and unestimated biases or dynamic disturbance states. While some analysis algorithms are very narrow in scope and focus on a single type of error, other approaches are very general and can analyze systems with any combination of these model errors.

There are many ways such an analysis can be applied, depending on whether the model error was intentional or not. Intentional mismatch between a filter and a true system typically occurs in the context of simplification. By designing a filter with a lower-dimensional state or decoupled dynamics, the computational burden of the filter can be greatly reduced. Oversimplification, however, can lead to an unacceptable loss of accuracy. Consider covariance analysis aids the designer in finding the right balance between filter performance and complexity. In many other situations, however, the behavior of the true system is poorly understood and the estimator mismatch is unintentional. In these cases it is often possible to identify the likely types of errors, or to assume bounds on parameter uncertainty. A number of Consider covariance analysis applications are possible in such a scenario. For example, Consider analysis can be part of procedure that seeks a filter or smoother that is insensitive to variations in a particular set of uncertain parameters. Similarly, a given estimator design may be evaluated in combination with many different hypothetical "true" system models to determine the possible range of estimation performance. One can even investigate whether a filter is preferable to a smoother in a given scenario, or vice versa.

The Consider covariance analysis of Chapters 4 and 5 is capable of investigating many different classes of modeling errors, and its square-root formulation is compact and efficient. Its flexibility arises from the definition of a special "Consider form" of the system model equations. Once a system with model errors has been written in this proposed standard form, the same set of analysis algorithms can be directly applied, regardless of what types of errors are present. The core algorithms themselves resemble the existing standard algorithms for the square-root information filter and smoother, with dynamic propagation and measurement update steps. In addition to computing the filter or smoother's assumed square-root information matrix, the Consider algorithms also compute terms that specify how the model errors affect the estimation uncertainty.

The contributions of the new Consider covariance analysis are as follows: First, the combined filter and smoother analysis is the first general Consider covariance analysis in square-root information form. The filter analysis of Chapter 4 is the most general Consider covariance analysis for a square-root information filter, although there are several previous square-root algorithms that handle a subset of its error classes. The smoother analysis of Chapter 5 is the first general Consider covariance analysis for a square-root information smoother. Only one other generalized algorithm exists for Consider covariance analysis of a discrete RTS smoother, and it is in the covariance domain [4, 5]. The present derivation is believed to be more thorough, and the final form of the equations is more compact. This derivation can also be more intuitive once one understands its Consider analysis strategy. Second, the analysis derivations are followed by illustrative examples. These examples play two roles. They demonstrate how a number of major error classes can be placed into the special "Consider form" so that the core algorithms can be applied. Application of previous Consider algo-

rithms to certain kinds of errors is possible but non-obvious for inexperienced practitioners. The chosen examples also highlight some non-intuitive behaviors of mismatched estimators. These include situations where the true estimation error covariance is lower than that reported by a pessimistic filter and cases where smoother estimates may be less accurate than filter estimates. Consider covariance analysis is an efficient tool for revealing such unexpected characteristics.

CHAPTER 2

SOLUTION STRATEGIES FOR AN EXTENSION OF WAHBA'S PROBLEM TO A SPINNING SPACECRAFT

Several methods have been devised to estimate the optimal attitude quaternion and angular velocity for an axially-symmetric spacecraft given only a time series of vector attitude measurements and a rigid-body model of the spacecraft dynamics. These methods solve a generalization of Wahba's problem that is applicable to a low-complexity spinning spacecraft such as a sounding rocket. They have the advantages of enabling rate estimation without gyros and avoiding the divergence issues of an Extended Kalman Filter. Measurements at different times are combined via the closed-form quaternion state transition matrix that makes use of Euler's equations for a rigid body. The problem takes the form of six or more nonlinear equations in six or seven unknowns, and these equations can be partially solved analytically to facilitate numerical solution of the remainder. Representative simulation results illustrate both the merits of these solution strategies and some of the practical issues that arise during implementation.

2.1 Introduction

Many spacecraft mission objectives, such as positioning of a solar panel or telescope, are strongly dependent on the spacecraft attitude relative to some reference frame. Consequently, mission success often depends on efficient and reliable methods of estimating attitude from available measurements. Attitude estimation is complicated by dynamic motion, which necessitates determination

of a history of spacecraft attitudes. Even when accurate dynamics and measurement models are available, they are usually highly nonlinear. Sometimes these nonlinearities can lead to failure of traditional estimation techniques.

Historically, attitude estimation methods can be separated into several broad categories, each with its own strengths and weaknesses. For an overview of the key developments, see Ref. [6]. First, there is the family of techniques formulated to solve what is known as "Wahba's problem" [2]: Given two or more vector measurements in body coordinates at a particular instant in time, corresponding to known reference frame vectors, find the orthonormal rotation matrix (or equivalent unit quaternion) that minimizes the measurement residuals in a weighted-least-squares sense by minimizing the cost function:

$$J\left(\boldsymbol{q}\right) = \frac{1}{2}\sum_{i=0}^{m-1}\frac{1}{\sigma_i^2}\left[\boldsymbol{b}_i - A\left(\boldsymbol{q}\right)\boldsymbol{r}_i\right]^{\mathrm{T}}\left[\boldsymbol{b}_i - A\left(\boldsymbol{q}\right)\boldsymbol{r}_i\right] \tag{2.1}$$

In this cost function, $\boldsymbol{q}$ is the unit-normalized quaternion that parameterizes the attitude, $A\left(\boldsymbol{q}\right)$ is the corresponding direction cosine matrix, $\boldsymbol{b}_i$ is the i^{th} unit vector measurement in body axes, and $\boldsymbol{r}_i$ is the corresponding reference frame unit vector. The measurements are weighted by the inverse square of the per-axis measurement error standard deviation σ_i. This non-standard weighting makes J a negative-log-likelihood cost. Standard solutions to Wahba's problem, such as Refs. [7–9], avoid errors due to linearization or any need for iteration, and they yield the global optimal solution. A major weakness of Wahba's problem is that it assumes all measurements are available at a single instant in time, or that the spacecraft attitude is stationary. Thus, it does not adapt easily to a spacecraft undergoing dynamic motion when the measurements are separated by time.

To address dynamic spacecraft scenarios, the second class of estimation

methods applies standard or adapted filtering techniques, such as various forms of the Extended Kalman Filter (EKF). Some examples include Refs. [10–12]. Filters need an initial guess of attitude (and often other quantities, such as angular velocity) and employ models of the dynamics and measurements to form an updated estimate. Filtering methods are powerful and capable of processing an arbitrary number of measurements either recursively or in a batch. Typically, however, they require linearization of the nonlinear models, and thus have a tendency to fail by divergence if the initial estimate is far from the truth. Nonlinear observers, though less common in attitude determination, also address dynamic scenarios [6].

A third family of attitude estimation methods seeks to maintain the optimal, globally convergent solutions to Wahba's problem within a filtering framework that incorporates attitude dynamics. Previous work in this category, however, has been restricted to systems with certain types of available measurements. Examples include Refs. [13–17]. Some approaches require two or more vector measurements at a time, such that a classical version of Wahba's problem can be solved at each step and these individual solutions combined in some fashion. Another version obtains rate information from the derivatives of the vector measurements in addition to the measurements themselves. Other methods assume the availability of rate gyros in order to measure non-zero angular velocity rather than estimating it. Such techniques are not applicable to spacecraft missions that eliminate rate gyros in order to reduce costs, weight, power, or some combination thereof.

A previous paper [18] proposes two generalized versions of Wahba's problem. They both seek to estimate the optimal initial quaternion and the space-

craft angular velocity given only a time series of vector attitude measurements. Reference [18] also presents a closed-form solution for a restricted version of one of these problems, and Ref. [19] solves a somewhat more general restricted version. Solutions to these proposed problems would enable both attitude and angular velocity estimation for a spinning spacecraft having only one or more vector sensors, such as narrow-field-of-view star cameras or magnetometers. The divergence issues associated with an EKF would be avoided. A sequence of batch filters based on solutions to these problems could be used for non-divergent attitude and rate estimation, or one of these batch solutions could be used to initialize an EKF with a good state estimate that would help it to avoid divergence. This chapter's algorithms are computationally complex; therefore, they may be more suited to ground-based post-processing than to real-time flight software.

The present chapter addresses a restricted form of the second problem proposed in Ref. [18]. In this restricted form, an axially-symmetric, rigid spacecraft with negligible nutation damping undergoes torque-free motion. The inertia matrix is known, and is assumed diagonal without loss of generality. Single vector measurements are available at distinct times, and the initial attitude quaternion and initial angular velocity vector are sought. Additionally, the number of measurements is restricted to three, which causes the problem to be fully determined but not overdetermined. This problem is useful because (nearly) axially-symmetric, spinning spacecraft are common for simple applications, and the inertia matrix can usually be well characterized in advance. Whereas the three-measurement case has limited applicability on its own, it is much more tractable than the case of an arbitrary number of measurements, and may even suggest some avenues towards the solution of this more difficult case. Furthermore, the

three-measurement solution could be used to initialize an EKF to avoid divergence problems.

The present chapter makes four original contributions beyond the initial work of Ref. [18]. First, it formulates a general global solution paradigm in which attitude and angular velocity are found by numerical solution of analytically reduced sets of nonlinear equations. Second, it derives several specific but representative sets of equations that can be used within the solution procedure. Third, it discusses numerical results obtained for the chosen sets of equations, as tested with simulated "truth" model data. Finally, it draws some conclusions relevant to general solution strategies for the restricted problem.

This chapter is organized as follows: Section II formulates the generalized Wahba's problems of Ref. [18], with emphasis on the restricted form of the second generalized problem, and it presents useful background information from the attitude determination and dynamics fields. In Section III, the general solution procedure is outlined, and several candidate sets of nonlinear equations are constructed such that their solutions solve the restricted problem. Section IV evaluates the results of numerical solutions to the nonlinear equations sets, and Section V draws some conclusions about the chapter.

2.2 Problem Formulation and Background

2.2.1 Problem Formulation

A well-known solution to Wahba's problem is the q-method [7, 20]. The q-method rearranges the original cost function (2.1) to produce an equivalent optimal estimation problem:

$$\text{find :} \qquad \boldsymbol{q} \tag{2.2a}$$

$$\text{to maximize :} \qquad p\left(\boldsymbol{q}\right) = \boldsymbol{q}^{\mathrm{T}} K \boldsymbol{q} \tag{2.2b}$$

$$\text{subject to :} \qquad \boldsymbol{q}^{\mathrm{T}} \boldsymbol{q} = 1 \tag{2.2c}$$

For the i^{th} measurement, a symmetric 4×4 matrix K_i is formed from the $\boldsymbol{b}_i$ and $\boldsymbol{r}_i$ vectors [20, p. 428]. Then the symmetric 4×4 matrix K is formed from the individual K_i matrices as follows:

$$K = \sum_{i=0}^{m-1} K_i \tag{2.3}$$

After adjoining the unit-normalization constraint of Eq. (2.2c) using the Lagrange multiplier λ and deriving the first-order optimality necessary condition, the q-method solution reduces to an eigenvalue problem. The optimal $\boldsymbol{q}$ is the eigenvector of K corresponding to its maximum eigenvalue, λ_{opt}: $K\boldsymbol{q}_{opt} = \lambda_{opt}\boldsymbol{q}_{opt}$.

The generalized Wahba's problems posed in Ref. [18] extend this formulation to the case of a spinning spacecraft with only a time series of vector measurements available. These problems also restrict the reference frame to be inertial. The first form considers a spacecraft spinning around its known major principal inertia axis with enough nutation damping that the spin direction and

rate can be assumed constant. In this problem, there are four unknowns: three to parameterize the initial attitude quaternion, and the unknown constant spin rate ω.

The second generalized problem, which this chapter attempts to solve in a restricted form, builds on the first problem. Instead of assuming a constant, known spin axis, the second problem allows all three components of the angular velocity vector $\boldsymbol{\omega}(t)$ to be unknown, but the moment-of-inertia matrix is assumed to be known, and the resulting nutational motion is assumed to be torque-free. If the initial attitude quaternion $\boldsymbol{q}_0$ and angular velocity $\boldsymbol{\omega}_0$ were known, then one could simultaneously integrate Euler's equation and the kinematic quaternion transition matrix differential equation forward in time in order to obtain the attitude and angular velocity at any time. Thus, six scalar unknowns fully parameterize the second problem: three associated with the initial quaternion and three associated with the initial angular velocity. A minimum of three vector measurements would be necessary to solve this problem. If the quaternion state transition matrix can be determined as a function of $\boldsymbol{\omega}_0$, then the problem solution can be written as a q-method maximization:

$$\text{find :} \quad \boldsymbol{q}_0 \text{ and } \boldsymbol{\omega}_0 \tag{2.4a}$$

$$\text{to maximize :} \quad \bar{p}\left(\boldsymbol{q}_0, \boldsymbol{\omega}_0\right) = \boldsymbol{q}_0^{\mathrm{T}} \left[\sum_{i=0}^{m-1} \Phi^{\mathrm{T}}(t_i, t_0; \boldsymbol{\omega}_0)\, K_i \Phi\left(t_i, t_0; \boldsymbol{\omega}_0\right) \right] \boldsymbol{q}_0 \tag{2.4b}$$

$$\text{subject to :} \quad \boldsymbol{q}_0^{\mathrm{T}} \boldsymbol{q}_0 = 1 \tag{2.4c}$$

where t_0 is the initial time at which $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$ apply, t_i is the time of the i^{th} measurement $\boldsymbol{b}_i$ and the i^{th} reference vector $\boldsymbol{r}_i$, and $\Phi\left(t_i, t_0; \boldsymbol{\omega}_0\right)$ is the quaternion transition matrix from time t_0 to time t_i, during which attitude motions occur as dictated by the value of $\boldsymbol{\omega}_0$. From the form of this maximization, it is clear that given $\boldsymbol{\omega}_0$ and the equations for the quaternion state transition matrix

$\Phi(t_i, t_0; \boldsymbol{\omega}_0)$, the K matrix could be computed and the optimal quaternion found by solving the corresponding q-method eigenvalue problem.

The problem form in Eqs. (2.4) suggests an inner-outer optimization procedure as discussed in Ref. [18]. An inner optimization would be performed on $\boldsymbol{q}_0$ to determine the optimal $\boldsymbol{q}_0$ as a function of $\boldsymbol{\omega}_0$, and an outer optimization would seek $\boldsymbol{\omega}_0$ to maximize the largest eigenvalue of the K matrix. This outer optimization is made challenging by the fact that the maximum eigenvalue of a symmetric matrix depends on the matrix elements in a complicated way. Gradient-based search methods are likely to fail because the problem contains an unconstrained three-dimensional search space of possible angular velocities in which the existence of multiple local maxima is possible, as are $\bar{p}$ gradient discontinuities where the largest and second-largest eigenvalues exchange places (as in Fig. 1 of Ref. [18]).

2.2.2 Background from Dynamics

The restricted form of the second problem of Ref. [18] addressed in this chapter considers an axially-symmetric rigid spacecraft undergoing torque-free motion with known inertia matrix I_{sc} given by

$$I_{sc} = \begin{bmatrix} I_T & 0 & 0 \\ 0 & I_T & 0 \\ 0 & 0 & I_S \end{bmatrix} \tag{2.5}$$

The nominal spin axis is the z-axis, and the two transverse axes have identical moments of inertia, consistent with the assumption of axial symmetry. Euler's equation of motion for this system is given in many texts [20, 21], and will not be repeated in full here, but several specific results are important for this problem.

First, the closed-form solution for angular velocity in body axes can be written as

$$\boldsymbol{\omega}(t) = \begin{bmatrix} \omega_x(t) \\ \omega_y(t) \\ \omega_z(t) \end{bmatrix} = \begin{bmatrix} \omega_T \cos\left[\omega_{nb}\left(t - t_0\right) + \phi_0\right] \\ \omega_T \sin\left[\omega_{nb}\left(t - t_0\right) + \phi_0\right] \\ \omega_{z0} \end{bmatrix} \tag{2.6}$$

where ω_T is the transverse component of angular velocity, which is constant in magnitude but circles around the z-axis, and ω_{z0} is the constant component along the z-axis. The rate at which ω_T circles the z-axis is the body-axes nutation rate, ω_{nb}, and the phase ϕ_0 is included to account for the location of the spin vector in the x-y plane at time t_0. In the inertial frame, the angular velocity nutates around the angular momentum vector $\boldsymbol{h}_{in}$ with the inertial nutation rate ω_{ns}. The angular momentum vector in spacecraft and inertial coordinates is related to angular velocity by the straightforward transformations

$$\boldsymbol{h}_{sc0} = I_{sc}\boldsymbol{\omega}_0 \;\text{ and }\; \boldsymbol{h}_{in} = A^{\mathrm{T}}(\boldsymbol{q}_0)\, I_{sc}\boldsymbol{\omega}_0 \tag{2.7}$$

Note that the diagonal structure of I_{sc} makes Eq. (2.7) easily invertible. The body-axes and inertial nutation rates are given by the expressions

$$\omega_{nb} = \left[1 - (I_S/I_T)\right]\omega_{z0} \tag{2.8}$$

$$\omega_{ns} = \frac{\|\boldsymbol{h}_{in}\|}{I_T} = \frac{\|\boldsymbol{h}_{sc0}\|}{I_T} = \frac{\|I_{sc}\boldsymbol{\omega}_0\|}{I_T} = \frac{\sqrt{I_T^2\omega_T^2 + I_S^2\omega_{z0}^2}}{I_T} \tag{2.9}$$

A closed-form attitude solution is also available for a spacecraft with axial symmetry. Throughout this chapter, attitude will be parameterized by quaternions; for more information on quaternion math see Refs. [20, 21]. The quaternion state transition matrix for this restricted case can be obtained as a function of $\boldsymbol{\omega}_0$, and it contains only simple trigonometric functions. Equivalently, one can find the quaternion for the rotation that transitions the spacecraft from its attitude at one time to its new attitude at another time. To find the transitional

rotation in the restricted case, it is useful to view the problem dynamics as a body cone rolling on a space cone [20, 21]. The angular velocity rotates around the body z-axis in body coordinates, and the body z-axis rotates around the angular momentum vector as viewed in inertial space in a way that keeps the body z-axis, the angular momentum vector, and the angular velocity vector all in the same instantaneous plane. The solution can be written in terms of the angular velocity at any time t. In the following development it is written in terms of the initial angular velocity or angular momentum, and the two quantities are treated as interchangeable in accordance with Eq. (2.7). The solution is given in Refs. [18, 20]. It takes the form:

$$\Phi\left(t,t_0;\boldsymbol{\omega}_0\right)=\begin{bmatrix} q_{\Phi4}\left(t-t_0,\boldsymbol{\omega}_0\right) & q_{\Phi3}\left(t-t_0,\boldsymbol{\omega}_0\right) & -q_{\Phi2}\left(t-t_0,\boldsymbol{\omega}_0\right) & q_{\Phi1}\left(t-t_0,\boldsymbol{\omega}_0\right) \\ -q_{\Phi3}\left(t-t_0,\boldsymbol{\omega}_0\right) & q_{\Phi4}\left(t-t_0,\boldsymbol{\omega}_0\right) & q_{\Phi1}\left(t-t_0,\boldsymbol{\omega}_0\right) & q_{\Phi2}\left(t-t_0,\boldsymbol{\omega}_0\right) \\ q_{\Phi2}\left(t-t_0,\boldsymbol{\omega}_0\right) & -q_{\Phi1}\left(t-t_0,\boldsymbol{\omega}_0\right) & q_{\Phi4}\left(t-t_0,\boldsymbol{\omega}_0\right) & q_{\Phi3}\left(t-t_0,\boldsymbol{\omega}_0\right) \\ -q_{\Phi1}\left(t-t_0,\boldsymbol{\omega}_0\right) & -q_{\Phi2}\left(t-t_0,\boldsymbol{\omega}_0\right) & -q_{\Phi3}\left(t-t_0,\boldsymbol{\omega}_0\right) & q_{\Phi4}\left(t-t_0,\boldsymbol{\omega}_0\right) \end{bmatrix} \tag{2.10}$$

where

$$\begin{bmatrix} q_{\Phi1}\left(t-t_0,\boldsymbol{\omega}_0\right) \\ q_{\Phi2}\left(t-t_0,\boldsymbol{\omega}_0\right) \\ q_{\Phi3}\left(t-t_0,\boldsymbol{\omega}_0\right) \\ q_{\Phi4}\left(t-t_0,\boldsymbol{\omega}_0\right) \end{bmatrix} = \boldsymbol{q}_\Phi\left(t-t_0,\boldsymbol{\omega}_0\right)=\boldsymbol{q}_s\left(t-t_0,\boldsymbol{\omega}_0\right)\otimes\boldsymbol{q}_h\left(t-t_0,\boldsymbol{\omega}_0\right) \tag{2.11}$$

The two quaternions $\boldsymbol{q}_s$ and $\boldsymbol{q}_h$ parameterize a sequence of two rotations that take the spacecraft from its attitude at t_0 to its attitude at t. $\boldsymbol{q}_s$ is a rotation about the body frame z-axis at the body-axes nutation rate ω_{nb}, and $\boldsymbol{q}_h$ is a rotation about $\boldsymbol{h}$ at the inertial nutation rate ω_{ns}. The "$\otimes$" symbol represents standard quaternion multiplication, with the convention that $A\left(\boldsymbol{q}_a\otimes\boldsymbol{q}_b\right)=A\left(\boldsymbol{q}_a\right)A\left(\boldsymbol{q}_b\right)$.

These two quaternions are given by:

$$\boldsymbol{q}_s\left(t-t_0, \boldsymbol{\omega}_0\right) = \begin{bmatrix} 0 \\ 0 \\ \sin\left\{\frac{1}{2}\left(t-t_0\right)\omega_{nb}\right\} \\ \cos\left\{\frac{1}{2}\left(t-t_0\right)\omega_{nb}\right\} \end{bmatrix} \tag{2.12}$$

$$\boldsymbol{q}_h\left(t-t_0, \boldsymbol{\omega}_0\right) = \begin{bmatrix} \frac{I_{sc}\boldsymbol{\omega}_0}{\|I_{sc}\boldsymbol{\omega}_0\|}\sin\left\{\frac{1}{2}\left(t-t_0\right)\omega_{ns}\right\} \\ \cos\left\{\frac{1}{2}\left(t-t_0\right)\omega_{ns}\right\} \end{bmatrix} \tag{2.13}$$

The same quaternion state transition matrix can be written in terms of angular momentum by substituting Eqs. (2.8) and (2.9) (or the component-wise equivalents of Eq. (2.9)) into Eqs. (2.12) and (2.13). The new expressions are:

$$\boldsymbol{q}_s\left(t-t_0, \boldsymbol{h}_{sc0}\right) = \begin{bmatrix} 0 \\ 0 \\ \sin\left\{\frac{(t-t_0)}{2}\left(\frac{I_T-I_S}{I_T I_S}\right)h_{sc03}\right\} \\ \cos\left\{\frac{(t-t_0)}{2}\left(\frac{I_T-I_S}{I_T I_S}\right)h_{sc03}\right\} \end{bmatrix} \tag{2.14}$$

$$\boldsymbol{q}_h\left(t-t_0, \boldsymbol{h}_{sc0}\right) = \begin{bmatrix} \frac{\boldsymbol{h}_{sc0}}{\|\boldsymbol{h}_{sc0}\|}\sin\left\{\frac{(t-t_0)}{2}\frac{\|\boldsymbol{h}_{sc0}\|}{I_T}\right\} \\ \cos\left\{\frac{(t-t_0)}{2}\frac{\|\boldsymbol{h}_{sc0}\|}{I_T}\right\} \end{bmatrix} \tag{2.15}$$

The resulting transition quaternion is of the form given in Ref. [18]:

$$\begin{bmatrix} q_{\Phi 1}\left(t-t_0, \boldsymbol{h}_{sc0}\right) \\ q_{\Phi 2}\left(t-t_0, \boldsymbol{h}_{sc0}\right) \\ q_{\Phi 3}\left(t-t_0, \boldsymbol{h}_{sc0}\right) \\ q_{\Phi 4}\left(t-t_0, \boldsymbol{h}_{sc0}\right) \end{bmatrix} = \begin{bmatrix} \hat{h}_{sc01}\cos\alpha\sin\beta + \hat{h}_{sc02}\sin\alpha\sin\beta \\ \hat{h}_{sc02}\cos\alpha\sin\beta - \hat{h}_{sc01}\sin\alpha\sin\beta \\ \hat{h}_{sc03}\cos\alpha\sin\beta + \sin\alpha\cos\beta \\ \cos\alpha\cos\beta - \hat{h}_{sc03}\sin\alpha\sin\beta \end{bmatrix} \tag{2.16}$$

where

$$\alpha = \frac{(t-t_0)}{2}\omega_{nb} = \frac{(t-t_0)}{2}\left(\frac{I_T-I_S}{I_T I_S}\right)h_{sc03} \tag{2.17}$$

$$\beta = \frac{(t-t_0)}{2}\omega_{ns} = \frac{(t-t_0)}{2}\frac{\|\boldsymbol{h}_{sc0}\|}{I_T} \tag{2.18}$$

$$\hat{h}_{sc0i} = \frac{h_{sc0i}}{\|\boldsymbol{h}_{sc0}\|} \tag{2.19}$$

These substitutions can be made in Eq. (2.16) in order to show the exact dependence of the rotation quaternion on $\boldsymbol{h}_{sc0}$, but the resulting complicated formula has been omitted because it provides no useful insights.

The solutions for angular velocity and attitude contained in this section are solutions to an initial value problem: given $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$, they provide $\boldsymbol{q}(t)$ and $\boldsymbol{\omega}(t)$ for any future time t. This chapter seeks to solve the inverse problem, which is much less tractable. Even though the quaternion state transition matrix contains only ordinary trigonometric functions, it would clearly be difficult to solve Eqs. (2.16)-(2.19) for $\boldsymbol{h}_{sc0}$ for a given $\boldsymbol{q}_\Phi$ on the left-hand side of Eq. (2.16). Other parameterizations of the state transition matrix exist. It is not clear that any of them provide a simpler means to solve the axially-symmetric restricted problem. The results given here will nevertheless be helpful in finding some partial analytical solutions in Section 2.3.

2.2.3 Background from Attitude Determination

In addition to the q-method performance metric of Eq. (2.4), one additional attitude result is required. This result is a parameterization that provides a simple way to relate each vector measurement to the corresponding attitude quaternion for the rotation from inertial coordinates to body coordinates. This parameterization is developed in Ref. [22]. If there are no errors associated with the unit-length measurement vector $\boldsymbol{b}_i$, then the single measurement/reference pair $(\boldsymbol{b}_i, \boldsymbol{r}_i)$ only partially constrains the quaternion solution at time t_i. Each

matrix K_i individually has two identical maximum eigenvalues, and two corresponding eigenvectors that are orthogonal to each other and that span the solution subspace. This same subspace is spanned by the two orthogonal, unit-normalized quaternions [22]

$$\boldsymbol{q}_{\alpha i} = \begin{cases} \frac{1}{\sqrt{2\left(1+\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{r}_i\right)}} \begin{bmatrix} \boldsymbol{b}_i \times \boldsymbol{r}_i \\ \left(1+\boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{r}_i\right) \end{bmatrix} & \text{if } -1 < \boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{r}_i \\ \begin{bmatrix} \boldsymbol{c}_i \\ 0 \end{bmatrix} & \text{if } -1 = \boldsymbol{b}_i^{\mathrm{T}}\boldsymbol{r}_i \end{cases} \tag{2.20a}$$

$$\boldsymbol{q}_{\beta i} = \begin{bmatrix} \boldsymbol{b}_i \\ 0 \end{bmatrix} \otimes \boldsymbol{q}_{\alpha i} \tag{2.20b}$$

where $\boldsymbol{c}_i$ in Eq. (2.20a) is any unit vector orthogonal to $\boldsymbol{b}_i$. Any quaternion that represents a valid attitude solution given the single measurement pair can be written without loss of generality as the following linear combination of these two quaternions:

$$\boldsymbol{q}_{opt,i} = \boldsymbol{q}_{\alpha i} \cos\theta_i + \boldsymbol{q}_{\beta i} \sin\theta_i \tag{2.21}$$

Therefore, for each vector measurement, the set of possible attitude solutions is parameterized by the angle θ_i, which equals half of the only remaining unknown attitude rotation given $\boldsymbol{r}_i$ and $\boldsymbol{b}_i$. As parameterized here, it is a rotation about $\boldsymbol{b}_i$. This interpretation of θ_i is even more clearly seen by re-writing Eq. (2.21) as a quaternion product (sequence of rotations):

$$\boldsymbol{q}_{opt,i} = \begin{bmatrix} \boldsymbol{b}_i \sin\theta_i \\ \cos\theta_i \end{bmatrix} \otimes \boldsymbol{q}_{\alpha i} \tag{2.22}$$

It is helpful to introduce two additional eigenvectors of the K_i matrix, which correspond to the two non-maximum eigenvalues. Call them $\boldsymbol{q}_{ci}$ and $\boldsymbol{q}_{di}$. They are orthogonal to each other and to $\boldsymbol{q}_{\alpha i}$ and $\boldsymbol{q}_{\beta i}$. Therefore, these two quaternions are orthogonal to every possible solution quaternion. $\boldsymbol{q}_{ci}$ and $\boldsymbol{q}_{di}$ can be

computed by substituting the vector pair $(-\boldsymbol{b}_i, \boldsymbol{r}_i)$ into Eqs. (2.20), as described in Ref. [22]. Alternatively, one can use singular value decomposition or orthogonal/upper triangular factorization (QR factorization). All of these methods are able to produce a $(\boldsymbol{q}_{ci}, \boldsymbol{q}_{di})$ pair such that the matrix $[\boldsymbol{q}_{\alpha i},\ \boldsymbol{q}_{\beta i},\ \boldsymbol{q}_{ci},\ \boldsymbol{q}_{di}]$ is orthonormal.

2.3 Solutions to the Restricted Problem

As discussed in Section 2.2.1, the second generalized problem contains six scalar pieces of information, and thus requires a minimum of three vector measurements for a solution. With exactly three measurements, the original optimization problem reduces to nonlinear equation solving. Recall that the original optimization problem arises from an overdetermined system of measurement equations. In the present problem, the equations are fully determined but not overdetermined, which makes equation solving reasonable. Assuming all measurement errors are zero, Ref. [18] demonstrates that the general case is locally observable. There are multiple ways to approach the equation-solving problem, but the key issue of aliasing must be addressed first.

2.3.1 Aliasing

Any technique that relies on measurements obtained at different times must address the issue of aliasing. For a given angular velocity, the measurements must be taken with some minimum frequency in order to capture all information necessary for a solution. In the restricted case addressed in this chapter, the three

measurement times t_0, t_1, and t_2 are given and the angular velocity is unknown. Thus the issue of aliasing can be restated: For a given sampling rate, what is the maximum angular velocity (or equivalently, angular momentum) for which there are no aliasing concerns?

The Nyquist-Shannon sampling theorem states that the sampling frequency must be at least twice the value of any frequency component in the original signal that one wishes to reconstruct. In the problem considered here, the spacecraft rotation is characterized by two frequency components: ω_{nb} from Eq. (2.8) and ω_{ns} from Eq. (2.9). To avoid aliasing, therefore, the following constraints must hold:

$$|\omega_{nb}| < \omega_{max} \tag{2.23a}$$

$$\omega_{ns} < \omega_{max} \tag{2.23b}$$

where $\omega_{max} = \pi/\Delta t_{max}$ is the maximum allowable frequency, which is based on the maximum time between any two consecutive measurements, thereby making it conservative. Note that the definitions of the two nutations in Eqs. (2.8) and (2.9) allow ω_{nb}, but not ω_{ns}, to be negative. The conditions of Eqs. (2.23) constrain only the magnitudes of the two nutation rates.

Often in attitude determination scenarios, some amount of *a priori* information is available. If a spacecraft has been designed to spin at a certain rate, the actual angular velocity is likely to be at least of the same order of magnitude, barring extraordinary circumstances. Thus, it is often possible to either assume that aliasing has not occurred or to account for it explicitly in a solution. This chapter's methods generally assume, for the sake of simplicity, that aliasing has not occurred, but instances where such an assumption is necessary will be explicitly noted. If the angular velocity is so uncertain that aliasing cannot be

explicitly incorporated or eliminated in a solution, the practitioner should be very cautious about results.

2.3.2 Solution Paradigm

When approaching the restricted Wahba's problem from the perspective of equation solving, one constructs a system of equations that require the actual attitude and angular velocity histories to be perfectly consistent with all three vector measurements, as well as with the constraints of dynamics. For example, suppose the unknowns are $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$. As $\boldsymbol{q}_0$ has unit length, there are really only six independent scalar variables. Suppose the three vector measurements occur at times t_0, t_1, and t_2. To construct a system of equations with these unknowns, one uses the closed-form solutions for attitude and angular velocity in Section 2.2.2 to propagate the system forward in time and get expressions for $\boldsymbol{q}(t_1)$ and $\boldsymbol{q}(t_2)$ as functions of the elements of $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$.

Two scalar equations arise from the requirement that the attitude at each time t_i be consistent with the measurement vector pair $(\boldsymbol{b}_i, \boldsymbol{r}_i)$. As there are three measurements, the result is a total of six scalar equations. For each measurement pair $(\boldsymbol{b}_i, \boldsymbol{r}_i)$, use any appropriate method to compute $\boldsymbol{q}_{ci}$ and $\boldsymbol{q}_{di}$, as described in Section 2.2.3. Recall that these quaternions are orthogonal to every possible solution quaternion at time t_i. Thus, the computed expressions for attitude must satisfy

$$\boldsymbol{q}_{ci}^{\mathrm{T}}\boldsymbol{q}\left(t_i; \boldsymbol{q}_0, \boldsymbol{\omega}_0\right) = 0 \tag{2.24a}$$

$$\boldsymbol{q}_{di}^{\mathrm{T}}\boldsymbol{q}\left(t_i; \boldsymbol{q}_0, \boldsymbol{\omega}_0\right) = 0 \tag{2.24b}$$

for $i = 0, 1, 2$.

After stacking, the end result is a system of six highly nonlinear scalar equations of the form

$$\boldsymbol{f}\left(\boldsymbol{q}_0, \boldsymbol{\omega}_0\right) = 0 \tag{2.25}$$

Whereas an analytical solution to this system is unlikely, a numerical approach may be feasible. One of the most obvious such numerical approaches is Newton's method for nonlinear equation solving (as outlined in Section 4.7.6 of Ref. [23]). As applied to Eq. (2.25), Newton's method consists of starting with a first guess of $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$, linearizing Eq. (2.25) about that guess, solving the linearized equation for increments $\Delta\boldsymbol{q}_0$ and $\Delta\boldsymbol{\omega}_0$, using those increments to improve the guesses of $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$, and iterating. Of course, the unit normalization of $\boldsymbol{q}_0$ must be preserved in this process, perhaps by adding the constraint as a seventh equation. Note also that analytical derivatives have been used in all Newton-method linearizations of this chapter, as opposed to numerical derivatives.

Newton's method can be forced to converge by considering the cost function

$$J_{\boldsymbol{f}}\left(\boldsymbol{q}_0, \boldsymbol{\omega}_0\right) = \frac{1}{2}\boldsymbol{f}^{\mathrm{T}}\boldsymbol{f} \tag{2.26}$$

This cost function is non-negative, and any solution that causes this cost function to equal zero will exactly satisfy Eq. (2.25). It will also maximize the q-method performance metric of Eq. (2.4), or minimize the equivalent cost function of the form in Eq. (2.1).

Newton's method can be augmented with a step size selection along the search direction $\left[\Delta\boldsymbol{q}_0 \ \Delta\boldsymbol{\omega}_0\right]^{\mathrm{T}}$ in order to guarantee convergence to a solution that locally minimizes the cost function in Eq. (2.26), but this approach may fail to globally minimize the cost function. Even if it does reach the global minimum of $J_{\boldsymbol{f}} = 0$, there may be multiple attitude solutions that are globally optimal and

that fit the data, as was the case for the solution obtained in Ref. [18] for the first extended Wahba's problem. The questions of why and how multiple solutions arise are interesting and important, but beyond the scope of this chapter.

Although the system of equations derived above is straightforward, it may not be the best choice for numerical solution. By exploiting knowledge of the special dynamics of an axially-symmetric spacecraft, one can reduce this system of six equations and six unknowns to a simpler system, containing only three unknowns and three or more equations. As a general solution paradigm, then, this chapter proposes the following: First, algebraically manipulate expressions from attitude dynamics and kinematics to obtain a system of equations with good numerical properties. In particular, systems with a small number of equations or weaker nonlinearities that enable rapid convergence of gradient-based methods are sought. Second, apply Newton's method to search for solutions. With a sufficiently large number of initial guesses, some of them should converge to the correct solution. In the following subsection, the first part of this solution strategy is carried out by presenting two candidate sets of simplified equations. Section 2.4 completes the procedure by applying Newton's method to those equation sets for two simulated attitude scenarios. Of course, one could apply other equation-solving methods, such as non-gradient-based nonlinear least squares. This chapter focuses on Newton gradient methods because of project constraints and the author's familiarity with them.

An alternative solution paradigm directly solves six equations in six unknowns using numerical techniques. The present strategy uses analytical techniques to reduce the number of equations and especially unknowns. This reduction offers two advantages. First, global solution strategies require multiple

independent solutions that start from multiple guesses of the unknowns. A reduced dimensionality of the set of unknowns can greatly reduce the needed number of starting guesses. Second, the computational burden per Newton step scales as the cube of the number of unknowns.

2.3.3 Equation Sets

The two sets of equations presented here are representative of two major categories of such sets. The first category is distinguished by its use of attitude variables (and specifically the "θ" parameters of Section 2.2.3) as the reduced set of unknowns. For the second category, the unknowns are more closely related to angular rates than to attitudes; the method presented here uses the three components of the angular momentum vector in body axes at time t_1.

Attitude Variables Method

This approach uses the angles θ_0, θ_1, and θ_2, the angles used in attitude solutions in the form of Eq. (2.21), to form attitude quaternions. The basic idea is that given expressions for the attitude quaternions at the three times, one can compute two spacecraft rotations, one from t_0 to t_1 and the other from t_1 to t_2. There is an implicit relationship between these rotations and the angular momentum that caused them. The equations developed in this section are based on these relationships and the conservation of angular momentum, and their only unknowns are θ_0, θ_1, and θ_2.

Given values for the angles θ_0, θ_1, and θ_2, these angles can be used to produce

attitude quaternions at the three sample times:

$$\boldsymbol{q}\left(t_0\right) = \boldsymbol{q}_{\alpha 0}\cos\theta_0 + \boldsymbol{q}_{\beta 0}\sin\theta_0 \tag{2.27a}$$

$$\boldsymbol{q}\left(t_1\right) = \boldsymbol{q}_{\alpha 1}\cos\theta_1 + \boldsymbol{q}_{\beta 1}\sin\theta_1 \tag{2.27b}$$

$$\boldsymbol{q}\left(t_2\right) = \boldsymbol{q}_{\alpha 2}\cos\theta_2 + \boldsymbol{q}_{\beta 2}\sin\theta_2 \tag{2.27c}$$

The two transitional rotation quaternions can be computed as:

$$\boldsymbol{q}_{rot01}\left(\theta_0, \theta_1\right) = \boldsymbol{q}\left(t_0\right) \otimes \boldsymbol{q}^{-1}(t_1) \tag{2.28a}$$

$$\boldsymbol{q}_{rot21}\left(\theta_2, \theta_1\right) = \boldsymbol{q}\left(t_2\right) \otimes \boldsymbol{q}^{-1}(t_1) \tag{2.28b}$$

where the quaternion inverse $\boldsymbol{q}^{-1}$ is formed by negating the first three elements of $\boldsymbol{q}$. The form of the rotation quaternions is also known as a function of $\boldsymbol{h}_{sc1}$ from the quaternion state transition matrix solution in Eq. (2.16). Thus, one obtains the two equations

$$\boldsymbol{q}_{\Phi}\left(t_0 - t_1, \boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{rot01}\left(\theta_0, \theta_1\right) \tag{2.29a}$$

$$\boldsymbol{q}_{\Phi}\left(t_2 - t_1, \boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{rot21}\left(\theta_2, \theta_1\right) \tag{2.29b}$$

Now consider the form of the transition quaternion $\boldsymbol{q}_{\Phi}$, which comes from the application of Eq. (2.16) to the time steps $t_1 \to t_0$ and $t_1 \to t_2$:

$$\begin{bmatrix} q_{\Phi 1}\left(t_0 - t_1, \boldsymbol{h}_{sc1}\right) \\ q_{\Phi 2}\left(t_0 - t_1, \boldsymbol{h}_{sc1}\right) \\ q_{\Phi 3}\left(t_0 - t_1, \boldsymbol{h}_{sc1}\right) \\ q_{\Phi 4}\left(t_0 - t_1, \boldsymbol{h}_{sc1}\right) \end{bmatrix} = \begin{bmatrix} \hat{h}_{sc11}\cos\alpha_{01}\sin\beta_{01} + \hat{h}_{sc12}\sin\alpha_{01}\sin\beta_{01} \\ \hat{h}_{sc12}\cos\alpha_{01}\sin\beta_{01} - \hat{h}_{sc11}\sin\alpha_{01}\sin\beta_{01} \\ \hat{h}_{sc13}\cos\alpha_{01}\sin\beta_{01} + \sin\alpha_{01}\cos\beta_{01} \\ \cos\alpha_{01}\cos\beta_{01} - \hat{h}_{sc13}\sin\alpha_{01}\sin\beta_{01} \end{bmatrix} \tag{2.30a}$$

$$\begin{bmatrix} q_{\Phi 1}\left(t_2 - t_1, \boldsymbol{h}_{sc1}\right) \\ q_{\Phi 2}\left(t_2 - t_1, \boldsymbol{h}_{sc1}\right) \\ q_{\Phi 3}\left(t_2 - t_1, \boldsymbol{h}_{sc1}\right) \\ q_{\Phi 4}\left(t_2 - t_1, \boldsymbol{h}_{sc1}\right) \end{bmatrix} = \begin{bmatrix} \hat{h}_{sc11}\cos\alpha_{21}\sin\beta_{21} + \hat{h}_{sc12}\sin\alpha_{21}\sin\beta_{21} \\ \hat{h}_{sc12}\cos\alpha_{21}\sin\beta_{21} - \hat{h}_{sc11}\sin\alpha_{21}\sin\beta_{21} \\ \hat{h}_{sc13}\cos\alpha_{21}\sin\beta_{21} + \sin\alpha_{21}\cos\beta_{21} \\ \cos\alpha_{21}\cos\beta_{21} - \hat{h}_{sc13}\sin\alpha_{21}\sin\beta_{21} \end{bmatrix} \tag{2.30b}$$

where $\alpha_{ij} = 0.5(t_i - t_j)\omega_{nb}$ and $\beta_{ij} = 0.5(t_i - t_j)\omega_{ns}$, as in Eqs. (2.17) and (2.18). By inspection, the first two elements of the transition quaternions of Eqs. (2.30a) and (2.30b) can be rewritten:

$$\begin{bmatrix} q_{\Phi 1}\left(t_0 - t_1, \boldsymbol{h}_{sc1}\right) \\ q_{\Phi 2}\left(t_0 - t_1, \boldsymbol{h}_{sc1}\right) \end{bmatrix} = \sin\beta_{01} \begin{bmatrix} \cos\alpha_{01} & \sin\alpha_{01} \\ -\sin\alpha_{01} & \cos\alpha_{01} \end{bmatrix} \begin{bmatrix} \hat{h}_{sc11} \\ \hat{h}_{sc12} \end{bmatrix} \tag{2.31a}$$

$$\begin{bmatrix} q_{\Phi 1}\left(t_2 - t_1, \boldsymbol{h}_{sc1}\right) \\ q_{\Phi 2}\left(t_2 - t_1, \boldsymbol{h}_{sc1}\right) \end{bmatrix} = \sin\beta_{21} \begin{bmatrix} \cos\alpha_{21} & \sin\alpha_{21} \\ -\sin\alpha_{21} & \cos\alpha_{21} \end{bmatrix} \begin{bmatrix} \hat{h}_{sc11} \\ \hat{h}_{sc12} \end{bmatrix} \tag{2.31b}$$

In this form, one can see that the first two elements of each rotation quaternion are just the first two elements of the unit angular momentum vector, scaled by $\sin\beta$ and rotated by an angle α. The rotation in both cases occurs around the body-frame z-axis and within the body-frame x-y plane. The direction of the vector in Eq. (2.31a) lies along the projection of the angular momentum vector onto the x-y plane at time $(t_0 + t_1)/2$. Similarly, the vector in Eq. (2.31b) is parallel to the x-y projection of the angular momentum vector at time $(t_1 + t_2)/2$. This fact provides a means to determine the body axes nutation rate ω_{nb}, because these projections rotate at that constant rate. One simply computes the total rotation angle between the two vectors, which is equivalent to the angular rate multiplied by the time interval:

$$\alpha_{20} = \pm\arccos\left[\pm\left(\hat{\boldsymbol{q}}_{01} \cdot \hat{\boldsymbol{q}}_{21}\right)\right] = \alpha_{21} - \alpha_{01} = \omega_{nb}\frac{(t_2 - t_0)}{2} \tag{2.32}$$

where $\hat{\boldsymbol{q}}_{01}$ is the unit vector parallel to the first two elements of $\boldsymbol{q}_{\Phi}(t_0 - t_1, \boldsymbol{h}_{sc1})$, and $\hat{\boldsymbol{q}}_{21}$ is the unit vector parallel to the first two elements of $\boldsymbol{q}_{\Phi}(t_2 - t_1, \boldsymbol{h}_{sc1})$.

