THE ADAPTIVE KALMAN FILTER AND MICROMECHANICAL INERTIAL INSTRUMENT PERFORMANCE

by

Jonathan Andrew Kossuth

S.B., Massachusetts Institute of Technology (1993)

Submitted to the Department of Aeronautics and Astronautics
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE in AERONAUTICS AND ASTRONAUTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1993

© Jonathan Andrew Kossuth, 1993. All Rights Reserved

Signature of Author Jonathan Kossuth
Department of Aeronautics and Astronautics
May 7, 1993

Approved by Paul Steranka
Paul Steranka
Principal Member Technical Staff, Charles Stark Draper Laboratory
Thesis Supervisor

Certified by Wallace E Vander Velde
Professor Wallace E. Vander Velde
Department of Aeronautics and Astronautics
Thesis Advisor

Accepted by Harold Y. Wachman
Professor Harold Y. Wachman
Chairman, Department Graduate Committee

Aero

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 08 1993

LIBRARIES

THE ADAPTIVE KALMAN FILTER
AND
MICROMECHANICAL INERTIAL INSTRUMENT PERFORMANCE

by

Jonathan Andrew Kossuth

Submitted to the Department of Aeronautics and Astronautics
on May 7, 1993
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Aeronautics and Astronautics

ABSTRACT

As newer technologies in inertial instruments emerge, the need for more powerful data analysis techniques is increasing because both unmodeled errors and unknown factors may exist, and must be accounted for, in the instrument data. An online adaptive Kalman filter has been developed for analysis of these inertial instruments using the innovation sequence of the Kalman filter in order to determine the optimal filter parameters.

Performance of the filter is improved by determining the proper system model, by identifying any unknown parameters in the system matrices, or by identifying any periodic noise in the signal. A fault tolerant algorithm is included in the filter. Maximum likelihood techniques are used to estimate the parameters of the filter matrices by choosing those parameter values that maximize the likelihood function of the parameters. Correlation methods are used to determine the proper system model by comparing the innovation sequence to a known signal. Power spectral density analysis is used to identify periodic signals by examining the PSD of the innovation sequence.

Simulations to verify the filter performance have been run successfully for a number of cases. In each of these simulations, the adaptive Kalman filter was able to modify the filter matrices to achieve optimality. The adaptive Kalman filter has been applied to data from various micromechanical gyroscope tests, including a stationary drift test and a commanded rate test. These filters have been used for both raw data reduction and reduced data analysis. By applying these adaptive Kalman filters to micromechanical gyroscope data, real improvements in data analysis have been shown. The raw data reduction adaptive filter improved data decimation by up to 50 percent over a triangular filter. The reduced data analysis adaptive filter produced filter estimates twice as accurate as the traditional Kalman filter, and has identified a pressure sensitivity and a rate squared sensitivity in the micromechanical gyroscope.

Thesis Supervisor: Paul Steranka
Principal Member Technical Staff, C. S. Draper Laboratory

Thesis Advisor: Professor Wallace E. Vander Velde
Department of Aeronautics and Astronautics

Acknowledgments

I would like to thank Charles Kochakian for his support over the past three years. My sincerest gratitude is extended to my thesis supervisor, Paul Steranka, for his guidance and assistance for the past year. I would like to thank my thesis advisor, Professor Vander Velde, for his insightful comments that helped this thesis come together. I would also like to thank Anthony Kourepenis and Thomas Farnkoff for their assistance in acquiring data, as well as their explanations of micromechanical inertial instrument technology.

I thank my friends at MIT for the help and support that they have given me and for making the past five years at MIT enjoyable. I also thank the people at Draper Lab that made my day to day life interesting. My family, especially my parents, George and Elizabeth Kossuth, deserves tremendous thanks for encouraging me to do my best over the past 23 years. I give special thanks to Kimberly Santagate for her love and friendship over the past five years.

This thesis was researched and written at the Charles Stark Draper Laboratory under Contract IR&D 93-1-357-1.

Publication of this thesis does not constitute approval by the Laboratory of the findings or conclusions contained herein. It is published for the exchange and stimulation of ideas.

I hereby assign my copyright of this thesis to the Charles Stark Draper Laboratory, Inc., of Cambridge, Massachusetts.

Jonathan Andrew Kossuth
May 7, 1993

Permission is hereby granted by the Charles Stark Draper Laboratory, Inc. to the Massachusetts Institute of Technology to reproduce and to distribute copies of this thesis document in whole or in part.

Table of Contents

Abstract 3
Acknowledgments 5
Table of Contents 7
List of Figures 9
List of Tables 11
Chapter 1 Introduction 13
1.1 Motivation 13
1.2 Adaptive Filter Concepts 14
1.2.1 Kalman Filter 14
1.2.2 Maximum Likelihood Estimator 14
1.2.3 Model Adaptive Filter 15
1.2.4 Periodic Signal Identification 15
1.2.5 Data Analysis Filters 15
1.3 Data Analysis 16
1.3.1 Micromechanical Inertial Instruments 16
1.3.2 Data Analysis 16
1.3.3 Future Work 16
Chapter 2 Kalman Filter Theory 17
2.1 History of Estimators 17
2.2 Kalman Filter Theory 17
2.3 The Innovation Property of an Optimal Filter 19
2.3.1 Innovation Sequence as White Noise in an Optimal Filter 19
2.3.2 White Noise Test for Optimality of Kalman Filter 21
2.4 Limitations of Kalman Filter 23
Chapter 3 Maximum Likelihood Estimator 25
3.1 Motivation for Maximum Likelihood Estimator 25
3.2 Maximum Likelihood Estimator 25
3.3 Maximum Likelihood Estimator Theory 26
3.4 Derivation of the Maximum Likelihood Kalman Filter 31
3.5 Equations for Maximum Likelihood Kalman Filter 34
3.5.1 Approximations for Online Operation 40
3.6 Verification of Maximum Likelihood Estimator 41
3.6.1 Generation of Simulated Data 42
3.6.1.1 Generation of White Noise Sequences 42
3.6.1.2 Generation of Random Walk Sequences 42
3.7 Results of Simulated Data Analysis 43
3.8 Conclusions 45
Chapter 4 Model Adaptive Filter Using Correlation Methods 47
4.1 Model Adaptive Filter 47
4.2 Theory of Correlation Functions 47
4.3 Derivation of Model Adaptive Filter 48
4.3.1 Implementation of Model Adaptive Filter 50
4.4 Verification of Model Adaptive Filter 51
4.5 An Example of the Model Adaptive Filter in a Historical Context 56

4.6 Conclusions 59
Chapter 5 Power Spectral Density Analysis 61
5.1 Motivation for PSD Analysis 61
5.2 Power Spectral Density Theory 61
5.2.1 Fourier Transforms 62
5.2.2 Power Spectral Density Function for Finite Data Sets 63
5.2.3 Discrete Fourier Transforms 65
5.2.4 Nyquist Criterion and Aliasing 65
5.3 Filter Design 67
5.4 Filter Verification 67
5.4.1 Single and Multiple Frequency Noise 67
5.4.2 Distributed Frequency Noise 71
5.4.2.1 Motivation 71
5.4.2.2 Approach 72
5.4.2.3 Results 72
5.5 Conclusions 77
Chapter 6 Adaptive Data Analysis Filter Development 79
6.1 Introduction 79
6.2 Raw Data Reduction Adaptive Filter 79
6.2.1 Verification of Raw Data Reduction Filter 79
6.2.2 Raw Data Reduction Adaptive Filter vs. Triangular Filter 83
6.3 Reduced Data Analysis Adaptive Filter 85
6.4 Conclusions on Adaptive Filters 91
Chapter 7 Micromechanical Inertial Instruments 93
7.1 Introduction 93
7.2 Micromechanical Gyroscopes 93
7.2.1 Vibratory Gyroscope 93
7.2.2 Noise Sources 96
7.2.3 Inverted Gyroscope 96
7.2.4 Tuning Fork Gyroscope 97
7.3 Conclusions 98
Chapter 8 Micromechanical Gyroscope Data Analysis 99
8.1 Approach 99
8.2 Stationary Drift Test with Original Gyro Design, Test 8A 99
8.2.1 Raw Data Reduction of 3 Hz Stationary Drift Test 99
8.2.2 Reduced Data Analysis of 3 Hz Stationary Drift Test 101
8.3 Scale Factor Test with Inverted Gyroscope Design, Test 8B 102
8.3.1 Reduced Data Analysis of Commanded Rate Test 103
8.4 Stationary Drift Test with Inverted Design, Test 8C 106
8.4.1 Raw Data Reduction of 1500 Hz Stationary Drift Test 107
8.4.2 Reduced Data Analysis of 1500 Hz Stationary Drift Test 107
8.5 Pressure Variation Test for Inverted Design 109
8.6 Commanded Rate Test for Original Design 111
8.7 Conclusions 112
Chapter 9 Conclusions 113
9.1 Results 113
9.2 Implementation of Adaptive Filters 114
9.3 Recommendations for Future Work 114
References 115

List of Figures

Figure 2.1. Descriptions of White Noise Properties 21
Figure 3.1a. Estimate Intervals at Every Sample Time 26
Figure 3.1b. Estimate Intervals at Every N Sample Times 26
Figure 3.2. Flowchart for Maximum Likelihood Estimation 35
Figure 3.3. Typical Parameter Estimate Convergence 44
Figure 3.4. Typical Parameter Estimate Standard Deviation 44
Figure 3.5. Convergence of Measurement Noise Estimate 45
Figure 4.1. Autocorrelation of $P(t)$ 49
Figure 4.2. Crosscorrelation of $S(t)$ and $P(t)$ with Scaling and Delay 49
Figure 4.3. Unmodeled Trend Innovation Sequence, Test 4A1 53
Figure 4.4. Modified Innovation Sequence, Test 4A1 53
Figure 4.5. Unmodeled Scale Factor Innovation Sequence, Test 4A3 54
Figure 4.6. Modified Innovation Sequence, Test 4A3 54
Figure 4.7. Unmodeled Rate^2 Term Innovation Sequence, Test 4A5 55
Figure 4.8. Modified Innovation Sequence, Test 4A5 55
Figure 4.9. Gas Bearing Gyro Output from Shaker Test, Test 4B 56
Figure 4.10. D_{IS} System Model Innovation Sequence, Test 4B1 58
Figure 4.11. Model Adaptive Filter Innovation Sequence, Test 4B2 58
Figure 5.1. Flowchart for PSD Adaptive Filter 66
Figure 5.2. Innovation Sequence for Test 5A1 69
Figure 5.3. Innovation Sequence for Test 5A3 69
Figure 5.4. Innovation Sequence for Test 5A4 70
Figure 5.5. Innovation Sequence for Test 5A5 70
Figure 5.6. Probability of Detecting Frequency and Noise 73
Figure 5.7. Aliased Frequency vs. Sampling Rate for 4400 Hz Signal 74
Figure 5.8. Probabilities for 2200 Hz Signal 75
Figure 5.9. Probabilities for 6600 Hz Signal 76
Figure 6.1. Innovation Sequence for Test 6A1 81
Figure 6.2. Innovation Sequence for Test 6A2 81
Figure 6.3. PSD of Original Raw Data for Test 6A2 82
Figure 6.4. PSD of Innovation Sequence for Test 6A2 82
Figure 7.1. Vibratory Micromechanical Gyroscope 94
Figure 7.2. Schematic of Micromechanical Gyroscope Operation 95

Figure 7.3. Schematic of Tuning Fork Gyroscope 97
Figure 7.4. Tuning Fork Gyro Operation 98
Figure 8.1. Power Spectral Density for Test 8A 100
Figure 8.2. Standard Deviation vs. Bandwidth for Test 8A1 100
Figure 8.3. Commanded Rate Profile for Test 8B 102
Figure 8.4. Inverted Gyro Output for Test 8B 103
Figure 8.5. Innovation Sequence of Kalman Filter for Test 8B1 104
Figure 8.6. Innovation Sequence of Adaptive Filter for Test 8B2 104
Figure 8.7. Bias Estimate of Kalman Filter for Test 8B1 105
Figure 8.8. Bias Estimate of Adaptive Filter for Test 8B2 105
Figure 8.9. PSD of Gyro Output for Stationary Drift Test, Test 8C 106
Figure 8.10. Standard Deviation vs. Bandwidth for Test 8C1 107
Figure 8.11. Gyroscope Output for Test 8D 109
Figure 8.12. Modified Innovation Sequence, Test 8D2 110

List of Tables

Table 3.1. Results of Maximum Likelihood Estimator Verification 43
Table 4.1. Results of Model Adaptive Filter Verification 51
Table 4.2. Model Adaptive Filter Analysis of Gas Bearing Gyro Data 59
Table 5.1. Results of Periodic Noise Identification 68
Table 5.2. Probabilities for Aliased 4400 Hz Frequency Signal 73
Table 5.3. Probabilities for Aliased 2200 Hz Frequency Signal 75
Table 5.4. Probabilities for Aliased 6600 Hz Frequency Signal 76
Table 6.1. Results of Raw Data Reduction Filter Verification 80
Table 6.2. Results of Reduced Data Analysis Filter Verification 87
Table 6.3. Kalman Filter versus Reduced Data Analysis Filter 89
Table 8.1. Data Analysis on Stationary Drift Data, Test 8A2 101
Table 8.2. Data Analysis on Stationary Drift Data, Test 8A3 101
Table 8.3. Data Analysis on Stationary Drift Data, Test 8A4 101
Table 8.4. Results for Commanded Rate Test, Test 8B 103
Table 8.5. Data Analysis on High Frequency Drift Data, Test 8C2 108
Table 8.6. Data Analysis on High Frequency Drift Data, Test 8C3 108
Table 8.7. Data Analysis on High Frequency Drift Data, Test 8C4 108
Table 8.8. Results for Pressure Variation Test, Test 8D 110
Table 8.9. Results for Commanded Rate Test, Test 8E 111

Chapter 1

Introduction

1.1 Motivation

As newer inertial instrument technologies emerge, better data processing is needed for characterizing these instruments. These newer instruments are more sophisticated and their behavior challenges traditional data processing techniques. Traditional data processing has commonly used a triangular filter for reducing raw data and a Kalman filter for analyzing this reduced data according to a prescribed instrument model. These filters, however, can result in a less-than-optimum indication of instrument quality because both unmodeled error sources in the original signal, such as high frequency noise, and unknown factors in the instrument performance, such as temperature effects, are not accounted for in either aspect of the data processing. Because it gives more flexibility to data analysis, the need for adaptive filters has become significant. Instead of using fixed values, the adaptive filter seeks out the true values of the system so that both filter performance and instrument characterization are optimized.

Adaptive filter concepts can be used to develop filters for two aspects of data analysis. Raw data reduction is an important aspect of data processing. By decimating data, the computing time of the filter is reduced without sacrificing information on instrument performance. The traditional triangular filters create a weighted average of the raw data, but if noise in the signal can be eliminated before decimation of the data, then this analysis can be improved.

Analysis of this decimated data is crucial to determining instrument performance. To obtain optimal instrument performance, an adequate model of both the state vector and all noise sources is necessary. Often, though, the noise model is poorly understood, and unmodeled terms may exist in the system. A Kalman filter only analyzes the data with

respect to given parameters and a given state vector. With an adaptive filter, both the state vector and noise parameters can be adequately determined during data processing.

1.2 Adaptive Filter Concepts

In the next four chapters, adaptive Kalman filter concepts are developed and verified independently. First, the Kalman filter is reviewed. Next, a maximum likelihood estimator is developed. Correlation methods are then used to identify unmodeled state terms. Finally, power spectral density analysis is used to locate periodic noise sources.

1.2.1 Kalman Filter

Chapter 2 presents the Kalman filter. The innovation sequence (also known as the residual) is discussed, and its importance in adaptive filters is shown. In an optimal filter, the innovation sequence is a Gaussian white noise process. However, if the innovation sequence is not white noise, then information can be extracted from the sequence to optimize the filter. A white noise test procedure using correlation methods is presented. Some limits of the Kalman filter are discussed, including its inability to change parameters to reflect the data more accurately, and its inability to adapt the state model to achieve an optimal model.

1.2.2 Maximum Likelihood Estimator

In Chapter 3, maximum likelihood estimation of system parameters is presented. With this technique, unknown parameters in the Kalman filter matrices, such as the measurement noise standard deviation, can be estimated by maximizing the likelihood function of these parameters. By assuming that parameter values are slowly changing with respect to the state estimates, an online estimator is developed.

The maximum likelihood estimator was chosen for several reasons. It is capable of obtaining an unbiased efficient estimate, and it converges to the true values of the parameters as the number of samples grows without bound. This estimator is capable of estimating the true values of both constant and slowly varying parameter values.

The scoring method is used for the maximum likelihood estimator, because this method has advantages over both a Newton-Raphson method and a conventional gradient method. The scoring method converges over a wider range than the Newton-Raphson method, and it converges more rapidly than a gradient method. The error in the scoring method is of order $1/N$, where N is the number of samples. Verification of this approach is given with several analyses of simulated data.

1.2.3 Model Adaptive Filter

Chapter 4 discusses a model adaptive filter using correlation methods. By examining the relationship between the autocorrelation function of a random process, such as the instrument temperature, and the crosscorrelation function between that process and another random process, in this case the innovation sequence, a relationship between the two processes can be identified. A random process that is a function of another random process will have a crosscorrelation function that is equivalent to the autocorrelation function of the original signal, except for scaling and a time delay. If such a relationship does exist, the state vector of the Kalman filter is modified to account for this previously unmodeled term. The theory for this approach is derived, and verification of this filter using simulated data is shown. Real data is then used to show the effectiveness of the this filter in identifying a component of gyro drift error of an electromechanical gyroscope model due to g^2 terms, D_{OS}. This accomplishment is then placed in an historical context.

1.2.4 Periodic Signal Identification

Chapter 5 develops an approach for identifying unmodeled periodic signals in the innovation sequence. By using the Fast Fourier Transform, power spectral density analysis can identify any periodic signals in the innovation sequence. These unmodeled periodic signals can be included in the state model either as noise or as state terms; in both cases the filter performance will be improved. Although a periodic signal may be aliased, it can still be removed from the data. Once again, the theory behind this filter is given; verification of its effectiveness is demonstrated using simulated data for multiple frequency noise signals and also for variations in the frequency of the noise.

1.2.5 Data Analysis Filters

In Chapter 6, the three adaptive filter concepts are combined into two filters. The first adaptive filter, used for raw data reduction, is a combination of the maximum likelihood estimator and the PSD filter. By optimizing the filter parameters and by eliminating periodic noise, the best estimates of the data can be used for data decimation. Verification of the operation of this filter is given, and the performance of this filter is compared with that of a triangular filter for data decimation.

The second filter developed in Chapter 6 is a combination of all three adaptive filter concepts. This filter is used for reduced data analysis. By using all three techniques, the optimal estimates of the matrix parameters, as well as an accurate state model, can be

obtained. Verification of this filter is shown, and the improvements of this filter over a traditional Kalman filter are shown for simulated typical gyro outputs.

1.3 Data Analysis

After developing these two adaptive filters, micromechanical inertial instrument technology is discussed, and potential areas for the use of these filters are identified. Data files are then analyzed from several test runs, and improvements in instrument performance are shown.

1.3.1 Micromechanical Inertial Instruments

In Chapter 7, micromechanical inertial instrument technology is introduced, and micromechanical gyroscopes are briefly surveyed. The advantages and disadvantages of different gyroscope designs are presented. Also, the limitations and noise sources of these instruments are discussed, and possible solutions using adaptive filters are presented.

1.3.2 Data Analysis

In Chapter 8, adaptive filters are applied to outputs from various micromechanical gyroscopes. Comparisons between the adaptive filters and the traditional triangular and Kalman filter approach are made for all data analyzed. Various tests, including stationary drift tests and commanded rate tests, were run on different micromechanical gyroscopes, both the original micromechanical gyro design and the inverted gyro design. Improvements in data analysis were shown in both raw data reduction and reduced data analysis. The raw data reduction filter improvement ranged from 10 percent to 50 percent better than a triangular filter. The reduced data analysis filter is capable of reducing the RMS of the innovation sequence to half the RMS of the Kalman filter. A pressure sensitivity and a rate squared sensitivity were identified by the adaptive filter.