Note the two instances of the "$\pm$" symbol in Eq. (2.32). These represent two ambiguities that enter the equation set at this point. First, the outer "$\pm$" captures the fact that the arccosine function is only capable of specifying positive angles between 0 and π radians. In other words, this function outputs only the magnitude of the angle between the vectors, and not its sign. The meaning of

the inner "$\pm$" is more subtle, and relates to the mechanics of the quaternion parameterization of attitude. Although the quaternion parameterization is global and non-singular, it is not one-to-one; the same physical rotation is represented by $\boldsymbol{q}$ and $-\boldsymbol{q}$. Thus, it is possible to mean two different things when equating quaternions: that all four elements of the two quaternions are equal, or that the quaternions represent the same rotation. In Eq. (2.29), only the second (weaker) equivalence holds. Consequently, it is possible that one or both of the vectors $\hat{\boldsymbol{q}}_{01}$ and $\hat{\boldsymbol{q}}_{21}$ could have the wrong sign, and this would cause the magnitude of the computed arccosine to be in error. In practice, there are four different possible values of ω_{nb}, and only one of them will cause the quadratic cost function in the form of Eq. (2.26) to be zero, even when the variables have the correct values. This difficulty will be discussed in more detail later; for now, it is sufficient that the remainder of this derivation is not affected.

Because the body-axes nutation rate ω_{nb} is constant, Equation (2.32) can be used to compute α_{01} and α_{21}:

$$\alpha_{01} = \alpha_{20}\left(\tfrac{t_0 - t_1}{t_2 - t_0}\right) \text{ and } \alpha_{21} = \alpha_{20}\left(\tfrac{t_2 - t_1}{t_2 - t_0}\right) \tag{2.33}$$

These angles, in turn, enter into the quaternions parameterizing the rotations about the body z-axis.

$$\boldsymbol{q}_{s01}\left(t_0 - t_1; \theta_0, \theta_1, \theta_2\right) = \begin{bmatrix} 0 \\ 0 \\ \sin\alpha_{01} \\ \cos\alpha_{01} \end{bmatrix} \tag{2.34a}$$

$$\boldsymbol{q}_{s21}\left(t_2 - t_1; \theta_0, \theta_1, \theta_2\right) = \begin{bmatrix} 0 \\ 0 \\ \sin\alpha_{21} \\ \cos\alpha_{21} \end{bmatrix} \tag{2.34b}$$

As discussed in Section 2.2.2, the change in spacecraft attitude between any two times is known to be composed of two rotations, one about the body spin axis, here given by Eqs. (2.34), and one about the angular momentum vector. This second rotation about the angular momentum vector can be obtained for each time interval from Eq. (2.11), by using the inverses of the above quaternions and quaternion multiplication.

$$\boldsymbol{q}_{h01}\left(t_0-t_1;\theta_0,\theta_1,\theta_2\right)=\boldsymbol{q}_{s01}^{-1}(t_0-t_1;\theta_0,\theta_1,\theta_2)\otimes\boldsymbol{q}_{rot01}\left(t_0-t_1;\theta_0,\theta_1,\theta_2\right) \tag{2.35a}$$

$$\boldsymbol{q}_{h21}\left(t_2-t_1;\theta_0,\theta_1,\theta_2\right)=\boldsymbol{q}_{s21}^{-1}(t_2-t_1;\theta_0,\theta_1,\theta_2)\otimes\boldsymbol{q}_{rot21}\left(t_2-t_1;\theta_0,\theta_1,\theta_2\right) \tag{2.35b}$$

Equations (2.35) can be equated to the expression for $\boldsymbol{q}_h$ given in Eq. (2.15), applied to the appropriate time intervals.

$$\boldsymbol{q}_{h01}\left(t_0-t_1;\theta_0,\theta_1,\theta_2\right)=\boldsymbol{q}_{h01}\left(t_0-t_1;\boldsymbol{h}_{sc1}\right)=\begin{bmatrix}\left(\frac{\boldsymbol{h}_{sc1}}{\|\boldsymbol{h}_{sc1}\|}\right)\sin\left\{\frac{(t_0-t_1)}{2}\frac{\|\boldsymbol{h}_{sc1}\|}{I_T}\right\}\\ \cos\left\{\frac{(t_0-t_1)}{2}\frac{\|\boldsymbol{h}_{sc1}\|}{I_T}\right\}\end{bmatrix} \tag{2.36a}$$

$$\boldsymbol{q}_{h21}\left(t_2-t_1;\theta_0,\theta_1,\theta_2\right)=\boldsymbol{q}_{h21}\left(t_2-t_1;\boldsymbol{h}_{sc1}\right)=\begin{bmatrix}\left(\frac{\boldsymbol{h}_{sc1}}{\|\boldsymbol{h}_{sc1}\|}\right)\sin\left\{\frac{(t_2-t_1)}{2}\frac{\|\boldsymbol{h}_{sc1}\|}{I_T}\right\}\\ \cos\left\{\frac{(t_2-t_1)}{2}\frac{\|\boldsymbol{h}_{sc1}\|}{I_T}\right\}\end{bmatrix} \tag{2.36b}$$

Finally, the theoretical expressions given by Eqs. (2.36) allow the formation of several equations in terms of the three θ angles. The first three elements of each quaternion constitute a unit vector in the direction of the angular momentum vector, scaled by the sine expression. Because the vectors are parallel to each other,

$$[\boldsymbol{q}_{h01}]_{1:3}\times[\boldsymbol{q}_{h21}]_{1:3}=\begin{bmatrix}0\\0\\0\end{bmatrix} \tag{2.37}$$

This expression comprises three equations that can be used to form part of the Newton's method cost function, but they are not all independent. Another equation comes from the arguments of the sines and cosines within $\boldsymbol{q}_{h01}$ and $\boldsymbol{q}_{h21}$. These arguments, called β_{01} and β_{21}, can be extracted via an arctangent function on the quaternion elements, and they are related to each other by

$$\beta_{01}\left(t_2-t_1\right)=\beta_{21}\left(t_0-t_1\right) \tag{2.38}$$

One final pair of equations constrains the α and β values to be consistent with each other. The correct relationships are

$$\frac{\alpha_{01}}{\beta_{01}}=\left(\frac{I_T-I_S}{I_S}\right)\hat{h}_{sc13}=\left(\frac{I_T-I_S}{I_S}\right)\frac{\left[\boldsymbol{q}_{h01}\left(t_0-t_1;\theta_0,\theta_1,\theta_2\right)\right]_3}{\left\|\left[\boldsymbol{q}_{h01}\left(t_0-t_1;\theta_0,\theta_1,\theta_2\right)\right]_{1:3}\right\|} \tag{2.39a}$$

$$\frac{\alpha_{21}}{\beta_{21}}=\left(\frac{I_T-I_S}{I_S}\right)\hat{h}_{sc13}=\left(\frac{I_T-I_S}{I_S}\right)\frac{\left[\boldsymbol{q}_{h21}\left(t_2-t_1;\theta_0,\theta_1,\theta_2\right)\right]_3}{\left\|\left[\boldsymbol{q}_{h21}\left(t_2-t_1;\theta_0,\theta_1,\theta_2\right)\right]_{1:3}\right\|} \tag{2.39b}$$

Equations (2.37), (2.38), and (2.39) constitute six scalar equations in three scalar unknowns. Although it would seem that more than three equations are unnecessary when solving for just three unknowns, recall that not all the equations are independent. Unfortunately, it can be challenging to guarantee independence of particular subsets of three of these equations. In practice, one can safely delete the third row of Eq. (2.37), thereby reducing to five equations. Experience indicates that Eq. (2.38) and at least one of Eqs. (2.39a) or (2.39b) must be retained. Newton's method is capable of solving systems with more equations than unknowns, provided the equations form a self-consistent set. Therefore, it is reasonable to retain five equations.

At this point it is appropriate to address once again the multiplicity of ω_{nb} values calculated by Eq. (2.32). There is no theoretical way to choose the correct ω_{nb} and reject the incorrect values. An ad hoc approach has, however, been

successful. After computing all four possible values, the remainder of the algorithm is executed separately for each possible value. The quadratic cost $J_{\boldsymbol{f}}$ is evaluated for each of the four instances, and then the correcting step for the next iteration of Newton's method is computed only for the value of ω_{nb} associated with the lowest cost.

The discerning reader may have noticed one other problem with this derivation. Both the arccosine and arctangent functions are only capable of producing outputs within certain limited ranges of angles. In addition to the ambiguities associated with Eq. (2.32), the actual argument of the cosine or tangent could differ by an integer multiple of 2π. In both cases, however, adding or subtracting any multiple of 2π would imply the violation of the anti-aliasing conditions established by Eqs. (2.23a) and (2.23b). Thus, no ambiguities arise if the no-aliasing assumption is enforced.

Angular Momentum Variables Method

The second approach defines a system of equations with the three body-axes components of angular momentum at time t_1 as unknowns. One additional attitude parameter, θ_1, also enters the derivation at intermediate stages, but it can be eliminated easily from the final equations. Conceptually, this method starts with an attitude and angular momentum at the second of the three measurement times, t_1. It then propagates the system dynamically backward and forward in time, and requires that the resulting attitudes be consistent with the first and third measurements. This parameterization automatically satisfies the attitude measurement equations at time t_1. The remaining attitude measurement equations at times t_0 and t_2 constitute the system that this method solves

to determine the unknown angular momentum.

Assume that angular momentum at time t_1 is given by

$$\boldsymbol{h}_{sc1} = \boldsymbol{h}_{sc}\left(t_1\right) = \begin{bmatrix} h_{sc11} & h_{sc12} & h_{sc13} \end{bmatrix}^{\mathrm{T}} \tag{2.40}$$

Furthermore, the optimal quaternion at t_1 can be written according to Eq. (2.21) as

$$\boldsymbol{q}\left(t_1\right) = \boldsymbol{q}_{\alpha 1}\cos\theta_1 + \boldsymbol{q}_{\beta 1}\sin\theta_1 \tag{2.41}$$

Compute the quaternion rotations $\boldsymbol{q}_{\Phi}(t_0 - t_1, \boldsymbol{h}_{sc1})$ from time t_1 to time t_0 and $\boldsymbol{q}_{\Phi}(t_2 - t_1, \boldsymbol{h}_{sc1})$ from time t_1 to time t_2 by employing Eq. (2.16), and use them to find $\boldsymbol{q}(t_0)$ and $\boldsymbol{q}(t_2)$ as functions of $\boldsymbol{h}_{sc1}$ and θ_1:

$$\boldsymbol{q}\left(t_0\right) = \boldsymbol{q}_{\Phi}\left(t_0 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \left[\boldsymbol{q}_{\alpha 1}\cos\theta_1 + \boldsymbol{q}_{\beta 1}\sin\theta_1\right] \tag{2.42a}$$

$$\boldsymbol{q}\left(t_2\right) = \boldsymbol{q}_{\Phi}\left(t_2 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \left[\boldsymbol{q}_{\alpha 1}\cos\theta_1 + \boldsymbol{q}_{\beta 1}\sin\theta_1\right] \tag{2.42b}$$

Next, compute $\boldsymbol{q}_{c0}$ and $\boldsymbol{q}_{d0}$ as described at the end of Section 2.2.3. Likewise, compute $\boldsymbol{q}_{c2}$ and $\boldsymbol{q}_{d2}$. Both $\boldsymbol{q}_{c0}$ and $\boldsymbol{q}_{d0}$ are orthogonal to any quaternion solution at t_0, and both $\boldsymbol{q}_{c2}$ and $\boldsymbol{q}_{d2}$ are orthogonal to any quaternion solution at t_2. Consequently, they can be used to form the four equations

$$0 = \boldsymbol{q}_{c0}^{\mathrm{T}}\,\boldsymbol{q}\left(t_0\right) \tag{2.43a}$$

$$0 = \boldsymbol{q}_{d0}^{\mathrm{T}}\,\boldsymbol{q}\left(t_0\right) \tag{2.43b}$$

$$0 = \boldsymbol{q}_{c2}^{\mathrm{T}}\,\boldsymbol{q}\left(t_2\right) \tag{2.43c}$$

$$0 = \boldsymbol{q}_{d2}^{\mathrm{T}}\,\boldsymbol{q}\left(t_2\right) \tag{2.43d}$$

where $\boldsymbol{q}(t_0)$ and $\boldsymbol{q}(t_2)$ are the functions of h_{sc11}, h_{sc12}, h_{sc13}, and θ_1 defined in Eqs. (2.42). Satisfaction of these two equations will guarantee that the attitudes determined at times t_0 and t_2 are consistent with the measurements associated

with those times. One could also write these two equations in matrix form as

$$\begin{bmatrix} \mu_1\left(\boldsymbol{h}_{sc1}\right) & \eta_1\left(\boldsymbol{h}_{sc1}\right) \\ \mu_2\left(\boldsymbol{h}_{sc1}\right) & \eta_2\left(\boldsymbol{h}_{sc1}\right) \\ \mu_3\left(\boldsymbol{h}_{sc1}\right) & \eta_3\left(\boldsymbol{h}_{sc1}\right) \\ \mu_4\left(\boldsymbol{h}_{sc1}\right) & \eta_4\left(\boldsymbol{h}_{sc1}\right) \end{bmatrix} \begin{bmatrix} \cos\theta_1 \\ \sin\theta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.44}$$

where the μ_j and η_j functions are defined according to

$$\mu_1\left(\boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{c0}^{\mathrm{T}}[\boldsymbol{q}_{\Phi}\left(t_0 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \boldsymbol{q}_{\alpha 1}] \tag{2.45a}$$

$$\mu_2\left(\boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{d0}^{\mathrm{T}}[\boldsymbol{q}_{\Phi}\left(t_0 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \boldsymbol{q}_{\alpha 1}] \tag{2.45b}$$

$$\mu_3\left(\boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{c2}^{\mathrm{T}}[\boldsymbol{q}_{\Phi}\left(t_2 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \boldsymbol{q}_{\alpha 1}] \tag{2.45c}$$

$$\mu_4\left(\boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{d2}^{\mathrm{T}}[\boldsymbol{q}_{\Phi}\left(t_2 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \boldsymbol{q}_{\alpha 1}] \tag{2.45d}$$

and

$$\eta_1\left(\boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{c0}^{\mathrm{T}}\left[\boldsymbol{q}_{\Phi}\left(t_0 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \boldsymbol{q}_{\beta 1}\right] \tag{2.46a}$$

$$\eta_2\left(\boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{d0}^{\mathrm{T}}\left[\boldsymbol{q}_{\Phi}\left(t_0 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \boldsymbol{q}_{\beta 1}\right] \tag{2.46b}$$

$$\eta_3\left(\boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{c2}^{\mathrm{T}}\left[\boldsymbol{q}_{\Phi}\left(t_2 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \boldsymbol{q}_{\beta 1}\right] \tag{2.46c}$$

$$\eta_4\left(\boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_{d2}^{\mathrm{T}}\left[\boldsymbol{q}_{\Phi}\left(t_2 - t_1; \boldsymbol{h}_{sc1}\right) \otimes \boldsymbol{q}_{\beta 1}\right] \tag{2.46d}$$

At this point, the variable θ_1 can be eliminated from Eq. (2.44) by using any of the rows of the 4×2 matrix of μ_j and η_j functions in conjunction with an arctangent function. Alternatively, one can recognize the form of Eq. (2.44) as a nullspace problem. In order for the matrix on the left to have a nullspace, $\boldsymbol{h}_{sc1}$ must be such that the two matrix columns are parallel. One can thus construct multiple equations of the form

$$\mu_j(\boldsymbol{h}_{sc1})\eta_l(\boldsymbol{h}_{sc1}) - \eta_j(\boldsymbol{h}_{sc1})\mu_l(\boldsymbol{h}_{sc1}) = 0; \quad j, l \in \{1, 2, 3, 4\},\ j \neq l \tag{2.47}$$

There are six possible equations of this form, and only three unknowns. Any three equations containing terms from all four rows of the matrix will be an independent set, and the remaining three possible equations can be formed from these by algebraic substitution. In practice, however, it has been found that at least four equations must be employed in order to avoid a singularity when one or more of the rows $[\mu_j,\ \eta_j]$ may equal $[0,\ 0]$. Additionally, normalization of the $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ columns may improve numerical properties of the problem.

The equation form of (2.47) is not the only way to deal with Eq. (2.44). For instance, another approach solves for the optimal θ_1 by taking the derivative with respect to θ_1 of the sum of the squared residuals of Eq. (2.44) and setting it equal to zero. After applying several trigonometric identities, this procedure yields an analytical solution for θ_1. This solution can then be substituted back into Eq. (2.44) to obtain four equations in only the $\boldsymbol{h}_{sc1}$ variables.

Examination of the nullspace equation (2.44) suggests some interesting connections between this set of equations with angular momentum unknowns and the original statement of the q-method solution in Eq. (2.4). To see this relationship more clearly, augment Eqs. (2.43) with two trivial equations relating the attitude at time t_1 in Eq. (2.41) to quaternions $\boldsymbol{q}_{c1}$ and $\boldsymbol{q}_{d1}$, which by definition are orthogonal to every possible solution $\boldsymbol{q}(t_1)$.

$$0 = \boldsymbol{q}_{c1}^{\mathrm{T}}\, \boldsymbol{q}\,(t_1) \tag{2.48a}$$

$$0 = \boldsymbol{q}_{d1}^{\mathrm{T}}\, \boldsymbol{q}\,(t_1) \tag{2.48b}$$

By combining Eqs. (2.43) and (2.48) and expanding them according to Eqs. (2.42a) and (2.42b), the nullspace equation (2.44) can be rewritten in an

equivalent form:

$$\boldsymbol{f}\left(\boldsymbol{h}_{sc1},\boldsymbol{q}_1\right)=\begin{bmatrix}\boldsymbol{q}_{c0}^{\mathrm{T}}\,\Phi\left(t_0-t_1;\boldsymbol{h}_{sc1}\right)\\ \boldsymbol{q}_{d0}^{\mathrm{T}}\,\Phi\left(t_0-t_1;\boldsymbol{h}_{sc1}\right)\\ \boldsymbol{q}_{c1}^{\mathrm{T}}\\ \boldsymbol{q}_{d1}^{\mathrm{T}}\\ \boldsymbol{q}_{c2}^{\mathrm{T}}\,\Phi\left(t_2-t_1;\boldsymbol{h}_{sc1}\right)\\ \boldsymbol{q}_{d2}^{\mathrm{T}}\,\Phi\left(t_2-t_1;\boldsymbol{h}_{sc1}\right)\end{bmatrix}\boldsymbol{q}_1=\begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix} \tag{2.49}$$

A quadratic cost function based on this system of equations then takes the form

$$\begin{aligned}J\left(\boldsymbol{q}_1,\boldsymbol{h}_{sc1}\right) &= \boldsymbol{f}\left(\boldsymbol{h}_{sc1},\boldsymbol{q}_1\right)^{\mathrm{T}}\boldsymbol{f}\left(\boldsymbol{h}_{sc1},\boldsymbol{q}_1\right)\\ &= \boldsymbol{q}_1^{\mathrm{T}}\left[\textstyle\sum_{i=0}^{2}\Phi^{\mathrm{T}}(t_i-t_1;\boldsymbol{h}_{sc1})\left(\boldsymbol{q}_{ci}\boldsymbol{q}_{ci}^{\mathrm{T}}+\boldsymbol{q}_{di}\boldsymbol{q}_{di}^{\mathrm{T}}\right)\Phi\left(t_i-t_1;\boldsymbol{h}_{sc1}\right)\right]\boldsymbol{q}_1\end{aligned} \tag{2.50}$$

where $\Phi\left(t_1-t_1;\boldsymbol{h}_{sc1}\right)$ is just the identity matrix I_4. Now compare this cost to the performance metric $\bar{p}\left(\boldsymbol{q}_0,\boldsymbol{\omega}_0\right)$ given in Eq. (2.4b). The cost in Eq. (2.50) is referenced to time t_1 rather than t_0, and it uses angular momentum rather than angular velocity as a variable. Both of these differences are purely matters of convenience, however. It is straightforward to show that the performance metric of Eq. (2.4b) is maximized by the same $\boldsymbol{q}_1$ that minimizes the cost of Eq. (2.50). A short proof of this result is included in the Appendix, along with citations of closely related derivations. Thus, in this form the equation set with angular momentum variables appears to be an implementation of the generalized q-method. In practice, however, there is a slight difference. For numerical solution, it has been found that equations based on Eq. (2.44) perform better than equations based on the modified form of Eq. (2.49). Effectively, this means that for a given guess of the $\boldsymbol{h}_{sc1}$ variables, the quadratic cost associated with the t_1 term of Eq. (2.50) is forced to be zero, and then the remaining two terms are minimized. This effective cost converges to the true Wahba-equivalent cost with equally weighted measurements only in the limit as the $\boldsymbol{h}_{sc1}$ variables approach the true angular momentum.

Other Methods

Several other sets of equations based on different variables have been developed, in addition to the two given in detail in the preceding sections. One of these employs a hybrid set of both attitude and rate variables as its unknowns, and its equations are similar to those derived in sections 2.3.3 and 2.3.3. Another approach does not reduce the equations analytically, but instead divides its numerical solution procedure into two stages that simplify the calculations. This approach bears some similarity to the inner-outer optimization procedure described in Section 2.2.1 and in Ref. [18]. Although both of these alternative methods are successful, they do not provide any obvious advantages over the two approaches already discussed. Therefore the numerical results presented in Section 2.4 are restricted to the attitude-variables and angular-momentum-variables methods for conciseness.

2.3.4 Initialization Procedures

Before presenting numerical results, it is important to consider the initialization procedures used to generate first guesses for Newton's method in the numerical solutions of the nonlinear equation sets. The second generalized Wahba's problem was shown to be locally observable in Ref. [18]. This means that for an initial guess that is "close enough", Newton's method will converge to the true solution. It is not generally possible, however, to determine analytically a distance in the variable space that is "close enough". Thus, the procedure that chooses initial guesses is crucial to the success of the algorithm.

The initialization procedures used in this chapter's implementations are spe-

cific to particular sets of nonlinear equations, but they are ad hoc in many respects. Two constraints must be respected: The initialization procedure must generate only initial guesses that are physically reasonable, and it must generate a sufficient number of initial guesses so that Newton's method will converge to every solution point. Experience dictates that about 1000 randomly generated, physically realistic first guesses are usually enough to satisfy the second of these requirements for each of the equation sets. If too few initial guesses are used, then the algorithm will miss some solutions, as is well-known in the field of nonlinear optimization and equation solving. The needed number of guesses is problem-dependent; it has been determined by trial and error in this study. The choice of physically realistic ranges naturally depends on what the solution variables represent. The practical effect of this restriction is to avoid wasteful computation.

For the equation set with the three angular θ parameters as its variables, the units are radians. These angles only enter the equations as the arguments of trigonometric functions, so initial guesses outside of the range $[-\pi, \pi]$ are redundant. Furthermore, given a particular measurement pair and the corresponding quaternions $\boldsymbol{q}_{\alpha i}$ and $\boldsymbol{q}_{\beta i}$ from Eqs. (2.20a) and (2.20b), application of Eq. (2.21) shows that $\boldsymbol{q}(\theta_i + \pi) = -\boldsymbol{q}(\theta_i)$, which is physically equivalent to $\boldsymbol{q}(\theta_i)$. Consequently, it is possible to represent every attitude by an angle $\theta \in [0, \pi)$. In practice, however, it seems preferable to use the entire range $[-\pi, \pi]$ for the θ variables, and recognize the equivalence of some of the solutions produced. Thus, the initial guesses can be drawn randomly from a cube with side lengths of 2π, and throughout the Newton's method procedure each iteration can be constrained to fall within this box by adding or subtracting integer multiples of 2π when necessary.

When the components of angular momentum $\boldsymbol{h}_{sc1}$ are used as variables, there is no such "natural" limit on their magnitude. As developed in this chapter, the particular choice of units of angular momentum is irrelevant, provided it is consistent with the units used for other physical quantities. A reasonable angular momentum is one that complies with the no-aliasing conditions of Eqs. (2.23a) and (2.23b). Application of these constraints bounds the magnitude of $\boldsymbol{h}_{sc1}$ in accordance with Eqs. (2.8) and (2.9), which relate the nutation rates to angular momentum. In contrast, the no-aliasing assumption is upheld for the attitude-variables equations by the structure of the equations themselves, and specifically by the way arccosine and arctangent ambiguities are treated in the algorithm.

The constraints on the magnitudes of individual elements of $\boldsymbol{h}_{sc1}$ are not independent. Substitution into Eq. (2.23a) yields a maximum allowable magnitude for the third angular momentum component h_{sc13}:

$$|h_{sc13}| < \left| \frac{I_T I_S}{I_T - I_S} \right| \frac{\pi}{\Delta t_{max}} \tag{2.51}$$

Likewise, substitution of Eq. (2.9) into Eq. (2.23b) bounds the magnitude of the entire angular momentum vector according to

$$\|\boldsymbol{h}_{sc1}\| < I_T \frac{\pi}{\Delta t_{max}} \tag{2.52}$$

Depending on the relative values of I_T and I_S, it may be possible to ignore Eq. (2.51) altogether because Eq. (2.52) is more restrictive. In this case, the valid solution space is a sphere, and initial guesses can be drawn randomly from within that sphere. Alternatively, it may be necessary to consider both of the bounding conditions. The resulting solution space is a sphere truncated by two parallel planes, equidistant from the plane defined by $z = 0$.

Of course, the defined solution spaces can be used not only for initialization but also as limits that halt the iterations of Newton's method when necessary to prevent it from diverging without bound. Furthermore, because the non-aliasing constraints for this method are imposed externally rather than by the equations themselves, it is possible to relax the non-aliasing assumption by defining a larger valid solution space. This will generally have the effect of producing multiple physically feasible solutions at different "energy levels", which cannot be distinguished by the vector measurements alone.

Given a willingness to do computations 1000 times, one might try a simpler or more standard solution method. One approach might evaluate the sum of the squared equation errors 1000 times, as in Eq. (2.26), and choose the guess with the lowest squared error as the solution. Unfortunately, this approach would require millions rather than thousands of guesses to achieve degree-level accuracy. Alternatively, one might run 1000 parallel EKFs with different initial conditions, similar to the approach of Ref. [11]. This strategy, however, does not have the convergence guarantee of the Gauss-Newton method. Furthermore, the computational burden might be equivalent to or greater than the burden of this chapter's methods.

2.4 Results

The numerical techniques described above, using Newton's method and linearizing the equations at each step, have been applied to each of the sets of reduced nonlinear equations derived in Section 2.3.3. For simplicity, only a subset of representative results is presented here, but all are qualitatively similar.

The following procedure has been used: First, a "truth" model simulation has been run to create a simulated sequence of vector measurements and the corresponding "truth" time histories of attitude and angular rate. Two dynamic scenarios have been studied, with the spacecraft symmetric about the major axis and the minor axis, respectively. Table 2.1 lists some of the model characteristics used in the "truth" model simulations. The parameters in the middle column

Table 2.1: Characteristics of "truth" model scenarios.

	Scenario #1	Scenario #2
Symmetric about	Major axis	Minor axis
I_T/I_S ratio	0.5227	4.0417
$\|\|\boldsymbol{\omega}_{sc}\|\|$ (rad/s)	0.0967	0.0764
$2\pi/\|\|\boldsymbol{\omega}_{sc}\|\|$, spin period (s)	65.0	82.2
$2\pi/\|\omega_{nb}\|$, body-axes nutation period (s)	71.2	1382
$2\pi/\omega_{ns}$, inertial nutation period (s)	34.0	82.5
Δt (s)	15.0	27.4
$\angle$ between $\boldsymbol{h}_{sc}$ & z-axis (deg)	0.670	88.9

correspond to major-axis spin with small nutations. The far right column is for minor-axis spin with very large nutations, characteristic of the flat spin that might occur at the end of a sounding rocket flight such as in Ref. [24]. Note how the sample time Δt is significantly shorter than any of the characteristic periods, thereby avoiding aliasing. Next, a large set of initial guesses has been generated according to the principles established in Section 2.3.4. A distinct initialization is required for each combination of scenario and equation set. Newton's method has been applied starting from each initial guess until the solution parameters converge to a global minimum that solves all the equations, get trapped at a local minimum, or diverge to unreasonable parameter values. For those points which exactly solve the equations to the limits of numerical precision, additional post-processing has been performed to determine any other solution quantities,

such as angular rates or quaternions, not explicitly calculated for that particular equation set. These solutions are compared to the "truth" values generated at the beginning of the procedure.

In most cases, the numerical procedure finds more than one physically distinct solution. Even discounting multiplicity due to physical equivalence, there are global minima that solve the nonlinear equations but do not match the "truth" model. This behavior indicates that the restricted problem with exactly three measurements is only locally observable; often there is more than one possible scenario that explains the measurement sequence even when anti-aliasing constraints are imposed. It has been found that such solution ambiguities can be resolved easily with one additional measurement at time t_3. To incorporate this measurement, each of the candidate attitude solutions has been dynamically propagated to the new measurement time by means of the quaternion state transition matrix of Eq. (2.10). The nullspace quaternions $\boldsymbol{q}_{c3}$ and $\boldsymbol{q}_{d3}$ have been computed, and the true solution (if present) is the one for which both $\boldsymbol{q}_{c3}^{\mathrm{T}}\boldsymbol{q}(t_3) = 0$ and $\boldsymbol{q}_{d3}^{\mathrm{T}}\boldsymbol{q}(t_3) = 0$ are satisfied. The success of this technique indicates that the unrestricted problem (with more than three vector measurements) is globally observable, even if the restricted problem is not.

The results presented in the remainder of this section are organized as follows: Section 2.4.1 presents "baseline" results for the application of the attitude-variables and angular-momentum-variables methods to the two scenarios described in Table 2.1. No errors are included in the measurements for this first set of results. Section 2.4.2 analyzes the effects of measurement errors on the success and solution accuracy of this chapter's methods. Further discussion of some aspects of the various cases is contained in Section 2.4.3, and a prac-

tical implementation method is proposed in Section 2.4.4. Note that missing from this analysis is any comparison with algorithms developed by other practitioners, because the author is unaware of any methods that solve a sufficiently similar problem.

2.4.1 Baseline Cases

All of the baseline, no-error cases successfully solved the restricted problem to obtain attitude and angular velocity. These results are presented in Figs. 2.1-2.4.

Figures 2.1 and 2.2 give the solutions obtained for the two "truth" model sce-

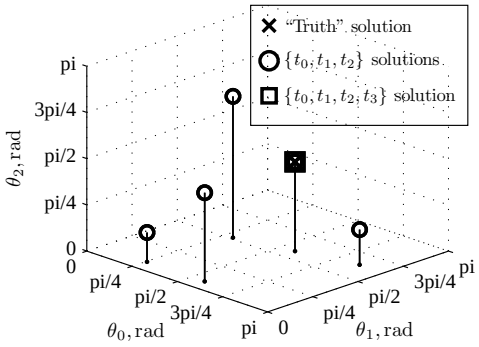

Figure 2.1: Attitude-variables method applied to major-axis spinner

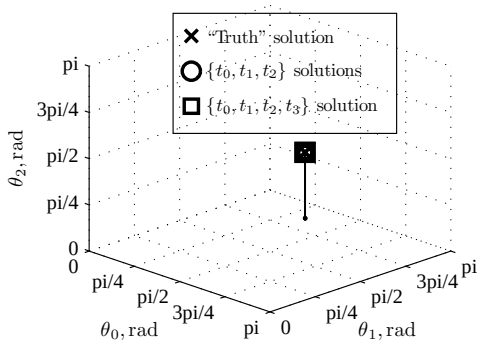

Figure 2.2: Attitude-variables method applied to minor-axis spinner

narios using the attitude-variables method of Section 2.3.3. Only solutions in the octant with three positive θ-angles are plotted to eliminate physical duplicates. Likewise, Figs. 2.3 and 2.4 display the solutions found by the angular-momentum-variables method of Section 2.3.3 for the two "truth" model scenarios. In each of these four figures, the round markers represent the solutions to the pertinent nonlinear equation set for the initial three measurements. The square markers designate solutions that, when propagated to a fourth measure-

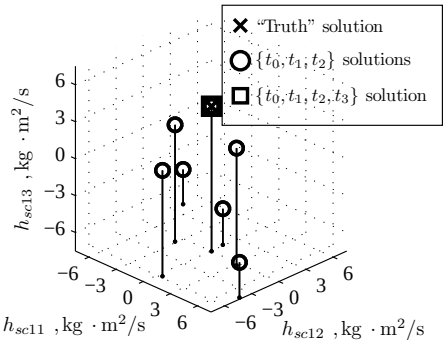

Figure 2.3: Angular-momentum-variables method applied to major-axis spinner

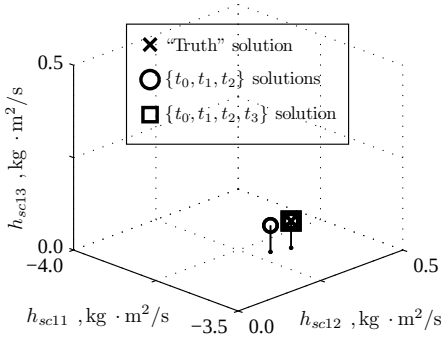

Figure 2.4: Angular-momentum-variables method applied to minor-axis spinner

ment time, also conform to the measurement at that time. In all cases, these four-measurement solutions match the "truth" solutions, indicated by X's, to numerical precision limits. Although Figs. 2.1 and 2.3 show round markers far away from the "truth" X markers, this does not indicate algorithm failure. Rather, it indicates that the underlying three-measurement problem does not have a unique solution. These ambiguities can be resolved only by considering additional measurements.

Note that Fig. 2.2 is somewhat misleading. It seems to show that only one solution was obtained, the true solution. In reality, two solutions have been obtained. They are so close that their circle markers lie on top of each other in Fig. 2.2. Only one, however, passes the t_3-measurement "truth" test.

Performance of successful algorithms can be compared based on how many of the initial guesses converge to solve the equations, and whether the true solution is likely to be among the set of numerical solutions for a given number of first guesses. These criteria are closely related to size of the "convergent neighborhood" surrounding each minimum of the sum-squared equation errors, or

how many random initial points are necessary in order to have one of them sufficiently close to the true solution to converge to it. Poor convergence may indicate problems with the numerical properties of the equation set, low observability of the underlying scenario, or both.

For the cases considered, the scenario with a major-axis spinner and small nutations tends to have a greater number of feasible solutions, as illustrated by the circles in Figs. 2.1 and 2.3. Also, a greater percentage of initial points converge to some solution of the equations. In contrast, the scenario with a greatly nutating minor-axis spinner yields only two feasible solutions, and only a small percentage of initial guesses converge to solutions at all. Figs. 2.2 and 2.4 demonstrate this behavior. For example, 11.3% of the initial guesses converged for Fig. 2.3 (major-axis), but only 1.4% for Fig. 2.4 (minor-axis). Comparison of the two equation sets suggests that the attitude-variables method tends to find a feasible solution for a greater percentage of its first guesses. This advantage, however, may be offset by its approximately fivefold increase in computational burden per guess.

2.4.2 Measurement Error Effects

The impacts of measurement errors are a necessary consideration for any practical implementation of this chapter's algorithms. Two distinct effects are possible. First, sufficiently large measurement errors may prevent the equation-solving procedures from converging to a solution near the "truth". This type of behavior depends at least in part on the system of equations being solved. Second, the inherent poor observability of a physical configuration may tend

to amplify errors in solutions that do converge. This second effect is in theory independent of the particular system of equations applied because each such system yields the same attitude solution for a given set of measurements.

To investigate the size of the measurement error effects, a Monte Carlo approximation of the estimation error covariance matrix has been computed for each of the scenarios in Table 2.1. For a particular measurement error standard deviation σ_{meas}, a large number of example problems have been generated with the same "truth" states, but different simulated random measurement errors. The measurement error levels have been chosen to correspond to typical attitude sensor errors, such as those from a magnetometer, sun sensor, or star camera. For each set of measurements a solution to the equations has been sought starting from the "truth" solution as an initial guess. Standard deviations of the resulting solutions relative to the "truth" $\boldsymbol{\omega}_0$ and $\boldsymbol{q}_0$ have been computed by standard techniques.

For a given physical scenario, these solution standard deviations grow linearly with the measurement error standard deviations. This trend is shown in Figures 2.5-2.6. The standard deviations of initial attitude and angular velocity for the major-axis spinner with small nutations are plotted in these figures. The quantity $\sigma_{\boldsymbol{q}0,k}$ is the angular initial attitude error standard deviation around the k-axis in degrees, and $\sigma_{\boldsymbol{\omega}0,k}$ is the standard deviation of the initial k-axis component of angular velocity error in degrees/s. In each figure, the standard deviation of measurement error is plotted on the horizontal axis, and the three components of the corresponding solution error standard deviation are plotted on the vertical axis. The slope of the line formed by each data set indicates the amount by which measurement errors tend to be magnified in the solution.

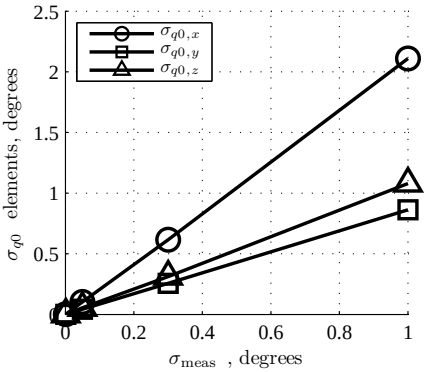

Figure 2.5: Standard deviations of attitude solution errors for major-axis spinner

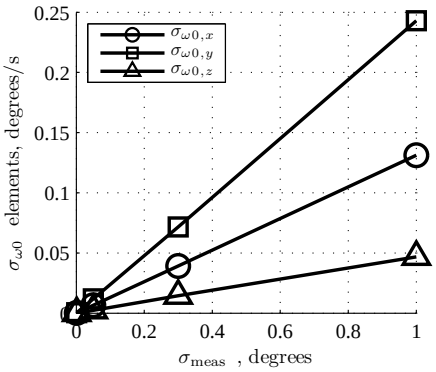

Figure 2.6: Standard deviations of angular velocity solution errors for major-axis spinner

Results for the minor-axis spinner are similar.

One other significant effect of measurement errors has been observed. If measurement errors are sufficiently large, the equation-solving procedure may fail to converge to a solution near the "truth" even when the "truth" solution is used as the initial guess. The severity of this effect depends not only on the physical scenario, but also on the particular set of inertial-frame measurements used and on the set of equations being solved. For instance, if the measurements at times t_1, t_2, and t_3 are used for the minor-axis spinner, as opposed to the measurements at times t_0, t_1, and t_2, the convergence rate for a given measurement error level is much higher. This fact appears to indicate improved problem observability when using measurement times t_1, t_2, and t_3. Furthermore, when initialized from the "truth" solution, the angular-momentum-variables method has a comparable or higher convergence rate in the presence of measurement errors than the attitude-variables method. This evidence suggests that the angular-momentum-variables method may be more robust in the presence of errors, despite the fact that the solutions it finds are no better in cases where both methods converge.

2.4.3 Further Discussion of Results

In addition to the quantitative results in the preceding sections, several general concepts have emerged from the various analyses. The first of these ideas concerns the role of vector measurement distribution and independence. To maintain theoretical local observability, it is only necessary that at least two of the $\boldsymbol{r}_i$ reference-frame vectors be linearly independent. The remaining reference-frame vector or any of the $\boldsymbol{b}_i$ body-frame vectors may be identical, except in certain perverse physical configurations. In practice, however, it has been found that both the body- and reference-frame vectors should be as independent as possible. A good geometric distribution of measurements tends to improve convergence to solutions, and also reduces the standard deviation of solutions for a given measurement error level. Unfortunately, this effect is coupled with dynamics in a way that makes it difficult to determine in advance the optimal measurement set for a given physical configuration.

Another important finding relates to the relative merits of different classes of equation sets. Some systems of equations, such as the method in Section 2.3.3, employ unknowns that store attitude information. Other equation sets may describe the same dynamic motion in terms of variables that capture angular rate information; an example is the method of Section 2.3.3. Several of the methods investigated but not derived in detail for this chapter use a hybrid approach, where some variables contain attitude information and others describe angular rates. The solution procedure developed in Ref. [18] for the first generalized Wahba problem developed equations based on attitude variables, and the success of this analytical technique suggested that something similar might be appropriate for the second generalized problem treated in this chapter. This chap-

ter's results indicate otherwise in two ways. First, aliasing is more easily dealt with by the angular-momentum-variables method. It is simpler to impose external magnitude constraints on angular rates or related quantities than to modify the structure of the equations, and by adjusting external constraints the practitioner gains the ability to solve the equations in high-velocity regimes where aliasing is a possibility. Second, the angular-momentum-variables method requires a larger number of first guesses in order to find the true solution for the scenarios considered. Its faster execution time per guess, however, causes its overall execution time to be lower on average.

2.4.4 A Practical Implementation Strategy

The necessity of applying this chapter's algorithms to a large set of initial guesses raises concerns about computational burden and suggests that these methods may be best suited for offline calculations. An adaptation of the solution strategy presented here could, however, mitigate some of these problems. This adaptation exploits the changing observability of the dynamic scenarios. For a given set of three measurements and a given time history of spacecraft motion, 1000 initial guesses might be required to ensure that the true solution is obtained. At a later time, the same spacecraft might have a very large region of convergence, meaning that only 10 initial guesses are necessary because a high percentage of initial guesses converge to the true solution. Rather than applying this chapter's algorithms 1000 times for each set of three measurements, the practitioner could try only a small number of initial guesses, such as 100 or even 10. If one or more solutions were obtained for those initializations, they could be individually tested by propagating them to a fourth measurement time

and comparing to the measurement at that time. If the true solution were not obtained, the oldest measurement would be discarded and a new one added, thus creating a new set of measurements applying to a shifted set of times. By iterating on this procedure, eventually one would expect to encounter a measurement set with good observability properties, and less computation would be wasted on scenarios with poor observability. This modified solution strategy has been implemented for both sets of equations described in Section 2.3.3, using just 10 initial guesses per time window. In most cases the true solution was obtained within two or three sets of measurements, i.e. using only 20-30 initial guesses.

2.5 Conclusions

This chapter has focused on the solution of a restricted form of an extended Wahba's problem that seeks to estimate the initial attitude and angular rate of a spinning spacecraft. Because this attitude estimation approach yields globally optimal estimates, it cannot diverge. The restricted problem considers the case of a rigid spacecraft with axial symmetry, known inertia properties, and zero external torque. It uses three vector attitude measurements at three distinct times. Axial symmetry allows the quaternion state transition matrix to be written in a closed form involving only trigonometric functions. The problem is equivalent to solving six nonlinear equations in six unknowns. Algebraic techniques have been used to reduce the restricted problem to sets of three or more nonlinear equations in only three unknowns. The assumption that aliasing has not occurred provides limits on the reasonable solution space. The use of Newton's method restarted many times from many initial guesses in the solution space

usually returns all globally optimal attitude and angular rate solutions for the given measurements. Alternatively, one can restart Newton's method only a small number of times for a given set of measurements, and one can repeatedly slide the time window of measurements forward until this procedure yields a valid solution.

"Truth" model simulations for two dynamic scenarios support the usefulness of the proposed solution strategies. Numerical solution of two of the reduced nonlinear equation sets succeed in identifying the "truth" solutions, although with only three measurements it is not possible to rule out other physically feasible solutions. A fourth measurement resolves the ambiguity and makes the system globally observable. Measurement error effects have been shown to depend strongly on the specific physical configuration and set of vector measurements, but it is often possible to obtain a reasonably accurate solution in the presence of moderate errors. Although all sets of equations result in the same attitude solution, some were found to converge more frequently to solutions or more easily handle aliasing situations.