1.3.3 Future Work

Chapter 9 presents the conclusions of this thesis, as well as some recommendations for future work. The algorithm used to compute the correlation functions can become computationally burdensome. A more efficient algorithm would increase online capabilities. This filter, if implemented on parallel processors, would easily be capable of real-time processing. By assigning the model adaptive and frequency location tests to secondary processors, the main processor could continue forward with data analysis, and the state model can be adjusted as necessary.

Chapter 2

Kalman Filter Theory

2.1 History of Estimators

The motivation to develop an effective estimator has existed for hundreds of years. In 1795, at the age of 18, Karl Friedrich Gauss developed the least squares method. His motivation was to find a method to determine the motion of the planets around the sun. In his work, *Theoria Motus Corporum Coelestium*, Gauss realized that single measurements were not accurate, but, by taking multiple measurements, the error of these measurements could be minimized. In the 1940s, Wiener and Kolmogorov both developed linear minimum mean-square estimators that spurred the development of the Kalman filter. Wiener developed an algorithm that minimized the mean-square error by choosing appropriate filter gains. Kalman then developed both the discrete and continuous versions of the Kalman filter in the early 1960s. The Kalman filter is basically a recursive solution to Gauss' original problem: the least squares estimator. Because of its computational efficiency and simplicity, the Kalman filter is an ideal choice for inertial instrument evaluation [45].

2.2 Kalman Filter Theory

A linear system can be expressed by the following stochastic difference equation,

$$\mathbf{x}(t_{i+1}) = \Phi(t_{i+1}, t_i)\mathbf{x}(t_i) + \mathbf{B}(t_i)\mathbf{u}(t_i) + \mathbf{G}(t_i)\mathbf{w}(t_i) \tag{2.1}$$

with available discrete-time measurements modeled by the linear relation,

$$\mathbf{z}(t_i) = \mathbf{H}(t_i)\mathbf{x}(t_i) + \mathbf{v}(t_i) \tag{2.2}$$

where,

$\mathbf{x}$ = the state vector = n x 1 vector

Φ = state transition matrix = n x n matrix

$\mathbf{u}$ = deterministic input vector = r x 1 vector

$\mathbf{B}$ = deterministic input matrix = n x r matrix

$\mathbf{G}$ = system plant noise input matrix = n x q matrix

$\mathbf{z}$ = the measurement vector = m x 1 vector

$\mathbf{H}$ = observability matrix = m x n matrix

t_i = a discrete measurement time

and,

$\mathbf{w}$ and $\mathbf{v}$ are independent, zero mean, white Gaussian noise processes with covariances

$$E\left\{\mathbf{w}(t_i)\mathbf{w}(t_j)^{\mathrm{T}}\right\} = \mathbf{Q}(t_i)\delta_{ij} \tag{2.3}$$

$$E\left\{\mathbf{v}(t_i)\mathbf{v}(t_j)^{\mathrm{T}}\right\} = \mathbf{R}(t_i)\delta_{ij} \tag{2.4}$$

where $\mathbf{Q}$ is positive semidefinite and $\mathbf{R}$ is positive definite for all discrete time t_i, $E\{\bullet\}$ is the expectation function, and δ_{ij} is the Kroneker delta. The initial conditions of the state vector are described by a Gaussian random vector with mean equal to Equation 2.5,

$$E\left\{\mathbf{x}(t_0)\right\} = \hat{\mathbf{x}}(t_0) \tag{2.5}$$

and covariance equal to Equation 2.6,

$$E\left\{\left[\mathbf{x}(t_0) - \hat{\mathbf{x}}(t_0)\right]\left[\mathbf{x}(t_0) - \hat{\mathbf{x}}(t_0)\right]^{\mathrm{T}}\right\} = \mathbf{P}(t_0) \tag{2.6}$$

The measurements, $\mathbf{z}$, are processed to produce the state estimates, $\hat{\mathbf{x}}$. The *state transition equations* are

$$\hat{\mathbf{x}}(t_i^-) = \Phi(t_i, t_{i-1})\hat{\mathbf{x}}(t_{i-1}^+) + \mathbf{B}(t_{i-1})\mathbf{u}(t_{i-1}) \tag{2.7}$$

$$\mathbf{P}(t_i^-) = \Phi(t_i, t_{i-1})\mathbf{P}(t_{i-1}^+)\Phi(t_i, t_{i-1})^{\mathrm{T}} + \mathbf{G}(t_{i-1})\mathbf{Q}(t_{i-1})\mathbf{G}(t_{i-1})^{\mathrm{T}} \tag{2.8}$$

$$\mathbf{K}(t_i) = \left[\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}}\right]\mathbf{S}(t_i)^{-1} \tag{2.9}$$

$$\mathbf{S}(t_i) = \mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}} + \mathbf{R}(t_i) \tag{2.10}$$

The *measurement incorporation equations* introduce the new measurement into the state vector estimates, as shown in Equation 2.11 through Equation 2.13.

$$\mathbf{r}_i = \mathbf{z}_i - \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-) \tag{2.11}$$

$$\hat{\mathbf{x}}(t_i^+) = \hat{\mathbf{x}}(t_i^-) + \mathbf{K}(t_i)\mathbf{r}(t_i) \tag{2.12}$$

$$\mathbf{P}(t_i^+) = \left[\mathbf{I} - \mathbf{K}(t_i)\mathbf{H}(t_i)\right]\mathbf{P}(t_i^-) \tag{2.13}$$

where,

$\hat{\mathbf{x}}$ = the estimate of **x** based upon the most recent measurement

r = innovation sequence

S = covariance of innovation sequence

and,

K = the Kalman gain matrix = n x m matrix.

The (+) and (-) superscripts on t_i indicate whether the value at a particular t_i is immediately before or immediately after the measurement incorporation at time t_i. The notation (t_{i+1}, t_i) on Φ means that the matrix Φ is the transition between the state vector at sample times t_i and t_{i+1}. In these recursive equations, Φ, **H**, **B**, and **G** may be time varying.

2.3 The Innovation Property of an Optimal Filter

In an optimal Kalman filter, the innovation sequence **r** (also known as the one step ahead predictor or residual shown in Equation 2.11) is a Gaussian white noise sequence with covariance **S** (Equation 2.10) [33, 34]. The innovation sequence is used to update $\hat{\mathbf{x}}(t_i^-)$ by incorporating the information from the newest measurement, as shown in Equation 2.12. By proving that **r** is white noise in an optimal filter, this property of **r** can then be used for filter optimization.

2.3.1 Innovation Sequence as White Noise in an Optimal Filter

First, define the error $\mathbf{e}(t_i)$ as the difference between the true value of $\mathbf{x}(t_i)$ and its estimate $\hat{\mathbf{x}}(t_i^-)$:

$$\mathbf{e}(t_i) = \mathbf{x}(t_i) - \hat{\mathbf{x}}(t_i^-) \tag{2.14}$$

This leads to a new definition of the innovation sequence,

$$\mathbf{r}_i = \mathbf{H}\mathbf{e}(t_i) + \mathbf{v}(t_i) \tag{2.15}$$

and its autocorrelation function,

$$E\left\{\mathbf{r}_i\mathbf{r}_j^{\mathrm{T}}\right\} = E\left\{\left(\mathbf{He}(t_i)+\mathbf{v}(t_i)\right)\left(\mathbf{He}(t_j)+\mathbf{v}(t_j)\right)^{\mathrm{T}}\right\} \tag{2.16}$$

with **H** and **v** defined above. For $i > j$, $\mathbf{v}(t_i)$ is independent of both $\mathbf{e}(t_j)$ and $\mathbf{v}(t_j)$, so that

$$\begin{aligned} E\left\{\mathbf{r}_i\mathbf{r}_j^{\mathrm{T}}\right\} &= E\left\{\mathbf{He}(t_i)\left(\mathbf{He}(t_j)+\mathbf{v}(t_j)\right)^{\mathrm{T}}\right\} \\ &= E\left\{\mathbf{He}(t_i)\left(\mathbf{z}_j - \mathbf{H}\hat{\mathbf{x}}(t_j^-)\right)^{\mathrm{T}}\right\} \end{aligned} \tag{2.17}$$

Also, the orthogonality principle [33] states that $\mathbf{e}(t_i)$ is orthogonal to $\mathbf{z}_k$, for $k < i$. Since $\hat{\mathbf{x}}(t_j^-)$ depends only on $\mathbf{z}_k$, for $k < j$, Equation 2.17 becomes

$$E\left\{\mathbf{r}_i\mathbf{r}_j^{\mathrm{T}}\right\} = 0 \quad \text{for } i > j \tag{2.18}$$

A similar exercise shows that this is true for all $i < j$. For $i = j$,

$$E\left\{\mathbf{r}_i\mathbf{r}_i^{\mathrm{T}}\right\} = \mathbf{S}(t_i) = \mathbf{H}(t_i)\mathbf{P}(t_i^-)\mathbf{H}(t_i)^{\mathrm{T}} + \mathbf{R}(t_i) \tag{2.19}$$

Therefore, the innovation sequence is a white noise process, with only one nonzero term in its autocorrelation sequence at $i = j$.

This can also be proved by examining Equation 2.11. The measurement $\mathbf{z}_i$ is a Gaussian random process, as is the state estimate $\hat{\mathbf{x}}(t_j^-)$. Therefore, the innovation sequence is also a Gaussian random process because it is a linear sum of Gaussian random variables. The innovation sequence is also a stationary process, as demonstrated by Mehra [33].

When the innovation sequence is white noise, then the filter is optimal. However, if the innovation sequence is not Gaussian white noise, then information required to optimize the filter can be extracted from the innovation sequence. For example, suppose a system is modeled with the following state vector

$$\hat{\mathbf{x}} = \left[\hat{x}_1\right] \tag{2.20}$$

But the measurements fit the model,

$$\mathbf{z}_i = \begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \mathbf{v}(t_i) \tag{2.21}$$

Because of the inadequate model for the state vector, the **H** matrix that is implemented in the Kalman filter is the 1 by 1 matrix [1], not the 1 by 2 matrix [1 1] that would be

expected from the measurement model. It is obvious that the state vector is an inadequate model for this measurement. This mismodeling will be stored in the innovation sequence,

$$\mathbf{r}_i = \mathbf{z}_i - \mathbf{H}(t_i)\hat{\mathbf{x}}(t_i^-) \tag{2.11}$$

which, in this example, becomes,

$$\mathbf{r}_i = x_2(t_i) + \left[x_1(t_i) - \hat{x}_1(t_i)\right] + \mathbf{v}(t_i) \tag{2.22}$$

In Equation 2.22, the difference term in x_1 is an unbiased Gaussian process. The innovation sequence **r** will contain the unmodeled term x_2, and **r** will no longer be a Gaussian white noise sequence. However, the information needed to optimize the filter can be extracted from the innovation sequence. By adapting the matrices **H** and **x** to include the x_2 term, the filter will be optimized.

2.3.2 White Noise Test for Optimality of Kalman Filter

A test is needed in order to exploit the innovation sequence for adaptive filtering. The white noise property of the innovation sequence in an optimal filter discussed in Section 2.3.1 will be used to develop an effective test. White noise has unique behavior for both the autocorrelation function and power spectral density analysis, as shown in Figure 2.1 [9]. Because a white noise process is independent, the autocorrelation sequence of white noise is an impulse at the zeroth lag term. Also, a white noise sequence is evenly distributed throughout the frequency domain, so that the power spectral density of white noise is a horizontal line. Either of these two properties of white noise could be used in a white noise test.

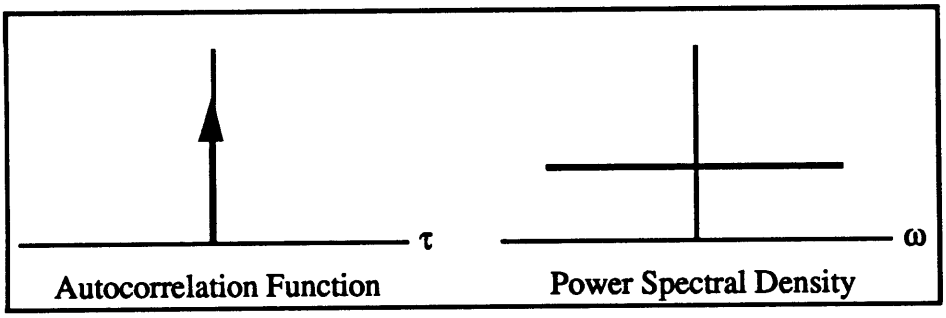

Figure 2.1. Descriptions of White Noise Properties

The autocorrelation function was chosen for identification of the innovation sequence as white noise for two reasons. First, Mehra [33] shows that this function can be used to determine the whiteness of a sequence. Second, a test using the autocorrelation function

can be implemented easily. After calculating the autocorrelation function, confidence limits for this process can be established. The standard deviation of the measurement of autocorrelation is $1/\sqrt{N}$, where N is the length of the innovation sequence. The confidence limits would be a given value times this standard deviation. For 95% confidence limits, the value is 1.96; for 99% confidence limits, the value is 2.58. Next, the number of points in the autocorrelation function that exceed these limits must be determined. If the percentage of points that exceed the confidence limits is greater than the confidence limit percentage, then the innovation sequence is not white noise. This test procedure is outlined in Equation 2.23 through Equation 2.25.

First, define $\mathbf{C}_k$, the autocorrelation matrix of $\mathbf{r}$, as shown in Equation 2.23.

$$\mathbf{C}_k \equiv E\left\{\mathbf{r}_i \mathbf{r}_{i-k}^{\mathrm{T}}\right\} \tag{2.23}$$

A biased estimate of $\mathbf{C}_k$ may be obtained using Equation 2.24. For a scalar measurement, the estimate of $\mathbf{C}_k$ will be a scalar value. Mehra argues that the biased estimate should be used for this white noise test [33].

$$\hat{\mathbf{C}}_k = \frac{1}{N}\sum_{i=k}^{N} \mathbf{r}_i \mathbf{r}_{i-k} \tag{2.24}$$

From this point, Mehra [33] shows that the test for optimality uses the normalized autocorrelation coefficients of the autocorrelation sequence, ρ_k. The normalized autocorrelation matrix can be estimated from Equation 2.25.

$$\left[\hat{\rho}_k\right]_{ij} = \frac{\left[\hat{\mathbf{C}}_k\right]_{ij}}{\left\{\left[\hat{\mathbf{C}}_k\right]_{ii}\left[\hat{\mathbf{C}}_k\right]_{jj}\right\}^{1/2}} \tag{2.25}$$

The diagonals of ρ_k are compared with the confidence limits, and if the required percentage of the diagonal values are within this region, then the innovation sequence is white noise.

This test for optimality was implemented on Matlab™ for the Macintosh. The command *xcorr* calculates the normalized autocorrelation sequence [30, 42]. A comparison test is then run to determine the percentage of the sequence outside the 95% confidence limits.

2.4 Limitations of Kalman Filter

Although the Kalman filter is a very efficient estimator, it has certain limitations that arise when the state vector cannot be fully described. The Kalman filter works with a given state vector; if this description of the state vector is not adequate, then errors will result. These errors will affect the state estimates, because the filter cannot improve the state model on its own. For example, a trend error may exist in the data, but the Kalman filter cannot compensate for this term.

Also, the Kalman filter operates with fixed parameters. For example, the measurement noise covariance **R** is constant throughout data analysis. The Kalman filter then makes the best possible state estimates based on this **R**; if this value of **R** is wrong, then the Kalman filter estimates may not be optimal. These limitations of the Kalman filter are the driving forces in developing adaptive filters that are capable of compensating for unknown errors.

The following chapters will explore different methods of exploiting the information contained in the innovation sequence. The three methods used are maximum likelihood techniques, to enhance the estimates of the system parameters; correlation functions, to improve the state model; and power spectral density, to isolate periodic noise signals. Each of these methods seeks different information from the innovation sequence, and all three methods can be used simultaneously, as will be shown later.

Chapter 3

Maximum Likelihood Estimator

3.1 Motivation for Maximum Likelihood Estimator

Often, there is inadequate information about certain parameters in the Kalman filter matrices Φ, **B**, **G**, **Q**, and **R**. For example, **R**, the measurement noise covariance matrix, may have uncertainty in its values. In order to obtain the optimal state estimates $\hat{\mathbf{x}}$, an accurate description of these parameter matrices is necessary. One way to do this is to use *a priori* statistical information in order to describe the parameter behavior for a particular system. However, it is usually not feasible to develop a probability density function model for the unknown variables, because the *a priori* estimates require previous information about the system. Therefore, another technique must be developed to determine these unknown parameters without previous data. The ideal approach would be able to adjust the unknown values based on the present data, and it would be able to do it with an efficient online method.

3.2 Maximum Likelihood Estimator

The maximum likelihood technique has been chosen for parameter estimation. This method will find an efficient estimate, if such an estimate exists. The efficient estimate is the unbiased (or at least negligibly biased) estimate with the lowest finite covariance of all unbiased estimates. Also, if an efficient estimate exists, the maximum likelihood estimator will have a unique solution that is equal to the efficient estimate. The maximum likelihood estimator is also consistent; it converges to the true value of the parameters as the number of sample elements grows without bound. This solution is both asymptotically Gaussian and asymptotically efficient. Also, given a single sufficient statistic for the estimated variable, the maximum likelihood estimate will be sufficient, and it will also be at least asymptotically efficient and unbiased [31].

To develop an online maximum likelihood estimator, an assumption is made that the parameters are *essentially constant over a given interval of N sample periods*, where N is arbitrary. This assumption can be implemented in one of two ways. At every sample time t_i, the estimates can be based on the previous N measurements (t_1, t_2, t_3, ...), as shown in Figure 3.1a. Or, at every N points, parameter estimates can be based on the previous N data points (t_N, t_{2N}, t_{3N}, ...), as shown in Figure 3.1b. However, in order to implement either maximum likelihood estimator, a likelihood function for the parameters must be chosen.

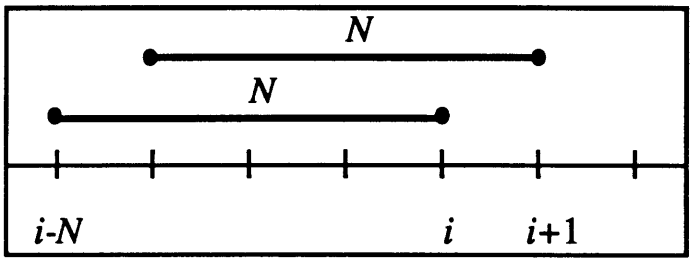

Figure 3.1a. Estimate Intervals at Every Sample Time

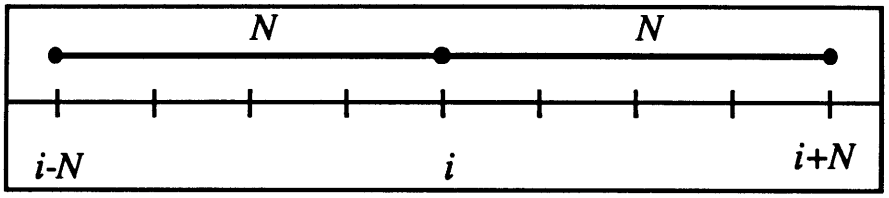

Figure 3.1b. Estimate Intervals at Every *N* Sample Times

3.3 Maximum Likelihood Estimator Theory

For a given likelihood function $L[\theta(t_i),\pounds_i]$, where $\theta(t_i)$ is the p length vector of variables (unknown parameters) and $\pounds_i$ is the set of realized values of the measurements to be used as data, the objective of the maximum likelihood estimator is to find a value of $\theta^*(t_i)$ that maximizes $L[\theta(t_i),\pounds_i]$ as a function of $\theta(t_i)$. This value can be found by solving Equation 3.1:

$$\left.\frac{\partial L[\theta(t_i),\pounds_i]}{\partial\theta(t_i)}\right|_{\theta(t_i)=\theta^*(t_i)} = \mathbf{0}^{\mathrm{T}} \tag{3.1}$$

Peter S. Maybeck, in Chapter 10 of *Stochastic Models, Estimation, and Control Volume 2*, argues that the preferred likelihood function is [31]

$$L = \ln f_{\mathbf{x}(t_i),\mathbf{z}(t_i)|\mathbf{a}}(\xi,\pounds_i|\alpha) \tag{3.2}$$

which can also be expressed as

$$L = \ln f_{\mathbf{x}(t_i),\mathbf{Z}_N(t_i)|\mathbf{Z}(t_{i-N}),\mathbf{a}}(\xi, £_{i,i-N+1}|£_{i-N}, \alpha) \tag{3.2a}$$

where,

$\mathbf{x}(t_i)$ = state vector at time t_i

$\mathbf{a}$ = the parameter vector

$\mathbf{Z}(t_i)$ = measurement history $\mathbf{z}(t_1), \mathbf{z}(t_2), \mathbf{z}(t_3), \ldots, \mathbf{z}(t_i)$

$\mathbf{Z}_N(t_i)$ = most recent N measurements $\mathbf{z}(t_{i-N+1}), \mathbf{z}(t_{i-N+2}), \mathbf{z}(t_{i-N+3}), \ldots, \mathbf{z}(t_i)$

$\mathbf{Z}(t_{i-N})$ = measurement history $\mathbf{z}(t_1), \mathbf{z}(t_2), \mathbf{z}(t_3), \ldots, \mathbf{z}(t_{i-N})$

f = probability density function

This likelihood function exploits all *a priori* information and yields an effective and computationally feasible estimator. Equation 3.2a is similar to Equation 3.2, but it provides a fixed-length memory parameter estimator.