2.6 Appendix: Equivalence of Wahba Performance Metric and Derived Cost Function

The following section demonstrates the equivalence of the quaternion that maximizes the performance metric of Eq. (2.4b) and the quaternion that minimizes the quadratic cost defined by Eq. (2.50). This equivalence has been asserted in Section 2.3.3. Repeated here for convenience and written in terms of attitude $\boldsymbol{q}_1$ and angular momentum $\boldsymbol{h}_{sc1}$, the generalized q-method seeks the quaternion $\boldsymbol{q}_1$

that maximizes

$$\bar{p}\left(\boldsymbol{q}_1, \boldsymbol{h}_{sc1}\right) = \boldsymbol{q}_1^{\mathrm{T}} \left[\sum_{i=0}^{2} \Phi^{\mathrm{T}}(t_i - t_1; \boldsymbol{h}_{sc1})\, K_i \Phi\left(t_i - t_1; \boldsymbol{h}_{sc1}\right) \right] \boldsymbol{q}_1 \tag{2.53}$$

Although it uses very different notation from that employed here, Ref. [22] shows that each q-method K_i matrix can be written in terms of its eigenvector quaternions as

$$K_i = \boldsymbol{q}_{\alpha i}\boldsymbol{q}_{\alpha i}^{\mathrm{T}} + \boldsymbol{q}_{\beta i}\boldsymbol{q}_{\beta i}^{\mathrm{T}} - \boldsymbol{q}_{ci}\boldsymbol{q}_{ci}^{\mathrm{T}} - \boldsymbol{q}_{di}\boldsymbol{q}_{di}^{\mathrm{T}} \tag{2.54}$$

Furthermore, the sum of the outer products of all these quaternions is just the identity matrix:

$$\boldsymbol{q}_{\alpha i}\boldsymbol{q}_{\alpha i}^{\mathrm{T}} + \boldsymbol{q}_{\beta i}\boldsymbol{q}_{\beta i}^{\mathrm{T}} + \boldsymbol{q}_{ci}\boldsymbol{q}_{ci}^{\mathrm{T}} + \boldsymbol{q}_{di}\boldsymbol{q}_{di}^{\mathrm{T}} = \mathrm{I}_4 \tag{2.55}$$

Combination of Eqs. (2.54) and (2.55) yields

$$K_i = \mathrm{I}_4 - 2\left(\boldsymbol{q}_{ci}\boldsymbol{q}_{ci}^{\mathrm{T}} + \boldsymbol{q}_{ci}\boldsymbol{q}_{ci}^{\mathrm{T}}\right) \tag{2.56}$$

Now substitute this new expression for K_i back into the performance metric of Eq. (2.53) and simplify slightly to get

$$\begin{aligned} \bar{p}\left(\boldsymbol{q}_1, \boldsymbol{h}_{sc1}\right) \;=\; \boldsymbol{q}_1^{\mathrm{T}} \Bigg[& \textstyle\sum_{i=0}^{2} \Phi^{\mathrm{T}}(t_i - t_1; \boldsymbol{h}_{sc1})\, \Phi\left(t_i - t_1; \boldsymbol{h}_{sc1}\right) \\ & -2 \textstyle\sum_{i=0}^{2} \Phi^{\mathrm{T}}(t_i - t_1; \boldsymbol{h}_{sc1}) \left(\boldsymbol{q}_{ci}\boldsymbol{q}_{ci}^{\mathrm{T}} + \boldsymbol{q}_{di}\boldsymbol{q}_{di}^{\mathrm{T}}\right) \Phi\left(t_i - t_1; \boldsymbol{h}_{sc1}\right) \Bigg] \boldsymbol{q}_1 \end{aligned} \tag{2.57}$$

The transpose of the state transition matrix Φ represents the inverse rotation, so the leftmost expression on the right-hand side of Eq. (2.57) is a sum of identity matrices and does not influence the optimal value of $\boldsymbol{q}_1$. As the second term on the right-hand side enters with a negative sign, the performance metric is maximized when this term is minimized. By inspection, the term to be minimized is a scaled version of the cost function defined by Eq. (2.50), thus proving the equivalence of the two criteria. For a closely related proof relating the q-method to the measurement update step of a Kalman filter that uses the quaternion as its state, see Ref. [25].

CHAPTER 3

ESTIMATION OF ATMOSPHERIC AND IONOSPHERIC EFFECTS IN MULTI-SATELLITE ORBIT DETERMINATION USING CROSSLINKS

A method is proposed for determining the orbits of multiple satellites in the same orbital plane by employing both downlink and crosslink radio-navigation signals. Such an orbit determination scheme must account for orbit perturbations due to unpredictable atmospheric drag and for ionospheric delay effects on single-frequency crosslink ranging measurements. The proposed method can improve orbit accuracy and sense variations in the space environment by simultaneously estimating satellite orbits and density distributions for the atmosphere and ionosphere. Satellites in the same orbital plane are expected to encounter roughly the same density features after a time delay corresponding to orbital separation, and this implies that errors due to both atmospheric drag and ionospheric delay may be spatially correlated. By exploiting rather than ignoring this correlation, it is hoped that orbit improvements associated with a given satellite will also benefit the orbit estimates for other satellites in the same plane. A linearized observability analysis indicates that the proposed estimation system is observable without *a priori* information. A specialized form of Consider covariance analysis shows that estimating atmospheric and ionospheric density rather than applying standard models significantly improves orbit determination accuracy.

3.1 Introduction

Accurate estimates of satellite trajectories, such as might be required for satellite navigation or scientific purposes, are obtained by the process of orbit determination. Satellite orbit determination requires accurate models, both of satellite dynamics and of the available measurements. In both areas, the greatest uncertainties often arise from unpredictable environmental effects. Several possible strategies for dealing with such model uncertainty are illustrated in this chapter.

This chapter focuses on the case of multiple satellites in the same Low-Earth Orbit (LEO) plane, such as might constitute a subset of a global constellation. The target satellite system's configuration is based on the Iridium satellite cell phone constellation. It employs radiocommunication signals in the form of crosslink ranging signals between adjacent satellites and downlink ranging measurements. Dynamic uncertainties are introduced by atmospheric drag, which directly perturbs the satellite orbits. Another source of systematic uncertainty is ionospheric delay on the single-frequency crosslink measurements. The density distributions of the upper atmosphere and ionosphere are highly variable and only imperfectly understood, so the resulting orbit determination errors may be significant. In the proposed configuration, however, each satellite would traverse a very similar orbit trajectory after some time delay, and the resulting correlated environmental perturbations might be exploited to mitigate errors in the overall orbit determination problem.

Atmospheric drag perturbations to satellite orbits have been widely studied, and a number of different strategies exist to cope with them. First, several different empirical models, largely based on past satellite orbit data, can predict

atmospheric density distributions [26–29]. Typical model standard deviations, however, are about 10-20% under normal conditions, and may be much higher depending on solar activity levels [30, 31]. A second approach employs highly accurate measurements or targeted sensors to directly remove non-gravitational perturbations; examples include spaceborne GPS receivers and high-quality accelerometers [32–35]. While these techniques are effective, they require expensive instrumentation and may not be available for preexisting satellites. A final method, known as atmospheric calibration, attempts to estimate atmospheric density parameters or corrections along with the orbits of a set of tracked satellites [36–38]. Calibration strategies have shown promise in reducing orbit determination fit errors for satellites not included in the calibration set, but there is still room for improvement.

Ionospheric delay is a function of signal frequency and of the total electron content (TEC) along the signal path. The proposed system's downlink signals traverse a relatively short path through the ionosphere, on the order of several hundred kilometers. It is anticipated that the delay on these signals can be adequately modeled and removed, for instance by a local ionosphere model that uses ground station dual-frequency GPS data as inputs [39]. This chapter focuses instead on crosslink ionospheric delay. Crosslinks can span a distance of thousands of kilometers through relatively dense regions of the ionosphere, and even for Ka-band signals the resulting measurement errors can be significant. Previous work addressing ionospheric delay on crosslink signals has been minimal. Some existing and proposed satellite constellations that employ crosslinks avoid the issue of ionospheric delay altogether. For instance, satellite altitudes may be sufficient that crosslinks never enter the dense region of the ionosphere, the path length or signal frequency may result in a delay

too short to matter, or the crosslinks may be exclusively used for data transfer rather than ranging [40–42]. Other satellite systems are able to directly measure and remove ionospheric effects by relating signals at two or more frequencies [43, 44]. One proposed technique for single-frequency crosslinks involves an *a priori* crosslink delay model based on the International Reference Ionosphere (IRI) density model [45, 46]. This model does not get updated based on actual crosslink data.

This chapter proposes a calibration-like scheme that simultaneously estimates all the satellite orbits and also local density distributions for the upper atmosphere and the ionosphere. Given the complexity of the driving physical processes, however, it is unclear what paramaterizations or dynamic models are suitable for estimating the environmental effects. An overly complex model for the density distributions would likely render the entire system unobservable, because too few measurements would be available to estimate all the model parameters. On the other hand, an observable model with poor correspondence to the true satellite environment could decrease orbit determination accuracy rather than improving it. To investigate these trade-offs, two related analyses are performed.

In its observability analysis, this chapter extends a previous work by the author on a similar orbit determination problem [47]. In Ref. [47], the satellites of interest comprised an entire constellation of six orbital planes. A calibration-like scheme was proposed to simultaneously estimate all the satellite orbits and the parameters of a global atmospheric density model, and the system was shown to be observable given a reasonable amount of *a priori* knowledge. This analysis, however, failed to account for ionospheric effects on crosslink measurements

because these effects had originally been assumed negligible. It was also discovered that the global atmospheric density parameterization was only weakly observable, in part because the measurements did not provide sufficient information about density between the orbital planes. The present chapter seeks to remedy these flaws in the observability analysis by incorporating a model of ionospheric density and its effect on satellite signals into the estimation scheme, and by restricting the analysis to a single plane of orbiting satellites.

In addition to the observability analysis, this chapter applies a special form of Consider covariance analysis to the proposed orbit determination problem. The Consider anlaysis computes the true estimation error covariance for a hypothetical filter that neglects to estimate the environmental parameters and uses *a priori* density distributions instead. By performing these calculations, this chapter seeks to determine whether the proposed orbit determination/ calibration scheme could significantly improve the satellite orbit estimates as compared to a simpler alternative.

The specific form of Consider covariance analysis used in this chapter is adapted from Ref. [3]. Although this form applies to a more general set of estimation model errors than a traditional Consider analysis, it is still relatively limited in possible applications. In particular, it is unable to deal with arbitrary dynamics in the atmospheric and ionospheric density distributions, or with the uncertainty in smoothed estimates. This orbit determination scenario, therefore, motivates the development of the multipurpose Consider covariance analysis techniques of Chapters 4 and 5.

This chapter offers three main contributions. First, it develops parameterizations for atmospheric and ionospheric density and models of how these den-

sity distributions relate to satellite drag and crosslink signal delay, respectively. These parameterizations are similar in some respects to the global spline-based atmospheric density parameterization developed in Ref. [47]. Second, a linearized observability analysis is performed to determine whether satellite orbits, atmospheric density, and ionospheric density are simultaneously observable for the single-orbital-plane system. Finally, a generalized Consider covariance analysis examines the significance to the orbit solution of estimating time-varying environmental parameters rather than assuming *a priori* models for the environmental effects.

The remainder of this chapter is organized as follows: The next section develops parameterizations for atmospheric and ionospheric density, and it models the effects of these densities on the satellite orbits and signals. Following this, Section 3.3 gives a brief review of the dynamics and measurement models for the estimation problem of Ref. [47], along with minor modifications for the present system. In Section 3.4, some observability analysis concepts are revisited, and the basic ideas of the Consider covariance analysis are presented. Next, observability and Consider covariance results are given and discussed in Section 3.5, followed by a summary of the chapter's findings and some conclusions in Section 3.6.

3.2 Atmospheric and Ionospheric Density Parameterizations

In order to estimate atmospheric and ionospheric density, appropriate parameterizations must be selected. The chosen parameterizations must be capable of representing the effects of realistic atmospheric drag and ionospheric delay.

These representations should also be confined to the region of interest within the orbital plane, to improve the chance that the measurements provide sufficient information to estimate the distributions. As a consequence of these criteria, this chapter's density distributions are similar to the global spline-based parameterization of Ref. [47], but not identical.

As in that previous work, the chosen parameterizations for atmospheric and ionospheric density both take exponential forms:

$$\rho_{atm}(r, \theta; \boldsymbol{p}_{atm}) = \rho_{0,atm}(\theta; \boldsymbol{p}_{atm}) \exp\left(-\frac{r - r_{0,atm}}{H_{atm}(\theta; \boldsymbol{p}_{atm})}\right) \tag{3.1a}$$

$$\rho_{ion}(r, \theta; \boldsymbol{p}_{ion}) = \rho_{0,ion}(\theta; \boldsymbol{p}_{ion}) \exp\left(-\frac{r - r_{0,ion}}{H_{ion}(\theta; \boldsymbol{p}_{ion})}\right) \tag{3.1b}$$

In these equations, the inputs are a position within the orbital plane in polar coordinates, r and θ, and the two parameter vectors $\boldsymbol{p}_{atm}$ and $\boldsymbol{p}_{ion}$. Note that the orbital plane is nearly fixed in inertial coordinates over short time intervals, so that diurnal variations in the density distributions at any given (θ, r) point will be minimal. The angle θ is measured relative to the ascending node. At each angle, the density distributions have values $\rho_{0,atm}$ and $\rho_{0,ion}$ at the two nominal radii $r_{0,atm}$ and $r_{0,ion}$, respectively. Density varies exponentially above and below these radii with respective scale heights H_{atm} and H_{ion}. Both the nominal densities and the scale heights are Fourier series functions of the angle θ, with series coefficients given by the parameter vectors $\boldsymbol{p}_{atm}$ and $\boldsymbol{p}_{ion}$. The choice of Fourier series functions rather than a cubic spline is a departure from the previous work, and it was motivated by the convenience of this representation within an orbital plane.

The exponential form is not a good general distribution for atmospheric density. It is, however, a reasonable model over the small range of altitudes encountered by the satellites in their nearly-circular orbits [47]. The choice of an expo-

nential distribution for ionospheric density is harder to justify for two reasons. First, electron density within the ionosphere is known to *not* decrease exponentially below and near its peak-density altitude. Even though the candidate satellite crosslinks are expected to pass mostly through the upper regions of the ionosphere where electron density is decreasing with altitude, these regions are poorly understood, and the decrease may not be exponential in nature. Second, the crosslinks for the candidate constellation pass through a 300-km range of altitudes, much larger than the range of altitudes encountered by the individual satellite orbits. Over such a large range, it is unlikely that a simple function like an exponential would fully capture the density behavior. Nevertheless, it is believed that the exponential model of ionospheric density is adequate for the purposes of this analysis. The International Reference Ionosphere (IRI), one of the best ionospheric models currently available, predicts electron densities within the range of interest that are approximately exponential under normal solar conditions [46]. Furthermore, even if this chapter's ionospheric density model is inaccurate, it may still correspond to crosslink TEC values which are physically realistic. Improved orbit determination requires a better estimate of the delay affecting the measurements, not of the physical phenomenon which produces that delay.

Satellite acceleration due to atmospheric drag is related to atmospheric density by the equation

$$\boldsymbol{a}_d = -\frac{1}{2}\left(C_d \frac{A_v}{m_s}\right)\rho_{atm} v_r^2 \boldsymbol{e}_v \tag{3.2}$$

where C_d is a drag coefficient, A_v is the cross-sectional area of the satellite in the direction of travel, m_s is the total spacecraft mass, v_r is the velocity magnitude relative to the ambient atmosphere, and $\boldsymbol{e}_v$ is a unit vector in the relative velocity direction. Although the cross-sectional area may be time-varying, it is assumed

that such variations are well-understood and can be removed by applying an appropriate scale factor. The scalar product $(C_d A_v/m_s)$, known as the inverse ballistic coefficient, will be estimated in addition to ρ_{atm}.

While drag acceleration depends on atmospheric density only at the satellite location, ionospheric delay is a function of TEC, the total number of electrons along a 1-m^2 cross-section line-of-sight path traversed by the crosslink signal. Specifically,

$$\delta t_{ion} = \frac{40.3\text{TEC}}{c f_{CL}^2} \tag{3.3}$$

where δt_{ion} is the ionospheric delay on the crosslink signal in seconds, c is the speed of light in m/s, and f_{CL} is the crosslink signal frequency in Hz. The units of TEC are electrons/m^2. TEC is typically computed by numerical integration through a density profile, but such integration is computationally expensive and inconvenient for estimation. An alternative technique is to approximate the density profile along the crosslink by an analytically integrable function.

To determine such a function and create an analytical model of the relationship between crosslink delay and ionospheric density, consider the crosslink geometry. As the signal travels between two satellites at approximately the same altitude, the altitude of the line-of-sight path decreases down to a minimum near the midpoint and then increases again. Electron density, on the other hand, is modeled as increasing exponentially with decreasing altitude. The resulting density profile versus crosslink distance has a shape similar to a Gaussian curve; it can be approximated well by an analytically integrable scaled hyperbolic secant function. The scaling of the hyperbolic secant is determined by evaluating the density model in the form of Eq. (3.1b) at just the endpoints and midpoint of the crosslink and stretching the hyperbolic secant to closely fit those points.

In this way, crosslink ionospheric delay can be determined by evaluating ρ_{ion} at just three points, rather than at the tens or hundreds of points required for numerical integration.

Both $\boldsymbol{p}_{atm}$ and $\boldsymbol{p}_{ion}$ can be initialized so as to best fit some arbitrary reference set of density data in a least-squares sense, as was done for atmospheric density in Ref. [47]. Such a reference set could be provided by a standard atmospheric or ionospheric density model, and the initialization would provide a physically realistic first guess for estimation purposes. This approach was found to perform well when initializing $\boldsymbol{p}_{atm}$ with data from the NRLMSISE-00 model [29]. In the case of $\boldsymbol{p}_{ion}$, however, fitting reference values of ρ_{ion} is less important than fitting reference values of δt_{ion}. To obtain such values, a pair of satellites were first simulated over the course of an orbit to obtain a history of the crosslink endpoints. The IRI-2007 model was then evaluated at many points along each crosslink for each simulated time step, and numerical integration was used to generate TEC. Ionospheric delay was obtained by evaluating Eq. (3.3). The parameter vector was then initialized by performing a least-squares fit to determine the $\boldsymbol{p}_{ion}$ that caused the Fourier series model and hyperbolic secant technique to most closely reproduce the IRI-model reference delays δt_{ion}. The residuals of this fit are shown in Figure 3.1, converted to equivalent centimeters of fit error. Figure 3.1 also displays ionospheric delay residuals for an independent satellite pair not used in the least-squares fit, but with the same reference IRI electron density distribution. For this particular initialization, $\boldsymbol{p}_{ion}$ was chosen to contain coefficients for the first ten Fourier frequencies. The parameter vector $\boldsymbol{p}_{ion}$ and the hyperbolic secant TEC calculation were able to fit the reference ionospheric delay values to better than 1 mm for the satellite crosslink employed in the initialization, and to better than 6 mm for the independent satellite crosslink. The

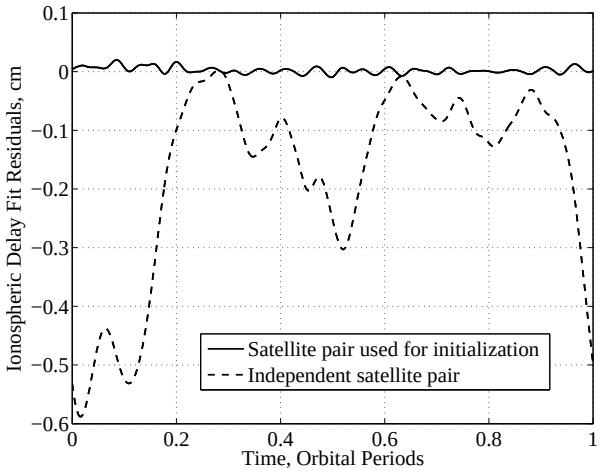

Figure 3.1: Residual errors of least-squares fit of ionospheric parameter vector and model to generated ionospheric delay data.

apparent bias in the residuals for the second pair of satellites is likely related to orbit differences. For instance, the two satellite pairs can have crosslinks with different mean minimum altitudes, such that the fit to the ionospheric model is optimized for a slightly different region of the ionosphere.

3.3 System Dynamics and Measurements

This chapter's dynamics and measurement models are essentially unchanged from those of Ref. [47], with the exceptions of the modified atmospheric parameterization and the new ionospheric states. Some of the notation has been updated, however, particularly to facilitate the Consider covariance analysis. This section provides a brief overview of the system models; for greater detail, see Ref. [47].

3.3.1 State Vector and Dynamics

The state vector for the proposed estimation scheme contains components related to satellites, measurements, and environmental effects. It is convenient to subdivide it as

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{p} \end{bmatrix} \tag{3.4}$$

where $\boldsymbol{x}$ contains all the states for all of the individual satellites and all measurement biases, and $\boldsymbol{p}$ contains the parameters related to the atmospheric and ionospheric density distributions. In its general discretized form, the nonlinear difference equation that describes the dynamics of the state vector can be written as

$$\begin{bmatrix} \boldsymbol{x}_{k+1} \\ \boldsymbol{p}_{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}_x\left(k; \boldsymbol{x}_k, \boldsymbol{p}_k, \boldsymbol{w}_{xk}\right) \\ \boldsymbol{f}_p\left(k; \boldsymbol{p}_k, \boldsymbol{w}_{pk}\right) \end{bmatrix} \tag{3.5}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{p}_k$ are the values of $\boldsymbol{x}$ and $\boldsymbol{p}$ at the k^{th} sample time t_k, and $\boldsymbol{w}_{xk}$ and $\boldsymbol{w}_{pk}$ are the associated discrete-time process noise vectors. Note that the dynamics of the satellite states and measurement biases contained in $\boldsymbol{x}$ may depend on the parameters in $\boldsymbol{p}$, but not vice versa. The difference equations $\boldsymbol{f}_x$ and $\boldsymbol{f}_p$ are, in practice, computed by simultaneous numerical integration of the differential equations for each of the subcomponents of $\boldsymbol{x}$ and $\boldsymbol{p}$.

The first part of the state vector, $\boldsymbol{x}$, can be further subdivided into states associated with individual satellites, crosslink measurement biases, and downlink measurement biases.

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_s^1 \\ \boldsymbol{x}_s^2 \\ \vdots \\ \boldsymbol{x}_s^M \\ \boldsymbol{b}_{cl} \\ \boldsymbol{b}_{gs} \end{bmatrix} \tag{3.6}$$

In this vector, the states specific to the j^{th} satellite are contained in $\boldsymbol{x}_s^j$, M is the number of satellites in the plane, and $\boldsymbol{b}_{cl}$ and $\boldsymbol{b}_{gs}$ contain the crosslink and downlink biases, respectively. The first six elements of each satellite's $\boldsymbol{x}_s^j$ are the Cartesian position $\boldsymbol{r}_s^j$ and velocity $\boldsymbol{v}_s^j$ in an inertial frame; their behavior is governed by Newton's second law. Force models include gravitational attraction from the Earth, Sun, and Moon, as well as atmospheric drag and solar radiation pressure. Horizontal wind disturbances are not included in the present model. Other components of $\boldsymbol{x}_s^j$ include an inverse ballistic coefficient and solar radiation pressure coefficient, each modeled as a slowly-varying random walk, and two satellite clock states that parameterize the clock error using a standard model [48]. The two bias vectors $\boldsymbol{b}_{cl}$ and $\boldsymbol{b}_{gs}$ are both modeled as constants.

The combined parameter vector $\boldsymbol{p}$ contains both the atmospheric and the ionospheric density parameters:

$$\boldsymbol{p} = \begin{bmatrix} \boldsymbol{p}_{atm} \\ \boldsymbol{p}_{ion} \end{bmatrix} \tag{3.7}$$

In the absence of reliable physics-based models, the two parameter vectors are modeled as some nominal, possibly time-varying vectors $\overline{\boldsymbol{p}_{atm}}$ and $\overline{\boldsymbol{p}_{ion}}$, with dynamic, discrete first-order Markov process perturbations from those nominal vectors. The nominal parameter vectors could be obtained by an initialization using a standard model, as described in Section 3.2. The i^{th} Fourier series coefficients in $\boldsymbol{p}_{atm}$ and $\boldsymbol{p}_{ion}$ vary according to

$$\left[p_{atm,i}\left(k+1\right) - \overline{p_{atm,i}}\left(k+1\right)\right] = \mathrm{e}^{-\left(\frac{t_{k+1}-t_k}{\tau_{atm,i}}\right)}\left[p_{atm,i}\left(k\right) - \overline{p_{atm,i}}\left(k\right)\right] + \gamma_{atm,i} w_{atm,i}\left(k\right) \tag{3.8a}$$

$$\left[p_{ion,i}\left(k+1\right) - \overline{p_{ion,i}}\left(k+1\right)\right] = \mathrm{e}^{-\left(\frac{t_{k+1}-t_k}{\tau_{ion,i}}\right)}\left[p_{ion,i}\left(k\right) - \overline{p_{ion,i}}\left(k\right)\right] + \gamma_{ion,i} w_{ion,i}\left(k\right) \tag{3.8b}$$

The discrete process noises $w_{atm,i}$ and $w_{ion,i}$ are white with zero mean and unit

variance. Proper tuning of the Markov time constants $\tau_{atm,i}$ and $\tau_{ion,i}$ and the process noise influence elements $\gamma_{atm,i}$ and $\gamma_{ion,i}$ will allow these parameter perturbations to be estimated as non-zero quantities if the data dictate such values.

3.3.2 Measurement Models

Three types of measurements are available for the proposed orbit determination scheme. First, there are dual-one-way-ranging measurements between the adjacent satellites in the plane, called crosslinks. Each satellite of a pair measures a code-phase-based accumulated delta range relative to the other satellite, with a unique unknown code-phase bias and ionospheric delay in each direction. Next, there is a carrier-phase-based accumulated delta range between each ground station and every satellite it tracks at a given time, and this downlink measurement also includes an unknown bias. Finally, there are GPS-like pseudorange measurements between the ground stations and satellites. Although the pseudoranges are not biased, they are much noisier than the corresponding accumulated delta ranges.

Between transmitting satellite ℓ and receiving satellite j, the crosslink accumulated delta range is given by

$$\lambda\phi^{j\ell} = \sqrt{\left(\boldsymbol{r}_s^j - \boldsymbol{r}_s^\ell\right)^{\mathrm{T}}\left(\boldsymbol{r}_s^j - \boldsymbol{r}_s^\ell\right)} + c\left(\delta t_s^j - \delta t_s^\ell\right) + \lambda b^{j\ell} + c\delta t_{ion}^{j\ell} + \eta^{j\ell} \tag{3.9}$$

The first term on the right-hand side of Eq. (3.9) is the range between satellite ℓ at its time of transmission and satellite j at its time of reception. Likewise, the second term is the difference between the receiving and transmitting satellite clock errors δt_s^j and δt_s^ℓ at the times of reception and transmission, respectively. The remaining terms are the phase bias $b^{j\ell}$, the ionospheric delay $\delta t_{ion}^{j\ell}$, and the

measurement noise $\eta^{j\ell}$, all multiplied by appropriate factors to yield distance units in Eq. (3.9). The ionospheric delay $\delta t_{ion}^{j\ell}$ is, of course, a function of the ionospheric parameters in $\boldsymbol{p}_{ion}$. It enters Eq. (3.9) positively because $\phi^{j\ell}$ is a code phase measured based on group delay rather than a carrier phase. Delays due to neutral atmosphere are negligible even at the minimum crosslink altitudes. Note that the crosslink measurement in the opposite direction, from satellite j to satellite ℓ, is not identical to this one due to the unique biases and the different times of transmission and reception.

The downlink accumulated delta range measurements are given by a very similar equation. From satellite ℓ to ground station j, the beat carrier phase is

$$\lambda\phi_j^\ell = \sqrt{(\boldsymbol{r}_{gj} - \boldsymbol{r}_s^\ell)^{\mathrm{T}}(\boldsymbol{r}_{gj} - \boldsymbol{r}_s^\ell)} - c\delta t_s^\ell + \lambda b_j^\ell + \eta_j^\ell \tag{3.10}$$

In Eq. (3.10), $\boldsymbol{r}_{gj}$ is the Cartesian position of ground station j at the time of reception, b_j^ℓ is the downlink bias, and η_j^ℓ is the measurement noise on the downlink. Note that the reference station clocks are assumed to be disciplined by GPS, so that their errors are negligible. The pseudorange downlink measurement is almost identical, except with no bias and different noise:

$$P_j^\ell = \sqrt{(\boldsymbol{r}_{gj} - \boldsymbol{r}_s^\ell)^{\mathrm{T}}(\boldsymbol{r}_{gj} - \boldsymbol{r}_s^\ell)} - c\delta t_s^\ell + \mu_j^\ell \tag{3.11}$$

Although both downlink equations would include effects from ionospheric and neutral atmosphere delay, it is assumed for this chapter's analyses that these components can be modeled well enough to be corrected and removed from the ground-based measurements [39].

While these measurement models are relatively simple to define, they are more difficult to evaluate and use. The satellite positions and clock errors are only conveniently available at discrete times corresponding to the endpoints of

the numerical integration intervals, whereas these equations require these states at intermediate, implicitly-defined times. Readers interested in the subtleties of these implicit calculations are encouraged to refer to Refs. [47] and [49]. For this chapter's analyses, it is recognized that the implicitly defined satellite states at the times of transmission and reception can be written as nonlinear functions of the satellite states at the endpoints of the integration interval during which the measurement arrived. That is, each measurement model takes the generic form

$$y_{k+1} = h\left(k; \boldsymbol{x}_k, \boldsymbol{x}_{k+1}, \boldsymbol{p}_k\right) + \nu_{k+1} \tag{3.12}$$

where the dependence on the parameters in $\boldsymbol{p}_k$ only applies for the crosslink measurements, and ν_{k+1} is any of the measurement noise components defined above. Note that the parameter states are assumed to be constant over the interval $[t_k, t_{k+1})$, so that crosslink measurements depend only on $\boldsymbol{p}_k$ and not on $\boldsymbol{p}_{k+1}$.

All the crosslink accumulated delta range measurements and downlink accumulated delta range and pseudorange measurements that arrived during the interval $[t_k, t_{k+1})$ are stacked together as one nonlinear vector measurement equation:

$$\boldsymbol{y}_{k+1} = \boldsymbol{h}\left(k; \boldsymbol{x}_k, \boldsymbol{x}_{k+1}, \boldsymbol{p}_k\right) + \boldsymbol{\nu}_{k+1} \tag{3.13}$$

This is the general, but non-standard, form that will be used in this chapter's observability and Consider covariance analyses.

3.3.3 Linearization

The two analyses described in the next section both operate on linearizations of this section's dynamics and measurement models. The linearization is about a

"truth" state history, in a manner analogous to an extended Kalman filter (EKF). The nominal satellite state trajectory is given by $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ and the environmental parameters are linearized about the nominal $\overline{\boldsymbol{p}_{atm}}$ and $\overline{\boldsymbol{p}_{ion}}$. Nominal process noise is zero.

The linearized dynamics equation used for this chapter's analyses, which is essentially a linearization of Eq. (3.5), is given by

$$\begin{bmatrix} \Delta\boldsymbol{x}_{k+1} \\ \Delta\boldsymbol{p}_{k+1} \end{bmatrix} = \begin{bmatrix} \Phi_{xk} & \Gamma_{xpk} \\ 0 & \Phi_{pk} \end{bmatrix} \begin{bmatrix} \Delta\boldsymbol{x}_k \\ \Delta\boldsymbol{p}_k \end{bmatrix} + \begin{bmatrix} \Gamma_{xk} & 0 \\ 0 & \Gamma_{pk} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_{xk} \\ \boldsymbol{w}_{pk} \end{bmatrix} \tag{3.14}$$

where the quantities $\Delta\boldsymbol{x}_k$ and $\Delta\boldsymbol{p}_k$ are perturbations from the nominal linearizing values of the states. The state transition matrix block Φ_{xk} is given by the partial derivative $\Phi_{xk} = \partial\boldsymbol{f}_x/\partial\boldsymbol{x}$, evaluated at the nominal linearizing values, and the block matrix Φ_{pk} is diagonal with elements $\mathrm{e}^{-(t_{k+1}-t_k)/\tau_{atm,i}}$ and $\mathrm{e}^{-(t_{k+1}-t_k)/\tau_{ion,i}}$ in accordance with Eq. (3.8). Likewise, $\Gamma_{xpk} = \partial\boldsymbol{f}_x/\partial\boldsymbol{p}$ and $\Gamma_{xk} = \partial\boldsymbol{f}_x/\partial\boldsymbol{w}_x$, both evaluated along the nominal state trajectory. Γ_{pk} is a diagonal matrix formed from the $\gamma_{atm,i}$ and $\gamma_{ion,i}$ elements of Eq. (3.8). When convenient, one can write Eq. (3.14) more compactly as the linearized dynamics equation for the total state $\boldsymbol{X}$ by making the appropriate block matrix and vector definitions.

$$\Delta\boldsymbol{X}_{k+1} = \Phi_k\Delta\boldsymbol{X}_k + \Gamma_k\boldsymbol{w}_k \tag{3.15}$$

The linearization of the measurement model in Eq. (3.13) gives

$$\Delta\boldsymbol{y}_{k+1} = \widehat{H}_{xk}\Delta\boldsymbol{x}_k + \overline{H}_{xk+1}\Delta\boldsymbol{x}_{k+1} + \widehat{H}_{pk}\Delta\boldsymbol{p}_k + \boldsymbol{\nu}_{k+1} \tag{3.16}$$

In this equation, $\Delta\boldsymbol{y}_{k+1}$ is the difference between $\boldsymbol{y}_{k+1}$ and $\boldsymbol{h}$ evaluated at the nominal values $\boldsymbol{x}_k$, $\boldsymbol{x}_{k+1}$, and $\overline{\boldsymbol{p}}_k$. The measurement matrices are formed by the partial derivatives $\widehat{H}_{xk} = \partial\boldsymbol{h}/\partial\boldsymbol{x}_k$, $\overline{H}_{xk+1} = \partial\boldsymbol{h}/\partial\boldsymbol{x}_{k+1}$, and $\widehat{H}_{pk} = \partial\boldsymbol{h}/\partial\boldsymbol{p}_k$, all evaluated at the nominal linearizing state values. Without loss of generality, this

linearized measurement equation is assumed to have been normalized such that its noise $\boldsymbol{\nu}_{k+1}$ has zero mean and identity covariance. Such a normalization is standard in many estimation techniques. As with the linearized dynamics, a condensed linearized measurement equation in terms of the total state vector $\boldsymbol{X}$ can be derived from Eq. (3.16) by defining the appropriate block matrices. It has the form

$$\Delta \boldsymbol{y}_{k+1} = \widehat{H}_k \Delta \boldsymbol{X}_k + \overline{H}_{k+1} \Delta \boldsymbol{X}_{k+1} + \boldsymbol{\nu}_{k+1} \tag{3.17}$$

Equations (3.14) and (3.16), and their condensed forms in Eqs. (3.15) and (3.17), form the starting point for the analyses that follow.

3.4 Observability and Consider Analysis Theory

3.4.1 Linearized Observability Analysis

Conceptually, a system is observable when it is possible to uniquely determine the initial state of the system from some finite number of measurements. It was shown in Ref. [47] that the full constellation system and atmospheric density model are at least weakly observable given reasonable *a priori* information. The present system has fewer satellite orbits to estimate and a lower-dimensional atmospheric model, but it has an additional ionospheric model that directly influences the crosslink measurements. Therefore it is worthwhile to reformulate the analysis for the system comprising the satellite states, atmospheric model parameters, and ionospheric model parameters, and determine whether all these states are simultaneously observable in some finite time.

Asking whether a system is observable is equivalent to asking whether a cost

function based on measurement error residuals has a unique global minimum. Many nonlinear systems have a distinct global minimum, but may also have additional distinct local minima. This chapter examines the local uniqueness, i.e. the isolation, of the global minimum at the true state. It does not address the question of whether other distinct local minima exist. This type of analysis is called a *linearized* observability analysis, and it operates on the linearized dynamics and measurement models obtained in Section 3.3.

In Ref. [47], the author demonstrated that the system defined by the linearized Eqs. (3.14) and (3.16) can be transformed into a large batch measurement equation of the form

$$\Delta \boldsymbol{y}_{batch} = \mathcal{H} \begin{bmatrix} \Delta \boldsymbol{x}_N \\ \Delta \boldsymbol{p}_N \end{bmatrix} + \boldsymbol{\nu}_{batch} \tag{3.18}$$

In this equation, N linearized measurement vectors have been stacked together. The system dynamics have been employed to write all these measurements in terms of their dependence on the states at the end of the batch, $\Delta \boldsymbol{x}_N$ and $\Delta \boldsymbol{p}_N$. Finally, the batch equation has been transformed so that the noise $\boldsymbol{\nu}_{batch}$ has zero mean and identity covariance. All these algebraic manipulations and scaling operations are reflected in the contents of the large batch measurement matrix $\mathcal{H}$. The least-squares solution for this batch problem is

$$\begin{bmatrix} \Delta \boldsymbol{x}_N \\ \Delta \boldsymbol{p}_N \end{bmatrix} = \left(\mathcal{H}^{\mathrm{T}} \mathcal{H} \right)^{-1} \mathcal{H}^{\mathrm{T}} \Delta \boldsymbol{y}_{batch} \tag{3.19}$$

Thus, the system is said to be locally observable if the matrix $\mathcal{H}^{\mathrm{T}}\mathcal{H}$ is invertible.

This type of batch observability analysis is equivalent to a properly formulated square-root information filter (SRIF) that performs only the covariance part of the filtering calculations. The SRIF form of the analysis is preferable because it does not require one to form the unwieldy batch matrix $\mathcal{H}$, and it al-

lows for a possible incorporation of process noise. It also lends itself readily to adaptation for the Consider covariance analysis calculations, as will be shown.

For a linear system, it is possible to completely decouple the covariance portion of the SRIF from state estimation, so that no estimates need actually be formed. A nonlinear system such as this one, however, requires some accurate states about which to linearize so as to compute the appropriate linearized system matrices. "Truth" model states rather than estimates are used for linearization in this chapter's analysis, similar to what was done in Ref. [47].

The SRIF stores both the state estimate and the estimation error covariance in a square-root information equation which takes the form

$$\begin{bmatrix} \widehat{\mathcal{R}}_{xxk} & \widehat{\mathcal{R}}_{xpk} \\ 0 & \widehat{\mathcal{R}}_{ppk} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{x}_k \\ \Delta \boldsymbol{p}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}_{xk} \\ \boldsymbol{z}_{pk} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\nu}_{xk} \\ \boldsymbol{\nu}_{pk} \end{bmatrix} \tag{3.20}$$

This equation can also be written more compactly as

$$\widehat{\mathcal{R}}_{XXk} \Delta \boldsymbol{X}_k = \boldsymbol{z}_{Xk} - \boldsymbol{\nu}_{Xk} \tag{3.21}$$

The matrix $\widehat{\mathcal{R}}_{XXk}$, or its equivalent containing the $\widehat{\mathcal{R}}_{xxk}$, $\widehat{\mathcal{R}}_{xpk}$, and $\widehat{\mathcal{R}}_{ppk}$ blocks, is the square-root information matrix for the combined state $\Delta \boldsymbol{X}_k = \left[\Delta \boldsymbol{x}_k^{\mathrm{T}}, \Delta \boldsymbol{p}_k^{\mathrm{T}}\right]^{\mathrm{T}}$. The error between the estimate and the true state is characterized by the zero-mean, identity covariance white noise sequence $\boldsymbol{\nu}_{Xk}$. The quantity $\boldsymbol{z}_{Xk} = \left[\boldsymbol{z}_{xk}^{\mathrm{T}}, \boldsymbol{z}_{pk}^{\mathrm{T}}\right]^{\mathrm{T}}$ is called the information state, and it is related to the current state estimate by the equation

$$\Delta \widehat{\boldsymbol{X}}_k = \widehat{\mathcal{R}}_{XXk}^{-1} \boldsymbol{z}_{Xk} \tag{3.22}$$

Likewise, the square-root information matrix $\widehat{\mathcal{R}}_{XXk}$ is related to the estimation error covariance by

$$\widehat{P}_{XXk} = \mathrm{cov}\left(\Delta \widehat{\boldsymbol{X}}_k\right) = \widehat{\mathcal{R}}_{XXk}^{-1} \widehat{\mathcal{R}}_{XXk}^{-\mathrm{T}} \tag{3.23}$$

The process of square-root information filtering consists of recursively updating Eq. (3.21) so as to incorporate information from the dynamic propagation model and the new measurements. As this analysis is only concerned with the estimation error covariance and not with the state estimates themselves, only the square-root information matrix $\widehat{\mathcal{R}}_{XXk}$ must be updated. Also, while a traditional SRIF performs the dynamic propagation and measurement update steps in separate stages, this division is unnecessary. Both update stages can be performed simultaneously. For this form of the SRIF, the combined update is given by

$$Q_{k+1}\begin{bmatrix}\widehat{\mathcal{R}}_{wwk} & \widehat{\mathcal{R}}_{wXk+1}\\ 0 & \widehat{\mathcal{R}}_{XXk+1}\\ 0 & 0\end{bmatrix} = \begin{bmatrix}\mathcal{R}_{wwk} & 0\\ -\widehat{\mathcal{R}}_{XXk}\Phi_k^{-1}\Gamma_k & \widehat{\mathcal{R}}_{XXk}\Phi_k^{-1}\\ -\widehat{H}_k\Phi_k^{-1}\Gamma_k & \widehat{H}_k\Phi_k^{-1}+\overline{H}_{k+1}\end{bmatrix} \tag{3.24}$$

In this equation, the matrices Φ_k, Γ_k, $\widehat{H}_k$, and $\overline{H}_{k+1}$ are taken from the system's linearized dynamics and measurement models given by Eqs. (3.15) and (3.17). The matrix $\mathcal{R}_{wwk}$ is the inverse-square-root of the covariance of the process noise $\boldsymbol{w}_k$. The calculation proceeds by orthonormal/upper triangular (QR) factorization of the block matrix on the right, and it yields the updated square-root information matrix $\widehat{\mathcal{R}}_{XXk+1}$ as one of the sub-blocks of the upper triangular matrix on the left.

Observability after N measurements is determined by the matrix $\widehat{\mathcal{R}}_{XXN}$, which must be full-rank. Equivalently, $\widehat{\mathcal{R}}_{XXN}$ must be numerically invertible, and the covariance $\widehat{P}_{XXN} = \widehat{\mathcal{R}}_{XXN}^{-1}\widehat{\mathcal{R}}_{XXN}^{-\mathrm{T}}$ must have finite values on its diagonal, signifying a finite amount of uncertainty in each of the state estimates. The matrix $\widehat{P}_{XXN}$ is equivalent to the matrix $\left(\mathcal{H}^{\mathrm{T}}\mathcal{H}\right)^{-1}$ from the batch formulation of Eq. (3.19) if and only if the SRIF calculations are initiated with $\widehat{\mathcal{R}}_{XX0} = 0$. This is the condition of no *a priori* information, consistent with the idea of system observability based on measurement data alone. Note that the inclusion of pro-

cess noise in this observability analysis, although similar to what was done in Ref. [47], is non-standard. It has no impact on the rank of $\widehat{\mathcal{R}}_{XXN}$, but it does affect the uncertainty levels in $\widehat{P}_{XXN}$.

3.4.2 Consider Covariance Analysis

The goal of a Consider covariance analysis is to investigate the estimation error covariance of a filter that mismodels the system noise covariances or neglects to estimate some subset of a system's state vector. In the latter situation, some of the states are not part of the estimation process but their effects on estimation accuracy are nevertheless "considered" [3, 50–52]. Such a filter would tend to have larger estimation errors for the remaining (estimated) states, and its own estimate of its estimation error covariance would be incorrect and often optimistic. By performing a Consider analysis, one can determine how detrimental it may be to operate on a simplified system rather than the full system, especially in situations where the Considered states are only poorly understood or modeled.

The most common form of Consider covariance analysis restricts itself to the problem of unmodeled constant biases, such as might result from an imperfectly known ground station location or antenna phase center [51, 52]. This chapter's special form of Consider analysis is adapted from Ref. [3], which also allows one to Consider the effects of process or measurement noise with incorrect statistics, incorrect *a priori* state estimates, or unmodeled colored process noise. The colored process noise, in particular, is really just a first-order Markov process, so the same formulation is used here to Consider the effects of unmodeled atmo-

spheric and ionospheric perturbations. Analysis of a more complicated model for atmospheric or ionospheric density might require the more general Consider covariance analysis of Chapters 4 and 5.

Conceptually, this chapter's Consider analysis answers the question: To what degree will estimation accuracy be degraded if the proposed orbit determination scheme includes only *a priori* models of the atmospheric and ionospheric density distributions, and does not estimate the first-order Markov process perturbations? In other words, does this chapter's simultaneous estimation of atmospheric and ionospheric parameters along with satellite orbits have a significant impact on orbit accuracy, or could a simpler approach be just as good?

While Consider covariance analysis is a powerful technique, care must be taken not to misinterpret its significance. The analysis demonstrates the accuracy difference of orbit estimates with and without simultaneous estimation of environmental parameters, but it does so *assuming that the mathematical models of those environmental parameters are correct*. Thus the value of the Consider analysis will depend on the degree to which the real physical atmosphere and ionosphere can be described by the proposed models. It is hoped that this chapter's results will contribute to the qualitative understanding of the proposed estimation scheme, even if their quantitative significance is uncertain. A more complicated analysis could be performed using the algorithms of Chapters 4 and 5. For instance, one could examine the performance of an estimation scheme employing one set of environmental models in the presence of a different set of "true" models. Such an analysis is beyond the scope of the present chapter.

Following the approach of Ref. [3], this chapter's Consider covariance analysis takes a square-root form that closely resembles the linearized observability analysis. As before, it performs only covariance calculations and does not compute estimates. Because this algorithm focuses on the case of a filter that does not compute estimates of the dynamic parameters in $\Delta\boldsymbol{p}_k$, its information equation is just the subset of Eq. (3.20) pertaining to the information state vector $\boldsymbol{z}_{xk}$ (and neglecting the $\boldsymbol{z}_{pk}$ part):

$$\begin{aligned}\widehat{\mathcal{R}}_{xxk}\Delta\boldsymbol{x}_k &= \boldsymbol{z}_{xk} - \widehat{\mathcal{R}}_{xpk}\Delta\boldsymbol{p}_k - \boldsymbol{\nu}_{xk} \\ &= \boldsymbol{z}_{xk} - \boldsymbol{\eta}_k \end{aligned} \tag{3.25}$$

From the filter's perspective, the entire quantity $\left(\widehat{\mathcal{R}}_{xpk}\Delta\boldsymbol{p}_k + \boldsymbol{\nu}_{xk}\right)$, which has been defined as the error state $\boldsymbol{\eta}_k$, is treated as zero-mean white noise with identity covariance. The filter's covariance estimate will deviate from the true estimation error covariance because these statistics no longer hold. Instead, the covariance of $\boldsymbol{\eta}_k$ can be written as

$$\begin{aligned}\mathrm{E}\left[\boldsymbol{\eta}_k\boldsymbol{\eta}_k^{\mathrm{T}}\right] = \widehat{\mathcal{R}}_{xpk}\mathrm{E}\left[\Delta\boldsymbol{p}_k\Delta\boldsymbol{p}_k^{\mathrm{T}}\right]\widehat{\mathcal{R}}_{xpk}^{\mathrm{T}} + \mathrm{E}\left[\boldsymbol{\nu}_{xk}\boldsymbol{\nu}_{xk}^{\mathrm{T}}\right] \\ + \widehat{\mathcal{R}}_{xpk}\mathrm{E}\left[\Delta\boldsymbol{p}_k\boldsymbol{\nu}_{xk}^{\mathrm{T}}\right] + \mathrm{E}\left[\boldsymbol{\nu}_{xk}\Delta\boldsymbol{p}_k^{\mathrm{T}}\right]\widehat{\mathcal{R}}_{xpk}^{\mathrm{T}}\end{aligned} \tag{3.26}$$

The Markov process covariance $\mathrm{E}\left[\Delta\boldsymbol{p}_k\Delta\boldsymbol{p}_k^{\mathrm{T}}\right]$ is defined *a priori* by the specification of the time constants and process noise influence parameters in Eq. (3.8). Although $\boldsymbol{\nu}_{xk}$ may initially be zero-mean, identity-covariance, and uncorrelated with the environmental parameters, those properties may change as the filter continues to perform updates based on incorrect information.

The analysis proceeds by a series of QR-factorization updates of Eq. (3.25), very similar to the observability analysis update of Eq. (3.24). In addition to the square-root information matrix $\widehat{\mathcal{R}}_{xxk}$, however, it also propagates and updates

the various statistics of the information error $\boldsymbol{\eta}_k$, including both covariance and cross-covariance terms. For more details of the mechanization, the interested reader is encouraged to refer to Ref. [3].

State estimates for the hypothetical filter, neglecting the dynamic perturbations to the atmospheric and ionospheric models, would be computed as

$$\Delta\hat{\boldsymbol{x}}_k = \widehat{\mathcal{R}}_{xxk}^{-1}\boldsymbol{z}_{xk} \tag{3.27}$$

Likewise, the filter would compute the incorrect estimation error covariance

$$\widehat{P}_{fxxk} = \widehat{\mathcal{R}}_{xxk}^{-1}\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}} \tag{3.28}$$

In contrast, the "Consider covariance", which represents the true covariance for a filter with this type of mismodeling errors, is formulated as

$$\begin{aligned}\widehat{P}_{xxk} &= \operatorname{cov}\left(\Delta\hat{\boldsymbol{x}}_k\right) = \mathrm{E}\Big[\left(\Delta\hat{\boldsymbol{x}}_k - \Delta\boldsymbol{x}_k\right)\left(\Delta\hat{\boldsymbol{x}}_k - \Delta\boldsymbol{x}_k\right)^{\mathrm{T}}\Big] \\ &= \mathrm{E}\left[\left(\widehat{\mathcal{R}}_{xxk}^{-1}\boldsymbol{\eta}_k\right)\left(\widehat{\mathcal{R}}_{xxk}^{-1}\boldsymbol{\eta}_k\right)^{\mathrm{T}}\right] = \widehat{\mathcal{R}}_{xxk}^{-1}\mathrm{E}\big[\boldsymbol{\eta}_k\boldsymbol{\eta}_k^{\mathrm{T}}\big]\,\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}}\end{aligned} \tag{3.29}$$

where the covariance of $\boldsymbol{\eta}_k$ comes from Eq. (3.26).

3.5 Results

This chapter's results were obtained as follows. First, a "truth" model was used to generate a representative system state history. Next, the dynamics and measurement models were linearized around these "truth" states as in Eqs. (3.14) and (3.16). The observability analysis and Consider analysis were both performed for this same set of "truth" states, so that the results could be more easily compared. The SRIF covariance calculations for the observability analysis

yielded $\widehat{\mathcal{R}}_{XXN}$, the square-root information matrix for the combined state vector (including the parameters) after N measurements. Likewise, the SRIF-style Consider covariance calculations produced $\widehat{\mathcal{R}}_{xxN}$, the square-root information matrix for the hypothetical filter that neglects to estimate the environmental parameters, and also the true statistics of the information error state $\boldsymbol{\eta}_N$.

From $\widehat{\mathcal{R}}_{XXN}$, the square-root information matrix of the observability analysis, the estimation error covariance $\widehat{P}_{XXN}$ was determined as in Eq. (3.23). This matrix specifies the covariance of the orbit and parameter estimate errors for a filter that estimates both the satellite states and the environmental parameter variations. Likewise, the outputs of the Consider analysis were used in Eqs. (3.28) and (3.29) to compute $\widehat{P}_{fxxN}$ and $\widehat{P}_{xxN}$, respectively. The matrix $\widehat{P}_{fxxN}$ is the incorrect covariance computed by a filter that does not estimate the environmental parameter variations in the vector $\Delta\boldsymbol{p}_k$. In contrast, $\widehat{P}_{xxN}$ is the Consider covariance, the true estimation error covariance for such an incorrect filter. The square roots of the diagonal elements of each covariance matrix are estimation error standard deviations for individual state elements, which are convenient metrics for interpreting results.

3.5.1 "Truth" Model

This chapter's sub-constellation contains 11 satellites in a nearly polar orbital plane, each with a near-circular orbit and a mean orbit altitude of approximately 790 km. The satellites each have an inverse ballistic coefficient of about 0.0134 m^2/kg and a solar radiation pressure coefficient of about 0.0196 m^{2}kg, with some small variations between satellites. Each satellite clock has noise

typical of a good oven-controlled crystal oscillator (OCXO) under benign conditions [48].

The satellites are tracked by 12 globally-distributed ground reference stations, although at any given time only a few of those reference stations have one or more satellites in view. An elevation mask of 10° is used to determine visibility. All the measurements have sampling rates of 5 sec. The accumulated delta range downlink measurements have measurement error standard deviations of 0.3 mm, and the pseudorange downlink signals have measurement error standard deviations of 30 m. The crosslink accumulated delta range measurements have measurement error standard deviations of 3.4 cm. These accuracies would be unrealistic for individual measurements, but they actually represent the accuracy obtained by averaging less accurate, higher frequency data over 5-sec intervals.

Values for the nominal, linearizing parameter vectors $\overline{\boldsymbol{p}_{atm}}$ and $\overline{\boldsymbol{p}_{ion}}$ were obtained by fitting to generated data sets from standard models. The nominal atmospheric Fourier series coefficients were initialized with density data from the NRLMSISE-00 model under normal solar conditions. The IRI-2007 model was evaluated under moderate solar conditions along many simulated crosslinks to obtain simulated TEC data, which were in turn used to initialize the nominal ionospheric density coefficients. Each first-order Markov process had a modeled time constant of 3 hours. The intensity of the driving process noise for each Markov process was chosen so as to achieve an equivalent normalized atmospheric density standard deviation of approximately 40%, and an equivalent normalized ionospheric density standard deviation of nearly 100%.

The ground station locations, satellite clocks and dynamic parameters, and

signals are the same as those of Ref. [47], to facilitate comparison with that paper's results. Both the observability analysis and the Consider covariance analysis are initialized with zero *a priori* information. The standard simulated batch length is 36 hours, and each analysis takes about 3 hours to complete on a desktop computer.

3.5.2 Observability Results

The system was found to be observable with no *a priori* information. For some of the parameter states, however, it took nearly the full simulated 36 hours for the estimation error standard deviations to decrease to reasonable values. This behavior is characteristic of weak observability, and the inclusion of some *a priori* information would likely decrease the time to convergence significantly.

The estimation error standard deviations for a typical satellite's position history are shown in Figure 3.2. The standard deviations in this figure are resolved into the along-track/altitude/cross-track directions. Overall 1-σ position uncertainty levels decrease and eventually stay below about 3 cm. This high accuracy level is believed to result from the richness of the data set, and in particular the crosslink data. It is not expected that such accuracy could be achieved if a filter based on these dynamics and measurement models were applied to real data. Notably, these models do not yet incorporate the effects of the ionosphere or neutral atmosphere on the downlink measurements, and the gravity model for a real filter would require a higher degree and order. Many simplifications have been made in this analysis to facilitate the determination of overall system observability and to explore the significance of including an ionospheric delay

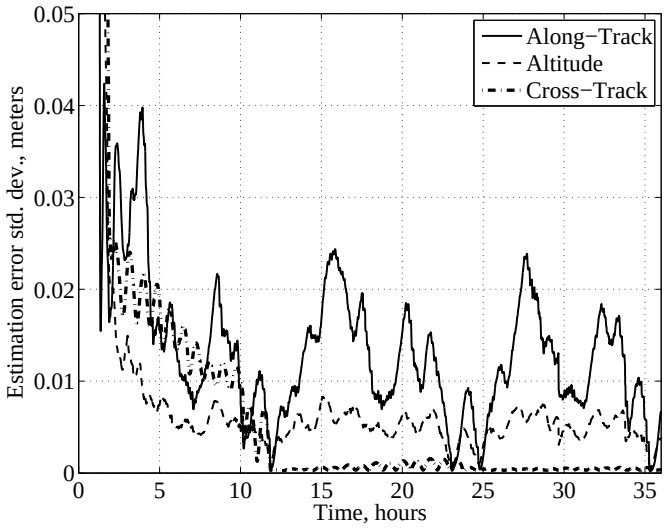

Figure 3.2: Estimation error standard deviations for a typical satellite's position states (along-track/altitude/cross-track directions).

model on the crosslink measurements. Some general observations can be made, however.

Among the three directions, the cross-track positions are slow to converge because there are fewer measurements in this direction. Once they do converge, however, they remain very accurate because there are also few perturbing forces normal to the satellite plane. In contrast, the standard deviations in the along-track direction have the most variation and they do not decay significantly after the first few hours. This is unsurprising since the atmospheric perturbations act in the along-track direction, and the ionospheric uncertainties effect the crosslink measurements, which are mostly in the along-track direction. The prominent 12-hour period in the standard deviations corresponds to the amount of time for the orbital plane to rotate relative to the Earth so that the satellites return to approximately the same ground tracks and are tracked by the same ground stations.

While the standard deviations of the atmospheric and ionospheric parameter states could be plotted as in Figure 3.2, such a plot would not be very meaningful. The uncertainty associated with individual Fourier series elements is less important than the corresponding uncertainty in atmospheric density, ρ_{atm}, and ionospheric density, ρ_{ion}. For this reason, the covariance of the $\Delta \boldsymbol{p}_{atm}$ and $\Delta \boldsymbol{p}_{ion}$ estimates have been transformed into standard deviations of density by employing the partial derivatives $\partial \rho_{atm} / \partial \Delta \boldsymbol{p}_{atm}$ and $\partial \rho_{ion} / \partial \Delta \boldsymbol{p}_{ion}$ in the equations

$$\sigma_{\rho atm} = \sqrt{\left(\frac{\partial \rho_{atm}}{\partial \Delta \boldsymbol{p}_{atm}} \right) \widehat{P}_{\boldsymbol{p}_{atm} \boldsymbol{p}_{atm}} \left(\frac{\partial \rho_{atm}}{\partial \Delta \boldsymbol{p}_{atm}} \right)^{\mathrm{T}}} \tag{3.30a}$$

$$\sigma_{\rho ion} = \sqrt{\left(\frac{\partial \rho_{ion}}{\partial \Delta \boldsymbol{p}_{ion}} \right) \widehat{P}_{\boldsymbol{p}_{ion} \boldsymbol{p}_{ion}} \left(\frac{\partial \rho_{ion}}{\partial \Delta \boldsymbol{p}_{ion}} \right)^{\mathrm{T}}} \tag{3.30b}$$

where $\sigma_{\rho atm}$ and $\sigma_{\rho ion}$ are the density estimation error standard deviations for the location at which the partial derivatives were evaluated, and $\widehat{P}_{\boldsymbol{p}_{atm} \boldsymbol{p}_{atm}}$ and $\widehat{P}_{\boldsymbol{p}_{ion} \boldsymbol{p}_{ion}}$ are the blocks of the covariance matrix $\widehat{P}_{XXN}$ corresponding to the elements of $\Delta \boldsymbol{p}_{atm}$ and $\Delta \boldsymbol{p}_{ion}$, respectively. These equations have been evaluated at a number of constant altitude points around the orbital plane, and normalized by the local densities ρ_{atm} and ρ_{ion} to give percent uncertainties. The normalized density standard deviations at 12-hour intervals are shown in Figures 3.3 and 3.4. Also shown in these figures are the density-equivalent nominal standard deviations of the first-order Markov processes for the atmospheric and ionospheric parameters. The significance of the nominal values is that, given no measurements to dictate otherwise, these are the levels to which the density uncertainty is expected to converge eventually. The fact that the density standard deviations eventually converge to levels below the nominal values for both the atmosphere and ionosphere indicates some degree of observability of the corresponding parameters. While both density distributions eventually reach similar uncertainty levels, ionospheric density does so much more rapidly. The atmo-

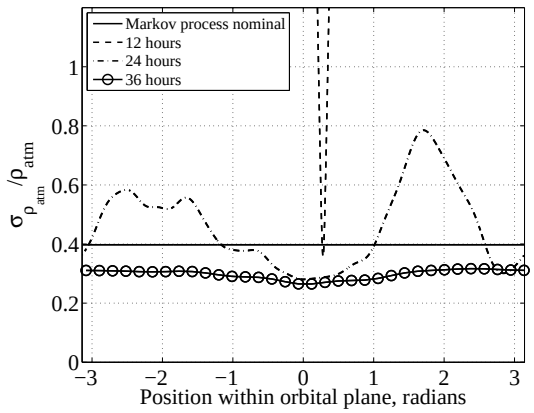

Figure 3.3: Atmospheric density normalized standard deviations.

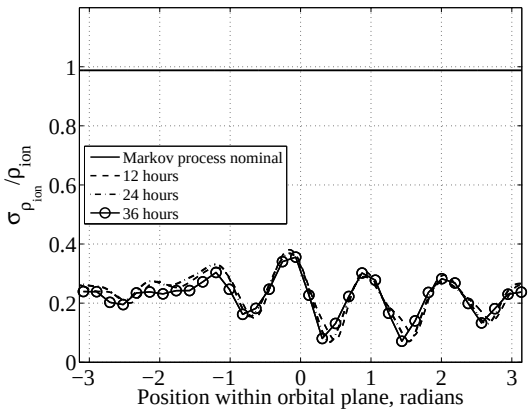

Figure 3.4: Ionospheric density normalized standard deviations.

spheric density distribution in Figure 3.3 requires nearly the full 36-hour simulation to converge to a normalized standard deviation of about 30%, whereas the ionospheric density in Figure 3.4 converges to normalized standard deviations between about 10% and 30% within the first 12 hours of simulated time. It is conjectured that this quick convergence is related to the high number of crosslink measurements, each of which is directly affected by ionospheric delay.

3.5.3 Consider Analysis Results

The Consider covariance analysis explores the importance of estimating dynamic variations in the two density distributions, rather than just applying *a priori* models. To interpret the results of this analysis, one must compare the standard deviations corresponding to the Consider covariance to the estimation error standard deviations of a system that estimates the environmental parameters. If the Consider standard deviations relating to the satellite orbit states are significantly greater, then neglecting to estimate the environmental parameters has a very detrimental effect on orbit determination accuracy. If, on the other

hand, both sets of standard deviations remain at similar levels throughout the analysis, it might be possible to safely remove estimation of the atmospheric and ionospheric distributions and apply a standard *a priori* model instead.

This chapter's Consider analysis results are fairly compelling. Figure 3.5 displays the position error standard deviations for a typical satellite, resolved onto the along-track/altitude/cross-track directions. These standard deviations rep-

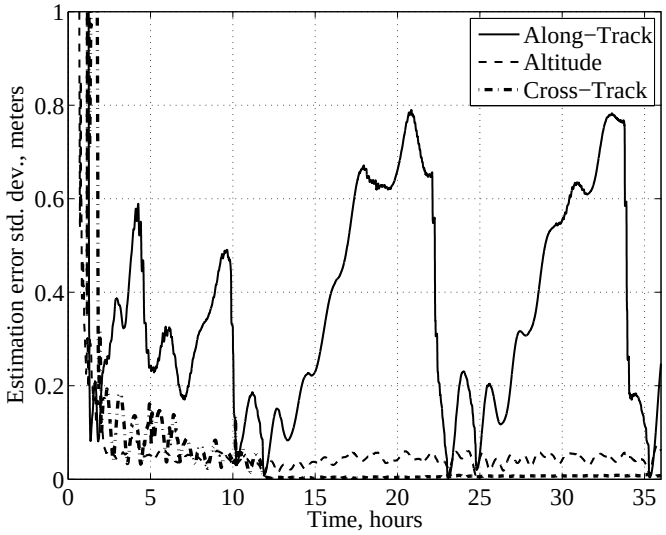

Figure 3.5: Position estimation error standard deviations corresponding to Consider covariance for a typical satellite

resent the level of estimation uncertainty that would result in a filter that uses only nominal *a priori* atmospheric and ionospheric distributions and does not estimate dynamic perturbations to those distributions. The uncertainties in Figure 3.5 can be compared directly to those of Figure 3.2, because the only differences between the two cases stem from the presence or absence of atmospheric and ionospheric parameters as estimated states. At first glance, the Consider standard deviations look only slightly higher than their counterparts in Figure 3.2. However, the vertical scales of the two figures differ by a factor of 20; the Consider standard deviations for the satellite positions are more than 20 times

greater, on average. These results indicate that simultaneous estimation of atmospheric and ionospheric variations along with the satellite orbits can greatly improve accuracy, at least for situations where the atmospheric and ionospheric density distributions can be adequately modeled as described. Note also that the improvements in orbit accuracy do not require a corresponding accuracy level in the parameter estimates. It is sufficient that the filter has a model of the types of perturbations that may occur and how they influence the system dynamics and measurements.

3.6 Conclusions

An orbit estimation scheme has been developed for a ring of satellites that are connected by crosslink ranging signals within a single orbital plane. The proposed method seeks to exploit the geographic correlation of dynamics and measurement perturbations that influence the satellite motion and ranging signals by simultaneously estimating density distributions within the plane for the atmosphere and ionosphere. It leverages a rich data set of radiocommunication signals and does not require any dedicated sensors for satellite tracking.

Two related types of analyses have been performed for the proposed system that consists of the satellites, signals, and density distributions for the atmosphere and ionosphere. First, a linearized observability analysis has sought to determine whether the given data and data rates provide enough information to estimate all the system states, even in the presence of the dynamically varying environmental effects. The proposed system is observable given a sufficiently long measurement batch, although standard deviations associated with

the satellite orbit states converge much more quickly and achieve greater accuracy levels than those associated with atmospheric and ionospheric parameters.

A Consider covariance analysis complements the observability analysis. Rather than asking if the system is observable, it explores the impact of estimating dynamically varying atmospheric and ionospheric parameters versus resorting to *a priori* models of their behavior. It demonstrates that the expected improvements in the system with estimated parameters are quite significant, provided the dynamic models for atmospheric and ionospheric density are sufficiently good approximations of the true distributions. These results suggest that the proposed simultaneous estimation scheme is worthwhile for orbit determination, even if the resulting estimates of atmospheric and ionospheric density retain large levels of uncertainty. Additionally, the usefulness of the Consider analysis for this somewhat complex scenario motivates the development of more generally applicable Consider covariance analysis tools.

CHAPTER 4

A MULTIPURPOSE SQUARE-ROOT CONSIDER COVARIANCE ANALYSIS FOR LINEAR FILTERS

A new form of Consider covariance analysis suitable for application to a wide variety of filtering problems is presented and demonstrated. A special system formulation is employed, and the analysis draws on the algorithms of square-root information filtering to provide generality and compactness. This analysis enables one to investigate the estimation errors that arise when a filter's dynamics model, measurement model, assumed statistics, or some combination of these is incorrect. Such an investigation can improve filter design or characterize an existing filter's true accuracy. Areas of application include incorrect initial state covariance; incorrect, colored, or correlated noise statistics; unestimated states; and erroneous system matrices. Several simple, yet practical, examples are developed, and the Consider analysis results for these examples are shown to agree closely with Monte Carlo simulations.

4.1 Introduction

The field of estimation addresses the question of how to form the best possible estimates of a system's state given the available, imperfect models of that system. For most applications, it is equally important to determine the statistical uncertainty of the estimates. The Kalman filter (KF) family of estimation schemes, including its various extended, square-root, and unscented formulations, generates both optimal state estimates and their error covariances, or an equivalent. These methods lend themselves naturally to the study of estimation

uncertainty by means of covariance analysis.

In covariance analysis, the covariance properties of the estimation error provide information about the filter and its accuracy. For example, one can examine the observability of particular subsets of states, or determine which measurements are most critical to filter performance. The covariance produced by a filter, however, is only a reliable metric of estimation error if the filter's model perfectly describes the true underlying system. When mismatch occurs between the filter model and the "truth" model, a more specialized alternative known as Consider covariance analysis can sometimes provide valuable insight.

A Consider covariance analysis allows one to "consider" the effects on estimation accuracy that result from certain types of mismatch between a filter's system model and the "truth" system. One can study the effects of model mismatch by computing the true covariance of the error between the filter's estimates and the "truth" states, which is distinct from the filter's reported covariance. The true covariances provide a metric for the accuracy of estimates produced by the mismodeled filter. In the context of this chapter and Chapter 5, the following definition for the phrase "Consider covariance analysis" is adopted: A Consider covariance analysis examines the errors of a filter that is based on an incorrect system model, specifically by *computing the "true" error covariance as dictated by an assumed "true" system model*. Note that the analysis is inherently hypothetical because it requires one to assume a specific "truth" model that may have errors of its own. The true covariances that the Consider covariance analysis computes are relative to that assumed "truth" model. They are true in the sense that *if* the given filter were applied to a system with the proposed "true" system model, then the errors in the resulting estimates

would have those covariances. While some Consider analysis methods compute true covariance directly, others produce some equivalent from which covariance may be calculated, such as the matrix difference between the true and filter-reported covariances. Some of this chapter's references prefer the name "sensitivity analysis", because such methods enable one to explore the sensitivity of a filter to various kinds of model error. Still others employ the generic name "covariance analysis" regardless of whether the filter model is assumed to match the true system model. Both alternate names refer to analyses that fall within this chapter's broad definition of Consider covariance analysis. Contrary to common practice, this chapter and Chapter 5 capitalize the word "Consider" in the covariance analysis context in order to more clearly distinguish between technical and colloquial usage.

Consider covariance analysis is a flexible tool that can be applied in different ways depending on the source of the filter model errors. Errors in filter models may occur unintentionally, especially in situations where the dynamics or disturbances are only poorly understood. They may also be intentional, as when a reduced-order filter or simplified equations are employed to mitigate computational burden. In the latter scenario, Consider covariance analysis is an important step in the design process that indicates whether the suboptimal filter performance is likely to be acceptable. Different candidate filters may be evaluated and compared. For example, the true system may have many constant random bias states which are known to be jointly unobservable. Estimation of these biases may be possible by assuming some reasonable level of *a priori* information, but the resulting bias estimates may be highly suspect. Consider covariance analysis can indicate which of these biases should be estimated as states in the filter so as to improve the estimates of the remaining quantities of

interest [3, 51]. The choice of biases to estimate is obtained by designing several candidate filters, each of which only estimate a subset of the biases, and comparing the true estimation error standard deviations associated with the states of interest for each filter. The analysis is not adversely affected if the bias states themselves remain poorly observable and the filter's estimates of those states do not converge to useful values. Estimation of poorly observable "nuisance" states can sometimes improve the estimation accuracy of other states of interest. Chapter 3 performs a similar analysis, except that it concerns weakly observable dynamic disturbances rather than biases.

When model errors result from high uncertainty in the dynamics, the practitioner typically knows that model errors are likely, but he or she does not know exactly what they are. Consequently, there is no single "truth" model standard with respect to which filters may be analyzed. A common application of Consider covariance analysis iteratively evaluates proposed filters, seeking one that is insensitive to changes in the uncertain system parameters. Such a filter is generally suboptimal for any particular realization of the "true" system, but maintains acceptable performance characteristics over the range of possible "true" systems [53, 54]. Similarly, an outer Monte Carlo analysis may be wrapped around a Consider covariance analysis to characterize the performance of a single proposed filter over some statistical distribution of hypothetical "truth" models. Alternatively, Consider analysis can compare the relative magnitude of different potential error effects, so as to develop an error budget or investigate discrepancies in solutions obtained from multiple sources [55, 56]. A final suggested application relates to multiple model filtering. This filtering approach assumes that the true system model is unknown and forms estimates as a weighted sum of the estimates corresponding to a set of candidate models.

The philosophical similarities between Consider covariance analysis and multiple model filtering hint at the possibility of using the former as a tool in the analysis or design of the latter. The exact form such an analysis would take is unclear.

Many previous Consider covariance analyses are narrow in scope, and address only a few particular classes of filter model errors. One of the simplest forms of Consider analysis focuses on filters with unmodeled constant random biases, particularly dynamics model parameter biases or measurement biases [51, 52, 57]. Another very simple analysis treats incorrect covariances of the process noise, measurement noise, or initial estimate [52, 58, 59]. Some less narrow efforts draw on both of these simple analyses to examine the effects of unmodeled coloring of the process noise or other unestimated dynamically-varying disturbance states [3, 50, 60]. This type of error is often encountered in the context of intentional model simplification and reduced-order filtering. Finally, a few references investigate the case where both the noise covariances and some or all of the state-space system matrices used by the filter are incorrect [4, 5, 61–63]. These sources do not explicitly allow for situations where some elements of the "truth" state vector are unestimated, although an experienced analyst could modify the algorithms to accommodate that case.

Algorithms that handle a subset of error classes can be useful, but a general approach that can simultaneously study all of these errors is preferable. Such general Consider covariance analyses are available in Refs. [54, 64, 65]. These algorithms, which are implemented in the covariance domain, augment the state vector of a traditional Kalman filter and perform several additional intermediate calculations. All of the previously-mentioned error classes can be

studied, including incorrect process and measurement noise covariances, incorrect initial estimate covariance, unestimated random biases or dynamics, and erroneous system matrices. Notably, none of the previous Consider covariance analyses is able to directly handle the situation of cross-correlated process and measurement noise. The single exception is Ref. [63], which suggests how such an analysis might be performed but does not derive the equations for that case. Many of the more general versions of Consider analysis could, however, incorporate the analysis of non-independent process and measurement noise with some additional pre-processing steps.

None of the general Consider analyses operate in the information domain or directly analyze a square-root information filter (SRIF). Square-root information filtering is theoretically equivalent to traditional Kalman filtering under the correct assumptions. The SRIF, however, is preferred in some applications that require good numerical properties or an infinite initial estimation error covariance. Of the aforementioned Consider covariance methods, Refs. [3, 50] are the most capable analyses in square-root information form. Neither of these works can study systems where the unestimated states have dynamics with a non-invertible state transition matrix, and they do not develop the algorithms for incorrect system matrices.

This chapter presents a new discrete square-root information form of generalized Consider covariance analysis. The following contributions are made: First, this chapter's analysis is the most general Consider covariance analysis for square-root information filters, and a single unifying framework encompasses filters with all the error classes discussed above. Second, the algorithms are demonstrated using simple numerical examples. These examples are se-

lected to illustrate the application of the new analysis to various types of filter errors. Consequently, the inexperienced practitioner does not have to derive non-obvious implementation steps independently. As an added benefit, the examples highlight some situations in which Consider covariance analysis may prove useful, such as when a mismodeled filter behaves in a counterintuitive way.

Note that this chapter restricts its focus to Consider *analysis*. Neither this chapter nor Chapter 5 addresses Consider *filtering*, which adjusts the filter estimates based on the Consider covariance calculations [48, 66, 67]. It is beyond the scope of this work to define or develop the various reasonable classes of Consider filters for this new framework.

Readers not familiar with the square-root information filter may struggle with this chapter's Consider covariance analysis, because of its heavy dependence on SRIF techniques. The generality and relative simplicity of this approach may outweigh such difficulties for many kinds of model error analyses. References [3] and [68] are recommended to the interested reader for gaining familiarity with SRIF methods.

The remainder of this chapter is organized as follows: Section 4.2 briefly reviews a simple example of Consider covariance analysis to illustrate the basic concept. Section 4.3 presents a special Consider form of the state-space system upon which this chapter's algorithms operate. In Section 4.4, the Consider analysis algorithms are developed. Section 4.5 demonstrates how some useful example problems can be translated into the Consider form of Section 4.3, and the Consider algorithms are validated for these examples by Monte Carlo analysis. Conclusions are drawn in Section 4.6.

4.2 Introductory Example

This section's simple example illustrates the principles of Consider covariance analysis in a more traditional form. The goal of this example is to help the reader to grasp the "big picture" of Consider analysis before moving on to more complicated systems and algorithms.

The estimation scenario of interest is a basic static system with a batch measurement equation given by

$$\boldsymbol{y} = H\boldsymbol{x} + \boldsymbol{\nu} \tag{4.1}$$

where $\boldsymbol{x}$ is a static vector that is to be estimated, $\boldsymbol{y}$ is a vector of measurements, H is the matrix relating the measurements to the unknown $\boldsymbol{x}$ vector, and $\boldsymbol{\nu}$ is the measurement noise. The measurement noise has been *incorrectly* assumed to have covariance $P_{\nu\nu a}$, whereas the "truth" measurement noise covariance is given by $\mathrm{E}\left[\boldsymbol{\nu}\boldsymbol{\nu}^{\mathrm{T}}\right] = P_{\nu\nu}$.

The standard weighted least-squares solution to the static batch problem of Eq. (4.1) is well-known; it takes the form

$$\hat{\boldsymbol{x}} = \left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1}H^{\mathrm{T}}P_{\nu\nu a}^{-1}\boldsymbol{y} \tag{4.2}$$

where the measurements have been weighted by the inverse of the assumed noise covariance $P_{\nu\nu a}$. The estimation error in this case is

$$\begin{aligned}(\hat{\boldsymbol{x}} - \boldsymbol{x}) &= \left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1}H^{\mathrm{T}}P_{\nu\nu a}^{-1}\boldsymbol{y} - \boldsymbol{x} \\ &= \left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1}H^{\mathrm{T}}P_{\nu\nu a}^{-1}\left(H\boldsymbol{x} + \boldsymbol{\nu}\right) - \boldsymbol{x} \\ &= \left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1}H^{\mathrm{T}}P_{\nu\nu a}^{-1}\boldsymbol{\nu}\end{aligned} \tag{4.3}$$

The expression for estimation error in Eq. (4.3) can be used to derive the estima-

tion error covariance, which should be small if $\hat{\boldsymbol{x}}$ is a good estimate of $\boldsymbol{x}$:

$$\mathrm{E}\Big[(\hat{\boldsymbol{x}}-\boldsymbol{x})(\hat{\boldsymbol{x}}-\boldsymbol{x})^{\mathrm{T}}\Big] = \left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1}H^{\mathrm{T}}P_{\nu\nu a}^{-1}\mathrm{E}\big[\boldsymbol{\nu}\boldsymbol{\nu}^{\mathrm{T}}\big]\,P_{\nu\nu a}^{-1}H\left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1} \quad (4.4)$$

In Eq. (4.4), it has been recognized that covariance matrices are symmetric so that there is no need to write $P_{\nu\nu a}^{-\mathrm{T}}$ in place of $P_{\nu\nu a}^{-1}$. Note, $(\,)^{-\mathrm{T}}$ refers to the transpose of the inverse of the matrix in question.

At this point, the distinction between covariance analysis and Consider covariance analysis can be demonstrated. A traditional covariance analysis would assume that the batch estimator's model of measurement noise covariance, $P_{\nu\nu a}$, was correct. It would then use this assumed value in Eq. (4.4) to compute the estimator's assumed estimation error covariance, P_{xxa}:

$$P_{xxa} = \left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1}H^{\mathrm{T}}P_{\nu\nu a}^{-1}P_{\nu\nu a}P_{\nu\nu a}^{-1}H\left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1} = \left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1} \quad (4.5)$$

The covariance P_{xxa} is useful in its own right. It gives an indication of how accurate the estimate $\hat{\boldsymbol{x}}$ is expected to be, given correct assumptions about the measurements. It is not, however, the true covariance of the estimation errors. The true covariance P_{xx} is computed by a simple form of Consider analysis, substituting the "truth" measurement noise covariance $P_{\nu\nu}$ into Eq. (4.4) to form

$$P_{xx} = \left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1}H^{\mathrm{T}}P_{\nu\nu a}^{-1}P_{\nu\nu}P_{\nu\nu a}^{-1}H\left(H^{\mathrm{T}}P_{\nu\nu a}^{-1}H\right)^{-1} \quad (4.6)$$

This expression does not simplify in the same way as Eq. (4.5). It provides a metric for the actual level of error expected from the batch least squares problem in the presence of incorrectly modeled noise.

Note, it is possible for P_{xx} to be greater than P_{xxa}, less than P_{xxa}, or to have some components that are greater while others are less. The definitions of "greater" and "less" relate to the positive definiteness of $P_{xx} - P_{xxa}$. For ex-

ample, if $P_{\nu\nu a} = \rho P_{\nu\nu}$ for some positive scalar ρ, then $P_{xx} - P_{xxa}$ is positive definite if $\rho < 1$ and negative definite if $\rho > 1$.

4.3 Consider System Formulation

One of the primary challenges in deriving highly general algorithms for Consider covariance analysis is that different kinds of filter model errors enter the system equations in different ways. This section presents a new standardized system formulation for Consider analysis by which all the discussed varieties of filter modeling errors can be represented. The Consider algorithms derived in Section 4.4 manipulate this standard set of equations. Therefore, any system that can be cast into this framework can be analyzed using the general algorithms of Section 4.4, without further customization.

It may be difficult to see how an arbitrary problem can be rewritten in the defined Consider form. In order to bridge this gap between theory and practice, Section 4.5 casts several simple yet widely applicable examples into this section's standard form.

Any Kalman-filter based estimation scheme requires three mathematical components. First, it must have state-space models of the system's dynamic behavior and of the measurements that are to be used for estimation. Second, it must have a complete statistical description of the process and measurement noise, including any correlation between noise components or between noise at different times. Finally, it must have initializations of both the state estimate and the uncertainty in that estimate. Square-root information filters differ from traditional Kalman filters in their representation of process noise and in their ini-

tialization, but they retain the same dynamics and measurement models. The Consider form defined here must also have these three components, so that it can represent the "true" behavior of the system as well as the filter's estimate of that behavior.

Before any consideration of Consider covariance analysis, the filter assumes that the system dynamics and measurement models are given by:

$$\boldsymbol{x}_{k+1} = \Phi_{fk}\boldsymbol{x}_k + \Gamma_{fk}\boldsymbol{w}_k \tag{4.7a}$$

$$\boldsymbol{y}_k = H_{fk}\boldsymbol{x}_k + \boldsymbol{\nu}_k \tag{4.7b}$$

In these equations, $\boldsymbol{x}_k$ is the state vector at discrete sample time k, Φ_{fk} is the state transition matrix from sample k to sample $k+1$, $\boldsymbol{w}_k$ is process noise, and Γ_{fk} is the process noise influence matrix. The vector $\boldsymbol{y}_k$ contains the measurements that apply at sample k, H_{fk} is the measurement sensitivity matrix, and the measurement noise vector is given by $\boldsymbol{\nu}_k$. The measurements are assumed to have been normalized such that $\boldsymbol{\nu}_k$ is a zero-mean, identity-covariance, white, Gaussian random vector:

$$\mathrm{E}[\boldsymbol{\nu}_k] = 0, \quad \mathrm{E}\left[\boldsymbol{\nu}_k\boldsymbol{\nu}_k^{\mathrm{T}}\right] = R_{fk} = \mathrm{I}, \quad \mathrm{E}\left[\boldsymbol{\nu}_k\boldsymbol{\nu}_j^{\mathrm{T}}\right] = 0 \quad \forall\, k \neq j \tag{4.8}$$

Such normalization is always possible via suitable transformation of raw measurements [68]. The Gaussian process noise $\boldsymbol{w}_k$ is also zero-mean and white, with nominal covariance given by

$$\mathrm{E}\left[\boldsymbol{w}_k\boldsymbol{w}_k^{\mathrm{T}}\right] = Q_{fk} \tag{4.9}$$

Process noise and measurement noise are further assumed to be uncorrelated with each other.

According to standard SRIF nomenclature, the filter's process noise information equation, which encapsulates the information about the process noise

statistics in a convenient form, is given by

$$\mathcal{R}_{fwwk}\boldsymbol{w}_k = -\boldsymbol{\nu}_{wk} \tag{4.10}$$

where $\boldsymbol{\nu}_{wk}$ is a zero-mean, identity-covariance, Gaussian random vector, and $\mathcal{R}_{fwwk}$ is related to the nominal process noise covariance according to

$$\mathcal{R}_{fwwk}^{-1}\mathcal{R}_{fwwk}^{-\mathrm{T}} = Q_{fk} \tag{4.11}$$

The notation "$\mathcal{R}$" is used throughout this chapter and Chapter 5 for all square-root information matrices. This differs from the standard notation of Ref. [3], which uses "R" for square-root information matrices. The symbol "R" is reserved for the measurement noise covariance in the present chapter, consistent with Eq. (4.8). The square-root information filter's estimate and covariance are initialized at $k = 0$ by assuming the standard *a priori* state information equation:

$$\bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_0 = \bar{\boldsymbol{z}}_0 - \boldsymbol{\nu}_{x0} \tag{4.12}$$

In this equation, $\boldsymbol{\nu}_{x0}$ is a zero-mean, identity-covariance, Gaussian random vector, and the information state $\bar{\boldsymbol{z}}_0$ parameterizes the initial state estimate according to the equation

$$\bar{\boldsymbol{x}}_0 = \bar{\mathcal{R}}_{fxx0}^{-1}\bar{\boldsymbol{z}}_0 \tag{4.13}$$

The matrix $\bar{\mathcal{R}}_{fxx0}$ is the square-root information matrix for the initial state estimate; it is related to the initial estimation error covariance by

$$\bar{\mathcal{R}}_{fxx0}^{-1}\bar{\mathcal{R}}_{fxx0}^{-\mathrm{T}} = \overline{P}_{fxx0} \tag{4.14}$$

The matrices Φ_{fk}, Γ_{fk}, and H_{fk} are specifically the versions of those system matrices assumed by the filter's model. They may differ from the "truth" system matrices Φ_k, Γ_k, and H_k. In many cases, however, the two sets of matrices

will be identical. In the remainder of this chapter, the symbols Φ_k, Γ_k, and H_k will be used for both the filter and "truth" system whenever the corresponding matrices are the same. Likewise, the covariance and square-root information matrices Q_{fk}, R_{fk}, $\overline{P}_{fxx0}$, $\mathcal{R}_{fwwk}$, $\mathcal{R}_{f\nu\nu k}$, and $\bar{\mathcal{R}}_{fxx0}$ are specifically the versions of those matrices used by the filter. The filter's measurement noise covariance and square-root information matrices R_{fk} and $\mathcal{R}_{f\nu\nu k}$ are always assumed to be identity matrices. The "truth" covariance and square-root information matrices are designated by Q_k, R_k, $\overline{P}_{xx0}$, $\mathcal{R}_{wwk}$, $\mathcal{R}_{\nu\nu k}$, and $\bar{\mathcal{R}}_{xx0}$. Except where they differ or could potentially differ, the symbols for the "truth" matrices will be used.

The Consider system form allows the user to describe a large variety of modeling errors within a common framework. It does this by keeping track of the true statistics of the system uncertainty. Uncertainty can arise from the correctly- or incorrectly-modeled process noise, measurement noise, or initial estimation error covariance. Additional sources of uncertainty include errors in the matrices that specify the dynamics or measurement models used by the filter and unestimated biases or time-varying disturbances. All of these sources of uncertainty, with the exception of any deterministic unestimated biases, are contained in the components of a single "Consider state vector" $\boldsymbol{x}_{\boldsymbol{c}k}$. The Consider analysis computes the contribution of $\boldsymbol{x}_{\boldsymbol{c}k}$ to the true estimation error covariance by tracking its covariance and the sensitivity matrices that specify how it enters the filter equations. All stochastic contributions are contained within this single vector. Therefore, the analysis does not have to additionally compute any cross-correlations, and computationally efficient techniques can be applied to the statistical calculations.

The Consider system begins by defining perturbed versions of the filter's

dynamics and measurement equations:

$$\boldsymbol{x}_{k+1} = \Phi_{fk}\boldsymbol{x}_k + \Gamma_{fk}\boldsymbol{w}_k + \Gamma_{xck}\boldsymbol{x}_{\boldsymbol{c}k} + \boldsymbol{b}_{xk} \tag{4.15a}$$

$$\boldsymbol{y}_k = H_{fk}\boldsymbol{x}_k + H_{ck}\boldsymbol{x}_{\boldsymbol{c}k} + \boldsymbol{b}_{yk} \tag{4.15b}$$

The matrix Γ_{xck} describes the influence of the Consider state vector on the filter's dynamics model, and the matrix H_{ck} describes the influence of the Consider state vector on the measurements. Another difference is the presence of the bias vectors $\boldsymbol{b}_{xk}$ and $\boldsymbol{b}_{yk}$. These are deterministic non-zero values that disturb the system from what the filter assumes. They may be time-varying, but are not driven by stochastic noise. In many cases, these biases are zero and can be neglected in the equations.

Equations (4.15a) and (4.15b) represent the Consider generalizations of Eqs. (4.7a) and (4.7b). The generalized dynamics in Eq. (4.15a) simply adds the two new terms $\Gamma_{xck}\boldsymbol{x}_{\boldsymbol{c}k}$ and $\boldsymbol{b}_{xk}$ to the right-hand side. The generalized measurement equation (4.15b) replaces the zero-mean, identity-covariance measurement noise $\boldsymbol{\nu}_k$ with the sum of the two terms $H_{ck}\boldsymbol{x}_{\boldsymbol{c}k}$ and $\boldsymbol{b}_{yk}$. All of the random part of the "truth" measurement noise is modeled within $H_{ck}\boldsymbol{x}_{\boldsymbol{c}k}$.

The vector of Consider states, $\boldsymbol{x}_{\boldsymbol{c}k}$, is defined differently for each individual system depending on what types of modeling errors are present. It has its own dynamics model given by

$$\boldsymbol{x}_{\boldsymbol{c}k+1} = \Phi_{ck}\boldsymbol{x}_{\boldsymbol{c}k} + \Gamma_{cck}\boldsymbol{w}_{\boldsymbol{c}k} \tag{4.16}$$

In Eq. (4.16), Φ_{ck} is the state transition matrix for the Consider state vector dynamics. It need not be invertible or even square, as the number of elements in $\boldsymbol{x}_{\boldsymbol{c}k}$ may change. The matrix Γ_{cck} describes the influence of the driving Consider process noise vector $\boldsymbol{w}_{\boldsymbol{c}k}$ on the Consider state vector $\boldsymbol{x}_{\boldsymbol{c}k+1}$. Without loss

of generality, both $\boldsymbol{w}_{\boldsymbol{c}k}$ and the initial Consider state $\boldsymbol{x}_{\boldsymbol{c}0}$ are constrained to be zero-mean, identity-covariance, Gaussian random vectors, $\boldsymbol{w}_{\boldsymbol{c}k}$ is a white noise sequence, and $\boldsymbol{x}_{\boldsymbol{c}0}$ is uncorrelated with that sequence.

By means of the Consider vector $\boldsymbol{x}_{\boldsymbol{c}k}$, it is possible to represent process noise and measurement noise with statistics that do not conform to the filter's assumptions. The noise intensities may be different from the nominal values, the process and measurement noise may be correlated, or the noise may be colored, i.e., correlated with noise at other times. Noise terms with non-zero mean values can be modeled by proper use of the bias vectors $\boldsymbol{b}_{xk}$ and $\boldsymbol{b}_{yk}$.

In place of Eq. (4.10), the Consider analysis models the process noise statistics using the following square-root information equation:

$$\mathcal{R}_{fwwk}\boldsymbol{w}_k = -S_{wck}\boldsymbol{x}_{\boldsymbol{c}k} - \boldsymbol{b}_{wk} \tag{4.17}$$

In this equation, S_{wck} is a sensitivity matrix, and $\boldsymbol{b}_{wk}$ is a bias. Note how Eq. (4.17) replaces the simple noise term $\boldsymbol{\nu}_{wk}$ of Eq. (4.10) by the sum of the terms $S_{wck}\boldsymbol{x}_{\boldsymbol{c}k}$ and $\boldsymbol{b}_{wk}$. This replacement enables the Consider process noise model to be non-white, biased, or correlated with the measurement noise. It also allows the "truth" process noise covariance to differ from the filter's assumed value.

In the same way, the initial state information equation of the Consider analysis differs from the filter's Eq. (4.12):

$$\bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_0 = \bar{\boldsymbol{z}}_0 - \bar{S}_{xc0}\boldsymbol{x}_{\boldsymbol{c}0} - \bar{\boldsymbol{b}}_{c0} \tag{4.18}$$

That is, $\boldsymbol{\nu}_{x0}$ of Eq. (4.12) is replaced by $\bar{S}_{xc0}\boldsymbol{x}_{\boldsymbol{c}0} + \bar{\boldsymbol{b}}_{c0}$. The bias $\bar{\boldsymbol{b}}_{c0}$ and sensitivity matrix $\bar{S}_{xc0}$ model the manner in which the "truth" initial mean and covariance of the state differ from the filter's assumed initial mean and covariance.