Applying Bayes' rule

$$f_{A_i|B_j} = \frac{f_{B_j|A_i} f_{A_i}}{f_{B_j}} \tag{3.3}$$

to Equation 3.2 yields the following relation of probability densities:

$$\begin{aligned} f_{\mathbf{x}(t_i),\mathbf{Z}(t_i)|\mathbf{a}} &= f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i),\mathbf{a}} f_{\mathbf{Z}(t_i)|\mathbf{a}} \\ &= f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i),\mathbf{a}} f_{\mathbf{z}(t_i)|\mathbf{Z}(t_{i-1}),\mathbf{a}} f_{\mathbf{Z}(t_{i-1})|\mathbf{a}} \\ &\vdots \\ &= f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i),\mathbf{a}} \prod_{j=1}^{i} f_{\mathbf{z}(t_j)|\mathbf{Z}(t_{j-1}),\mathbf{a}} \end{aligned} \tag{3.4}$$

Each of these densities can then be written as

$$\begin{aligned} f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i),\mathbf{a}}(\xi|£_i, \alpha) &= \frac{1}{(2\pi)^{n/2}|\mathbf{P}(t_i^+)|^{1/2}} \exp\{\cdot\} \\ \{\cdot\} &= \left\{-\tfrac{1}{2}\left[\xi - \hat{\mathbf{x}}(t_i^+)\right]^{\mathrm{T}} \mathbf{P}(t_i^+)^{-1}\left[\xi - \hat{\mathbf{x}}(t_i^+)\right]\right\} \end{aligned} \tag{3.5}$$

where $\hat{\mathbf{x}}(t_i^+)$ and $\mathbf{P}(t_i^+)$ are implicitly dependent upon the parameters (i.e. α) upon which the density is conditioned, and n is the length of $\hat{\mathbf{x}}$. Also, the probability densities in the product of Equation 3.4 are

$$f_{\mathbf{z}(t_j)|\mathbf{Z}(t_{j-1}),\mathbf{a}}\left(\zeta_j|£_{j-1},\alpha\right)=\frac{1}{(2\pi)^{m/2}\left|\mathbf{S}(t_j)\right|^{1/2}}\exp\{\cdot\}$$
$$\{\cdot\}=\left\{-\tfrac{1}{2}\left[\zeta_j-\mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\left[\zeta_j-\mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]\right\} \tag{3.6}$$

where $\mathbf{S}(t_i)$ is defined in Equation 2.1; $\hat{\mathbf{x}}(t_i^-)$ $\mathbf{P}(t_i^-)$, and $\mathbf{S}(t_i)$ are again implicitly dependent upon the parameters, and m is the length of the measurement vector $\mathbf{z}$.

If Equations 3.5 and 3.6 are substituted into Equation 3.4, the likelihood function L then becomes

$$\begin{aligned}&\ln f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i),\mathbf{a}}\left(\xi|£_i,\alpha\right)\\&=-\frac{n+im}{2}\ln(2\pi)-\frac{1}{2}\ln\left(\left|\mathbf{P}(t_i^+)\right|\right)-\frac{1}{2}\sum_{j=1}^{i}\ln\left(\left|\mathbf{S}(t_j)\right|\right)\\&-\frac{1}{2}\left[\xi-\hat{\mathbf{x}}(t_i^+)\right]^{\mathrm{T}}\mathbf{P}(t_i^+)^{-1}\left[\xi-\hat{\mathbf{x}}(t_i^+)\right]\\&-\frac{1}{2}\sum_{j=1}^{i}\left[\zeta_j-\mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\left[\zeta_j-\mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]\end{aligned} \tag{3.7}$$

where n and m are the lengths of the vectors $\mathbf{x}$ and $\mathbf{z}$, respectively.

Inserting Equation 3.2 into Equation 3.1 yields the following pair of simultaneous equations:

$$\frac{\partial}{\partial\xi}\left[\ln f_{\mathbf{x}(t_i),\mathbf{Z}(t_i)|\mathbf{a}}(\xi,£_i|\alpha)\right]\Bigg|_{\substack{\xi=\mathbf{x}^*(t_i)\\\alpha=\mathbf{a}^*(t_i)}}=\mathbf{0}^{\mathrm{T}} \tag{3.8}$$

$$\frac{\partial}{\partial\alpha}\left[\ln f_{\mathbf{x}(t_i),\mathbf{Z}(t_i)|\mathbf{a}}(\xi,£_i|\alpha)\right]\Bigg|_{\substack{\xi=\mathbf{x}^*(t_i)\\\alpha=\mathbf{a}^*(t_i)}}=\mathbf{0}^{\mathrm{T}} \tag{3.9}$$

Substituting Equation 3.7 into Equation 3.8 leads to

$$-\left[\xi-\hat{\mathbf{x}}(t_i^+)\right]^{\mathrm{T}}\mathbf{P}(t_i^+)^{-1}\Bigg|_{\substack{\xi=\mathbf{x}^*(t_i)\\\alpha=\mathbf{a}^*(t_i)}}=\mathbf{0}^{\mathrm{T}} \tag{3.10}$$

For which the solution is

$$\mathbf{x}^*(t_i) = \hat{\mathbf{x}}(t_i^+)\Big|_{\alpha=\hat{\alpha}^*(t_i)} \tag{3.11}$$

Equation 3.11 demonstrates that the maximum likelihood estimate of the state vector is given by the Kalman filter equations when the unknown parameters are replaced by the best estimates of these parameters at time t_i.

However, when Equation 3.7 is substituted into Equation 3.9, considerably more work is necessary to find a solution. The derivative of L with respect to the parameter α is given in Equation 3.12 [31].

$$\begin{aligned}
&-2\frac{\partial}{\partial\alpha_k}\left\{\ln f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i),\hat{\alpha}}(\xi,\pounds_i|\alpha)\right\} \\
&\quad = \mathrm{tr}\left\{\mathbf{P}(t_i^+)^{-1}\frac{\partial\mathbf{P}(t_i^+)}{\partial\alpha_k}\right\} \\
&\quad -2\frac{\partial\hat{\mathbf{x}}(t_i^+)^{\mathrm{T}}}{\partial\alpha_k}\mathbf{P}(t_i^+)^{-1}\left[\xi-\hat{\mathbf{x}}(t_i^+)\right] \\
&\quad -\left[\xi-\hat{\mathbf{x}}(t_i^+)\right]^{\mathrm{T}}\mathbf{P}(t_i^+)^{-1}\frac{\partial\mathbf{P}(t_i^+)}{\partial\alpha_k}\mathbf{P}(t_i^+)^{-1}\left[\xi-\hat{\mathbf{x}}(t_i^+)\right] \\
&\quad +\sum_{j=1}^{i}\mathrm{tr}\left\{\mathbf{S}(t_j)^{-1}\frac{\partial\mathbf{S}(t_j)}{\partial\alpha_k}\right\} \\
&\quad -2\sum_{j=1}^{i}\frac{\partial\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial\alpha_k}\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\left[\zeta_j-\mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right] \\
&\quad -\sum_{j=1}^{i}\left[\zeta_j-\mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\frac{\partial\mathbf{S}(t_j)}{\partial\alpha_k}\mathbf{S}(t_j)^{-1}\left[\zeta_j-\mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]
\end{aligned} \tag{3.12}$$

Some important definitions of matrix differentiation that were needed for this problem are given in Equations 3.13 and 3.14.

$$\frac{\partial\ln|\mathbf{X}|}{\partial\alpha_k} = \frac{\partial\ln|\mathbf{X}|}{\partial|\mathbf{X}|}\frac{\partial|\mathbf{X}|}{\partial\alpha_k} = \frac{1}{|\mathbf{X}|}\frac{\partial|\mathbf{X}|}{\partial\alpha_k} = \mathrm{tr}\left\{\mathbf{X}^{-1}\frac{\partial|\mathbf{X}|}{\partial\alpha_k}\right\} \tag{3.13}$$

$$\frac{\partial \mathbf{X}^{-1}}{\partial \alpha_k} = -\mathbf{X}^{-1} \frac{\partial \mathbf{X}}{\partial \alpha_k} \mathbf{X}^{-1} \tag{3.14}$$

where $\mathbf{X}$ is a square matrix and α_k is the kth component of α.

But, as shown in Equation 3.11, ξ is equal to $\hat{\mathbf{x}}(t_i^+)$, so that the term $\xi - \hat{\mathbf{x}}(t_i^+)$ is equal to $\mathbf{0}$. Also, for two vectors $\mathbf{f}$ and $\mathbf{g}$, $\mathbf{f}^T\mathbf{g} = \mathrm{tr}\{\mathbf{f}\mathbf{g}^T\} = \mathrm{tr}\{\mathbf{g}\mathbf{f}^T\}$. Making these substitutions into Equation 3.12, the derivative of the likelihood function becomes

$$\mathrm{tr}\left\{\mathbf{P}(t_i^+)^{-1} \frac{\partial \mathbf{P}(t_i^+)}{\partial \alpha_k}\right\} - 2\sum_{j=1}^{i} \frac{\partial \hat{\mathbf{x}}(t_j^-)^T}{\partial \alpha_k} \mathbf{H}(t_j)^T \mathbf{S}(t_j)^{-1}\left[\mathbf{z}_j - \mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]$$

$$+\sum_{j=1}^{i} \mathrm{tr}\left\{\left[\mathbf{S}(t_j)^{-1} - \mathbf{S}(t_j)^{-1}\left[\mathbf{z}_j - \mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]\left[\mathbf{z}_j - \mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-)\right]^T \mathbf{S}(t_j)^{-1}\right]\right.$$

$$\left.\times \frac{\partial \mathbf{S}(t_j)}{\partial \alpha_k}\right\}\Bigg|_{\alpha=\mathbf{a}^*(t_i)} = 0 \tag{3.15}$$

The optimal solution for α solves this equation. However, there is no closed form solution to this problem; an iterative procedure must be used to find the optimal parameter vector, $\mathbf{a}^*(t_i)$.

Using Equation 3.2a, a fixed-length memory parameter estimator can be developed with arguments similar to those shown in Equation 3.4. First, write the probability densities for the likelihood function,

$$\begin{aligned} f_{\mathbf{x}(t_i),\mathbf{Z}_N(t_i)|\mathbf{Z}(t_{i-N}),\mathbf{a}} &= f_{\mathbf{x}(t_i)|\mathbf{Z}_N(t_i),\mathbf{Z}(t_{i-N}),\mathbf{a}} f_{\mathbf{Z}_N(t_i)|\mathbf{Z}(t_{i-N}),\mathbf{a}} \\ &= f_{\mathbf{x}(t_i)|\mathbf{Z}_N(t_i),\mathbf{Z}(t_{i-N}),\mathbf{a}} \frac{f_{\mathbf{Z}_N(t_i),\mathbf{Z}(t_{i-N})|\mathbf{a}}}{f_{\mathbf{Z}(t_{i-N})|\mathbf{a}}} \\ &= f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i),\mathbf{a}} \frac{f_{\mathbf{Z}(t_i)|\mathbf{a}}}{f_{\mathbf{Z}(t_{i-N})|\mathbf{a}}} \\ &= f_{\mathbf{x}(t_i)|\mathbf{Z}(t_i),\mathbf{a}} \prod_{j=i-N+1}^{i} f_{\mathbf{z}(t_j)|\mathbf{Z}(t_{j-1}),\mathbf{a}} \end{aligned} \tag{3.16}$$

In Equation 3.4, the lower limit on the product is 1. In Equation 3.16, the lower limit is i-N+1. Once again, the estimator yields the same state vector estimate as shown in Equation 3.11. However, in this case, the parameter equation is

$$\operatorname{tr}\left\{\mathbf{P}(t_i^+)^{-1}\frac{\partial \mathbf{P}(t_i^+)}{\partial \alpha_k}\right\} - 2\sum_{j=i-N+1}^{i}\frac{\partial \hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial \alpha_k}\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\mathbf{r}_j$$
$$+\sum_{j=i-N+1}^{i}\operatorname{tr}\left\{\left[\mathbf{S}(t_j)^{-1}-\mathbf{S}(t_j)^{-1}\mathbf{r}_j\mathbf{r}_j^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\right]\frac{\partial \mathbf{S}(t_j)}{\partial \alpha_k}\right\}\Bigg|_{\alpha=\alpha^*(t_i)} = 0 \tag{3.17}$$

where **r** is the innovation sequence defined in Equation 2.11 and **S** is the covariance of **r**.

3.4 Derivation of the Maximum Likelihood Kalman Filter

Using the ideas presented in the Section 3.3 and Chapter 2, a Kalman filter with maximum likelihood estimation of the parameters can now be developed [31]. The recursive formula for Equation 3.1 can be written using the Newton–Raphson method,

$$\hat{\theta}^*(t_i) = \hat{\theta}_*(t_i) - \left[\frac{\partial^2 L\left[\hat{\theta}_*(t_i),\mathbf{Z}_i\right]}{\partial \theta^2}\right]^{-1}\frac{\partial L\left[\hat{\theta}_*(t_i),\mathbf{Z}_i\right]^{\mathrm{T}}}{\partial \theta} \tag{3.18}$$

where $\hat{\theta}^*(t_i)$ is the new estimate of the parameters at t_i and $\hat{\theta}_*(t_i)$ is the previous estimate of the parameters. The first derivative matrix in Equation 3.18 can also be written as

$$\frac{\partial L\left[\hat{\theta}_*(t_i),\mathbf{Z}_i\right]^{\mathrm{T}}}{\partial \theta} = \frac{\partial L[\theta,\mathbf{Z}_i]^{\mathrm{T}}}{\partial \theta}\Bigg|_{\theta=\hat{\theta}_*(t_i)} \tag{3.19}$$

and is called the *gradient* or *score vector*. The second derivative matrix is named the Hessian matrix, and it must be of full rank in order to be inverted. It is computationally burdensome to use the Hessian, so the following approximation to the Hessian is used

$$\frac{\partial^2 L\left[\hat{\theta}_*(t_i),\mathbf{Z}_i\right]}{\partial \theta^2} \cong -\mathbf{J}\left[t_i,\hat{\theta}_*(t_i)\right] \tag{3.20}$$

In Equation 3.20, **J** is defined as

$$\mathbf{J}\left[t_i,\hat{\theta}_*(t_i)\right] \equiv E\left\{\frac{\partial L[\theta,\mathbf{Z}(t_i)]^{\mathrm{T}}}{\partial \theta}\frac{\partial L[\theta,\mathbf{Z}(t_i)]}{\partial \theta}\Big|\theta=\hat{\theta}_*(t_i)\right\} \tag{3.21}$$

and it is called the *conditional information matrix* [31]. It can be shown that

$$-\mathbf{J}\left[t_i, \hat{\theta}_*(t_i)\right] = E\left\{\frac{\partial^2 L[\theta, \mathbf{Z}(t_i)]}{\partial \theta^2}\middle|\theta = \hat{\theta}_*(t_i)\right\} \tag{3.22}$$

This approximation states that the second derivative matrix, the Hessian, for a particular realization of $\mathbf{Z}_i$, can be adequately represented by its ensemble average over all possible measurement time histories. With this approximation, only first order information about the likelihood function is required, and Equation 3.18 can then be expressed as

$$\hat{\theta}^*(t_i) = \hat{\theta}_*(t_i) + \mathbf{J}\left[t_i, \hat{\theta}_*(t_i)\right]^{-1} \frac{\partial L\left[\hat{\theta}_*(t_i), \mathbf{Z}_i\right]^{\mathrm{T}}}{\partial \theta} \tag{3.23}$$

The scoring approximation, Maybeck argues, is a superior algorithm to both the Newton–Raphson method, which uses the Hessian matrix, and the conventional gradient algorithm [31]. With the conditional information matrix approximation, the error is of the order $1/N$ for large N. Also for large N, the rate of convergence approaches that of the Newton–Raphson method. Although scoring converges less rapidly near the solution than Newton–Raphson, it does converge over a larger region. The computational advantages over Newton–Raphson are also quite considerable. Scoring converges more rapidly than a gradient algorithm, but it requires slightly more computation to do so.

There are some problems with the scoring method, however. The initial $\mathbf{J}$ entries may be small, so that its inverse has large entries; this problem can be corrected by using a precomputed $\mathbf{J}$ value based on simulations or previous analysis. Also, $\mathbf{J}$ and $\mathbf{J}^{-1}$ must be constantly recomputed. However, after an initial transient, these values may be computed infrequently since $\mathbf{J}$ does not vary much after this transient has vanished. In this thesis, no problems with inverting $\mathbf{J}$ were encountered.