In summary, the Consider model is a generalized form of the filter model. It consists of Eqs. (4.15)-(4.18). The Consider state dynamics model, Eq. (4.16), is a completely new equation. All of the other Consider model equations are generalizations of the original filter model equations, Eqs. (4.7), (4.10), and (4.12). An important feature of Eq. (4.16) is the complete flexibility with which Φ_{ck} can be defined. In particular, Φ_{ck} can be non-square and can have a non-trivial nullspace, e.g., columns of zeros. This feature allows components of $\boldsymbol{x}_{\boldsymbol{c}k}$ to be white noise, as would be required to reproduce the original filter model.

The Consider covariance analysis is driven by a simple philosophy: In each filter equation, replace the filter's simplistic random terms with the "truth" random terms from the Consider analysis. This has been done explicitly in Consider model Eqs. (4.15b), (4.17), and (4.18), and implicitly in Eq. (4.15a). The same philosophy carries over to the square-root information equations that are recursively updated by the Consider analysis. At each sample time, the filter forms the state information equation

$$\mathcal{R}_{xxk}\boldsymbol{x}_k = \boldsymbol{z}_k - \boldsymbol{\nu}_{xk} \tag{4.19}$$

In the Consider analysis, the simplistic random term $\boldsymbol{\nu}_{xk}$ is replaced by a more complicated expression related to the history of the Consider vector $\boldsymbol{x}_{\boldsymbol{c}k}$. This expression's "truth" mean and covariance replace the zero mean and identity covariance of $\boldsymbol{\nu}_{xk}$ in the Consider analysis covariance calculations. These calculations are given in detail in Section 4.4.

4.4 Consider Analysis Algorithms

The Consider covariance analysis algorithms presented here largely follow standard SRIF procedures. They differ from standard filtering equations in that they keep track of some additional sensitivity matrices and perform a few extra factorizations. The analysis performs the original SRIF calculations and appends new calculations that capture all the model error effects. In most cases, only minimal modification is necessary from the original filter implementation. All the derivations in this section assume that the system has already been placed in the Consider form defined in Section 4.3.

4.4.1 Initialization and First Steps

The first part of the Consider analysis requires several special initialization steps. After initialization, the analysis can proceed recursively with a generic form for each dynamic propagation and measurement update. The analysis is presumed to start from the state information equation, Eq. (4.18), repeated here:

$$\bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_0 = \bar{\boldsymbol{z}}_0 - \bar{S}_{xc0}\boldsymbol{x}_{\boldsymbol{c}0} - \bar{\boldsymbol{b}}_{c0} \tag{4.20}$$

A measurement is assumed to be available at time $k = 0$, as per Eq. (4.15b):

$$\boldsymbol{y}_0 = H_{f0}\boldsymbol{x}_0 + H_{c0}\boldsymbol{x}_{\boldsymbol{c}0} + \boldsymbol{b}_{y0} \tag{4.21}$$

Per standard square-root information algorithms, the measurement update proceeds by rearranging and stacking Eqs. (4.20) and (4.21).

$$\begin{bmatrix} \bar{\mathcal{R}}_{fxx0} \\ H_{f0} \end{bmatrix} \boldsymbol{x}_0 = \begin{bmatrix} \bar{\boldsymbol{z}}_0 \\ \boldsymbol{y}_0 \end{bmatrix} - \begin{bmatrix} \bar{S}_{xc0} \\ H_{c0} \end{bmatrix} \boldsymbol{x}_{\boldsymbol{c}0} - \begin{bmatrix} \bar{\boldsymbol{b}}_{c0} \\ \boldsymbol{b}_{y0} \end{bmatrix} \tag{4.22}$$

Standard orthonormal/upper-triangular (QR) factorization is used to compute an orthonormal transformation matrix $\widehat{T}_0$ such that:

$$\widehat{T}_0 \begin{bmatrix} \bar{\mathcal{R}}_{fxx0} \\ H_{f0} \end{bmatrix} = \begin{bmatrix} \widehat{\mathcal{R}}_{xx0} \\ 0 \end{bmatrix} \tag{4.23}$$

where $\widehat{\mathcal{R}}_{xx0}$ is a square, upper-triangular matrix. The matrices $\widehat{T}_0$ and $\widehat{\mathcal{R}}_{xx0}$ can be computed using the "qr" function that is discussed in the Appendix. The input to this function will be the block matrix on the left-hand side of Eq. (4.23), and the $\widehat{T}_0$ matrix will be the transpose of the function's "Q" matrix output. The function's "R" output will be the entire block matrix on the right-hand side of Eq. (4.23).

The matrix $\widehat{T}_0$ effectively compresses all the information about the state $\boldsymbol{x}_0$, both the *a priori* information and measurement information, into the block $\widehat{\mathcal{R}}_{xx0}$. After computing $\widehat{T}_0$, every term in Eq. (4.22) is premultiplied by this transformation to accomplish the measurement update. The resulting block matrices and vectors are renamed for convenience to obtain the new, updated information equation:

$$\begin{bmatrix} \widehat{\mathcal{R}}_{xx0} \\ 0 \end{bmatrix} \boldsymbol{x}_0 = \begin{bmatrix} \hat{\boldsymbol{z}}_0 \\ \boldsymbol{z}_{r0} \end{bmatrix} - \begin{bmatrix} \widehat{S}_{xc0} \\ S_{rc0} \end{bmatrix} \boldsymbol{x}_{\boldsymbol{c}0} - \begin{bmatrix} \hat{\boldsymbol{b}}_{c0} \\ \boldsymbol{b}_{r0} \end{bmatrix} \tag{4.24}$$

The new vectors $\hat{\boldsymbol{z}}_0$, $\boldsymbol{z}_{r0}$, $\hat{\boldsymbol{b}}_{c0}$, and $\boldsymbol{b}_{r0}$ and the new matrices $\widehat{S}_{xc0}$ and S_{rc0} are the results of this premultiplication. Note that $\hat{\boldsymbol{z}}_0$, $\widehat{S}_{xc0}$, and $\hat{\boldsymbol{b}}_{c0}$ all have row dimensions of n_x, the number of elements in $\boldsymbol{x}$. The bottom rows of this block equation contain information about the measurement residuals. For purposes of this derivation, only the top rows are necessary to continue the analysis. They comprise the *a posteriori* state information equation at $k = 0$:

$$\widehat{\mathcal{R}}_{xx0}\boldsymbol{x}_0 = \hat{\boldsymbol{z}}_0 - \widehat{S}_{xc0}\boldsymbol{x}_{\boldsymbol{c}0} - \hat{\boldsymbol{b}}_{c0} \tag{4.25}$$

So far, the analysis looks exactly like the measurement update for a square-

root information filter, except that the term $\left(\widehat{S}_{xc0}\boldsymbol{x}_{\boldsymbol{c}0} + \hat{\boldsymbol{b}}_{c0}\right)$ takes the place of the previous zero-mean, identity covariance error term $\boldsymbol{\nu}_{x0}$. A naïve approach might continue along these lines, performing dynamic propagation and measurement update steps that keep track of the sensitivity matrices that multiply the Consider state $\boldsymbol{x}_{\boldsymbol{c}k}$.

This approach is not sustainable, however. Successive dynamics and measurement equations introduce a dependence on the Consider state vector at times $k = 1, 2, 3, \ldots$ without eliminating the dependence on the Consider states at previous times. Neglecting any deterministic biases $\boldsymbol{b}_{ck}$, the error term would expand so that the information equation at time k would become

$$\widehat{\mathcal{R}}_{xxk}\boldsymbol{x}_k = \hat{\boldsymbol{z}}_k - \begin{bmatrix} \widehat{S}_{xc0,k} & \widehat{S}_{xc1,k} & \cdots & \widehat{S}_{xck} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_{\boldsymbol{c}0} \\ \boldsymbol{x}_{\boldsymbol{c}1} \\ \vdots \\ \boldsymbol{x}_{\boldsymbol{c}k} \end{bmatrix} \tag{4.26}$$

The sensitivity matrix blocks $\widehat{S}_{xcj,k}$ would be recursively computed at each stage of the analysis.

There are two problems associated with this result. First, the large block sensitivity matrix that multiplies the augmented vector of Consider states keeps growing as k increases. Eventually, this may cause computer storage problems. This long vector of Consider states is unnecessary, because only a subspace of that vector has any effect on the estimation covariance. The dimension of this subspace is less than or equal to n_x, the dimension of the state $\boldsymbol{x}$. In other words, it is theoretically possible to represent all the parts of the error term that actually affect uncertainty with a vector that is the same length as $\boldsymbol{x}_k$.

A second difficulty with Eq. (4.26) is that the Consider analysis would re-

quire one to compute the covariance of the vector containing $\boldsymbol{x}_{\boldsymbol{c}0}$, $\boldsymbol{x}_{\boldsymbol{c}1}$, ..., $\boldsymbol{x}_{\boldsymbol{c}k}$. This computation is not easy. The initial covariance of $\boldsymbol{x}_{\boldsymbol{c}0}$ is constrained to be the identity matrix, but the covariance of each successive $\boldsymbol{x}_{\boldsymbol{c}k}$ must be found by propagating the previous covariance through the Consider dynamics equation. Additionally, the cross-terms of the covariance matrix, of the form $\mathrm{E}\left[\boldsymbol{x}_{\boldsymbol{c}k}\boldsymbol{x}_{\boldsymbol{c}j}^{\mathrm{T}}\right]$, would need to be computed at each step. It is preferable to represent the error term in the information equation in a way that makes these covariance calculations trivial.

After recognizing the weaknesses of the preceding naïve approach, one might be tempted to try to reduce the size of the error term by employing the Consider dynamics equation, Eq. (4.16), to write each Consider state $\boldsymbol{x}_{\boldsymbol{c}k}$ in terms of the Consider state at the initial time, $\boldsymbol{x}_{\boldsymbol{c}0}$. This new strategy, however, introduces the Consider process noise $\boldsymbol{w}_{\boldsymbol{c}k}$ at successive times, so that the information equation at sample time k would become

$$\widehat{\mathcal{R}}_{xxk}\boldsymbol{x}_k = \hat{\boldsymbol{z}}_k - \begin{bmatrix}\widehat{S}_{xc0,k} & \widehat{S}_{xwc0,k} & \cdots & \widehat{S}_{xwck-1,k}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_{\boldsymbol{c}0}\\ \boldsymbol{w}_{\boldsymbol{c}0}\\ \vdots\\ \boldsymbol{w}_{\boldsymbol{c}k-1}\end{bmatrix} \tag{4.27}$$

This result is superior to Eq. (4.26) when it comes to covariance calculations: Each of the elements of the right-most vector has zero mean and identity covariance, and its elements are uncorrelated:

$$\begin{bmatrix}\boldsymbol{x}_{\boldsymbol{c}0}^{\mathrm{T}} & \boldsymbol{w}_{\boldsymbol{c}0}^{\mathrm{T}} & \cdots & \boldsymbol{w}_{\boldsymbol{c}k-1}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \sim \mathcal{N}\left(0, \mathrm{I}\right) \tag{4.28}$$

In other words, the covariance of the entire vector is the identity matrix. However, this form of the analysis still has the problem of the growing dimension of its block sensitivity matrix. Furthermore, the calculations necessary to map the

Consider state back to sample $k = 0$ are nontrivial; they take the form

$$\boldsymbol{x}_{\boldsymbol{c}k+1} = \left(\prod_{j=0}^{k} \Phi_{ck-j} \right) \boldsymbol{x}_{\boldsymbol{c}0} + \sum_{j=0}^{k-1} \left(\prod_{i=0}^{k-j-1} \Phi_{ck-i} \right) \Gamma_{ccj} \boldsymbol{w}_{\boldsymbol{c}j} + \Gamma_{cck} \boldsymbol{w}_{\boldsymbol{c}k} \tag{4.29}$$

In light of the difficulties associated with these two naïve approaches, a transformation has been introduced to achieve two important goals: A minimal-length error effects vector in the information equation and simple statistical properties of that vector. The operations described are computationally similar to some of the orthogonal transformations performed in Ref. [3] to reduce the dimensions of some sensitivity matrix blocks, but with additional statistical benefits in the present formulation.

One version of this transformation applies only at sample $k = 0$. It starts with the information equation error term $\widehat{S}_{xc0}\boldsymbol{x}_{\boldsymbol{c}0}$, where $\boldsymbol{x}_{\boldsymbol{c}0} \sim \mathcal{N}(0, \mathrm{I})$ by construction. Recall that the entire error term can be represented by a vector with dimension less than or equal to n_x, whereas $\boldsymbol{x}_{\boldsymbol{c}0}$ has dimension n_{xc0}, which may already exceed this minimum necessary length. An orthonormal transformation of the vector $\boldsymbol{x}_{\boldsymbol{c}0}$ is performed to separate it into two parts: One part, $\boldsymbol{\alpha}_0$, that has an effect on the information equation for $\boldsymbol{x}_0$, and one part, $\boldsymbol{\beta}_0$, that has no effect on the information equation at the current sample time. Conceptually, these newly-defined "error effects variables" bear some resemblance to those of Refs. [50, 63], although the exact definitions and implementation details are distinct. The present transformation is accomplished by means of LQ factorization, which is analogous to QR factorization except that it decomposes a given matrix into a product of a lower-triangular matrix with possible columns of zeros and an orthonormal matrix. More information about the LQ factorization is given in the Appendix. The LQ factorization is applied here to the matrix $\widehat{S}_{xc0}$:

$$\widehat{S}_{xc0} = \begin{bmatrix} S_{x\alpha 0} & 0 \end{bmatrix} C_0 \tag{4.30}$$

in order to compute the right-hand-side terms as per the Appendix. In this equation, the computed $S_{x\alpha 0}$ is lower-triangular with at most n_x columns. The computed matrix C_0 is orthonormal with n_{xc0} columns and rows. The matrix C_0 transforms $\boldsymbol{x}_{\boldsymbol{c}0}$ to yield

$$\begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\beta}_0 \end{bmatrix} = C_0 \boldsymbol{x}_{\boldsymbol{c}0} \tag{4.31}$$

Note, if $n_{xc0} \leq n_x$, then $\boldsymbol{\beta}_0$ will be an empty vector because there is no way to produce an $\boldsymbol{\alpha}_0$ with fewer elements than $\boldsymbol{x}_{\boldsymbol{c}0}$. It is sometimes convenient to define $[L_{\alpha 0} \quad L_{\beta 0}] \equiv C_0^{\mathrm{T}}$, so that

$$\boldsymbol{x}_{\boldsymbol{c}0} = C_0^{\mathrm{T}} \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\beta}_0 \end{bmatrix} = \begin{bmatrix} L_{\alpha 0} & L_{\beta 0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\beta}_0 \end{bmatrix} = L_{\alpha 0} \boldsymbol{\alpha}_0 + L_{\beta 0} \boldsymbol{\beta}_0 \tag{4.32}$$

As C_0 and its inverse C_0^{T} are both orthonormal, transformation of $\boldsymbol{x}_{\boldsymbol{c}0}$ by either of these matrices preserves its property of identity covariance. This result will be important for calculation of the Consider analysis covariances. One can now use Eqs. (4.30) and (4.31) to rewrite the state information equation, Eq. (4.25), so that $\widehat{S}_{xc0}$ is replaced by the potentially lower-dimensional sensitivity matrix $S_{x\alpha 0}$.

$$\begin{aligned} \widehat{\mathcal{R}}_{xx0} \boldsymbol{x}_0 &= \hat{\boldsymbol{z}}_0 - \widehat{S}_{xc0} \boldsymbol{x}_{\boldsymbol{c}0} - \hat{\boldsymbol{b}}_{c0} \\ &= \hat{\boldsymbol{z}}_0 - \begin{bmatrix} S_{x\alpha 0} & 0 \end{bmatrix} C_0 \boldsymbol{x}_{\boldsymbol{c}0} - \hat{\boldsymbol{b}}_{c0} \\ &= \hat{\boldsymbol{z}}_0 - \begin{bmatrix} S_{x\alpha 0} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\beta}_0 \end{bmatrix} - \hat{\boldsymbol{b}}_{c0} \\ &= \hat{\boldsymbol{z}}_0 - S_{x\alpha 0} \boldsymbol{\alpha}_0 - \hat{\boldsymbol{b}}_{c0} \end{aligned} \tag{4.33}$$

The final line of Eq. (4.33) is the generic Consider analysis form of the *a posteriori* SRIF information equation. It will be used throughout the remaining analysis.

4.4.2 Main Algorithm

At this point, it is desirable to shift away from derivations specific to a particular sample so that the full Consider analysis can be developed in a manner analogous to mathematical induction. To that end, assume that Eq. (4.33) is available for some generic sample time k:

$$\widehat{\mathcal{R}}_{xxk}\boldsymbol{x}_k = \hat{\boldsymbol{z}}_k - S_{x\alpha k}\boldsymbol{\alpha}_k - \hat{\boldsymbol{b}}_{ck} \tag{4.34}$$

Note that the previously derived Eq. (4.33) can be recovered by plugging in $k = 0$. Assume also that $\boldsymbol{x_c}_k$ is related to $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ by some transformation of the form:

$$\boldsymbol{x_c}_k = \begin{bmatrix} L_{\alpha k} & L_{\beta k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} = L_{\alpha k}\boldsymbol{\alpha}_k + L_{\beta k}\boldsymbol{\beta}_k \tag{4.35}$$

where the vector $\begin{bmatrix} \boldsymbol{\alpha}_k^{\mathrm{T}} & \boldsymbol{\beta}_k^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ has zero mean and identity covariance. Note, however, that $\boldsymbol{x_c}_k$ is not required to have identity covariance for $k > 0$, and consequently it is no longer required that $[L_{\alpha k} \quad L_{\beta k}]$ be orthonormal.

To complete the Consider covariance analysis algorithm, it is necessary to develop dynamic propagation and measurement update steps that transition these equations from sample k to sample $k + 1$. Initialization of the mathematical induction occurs at sample $k = 0$ because Eqs. (4.33) and (4.32) constitute Eqs. (4.34) and (4.35) when $k = 0$ in the latter equations.

The development starts by rewriting the dynamics, measurement, and process-noise information equations in terms of the $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ vectors instead of $\boldsymbol{x_c}_k$. First, the system dynamics and measurement equations, Eqs. (4.15a) and (4.15b), become

$$\boldsymbol{x}_{k+1} = \Phi_{fk}\boldsymbol{x}_k + \Gamma_{fk}\boldsymbol{w}_k + \Gamma_{xck} \begin{bmatrix} L_{\alpha k} & L_{\beta k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} + \boldsymbol{b}_{xk} \tag{4.36a}$$

$$\boldsymbol{y}_k = H_{fk}\boldsymbol{x}_k + H_{ck}\begin{bmatrix} L_{\alpha k} & L_{\beta k} \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} + \boldsymbol{b}_{yk} \tag{4.36b}$$

Next, the Consider dynamics of Eq. (4.16) can be written as

$$\boldsymbol{x}_{\boldsymbol{c}k+1} = \Phi_{ck}\begin{bmatrix} L_{\alpha k} & L_{\beta k} \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} + \Gamma_{cck}\boldsymbol{w}_{\boldsymbol{c}k} = \begin{bmatrix} \Phi_{ck}L_{\alpha k} & \Phi_{ck}L_{\beta k} & \Gamma_{cck} \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} \tag{4.37}$$

Because the Consider process noise $\boldsymbol{w}_{\boldsymbol{c}k}$ is independent of the Consider state at time k, the augmented right-most vector in Eq. (4.37) is zero-mean, identity-covariance, and Gaussian:

$$\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} \sim \mathcal{N}\left(0, \mathrm{I}\right) \tag{4.38}$$

Finally, the process noise information equation, Eq. (4.17), now takes the form

$$\mathcal{R}_{wwk}\boldsymbol{w}_k = -S_{wck}\begin{bmatrix} L_{\alpha k} & L_{\beta k} \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} - \boldsymbol{b}_{wk} \tag{4.39}$$

Dynamic propagation proceeds as usual for a square-root information filter, except that in place of the information equation error vector $\boldsymbol{\nu}_{xk}$, one manipulates the terms involving $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$, and $\hat{\boldsymbol{b}}_{ck}$. As in a standard SRIF, the propagation solves the dynamics equation, Eq. (4.36a), for $\boldsymbol{x}_k$ and substitutes this expression into information Eq. (4.34). The result is stacked together with process noise information Eq. (4.39) to yield:

$$\begin{aligned}\begin{bmatrix} \mathcal{R}_{wwk} & 0 \\ -\widehat{\mathcal{R}}_{xxk}\Phi_{fk}^{-1}\Gamma_{fk} & \widehat{\mathcal{R}}_{xxk}\Phi_{fk}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{x}_{k+1} \end{bmatrix} &= \begin{bmatrix} 0 \\ \hat{\boldsymbol{z}}_k \end{bmatrix} \\ &- \begin{bmatrix} S_{wck}L_{\alpha k} & S_{wck}L_{\beta k} \\ \left(S_{x\alpha k} - \widehat{\mathcal{R}}_{xxk}\Phi_{fk}^{-1}\Gamma_{xck}L_{\alpha k}\right) & \left(-\widehat{\mathcal{R}}_{xxk}\Phi_{fk}^{-1}\Gamma_{xck}L_{\beta k}\right) \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} \\ &- \begin{bmatrix} \boldsymbol{b}_{wk} \\ \hat{\boldsymbol{b}}_{ck} - \widehat{\mathcal{R}}_{xxk}\Phi_{fk}^{-1}\boldsymbol{b}_{xk} \end{bmatrix}\end{aligned} \tag{4.40}$$

Propagation is completed by using QR factorization to compute an orthonormal transformation $\overline{T}_{k+1}$ that triangularizes the block matrix on the left. This operation yields the relationship

$$\overline{T}_{k+1}\begin{bmatrix} \mathcal{R}_{wwk} & 0 \\ -\widehat{\mathcal{R}}_{xxk}\Phi_{fk}^{-1}\Gamma_{fk} & \widehat{\mathcal{R}}_{xxk}\Phi_{fk}^{-1} \end{bmatrix} = \begin{bmatrix} \bar{\mathcal{R}}_{wwk} & \bar{\mathcal{R}}_{wxk+1} \\ 0 & \bar{\mathcal{R}}_{xxk+1} \end{bmatrix} \tag{4.41}$$

where the square, upper-triangular matrices $\bar{\mathcal{R}}_{wwk}$ and $\bar{\mathcal{R}}_{xxk+1}$ and the general matrix $\bar{\mathcal{R}}_{wxk+1}$ are also computed by the factorization. Thus, left matrix multiplication of Eq. (4.40) by $\overline{T}_{k+1}$ yields:

$$\begin{bmatrix} \bar{\mathcal{R}}_{wwk} & \bar{\mathcal{R}}_{wxk+1} \\ 0 & \bar{\mathcal{R}}_{xxk+1} \end{bmatrix}\begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{x}_{k+1} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{z}}_{wk} \\ \bar{\boldsymbol{z}}_{k+1} \end{bmatrix} - \begin{bmatrix} \bar{S}_{w\alpha k} & \bar{S}_{w\beta k} \\ \bar{S}_{x\alpha k} & \bar{S}_{x\beta k} \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} - \begin{bmatrix} \bar{\boldsymbol{b}}_{wk} \\ \bar{\boldsymbol{b}}_{ck+1} \end{bmatrix} \tag{4.42}$$

The vectors $\bar{\boldsymbol{z}}_{wk}$ and $\bar{\boldsymbol{z}}_{k+1}$ are the standard SRIF *a priori* information vectors that result from transforming the first term on the right-hand side of Eq. (4.40). The sensitivity matrices $\bar{S}_{w\alpha k}$, $\bar{S}_{w\beta k}$, $\bar{S}_{x\alpha k}$, and $\bar{S}_{x\beta k}$ come from transformation of the large block matrix on the right-hand side of Eq. (4.40). The vectors $\bar{\boldsymbol{b}}_{wk}$ and $\bar{\boldsymbol{b}}_{ck+1}$ are biases obtained by transforming the right-most term of Eq. (4.40).

In Eq. (4.42), the upper blocks relating to process noise are required for Consider analysis of a smoother for the given system. Consider analysis of a smoother can be accomplished by extending the present techniques, as is done in Chapter 5. For the remainder of the filter analysis derived here, only the lower blocks of Eq. (4.42) are retained:

$$\bar{\mathcal{R}}_{xxk+1}\boldsymbol{x}_{k+1} = \bar{\boldsymbol{z}}_{k+1} - \begin{bmatrix} \bar{S}_{x\alpha k} & \bar{S}_{x\beta k} \end{bmatrix}\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} - \bar{\boldsymbol{b}}_{ck+1} \tag{4.43}$$

This equation is just the Consider version of the *a priori* state information equation at sample $k+1$. Note that its error vector is still written in terms of $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$, rather than $\boldsymbol{\alpha}_{k+1}$ and $\boldsymbol{\beta}_{k+1}$.

The measurement update step at sample $k+1$ closely resembles the procedure of Section 4.4.1, except that the composite random vector $\begin{bmatrix} \boldsymbol{\alpha}_k^{\mathrm{T}} & \boldsymbol{\beta}_k^{\mathrm{T}} & \boldsymbol{w}_{\boldsymbol{c}k}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$

appears instead of $\boldsymbol{x}_{\boldsymbol{c}k+1}$. After stacking *a priori* information Eq. (4.43) and the measurement equation, one obtains an analog of Eq. (4.22):

$$\begin{bmatrix} \bar{\mathcal{R}}_{xxk+1} \\ H_{fk+1} \end{bmatrix} \boldsymbol{x}_{k+1} = \begin{bmatrix} \bar{\boldsymbol{z}}_{k+1} \\ \boldsymbol{y}_{k+1} \end{bmatrix} - \begin{bmatrix} \bar{S}_{x\alpha k} & \bar{S}_{x\beta k} & 0 \\ (H_{ck+1}\Phi_{ck}L_{\alpha k}) & (H_{ck+1}\Phi_{ck}L_{\beta k}) & (H_{ck+1}\Gamma_{cck}) \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} - \begin{bmatrix} \bar{\boldsymbol{b}}_{ck+1} \\ \boldsymbol{b}_{yk+1} \end{bmatrix} \tag{4.44}$$

The top line of this equation is just Eq. (4.43). The bottom line is a modified version of measurement Eq. (4.15b) at sample $k+1$. In this measurement equation, $\boldsymbol{x}_{\boldsymbol{c}k+1}$ is replaced by the expression on the extreme right-hand side of Eq. (4.37). As in the $k = 0$ case, the orthonormal transformation matrix $\widehat{T}_{k+1}$ is computed via QR factorization of the block matrix on the left, such that

$$\widehat{T}_{k+1} \begin{bmatrix} \bar{\mathcal{R}}_{xxk+1} \\ H_{fk+1} \end{bmatrix} = \begin{bmatrix} \widehat{\mathcal{R}}_{xxk+1} \\ 0 \end{bmatrix} \tag{4.45}$$

where the additional factorization output $\widehat{\mathcal{R}}_{xxk+1}$ is square and upper triangular. Left matrix multiplication of Eq. (4.44) by $\widehat{T}_{k+1}$ yields:

$$\begin{bmatrix} \widehat{\mathcal{R}}_{xxk+1} \\ 0 \end{bmatrix} \boldsymbol{x}_{k+1} = \begin{bmatrix} \hat{\boldsymbol{z}}_{k+1} \\ \boldsymbol{z}_{rk+1} \end{bmatrix} - \begin{bmatrix} \tilde{S}_{x\alpha k} & \tilde{S}_{x\beta k} & \tilde{S}_{xwck} \\ \tilde{S}_{r\alpha k} & \tilde{S}_{r\beta k} & \tilde{S}_{rwck} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} - \begin{bmatrix} \hat{\boldsymbol{b}}_{ck+1} \\ \boldsymbol{b}_{rk+1} \end{bmatrix} \tag{4.46}$$

Equation (4.46) is a transformed version of Eq. (4.44) in the same way as Eq. (4.42) is a transformed version of Eq. (4.40). Therefore, the various new vectors and matrices on the right-hand side of Eq. (4.46) are computed from the corresponding terms in Eq. (4.44) in a manner analogous to the Eq. (4.42) derivation. The lower blocks of Eq. (4.46) are related to measurement residuals. Neglecting these, the *a posteriori* state information equation at sample time $k+1$

is given by

$$\widehat{\mathcal{R}}_{xxk+1}\boldsymbol{x}_{k+1} = \hat{\boldsymbol{z}}_{k+1} - \begin{bmatrix}\tilde{S}_{x\alpha k} & \tilde{S}_{x\beta k} & \tilde{S}_{xwck}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{\boldsymbol{c}k}\end{bmatrix} - \hat{\boldsymbol{b}}_{ck+1} \tag{4.47}$$

Recall that the main algorithm derivation began with Eq. (4.34), the *a posteriori* state information equation at sample k. The current information equation is not yet in the same form, however, because it depends on the three vectors $\boldsymbol{\alpha}_k$, $\boldsymbol{\beta}_k$, and $\boldsymbol{w}_{\boldsymbol{c}k}$ rather than the single vector $\boldsymbol{\alpha}_{k+1}$. At this point, an LQ factorization can be applied to transform the composite vector $\begin{bmatrix}\boldsymbol{\alpha}_k^{\mathrm{T}} & \boldsymbol{\beta}_k^{\mathrm{T}} & \boldsymbol{w}_{\boldsymbol{c}k}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ and its coefficient matrix in Eq. (4.47). This process will yield the desired $\boldsymbol{\alpha}_{k+1}$ as a component of the transformed vector and $S_{x\alpha k+1}$ as the only non-zero block of the transformed matrix, similar to the technique used in Eqs. (4.30)-(4.33) for sample $k = 0$.

The recursion process requires an updated version of Eq. (4.35) that applies at sample $k + 1$. This new equation will express $\boldsymbol{x}_{\boldsymbol{c}k+1}$ as a linear combination of $\boldsymbol{\alpha}_{k+1}$ and a newly-defined $\boldsymbol{\beta}_{k+1}$. It will be derived from the Consider state dynamics using the dynamics form in Eq. (4.37). The derivation also uses an LQ factorization that transforms the vector $\begin{bmatrix}\boldsymbol{\alpha}_k^{\mathrm{T}} & \boldsymbol{\beta}_k^{\mathrm{T}} & \boldsymbol{w}_{\boldsymbol{c}k}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$. It is convenient to LQ factorize the relevant coefficient matrices from Eqs. (4.47) and (4.37) simultaneously. The required LQ factorization computes another orthonormal matrix C_{k+1} such that the following relationship holds:

$$\begin{bmatrix}\tilde{S}_{x\alpha k} & \tilde{S}_{x\beta k} & \tilde{S}_{xwck} \\ \Phi_{ck}L_{\alpha k} & \Phi_{ck}L_{\beta k} & \Gamma_{cck}\end{bmatrix} = \begin{bmatrix}S_{x\alpha k+1} & 0 & 0 \\ L_{\alpha k+1} & L_{\beta k+1} & 0\end{bmatrix} C_{k+1} \tag{4.48}$$

The input to this LQ factorization is the block matrix on the left-hand side of Eq. (4.48). The upper block row is the coefficient of $\begin{bmatrix}\boldsymbol{\alpha}_k^{\mathrm{T}} & \boldsymbol{\beta}_k^{\mathrm{T}} & \boldsymbol{w}_{\boldsymbol{c}k}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ from Eq. (4.47), and the lower row is the coefficient of the same vector in Eq. (4.37). In

addition to C_{k+1}, the square lower-triangular matrices $S_{x\alpha k+1}$ and $L_{\beta k+1}$ and the general matrix $L_{\alpha k+1}$ are outputs of the factorization. The orthonormal matrix C_{k+1} is used to define a transformation of variables from the vector at sample k to a different vector that will apply at sample $k+1$:

$$\begin{bmatrix} \boldsymbol{\alpha}_{k+1} \\ \boldsymbol{\beta}_{k+1} \\ \boldsymbol{\gamma}_{k+1} \end{bmatrix} \equiv C_{k+1} \begin{bmatrix} \boldsymbol{\alpha}_{k} \\ \boldsymbol{\beta}_{k} \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} \tag{4.49}$$

where $\begin{bmatrix} \boldsymbol{\alpha}_{k+1}^{\mathrm{T}} & \boldsymbol{\beta}_{k+1}^{\mathrm{T}} & \boldsymbol{\gamma}_{k+1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \sim \mathcal{N}(0, \mathrm{I})$ because the orthonormal transformation preserves the vector's identity covariance.

When this transformation is applied to the Consider dynamics equation by substituting Eqs. (4.48) and (4.49) into Eq. (4.37), it yields

$$\begin{aligned} \boldsymbol{x}_{\boldsymbol{c}k+1} &= \begin{bmatrix} \Phi_{ck} L_{\alpha k} & \Phi_{ck} L_{\beta k} & \Gamma_{cck} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{k} \\ \boldsymbol{\beta}_{k} \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} = \begin{bmatrix} L_{\alpha k+1} & L_{\beta k+1} & 0 \end{bmatrix} C_{k+1} \begin{bmatrix} \boldsymbol{\alpha}_{k} \\ \boldsymbol{\beta}_{k} \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} \\ &= \begin{bmatrix} L_{\alpha k+1} & L_{\beta k+1} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{k+1} \\ \boldsymbol{\beta}_{k+1} \\ \boldsymbol{\gamma}_{k+1} \end{bmatrix} = \begin{bmatrix} L_{\alpha k+1} & L_{\beta k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{k+1} \\ \boldsymbol{\beta}_{k+1} \end{bmatrix} \end{aligned} \tag{4.50}$$

The final result is a copy of Eq. (4.35), referenced to sample $k+1$ instead of sample k. Likewise, one can substitute Eqs. (4.48) and (4.49) into Eq. (4.47) to

get

$$\begin{aligned}
\widehat{\mathcal{R}}_{xxk+1}\boldsymbol{x}_{k+1} &= \hat{\boldsymbol{z}}_{k+1} - \begin{bmatrix}\tilde{S}_{x\alpha k} & \tilde{S}_{x\beta k} & \tilde{S}_{xwck}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{ck}\end{bmatrix} - \hat{\boldsymbol{b}}_{ck+1} \\
&= \hat{\boldsymbol{z}}_{k+1} - \begin{bmatrix}S_{x\alpha k+1} & 0 & 0\end{bmatrix} C_{k+1}\begin{bmatrix}\boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{ck}\end{bmatrix} - \hat{\boldsymbol{b}}_{ck+1} \\
&= \hat{\boldsymbol{z}}_{k+1} - \begin{bmatrix}S_{x\alpha k+1} & 0 & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_{k+1} \\ \boldsymbol{\beta}_{k+1} \\ \boldsymbol{\gamma}_{k+1}\end{bmatrix} - \hat{\boldsymbol{b}}_{ck+1} \\
&= \hat{\boldsymbol{z}}_{k+1} - S_{x\alpha k+1}\boldsymbol{\alpha}_{k+1} - \hat{\boldsymbol{b}}_{ck+1}
\end{aligned} \tag{4.51}$$

The final line of Eq. (4.51) is a copy of Eq. (4.34), referenced to sample $k+1$ rather than sample k. At this point, the algorithms have come full circle and the dynamic propagation from $k+1$ to $k+2$ can commence.

4.4.3 Consider Covariance Calculations

One of the distinguishing features of the SRIF as opposed to the Kalman filter is that the filter does not compute its own covariance estimate at each step. Whenever the covariance is needed, however, it can be reconstructed from the square-root information matrices. This chapter's Consider covariance analysis operates in the same way; at any point the analyst may compute both the filter's presumed error covariance and the true covariance of the estimation error. More generally, the quantity of interest in a Consider analysis will be the matrix mean square error (MSE), which consists of a covariance plus a rank-one term due to biases. These calculations can occur at the *a priori* or *a posteriori* stages of the analysis with equal simplicity.

To compute covariance, one starts with the *a priori* or *a posteriori* state information equation, both repeated here for convenience:

$$\bar{\mathcal{R}}_{xxk}\boldsymbol{x}_k = \bar{\boldsymbol{z}}_k - \begin{bmatrix}\bar{S}_{x\alpha k-1} & \bar{S}_{x\beta k-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_{k-1}\\ \boldsymbol{\beta}_{k-1}\end{bmatrix} - \bar{\boldsymbol{b}}_{ck} \tag{4.52a}$$

$$\widehat{\mathcal{R}}_{xxk}\boldsymbol{x}_k = \hat{\boldsymbol{z}}_k - S_{x\alpha k}\boldsymbol{\alpha}_k - \hat{\boldsymbol{b}}_{ck} \tag{4.52b}$$

The filter's *a priori* and *a posteriori* state estimates are $\bar{\boldsymbol{x}}_k = \bar{\mathcal{R}}_{xxk}^{-1}\bar{\boldsymbol{z}}_k$ and $\hat{\boldsymbol{x}}_k = \widehat{\mathcal{R}}_{xxk}^{-1}\hat{\boldsymbol{z}}_k$, respectively. Errors in the filter's presumed covariances arise from the filter's incorrect assumption that the errors in the information equations have zero mean and identity covariance. Under this assumption, one can compute the filter's presumed covariances. They are

$$\overline{P}_{fxxk} = \mathrm{E}\Big[(\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k)(\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k)^{\mathrm{T}}\Big] = \bar{\mathcal{R}}_{xxk}^{-1}\mathrm{E}\big[\bar{\boldsymbol{\nu}}_{xk}\bar{\boldsymbol{\nu}}_{xk}^{\mathrm{T}}\big]\,\bar{\mathcal{R}}_{xxk}^{-\mathrm{T}} = \bar{\mathcal{R}}_{xxk}^{-1}\bar{\mathcal{R}}_{xxk}^{-\mathrm{T}} \tag{4.53a}$$

$$P_{fxxk} = \mathrm{E}\Big[(\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)(\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)^{\mathrm{T}}\Big] = \widehat{\mathcal{R}}_{xxk}^{-1}\mathrm{E}\big[\boldsymbol{\nu}_{xk}\boldsymbol{\nu}_{xk}^{\mathrm{T}}\big]\,\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}} = \widehat{\mathcal{R}}_{xxk}^{-1}\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}} \tag{4.53b}$$

The true matrix MSEs of the filter's estimates are only slightly more difficult to compute and require no additional matrix inversions. Starting from Eq. (4.52a) for the *a priori* case, the true matrix MSE is given by

$$\begin{aligned}
\overline{P}_{xxk} &= \mathrm{E}\Big[(\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k)(\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k)^{\mathrm{T}}\Big] \\
&= \bar{\mathcal{R}}_{xxk}^{-1}\mathrm{E}\left[\left(\begin{bmatrix}\bar{S}_{x\alpha k-1} & \bar{S}_{x\beta k-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_{k-1}\\ \boldsymbol{\beta}_{k-1}\end{bmatrix} + \bar{\boldsymbol{b}}_{ck}\right)\right. \\
&\qquad\left.\times\left(\begin{bmatrix}\bar{S}_{x\alpha k-1} & \bar{S}_{x\beta k-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_{k-1}\\ \boldsymbol{\beta}_{k-1}\end{bmatrix} + \bar{\boldsymbol{b}}_{ck}\right)^{\mathrm{T}}\right]\bar{\mathcal{R}}_{xxk}^{-\mathrm{T}} \\
&= \bar{\mathcal{R}}_{xxk}^{-1}\begin{bmatrix}\bar{S}_{x\alpha k-1} & \bar{S}_{x\beta k-1}\end{bmatrix}\mathrm{E}\left\{\begin{bmatrix}\boldsymbol{\alpha}_{k-1}\\ \boldsymbol{\beta}_{k-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_{k-1}^{\mathrm{T}} & \boldsymbol{\beta}_{k-1}^{\mathrm{T}}\end{bmatrix}\right\}\begin{bmatrix}\bar{S}_{x\alpha k-1}^{\mathrm{T}}\\ \bar{S}_{x\beta k-1}^{\mathrm{T}}\end{bmatrix}\bar{\mathcal{R}}_{xxk}^{-\mathrm{T}} \\
&\qquad + \bar{\mathcal{R}}_{xxk}^{-1}\bar{\boldsymbol{b}}_{ck}\bar{\boldsymbol{b}}_{ck}^{\mathrm{T}}\bar{\mathcal{R}}_{xxk}^{-\mathrm{T}} \\
&= \bar{\mathcal{R}}_{xxk}^{-1}\begin{bmatrix}\bar{S}_{x\alpha k-1} & \bar{S}_{x\beta k-1}\end{bmatrix}\begin{bmatrix}\bar{S}_{x\alpha k-1}^{\mathrm{T}}\\ \bar{S}_{x\beta k-1}^{\mathrm{T}}\end{bmatrix}\bar{\mathcal{R}}_{xxk}^{-\mathrm{T}} + \bar{\mathcal{R}}_{xxk}^{-1}\bar{\boldsymbol{b}}_{ck}\bar{\boldsymbol{b}}_{ck}^{\mathrm{T}}\bar{\mathcal{R}}_{xxk}^{-\mathrm{T}}
\end{aligned} \tag{4.54}$$

Likewise, for the *a posteriori* case, starting from Eq. (4.52b),

$$\begin{aligned} P_{xxk} &= \mathrm{E}\Big[(\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)\,(\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k)^{\mathrm{T}}\Big] \\ &= \widehat{\mathcal{R}}_{xxk}^{-1}\mathrm{E}\left[\left(S_{x\alpha k}\boldsymbol{\alpha}_k + \hat{\boldsymbol{b}}_{ck}\right)\left(S_{x\alpha k}\boldsymbol{\alpha}_k + \hat{\boldsymbol{b}}_{ck}\right)^{\mathrm{T}}\right]\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}} \\ &= \widehat{\mathcal{R}}_{xxk}^{-1}S_{x\alpha k}\mathrm{E}\big[\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^{\mathrm{T}}\big]\,S_{x\alpha k}^{\mathrm{T}}\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}} + \widehat{\mathcal{R}}_{xxk}^{-1}\hat{\boldsymbol{b}}_{ck}\hat{\boldsymbol{b}}_{ck}^{\mathrm{T}}\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}} \\ &= \widehat{\mathcal{R}}_{xxk}^{-1}S_{x\alpha k}S_{x\alpha k}^{\mathrm{T}}\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}} + \widehat{\mathcal{R}}_{xxk}^{-1}\hat{\boldsymbol{b}}_{ck}\hat{\boldsymbol{b}}_{ck}^{\mathrm{T}}\widehat{\mathcal{R}}_{xxk}^{-\mathrm{T}} \end{aligned} \tag{4.55}$$

Thus, the true matrix MSE calculations differ from the filter's covariance calculations in that they are weighted by a product of sensitivity matrices and may have an additional bias term. Note that in the final lines of Eqs. (4.54) and (4.55), the first term is the true estimation error covariance, and the second term constitutes the bias effect. The sum of these two terms is designated "P_{xx}" even though it is a matrix MSE rather than a covariance. This non-standard notation emphasizes the fact that the Consider quantities $\overline{P}_{xxk}$ and P_{xxk} are most directly comparable to the presumed filter covariances $\overline{P}_{fxxk}$ and P_{fxxk}.

4.5 Examples

To demonstrate the capabilities of the proposed Consider analysis and provide a degree of confidence in the derived algorithms, several concrete examples have been developed. These examples were chosen because they are sufficiently simple to explain, yet they address some realistic engineering problems. One of the biggest practical challenges of this chapter's Consider covariance analysis is the process of rewriting each problem in the defined Consider form of Section 4.3. Once the examples are in Consider form, the analysis algorithms of Section 4.4 can be implemented.

The results of the Consider analysis for each example are compared to the results of Monte Carlo simulations. These simulations verify that the true matrix MSEs of the incorrectly-modeled filters are indeed those predicted by the Consider covariance analysis. The simulations also prove to be a useful debugging tool for the code that implements the algorithms.

The remainder of this section is organized as follows: Sections 4.5.1 and 4.5.2 lay out the two examples, one of mismodeled noise and the other with incorrect system matrices. Each section begins with a scenario description and then shows step-by-step how to rewrite the given systems in Consider form. The Monte Carlo simulations are explained, and the simulation results are presented and discussed.

4.5.1 Example: Incorrectly Modeled Noise

Scenario Description and Consider Form Setup

The dynamics and measurement models for the first example are given by

$$\begin{bmatrix} r_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_k \\ v_k \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_k \tag{4.56a}$$

$$y_k = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} r_k \\ v_k \end{bmatrix} + \nu_k \tag{4.56b}$$

The states r_k and v_k are thought of as position and velocity for some one-dimensional motion, and the sample interval Δt is assumed to be 0.5 s. The variables w_k and ν_k are scalar process and measurement noise, respectively, and y_k is a scalar measurement. The initial state estimate and initial state error co-

variance are

$$\bar{\boldsymbol{x}}_0 = \begin{bmatrix} \bar{r}_0 \\ \bar{v}_0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \overline{P}_{xx0} = \begin{bmatrix} \sigma_{r0}^2 & \sigma_{rv0} \\ \sigma_{rv0} & \sigma_{v0}^2 \end{bmatrix} = \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix} \tag{4.57}$$

The filter operates on measurements available once per sample interval Δt for a total of 50 s (100 discrete-time samples).

One of the most practical uses for Consider covariance analysis is to deal with the situation of incorrect process and measurement noise covariances. Specifically, in this example the filter process noise covariance Q_{fk} and measurement noise covariance R_{fk} are assumed to be

$$Q_{fk} = 1, \quad R_{fk} = 1 \tag{4.58}$$

The "truth" process noise covariance Q_k and measurement noise covariance R_k are

$$Q_k = 0.25, \quad R_k = 2.25 \tag{4.59}$$

Note that the "truth" process noise covariance is smaller and the "truth" measurement noise covariance larger than modeled by the filter.

In this example, the system matrices Φ_k, Γ_k, and H_k can be extracted directly from Eqs. (4.56a) and (4.56b) for both the filter and the Consider form. The initial state square-root information matrix $\bar{\mathcal{R}}_{xx0}$ can be computed from the initial estimation error covariance $\overline{P}_{xx0}$ as in Eq. (4.14) by using Cholesky factorization or some other standard matrix square-root routine. The initial information state $\bar{\boldsymbol{z}}_0$ can be obtained by substituting the values for $\bar{\mathcal{R}}_{xx0}$ and $\bar{\boldsymbol{x}}_0$ into the inverse of Eq. (4.13). These calculations yield

$$\bar{\mathcal{R}}_{xx0} = \begin{bmatrix} 1/\sqrt{10} & 0 \\ 0 & 1/\sqrt{5} \end{bmatrix}, \quad \bar{\boldsymbol{z}}_0 = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{5} \end{bmatrix} \tag{4.60}$$

The filter's process noise information matrix $\mathcal{R}_{fwwk}$ is the inverse square root of Q_{fk}; it can also be computed by Cholesky factorization, but in this scalar case will just be $1/\sqrt{1} = 1$. Likewise, the "truth" inverse square roots for process and measurement noise, which will be needed later, are just $\mathcal{R}_{wwk} = 1/\sqrt{0.25} = 2$ and $\mathcal{R}_{\nu\nu k} = 1/\sqrt{2.25} = 2/3$.

So far, the quantities defined or computed hold for both the filter's equations and the system's Consider form. In order to put each system into Consider form, the user must correctly define the matrices Γ_{xck}, Φ_{ck}, Γ_{cck}, S_{wck}, H_{ck}, and $\bar{S}_{xc0}$, which appear in Eqs. (4.15)-(4.18). One must also compute any deterministic biases, but for this example there are none. Typically, the design of the matrices is not unique, as it depends on the particular choice and ordering of elements of the Consider state vector $\boldsymbol{x_{ck}}$.

A rough procedure for transforming a given system to Consider form is as follows: First, one defines $\boldsymbol{x_{ck}}$ so that it contains as elements all the sources of random uncertainty or modeling error in the system, either directly or as some transformed version. Particular care must be taken in defining the initial $\boldsymbol{x_{c0}}$ in a way that also includes initial estimation uncertainty. In addition, it must have zero mean and identity covariance. Next, one writes the equations that describe the dynamics and statistics of the uncertainty components contained in $\boldsymbol{x_{ck}}$. This step typically involves designing the matrices Φ_{ck} and Γ_{cck}, as well as precomputing any deterministic bias disturbances. Third, one specifies how the uncertainty enters the system dynamics and measurements by creating the matrices Γ_{xck} and H_{ck}. Finally, one writes expressions for the sensitivity matrices S_{wck} and $\bar{S}_{xc0}$ that determine how the Consider state vector $\boldsymbol{x_{ck}}$ influences the filter's process noise equation and its initial state information equation. Al-

though all of these steps are necessary to define a Consider system, it may not always be convenient to perform them in this order. Trial and error may help to determine the most useful of the system description alternatives.

All of the errors in this example are related to noise statistics, and the only additional uncertainty that must be captured by the Consider vector $\boldsymbol{x}_{\boldsymbol{c}k}$ is the correctly-modeled uncertainty in the initial estimate. One suitable definition is

$$\boldsymbol{x}_{\boldsymbol{c}k} \equiv \begin{cases} \begin{bmatrix} \boldsymbol{\nu}_{x0} \\ \mathcal{R}_{ww0} w_0 \\ \mathcal{R}_{\nu\nu 0} \nu_0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} w_k \\ \nu_k \end{bmatrix} & k > 0 \end{cases} \tag{4.61}$$

where the premultiplication at time $k = 0$ by the various square-root information matrices constrains $\boldsymbol{x}_{\boldsymbol{c}0}$ to have identity covariance.

Based on this definition, the Consider dynamics equation can be constructed. The first component of $\boldsymbol{x}_{\boldsymbol{c}0}$ is related to uncertainty in the initial estimate; it does not enter the system directly at any later times and thus has no dynamic behavior to describe. The process and measurement noise elements, w_k and ν_k, are zero-mean white stochastic random variables. They do not have any true dynamics because samples at different times are uncorrelated. Thus, Φ_{ck} is an appropriately-dimensioned matrix of zeros for all samples k. Note that Φ_{c0} will have fewer rows than columns in order to omit the component of $\boldsymbol{x}_{\boldsymbol{c}0}$ related to initial estimate uncertainty from all later $\boldsymbol{x}_{\boldsymbol{c}k}$ Consider states. Even though the dynamics of w_k and ν_k are trivial, the Consider dynamics equation is used to describe their "truth" statistical behavior by means of the Γ_{cck} matrix, which applies both at $k = 0$ and later sample times:

$$\Gamma_{cck} = \begin{bmatrix} \mathcal{R}^{-1}_{wwk+1} & 0 \\ 0 & \mathcal{R}^{-1}_{\nu\nu k+1} \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 1.5 \end{bmatrix} \tag{4.62}$$

No unestimated disturbances enter the state dynamics equation, so Γ_{xck} is a matrix of zeros. In the measurement equation, the matrix H_{ck} is a block matrix which extracts the measurement noise component of $\boldsymbol{x_{ck}}$:

$$H_{ck} = \begin{cases} \begin{bmatrix} 0 & 0 & \mathcal{R}_{\nu\nu 0}^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1.5 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & 1 \end{bmatrix} & k > 0 \end{cases} \tag{4.63}$$

The first component of $\boldsymbol{x_{ck}}$ (or the second component when $k = 0$) is the "truth" system process noise. The Consider analysis knows how the "truth" process noise enters the filter's process noise information equation, Eq. (4.17). It designs S_{wck} to yield the "truth" process noise covariance, taking into account the Consider dynamics model:

$$S_{wck} = \begin{cases} \begin{bmatrix} 0 & -\left(\mathcal{R}_{fww0}\mathcal{R}_{ww0}^{-1}\right) & 0 \end{bmatrix} = \begin{bmatrix} 0 & -0.5 & 0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} -\mathcal{R}_{fwwk} & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \end{bmatrix} & k > 0 \end{cases} \tag{4.64}$$

Finally, the sensitivity matrix $\bar{S}_{xc0}$ effectively contains the information about the way in which the "truth" initial estimation error covariance differs from what the filter assumes. In this example, of course, there is no difference.

$$\bar{S}_{xc0} = \begin{bmatrix} \left(\bar{\mathcal{R}}_{fxx0}\bar{\mathcal{R}}_{xx0}^{-1}\right) & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{4.65}$$

Monte Carlo Simulations

At this point, all the Consider form matrices have been defined, and the algorithms of Section 4.4 can be applied to compute both filter and true covariances analytically. True covariance can also be estimated numerically by means of Monte Carlo simulations. This technique computes sample mean and covariance from a large number of simulated estimation errors, which can be thought

of as samples of a random process. As the number of trials increases, the estimated covariance approaches the true covariance. This chapter's examples each use 5000 trials in their Monte Carlo simulations.

To implement the simulations, a "truth" model is created for the scenario based on the defined system matrices and "truth" noise covariances. The "truth" model takes as inputs the initial state and histories of measurement and process noise, which are constrained to have the statistical properties specified in the example. It outputs a history of the state vector and a series of noisy measurements. None of the "truth" model calculations use the special Consider model form developed in the previous subsection. These "truth" measurements are fed into the incorrectly-modeled filter, which outputs histories of state estimates. By differencing the "truth" model's states and the filter's state estimates, a history of estimation error is generated.

This entire procedure is repeated many times, each with different samples from the same "truth" distributions of the initial state, process noise, and measurement noise. The result is a large number of estimation error histories, each of which is an independent instance of the same random process. Sample means, covariances, and standard deviations are computed at each time step, and compared with the analytical results from a single run of the Consider analysis.

The Monte-Carlo covariance matrices agree closely with the covariances computed by the Consider analysis. This result holds for the elements corresponding to both the position and velocity states, as well as the cross-covariances of position and velocity. Additionally, confidence bounds on the estimated covariances have been computed, and it has been verified that the

differences between the Monte Carlo and Consider analysis covariances generally fall well within the expected bounds.

To illustrate, Fig. 4.1 plots three related quantities: The standard deviations of error in the *a posteriori* position estimates as computed by the incorrect filter, by the Consider analysis, and by the Monte Carlo simulations. In this figure, the

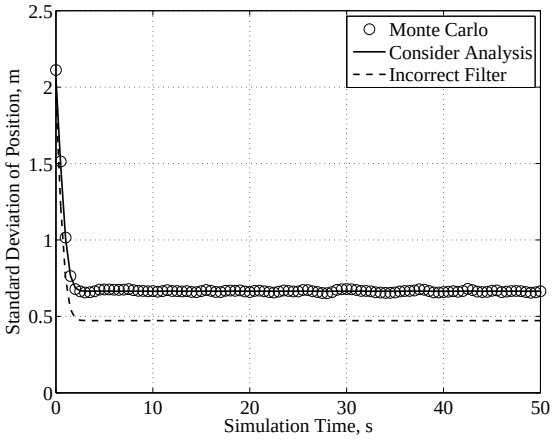

Figure 4.1: Standard deviations of *a posteriori* position error for incorrect noise example.

circles representing the Monte Carlo estimates of standard deviations clearly follow the solid line of the analytical Consider standard deviations. Both are greater than the filter's computed standard deviations by an approximately-constant offset after the filter reaches a steady state. This offset represents the additional error one could expect in the estimates produced by a filter that used incorrect noise covariances as described. Note that the Monte Carlo confidence bounds are not shown on the plot because they tend to add clutter.

Although filter modeling errors in general are likely to *increase* the estimation error standard deviations as shown in Fig. 4.1, this is not always the case. In some scenarios, the true covariance could actually be lower than the filter's computed covariance. This is in fact the case for the standard deviations asso-

ciated with the *a priori* velocity estimates for this example. The reason for this is simple: the "truth" process noise intensity is lower than that assumed by the filter, so immediately after a dynamic propagation step the true estimation error standard deviations are lower than those predicted by the filter. After each measurement update, however, the higher-than-modeled measurement noise intensity reverses the situation so that the true *a posteriori* standard deviations are higher than those of the filter. Even though the filter would have a lower *a priori* velocity estimation error than predicted, its estimates would be computed based on suboptimal filter gains. In other words, the true error would be lower than the filter predicts but not as low as it could be with an optimal filter.

It is also important to note that while Monte Carlo simulations can yield the same covariance results as the Consider analysis and are conceptually simpler to program, they incur a far greater computational burden. Each Monte Carlo trial must re-run the entire "truth" model and filter, and a large number of trials may be required for sufficient accuracy. In contrast, the Consider analysis runs just once, with only moderate computational increase beyond that of a single filter run. The Consider analysis also has no need to compute the specific estimates associated with a particular measurement history realization, and so does not require a "truth"-model run to produce measurements.

4.5.2 Example: Incorrect System Matrices

Scenario Description and Consider Form Setup

The filter used in the second example is identical to that of the first example. As before, the dynamics and measurement models are given by Eqs. (4.56a)

and (4.56b), with the same position and velocity states r_k and v_k and time step $\Delta t = 0.5$ s. The initial estimate and covariance of Eq. (4.57) are also used, and the nominal process and measurement noise covariances are those of Eq. (4.58). From these equations, the filter matrices Φ_{fk}, Γ_{fk}, H_{fk}, $\bar{\mathcal{R}}_{fxx0}$, and $\mathcal{R}_{fwwk}$ can be extracted.

In contrast to the previous example, the "truth" process and measurement noise covariances are both identical to those of the filter, such that $Q_k = Q_{fk}$ and $R_k = R_{fk}$. Once again, the filter's initial covariance $\overline{P}_{fxx0}$ is assumed to be correct.

The "truth" model system equations, however, have slightly different matrices than those assumed by the filter, such that $\Phi_k \neq \Phi_{fk}$, $\Gamma_k \neq \Gamma_{fk}$, and $H_k \neq H_{fk}$. The "truth" dynamics and measurement models are given by

$$\begin{bmatrix} r_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} 0.95 & 1.01\Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_k \\ v_k \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.9 \end{bmatrix} w_k \tag{4.66a}$$

$$y_k = \begin{bmatrix} 0.95 & 1.05 \end{bmatrix} \begin{bmatrix} r_k \\ v_k \end{bmatrix} + \nu_k \tag{4.66b}$$

The generalized Consider system defined by Eqs. (4.15a) and (4.15b) writes the "truth" system dynamics as a perturbed version of the filter's dynamics equation, and writes the "truth" system measurement model as a perturbed version of the filter's measurement model. To carry out this procedure for the current example, one need only see what terms must be added to the filter system equations in order to make them equal the "truth" system equations. Specifically,

$$\boldsymbol{x}_{k+1} = \Phi_{fk}\boldsymbol{x}_k + \Gamma_{fk}\boldsymbol{w}_k + (\Phi_k - \Phi_{fk})\,\boldsymbol{x}_k + (\Gamma_k - \Gamma_{fk})\,\boldsymbol{w}_k \tag{4.67a}$$

$$\boldsymbol{y}_k = H_{fk}\boldsymbol{x}_k + \boldsymbol{\nu}_k + (H_k - H_{fk})\,\boldsymbol{x}_k \tag{4.67b}$$

The perturbation to the filter's dynamics must next be written as a sum of a zero-mean stochastic part, $\Gamma_{xck}\boldsymbol{x}_{\boldsymbol{c}k}$, and a deterministic time-varying bias part $\boldsymbol{b}_{xk}$. Likewise, the perturbation to the filter's measurement equation must be decomposed into a zero-mean stochastic part, $H_{ck}\boldsymbol{x}_{\boldsymbol{c}k}$, and a deterministic time-varying bias part $\boldsymbol{b}_{yk}$.

A significant difficulty for the current example arises because both perturbations depend on $\boldsymbol{x}_k$. This state vector is driven by the stochastic noise $\boldsymbol{w}_k$, but has a time-varying, deterministic, non-zero mean due to non-zero $\bar{\boldsymbol{x}}_0$. The effect of the non-zero mean can be resolved by recognizing that $\boldsymbol{x}_k$ can be written as

$$\boldsymbol{x}_k = \tilde{\boldsymbol{x}}_k + \boldsymbol{x}_{bk} \tag{4.68}$$

where $\tilde{\boldsymbol{x}}_k$ is the zero-mean, stochastic part of $\boldsymbol{x}_k$ driven by $\boldsymbol{w}_k$, and $\boldsymbol{x}_{bk}$ is the deterministic part of $\boldsymbol{x}_k$. The component $\boldsymbol{x}_{bk}$ can be pre-computed for any sample k by propagating the filter's initial estimate (which can be thought of as $\boldsymbol{x}_{b0} = \bar{\boldsymbol{x}}_0$) through the "truth" dynamics equation while setting the process noise to zero. In other words,

$$\boldsymbol{x}_{bk} = \Phi_{k-1}\Phi_{k-2}\cdots\Phi_0\boldsymbol{x}_{b0} \tag{4.69}$$

The stochastic component $\tilde{\boldsymbol{x}}_k$, on the other hand, is initially drawn from a zero-mean distribution with covariance specified by $\overline{P}_{xx0}$. It has dynamics modeled by the system's "truth" dynamics equation:

$$\tilde{\boldsymbol{x}}_{k+1} = \Phi_k\tilde{\boldsymbol{x}}_k + \Gamma_k\boldsymbol{w}_k \tag{4.70}$$

It is now possible to write the perturbation to the filter's dynamics as a random part, $(\Phi_k - \Phi_{fk})\,\tilde{\boldsymbol{x}}_k + (\Gamma_k - \Gamma_{fk})\,\boldsymbol{w}_k$, and a deterministic part, $(\Phi_k - \Phi_{fk})\,\boldsymbol{x}_{bk}$. The corresponding random part of the measurement model's perturbation is $(H_k - H_{fk})\,\tilde{\boldsymbol{x}}_k$, and its deterministic part is $(H_k - H_{fk})\,\boldsymbol{x}_{bk}$.

The Consider vector $\boldsymbol{x}_{ck}$ must capture the dynamical and statistical behavior of the random parts of the system. In this example, those random parts include process noise, measurement noise, initial uncertainty, and a copy of the random part of $\boldsymbol{x}_k$ itself. A suitable choice for the Consider vector is therefore:

$$\boldsymbol{x}_{ck} \equiv \begin{cases} \begin{bmatrix} \bar{\mathcal{R}}_{xx0}\tilde{\boldsymbol{x}}_0 \\ \mathcal{R}_{ww0}w_0 \\ \mathcal{R}_{\nu\nu0}\nu_0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} \tilde{\boldsymbol{x}}_k \\ w_k \\ \nu_k \end{bmatrix} & k > 0 \end{cases} \tag{4.71}$$

As before, the Consider vector at sample $k = 0$ is defined such that it has identity covariance. In contrast to the preceding example, the Consider state at sample $k = 0$ has no more elements than the Consider states at all samples $k > 0$. With this definition of $\boldsymbol{x}_{ck}$, one can write the matrices for the Consider dynamics equation for $k > 0$:

$$\Phi_{ck} = \begin{bmatrix} \Phi_k & \Gamma_k & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.95 & 1.01\Delta t & 0.1 & 0 \\ 0 & 1 & 0.9 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.72a}$$

$$\Gamma_{cck} = \begin{bmatrix} 0 & 0 \\ \mathcal{R}^{-1}_{wwk+1} & 0 \\ 0 & \mathcal{R}^{-1}_{\nu\nu k+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{4.72b}$$

Note how the Consider dynamics matrices model the dynamic behavior of $\tilde{\boldsymbol{x}}_k$, the statistical relationship between $\tilde{\boldsymbol{x}}_k$ and w_k, and the covariances of w_k and ν_k. This definition of the Consider vector allows the Consider dynamics equation to model the "truth" system dynamics, which the filter cannot do with its incorrect system matrices Φ_{fk} and Γ_{fk}. At $k = 0$, the Γ_{cck} matrix is unchanged, but Φ_{ck}

becomes

$$\Phi_{c0} = \begin{bmatrix} \Phi_0 \bar{\mathcal{R}}_{xx0}^{-1} & \Gamma_0 \mathcal{R}_{ww0}^{-1} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.95\sqrt{10} & 1.01\sqrt{5}\Delta t & 0.1 & 0 \\ 0 & \sqrt{5} & 0.9 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.73}$$

in order to transition from the original identity-covariance Consider state vector at $k = 0$ to its new, non-identity-covariance form for $k > 0$.

In contrast to the previous example, Γ_{xck} is not a matrix of zeros, and $\boldsymbol{b}_{xk}$ is not a vector of zeros. The matrix Γ_{xck}, in combination with the defined $\boldsymbol{x}_{\boldsymbol{c}k}$, specifies the random part of the perturbation to the filter's dynamics equation. This matrix is given by

$$\Gamma_{xck} = \begin{bmatrix} (\Phi_k - \Phi_{fk}) & (\Gamma_k - \Gamma_{fk}) & 0 \end{bmatrix} = \begin{bmatrix} -0.05 & 0.01\Delta t & 0.1 & 0 \\ 0 & 0 & -0.1 & 0 \end{bmatrix} \tag{4.74}$$

for $k > 0$ and by

$$\Gamma_{xc0} = \begin{bmatrix} (\Phi_0 - \Phi_{f0}) \bar{\mathcal{R}}_{xx0}^{-1} & (\Gamma_0 - \Gamma_{f0}) \mathcal{R}_{ww0}^{-1} & 0 \end{bmatrix} = \begin{bmatrix} -0.05\sqrt{10} & 0.01\sqrt{5}\Delta t & 0.1 & 0 \\ 0 & 0 & -0.1 & 0 \end{bmatrix} \tag{4.75}$$

for the special case of $k = 0$.

The deterministic disturbance $\boldsymbol{b}_{xk}$ can be computed in advance by employing the formula

$$\boldsymbol{b}_{xk} = (\Phi_k - \Phi_{fk}) \boldsymbol{x}_{bk} = (\Phi_k - \Phi_{fk}) \Phi_{k-1} \Phi_{k-2} \cdots \Phi_0 \boldsymbol{x}_{b0} \tag{4.76}$$

where $\boldsymbol{x}_{b0}$ is the initial estimate $\begin{bmatrix} 3 & 1 \end{bmatrix}^{\mathrm{T}}$.

Just as Γ_{xck} models the sensitivity of the filter's dynamics equation to $\boldsymbol{x}_{\boldsymbol{c}k}$, the Consider measurement sensitivity matrix H_{ck} models the way in which $\boldsymbol{x}_{\boldsymbol{c}k}$ perturbs the filter's measurements. Additionally, it must include the effects of

the correctly-modeled measurement noise ν_k. This matrix is given by

$$H_{ck} = \begin{cases} \begin{bmatrix} (H_0 - H_{f0})\,\bar{\mathcal{R}}_{xx0}^{-1} & 0 & \mathcal{R}_{\nu\nu 0}^{-1} \end{bmatrix} = \begin{bmatrix} -0.05\sqrt{10} & 0.05\sqrt{5} & 0 & 1 \end{bmatrix} & k = 0 \\ \begin{bmatrix} (H_k - H_{fk}) & 0 & 1 \end{bmatrix} = \begin{bmatrix} -0.05 & 0.05 & 0 & 1 \end{bmatrix} & k > 0 \end{cases} \tag{4.77}$$

The effect of the deterministic part of $\boldsymbol{x}_k$ on the measurements is the Consider measurement bias $\boldsymbol{b}_{yk}$, which is just

$$\boldsymbol{b}_{yk} = (H_k - H_{fk})\,\boldsymbol{x}_{bk} = (H_k - H_{fk})\,\Phi_{k-1}\Phi_{k-2}\cdots\Phi_0\boldsymbol{x}_{b0} \tag{4.78}$$

The matrix S_{wck} selects the components of $\boldsymbol{x}_{\boldsymbol{c}k}$ related to process noise. Because the "truth" and modeled process noise are identical in this example, it is relatively simple:

$$S_{wck} = \begin{cases} \begin{bmatrix} 0 & -\left(\mathcal{R}_{fww0}\mathcal{R}_{ww0}^{-1}\right) & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & 0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & -\mathcal{R}_{fwwk} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & 0 \end{bmatrix} & k > 0 \end{cases} \tag{4.79}$$

Finally, the matrix $\bar{S}_{xc0}$ specifies that the initial estimation error covariance is correct.

$$\bar{S}_{xc0} = \begin{bmatrix} -\mathrm{I} & 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \tag{4.80}$$

Likewise, the initial estimation error has a mean of zero, so the bias $\bar{\boldsymbol{b}}_{c0}$ is zero. At this point, the system has been written in Consider form, and the analysis algorithms can proceed in a standard fashion.

Monte Carlo Simulations

Once again, all elements of the Monte Carlo matrix MSEs agree closely with the analytical matrix MSEs found by the Consider analysis, and they fall within the

computed confidence bounds. Figure 4.2, the Example 2 equivalent of Fig. 4.1, shows the root mean square (RMS) estimation errors for the *a posteriori* position estimates. The close agreement between the solid curve and the circles in Fig. 4.2 shows the correctness of the Consider analysis.

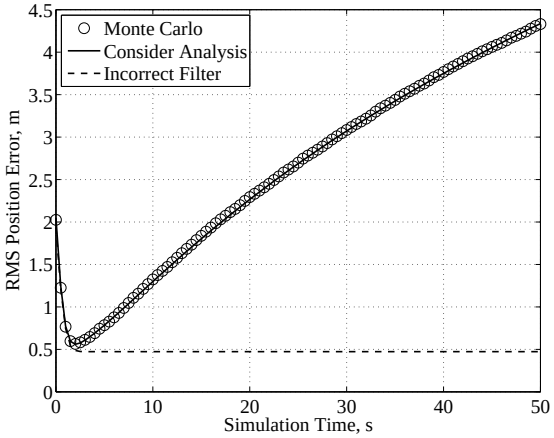

Figure 4.2: RMS *a posteriori* position errors for incorrect system matrices example.

Because this example includes biases, i.e. $\boldsymbol{b}_{xk}$ and $\boldsymbol{b}_{yk}$, the matrix MSE is the Consider analysis quantity of interest. True error covariances could be computed, but they are less interesting than the matrix MSE because they only constitute a portion of the total error.

In the incorrect noise example of Fig. 4.1, the estimation error standard deviations leveled off to a constant value after some transient behavior. For this example, in contrast, the true RMS errors diverge steadily from those reported by the filter. Because the filter is not modeling the correct system dynamics or measurements, it is unable to maintain trustworthy state estimates.

This second example examined filter performance for a particular set of incorrect system matrices. Given the possibility of uncertainty in the "truth" system matrices, a more valuable analysis might compute the distribution of RMS

estimation errors resulting from a chosen statistical distribution of system matrix errors. Such results can be obtained by wrapping an outer Monte Carlo analysis around a Consider covariance analysis. Many different particular "truth" models would be generated for the given filter, and a Consider analysis of each case would produce the corresponding RMS estimation error history. The result would be a Monte Carlo distribution of RMS estimation error histories. This distribution would characterize the performance range of a given filter design. Significant computational savings can be realized from this analysis approach. The Consider covariance analysis that computes the true matrix MSE for each "truth"/filter combination replaces a burdensome inner Monte Carlo analysis. Computational efficiency is thus increased by two or three orders of magnitude. Additional savings are possible by optimizing the Consider analysis algorithms for this application. Many of the calculations are specific to the filter being analyzed and are independent of the "truth" model. By performing such calculations only once rather than repeating them for each Monte Carlo "truth" model trial, approximately 25% of the remaining computational burden can be eliminated.

Note how this example and the preceding example are both linear, consistent with all of this chapter's Consider derivations. It should be possible, however, to extend these techniques to some nonlinear systems, similar to what is done in Chapter 3. The obvious extension would use linearizations of the system dynamics and measurement models. As in the case of an Extended Kalman Filter, such an approach should succeed if neglected higher-order terms are sufficiently small.

4.6 Conclusions

This chapter presents a new form of Consider covariance analysis that can be applied to square-root information filters with many different forms of filter modeling errors. The analysis computes the true estimation error covariance or the true matrix mean square error, whichever is most relevant. It is suitable for studying filter design choices, for analyzing potential filter performance, or for calculating the uncertainty of previously computed estimates after a filter model error has been discovered. The new Consider analysis begins by casting each problem into a predefined general form. This problem form makes it possible for the key algorithms to operate on a wide class of systems without customization.

Two specific examples demonstrate the power of the new method: One with incorrectly modeled noise and the other with erroneous system matrices. For these cases, Monte Carlo simulations provide independent verification of the Consider analysis equations. Comparison of the resulting true and mismodeled root mean square estimation errors illustrates several of the types of degraded filter behavior that can be studied in this manner.

4.7 Appendix: QR and LQ Factorization Calculations

One well-known technique for matrix calculations is QR factorization, which decomposes any given matrix into a product of an orthonormal matrix "Q" on the left and an upper triangular matrix "R" on the right. Both orthonormal and triangular matrices have convenient theoretical and computational properties.

Among other applications, QR factorization is a central feature of the standard SRIF algorithms.

The LQ factorization is also exploited in this chapter's Consider covariance analysis. It decomposes a given matrix into a lower triangular matrix "L" on the left and an orthonormal matrix "Q" on the right.

The computational routines necessary to perform QR factorization are available in many software packages including MATLAB; they will not be discussed here. The less-common LQ factorization can be obtained from a QR-factorization routine by implementing the following pseudocode, written with MATLAB-style syntax:

$$
\begin{aligned}
&\texttt{function}\ [L, Q] = \texttt{lq}(A) \\
&\qquad [\tilde{Q}, \tilde{R}] = \texttt{qr}\left(A^{\mathrm{T}}\right) \\
&\qquad Q = \tilde{Q}^{\mathrm{T}} \\
&\qquad L = \tilde{R}^{\mathrm{T}} \\
&\texttt{end function}
\end{aligned}
$$

This algorithm yields a lower-triangular matrix L and an orthonormal matrix Q such that $LQ = A$ for any square or rectangular matrix A. Of course, this pseudocode assumes that the statement $[Q, R] = \texttt{qr}(A)$ produces an orthonormal matrix Q and an upper-triangular matrix R such that $QR = A$. The function "`qr`" is standard in MATLAB.

CHAPTER 5

A MULTIPURPOSE SQUARE-ROOT CONSIDER COVARIANCE ANALYSIS FOR LINEAR SMOOTHERS

A new form of Consider covariance analysis for studying incorrectly-modeled square-root information smoothers is presented and demonstrated. The value of this technique is its ability to compute the true estimation error covariance when the smoother has an incorrect dynamics model, an incorrect measurement model, or an incorrect statistical model. The new smoother analysis casts systems with a wide range of possible modeling errors into a special form and exploits square-root techniques to provide both generality and compactness. A Consider covariance analysis can improve smoother design or characterize an existing smoother's true accuracy. Areas of application include incorrect initial state covariance, colored or correlated noise statistics, unestimated disturbance states, and erroneous system matrices. Several simple examples are developed to illustrate the application of the Consider smoother analysis to models with a variety of error types, and these examples are independently validated by Monte Carlo simulation.

5.1 Introduction

Dynamic estimation methods compute two main outputs: Estimates of a system state vector and some metric of uncertainty associated with those estimates. This general statement holds for Kalman filters, square-root information filters (SRIFs), unscented filters, and particle filters, among other filter varieties. Most commonly, estimation error covariance is the chosen uncertainty metric. Filters

typically form estimates and estimation error covariances recursively as measurements arrive, and thus they only use measurements up to and including the sample time at which the estimate applies. Conversely, smoothers employ measurements from both past and future times to compute better estimates and reduce estimation error covariance by "smoothing out" the uncertainty in the system. In particular, fixed-interval smoothers form estimates over an interval based on all the measurements obtained during that interval. When a filter or smoother is based on an incorrect system model, both the estimates and the estimation error covariance may be affected. The estimates may be degraded in some way, and the estimator will report an estimation error covariance that does not correspond to the true level of error.

One approach to the problem of estimators with incorrect models is Consider covariance analysis. The name "Consider covariance analysis" stems from the analysis' ability to "consider" the effects of various kinds of modeling errors on filter and smoother behavior. The analysis does this by computing the "true" estimation error covariance of the mismodeled estimator. Note that the resulting estimation error covariance is *true* relative to a particular mismodeled estimator and a particular assumed "truth" model. It is not the true covariance in an absolute sense unless the "truth" model perfectly describes the real system. Throughout this chapter, the word "Consider" is capitalized when it is used in a technical sense, to distinguish from colloquial usage. This is a departure from previous works on this subject. Many of this chapter's references give the name "covariance analysis" to any analysis that investigates covariances, whether or not the filter or smoother is based on an incorrect system model. Others use the name "sensitivity analysis", because comparison of the optimal and actual estimation error covariances provides a metric of the sensitivity of the system

to modeling error. This chapter uses the phrase "Consider covariance analysis" for all such analyses.

Filter and smoother model errors may result from intentional simplifications, or they may be unavoidable due to poor understanding of the system dynamics or measurements. When simplifications have been made, such as model order reduction or decoupling of dynamics, Consider covariance analysis is valuable as a design tool. For each candidate simplified estimator, a Consider analysis can indicate its expected estimation accuracy. In other situations where the true system characteristics are not well-known, even the most detailed filter or smoother is likely to deviate from the true system significantly. A simpler estimator design, while suboptimal, may be more robust and less sensitive to the potential errors. One can explore these trade-offs by an iterative Consider analysis process, evaluating multiple combinations of proposed filters and hypothetical "truth" models. Consider covariance analysis may even be helpful in determining whether to implement a filter only or a filter/smoother for a given system. In the presence of model mismatch, smoothing may not improve estimation accuracy as much as it otherwise would. Alternatively, depending on the system and type of error, smoothing may eliminate most of the error effects and restore much of the performance of the optimal system.

As defined here, Consider covariance analysis can be used to study the effects of many different classes of model errors. One of the most commonly addressed error classes is that of unmodeled constant biases in the dynamics or measurements [51, 52, 57]. Another function of Consider analysis is to investigate systems with incorrect *a priori* state estimate covariance, incorrect process noise, or incorrect measurement noise [3, 50, 58, 59]. This includes the situation

where the noise is colored rather than white. One may also study systems that have unestimated dynamic states [54, 60, 64, 65]. Finally, some analyses examine the situation where some or all of the state-space system matrices used by the estimator have errors [4, 5, 54, 61–65]. Some methods for Consider covariance analysis are narrow algorithms that focus on a single class of filter errors; others are general approaches that can analyze all of these filter error types simultaneously.

The preceding chapter, Chapter 4, presents a new square-root information form of Consider covariance analysis. This method treats all of the discussed error classes within a single framework. The present chapter extends the Consider covariance analysis of Chapter 4 to analyze a discrete, fixed-interval square-root information smoother (SRIS). In keeping with common smoother practice, this chapter's Consider smoother analysis operates on a Rauch-Tung-Striebel (RTS) square-root information smoother [69]. That is, the analysis requires that there is first a forward filtering pass using SRIF techniques, followed by a backward smoothing pass also using square-root techniques. The Consider smoother analysis uses the same architecture: a forward-pass Consider analysis of the SRIF, followed by a backward-pass Consider analysis of the SRIS. The forward filtering pass is analyzed via the algorithms of Chapter 4. The present chapter develops the backward-pass Consider covariance analysis using data generated during the forward pass, in the spirit of the RTS method.

Several of the previous Consider covariance analyses have investigated the smoother problem. Reference [54] analyzes a continuous-time RTS smoother and allows general model errors, but it does not present an equivalent discrete-time analysis. Both [63] and [70] operate on different non-standard two-filter

smoothers rather than RTS-style smoothers. To the author's knowledge, the only general Consider analysis of a discrete-time RTS smoother is given in Refs. [4, 5]. The derivation, which is performed in the covariance domain, is given only in Ref. [4], and the provided algorithms in both sources are very complicated with many intermediate calculations. While neither of Refs. [4, 5] explicitly allows for unestimated state vector elements or mutually-correlated process and measurement noise, the provided algorithms could be modified by an experienced analyst to accommodate these cases.

All of the above Consider smoother algorithms operate in the covariance domain. In the information domain, Refs. [3, 57] analyze RTS square-root information smoothers, but only for systems with unestimated random biases. These sources do not address smoothers with other types of model error, such as unestimated dynamic disturbances, incorrect noise covariances, or erroneous system matrices.

This chapter makes two main contributions. First, the new Consider covariance analysis of Chapter 4 is extended to analyze fixed-interval smoothers with incorrect system models. The resulting smoother analysis is the only generalized Consider covariance analysis for discrete RTS square-root information smoothers. The analysis derivation, which is analogous to the derivation of the SRIS equations, is believed to be more understandable than that of Ref. [4], and the final form is simpler and more compact. Second, this chapter provides several simple, concrete numerical examples beyond those of Chapter 4. These examples demonstrate how the smoother analysis procedure can be applied to common modeling error situations. Some of these types of errors, such as non-independent process and measurement noise, cannot be handled by the algo-

rithms of Refs. [4, 5] without significant pre-processing/model augmentation. Monte Carlo techniques validate the Consider analyses of these examples.

Implementation of this chapter's algorithms requires two building blocks. The first building block is the Consider filter analysis of Chapter 4, which is used for the forward-filtering-pass stage of the present chapter's analysis. The second building block is the basic method of square-root information smoothing. Readers unfamiliar with SRIS techniques are encouraged to refer to Refs. [3, 71] for more background on the subject.

The remainder of this chapter is organized as follows: Section 5.2 summarizes some of the results of Chapter 4 that are needed for this chapter's analysis. In Section 5.3, the algorithms for the Consider covariance analysis of a smoother backward pass are developed and explained. Section 5.4 discusses several interesting example problems and validates their Consider analyses using Monte Carlo simulations. Conclusions are drawn in Section 5.5.

5.2 Background: Consider System Model & Forward-Pass Filter Analysis

A number of key ideas and results from Chapter 4 are necessary for the development of this chapter's smoothing algorithms. These concepts are briefly reviewed in this section. Section 5.2.1 describes a special "Consider form" of the system equations. Select equations from the forward-pass filter analysis are presented in Section 5.2.2.

5.2.1 Consider Form

The Consider analysis developed in this chapter and in Chapter 4 begins by casting each problem into a special, pre-defined Consider form. The algorithms of Section 5.3 operate directly on matrices and equations in this special form without regard to how the problem statement was originally framed. As a result, a wide variety of model errors can be studied without modifying the core algorithms.

The philosophy of the Consider form is simple. For each set of filter/smoother modeling errors, one can write both the estimator's assumed model equations and the "truth" model equations. All but one Consider form equation is just a version of the corresponding "truth" equation, rewritten as the filter/smoother equation with modified noise terms or an additive perturbation. The equation pairs are all listed for convenience in Table 5.1, which shows the parallel structure of the two model formulations. The filter/smoother equations in

Table 5.1: Consider system model summary.

Equation	Filter/Smoother Version		Consider Version
State dynamics	$\boldsymbol{x}_{k+1} = \Phi_{fk}\boldsymbol{x}_k + \Gamma_{fk}\boldsymbol{w}_k$	$\Rightarrow$	$\boldsymbol{x}_{k+1} = \Phi_{fk}\boldsymbol{x}_k + \Gamma_{fk}\boldsymbol{w}_k + \Gamma_{xck}\boldsymbol{x_c}_k + \boldsymbol{b}_{xk}$
Measurement	$\boldsymbol{y}_k = H_{fk}\boldsymbol{x}_k + \boldsymbol{\nu}_k$	$\Rightarrow$	$\boldsymbol{y}_k = H_{fk}\boldsymbol{x}_k + H_{ck}\boldsymbol{x_c}_k + \boldsymbol{b}_{yk}$
Consider dynamics			$\boldsymbol{x_c}_{k+1} = \Phi_{ck}\boldsymbol{x_c}_k + \Gamma_{cck}\boldsymbol{w_c}_k$
Process noise information	$\mathcal{R}_{fwwk}\boldsymbol{w}_k = -\boldsymbol{\nu}_{wk}$	$\Rightarrow$	$\mathcal{R}_{fwwk}\boldsymbol{w}_k = -S_{wck}\boldsymbol{x_c}_k - \boldsymbol{b}_{wk}$
Initial state information	$\bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_0 = \bar{\boldsymbol{z}}_0 - \bar{\boldsymbol{\nu}}_{x0}$	$\Rightarrow$	$\bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_0 = \bar{\boldsymbol{z}}_0 - \bar{S}_{xc0}\boldsymbol{x_c}_0 - \bar{\boldsymbol{b}}_{c0}$

the center column of Table 5.1 are those of a standard square-root information filter/smoother (SRIF/S). In the dynamics equation, the state vector $\boldsymbol{x}_k$ evolves as dictated by the smoother's state transition matrix Φ_{fk} and process noise influence matrix Γ_{fk}. The dynamics are driven by the stochastic process noise vector $\boldsymbol{w}_k$. Likewise, the smoother's assumed measurement sensitivity matrix

H_{fk} is standard. Here and throughout this chapter, the subscript "f " designates the versions of the given quantities assumed by the filter/smoother, and symbols without this subscript represent the corresponding "truth" quantities. This notation is chosen to be consistent with that of Chapter 4. Without loss of generality, the smoother's measurement equation is assumed to have been normalized so that the filter models $\boldsymbol{\nu}_k$ as a zero-mean, identity-covariance, white, Gaussian random vector, which is uncorrelated with the process noise. The smoother's assumed statistics for the process noise $\boldsymbol{w}_k$ are captured in the third equation of the center column, the process noise information equation. Its matrix $\mathcal{R}_{fwwk}$ is the inverse-square-root of the smoother's assumed process noise covariance matrix Q_{fk}. Because the process noise is also assumed to be Gaussian and white with zero mean, the process noise error term $\boldsymbol{\nu}_{wk}$ has zero mean and identity covariance. Finally, the fourth equation in the center column of Table 5.1 parameterizes the filter/smoother's initial estimation uncertainty. The vector $\bar{\boldsymbol{z}}_0$ is the *a priori* information state at sample $k = 0$, and $\bar{\mathcal{R}}_{fxx0}$ is the *a priori* inverse-square-root of the assumed initial estimation error covariance $\overline{P}_{fxx0}$. By construction, the state information error $\bar{\boldsymbol{\nu}}_{x0}$ is assumed to be zero-mean, identity-covariance, and Gaussian.

The Consider dynamics equation, the third equation in the right-hand column, is unique to the new Consider system form; it has no filter/smoother counterpart in the center column of Table 5.1. It describes the behavior of the "Consider state" $\boldsymbol{x}_{\boldsymbol{c}k}$ using its own state transition matrix Φ_{ck}, process noise influence matrix Γ_{cck}, and Consider process noise $\boldsymbol{w}_{\boldsymbol{c}k}$. The Consider vector $\boldsymbol{x}_{\boldsymbol{c}k}$ captures in one quantity all the sources of randomness and uncertainty contained within the "truth" system. This typically includes the "truth" process and measurement noise, as well as the "truth" initial state uncertainty at sam-

ple $k = 0$. Also included in the Consider vector are stochastic quantities that are not modeled by the filter/smoother, such as unmodeled zero-mean stochastic biases and unmodeled dynamic states that influence the "truth" system dynamics. It is possible to model quantities such as biases or white noise sequences using the Consider dynamics equation by including rows of zeros in Γ_{cck} or in Φ_{ck}. The Consider state transition matrix Φ_{ck} need not be invertible or even square, and this flexibility allows for rows of zeros or a changing number of Consider state elements. Without loss of generality, the Consider process noise $\boldsymbol{w}_{\boldsymbol{c}k}$ is constrained to be zero-mean, identity-covariance, white, and Gaussian. The Consider analysis algorithms require, again without loss of generality, that the initial Consider vector $\boldsymbol{x}_{\boldsymbol{c}0}$ be zero-mean, identity-covariance, Gaussian, and uncorrelated with the $\boldsymbol{w}_{\boldsymbol{c}k}$ noise sequence.

Each of the remaining four Consider equations in the right-hand column of Table 5.1 differs from the corresponding filter/smoother version by its last two terms. These terms have the same structure in all equations: a zero-mean stochastic part plus a deterministic, possibly-time-varying bias part. In this context, the term "bias" is used in its statistical sense, as a known value that perturbs the mean from zero. It is not necessarily a dynamical bias in the sense of being a constant independent of k. To make matters even more confusing, elements of $\boldsymbol{x}_{\boldsymbol{c}k}$ sometimes represent constant biases that are sampled from zero-mean distributions, as will be apparent in one of the examples.

Each zero-mean stochastic component in the Consider model is written as a coefficient matrix multiplying the Consider state vector $\boldsymbol{x}_{\boldsymbol{c}k}$. As previously stated, $\boldsymbol{x}_{\boldsymbol{c}k}$ contains all the sources of randomness and uncertainty in the "truth" system. The coefficient matrices Γ_{xck}, H_{ck}, S_{wck}, and $\bar{S}_{xc0}$ have two roles. They

select those portions of $\boldsymbol{x}_{ck}$ relating to a particular equation's noise terms, and they weight the stochastic perturbation appropriately so that the "truth" noise or error enters the equation in place of the assumed uncertainty. The deterministic biases $\boldsymbol{b}_{xk}$, $\boldsymbol{b}_{yk}$, $\boldsymbol{b}_{wk}$, and $\bar{\boldsymbol{b}}_{c0}$ do not depend on a particular noise or state estimate history, and can be precomputed. Among other roles, they allow the user to Consider non-zero-mean noise.

An intuitive understanding of the Consider form and the Consider state vector is most easily gained by studying the examples in Section 5.4 and Chapter 4. These examples demonstrate how specific filter/smoother error classes can be represented using the equations of Table 5.1. In particular, one of the examples in Chapter 4 shows how to handle the case where the "truth" system matrices Φ_k, Γ_k, and H_k differ from those assumed by the filter/smoother. This analysis is possible even though the Consider form equations explicitly contain only the assumed system matrices Φ_{fk}, Γ_{fk}, and H_{fk}.

5.2.2 Needed Outputs from Forward-Pass Consider Filter Analysis

In a traditional RTS smoother, select quantities computed during the forward filter pass must be stored for later use in the backward smoother pass. The Consider smoother analysis likewise requires some matrices and vectors from the forward Consider filter analysis to complete its algorithm. Specifically, two sets of quantities are required: There is a set of quantities related to the Consider state vector $\boldsymbol{x}_{ck}$ and the way it propagates through the Consider analysis, and a set that is the Consider-analysis analog of the quantities stored for a tradi-

tional SRIF/S. The matrices Φ_{fk}, Γ_{fk}, and Γ_{xck} and the vector bias $\boldsymbol{b}_{xk}$ from the Consider-form state dynamics equation of Table 5.1 are also needed.