To generate parameter estimates, the score vector, $\{\partial L[\hat{\mathbf{x}}_*(t_i), \hat{\mathbf{a}}_*(t_i), \mathbf{Z}_i]/\partial \mathbf{a}\}^{\mathrm{T}}$, and the conditional information matrix $\mathbf{J}[t_i, \hat{\mathbf{x}}_*(t_i), \hat{\mathbf{a}}_*(t_i)]$ must be computed. The score is equal to Equation 3.17 times $-\frac{1}{2}$, and the equation is evaluated with the actual estimate $\hat{\mathbf{a}}_*(t_i)$, not with the maximum likelihood estimate $\mathbf{a}^*(t_i)$. The score is composed of the most recent N measurements. The score can be broken down into the sum of the N most recent *single measurement scores*, $\mathbf{s}_k^1[\mathbf{Z}_j, \hat{\mathbf{a}}_*(t_i)]$, and a final term $\gamma[\mathbf{Z}_i, \hat{\mathbf{a}}_*(t_i)]$:

$$\frac{\partial L}{\partial a_k}\left[\hat{\mathbf{x}}_*(t_i), \hat{\mathbf{a}}_*(t_i), \mathbf{Z}_i\right] = \gamma_k\left[\mathbf{Z}_i, \hat{\mathbf{a}}_*(t_i)\right] + \sum_{j=i-N+1}^{i} s_k{}^1\left[\mathbf{Z}_j, \hat{\mathbf{a}}_*(t_i)\right] \tag{3.24}$$

where

$$s_k^1[\mathbf{Z}_j,\mathbf{a}] = \frac{\partial \hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_k}\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\mathbf{r}_j - \frac{1}{2}\mathrm{tr}\left\{\left[\mathbf{S}(t_j)^{-1} - \mathbf{S}(t_j)^{-1}\mathbf{r}_j\mathbf{r}_j^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\right]\frac{\partial \mathbf{S}(t_j)}{\partial a_k}\right\} \tag{3.25a}$$

and

$$\gamma_k[\mathbf{Z}_i,\mathbf{a}] = -\frac{1}{2}\mathrm{tr}\left\{\mathbf{P}(t_i^+)^{-1}\frac{\partial \mathbf{P}(t_i^+)}{\partial a_k}\right\} \tag{3.25b}$$

The conditional information matrix can similarly be expressed in terms of a sum over the N most recent terms and a final term, if the parameter value is allowed to assume a true, but unknown value, $\mathbf{a}_t$:

$$J_{kl}[t_i,\hat{\mathbf{x}}_*(t_i^+),\mathbf{a}_t] = E\left\{\gamma_k[\mathbf{Z}(t_i),\mathbf{a}]\gamma_l[\mathbf{Z}(t_i),\mathbf{a}]\middle|\mathbf{a}=\mathbf{a}_t\right\} + \sum_{j=i-N+1}^{i} E\left\{s_k^1[\mathbf{Z}(t_j),\mathbf{a}]s_l^1[\mathbf{Z}(t_j),\mathbf{a}]\middle|\mathbf{a}=\mathbf{a}_t\right\} \tag{3.26}$$

where

$$E\{s_k^1[\mathbf{Z}(t_j),\mathbf{a}]s_l^1[\mathbf{Z}(t_j),\mathbf{a}]|\mathbf{a}=\mathbf{a}_t\} = \frac{1}{2}\mathrm{tr}\left[\mathbf{S}(t_j)^{-1}\frac{\partial \mathbf{S}(t_j)}{\partial a_k}\mathbf{S}(t_j)^{-1}\frac{\partial \mathbf{S}(t_j)}{\partial a_l} + 2\mathbf{S}(t_j)^{-1}\mathbf{H}(t_j)E\left\{\frac{\partial \hat{\mathbf{x}}(t_j^-)}{\partial a_k}\frac{\partial \hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_l}\middle|\mathbf{a}=\mathbf{a}_t\right\}\mathbf{H}(t_j)^{\mathrm{T}}\right] \tag{3.27a}$$

and,

$$E\{\gamma_k[\mathbf{Z}(t_i),\mathbf{a}]\gamma_l[\mathbf{Z}(t_i),\mathbf{a}]|\mathbf{a}=\mathbf{a}_t\} = \frac{1}{2}\mathrm{tr}\left[\mathbf{P}(t_i^+)^{-1}\frac{\partial \mathbf{P}(t_i^+)}{\partial a_k}\mathbf{P}(t_i^+)^{-1}\frac{\partial \mathbf{P}(t_i^+)}{\partial a_l} + 2\mathbf{P}(t_i^+)^{-1}E\left\{\frac{\partial \hat{\mathbf{x}}(t_i^+)}{\partial a_k}\frac{\partial \hat{\mathbf{x}}(t_i^+)^{\mathrm{T}}}{\partial a_l}\middle|\mathbf{a}=\mathbf{a}_t\right\}\right] \tag{3.27b}$$

Equations 3.1 through 3.17 were derived without approximations. For Equations 3.18 through 3.27, the approximation has been made that the conditional information matrix **J** is equal to the Hessian; an approximation will also be made that these equations hold for values other than $\mathbf{a}_t$.

3.5 Equations for Maximum Likelihood Kalman Filter

The maximum likelihood parameter estimator assumes that the parameters are constant over a period of N sample times [31]. At a sample time t_i, the quantities at time t_{i-N} are unchangeable because nothing done in the present can affect a past estimate. Therefore, the initial conditions of the N-step recursion for parameter estimation are:

$$\hat{\mathbf{x}}(t_{i-N}^{+}) = \text{previously computed} \tag{3.28a}$$

$$\mathbf{P}(t_{i-N}^{+}) = \text{previously computed} \tag{3.28b}$$

$$\frac{\partial \hat{\mathbf{x}}(t_{i-N}^{+})}{\partial a_k} = \mathbf{0} \qquad \text{(for all } k\text{)} \tag{3.28c}$$

$$\frac{\partial \mathbf{P}(t_{i-N}^{+})}{\partial a_k} = \mathbf{0} \qquad \text{(for all } k\text{)} \tag{3.28d}$$

$$E\left\{ \frac{\partial \hat{\mathbf{x}}(t_{i-N}^{+})}{\partial a_k} \frac{\partial \hat{\mathbf{x}}(t_{i-N}^{+})^{\mathrm{T}}}{\partial a_l} \Big| \mathbf{a} = \hat{\mathbf{a}}_*(t_i) \right\} = \mathbf{0} \quad \text{(for all } k \text{ and } l\text{)} \tag{3.28e}$$

To solve for the parameter estimates, the score vector s^1 and conditional information matrix $\mathbf{J}$ must be calculated. Both of these matrices depend on the derivatives of both the state vector and its covariance with respect to the parameter vector. In turn, these derivatives depend upon how the matrices of the system vary with the parameters. In order to determine this dependence, the derivatives of the Kalman filter equations are taken with respect to the parameter vector $\mathbf{a}$. The Kalman filter is implemented as discussed in Chapter 2, but between each of the updates of the state vector, the derivatives of these vectors are also calculated. The procedure, then, for maximum likelihood estimation in a Kalman filter follows the flowchart in Figure 3.2. In this flowchart, the estimates of the parameters are made at every N points based on the previous N points, as shown in Figure 3.1b.

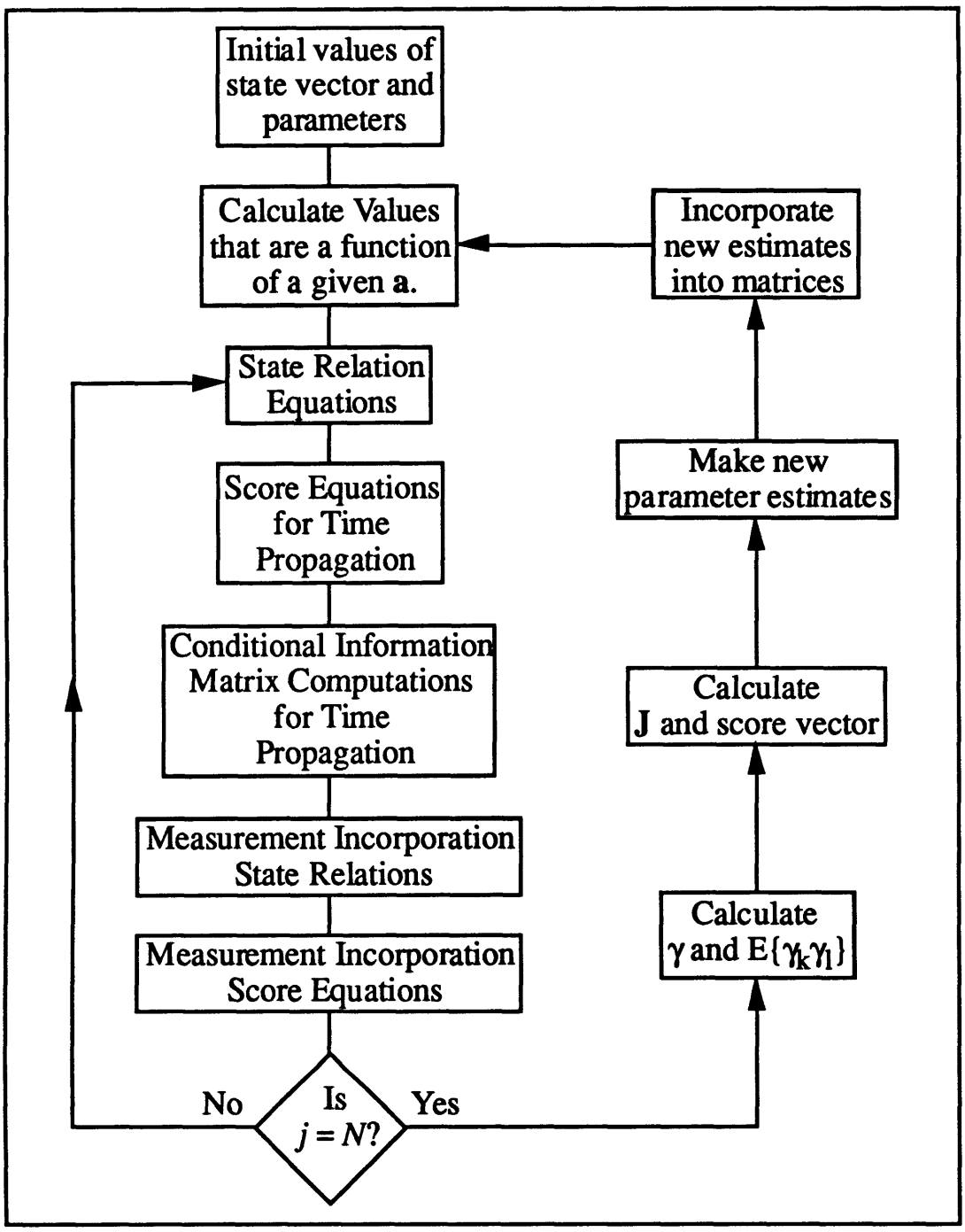

Figure 3.2. Flowchart for Maximum Likelihood Estimation

Before the filter can be implemented, certain values must be given. For example, the initial values of **x** and **P**, as well as the general form of the dependence of Φ, **B**, **Q**, **G** and **R** on the parameter vector **a**, such as $\partial\Phi/\partial a_k$, $\partial\mathbf{R}/\partial a_k$, etc., must be given. These matrices must be reevaluated for every new estimate of the parameters, $\hat{\mathbf{a}}^*(t_i)$. Once these functions of **a** are given, the following equations govern the filter.

First, the *state relation equations* from Chapter 2 are processed using the most recent parameter estimate $\hat{\mathbf{a}}^*(t_i)$ to evaluate Φ, **B**, **Q**, **G** and **R** as required:

$$\hat{\mathbf{x}}(t_j^-) = \Phi(t_j, t_{j-1})\hat{\mathbf{x}}(t_{j-1}^+) + \mathbf{B}(t_{j-1})\mathbf{u}(t_{j-1}) \tag{2.7}$$

$$\mathbf{P}(t_j^-) = \Phi(t_j, t_{j-1})\mathbf{P}(t_{j-1}^+)\Phi(t_j, t_{j-1})^{\mathrm{T}} + \mathbf{G}(t_{j-1})\mathbf{Q}(t_{j-1})\mathbf{G}(t_{j-1})^{\mathrm{T}} \tag{2.8}$$

$$\mathbf{K}(t_j) = \left[\mathbf{P}(t_j^-)\mathbf{H}(t_j)^{\mathrm{T}}\right]\mathbf{S}(t_j)^{-1} \tag{2.9}$$

$$\mathbf{S}(t_j) = \mathbf{H}(t_j)\mathbf{P}(t_j^-)\mathbf{H}(t_j)^{\mathrm{T}} + \mathbf{R}(t_j) \tag{2.10}$$

The *score equations* for the time propagation are given by the p set of "sensitivity system" equations. These equations, as well as all following equations, must be evaluated for k = 1, 2, ..., p, when the subscript k appears, (p is the length of the parameter vector **a**). First, the state vector estimate derivative is updated:

$$\frac{\partial \hat{\mathbf{x}}(t_j^-)}{\partial a_k} = \Phi(t_j, t_{j-1})\frac{\partial \hat{\mathbf{x}}(t_{j-1}^+)}{\partial a_k} + \frac{\partial \Phi(t_j, t_{j-1})}{\partial a_k}\hat{\mathbf{x}}(t_{j-1}^+) + \frac{\partial \mathbf{B}(t_{j-1})}{\partial a_k}\mathbf{u}(t_{j-1}) \tag{3.29}$$

Next, the derivative of the covariance matrix **P** is calculated:

$$\begin{aligned}
\frac{\partial \mathbf{P}(t_j^-)}{\partial a_k} &= \Phi(t_j, t_{j-1})\frac{\partial \mathbf{P}(t_{j-1}^+)}{\partial a_k}\Phi(t_j, t_{j-1})^{\mathrm{T}} \\
&+ \frac{\partial \Phi(t_j, t_{j-1})}{\partial a_k}\mathbf{P}(t_{j-1}^+)\Phi(t_j, t_{j-1})^{\mathrm{T}} \\
&+ \Phi(t_j, t_{j-1})\mathbf{P}(t_{j-1}^+)\frac{\partial \Phi(t_j, t_{j-1})^{\mathrm{T}}}{\partial a_k} \\
&+ \mathbf{G}(t_{j-1})\frac{\partial \mathbf{Q}(t_{j-1})}{\partial a_k}\mathbf{G}(t_{j-1})^{\mathrm{T}} + \frac{\partial \mathbf{G}(t_{j-1})}{\partial a_k}\mathbf{Q}(t_{j-1})\mathbf{G}(t_{j-1})^{\mathrm{T}} \\
&+ \mathbf{G}(t_{j-1})\mathbf{Q}(t_{j-1})\frac{\partial \mathbf{G}(t_{j-1})^{\mathrm{T}}}{\partial a_k}
\end{aligned} \tag{3.30}$$

For the innovation sequence covariance matrix, **S**, the derivative update equation is:

$$\frac{\partial \mathbf{S}(t_j)}{\partial a_k} = \mathbf{H}(t_j)\frac{\partial \mathbf{P}(t_j^-)}{\partial a_k}\mathbf{H}(t_j)^{\mathrm{T}} + \frac{\partial \mathbf{R}(t_j)}{\partial a_k} \tag{3.31}$$

Also needed are the *conditional information matrix computations for forward time propagation* (with time arguments t_{j-1} removed and $\mathbf{a} = \hat{\mathbf{a}}_*(t_i)$ reduced to $\hat{\mathbf{a}}_*$). These equations are used to solve $E\{s_k s_l\}$, and they are given in Equation 3.32 through 3.37. The expectation needed to calculate $E\{s_k s_l\}$ is shown in Equation 3.32.

$$
\begin{aligned}
&E\left\{\frac{\partial\hat{\mathbf{x}}(t_j^-)}{\partial a_k}\frac{\partial\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_l}\Big|\mathbf{a}=\hat{\mathbf{a}}_*(t_i)\right\}\\
&=\Phi E\left\{\frac{\partial\hat{\mathbf{x}}^+}{\partial a_k}\frac{\partial\hat{\mathbf{x}}^{+\mathrm{T}}}{\partial a_l}\Big|\hat{\mathbf{a}}_*\right\}\Phi^{\mathrm{T}}+\frac{\partial\Phi}{\partial a_k}E\left\{\hat{\mathbf{x}}^+\hat{\mathbf{x}}^{+\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\frac{\partial\Phi^{\mathrm{T}}}{\partial a_l}\\
&+\Phi E\left\{\frac{\partial\hat{\mathbf{x}}^+}{\partial a_k}\hat{\mathbf{x}}^{+\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\frac{\partial\Phi^{\mathrm{T}}}{\partial a_l}+\frac{\partial\Phi}{\partial a_k}E\left\{\hat{\mathbf{x}}^+\frac{\partial\hat{\mathbf{x}}^{+\mathrm{T}}}{\partial a_l}\Big|\hat{\mathbf{a}}_*\right\}\Phi^{\mathrm{T}}\\
&+\frac{\partial\mathbf{B}}{\partial a_k}E\left\{\mathbf{u}\mathbf{u}^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\frac{\partial\mathbf{B}^{\mathrm{T}}}{\partial a_l}\\
&+\frac{\partial\Phi}{\partial a_k}E\left\{\hat{\mathbf{x}}^+\mathbf{u}^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\frac{\partial\mathbf{B}^{\mathrm{T}}}{\partial a_l}+\frac{\partial\mathbf{B}}{\partial a_k}E\left\{\mathbf{u}\hat{\mathbf{x}}^{+\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\frac{\partial\Phi^{\mathrm{T}}}{\partial a_l}\\
&+\Phi E\left\{\frac{\partial\hat{\mathbf{x}}^+}{\partial a_k}\mathbf{u}^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\frac{\partial\mathbf{B}^{\mathrm{T}}}{\partial a_l}+\frac{\partial\mathbf{B}}{\partial a_k}E\left\{\mathbf{u}\frac{\partial\hat{\mathbf{x}}^{+\mathrm{T}}}{\partial a_k}\Big|\hat{\mathbf{a}}_*\right\}\frac{\partial\Phi^{\mathrm{T}}}{\partial a_l}
\end{aligned}
\tag{3.32}
$$

Equation 3.32 requires the following expectations in order to be solved:

$$
\begin{aligned}
E\left\{\hat{\mathbf{x}}(t_j^-)\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}\Big|\mathbf{a}=\hat{\mathbf{a}}_*(t_i)\right\}&=\Phi E\left\{\hat{\mathbf{x}}^+\hat{\mathbf{x}}^{+\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\Phi^{\mathrm{T}}+\mathbf{B}E\left\{\mathbf{u}\mathbf{u}^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\mathbf{B}^{\mathrm{T}}\\
&\quad+\Phi E\left\{\hat{\mathbf{x}}^+\mathbf{u}^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\mathbf{B}^{\mathrm{T}}+\mathbf{B}E\left\{\mathbf{u}\hat{\mathbf{x}}^{+\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\Phi^{\mathrm{T}}
\end{aligned}
\tag{3.33}
$$

$$
\begin{aligned}
&E\left\{\frac{\partial\hat{\mathbf{x}}(t_j^-)}{\partial a_k}\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}\Big|\mathbf{a}=\hat{\mathbf{a}}_*(t_i)\right\}=\\
&\Phi E\left\{\frac{\partial\hat{\mathbf{x}}^+}{\partial a_k}\hat{\mathbf{x}}^{+\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\Phi^{\mathrm{T}}+\frac{\partial\Phi}{\partial a_k}E\left\{\hat{\mathbf{x}}^+\hat{\mathbf{x}}^{+\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\Phi^{\mathrm{T}}\\
&+\Phi E\left\{\frac{\partial\hat{\mathbf{x}}^+}{\partial a_k}\mathbf{u}^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\mathbf{B}^{\mathrm{T}}+\frac{\partial\mathbf{B}}{\partial a_k}E\left\{\mathbf{u}\mathbf{u}^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\mathbf{B}^{\mathrm{T}}\\
&+\frac{\partial\Phi}{\partial a_k}E\left\{\hat{\mathbf{x}}^+\mathbf{u}^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\mathbf{B}^{\mathrm{T}}+\frac{\partial\mathbf{B}}{\partial a_k}E\left\{\mathbf{u}\hat{\mathbf{x}}^{+\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\}\Phi^{\mathrm{T}}
\end{aligned}
\tag{3.34}
$$

If $\mathbf{u}$ is precomputed, then $E\{\mathbf{u}()^{\mathrm{T}} \mid \hat{\mathbf{a}}_*\}$ can be written as $\mathbf{u}E\{()^{\mathrm{T}} \mid \hat{\mathbf{a}}_*\}$ With this assumption, Equation 3.35 and Equation 3.36 can be evaluated and used to solve Equation 3.32 through Equation 3.34.

$$E\left\{\hat{\mathbf{x}}(t_j^+)\middle|\hat{\mathbf{a}}_*\right\} = \Phi E\left\{\hat{\mathbf{x}}^+\middle|\hat{\mathbf{a}}_*\right\} + \mathbf{B}\mathbf{u} \tag{3.35}$$

And finally,

$$E\left\{\frac{\partial\hat{\mathbf{x}}(t_j^+)}{\partial a_k}\middle|\hat{\mathbf{a}}_*\right\} = \mathbf{D}(t_j)\left[\Phi E\left\{\frac{\partial\hat{\mathbf{x}}^+}{\partial a_k}\middle|\hat{\mathbf{a}}_*\right\} + \frac{\partial\Phi}{\partial a_k}E\left\{\hat{\mathbf{x}}^+\middle|\hat{\mathbf{a}}_*\right\} + \frac{\partial\mathbf{B}}{\partial a_k}\mathbf{u}\right] \tag{3.36}$$

where $\mathbf{D}(t_j)$ is defined in Equation 3.38.