During the forward filter pass, the Consider analysis must store matrices related to the Consider state vector $\boldsymbol{x}_{\boldsymbol{c}k}$. Within the filter analysis, $\boldsymbol{x}_{\boldsymbol{c}k}$ is replaced by a new vector $\begin{bmatrix}\boldsymbol{\alpha}_k^{\mathrm{T}} & \boldsymbol{\beta}_k^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$, which satisfies the relationship:

$$\boldsymbol{x}_{\boldsymbol{c}k} = \begin{bmatrix} L_{\alpha k} & L_{\beta k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} = L_{\alpha k}\boldsymbol{\alpha}_k + L_{\beta k}\boldsymbol{\beta}_k \tag{5.1}$$

The coefficient matrices $L_{\alpha k}$ and $L_{\beta k}$ are computed recursively at each filter sample k, and are required as inputs for this chapter's smoother analysis algorithms. The vectors $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ are never explicitly computed. Rather, they are mathematical constructs with certain useful properties. Equation (5.1) gives the Consider vector $\boldsymbol{x}_{\boldsymbol{c}k}$ as a linear combination of the two vectors $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$, but the relationship cannot be inverted. The quantities $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ do not depend on $\boldsymbol{x}_{\boldsymbol{c}k}$ alone, but rather on the entire Consider state history up to sample k: $\boldsymbol{x}_{\boldsymbol{c}0}, \boldsymbol{x}_{\boldsymbol{c}1}, \ldots, \boldsymbol{x}_{\boldsymbol{c}k}$. The vector $\boldsymbol{\alpha}_k$ can be thought of as the part of the Consider state history that has an effect on the estimate of the state $\boldsymbol{x}_k$ at sample k. Conversely, the vector $\boldsymbol{\beta}_k$ is the part of the Consider state history that has no effect on the estimate of $\boldsymbol{x}_k$ at sample k, while having potential to affect the filter's estimates at later sample times.

An advantage of this representation is compactness. By construction, the vector $\boldsymbol{\alpha}_k$ has at most n_x elements, the number of elements in the state vector $\boldsymbol{x}_k$. Similarly, the combined vector $\begin{bmatrix}\boldsymbol{\alpha}_k^{\mathrm{T}} & \boldsymbol{\beta}_k^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ has at most $n_x + n_{xck}$ elements, where n_{xck} is the number of elements in $\boldsymbol{x}_{\boldsymbol{c}k}$. Thus, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ parameterize the error effects of the Consider state history without requiring an increasing number of storage elements at every new sample.

One other significant feature makes $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ the preferred Consider analysis variables: The composite vector formed from these quantities and the Consider process noise $\boldsymbol{w}_{ck}$ has zero mean and identity covariance.

$$\begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{ck} \end{bmatrix} \sim \mathcal{N}(0, \mathrm{I}) \tag{5.2}$$

The composite vector in Eq. (5.2) is propagated according to:

$$\begin{bmatrix} \boldsymbol{\alpha}_{k+1} \\ \boldsymbol{\beta}_{k+1} \\ \boldsymbol{\gamma}_{k+1} \end{bmatrix} = C_{k+1} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{ck} \end{bmatrix} \tag{5.3}$$

where the matrix C_{k+1} is orthonormal, and is used extensively in the smoother algorithms. The vector $\boldsymbol{\gamma}_{k+1}$ is not needed in further calculations. This propagation equation is derived in the filter analysis from the *a posteriori* state information equation and from the dynamics of the Consider vector $\boldsymbol{x}_{ck}$.

The second set of stored filter analysis quantities is analogous to the set of quantities stored by a standard SRIF/S. Fixed-interval RTS square-root information smoothing algorithms run backwards from $k = N$, the final sample, to $k = 0$. This backward recursion relies on the post-dynamic-propagation version of the process noise information equation. This chapter's Consider smoother derivations require the Consider version of this propagated process noise information equation at every sample k. It takes the form:

$$\bar{\mathcal{R}}_{wwk}\boldsymbol{w}_k + \bar{\mathcal{R}}_{wxk+1}\boldsymbol{x}_{k+1} = \bar{\boldsymbol{z}}_{wk} - \begin{bmatrix} \bar{S}_{w\alpha k} & \bar{S}_{w\beta k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \end{bmatrix} - \bar{\boldsymbol{b}}_{wk} \tag{5.4}$$

This equation contains information about the process noise $\boldsymbol{w}_k$ that can be inferred from the estimate of the state vector $\boldsymbol{x}_{k+1}$. In the Consider smoother analysis, the matrix coefficients $\bar{S}_{w\alpha k}$ and $\bar{S}_{w\beta k}$ and the bias vector $\bar{\boldsymbol{b}}_{wk}$ also convey

information. They are used to smooth the calculations of the Consider error effects in the smoothed state information equations.

The Consider smoother analysis requires one additional equation from the forward-pass Consider filter analysis: the terminal-sample Consider form of the *a posteriori* state information equation. It is described in the next section in the context of the backward smoothing pass initialization.

5.3 Consider Smoother Analysis Algorithms

The Consider smoother analysis is based on a modification of a standard SRIF/S. This type of modification already has been portrayed in the second, fourth, and fifth rows of Table 5.1. In the center column, the noise terms $\boldsymbol{\nu}_k$, $\boldsymbol{\nu}_{wk}$, and $\bar{\boldsymbol{\nu}}_{x0}$ are each simple, zero-mean, identity-covariance Gaussian random vectors. Each of these simple noise terms is replaced in the right-hand column by its Consider-analysis counterpart, which can have non-zero mean and non-identity covariance. This generalization to arbitrary noise means and covariances flows through the entire Consider smoother analysis.

The Consider analysis derivations for the smoother proceed in a manner that resembles mathematical induction. That is, one begins by defining a particular form for the smoother state information equation at generic sample $k+1$ and by showing that this form holds for some specific value of $k+1$. Next, one shows how to derive the smoother state information equation at preceding sample k by processing the equation at sample $k+1$. This backward transition step also employs the Consider-form model equations and various stored matrix and vector quantities from the forward-pass filter analysis.

5.3.1 Initial Definitions and Setup

At each sample $k+1$, the Consider smoother analysis is assumed to have a state information equation of the generic form:

$$\mathcal{R}^*_{xxk+1}\boldsymbol{x}_{k+1} = \boldsymbol{z}^*_{k+1} - \begin{bmatrix} S^*_{x\alpha k+1} & S^*_{x\beta k+1} & S^*_{x\psi k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{k+1} \\ \boldsymbol{\beta}_{k+1} \\ \boldsymbol{\psi}_{k+1} \end{bmatrix} - \boldsymbol{b}^*_{ck+1} \tag{5.5}$$

Furthermore, the error effects vector, $\begin{bmatrix} \boldsymbol{\alpha}^{\mathrm{T}}_{k+1} & \boldsymbol{\beta}^{\mathrm{T}}_{k+1} & \boldsymbol{\psi}^{\mathrm{T}}_{k+1} \end{bmatrix}^{\mathrm{T}}$, is assumed to have zero mean and identity covariance at sample $k+1$. The task of the Consider smoother analysis derivation is to show how one can obtain a sample-k version of Eq. (5.5). This sample-k version must have an error effects vector $\begin{bmatrix} \boldsymbol{\alpha}^{\mathrm{T}}_{k} & \boldsymbol{\beta}^{\mathrm{T}}_{k} & \boldsymbol{\psi}^{\mathrm{T}}_{k} \end{bmatrix}^{\mathrm{T}}$ with zero mean and identity covariance.

The form of Eq. (5.5) is non-obvious and merits some discussion. The Consider version of the filter's *a posteriori* state information equation for sample $k+1$, as given in Chapter 4, takes the form:

$$\widehat{\mathcal{R}}_{xxk+1}\boldsymbol{x}_{k+1} = \hat{\boldsymbol{z}}_{k+1} - S_{x\alpha k+1}\boldsymbol{\alpha}_{k+1} - \hat{\boldsymbol{b}}_{ck+1} \tag{5.6}$$

This state information equation resembles the Consider form of the initial state information equation in Table 5.1, except that the stochastic part of its error has been written in terms of $\boldsymbol{\alpha}_{k+1}$ rather than $\boldsymbol{x}_{\boldsymbol{c}k+1}$. Recall that the vector $\boldsymbol{\alpha}_{k+1}$ captures the estimation error effects of the entire Consider state history up to and including sample $k+1$, i.e., the effects of $\boldsymbol{x}_{\boldsymbol{c}0}, \boldsymbol{x}_{\boldsymbol{c}1}, \ldots, \boldsymbol{x}_{\boldsymbol{c}k+1}$. Thus, Eq. (5.6) is adequate for the forward filter stage of the Consider analysis. The principles of smoothing, however, dictate that the smoothed estimate of $\boldsymbol{x}_{k+1}$ may be degraded by the uncertainty contained in the Consider state history over the entire interval $\boldsymbol{x}_{\boldsymbol{c}0}, \boldsymbol{x}_{\boldsymbol{c}1}, \ldots, \boldsymbol{x}_{\boldsymbol{c}N}$. This includes the effects up to sample $k+1$

represented by $\boldsymbol{\beta}_{k+1}$. It also includes the uncertainty introduced by the Consider process noise vector at later samples, i.e., $\boldsymbol{w}_{\boldsymbol{c}k+1}, \boldsymbol{w}_{\boldsymbol{c}k+2}, \ldots, \boldsymbol{w}_{\boldsymbol{c}N-1}$. By construction, the new vector $\boldsymbol{\psi}_{k+1}$ parameterizes the uncertainty contained in this Consider process noise sequence at samples *after* k.

Before propagating Eq. (5.5) from $k+1$ back to sample k, it is necessary to initialize the induction by showing that the equation is satisfied for some specific $k+1$. With suitable definitions, the $k+1 = N$ version of Eq. (5.6) provides the initialization for Eq. (5.5). Thus, the Consider filter analysis results at the terminal sample serve to initialize the Consider smoother analysis. This choice for the initialization of the backward-smoother Consider pass should not be surprising. In a traditional RTS SRIS, the backward-smoother pass starts from the final *a posteriori* state information equation that is produced by the filter. At the final sample $k+1 = N$, the filter has processed all of the measurements and the filter's estimate is identical to the smoothed estimate.

To show the equivalence of Consider smoother analysis Eq. (5.5) and Consider filter analysis Eq. (5.6) at sample $k+1 = N$, one must first equate the square-root information matrices, information state vectors, and deterministic bias vectors at this final sample time. One must also equate the coefficient matrix for the vector $\boldsymbol{\alpha}_N$:

$$\mathcal{R}^*_{xxN} = \widehat{\mathcal{R}}_{xxN}, \quad \boldsymbol{z}^*_N = \hat{\boldsymbol{z}}_N, \quad \boldsymbol{b}^*_{cN} = \hat{\boldsymbol{b}}_{cN}, \quad S^*_{x\alpha N} = S_{x\alpha N} \tag{5.7}$$

where the notation "$(\,)^*$" designates a smoother-analysis matrix or vector. The resulting version of Eq. (5.5) still has terms containing $\boldsymbol{\beta}_N$ and $\boldsymbol{\psi}_N$ that have no counterpart in filter-analysis Eq. (5.6). To complete the initialization, one sets the coefficient matrix $S^*_{x\beta N}$ to zero. The new vector $\boldsymbol{\psi}_N$ is defined as an empty vector for this terminal sample, and its coefficient matrix $S^*_{x\psi N}$ is defined as an

empty matrix:

$$S^*_{x\beta N} = 0, \quad , \boldsymbol{\psi}_N = [\,], \quad S^*_{x\psi N} = [\,] \tag{5.8}$$

The definition of $\boldsymbol{\psi}_N$ as an empty vector also satisfies the identity-covariance requirement for $k+1 = N$. The composite vector has the correct initial distribution because $\begin{bmatrix}\boldsymbol{\alpha}_N^{\mathrm{T}} & \boldsymbol{\beta}_N^{\mathrm{T}} & \boldsymbol{\psi}_N^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} = \begin{bmatrix}\boldsymbol{\alpha}_N^{\mathrm{T}} & \boldsymbol{\beta}_N^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$, and $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ were previously defined by the filter analysis to be uncorrelated with identity covariances for all samples k.

5.3.2 Backwards Propagation

The backwards propagation of Eq. (5.5) from sample $k+1$ to sample k smoothes the estimates by incorporating the process noise information of Eq. (5.4). Both are repeated here for convenience:

$$\mathcal{R}^*_{xxk+1}\boldsymbol{x}_{k+1} = \boldsymbol{z}^*_{k+1} - \begin{bmatrix} S^*_{x\alpha k+1} & S^*_{x\beta k+1} & S^*_{x\psi k+1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_{k+1}\\ \boldsymbol{\beta}_{k+1}\\ \boldsymbol{\psi}_{k+1}\end{bmatrix} - \boldsymbol{b}^*_{ck+1} \tag{5.9a}$$

$$\bar{\mathcal{R}}_{wwk}\boldsymbol{w}_k + \bar{\mathcal{R}}_{wxk+1}\boldsymbol{x}_{k+1} = \bar{\boldsymbol{z}}_{wk} - \begin{bmatrix}\bar{S}_{w\alpha k} & \bar{S}_{w\beta k}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_k\\ \boldsymbol{\beta}_k\end{bmatrix} - \bar{\boldsymbol{b}}_{wk} \tag{5.9b}$$

All vectors with index $k+1$ must be eliminated in these two equations in favor of their sample-k versions. In particular, $\boldsymbol{\alpha}_{k+1}$, $\boldsymbol{\beta}_{k+1}$, $\boldsymbol{\psi}_{k+1}$, and $\boldsymbol{x}_{k+1}$ must be replaced.

The zero-mean, identity covariance $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ have been propagated between samples by means of Eq. (5.3). This equation also involves the Consider process noise vector $\boldsymbol{w}_{ck}$. In greater detail with individual blocks of the ortho-

normal propagation matrix C_{k+1} shown, the equation is:

$$\begin{bmatrix} \boldsymbol{\alpha}_{k+1} \\ \boldsymbol{\beta}_{k+1} \\ \boldsymbol{\gamma}_{k+1} \end{bmatrix} = C_{k+1} \begin{bmatrix} \boldsymbol{\alpha}_{k} \\ \boldsymbol{\beta}_{k} \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} = \begin{bmatrix} C_{\alpha\alpha k+1} & C_{\alpha\beta k+1} & C_{\alpha w c k+1} \\ C_{\beta\alpha k+1} & C_{\beta\beta k+1} & C_{\beta w c k+1} \\ C_{\gamma\alpha k+1} & C_{\gamma\beta k+1} & C_{\gamma w c k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{k} \\ \boldsymbol{\beta}_{k} \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} \tag{5.10}$$

In order to eliminate $\boldsymbol{\alpha}_{k+1}$ and $\boldsymbol{\beta}_{k+1}$ from Eq. (5.9a), only the first two rows of Eq. (5.10) are required:

$$\begin{bmatrix} \boldsymbol{\alpha}_{k+1} \\ \boldsymbol{\beta}_{k+1} \end{bmatrix} = \begin{bmatrix} C_{\alpha\alpha k+1} & C_{\alpha\beta k+1} & C_{\alpha w c k+1} \\ C_{\beta\alpha k+1} & C_{\beta\beta k+1} & C_{\beta w c k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{k} \\ \boldsymbol{\beta}_{k} \\ \boldsymbol{w}_{\boldsymbol{c}k} \end{bmatrix} \tag{5.11}$$

Substitution of Eq. (5.11) into Eq. (5.9a) involves the multiplication of two large block matrices and the resulting block matrix is complicated. For the sake of compactness, it helps to first define an intermediate expression:

$$\begin{bmatrix} \tilde{S}^*_{x\alpha k} & \tilde{S}^*_{x\beta k} & \tilde{S}^*_{xwck} \end{bmatrix} \equiv \begin{bmatrix} S^*_{x\alpha k+1} & S^*_{x\beta k+1} \end{bmatrix} \begin{bmatrix} C_{\alpha\alpha k+1} & C_{\alpha\beta k+1} & C_{\alpha w c k+1} \\ C_{\beta\alpha k+1} & C_{\beta\beta k+1} & C_{\beta w c k+1} \end{bmatrix} \tag{5.12}$$

With this notation, Eq. (5.9a) can be written without reference to $\boldsymbol{\alpha}_{k+1}$ or $\boldsymbol{\beta}_{k+1}$:

$$\mathcal{R}^*_{xxk+1}\boldsymbol{x}_{k+1} = \boldsymbol{z}^*_{k+1} - \begin{bmatrix} \tilde{S}^*_{x\alpha k} & \tilde{S}^*_{x\beta k} & \tilde{S}^*_{xwck} & S^*_{x\psi k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{k} \\ \boldsymbol{\beta}_{k} \\ \boldsymbol{w}_{\boldsymbol{c}k} \\ \boldsymbol{\psi}_{k+1} \end{bmatrix} - \boldsymbol{b}^*_{ck+1} \tag{5.13}$$

Next, the state $\boldsymbol{x}_{k+1}$ can be written in terms of quantities referenced to sample k by means of the Consider version of the state dynamics equation in Table 5.1. In order to make the substitution, the Consider state $\boldsymbol{x}_{\boldsymbol{c}k}$ must first be replaced in this dynamics equation by $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ using the relationship defined in Eq. (5.1). The resulting state dynamics equation is:

$$\boldsymbol{x}_{k+1} = \Phi_{fk}\boldsymbol{x}_k + \Gamma_{fk}\boldsymbol{w}_k + \Gamma_{xck} \begin{bmatrix} L_{\alpha k} & L_{\beta k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_{k} \\ \boldsymbol{\beta}_{k} \end{bmatrix} + \boldsymbol{b}_{xk} \tag{5.14}$$

After substituting Eq. (5.14) into Eqs. (5.13) and (5.9b), rearranging terms, and stacking the results, one obtains:

$$\begin{bmatrix} \left(\bar{\mathcal{R}}_{wwk} + \bar{\mathcal{R}}_{wxk+1}\Gamma_{fk}\right) & \bar{\mathcal{R}}_{wxk+1}\Phi_{fk} \\ \mathcal{R}^*_{xxk+1}\Gamma_{fk} & \mathcal{R}^*_{xxk+1}\Phi_{fk} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{x}_k \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{z}}_{wk} \\ \boldsymbol{z}^*_{k+1} \end{bmatrix}$$

$$- \begin{bmatrix} \left(\bar{S}_{w\alpha k} + \bar{\mathcal{R}}_{wxk+1}\Gamma_{xck}L_{\alpha k}\right) & \left(\bar{S}_{w\beta k} + \bar{\mathcal{R}}_{wxk+1}\Gamma_{xck}L_{\beta k}\right) & 0 & 0 \\ \left(\tilde{S}^*_{x\alpha k} + \mathcal{R}^*_{xxk+1}\Gamma_{xck}L_{\alpha k}\right) & \left(\tilde{S}^*_{x\beta k} + \mathcal{R}^*_{xxk+1}\Gamma_{xck}L_{\beta k}\right) & \tilde{S}^*_{xwck} & S^*_{x\psi k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{\boldsymbol{c}k} \\ \boldsymbol{\psi}_{k+1} \end{bmatrix}$$

$$- \begin{bmatrix} \bar{\boldsymbol{b}}_{wk} + \bar{\mathcal{R}}_{wxk+1}\boldsymbol{b}_{xk} \\ \boldsymbol{b}^*_{ck+1} + \mathcal{R}^*_{xxk+1}\boldsymbol{b}_{xk} \end{bmatrix} \quad (5.15)$$

In Eq. (5.15), the left-hand side and the first term on the right-hand side are exactly the terms formed by a standard RTS SRIS prior to completing a backwards propagation step. The remaining terms take the place of a zero-mean, identity-covariance information error vector. They describe the combined effects of all of the filter/smoother model errors on the smoother's estimation error.

To complete the backwards propagation, one next uses standard orthonormal/upper-triangular (QR) factorization to compute an orthonormal matrix T^*_k such that the following relationship is satisfied:

$$T^*_k \begin{bmatrix} \left(\bar{\mathcal{R}}_{wwk} + \bar{\mathcal{R}}_{wxk+1}\Gamma_{fk}\right) & \bar{\mathcal{R}}_{wxk+1}\Phi_{fk} \\ \mathcal{R}^*_{xxk+1}\Gamma_{fk} & \mathcal{R}^*_{xxk+1}\Phi_{fk} \end{bmatrix} = \begin{bmatrix} \mathcal{R}^*_{wwk} & \mathcal{R}^*_{wxk} \\ 0 & \mathcal{R}^*_{xxk} \end{bmatrix} \quad (5.16)$$

In Eq. (5.16), the block matrix on the left-hand side is the input to the factorization, and T^*_k is the transpose of the orthonormal matrix output. The upper-triangular matrices $\mathcal{R}^*_{wwk}$ and $\mathcal{R}^*_{xxk}$ and the general matrix $\mathcal{R}^*_{wxk}$ are additional QR-factorization outputs. Every term in Eq. (5.15) is multiplied by the transfor-

mation T_k^*, and the resulting matrices and vectors are named to yield:

$$\begin{bmatrix} \mathcal{R}^*_{wwk} & \mathcal{R}^*_{wxk} \\ 0 & \mathcal{R}^*_{xxk} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_k \\ \boldsymbol{x}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}^*_{wk} \\ \boldsymbol{z}^*_k \end{bmatrix} - \begin{bmatrix} S^*_{w\alpha k} & S^*_{w\beta k} & S^*_{wwck} & S^*_{w\psi k+1} \\ S^*_{x\alpha k} & S^*_{x\beta k} & \bar{S}^*_{xwck} & \bar{S}^*_{x\psi k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{\boldsymbol{c}k} \\ \boldsymbol{\psi}_{k+1} \end{bmatrix} - \begin{bmatrix} \boldsymbol{b}^*_{wk} \\ \boldsymbol{b}^*_{ck} \end{bmatrix} \tag{5.17}$$

In this equation, the transformed matrix and vector terms appear in exactly the same order as the corresponding original terms from Eq. (5.15).

Neglecting the top row of Eq. (5.17), which is no longer needed, the bottom row is:

$$\mathcal{R}^*_{xxk}\boldsymbol{x}_k = \boldsymbol{z}^*_k - \begin{bmatrix} S^*_{x\alpha k} & S^*_{x\beta k} & \bar{S}^*_{xwck} & \bar{S}^*_{x\psi k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{w}_{\boldsymbol{c}k} \\ \boldsymbol{\psi}_{k+1} \end{bmatrix} - \boldsymbol{b}^*_{ck} \tag{5.18}$$

This information equation is almost in the desired form, which is identical to Eq. (5.9a) except that it is written in terms of quantities at sample k rather than sample $k+1$. The current form differs from the desired equation by the presence of the two vectors $\boldsymbol{w}_{\boldsymbol{c}k}$ and $\boldsymbol{\psi}_{k+1}$ rather than the single vector $\boldsymbol{\psi}_k$. To complete the recursion step, it is necessary to define a $\boldsymbol{\psi}_k$ and to compute an $S^*_{x\psi k}$ such that the following relationship holds:

$$S^*_{x\psi k}\boldsymbol{\psi}_k = \begin{bmatrix} \bar{S}^*_{xwck} & \bar{S}^*_{x\psi k+1} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_{\boldsymbol{c}k} \\ \boldsymbol{\psi}_{k+1} \end{bmatrix} \tag{5.19}$$

There is more than one possible $\left(\boldsymbol{\psi}_k, S^*_{x\psi k}\right)$ pair that satisfies Eq. (5.19). For any sensible choice, however, the vector $\boldsymbol{\psi}_k$ should not have more elements than necessary. To capture the estimation error effects of the Consider process noise vector $\boldsymbol{w}_{\boldsymbol{c}j}$ at all samples $j \geq k$, $\boldsymbol{\psi}_k$ requires at most n_x elements. Thus there are two cases to handle.

In the first and simplest case, the composite vector $\begin{bmatrix} \boldsymbol{w}_{\boldsymbol{c}k}^{\mathrm{T}} & \boldsymbol{\psi}_{k+1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ already has dimension less than or equal to the maximum dimension n_x. This is more likely to occur near the beginning of a smoothing pass since $\boldsymbol{\psi}_N$ is initialized as an empty vector. In this situation, the recursion can be completed by defining the composite vector and computing its matrix coefficient according to:

$$\boldsymbol{\psi}_k = \begin{bmatrix} \boldsymbol{w}_{\boldsymbol{c}k} \\ \boldsymbol{\psi}_{k+1} \end{bmatrix}, \quad S_{x\psi k}^* = \begin{bmatrix} \bar{S}_{xwck}^* & \bar{S}_{x\psi k+1}^* \end{bmatrix} \tag{5.20}$$

It is further possible to verify that the resulting composite error effects vector $\begin{bmatrix} \boldsymbol{\alpha}_k^{\mathrm{T}} & \boldsymbol{\beta}_k^{\mathrm{T}} & \boldsymbol{\psi}_k^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ has the desired identity covariance. The new $\boldsymbol{\psi}_k$ has identity covariance because $\boldsymbol{\psi}_{k+1}$ and $\boldsymbol{w}_{\boldsymbol{c}k}$ individually have identity covariance, and because $\boldsymbol{\psi}_{k+1}$ depends only on the history of the Consider process noise at future samples and so is uncorrelated with $\boldsymbol{w}_{\boldsymbol{c}k}$. As $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ depend on the history of the Consider state vector $\boldsymbol{x}_{\boldsymbol{c}k}$ up to and including sample k, they are uncorrelated with this $\boldsymbol{\psi}_k$ which depends only on Consider process noise at samples greater than or equal to k.

The second case is slightly more complicated. If $\begin{bmatrix} \boldsymbol{w}_{\boldsymbol{c}k}^{\mathrm{T}} & \boldsymbol{\psi}_{k+1}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$ has dimension greater than n_x, then the error effects contained in the composite vector can be compressed into a vector of dimension n_x by means of lower-triangular/orthonormal (LQ) factorization. This factorization procedure is somewhat less common than QR factorization, but is closely related. It is described in more detail in the Appendix of Chapter 4. In this situation, the LQ factorization computes an orthonormal transformation C_k^* that satisfies:

$$\begin{bmatrix} S_{x\psi k}^* & 0 \end{bmatrix} C_k^* = \begin{bmatrix} \bar{S}_{xwck}^* & \bar{S}_{x\psi k+1}^* \end{bmatrix} \tag{5.21}$$

The block sensitivity matrix on the right-hand side is the input to the LQ factorization. In addition to the orthonormal C_k^*, the block lower-triangular ma-

trix on the left-hand side is an output. The matrix $S^*_{x\psi k}$ is thus square and lower-triangular, with dimension n_x. When the composite vector $\begin{bmatrix}\boldsymbol{w_c}_k^{\mathrm{T}} & \boldsymbol{\psi}_{k+1}^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$ is transformed by C^*_k, the result is defined as:

$$\begin{bmatrix}\boldsymbol{\psi}_k \\ \boldsymbol{\zeta}_k\end{bmatrix} = C^*_k \begin{bmatrix}\boldsymbol{w_c}_k \\ \boldsymbol{\psi}_{k+1}\end{bmatrix} \tag{5.22}$$

The resulting $\boldsymbol{\psi}_k$ has the desired dimension n_x. The transformation given by Eqs. (5.21) and (5.22) can be shown to satisfy Eq. (5.19):

$$\begin{bmatrix}\bar{S}^*_{xwck} & \bar{S}^*_{x\psi k+1}\end{bmatrix}\begin{bmatrix}\boldsymbol{w_c}_k \\ \boldsymbol{\psi}_{k+1}\end{bmatrix} = \begin{bmatrix}S^*_{x\psi k} & 0\end{bmatrix} C^*_k \begin{bmatrix}\boldsymbol{w_c}_k \\ \boldsymbol{\psi}_{k+1}\end{bmatrix} = \begin{bmatrix}S^*_{x\psi k} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{\psi}_k \\ \boldsymbol{\zeta}_k\end{bmatrix} = S^*_{x\psi k}\boldsymbol{\psi}_k \tag{5.23}$$

Furthermore, the orthonormality of the transformation C^*_k preserves the identity covariance of $\boldsymbol{\psi}_k$, and it remains uncorrelated with $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$.

In either of the two cases represented by Eqs. (5.20)-(5.22), smoother state information Eq. 5.18 reduces to

$$\mathcal{R}^*_{xxk}\boldsymbol{x}_k = \boldsymbol{z}^*_k - \begin{bmatrix}S^*_{x\alpha k} & S^*_{x\beta k} & S^*_{x\psi k}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{\psi}_k\end{bmatrix} - \boldsymbol{b}^*_{ck} \tag{5.24}$$

which is just where the Consider smoother analysis started, except indexed to k instead of $k+1$. The RTS process is thus complete, and the next iteration of the backward recursion may begin.

5.3.3 Smoothed Estimation Error Covariances

In a typical square-root information smoother, a smoothed state information equation is computed at each step of the backward pass. It takes the form:

$$\mathcal{R}^*_{xxk}\boldsymbol{x}_k = \boldsymbol{z}^*_k - \boldsymbol{\nu}^*_{xk} \tag{5.25}$$

where $\mathcal{R}^*_{xxk}$ is the inverse-square-root of the smoothed estimation error covariance, $\boldsymbol{z}^*_k$ is the smoothed information vector, and the error vector $\boldsymbol{\nu}^*_{xk}$ is assumed to have zero mean and identity covariance. The smoothed estimate can be computed from Eq. (5.25) as:

$$\boldsymbol{x}^*_k = \mathcal{R}^{*-1}_{xxk}\boldsymbol{z}^*_k \tag{5.26}$$

and the smoothed estimation error is

$$\boldsymbol{x}^*_k - \boldsymbol{x}_k = \mathcal{R}^{*-1}_{xxk}\boldsymbol{\nu}^*_{xk} \tag{5.27}$$

The smoothed estimation error covariance is therefore:

$$P^*_{fxxk} = \mathrm{E}\Big[\left(\mathcal{R}^{*-1}_{xxk}\boldsymbol{\nu}^*_{xk}\right)\left(\mathcal{R}^{*-1}_{xxk}\boldsymbol{\nu}^*_{xk}\right)^{\mathrm{T}}\Big] = \mathcal{R}^{*-1}_{xxk}\mathrm{E}\big[\boldsymbol{\nu}^*_{xk}\boldsymbol{\nu}^{*\mathrm{T}}_{xk}\big]\,\mathcal{R}^{*-\mathrm{T}}_{xxk} = \mathcal{R}^{*-1}_{xxk}\mathcal{R}^{*-\mathrm{T}}_{xxk} \tag{5.28}$$

The Consider analysis version of this calculation substitutes its own complicated error term in place of the simplistic $\boldsymbol{\nu}^*_{xk}$. It starts with the Consider version of the smoothed information equation.

$$\mathcal{R}^*_{xxk}\boldsymbol{x}_k = \boldsymbol{z}^*_k - \boldsymbol{\eta}^*_k \tag{5.29}$$

where the error vector $\boldsymbol{\eta}^*_k$ replaces $\boldsymbol{\nu}^*_{xk}$ of the standard smoother and is defined as

$$\boldsymbol{\eta}^*_k \equiv \begin{bmatrix} S^*_{x\alpha k} & S^*_{x\beta k} & S^*_{x\psi k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{\psi}_k \end{bmatrix} + \boldsymbol{b}^*_{ck} \tag{5.30}$$

Equation (5.29) is just a more compact way of writing Eq. (5.24) from the end of the preceding subsection. Substitution of $\boldsymbol{\eta}^*_k$ and its statistics into the smoother

covariance calculations of Eq. (5.28) yields:

$$
\begin{aligned}
P^*_{xxk} &= \mathrm{E}\Big[\left(\mathcal{R}^{*-1}_{xxk}\boldsymbol{\eta}^*_k\right)\left(\mathcal{R}^{*-1}_{xxk}\boldsymbol{\eta}^*_k\right)^{\mathrm{T}}\Big] \\
&= \mathcal{R}^{*-1}_{xxk}\mathrm{E}\left[\left(\begin{bmatrix} S^*_{x\alpha k} & S^*_{x\beta k} & S^*_{x\psi k}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{\psi}_k\end{bmatrix} + \boldsymbol{b}^*_{ck}\right)\right. \\
&\qquad \left.\times\left(\begin{bmatrix} S^*_{x\alpha k} & S^*_{x\beta k} & S^*_{x\psi k}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{\psi}_k\end{bmatrix} + \boldsymbol{b}^*_{ck}\right)^{\mathrm{T}}\right]\mathcal{R}^{*-\mathrm{T}}_{xxk} \\
&= \mathcal{R}^{*-1}_{xxk}\begin{bmatrix} S^*_{x\alpha k} & S^*_{x\beta k} & S^*_{x\psi k}\end{bmatrix}\mathrm{E}\left\{\begin{bmatrix}\boldsymbol{\alpha}_k \\ \boldsymbol{\beta}_k \\ \boldsymbol{\psi}_k\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}^{\mathrm{T}}_k & \boldsymbol{\beta}^{\mathrm{T}}_k & \boldsymbol{\psi}^{\mathrm{T}}_k\end{bmatrix}\right\}\begin{bmatrix} S^{*\mathrm{T}}_{x\alpha k} \\ S^{*\mathrm{T}}_{x\beta k} \\ S^{*\mathrm{T}}_{x\psi k}\end{bmatrix}\mathcal{R}^{*-\mathrm{T}}_{xxk} \\
&\quad + \mathcal{R}^{*-1}_{xxk}\boldsymbol{b}^*_{ck}\boldsymbol{b}^{*\mathrm{T}}_{ck}\mathcal{R}^{*-\mathrm{T}}_{xxk} \\
&= \mathcal{R}^{*-1}_{xxk}\begin{bmatrix} S^*_{x\alpha k} & S^*_{x\beta k} & S^*_{x\psi k}\end{bmatrix}\begin{bmatrix} S^{*\mathrm{T}}_{x\alpha k} \\ S^{*\mathrm{T}}_{x\beta k} \\ S^{*\mathrm{T}}_{x\psi k}\end{bmatrix}\mathcal{R}^{*-\mathrm{T}}_{xxk} + \mathcal{R}^{*-1}_{xxk}\boldsymbol{b}^*_{ck}\boldsymbol{b}^{*\mathrm{T}}_{ck}\mathcal{R}^{*-\mathrm{T}}_{xxk}
\end{aligned}
\tag{5.31}
$$

The resulting quantity, P^*_{xxk}, is the true matrix mean square error (MSE) for the smoother. It is equal to the true smoothed estimation error covariance plus a term relating to the deterministic bias $\boldsymbol{b}^*_{ck}$. The true covariance is thus the left-hand term of the last line of Eq. (5.31). Although the use of "P_{xx} " to represent matrix MSE rather than covariance is non-standard, it emphasizes the role of P^*_{xxk} in the Consider analysis. The matrix MSE P^*_{xxk}, in contrast to the estimation error covariance, contains the full mismodeling effect from both random and deterministic errors. Consequently, it is the most appropriate quantity for comparison with the smoother-assumed covariance P^*_{fxxk} of Eq. 5.28.

5.4 Examples

In order to demonstrate the Consider smoother analysis, several concrete examples have been developed. These examples, while simple, illustrate several common classes of filter/smoother modeling errors. They complement Chapter 4 by addressing some error varieties not investigated by that chapter's examples. The examples in this section further clarify how one can pose Consider analyses in the Consider model form of Section 5.2.1. They also highlight Considered error effects and may suggest areas of application.

For each example, the techniques of Section 5.2.1 have been employed to write the "truth" system equations in the defined Consider form. This procedure is often the most challenging part of the Consider analysis. Once in the correct form, the equations can be manipulated to perform a forward Consider filter analysis and a backward Consider smoother analysis, as per the algorithms of Chapter 4 and Section 5.3 of the present chapter. The results of each Consider smoother analysis have been independently validated by Monte Carlo analysis.

The examples to be addressed are as follows: The first investigates the effects of a biased initial state estimate and incorrect initial covariance. The next example is a case with mutually correlated process and measurement noise, both with incorrect covariances. A third example examines a system that has dynamics perturbed by unmodeled colored process noise with a sinusoidal influence matrix. This case also has an unmodeled random bias in its measurements. All of these examples share the same incorrect filter/smoother for easy comparison, but the nature of its errors varies.

5.4.1 Filter/Smoother Assumed Model Description

The filter/smoother for all of the examples has assumed dynamics and measurement models given by

$$\begin{bmatrix} r_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_k \\ v_k \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_k \tag{5.32a}$$

$$y_k = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} r_k \\ v_k \end{bmatrix} + \nu_k \tag{5.32b}$$

This system can be thought of as describing some one-dimensional motion, with position state r_k and velocity state v_k. The sample interval, Δt, is assumed to be 0.5 s. The scalar y_k is a measurement. The filter/smoother's assumed system matrices Φ_{fk}, Γ_{fk}, and H_{fk} can be directly extracted from these equations.

Scalar process noise w_k and measurement noise ν_k are assumed to be zero-mean, Gaussian, and white, with joint covariance given by:

$$\mathrm{E}\left\{ \begin{bmatrix} w_k \\ \nu_k \end{bmatrix} \begin{bmatrix} w_k & \nu_k \end{bmatrix} \right\} = \begin{bmatrix} Q_{fk} & 0 \\ 0 & R_{fk} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.33}$$

where Q_{fk} and R_{fk} are the nominal process and measurement noise covariances, respectively. In other words, Eq. (5.33) specifies that the noise processes are assumed to be uncorrelated, and both are assumed to have unit variance. The filter/smoother's assumed square-root information matrices for process noise and measurement noise are the inverse square roots of the corresponding assumed covariances matrices. In the general case, such matrix square roots can be computed by a standard method such as Cholesky factorization. For this scalar case, however, they are just $\mathcal{R}_{fwwk} = 1/\sqrt{1} = 1$ and $\mathcal{R}_{f\nu\nu k} = 1/\sqrt{1} = 1$. Note that the assumed measurement model form in the center column of Table 5.1 requires, without loss of generality, that $\mathcal{R}_{f\nu\nu k}$ be the identity matrix in order to be consistent with standard SRIF/S practice.

The initial state estimate and initial state error covariance are assumed to be:

$$\bar{\boldsymbol{x}}_0 = \begin{bmatrix} \bar{r}_0 \\ \bar{v}_0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \overline{P}_{fxx0} = \begin{bmatrix} \sigma_{r0}^2 & \sigma_{rv0} \\ \sigma_{rv0} & \sigma_{v0}^2 \end{bmatrix} = \begin{bmatrix} 10 & 0 \\ 0 & 5 \end{bmatrix} \tag{5.34}$$

The initial estimation error is further assumed to be uncorrelated with process or measurement noise at any sample time. From Eq. (5.34), the assumed initial state square-root information matrix $\bar{\mathcal{R}}_{fxx0}$ and the initial information state $\bar{\boldsymbol{z}}_0$ can be computed. They are:

$$\bar{\mathcal{R}}_{fxx0} = \begin{bmatrix} 1/\sqrt{10} & 0 \\ 0 & 1/\sqrt{5} \end{bmatrix}, \quad \bar{\boldsymbol{z}}_0 = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{5} \end{bmatrix} \tag{5.35}$$

The filter/smoother operates on measurements available once per sample interval Δt for a total of 50 s (100 discrete-time samples). Note that all of the filter/smoother equations conform to the assumed forms in Table 5.1 of Section 5.2.1.

5.4.2 Example: Incorrect Initialization

The first example has error only in its initialization. Its dynamics and measurement models are assumed to be correctly given by the filter/smoother's models of Eqs. (5.32a) and (5.32b). Likewise, the Gaussian process and measurement noise are assumed to be correctly modeled as zero-mean, white, and uncorrelated, with unit variance. Process and measurement noise are uncorrelated with the initial estimation error.

Two kinds of error enter the initialization: An incorrect initial covariance and a deterministic estimation bias. The filter/smoother assumes that the initial estimation error has zero mean and covariance $\overline{P}_{fxx0}$. This example's "truth" system instead has an initial estimate that relates to the "truth" initial state ac-

cording to:

$$\bar{\boldsymbol{x}}_0 = \boldsymbol{x}_0 + \tilde{\boldsymbol{x}}_0 + \boldsymbol{x}_{b0} \tag{5.36}$$

where $\tilde{\boldsymbol{x}}_0$ is the zero-mean random component of the error in the initial estimate, and $\boldsymbol{x}_{b0}$ is a deterministic non-zero bias error. The covariance of the random error $\tilde{\boldsymbol{x}}_0$ and the value of the bias $\boldsymbol{x}_{b0}$ are given by:

$$\overline{P}_{xx0} = \mathrm{E}\big[\tilde{\boldsymbol{x}}_0 \tilde{\boldsymbol{x}}_0^{\mathrm{T}}\big] = \begin{bmatrix} 16 & 0 \\ 0 & 9 \end{bmatrix}, \quad \boldsymbol{x}_{b0} = \begin{bmatrix} -20 \\ 30 \end{bmatrix} \tag{5.37}$$

Note that the covariance of $\tilde{\boldsymbol{x}}_0$ is the "truth" initial state error covariance $\overline{P}_{xx0}$, which in this case is not equal to the filter/smoother's initial state error covariance $\overline{P}_{fxx0}$. The corresponding "truth" initial square-root information matrix for the state is

$$\bar{\mathcal{R}}_{xx0} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/3 \end{bmatrix} \tag{5.38}$$

To write this example's system in Consider form, one must first choose an appropriate Consider state vector $\boldsymbol{x}_{\boldsymbol{c}k}$. It must contain all the random uncertainty that enters via the initial estimate, the process noise, and the measurement noise. One such definition is:

$$\boldsymbol{x}_{\boldsymbol{c}k} \equiv \begin{cases} \begin{bmatrix} \bar{\mathcal{R}}_{xx0}\tilde{\boldsymbol{x}}_0 \\ \mathcal{R}_{ww0} w_0 \\ \mathcal{R}_{\nu\nu 0}\nu_0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} w_k \\ \nu_k \end{bmatrix} & k > 0 \end{cases} \tag{5.39}$$

Note how only the stochastic part of the initial estimation error enters this definition. Premultiplication by the various square-root information matrices at sample $k = 0$ is required to ensure that $\boldsymbol{x}_{\boldsymbol{c}0}$ has identity covariance.

The "truth" system matrices Φ_k, Γ_k, and H_k are identical to those of the filter, as are the "truth" noise square-root information matrices $\mathcal{R}_{wwk}$ and $\mathcal{R}_{\nu\nu k}$. There

is no dynamics bias or measurement bias, i.e., $\boldsymbol{b}_{xk} = 0$ and $\boldsymbol{b}_{yk} = 0$. Because there are no unmodeled disturbances to the system dynamics, the matrix Γ_{xck} is an appropriately-sized matrix of zeros for all samples k.

Next, one can define the dynamics of the Consider state $\boldsymbol{x_{ck}}$. The first component at $k = 0$, $\bar{\mathcal{R}}_{xx0}\tilde{\boldsymbol{x}}_0$, exists only at this sample and thus has no true dynamics. Likewise, the noise components w_k and ν_k are both white, and therefore they do not depend on noise at previous samples. Consequently, the Consider state transition matrix Φ_{ck} is also an appropriately-sized matrix of zeros for all samples k. In the special case of $k = 0$, the matrix Φ_{c0} is rectangular, with fewer rows than columns in order to omit the initial uncertainty component at the next sample. The Consider process noise influence matrix Γ_{cck} gives w_k and ν_k their proper statistics. Specifically,

$$\Gamma_{cck} = \begin{bmatrix} \mathcal{R}_{wwk+1}^{-1} & 0 \\ 0 & \mathcal{R}_{\nu\nu k+1}^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \forall\, k \tag{5.40}$$

The process noise information equation models the way the "truth" process noise statistics enter the system. Because the process noise is unbiased, $\boldsymbol{b}_{wk}$ is zero. As per the Consider-form process noise information equation of Table 5.1, the matrix S_{wck} is given by

$$S_{wck} = \begin{cases} \begin{bmatrix} \mathbf{0} & -\mathcal{R}_{fww0}\mathcal{R}_{ww0}^{-1} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & 0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} -\mathcal{R}_{fwwk} & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \end{bmatrix} & k > 0 \end{cases} \tag{5.41}$$

which picks out the properly weighted process noise component of $\boldsymbol{x_{ck}}$. Note that the $\mathbf{0}$ on the top line of Eq. (5.41) indicates a matrix of zeros with appropriate dimensions in all of this chapter's examples. The previous sections and Chapter 4 did not use this convention because no dimensional ambiguities arose for 0 matrices. In a similar manner to S_{wck}, H_{ck} extracts the properly weighted

measurement noise component of $\boldsymbol{x}_{\boldsymbol{c}k}$:

$$H_{ck} = \begin{cases} \begin{bmatrix} \mathbf{0} & 0 & \mathcal{R}_{\nu\nu 0}^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & 1 \end{bmatrix} & k > 0 \end{cases} \tag{5.42}$$

The remaining and most difficult part of the Consider form setup for this example relates to the initial state information equation, where the only Considered model errors enter. It is necessary to place the "truth" initial information equation into its Consider form, repeated here for convenience:

$$\bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_0 = \bar{\boldsymbol{z}}_0 - \bar{S}_{xc0}\boldsymbol{x}_{\boldsymbol{c}0} - \bar{\boldsymbol{b}}_{c0} \tag{5.43}$$

In order to determine the correct $\bar{S}_{xc0}$ and $\bar{\boldsymbol{b}}_{c0}$, one multiplies Eq. (5.36) by the filter/smoother's presumed initial square root information matrix $\bar{\mathcal{R}}_{fxx0}$, and one rearranges the result into something similar to Eq. (5.43).

$$\bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_0 = \bar{\mathcal{R}}_{fxx0}\bar{\boldsymbol{x}}_0 - \bar{\mathcal{R}}_{fxx0}\tilde{\boldsymbol{x}}_0 - \bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_{b0} \tag{5.44}$$

The first term on the right-hand side of Eq. (5.44) is by definition equal to the filter/smoother's initial *a priori* information state $\bar{\boldsymbol{z}}_0$. The second term is zero-mean and stochastic, so by comparison of Eqs. (5.43) and (5.44), $\bar{S}_{xc0}$ must be chosen such that $\bar{S}_{xc0}\boldsymbol{x}_{\boldsymbol{c}0} = \bar{\mathcal{R}}_{fxx0}\tilde{\boldsymbol{x}}_0$. Consistent with the first line of Eq. (5.39), this is accomplished by the definition:

$$\bar{S}_{xc0} = \begin{bmatrix} \left(\bar{\mathcal{R}}_{fxx0}\bar{\mathcal{R}}_{xx0}^{-1}\right) & \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \frac{4}{\sqrt{10}} & 0 & 0 & 0 \\ 0 & \frac{3}{\sqrt{5}} & 0 & 0 \end{bmatrix} \tag{5.45}$$

Likewise, the final terms in Eqs. (5.43) and (5.44) are both deterministic initial information biases. Setting them equal yields:

$$\bar{\boldsymbol{b}}_{c0} = \bar{\mathcal{R}}_{fxx0}\boldsymbol{x}_{b0} = \begin{bmatrix} -\frac{20}{\sqrt{10}} \\ \frac{30}{\sqrt{5}} \end{bmatrix} \tag{5.46}$$

At this point, all the Consider form quantities have been defined, and the filter/smoother Consider analysis algorithms of Chapter 4 and the present chapter's Section 5.3 can be implemented.

In addition to implementation of the analytical Consider algorithms, numerical Monte Carlo simulations have been used to independently verify the computed smoothed matrix mean square errors. This technique forms a large number of simulated estimation errors by comparing "truth" state vectors to estimated states produced by the mismodeled filter/smoother. Each simulated estimation error history is a sample of a random process, and sample means and matrix MSEs can be computed. As the number of trials increases, the estimated matrix MSE or covariance approaches the true value. For each of this chapter's examples, as in the preceding chapter, 5000 trials have been used.

To implement the Monte Carlo simulations, a "truth" model is created for the given example based on the defined "truth" system matrices and "truth" noise covariances. The "truth" model takes as inputs the initial state and histories of the measurement and process noise, which are constrained to have the "truth" statistical properties as specified in the example. It outputs a "truth" history of the state vector and a series of noisy measurements. The noisy measurements from the "truth" model are next used as the inputs to the filter/smoother described in Section 5.4.1, and it outputs histories of state estimates. Estimation error histories are computed as the difference of these estimates and the states from the "truth" model.

Note that the "truth" model implementation does not use the special Consider form defined in Section 5.2.1. This independence from the Consider model form is an important feature of the Monte Carlo tests. Close correspondence

between the Monte Carlo and Consider analysis results is unlikely unless the Consider analysis algorithms and the Consider model form both have been implemented correctly.

Figure 5.1 plots representative results from the Monte Carlo and Consider analyses for the incorrect initialization example. It displays three versions of the smoothed root mean square (RMS) position error, which is the square root of the diagonal element of the matrix MSE corresponding to position. The dashed

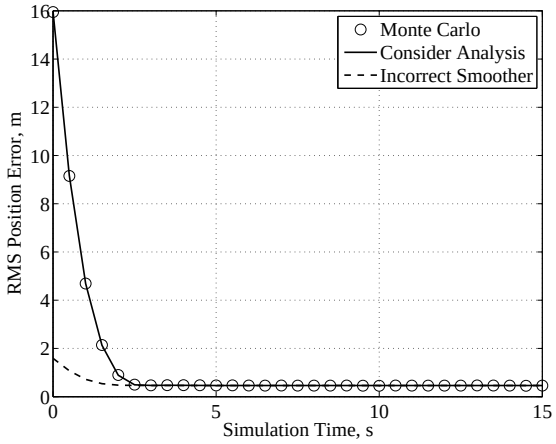

Figure 5.1: RMS smoother position errors for incorrect initialization example.

line is the smoothed RMS error reported by the mismodeled filter/smoother, and the solid line is the true RMS error for the smoother as computed by the Consider analysis. The circles are the smoothed RMS errors computed from the Monte Carlo simulations. Note that only the first 15 s of the 50 s simulation time are shown in Fig. 5.1. Where the Consider and nominal smoother errors diverge, the Monte Carlo results follow the Consider analysis closely. The biased initial estimate and incorrect initial covariance only affect estimation uncertainty for a relatively short time. After this initial transient behavior, the true filter/smoother accuracy returns to the same level that it would have had with no model errors. The other filtering and smoothing results for this example are

qualitatively similar to those of Fig. 5.1.

In addition to true matrix MSE, the distributions of estimation error from the Monte Carlo simulations have been used to compute confidence bounds on the results. Such confidence bounds can indicate whether a sufficiently large number of Monte Carlo trials has been used, so as to achieve a desired level of accuracy. For the present example, the Monte Carlo and Consider analysis results generally agree to well within the confidence bounds. The confidence bounds are not shown in Fig. 5.1 in order to reduce clutter.

5.4.3 Example: Correlated Noise

The second numerical example is that of correlated process and measurement noise. In this example, the "truth" system's dynamics and measurements are still described by the models of Eqs. (5.32a) and (5.32b). Also, the filter/ smoother's initial estimate and initial state error covariance are accurate, and the initial estimation error is uncorrelated with process or measurement noise. The process and measurement noise, however, both have much larger covariances than the assumed values, and they are correlated. Specifically,

$$\mathrm{E}\left\{ \begin{bmatrix} w_k \\ \nu_k \end{bmatrix} \begin{bmatrix} w_k & \nu_k \end{bmatrix} \right\} = \begin{bmatrix} Q_k & P_{w\nu k} \\ P_{w\nu k} & R_k \end{bmatrix} = \begin{bmatrix} 10 & -3 \\ -3 & 8 \end{bmatrix} \tag{5.47}$$

This case can be handled by defining the composite noise vector $\boldsymbol{\mu}_k = [w_k \quad \nu_k]^{\mathrm{T}}$. With this definition, the composite noise covariance matrix $P_{\mu\mu k}$ is just the rightmost matrix of Eq. (5.47). Its inverse square root, $\mathcal{R}_{\mu\mu k}$, is found by Cholesky factorization of $P_{\mu\mu k}^{-1}$ to be approximately:

$$\mathcal{R}_{\mu\mu k} = \begin{bmatrix} 0.3357 & 0.1259 \\ 0 & 0.3536 \end{bmatrix} \tag{5.48}$$

While one could compute the individual square-root information matrices for process and measurement noise, Q_k and R_k, they are less meaningful in this example than the joint square-root information matrix $\mathcal{R}_{\mu\mu k}$.

To describe this system in Consider form, the Consider state vector $\boldsymbol{x_{ck}}$ must contain the "true", correlated process and measurement noise. It must also capture the initial estimation error at sample $k = 0$. A suitable definition is:

$$\boldsymbol{x_{ck}} \equiv \begin{cases} \begin{bmatrix} \bar{\boldsymbol{\nu}}_{x0} \\ \mathcal{R}_{\mu\mu 0}\boldsymbol{\mu}_0 \end{bmatrix} & k = 0 \\ \\ \begin{bmatrix} \boldsymbol{\mu}_k \end{bmatrix} & k > 0 \end{cases} \tag{5.49}$$

where $\bar{\boldsymbol{\nu}}_{x0}$ is the initial information state error vector. It appears in the filter/smoother version of the initial information equation in Table 5.1, and in this scenario is unchanged in the "truth" system. The vector $\boldsymbol{\mu}_k$ is the previously defined composite noise vector containing w_k and ν_k. At sample $k = 0$, premultiplication of $\boldsymbol{\mu}_0$ by $\mathcal{R}_{\mu\mu 0}$ causes the Consider state vector $\boldsymbol{x_{c0}}$ to have identity covariance, as required by the Consider model form.

As in the previous example, the "truth" system matrices Φ_k, Γ_k, and H_k are the same as those used by the filter/smoother. Also, there is no dynamics disturbance so Γ_{xck} is an appropriately-dimensioned matrix of zeros. There are no non-zero deterministic biases for this example, so $\boldsymbol{b}_{xk}$, $\boldsymbol{b}_{yk}$, $\boldsymbol{b}_{wk}$, and $\bar{\boldsymbol{b}}_{c0}$ are all zero-valued.

Based on the definition of the Consider state in Eq. (5.49), the Consider state dynamics equation can be constructed. The noise contained in $\boldsymbol{\mu}_k$ is white, and the initial information error $\bar{\boldsymbol{\nu}}_{x0}$ is only present at sample $k = 0$, so Φ_{ck} is a matrix of zeros for all samples k. At $k = 0$, Φ_{c0} is rectangular with more columns than rows in order to transition to the new version of the Consider state vector with

fewer components. The matrix Γ_{cck} models the correlated statistics of w_k and ν_k. Thus for all samples k it takes the form:

$$\Gamma_{cck} = \mathcal{R}^{-1}_{\mu\mu k+1} = \begin{bmatrix} 2.9791 & -1.0607 \\ 0 & 2.8284 \end{bmatrix} \tag{5.50}$$

The matrix S_{wck} models how the "true", correlated process noise enters the filter/smoother. It extracts the first component of the composite noise vector $\boldsymbol{\mu}_k$ from $\boldsymbol{x_{ck}}$ and properly weights the result. In terms of the defined quantities, it is:

$$S_{wck} = \begin{cases} -\mathcal{R}_{fww0}\begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathcal{R}^{-1}_{\mu\mu 0} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -2.9791 & 1.0607 \end{bmatrix} & k = 0 \\ -\mathcal{R}_{fwwk}\begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \end{bmatrix} & k > 0 \end{cases} \tag{5.51}$$

In the same manner, H_{ck} selects the "true", correlated measurement noise by extracting the second component of $\boldsymbol{\mu}_k$ from $\boldsymbol{x_{ck}}$. The matrix that accomplishes this is:

$$H_{ck} = \begin{cases} \begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathcal{R}^{-1}_{\mu\mu 0} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 2.8284 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & 1 \end{bmatrix} & k > 0 \end{cases} \tag{5.52}$$

Finally, the initial estimation error covariance assumed by the filter/ smoother is correct. This is modeled by $\bar{S}_{xc0}$, which selects the unweighted $\bar{\boldsymbol{\nu}}_{x0}$ that forms the first component of $\boldsymbol{x_{c0}}$:

$$\bar{S}_{xc0} = \begin{bmatrix} \mathrm{I} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{5.53}$$

After applying the Consider filter/smoother algorithms to this Consider-form system, Monte Carlo simulations have been used to demonstrate the correctness of the calculated true smoother covariances. Figures 5.2 and 5.3 display the filtered (*a posteriori*) and smoothed velocity error standard deviations,

respectively. As in Fig. 5.1, the solid lines are the analytical results from the

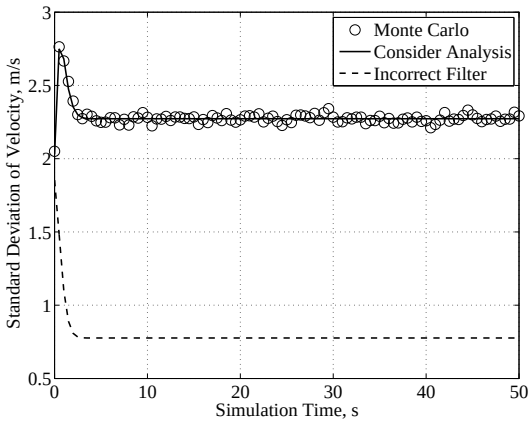

Figure 5.2: Standard deviations of filtered *a posteriori* velocity error for correlated noise example.

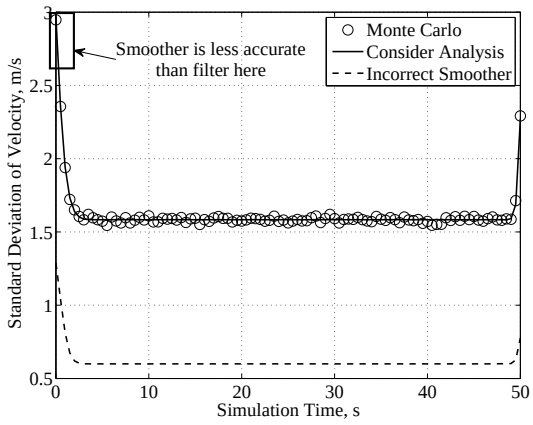

Figure 5.3: Standard deviations of smoothed velocity error for correlated noise example.

Consider analysis and the circles are the numerical results from the Monte Carlo simulations. The dashed lines show the incorrect standard deviations reported by the original filter/smoother.

This example's results demonstrate close agreement between the analytical and numerical techniques. They further illustrate an important and counterintuitive feature of the Consider analysis of a smoother. A traditional linear filter/smoother has the property that its smoothed covariance is always less than its filtered *a posteriori* covariance in a positive definite sense: $P^*_{fxxk} \leq P_{fxxk}$. One can understand this property in terms of the information used to form each estimate; information from future measurements always reduces uncertainty rather than increasing it. This principle no longer holds for the Consider analysis covariances: $P^*_{xxk} \lesseqgtr P_{xxk}$. The smoothed velocity standard deviations of Fig. 5.3 are higher than the *a posteriori* standard deviations of Fig. 5.2 near the beginning of the simulation. Although not shown, the same situation occurs for this example's position standard deviations. Not only does the smoother incorporate

information from future measurements into its estimates, but it also incorporates model errors that apply at future samples. For this example, the additional model error effects added to the early estimates are greater than the uncertainty eliminated by additional measurements. In other words, a smoother with the wrong kind of modeling error may degrade rather than enhance estimation quality.

5.4.4 Example: Unestimated Disturbances

As with the first two examples, the filter/smoother for the third example is the one described in Section 5.4.1. In this example, the deterministic biases $\boldsymbol{b}_{xk}$, $\boldsymbol{b}_{yk}$, $\boldsymbol{b}_{wk}$, and $\bar{\boldsymbol{b}}_{c0}$ are all zero. There are also no errors associated with initialization or noise statistics. That is, $\overline{P}_{xx0} = \overline{P}_{fxx0}$, $Q_k = Q_{fk}$, and $R_k = R_{fk}$. The noise elements and initial estimation error are uncorrelated, and both the process noise and the measurement noise are white and Gaussian.

Although the "truth" system matrices are identical to those assumed by the filter/smoother, the "truth" dynamics and measurement models are affected by additional unmodeled disturbances. The "truth" dynamics equation is disturbed by unmodeled, colored process noise that has a sinusoidally time-varying influence on the velocity state. The measurements contain a constant random bias. Neither of these effects is estimated by the filter/smoother. The "truth" dynamics and measurement models are given by:

$$\begin{bmatrix} r_{k+1} \\ v_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_k \\ v_k \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w_k + \begin{bmatrix} 0 \\ \sin\left(\frac{2\pi k \Delta t}{T_{per}}\right) \end{bmatrix} a_k \tag{5.54a}$$

$$y_k = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} r_k \\ v_k \end{bmatrix} + \nu_k + b_k \tag{5.54b}$$

where the period of the sinusoidal disturbance influence is $T_{per} = 11.15$ s and where the colored disturbance a_k is a first-order Markov process. The bias $b_k = b_0$ is assumed to be drawn from a zero-mean random distribution with standard deviation $\sigma_b = 2/3$, and the initial noise state a_0 is drawn from a zero-mean random distribution with standard deviation $\sigma_a = 4$. The dynamics of the joint disturbance vector $[a_k \quad b_k]^{\mathrm{T}}$ are

$$\begin{bmatrix} a_{k+1} \\ b_{k+1} \end{bmatrix} = \begin{bmatrix} \mathrm{e}^{-\frac{\Delta t}{\tau_a}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_k \\ b_k \end{bmatrix} + \begin{bmatrix} \gamma_a \\ 0 \end{bmatrix} w_{ak} \tag{5.55}$$

with Markov time constant $\tau_a = 16.725$ s and parameter $\gamma_a = \sigma_a \sqrt{1 - \mathrm{e}^{-2\Delta t/\tau_a}}$, or approximately $\gamma_a = 0.9636$. The process noise w_{ak} that drives the Markov process has zero mean and unit variance. Note, the steady-state standard deviation of a_k equals the initial standard deviation σ_a by construction.

To write this system in Consider form, the Consider state vector must include not only the correctly-modeled initial uncertainty, process noise, and measurement noise, but also the unestimated dynamically varying a_k and the random bias b_k. A suitable definition is

$$\boldsymbol{x_c}_k \equiv \begin{cases} \begin{bmatrix} \bar{\boldsymbol{\nu}}_{x0} \\ a_0/\sigma_a \\ b_0/\sigma_b \\ \mathcal{R}_{ww0} w_0 \\ \mathcal{R}_{\nu\nu 0} \nu_0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} a_k \\ b_k \\ w_k \\ \nu_k \end{bmatrix} & k > 0 \end{cases} \tag{5.56}$$

As before, the Consider vector at sample $k = 0$ contains an additional component for initial estimate uncertainty, and its elements are defined such that it has

identity covariance. With this definition of $\boldsymbol{x_{ck}}$, one can write the matrices for the Consider dynamics equation for $k > 0$:

$$\Phi_{ck} = \begin{bmatrix} e^{-\frac{\Delta t}{\tau_a}} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.9705 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.57a}$$

$$\Gamma_{cck} = \begin{bmatrix} \gamma_a & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \mathcal{R}^{-1}_{wwk+1} & 0 \\ 0 & 0 & \mathcal{R}^{-1}_{\nu\nu k+1} \end{bmatrix} = \begin{bmatrix} 0.9636 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.57b}$$

Note how the Consider dynamics matrices capture both the dynamic behavior of a_k and b_k and the statistical behavior of w_k and ν_k. At $k = 0$, the Γ_{cck} matrix is unchanged, but Φ_{ck} becomes

$$\Phi_{c0} = \begin{bmatrix} \mathbf{0} & \sigma_a e^{-\frac{\Delta t}{\tau_a}} & 0 & 0 & 0 \\ \mathbf{0} & 0 & \sigma_b & 0 & 0 \\ \mathbf{0} & 0 & 0 & 0 & 0 \\ \mathbf{0} & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 3.8822 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.58}$$

in order to transition from the original Consider vector at $k = 0$ to its new form for $k > 0$.

In contrast to the previous two examples, Γ_{xck} is a non-zero matrix. It specifies how the unmodeled disturbance affects the filter's state. It is given by:

$$\Gamma_{xck} = \begin{cases} \begin{bmatrix} \mathbf{0} & 0 & 0 & 0 & 0 \\ \mathbf{0} & \sigma_a \sin\left(\frac{2\pi k \Delta t}{T_{per}}\right) & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ \sin\left(\frac{2\pi k \Delta t}{T_{per}}\right) & 0 & 0 & 0 \end{bmatrix} & k > 0 \end{cases} \tag{5.59}$$

In the measurement equation, H_{ck} must select the component of $\boldsymbol{x_{ck}}$ corresponding to measurement noise ν_k. For this example, it must additionally select

the component corresponding to the random measurement bias b_k. Note that this analysis Considers the effects of a zero-mean random measurement bias rather than a non-zero deterministic one. In other words, it predicts the true covariance when a bias with a given statistical distribution is present, rather than the matrix MSE resulting from a specific deterministic bias. The matrix H_{ck} that extracts the measurement noise and bias is given by:

$$H_{ck} = \begin{cases} \begin{bmatrix} \mathbf{0} & 0 & \sigma_b & 0 & \mathcal{R}_{\nu\nu 0}^{-1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \frac{2}{3} & 0 & 1 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & 1 & 0 & 1 \end{bmatrix} & k > 0 \end{cases} \tag{5.60}$$

The matrix S_{wck} selects the components of $\boldsymbol{x_{ck}}$ related to process noise and indicates that "truth" and modeled process noise are identical for this scenario:

$$S_{wck} = \begin{cases} \begin{bmatrix} \mathbf{0} & 0 & 0 & -\mathcal{R}_{fww0}\mathcal{R}_{ww0}^{-1} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} & k = 0 \\ \begin{bmatrix} 0 & 0 & -\mathcal{R}_{fwwk} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & 0 \end{bmatrix} & k > 0 \end{cases} \tag{5.61}$$

Finally, the matrix $\bar{S}_{xc0}$ specifies that the error term in the initial state information equation is statistically correct.

$$\bar{S}_{xc0} = \begin{bmatrix} \mathrm{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.62}$$

At this point, the system has been written in Consider form, and the Consider algorithms and Monte Carlo analysis can proceed in standard fashion.

Figures 5.4 and 5.5 show the filtered and smoothed velocity estimation error standard deviations. They are labeled in the same manner as the corresponding results in Figs. 5.2 and 5.3. In Fig. 5.4, it is clear that the unestimated sinusoidal disturbance in the velocity state corresponds to an unpredicted, roughly sinusoidal variation of the velocity error standard deviation. The oscillation

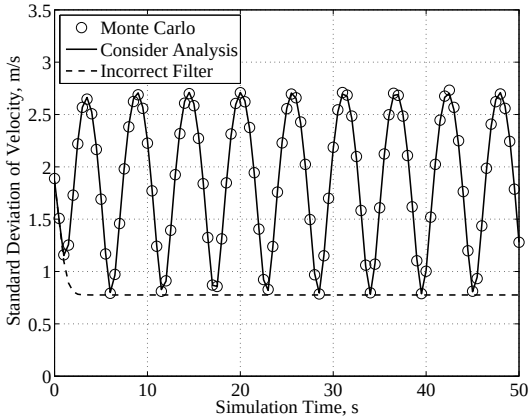

Figure 5.4: Standard deviations of filtered *a posteriori* velocity error for unestimated disturbances example.

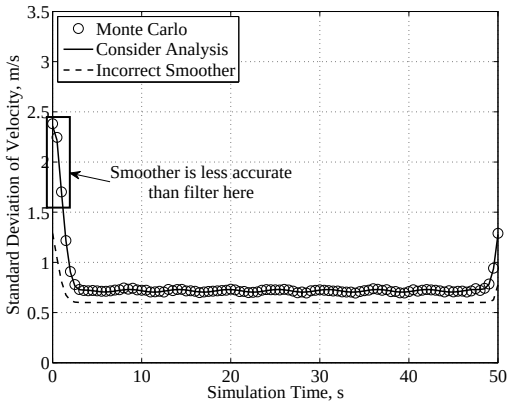

Figure 5.5: Standard deviations of smoothed velocity error for unestimated disturbances example.

frequency in this plot is approximately twice that of the disturbance, however. Both peaks and valleys of the sinusoidal disturbance result in high uncertainty, and the uncertainty drops to near its nominal value when the disturbance passes through zero.

Similar to the preceding correlated noise example, the smoothed standard deviations of Fig. 5.5 are mostly lower than the filtered *a posteriori* standard deviations of Fig. 5.4, but not at the beginning of the interval. It is significant that the sinusoidal variation in standard deviation is greatly reduced by the smoother. For this example, the Consider analyses show that the filter is very sensitive to the modeling error while the smoother is not: Note in Fig. 5.5 how the Consider analysis standard deviation is nearly equal to that of the incorrect smoother except at the beginning of the interval. This reduced smoother sensitivity stands in sharp contrast with the filter results of Fig. 5.4.

5.5 Conclusions

This chapter has developed a new form of Consider covariance analysis that can be applied to Rauch-Tung-Striebel discrete square-root information smoothers. Smoother analysis is accomplished by performing a backward analysis pass after having performed a forward-pass Consider filter analysis. The analysis computes the smoother's true estimation error covariance or matrix mean square error, whichever is most relevant. It generalizes to a wide variety of possible model errors. This generalization capability is enabled by the definition of a new standard system form. Once the system modeling errors have been written in this standard form, the analysis algorithms can be applied in a straightforward manner. A special feature of the Consider smoother analysis is its ability to "consider" the effects of both past and future model errors on the smoothed estimates.

Several concrete examples illustrate the power of the new method while clarifying its implementation. They include a system model with a biased initial estimate and an incorrect initial covariance, a model with cross-correlated process and measurement noise, and a model with dynamically-varying unestimated disturbance states. The ability to model three distinct varieties of errors in the same framework is thus demonstrated. Monte Carlo simulations provide independent verification of the smoother Consider analysis equations. Comparisons of the filter and smoother results for two of the examples show that the addition of a smoothing pass sometimes mitigates the effects of filter/smoother modeling errors, but can exacerbate model error effects in other instances.

CHAPTER 6

SUMMARY & CONCLUSIONS

This dissertation has developed several new estimation and analysis methods for situations where simple filtering algorithms may fail. Two strategies are proposed for specific applications in spacecraft attitude and orbit determination. These scenarios present challenges in the form of high nonlinearity and high model uncertainty. A more broadly applicable analysis technique is also derived to investigate the effects of modeling errors on square-root information filter and smoother performance.

Chapter 2 examines an extended version of Wahba's problem for spacecraft attitude determination. Specifically, it seeks a quaternion attitude and an angular velocity for a spinning spacecraft given only a sequence of unit-vector measurements at distinct times. Most previous efforts in this area have assumed that the estimator has a good initialization, that there are multiple vector measurements at each time step, or that some form of rate information is available. This chapter makes none of those assumptions, but it restricts its focus to an axially-symmetric spacecraft and processes sequences of only three measurements at a time. It proposes a solution procedure that formulates the three-measurement case as the solution of a set of coupled nonlinear equations. These equations are manipulated analytically to obtain a smaller number of equations in fewer unknowns, and the solution for the reduced equation set is obtained by gradient-based numerical methods. Two specific reduced nonlinear equation sets are derived, and "truth" model simulations demonstrate the application of the technique.

In the example "truth" model simulations, the proposed solution method successfully determines an initial attitude and angular velocity using only unit vector measurements. The simulations, however, also highlight several practical implementation challenges. First, the practitioner must know the magnitude of the angular velocity well enough in advance to avoid measurement aliasing, or else he or she must account for solution ambiguities. Second, multiple solutions may be possible for a given set of three measurements, so at least one additional measurement must be incorporated to verify consistency and eliminate the incorrect solutions. Finally, the convergence rate for the numerical solution procedures depends strongly on the particular dynamics and measurement set. For a given set of measurements, it may be necessary to try a large number of initial guesses in order to find the true solution. This difficulty can be partially alleviated by a strategy that uses a much smaller number of initial guesses for a given set of measurements, and then slides forward in time to a new measurement set and repeats the process until the solution is obtained. In this manner, little computation is wasted on time intervals where observability is poor. Despite its inherent challenges, the solution strategy of Chapter 2 may be a good candidate for spacecraft attitude determination when its assumptions hold. In particular, it is suitable for initialization of an extended Kalman filter to prevent filter divergence. It is much more efficient than the alternative strategy of running a large bank of extended Kalman filters from different initial attitudes and angular velocities until one of them converges to the correct solution.

Another challenging spacecraft estimation problem has been addressed in Chapter 3. An orbit solution is desired for a set of satellites occupying a single orbital plane, and a rich set of downlink and crosslink ranging measurements is available. The satellite orbits are disturbed by atmospheric drag, however,

and the crosslink signals are delayed by the ionosphere. Chapter 3 proposes a calibration-like scheme by which distributions of atmospheric and ionospheric density are estimated simultaneously with the satellite orbital states. In contrast to previous calibration methods, it exploits the ring-like satellite configuration geometry to more easily estimate spatially correlated distributions, and the modeled distributions are restricted to the region of the orbital plane actually traversed by the satellites.

Chapter 3's linearized observability analysis shows that the joint estimation problem is observable with no *a priori* information. Thus, it should be possible to implement the proposed scheme and obtain estimates of both the satellite orbits and the density distributions. The satellite orbit states are much more observable than the environmental parameter states, however, so the computed density distributions for the atmosphere and ionosphere should not be trusted beyond reason. A special form of Consider covariance analysis is also performed. It suggests that the additional filter complexity required to estimate the environmental parameters may be worthwhile. Even though the resulting density distributions are poorly observable and unreliable, their estimation is likely to improve the accuracy of the satellite orbit state estimates. How much the orbit accuracy would improve is uncertain, due to the lack of a reliable "truth" model for the upper atmosphere and ionosphere. If the true density distributions differ sufficiently from the modeled distributions, or if the assumption of spatial correlation of the disturbances is invalid, then the proposed calibration scheme would suffer. Its general framework, however, could be employed to investigate the effects of better dynamic models for the atmosphere and ionosphere as such models become available.

Chapters 4 and 5 present a new form of Consider covariance analysis for square-root information filters and Rauch-Tung-Striebel square-root information smoothers, respectively. Consider covariance analyses compute the true estimation error covariance or matrix mean square error for a filter or smoother corrupted by some form of modeling error. The new algorithms are capable of analyzing a large variety of model errors. A number of simple examples are included to show how systems with various common modeling errors can be rewritten in a common Consider form that has been developed in this dissertation. The core analysis algorithms can be directly applied to any system written in this form. Monte Carlo simulations provide independent verification of the algorithms.

The new Consider covariance analysis can be applied directly to an existing square-root information filter or smoother to analyze model errors. To accomplish an equivalent analysis using previously-existing techniques, one would need to first convert to a covariance-domain Kalman filter, and then do significant additional pre-processing/model augmentation in order to handle certain model error classes. Alternatively, the true estimation error covariances could be computed by Monte Carlo methods, but only by increasing the computational burden by several orders of magnitude. The derivation of the Consider analysis algorithms, because it is analogous to the square-root information filter/smoother derivation, may be intuitive for practitioners familiar with those estimation techniques. By following the procedures laid out in these chapters' examples, one can readily adapt the analysis algorithms to many different kinds of filter/smoother modeling errors. The results of the examples also provide insight into the behavior of mismatched estimators and some potential applications for Consider covariance analysis. For instance, one can use such an an-

alysis to identify situations in which the estimator may be too optimistic or too pessimistic about estimation accuracy. Consider covariance analysis can also inform the decision to implement a filter/smoother, as opposed to a filter only. In some situations with modeling errors, a smoother may mitigate the error effects, but in other situations smoother estimates may be less accurate than filter estimates. The possibility that a filter could ever be more accurate than a smoother is a counterintuitive result that cannot occur for correctly-modeled linear estimation problems. The examples demonstrate that Consider covariance analysis can be an efficient tool for investigating these and other scenarios.

REFERENCES

[1] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the ASME - Journal of Basic Engineering*, Vol. 82, No. Series D, 1960, pp. 35–45.

[2] Wahba, G., "A Least Squares Estimate of Satellite Attitude," *SIAM Review*, Vol. 7, No. 3, 1965, pp. 409.

[3] Bierman, G. J., *Factorization Methods for Discrete Sequential Estimation*, Academic Press, 1977, pp. 113–122, 135–145, 162–210.

[4] Griffin, R. E. and Sage, A. P., "Sensitivity Analysis of Discrete Filtering and Smoothing Algorithms," *AIAA Journal*, Vol. 7, No. 10, Oct. 1969, pp. 1890–1897.

[5] Sage, A. P. and Melsa, J. L., *Estimation Theory with Applications to Communications and Control*, McGraw-Hill Book Company, 1971, pp. 376–417.

[6] Crassidis, J. L., Markley, F. L., and Cheng, Y., "Survey of Nonlinear Attitude Estimation Methods," *Journal of Guidance, Control, and Dynamics*, Vol. 30, No. 1, Jan.-Feb. 2007, pp. 12–28.

[7] Keat, J., "Analysis of Least-Squares Attitude Determination Routine DOAOP," *Computer Sciences Corporation Report CSC/TM-77/6034*, Feb. 1977.

[8] Shuster, M. D. and Oh, S. D., "Three-Axis Attitude Determination from Vector Observations," *Journal of Guidance and Control*, Vol. 4, No. 1, Jan.-Feb. 1981, pp. 70–77.

[9] Markley, F. L., "Attitude Determination Using Vector Observations and the Singular Value Decomposition," *Journal of the Astronautical Sciences*, Vol. 36, No. 3, July-Sept. 1988, pp. 245–258.

[10] Crassidis, J. L. and Markley, F. L., "Predictive Filtering for Attitude Estimation Without Rate Sensors," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 3, 1997, pp. 522–527.

[11] Psiaki, M. L., "Global Magnetometer-Based Spacecraft Attitude and Rate Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 2, March-April 2004, pp. 240–250.

[12] Markley, F. L. and Sedlak, J. E., "Kalman Filter for Spinning Spacecraft Attitude Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 6, Nov.-Dec. 2008, pp. 1750–1760.

[13] Shuster, M. D., "A Simple Kalman Filter and Smoother for Spacecraft Attitude," *Journal of the Astronautical Sciences*, Vol. 37, No. 1, 1989, pp. 89–106.

[14] Bar-Itzhack, I. Y., "REQUEST: A Recursive QUEST Algorithm for Sequential Attitude Determination," *Journal of Guidance, Control, and Dynamics*, Vol. 19, No. 5, Sept.-Oct. 1996, pp. 1034–1038.

[15] Psiaki, M. L., "Attitude-Determination Filtering via Extended Quaternion Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 23, No. 2, March-April 2000, pp. 206–214.

[16] Challa, M., Natanson, G., and Ottenstein, N., "Magnetometer-Only Attitude and Rate Estimates for Spinning Spacecraft," *AIAA/AAS Astrodynamics Specialist Conference*, 14-17 Aug. 2000, pp. 311–321.

[17] Christian, J. A. and Lightsey, E. G., "The Sequential Optimal Attitude Recursion Filter," *Journal of Guidance, Control, and Dynamics*, Vol. 33, No. 6, Nov.-Dec. 2010, pp. 1787–1800.

[18] Psiaki, M. L., "Generalized Wahba Problems for Spinning Spacecraft Attitude and Rate Determination," *Journal of the Astronautical Sciences*, Vol. 57, No. 1&2, Jan.-June 2009, pp. 73–92.

[19] Natanson, G., Keat, J., and McLaughlin, S., "Sensor and Advanced Attitude Studies: Deterministic Attitude Computation Using Only Magnetometer Data," *Computer Sciences Corporation Report CSC/TM-91/6017, NASA/GSFC, Flight Dynamics Division, 554-FDD-91/010*, March 1991.

[20] Wertz, J. R., editor, *Spacecraft Attitude Determination and Control*, D. Reidel Pub. Co., 1978, pp. 426–428, 487–494, 511–512, 523–531, 758–759.

[21] Schaub, H. and Junkins, J. L., *Analytical Mechanics of Space Systems*, AIAA Education Series, AIAA, 2003, pp. 95–101, 136–138, 149–150.

[22] Reynolds, R. G., "Quaternion Parameterization and a Simple Algorithm for Global Attitude Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 21, No. 4, July-Aug. 1998, pp. 669–671.

[23] Gill, P. E., Murray, W., and Wright, M. H., *Practical Optimization*, Academic Press, 1981, pp. 59–82, 133–141.

[24] Psiaki, M. L., Klatt, E. M., Kintner, P. M., and Powell, S. P., "Attitude Estimation for a Flexible Spacecraft in an Unstable Spin," *Journal of Guidance, Control, and Dynamics*, Vol. 25, No. 1, Jan.-Feb. 2002, pp. 88–95.

[25] Choukroun, D., "Novel Results on Quaternion Modelling and Estimation

from Vector Observations," *AIAA Guidance, Navigation, and Control Conference*, 10-13 Aug. 2009, AIAA-2009-6315.

[26] Jacchia, L. G., "Thermospheric Temperature, Density, and Composition: New Models," *SAO Special Report*, Vol. 375, 1977.

[27] Roberts, C. E., "An Analytic Model for Upper Atmosphere Densities Based Upon Jacchia's 1970 Models," *Celestial Mechanics and Dynamical Astronomy*, Vol. 4, No. 3, 1971, pp. 368–377.

[28] Hedin, A. E., "Extension of the MSIS Thermosphere Model Into the Middle and Lower Atmosphere," *Journal of Geophysical Research*, Vol. 96, No. A2, 1991, pp. 1159–1172.

[29] Picone, J. M., Hedin, A. E., Drob, D. P., and Aikin, A. C., "NRLMSISE-00 Empirical Model of the Atmosphere: Statistical Comparisons and Scientific Issues," *Journal of Geophysical Research*, Vol. 107, No. A12, 2002, pp. 1468–1483.

[30] Marcos, F. A., Wise, J. O., Kendra, M. J., and Grossbard, N., "Advances in Satellite Drag Modeling," *42nd AIAA Aerospace Sciences Meeting and Exhibit*, 5-8 Jan. 2004, AIAA-2004-1254.

[31] Pardini, C. and Anselmo, L., "Comparison and Accuracy Assessment of Semi-empirical Atmosphere Models Through the Orbital Decay of Spherical Satellites," *Journal of Astronautical Sciences*, Vol. 49, No. 2, 2001, pp. 255–268.

[32] Bowman, B. R., Tobiska, W. K., Marcos, F. A., Huang, C. Y., Lin, C. S., and Burke, W. J., "A New Empirical Thermospheric Density Model JB2008 Us-

ing New Solar and Geomagnetic Indices," *AIAA/AAS Astrodynamics Specialist Conference*, 18-21 Aug. 2008, AIAA-2008-6438.

[33] Hedin, A. E. and Carignan, G. R., "Morphology of Thermospheric Composition Variations in the Quiet Polar Thermosphere from Dynamics Explorer Measurements," *Journal of Geophysical Research*, Vol. 90, No. A6, 1985, pp. 5269–5277.

[34] Yunck, T. P., "Coping with the Atmosphere and Ionosphere in Precise Satellite and Ground Positioning," *Environmental effects on spacecraft positioning and trajectories*, 1993, pp. 1–16.

[35] DeBra, D. B., "Drag-free Spacecraft as Platforms for Space Missions and Fundamental Physics," *Classical and Quantum Gravity*, Vol. 14, 1997, pp. 1549–1555.

[36] Nazarenko, A. I., Cefola, P. J., and Yurasov, V., "Estimating Atmospheric Density Variations to Improve LEO Orbit Prediction Accuracy," *AAS/AIAA Spaceflight Mechanics Meeting 1998*, 1998, pp. 1235–1264.

[37] Marcos, F. A., Kendra, M. J., Griffin, J. M., Bass, J. N., Larson, D. R., and Liu, J. J., "Precision Low Earth Orbit Determination Using Atmospheric Density Calibration," *Journal of the Astronautical Sciences*, Vol. 46, No. 4, 1998, pp. 395–409.

[38] Storz, M. F., Bowman, B. R., Branson, M. J. I., Casali, S. J., and Tobiska, W. K., "High Accuracy Satellite Drag Model (HASDM)," *Advances in Space Research*, Vol. 36, No. 12, 2005, pp. 2497–2505.

[39] Mitch, R. H. and Psiaki, M. L., "Local Ionosphere Model Estimation from

Dual-Frequency GNSS Observables," *ION GNSS 2010*, 21-24 Sept. 2010, pp. 301–312.

[40] Maine, K. P., Anderson, P., and Langer, J., "Crosslinks for the Next-Generation GPS," *IEEE Aerospace Conference*, Vol. 4, 8-15 March 2003, pp. 4–1589–4–1596.

[41] Gavish, B. and Kalvenes, J., "The Impact of Intersatellite Communication Links on LEOS Performance," *Telecommunication Systems*, Vol. 8, No. 2, 1997, pp. 159–190.

[42] Chan, V. W. S., "Optical Satellite Networks," *Journal of Lightwave Technology*, Vol. 21, No. 11, Nov. 2003, pp. 2811–2827.

[43] Stadter, P. A., Chacos, A. A., Heins, R. J., and Asher, M. S., "Enabling Distributed Spacecraft System Operations with the Crosslink Transceiver," *IEEE Aerospace Conference*, Vol. 2, 9-16 March 2002, pp. 2–743–2–754.

[44] Kim, J. and Tapley, B. D., "Error Analysis of a Low-Low Satellite-to-Satellite Tracking Mission," *Journal of Guidance, Control, and Dynamics*, Vol. 25, No. 6, Nov.-Dec. 2002, pp. 1100–1106.

[45] Olsen, E. A., Stadter, P. A., and Asher, M. S., "Long-Baseline Differential GPS Based Relative Navigation for Spacecraft with Crosslink Ranging Measurements," *ION GPS 2000*, 19-22 Sept. 2000, pp. 1612–1621.

[46] Bilitza, D. and Reinisch, B. W., "International Reference Ionosphere 2007: Improvements and New Parameters," *Advances in Space Research*, Vol. 42, 2008, pp. 599–609.

[47] Hinks, J. C. and Psiaki, M. L., "Simultaneous Orbit and Atmospheric Density Estimation for a Satellite Constellation," *AIAA/AAS Astrodynamics Specialist Conference*, 2-5 Aug. 2010, AIAA-2010-8258.

[48] Brown, R. G. and Hwang, P. Y. C., *Introduction to Random Signals and Applied Kalman Filtering with Matlab Exercises and Solutions*, John Wiley & Sons, 3rd ed., 1997, pp. 361–367, 428–431.

[49] Mohiuddin, S. and Psiaki, M. L., "Continuous-Time Kalman Filtering with Implicit Discrete Measurement Times," *Journal of Guidance, Control, and Dynamics*, Vol. 34, No. 1, 2011, pp. 148–163.

[50] Curkendall, D. W., *Problems in Estimation Theory with Applications to Orbit Determination*, Ph.D. thesis, UCLA School of Engineering and Applied Science, Sept. 1972, pp. 61–86.

[51] Montenbruck, O. and Gill, E., *Satellite Orbits: Models, Methods, and Applications*, Springer-Verlag, 2000, pp. 265–266, 297–299.

[52] Tapley, B. D., Schutz, B. E., and Born, G. H., *Statistical Orbit Determination*, Academic Press, 2004, pp. 387–438.

[53] D'Appolito, J. and Hutchinson, C., "Low Sensitivity Filters for State Estimation in the Presence of Large Parameter Uncertainties," *IEEE Transactions on Automatic Control*, Vol. 14, No. 3, 1969, pp. 310–312.

[54] Gelb, A., editor, *Applied Optimal Estimation*, The Analytical Sciences Corporation, The M.I.T. Press, 1974, pp. 229–276.

[55] Lichten, S. M. and Border, J. S., "Strategies for High-Precision Global Positioning System Orbit Determination," *Journal of Geophysical Research*, Vol. 92, No. B12, Nov. 1987, pp. 12751–12762.

[56] Schutz, B. E., Cheng, M. K., Shum, C. K., Eanes, R. J., and Tapley, B. D., "Analysis of Earth Rotation Solution from Starlette," *Journal of Geophysical Research*, Vol. 94, No. B8, Aug. 1989, pp. 10167–10174.

[57] Bierman, G. J., "The Treatment of Bias in the Square-Root Information Filter/Smoother," *Journal of Optimization Theory and Applications*, Vol. 16, No. 1, 1975, pp. 165–178.

[58] Nishimura, T., "On the a priori Information in Sequential Estimation Problems," *IEEE Transactions on Automatic Control*, Vol. 11, No. 2, 1966, pp. 197–204.

[59] Heffes, H., "The Effect of Erroneous Models on the Kalman Filter Response," *IEEE Transactions on Automatic Control*, Vol. 11, No. 3, July 1966, pp. 541–543.

[60] Huddle, J. and Wismer, D., "Degradation of Linear Filter Performance Due to Modeling Error," *IEEE Transactions on Automatic Control*, Vol. 13, No. 4, 1968, pp. 421–423.

[61] Fagin, S. L., "Recursive Linear Regression Theory, Optimal Filter Theory and Error Analysis of Optimal Systems," *IEEE international convention record*, Vol. 12, 1964, pp. 216–245.

[62] Griffin, R. E. and Sage, A. P., "Large and Small Scale Sensitivity Analysis of Optimum Estimation Algorithms," *IEEE Transactions on Automatic Control*, Vol. 13, No. 4, Aug. 1968, pp. 320–329.

[63] Morf, M., Verriest, E., Dobbins, J., and Kailath, T., "Square-Root Algorithms for Model Sensitivity Analysis," *Proc. of the 1977 Conf. on Information Sciences and Systems*, Vol. 11, Mar. 30-Apr. 1, 1977, pp. 64–69.

[64] Lewis, F. L., *Optimal Estimation*, John Wiley & Sons, 1986, pp. 205–248.

[65] Bar-Shalom, Y., Li, X. R., and Kirubarajan, T., *Estimation with Applications to Tracking and Navigation*, John Wiley & Sons, Inc., 2001, pp. 248–261.

[66] Schmidt, S. F., "Applications of State-Space Methods to Navigation Problems," *Advances in Control Systems*, edited by C. T. Leondes, Vol. 3, Academic Press, New York, 1966, pp. 293–340.

[67] Woodbury, D. P. and Junkins, J. L., "On the Consider Kalman Filter," *AIAA Guidance, Navigation, and Control Conference*, 2-5 Aug. 2010, AIAA-2010-7752.

[68] Dyer, P. and McReynolds, S., "Extension of Square-Root Filtering to Include Process Noise," *Journal of Optimization Theory and Applications*, Vol. 3, No. 6, 1969, pp. 444–458.

[69] Rauch, H. E., Tung, F., and Striebel, C. T., "Maximum Likelihood Estimates of Linear Dynamic Systems," *AIAA Journal*, Vol. 3, No. 8, Aug. 1965, pp. 1445–1450.

[70] Wall, J. E., Willsky, A. S., and Sandell, N. R., "On the Fixed-Interval Smoothing Problem," *Stochastics*, Vol. 5, No. 1-2, 1981, pp. 1–41.

[71] Bierman, G. J., "Sequential Square Root Filtering and Smoothing of Discrete Linear Systems," *Automatica*, Vol. 10, 1974, pp. 147–158.