After these calculations are complete, $E\{s_k s_l\}$ can be solved with Equation 3.37.

$$\begin{aligned} &E\{s_k{}^1[\mathbf{Z}(t_j),\mathbf{a}]s_l{}^1[\mathbf{Z}(t_j),\mathbf{a}]\big|\mathbf{a}=\hat{\mathbf{a}}_*(t_i)\} \\ &\quad = \frac{1}{2}\mathrm{tr}\left[\mathbf{S}(t_j)^{-1}\frac{\partial\mathbf{S}(t_j)}{\partial a_k}\mathbf{S}(t_j)^{-1}\frac{\partial\mathbf{S}(t_j)}{\partial a_l}\right. \\ &\quad \left. +2\mathbf{S}(t_j)^{-1}\mathbf{H}(t_j)E\left\{\frac{\partial\hat{\mathbf{x}}(t_j^-)}{\partial a_k}\frac{\partial\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_l}\middle|\mathbf{a}=\hat{\mathbf{a}}_*(t_i)\right\}\mathbf{H}(t_j)^{\mathrm{T}}\right] \end{aligned} \tag{3.37}$$

Next, the *measurement incorporation state relations* at time t_j are:

$$\mathbf{r}_j = \mathbf{z}_j - \mathbf{H}(t_j)\hat{\mathbf{x}}(t_j^-) \tag{2.11}$$

$$\mathbf{D}(t_j) = \mathbf{I} - \mathbf{K}(t_j)\mathbf{H}(t_j) \tag{3.38}$$

$$\hat{\mathbf{x}}(t_j^+) = \hat{\mathbf{x}}(t_j^-) + \mathbf{K}(t_j)\mathbf{r}_j \tag{2.12}$$

$$\mathbf{P}(t_j^+) = \left[\mathbf{I} - \mathbf{K}(t_j)\mathbf{H}(t_j)\right]\mathbf{P}(t_j^-) \tag{2.13}$$

The *measurement update score equations* are then be processed using these definitions:

$$\mathbf{n}_j = \mathbf{S}(t_j)^{-1}\mathbf{r}_j \tag{3.39}$$

$$\mathbf{C}(t_j) = \mathbf{S}(t_j)^{-1} - \mathbf{n}_j\mathbf{n}_j^{\mathrm{T}} \tag{3.40}$$

For $k = 1, 2, \ldots, p$, the *first measurement update score equation* is

$$s_k^1\left[\mathbf{Z}_j, \hat{\mathbf{a}}_*(t_i)\right] = \frac{\partial \hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_k}\left[\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{n}_j\right] - \frac{1}{2}\mathrm{tr}\left\{\mathbf{C}(t_j)\frac{\partial \mathbf{S}(t_j)}{\partial a_k}\right\} \tag{3.41}$$

The second equation of the *measurement update score equations* can be developed by taking the derivative of Equation 2.10, as shown in Equation 3.42.

$$\frac{\partial \hat{\mathbf{x}}(t_j^+)}{\partial a_k} = \frac{\partial \hat{\mathbf{x}}(t_j^-)}{\partial a_k} + \frac{\partial \mathbf{K}(t_j)}{\partial a_k}\mathbf{r}_j + \mathbf{K}(t_j)\frac{\partial \mathbf{r}_j}{\partial a_k} \tag{3.42}$$

where,

$$\frac{\partial \mathbf{K}(t_j)}{\partial a_k} = \frac{\partial \mathbf{P}(t_j^-)}{\partial a_k}\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{S}(t_j)^{-1} + \mathbf{P}(t_j^-)\mathbf{H}(t_j)^{\mathrm{T}}\frac{\partial \mathbf{S}(t_j)^{-1}}{\partial a_k} \tag{3.43}$$

and,

$$\frac{\partial \mathbf{r}_j}{\partial a_k} = -\mathbf{H}(t_j)\frac{\partial \hat{\mathbf{x}}(t_j^-)}{\partial a_k} \tag{3.44}$$

Substituting Equations 3.43 and 3.44 into Equation 3.42 leads to the *second measurement update score equation*:

$$\frac{\partial \hat{\mathbf{x}}(t_j^+)}{\partial a_k} = \mathbf{D}(t_j)\left\{\frac{\partial \hat{\mathbf{x}}(t_j^-)}{\partial a_k} + \frac{\partial \mathbf{P}(t_j^-)}{\partial a_k}\left[\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{n}_j\right]\right\} \tag{3.45}$$

A similar exercise with the derivative of Equation 2.13 gives the *third measurement update score equation*:

$$\frac{\partial \mathbf{P}(t_j^+)}{\partial a_k} = \mathbf{D}(t_j)\frac{\partial \mathbf{P}(t_j^-)}{\partial a_k}\mathbf{D}(t_j)^{\mathrm{T}} + \mathbf{K}(t_j)\frac{\partial \mathbf{R}(t_j)}{\partial a_k}\mathbf{K}(t_j)^{\mathrm{T}} \tag{3.46}$$

The *measurement update conditional information matrix equations* are

$$\begin{aligned} &E\left\{\frac{\partial \hat{\mathbf{x}}(t_j^+)}{\partial a_k}\frac{\partial \hat{\mathbf{x}}(t_j^+)^{\mathrm{T}}}{\partial a_l}\middle|\mathbf{a} = \hat{\mathbf{a}}_*(t_i)\right\} \\ &= \mathbf{D}(t_j)E\left\{\frac{\partial \hat{\mathbf{x}}(t_j^-)}{\partial a_k}\frac{\partial \hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_l}\middle|\hat{\mathbf{a}}_*\right\}\mathbf{D}(t_j)^{\mathrm{T}} \\ &+\mathbf{D}(t_j)\frac{\partial \mathbf{P}(t_j^-)}{\partial a_k}\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\mathbf{H}(t_j)\frac{\partial \mathbf{P}(t_j^-)}{\partial a_k}\mathbf{D}(t_j)^{\mathrm{T}} \end{aligned} \tag{3.47}$$

As well as,

$$E\left\{\hat{\mathbf{x}}(t_j^+)\hat{\mathbf{x}}(t_j^+)^{\mathrm{T}}\Big|_{\mathbf{a}=\hat{\mathbf{a}}_*(t_i)}\right\} = E\left\{\hat{\mathbf{x}}(t_j^-)\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}\Big|_{\mathbf{a}=\hat{\mathbf{a}}_*(t_i)}\right\} + \mathbf{K}(t_j)\mathbf{S}(t_j)\mathbf{K}(t_j)^{\mathrm{T}} \quad (3.48)$$

And,

$$\begin{aligned} &E\left\{\frac{\partial\hat{\mathbf{x}}(t_j^+)}{\partial a_k}\hat{\mathbf{x}}(t_j^+)^{\mathrm{T}}\Big|\mathbf{a}=\hat{\mathbf{a}}_*(t_i)\right\} \\ &= \mathbf{D}(t_j)\left[E\left\{\frac{\partial\hat{\mathbf{x}}(t_j^-)}{\partial a_k}\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}\Big|\hat{\mathbf{a}}_*\right\} + \frac{\partial\mathbf{P}(t_j^-)}{\partial a_k}\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{K}(t_j)^{\mathrm{T}}\right] \end{aligned} \quad (3.49)$$

At the end of processing N points, the score vector and conditional information matrix are computed, and a new parameter estimate is made using Equation 3.50.

$$\hat{\mathbf{a}}^*(t_i) = \hat{\mathbf{a}}_*(t_i) + \mathbf{J}\left[t_i, \hat{\mathbf{x}}_*(t_i), \hat{\mathbf{a}}_*(t_i)\right]^{-1}\left\{\frac{\partial L\left[\hat{\mathbf{x}}_*(t_i), \hat{\mathbf{a}}_*(t_i), \mathbf{Z}_i\right]}{\partial \mathbf{a}}\right\}^{\mathrm{T}} \quad (3.50)$$

This procedure is then repeated until all of the data has been processed.

3.5.1 Approximations for Online Operation

A quick look at Equation 3.32 and Equation 3.47 shows that a tremendous amount of calculations are required to construct the conditional information matrix, **J**. Maybeck suggests that the approximation

$$E\left\{\frac{\partial\hat{\mathbf{x}}(t_j^-)}{\partial a_k}\frac{\partial\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_l}\Big|\mathbf{a}=\hat{\mathbf{a}}_*(t_i)\right\} \cong \frac{\partial\hat{\mathbf{x}}(t_j^-)}{\partial a_k}\frac{\partial\hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_l} \quad (3.51)$$

can be made, greatly simplifying the equations [31]. This approximation asserts that the expectation of $E\left\{\partial\hat{\mathbf{x}}/\partial a_k\,\partial\hat{\mathbf{x}}^{\mathrm{T}}/\partial a_l\right\}$ over all possible noise sequences can be adequately represented by the value it would take due to a particular sequence that generated the measurement data. Maybeck argues that this approximation reintroduces the dependence of the Hessian matrix on the second derivative into the scoring approximation [31].

With this approximation, Equations 3.32 through 3.37 are replaced by Equation 3.52 for the conditional information matrix components:

$$E\{s_k^{\,1}[\mathbf{Z}(t_j),\mathbf{a}]\,s_l^{\,1}[\mathbf{Z}(t_j),\mathbf{a}]\Big|_{\mathbf{a}=\hat{\mathbf{a}}_*(t_i)}\} \cong$$
$$\frac{1}{2}\mathrm{tr}\left[\mathbf{S}(t_j)^{-1}\frac{\partial \mathbf{S}(t_j)}{\partial a_k}\mathbf{S}(t_j)^{-1}\frac{\partial \mathbf{S}(t_j)}{\partial a_l}\right] \tag{3.52}$$
$$+\frac{\partial \hat{\mathbf{x}}(t_j^-)^{\mathrm{T}}}{\partial a_k}\mathbf{H}(t_j)^{\mathrm{T}}\mathbf{S}(t_j)^{-1}\mathbf{H}(t_j)\frac{\partial \hat{\mathbf{x}}(t_j^-)}{\partial a_l}$$

and Equation 3.27b is replaced by Equation 3.53

$$E\{\gamma_k[\mathbf{Z}(t_i),\mathbf{a}]\,\gamma_l[\mathbf{Z}(t_i),\mathbf{a}]\Big|_{\mathbf{a}=\hat{\mathbf{a}}_*(t_i)}\} \cong$$
$$\frac{1}{2}\mathrm{tr}\left[\mathbf{P}(t_i^+)^{-1}\frac{\partial \mathbf{P}(t_i^+)}{\partial a_k}\mathbf{P}(t_i^+)^{-1}\frac{\partial \mathbf{P}(t_i^+)}{\partial a_l}\right] \tag{3.53}$$
$$+\frac{\partial \hat{\mathbf{x}}(t_i^+)^{\mathrm{T}}}{\partial a_l}\mathbf{P}(t_i^+)^{-1}\frac{\partial \hat{\mathbf{x}}(t_i^+)}{\partial a_k}$$

Maybeck has shown that this approximation is valid [31]. The simplicity of both Equation 3.52 and Equation 3.53 demonstrate the effectiveness of this approximation. Now, the propagation and state relations that are necessary for the state and score vector computations can also be used to evaluate **J**.

The conditional information matrix **J** is the covariance matrix of the parameter estimates, **a** [31, 43]. Therefore, the maximum likelihood estimator provides the optimal state estimates and covariances, as well as the optimal parameter estimates and covariances.

3.6 Verification of Maximum Likelihood Estimator

In order to verify the Matlab™ program code, as well as the theory, for the maximum likelihood estimator, several runs were made with simulated data. In these test cases, the true values of the parameters were known so that the estimator behavior could be studied. The test cases include unknown measurement noise standard deviation, unknown process noise standard deviation, and unknown measurement and process noises. Test cases were not run for unknown values in Φ, **B**, or **G**; these parameters were not significant in this thesis. However, the estimator should work for these cases also.

3.6.1 Generation of Simulated Data

To create white noise and random walk, the *rand* function on Matlab™ was used. The *rand* function creates a normally distributed Gaussian white noise sequence with a mean of zero and a variance of 1 [30].

3.6.1.1 Generation of White Noise Sequences

To generate white noise sequences with variances other than 1, the output of the *rand* function was multiplied by the desired standard deviation of the noise, as Equations 3.54 through 3.56 show. If N is a white noise sequence of n terms with zero mean and variance 1, let $R = \sigma N$, where R is also a white noise sequence, but with standard deviation σ. The variance of N is defined as

$$\operatorname{var} N = \frac{1}{n}\sum_{i=1}^{n} N_i^2 = 1 \tag{3.54}$$

Therefore,

$$\operatorname{var} R = \frac{1}{n}\sum_{i=1}^{n} R_i^2 \tag{3.55}$$

Substituting $R_i = \sigma N_i$ yields

$$\operatorname{var} R = \sigma^2 \frac{1}{n}\sum_{i=1}^{n} N_i^2 = \sigma^2 \operatorname{var} N = \sigma^2 \tag{3.56}$$

As shown, the variance of the white noise sequence can be altered by multiplying by the desired standard deviation.

3.6.1.2 Generation of Random Walk Sequences

Random walk is the integral of a white noise sequence. Therefore, a white noise sequence generated by the *rand* function can be used to create a random walk sequence. The following algorithm produces a random walk sequence [43]:

$$rw(t_i) = rw(t_{i-1}) + b\sqrt{\Delta t}N(t_i) \tag{3.57}$$

where rw is the random walk sequence, b is the random walk standard deviation parameter with units [$1/\sqrt{\text{sec}}$], Δt is the time step, and N is a white noise sequence of zero mean and unit variance.

3.7 Results of Simulated Data Analysis

Table 3.1 shows the results of these test cases and the uncertainties in the parameter estimates. In all three tests, the maximum likelihood estimate was able to converge closely to the actual values of the parameters. Extensive testing showed that the estimator is able to converge to the measurement noise from a significantly higher initial guess, but that the process noise initial guess should be lower than the expected value.

Table 3.1. Results of Maximum Likelihood Estimator Verification

Test Number	Parameters	Initial Guess	Actual Value	Estimate
3A1	Measurement Noise	1.0000	0.09935	0.09700 ± 0.00451
3A2	Process Noise	0.0100	0.09902	0.09028 ± 0.01916
3A3	Measurement Noise	1.0000	0.10036	0.09835 ± 0.00461
	Process Noise	0.0100	0.09922	0.07038 ± 0.01832

Figures 3.3 and 3.4 are plots that demonstrate the convergence of the parameter estimates, and the convergence of the uncertainties of the parameter estimates. Both of these plots are for typical parameter estimates, in this case the measurement noise standard deviation. In Figure 3.3, the estimate reaches the correct value of 0.1 units, although the initial estimate is an order of magnitude higher. The parameter estimates usually converge to the correct value after five estimation intervals.

An additional test was run with varying measurement noise to determine the ability of the maximum likelihood estimator to follow a varying noise process. The covariance of the measurement noise for this test was generated using an increasing exponential function. Figure 3.5 shows the estimated value of the measurement noise standard deviation as a solid line, and the actual value of the standard deviation as a dashed line.

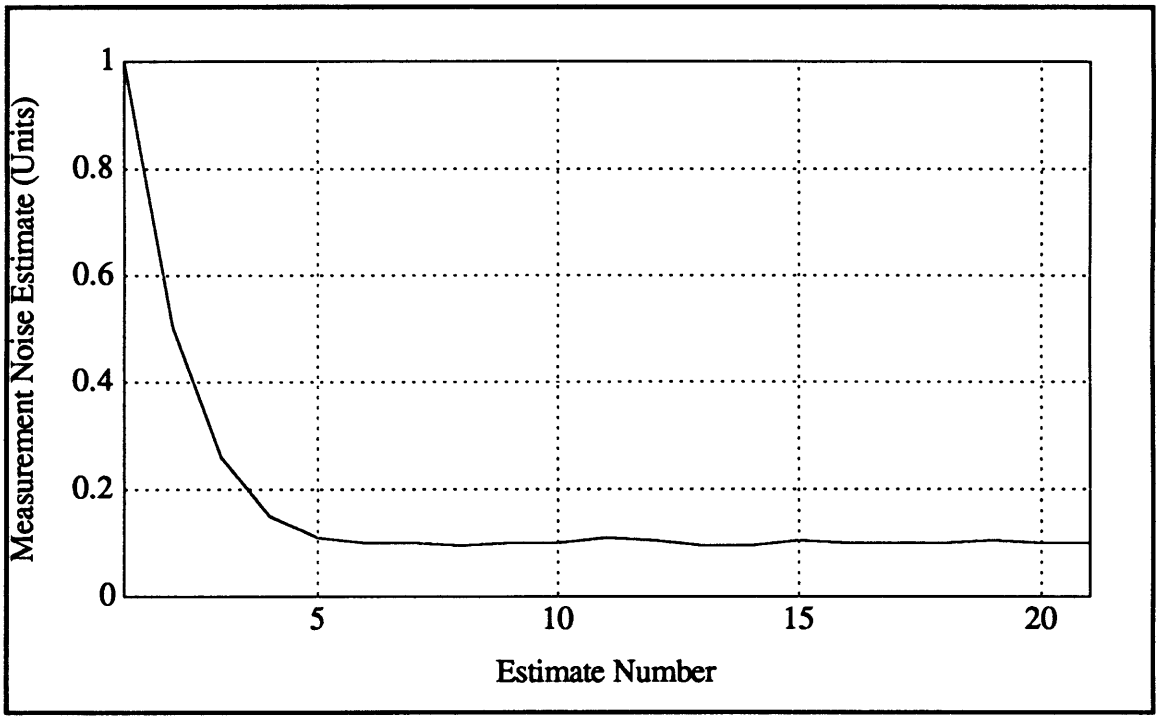

Figure 3.3. Typical Parameter Estimate Convergence

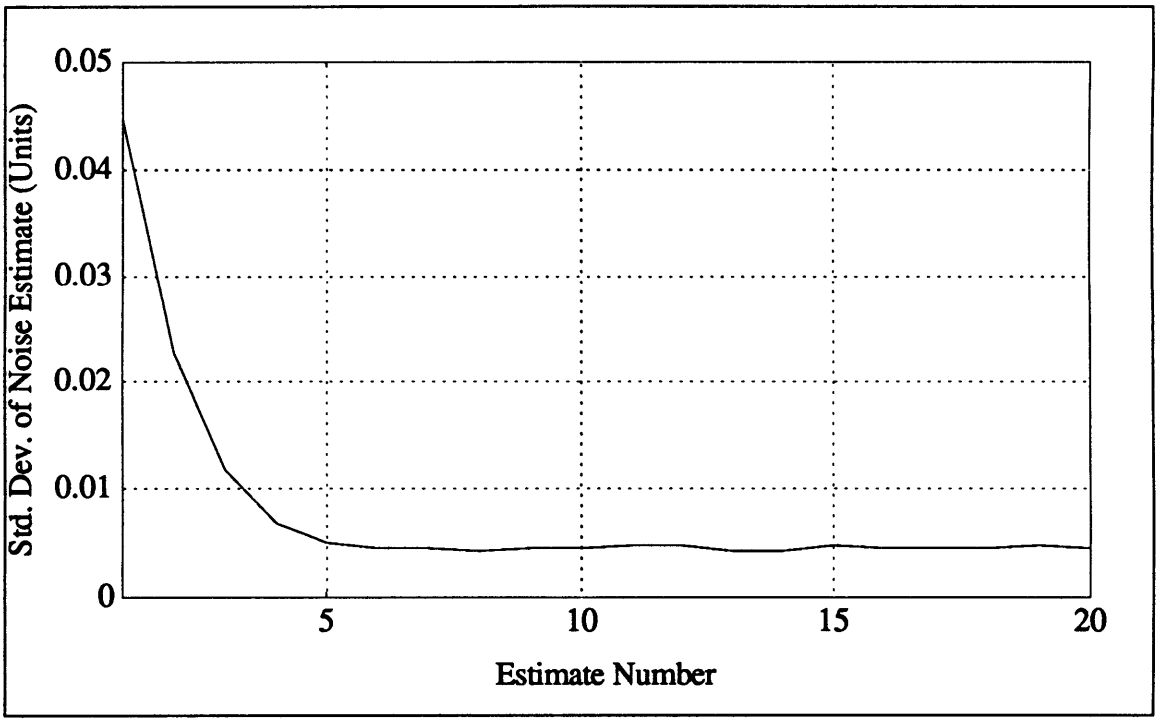

Figure 3.4. Typical Parameter Estimate Standard Deviation

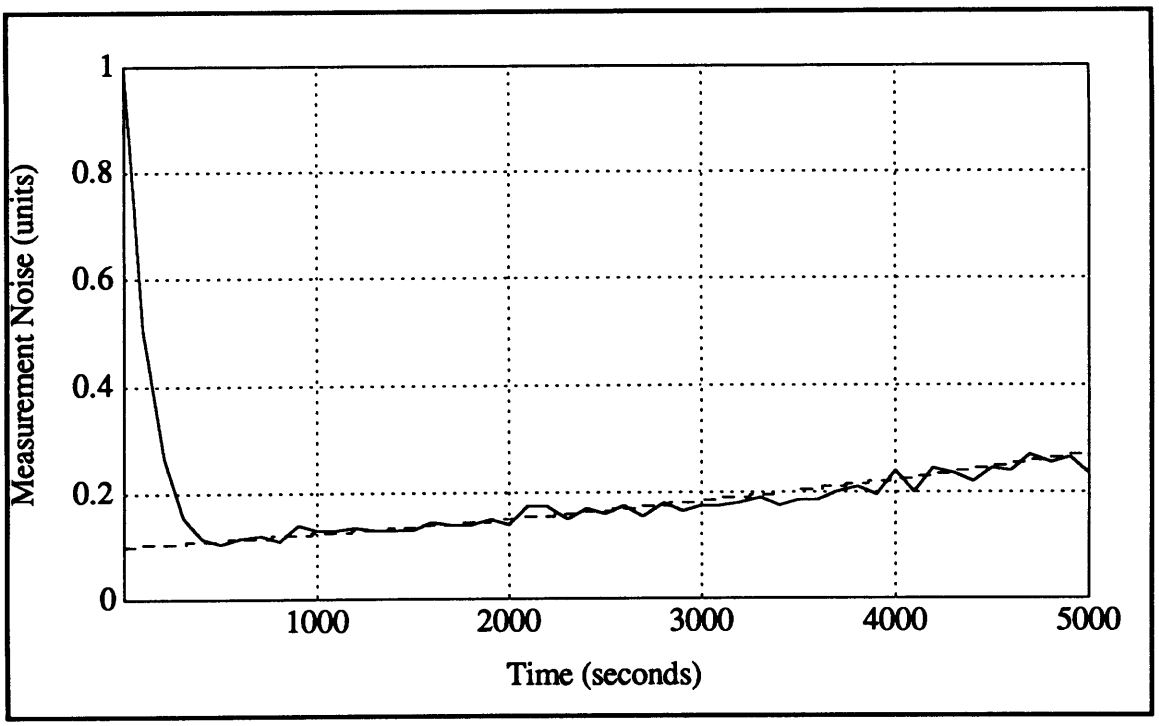

Figure 3.5. Convergence of Measurement Noise Estimate

3.8 Conclusions

These simulations demonstrate the effectiveness of the maximum likelihood technique for estimating unknown parameters in the Kalman filter. This estimator is capable of achieving accurate estimates of both **Q** and **R**. Figures 3.3 and 3.4 show that the estimate quickly converges from an initial guess an order of magnitude higher to approximately the correct value. Therefore, this filter quickly becomes an optimal filter. Figure 3.5 shows that the estimator can follow a changing noise signal, and will therefore remain optimal if the parameters of the system change. This achievement is significant; the estimator will prevent performance degradation due to slowly varying noise signals. Both the raw data reduction and reduced data analysis adaptive filters incorporate this estimator.

Chapter 4

Model Adaptive Filter Using Correlation Methods

4.1 Model Adaptive Filter

Often, when processing data, the system model which describes the behavior of the instrument has not been or cannot be completely described. In order to test the completeness of a given system model, a term could be added to the system model, and the measurement data could then be processed using this enhanced model. However, the results from this procedure may be misleading. When terms are added to a system model, the Kalman filter tries to determine the value of the term, not whether it is actually present. Therefore, terms that do not actually exist may appear to exist because of the given system model. Another approach is therefore necessary. By exploiting the information in the innovation sequence, the existence of unmodeled terms can be determined during the initial data processing. In this approach, the innovation sequence is reduced to white noise by enhancing the system model. Correlation methods are used to determine whether or not an additional term should be added to the system model.

4.2 Theory of Correlation Functions

Two random variables X and Y can be expressed in terms of their correlation [4],

$$\varphi_{XY}(t_1,t_2) = E\{X(t_1)Y(t_2)\} \tag{4.1}$$

in which φ is the correlation function between X and Y, and $E\{\bullet\}$ is the expectation operator. The crosscorrelation function can also be written as an integral:

$$\varphi_{XY}(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)dxdy \tag{4.2}$$

where f_{XY} is the joint probability density of X and Y.

The autocorrelation function of the random variable X is expressed as

$$\varphi_{XX}(t_1,t_2) = E\{X(t_1)X(t_2)\} \tag{4.3}$$

or in integral form as,

$$\varphi_{XX}(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1(t_1)x_2(t_2)f_{XX}(x_1,x_2)dx_1dx_2 \tag{4.4}$$

If the two random processes are stationary, that is the probability density does not change with time, then the correlation function can be simplified to be dependent only on the time difference $\tau = t_2 - t_1$, as shown in Equation 4.5.

$$\varphi_{XY}(\tau) = E\{X(t)Y(t+\tau)\} \tag{4.5}$$

4.3 Derivation of Model Adaptive Filter

Suppose that a random process $P(t)$, such as temperature, pressure, rate, etc. is known completely throughout the measurement interval [49]. If another random signal, $S(t)$, such as the innovation sequence, is an unknown function of this random process,

$$S(t) = kP(t-T) + N(t) \tag{4.6}$$

where k is an unknown constant, T is a time delay, and $N(t)$ is a white noise process, then the non-stationary crosscorrelation function between the signal and the random process is

$$\varphi_{SP}(t_1,t_2) = \overline{S(t_1)P(t_2)} \tag{4.7}$$

Substituting in for $S(t)$ gives

$$\varphi_{SP}(t_1,t_2) = \overline{[kP(t_1-T)+N(t_1)]P(t_2)} \tag{4.8}$$

Expanding the expectation yields

$$\varphi_{SP}(t_1,t_2) = k\overline{P(t_1-T)P(t_2)} + \overline{N(t_1)P(t_2)} \tag{4.9}$$

But the first term in the crosscorrelation is just the autocorrelation of the random process $P(t)$, and if $N(t)$ is a zero mean noise process and is independent of $P(t)$, then the crosscorrelation of $S(t)$ and $P(t)$ is,

$$\varphi_{SP}(t_1,t_2) = k\varphi_{PP}(t_1-T,t_2) \tag{4.10}$$

Equation 4.10 shows that the crosscorrelation function between $P(t)$ and $S(t)$ is directly proportional to the autocorrelation function of $P(t)$, which is usually known. Therefore, by taking the crosscorrelation of $S(t)$ and $P(t)$ and the autocorrelation of $P(t)$, a relationship between the signal $S(t)$ and $P(t)$, if it exists, can be found.

In a real system, however, the assumption that $N(t)$ and $P(t)$ are independent is not valid. Therefore, the crosscorrelation function becomes:

$$\varphi_{SP}(\tau) = k\varphi_{PP}(\tau - T) + \varphi_{NP}(\tau) \tag{4.11}$$

In this case, the scaling and delay can still be found, as shown in the next section.

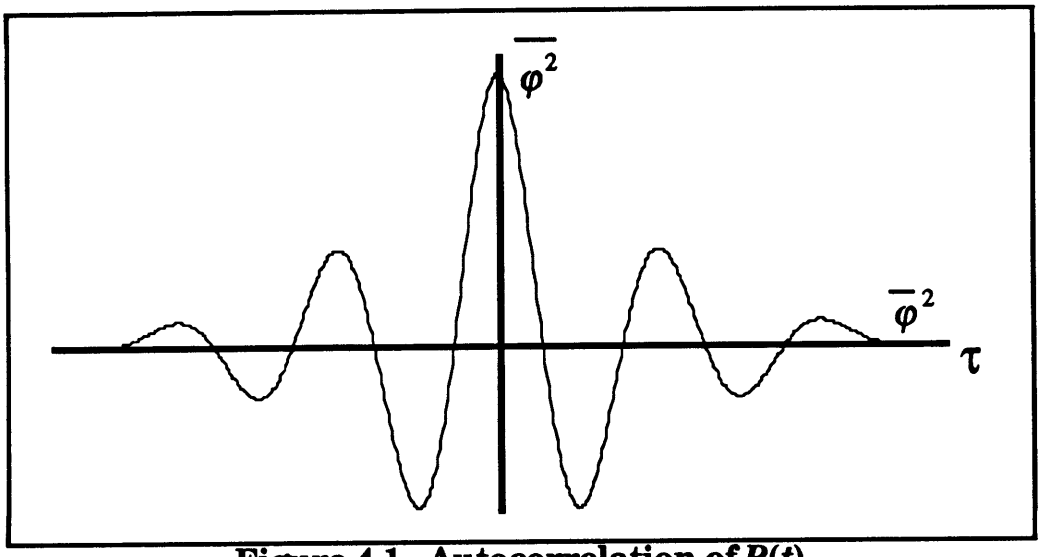

Figure 4.1. Autocorrelation of $P(t)$

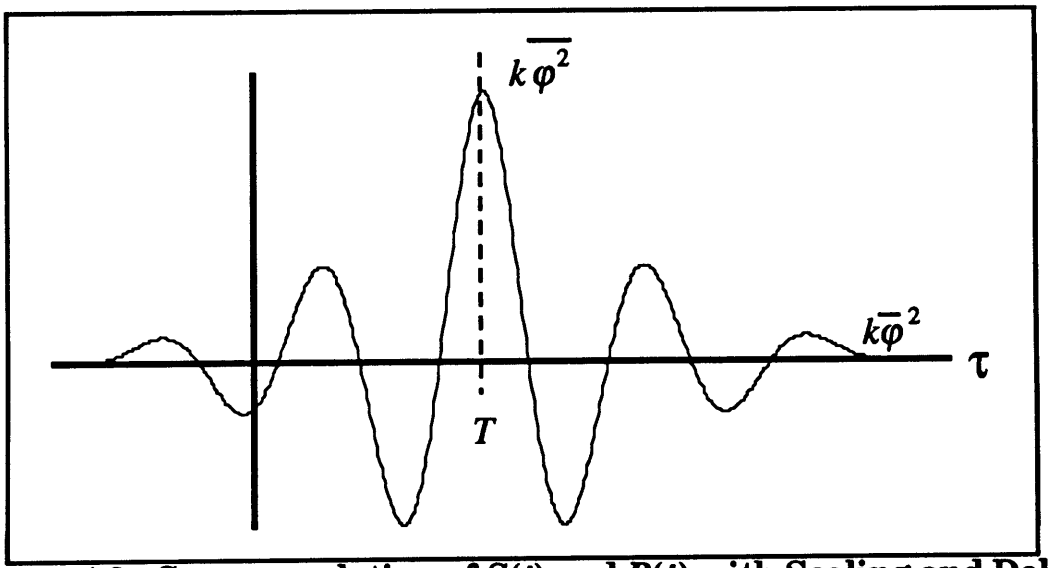

Figure 4.2. Crosscorrelation of $S(t)$ and $P(t)$ with Scaling and Delay

4.3.1 Implementation of Model Adaptive Filter

The system model is not modified unless the innovation sequence does not test as white noise. If the innovation sequence is colored noise, then the model adaptive filter implements the following procedure to identify possible unmodeled terms. First, the autocorrelation function of the random process P is determined. The crosscorrelation between the random process P and the innovation sequence **r** is then taken.

The delay, if it exists, must be determined next using the following steps. By calculating φ_{SP}, the delay will move the crosscorrelation with respect to the autocorrelation of $P(t)$. At the zeroth lag of the autocorrelation, the value will be a maximum. For the autocorrelation, this value is $\overline{\varphi^2}$, as shown in Figure 4.1. At t = ∞, the value of the autocorrelation function will be $\overline{\varphi}^2$. Therefore, the crosscorrelation maximum will correspond to a scaled value of $\overline{\varphi^2}$. The delay can be calculated by determining the time difference between the maxima of the crosscorrelation and autocorrelation. In this thesis, the delay was assumed to be zero for all processes.

The value of k must then be calculated. First, the autocorrelation is shifted by the delay T so that the maxima are aligned. After removing the delay, the error of the best estimate of k can be written as:

$$\varepsilon = \varphi_{SP}(\tau) - k\varphi_{PP}(\tau) \tag{4.12}$$

If the sum of the squares of this error is minimized with respect to k, then the best estimate of k can be found. First, rewrite ε as the sum of the squares of Equation 4.12.

$$\varepsilon = \sum_{j=1}^{N}\left(\varphi_{SP}(\tau_j) - k\varphi_{PP}(\tau_j)\right)^2 \tag{4.13}$$

The subscript j represents the discrete increments of the correlation functions. Next, minimize ε with respect to k:

$$\frac{\partial \varepsilon}{\partial k} = -2\sum_{j=1}^{N}\varphi_{PP}(\tau_j)\left(\varphi_{SP}(\tau_j) - k\varphi_{PP}(\tau_j)\right) \tag{4.14}$$

Now, set the derivative equal to zero, and solve for k.

$$k = \frac{\sum_{j=1}^{N}\varphi_{PP}(\tau_j)\varphi_{SP}(\tau_j)}{\sum_{j=1}^{N}\left(\varphi_{PP}(\tau_j)\right)^2} \tag{4.15}$$

The signal S is the innovation sequence, and the process P is the unmodeled variable. To determine whether the process P should be included in the system model, P is multiplied by the best estimate of k, and this product is subtracted from the innovation sequence for all past time, resulting in a difference sequence. This difference sequence represents the innovation sequence if the additional term was included in the original system model. The RMS of this difference sequence is then compared with the RMS of the innovation sequence. If the addition of this term improves the RMS by at least five percent, then the term is included in the system model. Otherwise, the term is not added.

4.4 Verification of Model Adaptive Filter

To demonstrate that this model adaptive theory is valid, gyro outputs were generated in Matlab™ for several cases of unmodeled terms: trend, scale factor, and rate squared. This data was then analyzed using the model adaptive filter, and results of these tests are shown in Table 4.1 and in Figures 4.3 through 4.8. Table 4.1 lists the actual model values, the initial model estimates, and the final model estimates determined by the filter.

Table 4.1. Results of Model Adaptive Filter Verification

Test Number	Modeled Terms	Actual Model	Initial Model	Final Model
4A1	Constant	2.0000	0.0000	2.0083 ± 0.0080
	Trend	2.0000	Not Modeled	1.9981 ± 0.0013
4A2	Constant	2.0000	0.0000	1.9986 ± 0.0033
	Trend	None	Not Modeled	Not Modeled
4A3	Bias	1.1000	0.0000	1.0996 ± 0.0039
	Scale Factor	0.0150	Not Modeled	0.0150 ± 8.94e-5
4A4	Bias	1.100	0.0000	1.1007 ± 0.0036
	Scale Factor	None	Not Modeled	Not Modeled
4A5	Bias	1.1000	0.0000	1.1008 ± 0.0051
	Scale Factor	0.0150	0.0000	0.0150 ± 1.06e-4
	Rate Squared	0.0003	Not Modeled	2.98e-4 ± 1.9e-6
4A6	Bias	1.1000	0.0000	1.1006 ± 0.0036
	Scale Factor	0.0150	0.0000	0.0149 ± 7.58e-5
	Rate Squared	None	Not Modeled	Not Modeled

For Tests 4A1 and 4A2, the data was generated at 100 Hz for 10 seconds. Using an estimation interval of 100 points, the trend term was added at the first interval in Test 4A1, as seen in Figure 4.3; it was never added in Test 4A2. In Figure 4.4, the dashed line represents the innovation sequence, and the solid line represents the innovation sequence after the trend (scaled by k) has been subtracted from the innovation sequence. The improvement in the RMS of the innovation sequence by removing the trend term was 62.8 percent. In Test 4A2, no trend term was identified.

In Tests 4A3 and 4A4, the data was generated at 100 Hz for 8.6 seconds, and an estimation interval of 100 points was again used. For these tests, the commanded rate profile was from 100 deg/sec to -100 deg/sec in 25 deg/sec increments. In Test 4A3, the scale factor term was added at the first estimation interval. Figure 4.5 shows the innovation sequence for Test 4A3 for all time. The innovation sequence appears to have a transient because the estimate of the bias is adjusting to compensate for the poor system model. Figure 4.6 compares the pre-modified (dashed line) and post-modified (solid line) residuals. In Test 4A3, the improvement by adding the rate dependent term was 72.3 percent. The scale factor term was never added in Test 4A4.

In Tests 4A5 and 4A6, the commanded rate profile ranged from $\pm$ 100 deg/sec to $\pm$ 25 deg/sec in 25 deg/sec increments. The first commanded rate was +100 deg/sec, the second was -100 deg/sec, the third was +75 deg/sec, and so on. In Test 4A5, the first rate squared term test, the term was added at the second interval, which was the first opportunity for the filter to differentiate between the rate and the square of the rate due to the sign of the rate. The term was added when the improvement in the residual was 24.2 percent. The rate squared term was never added in Test 4A6.

In the three tests that did not have unmodeled terms (Tests 4A2, 4A4, and 4A6) the estimate of the scaling term k was always of a small enough magnitude such that the suspected unmodeled term would not affect the innovation sequence. This prevented the inclusion of a nonexistent term accidentally. In these three tests, the post-modified residual was almost identical to the pre-modified residual.

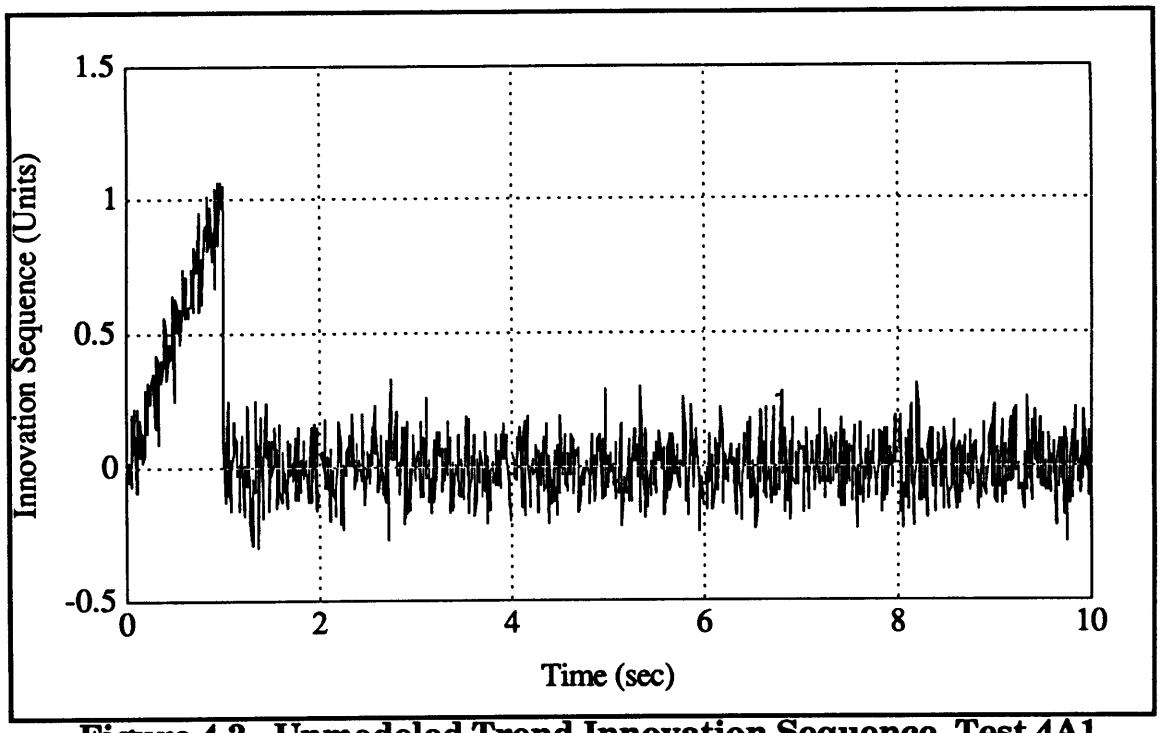

Figure 4.3. Unmodeled Trend Innovation Sequence, Test 4A1

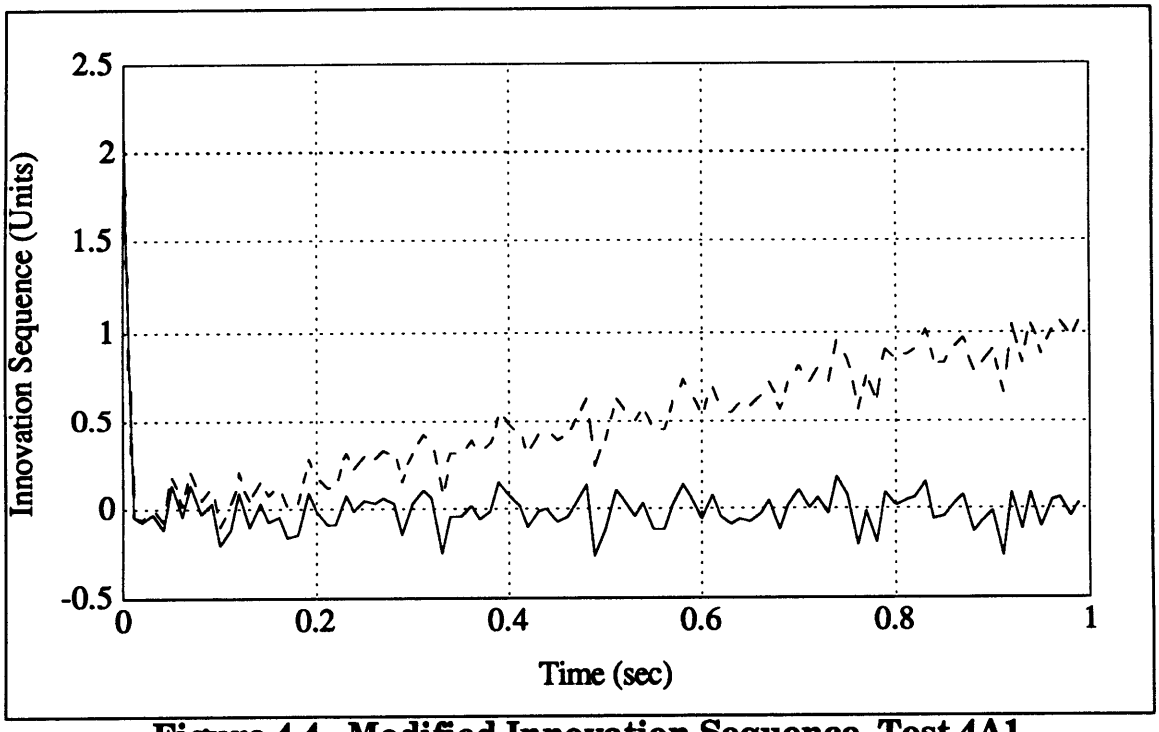

Figure 4.4. Modified Innovation Sequence, Test 4A1

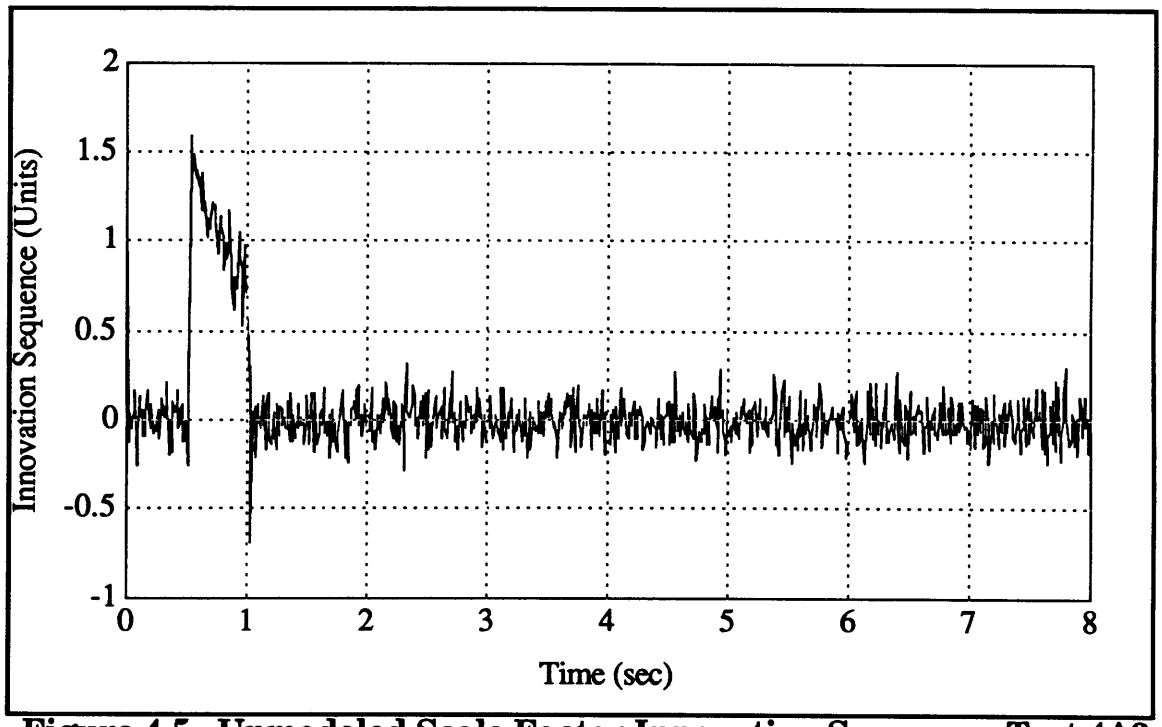

Figure 4.5. Unmodeled Scale Factor Innovation Sequence, Test 4A3

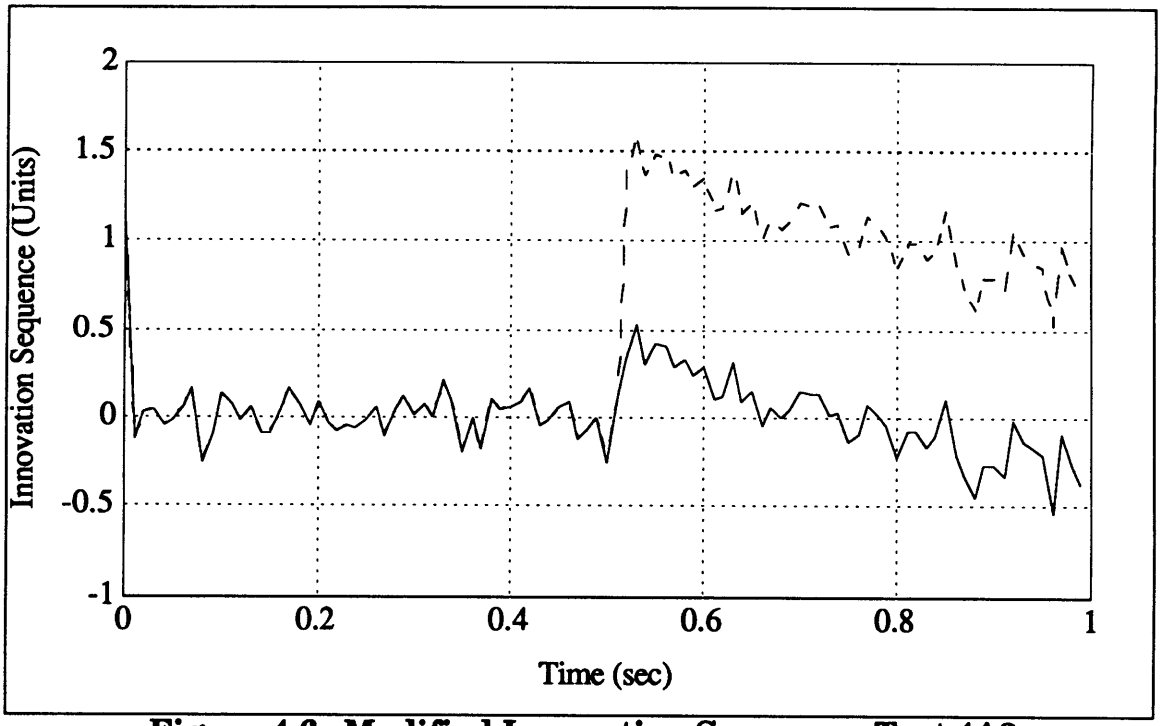

Figure 4.6. Modified Innovation Sequence, Test 4A3

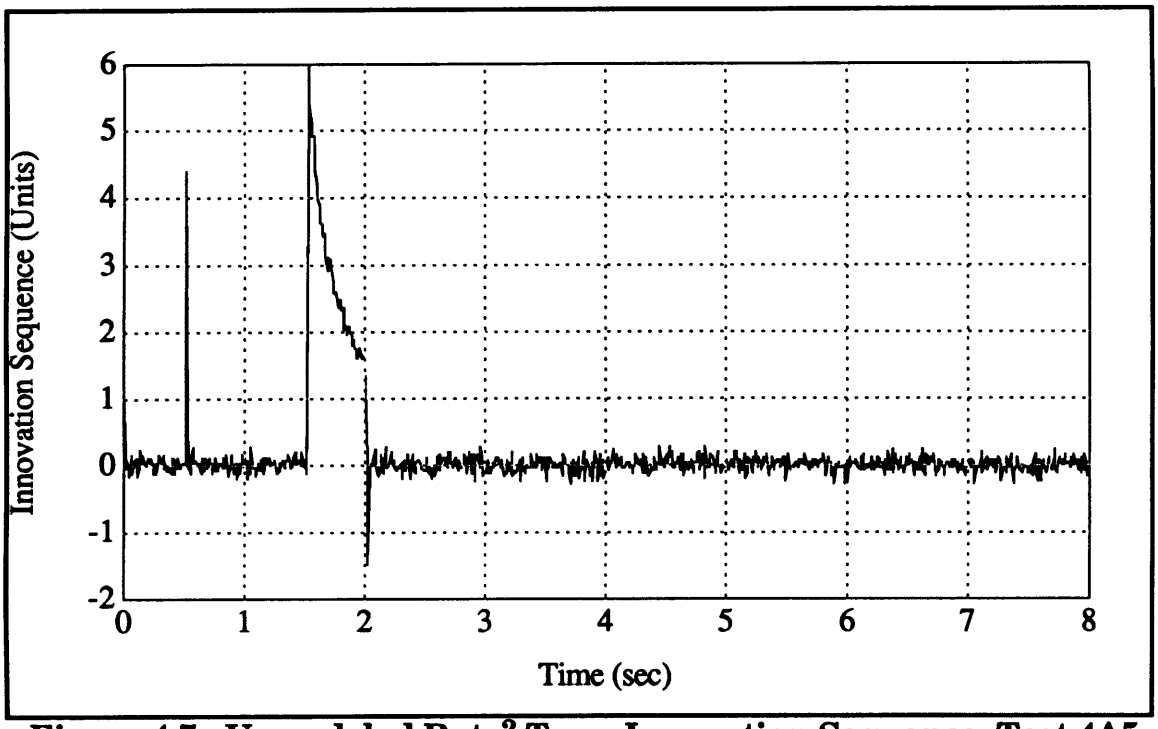

Figure 4.7. Unmodeled Rate^2 Term Innovation Sequence, Test 4A5

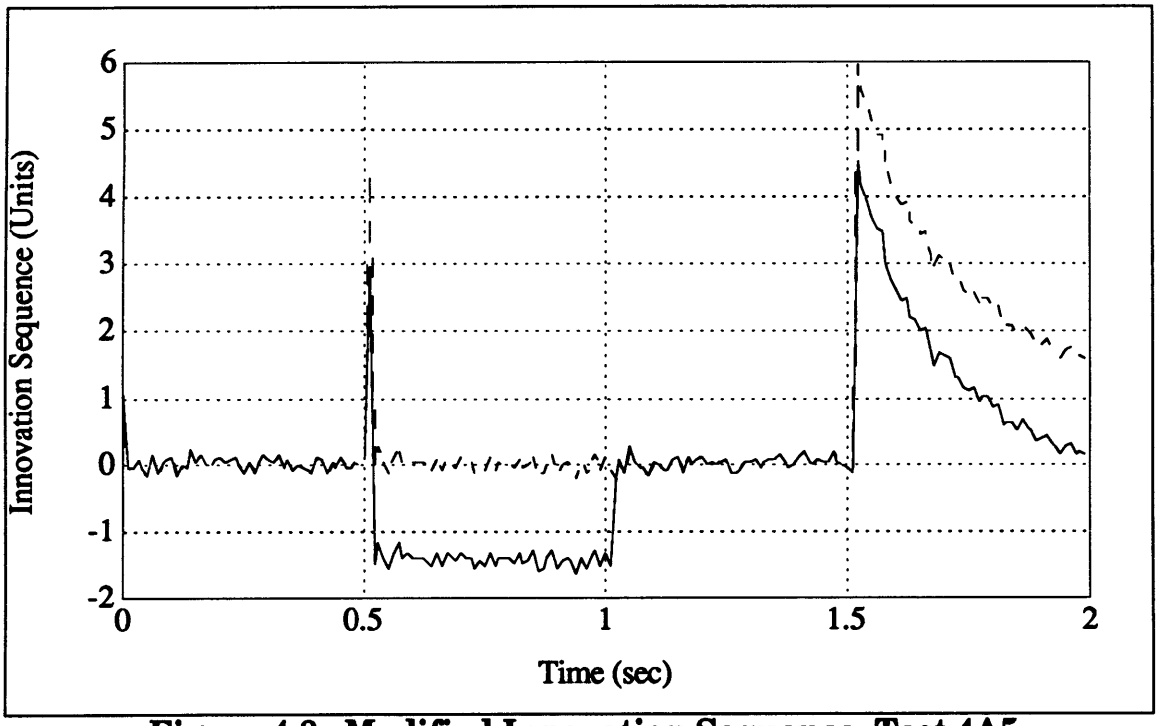

Figure 4.8. Modified Innovation Sequence, Test 4A5

4.5 An Example of the Model Adaptive Filter in a Historical Context

For electromechanical gyroscopes, accelerations along the principal axes can cause drift errors in the output of the gyroscope. The three first-order drift error terms are D_I, D_O, and D_S, which correspond to the accelerations along the input axis (IA), the output axis (OA) and the spin axis (SA), respectively. For the ball bearing gyroscopes, which were the first type of gyroscopes widely used, the error term D_{IS}, which corresponds to the drift error due to the product of accelerations along both IA and SA, was identified as the only second order error term. When gas bearing gyroscope technology emerged, another second order term, D_{OS}, corresponding to the error due to the product of the accelerations along OA and SA, was identified as an additional error source. This was only discovered after significant testing was done without the D_{OS} term included in the state model. If the ball bearing error model D_{IS} is used to analyze data from a gas bearing gyroscope, then D_{OS} effects will be unmodeled, and the results are suboptimal. However, if this term could be modeled during processing, then improvement in the gyroscope performance is rapidly achieved. The model adaptive filter can identify D_{OS} as the next significant error term in the gas bearing gyroscope error model, based on the output of a vibration test.

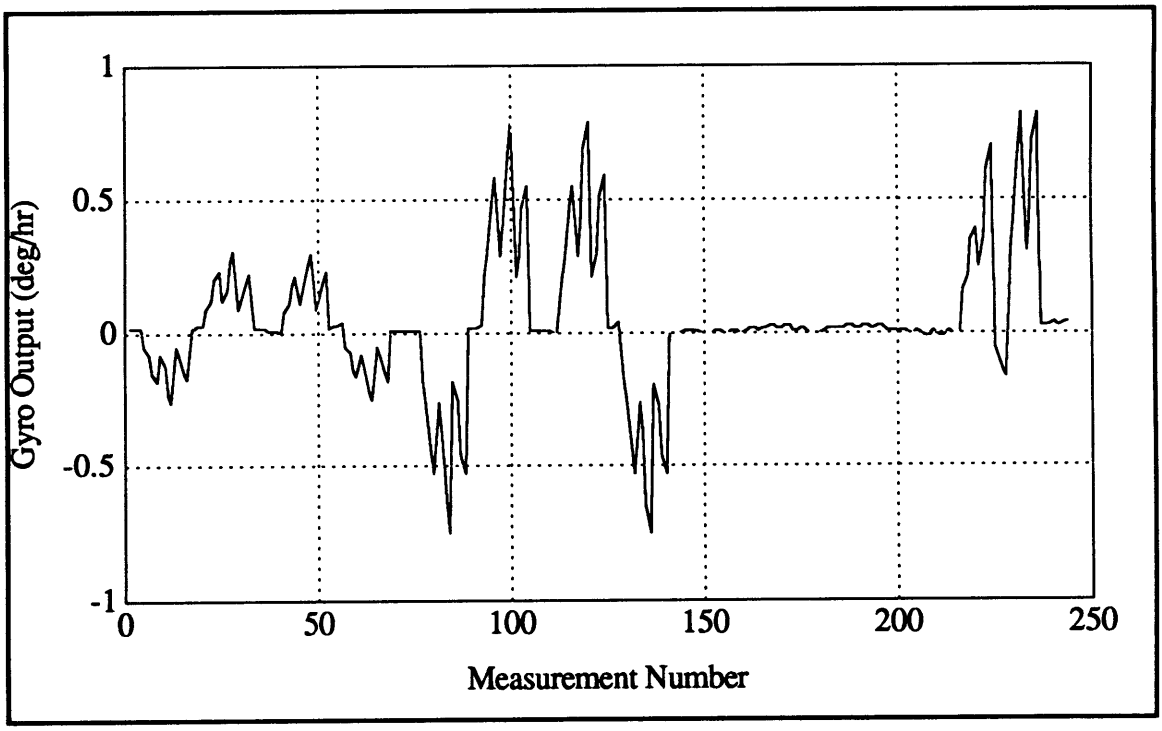

Figure 4.9. Gas Bearing Gyro Output from Shaker Test, Test 4B

For this application of the model adaptive filter, data from multiple position and g-level vibration testing at a certain frequency, consisting of 244 measurements, was used.

The output of the gyro for this test is shown in Figure 4.9. The data was reduced so that each measurement represented an average of the data over an integral number of vibrations. Each acceleration term is represented as a constant times the sine function, $A\sin(\omega t)$ where A is the amplitude of vibration and ω is the frequency of vibration. This data reduction technique eliminated the first order drift error terms from the data because the integral of the sine over one cycle is zero.

In Figure 4.9, measurement numbers 1 through 72 represent vibration with the input axis (IA) perpendicular to the earth's axis. Measurement numbers 73 through 144 represent vibration with the output axis (OA) parallel to the earth's axis, and measurements 145 through 216 represent vibration with the spin axis (SA) parallel to the earth's axis. In each of these 216 measurements, the vibration is orthogonal to the earth's axis, causing vibration along one or two of the gyroscope axes, depending on the gyroscope orientation. The remaining 28 measurement numbers are various orientations that were chosen to cause simultaneous acceleration along all three gyro axes.

Initially, the data was analyzed using only the D_{IS} model. The innovation sequence from that test is shown in Figure 4.10. Figure 4.10 shows that improvement in the state model can probably be made. In the second filter test, the model adaptive filter was used. The initial state vector was D_{IS} only. The D_{OS} term was identified as the term to analyze for possible inclusion. Using an estimation interval of 30 points, the model adaptive filter identified the D_{OS} term as significant, and successfully added this term to the state vector during the first estimation interval. By adding this term to the system model. the improvement in the RMS of the residual was 86.2 percent. Figure 4.11 shows that, at the first attempt to add to the system model, the model adaptive filter identified the D_{OS} term as a correlated variable, and then successfully added this term to the state vector. Both Figures 4.10 and 4.11 have a bad data point at Measurement Number 73. No explanation for this spike has been found. The improvement in the system model is easily seen by comparing Figure 4.10 to Figure 4.11, in which the D_{OS} term was never added.

Table 4.2 shows the true values of the D_{IS} and D_{OS} terms, the estimate of D_{IS} for the first filter test, and the estimates of D_{IS} and D_{OS} for the second filter test. It should be noted that, in the second filter test, the filter estimated the coefficient D_{OS}. However, in a proper model, the term would be $1/2D_{OS}$. Therefore, the estimate of D_{OS} must be multiplied by two in order to determine the true value of D_{OS}. This factor of 1/2 would be readily apparent during post processing mathematical analysis of the results and the new system model. For the second order terms, the expression for acceleration is squared, and the average of the sine squared over an integral number of cycles is 1/2.

The estimate for D_{OS} shown in Table 4.2 has been multiplied by 2 to determine the proper estimate.

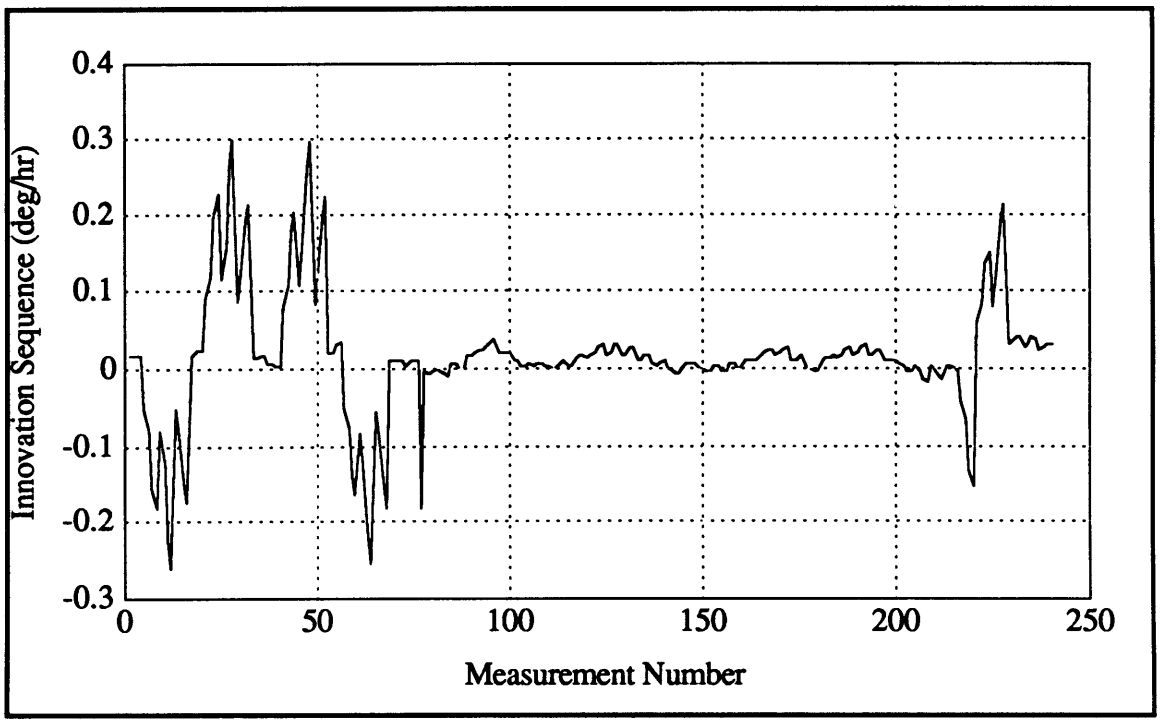

Figure 4.10. D_{IS} System Model Innovation Sequence, Test 4B1

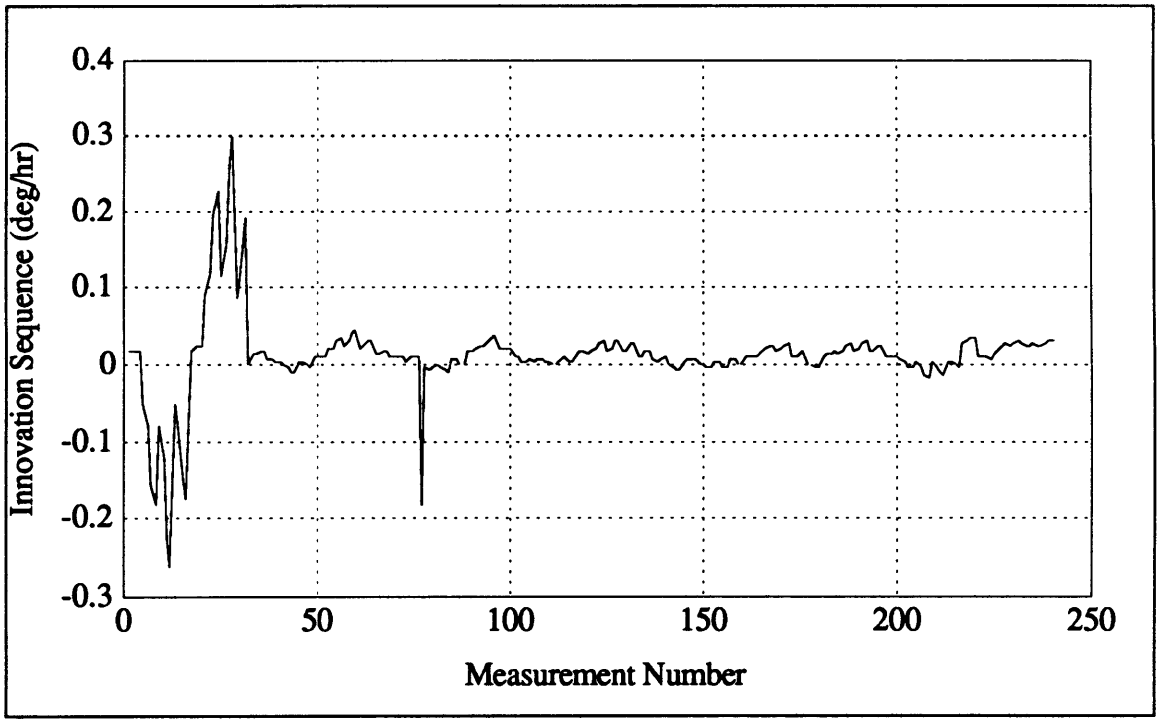

Figure 4.11. Model Adaptive Filter Innovation Sequence, Test 4B2

Table 4.2 shows that the RMS of the innovation sequence is improved, and the estimates of D_{IS} and D_{OS} are very close to their true values, which have been established through extensive testing. It should be noted that, if the D_{OS} term is included in the original system model, then the RMS of the innovation sequence is 0.0215 deg/hr.

Table 4.2. Model Adaptive Filter Analysis of Gas Bearing Gyro Data

Coefficient	Actual Value	Filter Test 4B1	Filter Test 4B2
D_{IS}	-0.0319 ± 2.389e-4	-0.0314 ± 1.67e-4	-0.0319 ± 6.679e-4
D_{OS}	0.0111 ± 2.397e-3	Not Modeled	0.0112 ± 4.686e-4
RMS of **r**	—	0.0822	0.0555

4.6 Conclusions

By using correlation methods, the model adaptive filter is able to identify unmodeled terms and include them in the system model during data processing. The data verification runs have demonstrated that this filter is capable of identifying a variety of unmodeled terms, from trend to an acceleration dependent error coefficient such as D_{OS}. These simulations have shown that by using the maximum likelihood technique to determine the scaling factor k, only significant terms are identified and added to the model. In none of the tests in which the system model was correct did the model adaptive filter add a term. In Chapter 6, this filter is incorporated into the reduced data analysis filter.

Chapter 5

Power Spectral Density Analysis

5.1 Motivation for PSD Analysis

Besides the unmodeled effects discussed in Chapter 4, periodic noise signals may exist in instrument outputs. A periodic noise signal may be due to stray electrical signals or a malfunction in the instrument hardware. For example, in the micromechanical gyroscope, the inertial element is vibrated at its natural frequency. Later, this sinusoidal signal is demodulated to obtain the DC signal. However, if there is a discrepancy between the demodulation frequency and the vibration frequency, then a periodic noise signal will result. This periodic noise signal may not be removed by the baseband filter, and could therefore exist in the gyroscope output. Another possible periodic noise source is the 60 Hz frequency of AC power sources. Poor circuit design may result in a 60 Hz signal in the output of an instrument. Fortunately, modeling these periodic noise signals is easily accomplished with a combination of sine and cosine terms. The identification of these periodic signals can be accomplished using power spectral density methods.

5.2 Power Spectral Density Theory

The power spectral density function (PSD) of a stationary random process determines the frequency content of that process. The power spectral density function can be defined in terms of the correlation function, as shown in Equation 5.1 [4].

$$\Phi_{XY}(\omega) = \int_{-\infty}^{\infty} \varphi_{XY}(\tau) e^{-j\omega\tau} d\tau \tag{5.1}$$

where ω is 2π times the frequency in Hertz, and $j = \sqrt{-1}$. The units of the power spectral density function are [units^2/Hertz]. The PSD may also be expressed in terms of Fourier Transforms, as discussed in the next section.

5.2.1 Fourier Transforms

The Fourier transform of a function $f(t)$ is expressed as

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{5.2}$$

The inversion formula from the frequency domain to the time domain is

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(j\omega)e^{j\omega t}d\omega \tag{5.3}$$

Using these two equations, it can be shown that the Fourier transform of a convolution, shown in Equation 5.4,

$$x(t) * y(t) = \int_{-\infty}^{\infty} x(\tau)y(t-\tau)d\tau \tag{5.4}$$

is equal to the product of the Fourier transforms of each of the variables, as shown in Equation 5.5.

$$\int_{-\infty}^{\infty} x(t) * y(t)e^{-j\omega t}dt = X(\omega)Y(\omega) \tag{5.5}$$

A similar approach can be taken with the correlation function, which was discussed in Chapter 4. For stationary random processes the correlation function is

$$\varphi_{XY}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{XY}(x,y)dxdy \tag{5.6}$$

The Fourier transform of a correlation function is

$$\int_{-\infty}^{\infty} \varphi_{XY}(\tau)e^{-j\omega\tau}d\tau = X(\omega)\overline{Y}(\omega) \tag{5.7}$$

where $\overline{Y}(\omega)$ is the complex conjugate of $Y(\omega)$.

5.2.2 Power Spectral Density Function for Finite Data Sets

Substituting Equation 5.7 into Equation 5.1 gives an expression of the PSD in terms of Fourier transforms:

$$\Phi_{XY}(\omega) = X(\omega)\overline{Y}(\omega) \tag{5.8}$$

However, this equation is only valid for all time, i.e. the bounds of the Fourier transform are -∞ to ∞. In real data analysis, a finite segment of data is used, $0 \le t \le T$. In order to analyze this data for a finite segment of data, $X_T(t)$ is defined as a segment of the random process $X(t)$. The periodogram of this truncated signal $X_T(t)$ is defined by

$$periodogram = \frac{1}{T}\left|F\{X_T(t)\}\right|^2 \tag{5.9}$$

where $F\{\bullet\}$ is the finite Fourier transform [4]. Taking the expectation of the periodogram yields:

$$\begin{aligned} & E\left\{\frac{1}{T}\left|F\{X_T(t)\}\right|^2\right\} \\ & = E\left\{\frac{1}{T}\int_0^T X(t)e^{-j\omega t}dt\int_0^T X(s)e^{j\omega s}ds\right\} \\ & = \frac{1}{T}\int_0^T\int_0^T E\{X(t)X(s)\}e^{-j\omega(t-s)}dt\,ds \end{aligned} \tag{5.10}$$

The subscript T on X has been dropped because the range of integration of X is now only over the interval T, and this is the definition of $X_T(t)$. If $X(t)$ is stationary, then the expectation becomes the autocorrelation function of $X(t)$, and Equation 5.10 becomes [4]

$$E\left\{\frac{1}{T}\left|F\{X_T(t)\}\right|^2\right\} = \frac{1}{T}\int_0^T\int_0^T \varphi_{XX}(t-s)e^{-j\omega(t-s)}ds\,dt \tag{5.11}$$

Now, let $\tau = t - s$. The double integral of Equation 5.11 can be rewritten as

$$\frac{1}{T}\int_0^T\int_0^T \varphi_{XX}(t-s)e^{-j\omega(t-s)}ds\,dt = -\frac{1}{T}\int_0^T\int_t^{t-T} \varphi_{XX}(\tau)e^{-j\omega\tau}d\tau\,dt \tag{5.12}$$

Next, change the order of integration, and break up the integral into two intervals:

$$E\left\{\frac{1}{T}\left|F\{X_T(t)\}\right|^2\right\} = \frac{1}{T}\int_{-T}^{0}\int_{0}^{\tau+T}\varphi_{XX}(\tau)e^{-j\omega\tau}dt\,d\tau + \frac{1}{T}\int_{0}^{T}\int_{\tau}^{T}\varphi_{XX}(\tau)e^{-j\omega\tau}dt\,d\tau \tag{5.13}$$

By integrating with respect to t, Equation 5.13 becomes

$$E\left\{\frac{1}{T}\left|F\{X_T(t)\}\right|^2\right\} = \frac{1}{T}\int_{-T}^{0}(\tau+T)\varphi_{XX}(\tau)e^{-j\omega\tau}d\tau + \frac{1}{T}\int_{0}^{T}(T-\tau)\varphi_{XX}(\tau)e^{-j\omega\tau}d\tau \tag{5.14}$$

which can be rewritten more compactly as

$$E\left\{\frac{1}{T}\left|F\{X_T(t)\}\right|^2\right\} = \int_{-T}^{T}\left(1-\frac{|\tau|}{T}\right)\varphi_{XX}(\tau)e^{-j\omega\tau}d\tau \tag{5.15}$$

As $T \to \infty$, the expectation becomes

$$E\left\{\frac{1}{T}\left|F\{X_T(t)\}\right|^2\right\} \Rightarrow \int_{-\infty}^{\infty}\varphi_{XX}(\tau)e^{-j\omega\tau}d\tau = \Phi_{XX}(\omega) \tag{5.16}$$

Therefore, an approximation to the power spectral density is:

$$\tilde{\Phi}_{XX}(\omega) = \frac{1}{T}\left|F\{X_T(t)\}\right|^2 \tag{5.17}$$

where

$$F\{X_T(t)\} = \int_{-T}^{T}X_T(t)e^{-j\omega t}dt \tag{5.18}$$

Therefore, over a finite span of data, the power spectral density of that data can be expressed using Equation 5.17.

5.2.3 Discrete Fourier Transforms

The derivation of the PSD approximation in Section 5.2.2 used the continuous Fourier transform. However, in signal processing, the data is usually discrete, and the derivation above must be adjusted for discrete data sets. If N is the number of samples in a batch of data, and g_k is the kth term of the sequence g, then the discrete Fourier transform of g is [4]

$$G_n = \frac{1}{N}\sum_{k=0}^{N-1} g_k e^{-j\frac{2\pi nk}{N}} \qquad n = 0, 1, \ldots, N-1 \quad (5.19)$$

with the inversion formula,

$$g_k = \sum_{n=0}^{N-1} G_n e^{j\frac{2\pi nk}{N}} \qquad k = 0, 1, \ldots, N-1 \quad (5.20)$$

In these equations, N is equal to $2T/\Delta t$. This approximation to the continuous Fourier transform may still be used in Equation 5.17 to determine the power spectral density function. If the value N is chosen to be a power of 2, then a Fast Fourier Transform (FFT) may be used [4]. The FFT greatly reduces the required number of calculations from N^2 to $N \log_2 N$. Numerous computer algorithms have been written for the FFT, and it will not be discussed further in this thesis.

5.2.4 Nyquist Criterion and Aliasing

The Nyquist criterion states that, if the sampling rate is $2f$ Hertz, then the only range in which frequency information will be available is 0 Hertz to f Hertz. In the discrete case, $2f$ is equal to $1/\Delta t$, which is the time interval between samples. If a frequency component is greater than f, then that frequency component will be aliased into the range 0 Hz to f Hz. For example, if the sampling rate is 100 Hz, a 1225 Hz signal would show up in the power spectral density function as a 25 Hz signal, and a 1985 Hz signal would show up as a 15 Hz signal.

If information is contained in a high frequency signal that is aliased because of the sampling rate, then that information is lost and cannot be reconstructed. However, if the high frequency signal is noise, then the fact that it is aliased by the sampling rate is not important. Whether the noise is aliased or not, it can be removed with a combination of sines and cosines at that frequency. The goal of this filter is to remove unwanted signals, not to determine the information in those signals.

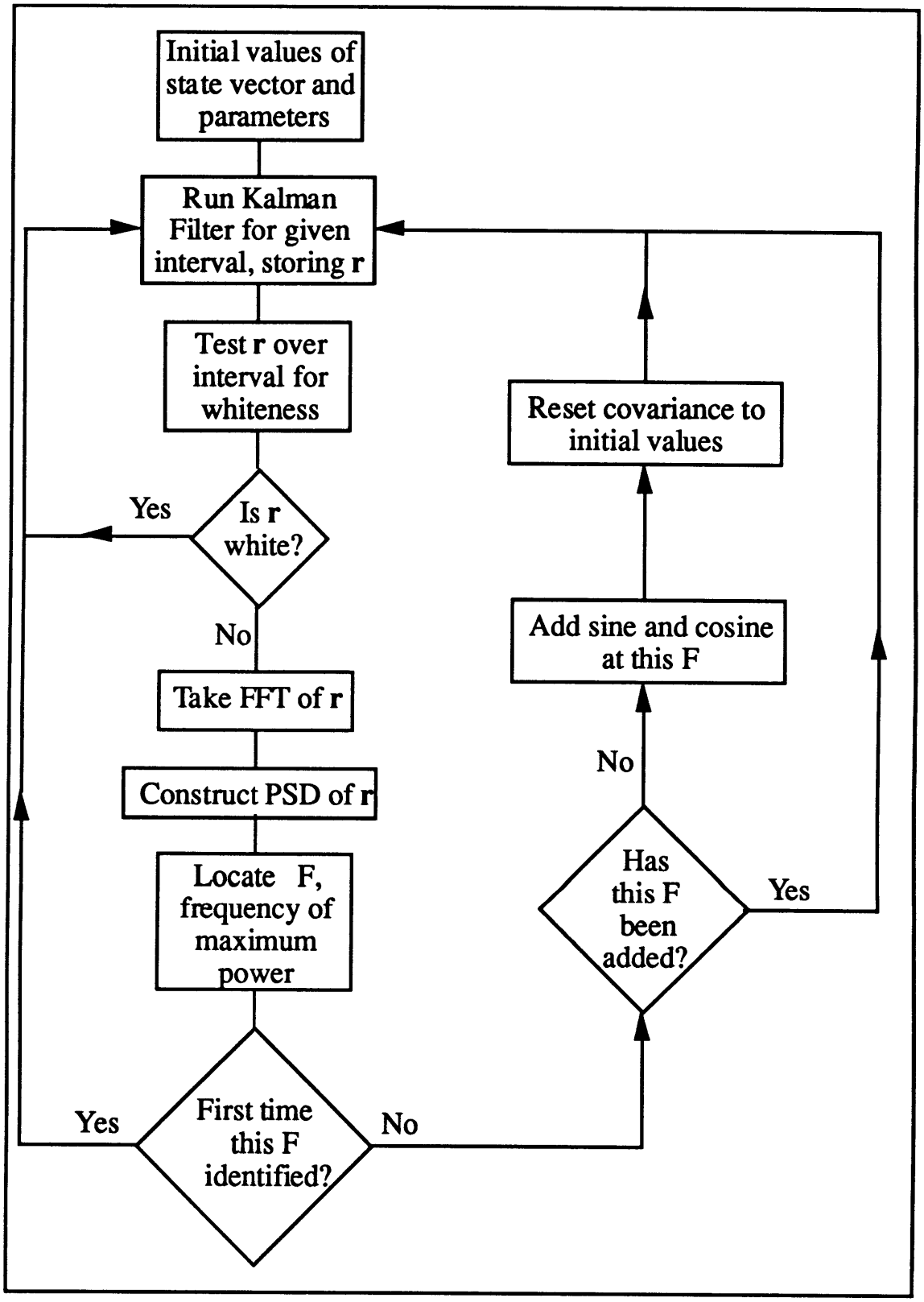

Figure 5.1. Flowchart for PSD Adaptive Filter

5.3 Filter Design

To implement the periodic noise identification filter using PSD methods, the following procedure, shown in Figure 5.1, was used. An estimation interval was established so that the innovation sequence could be examined over a given segment. Next, that segment of the innovation sequence was tested for whiteness. If the innovation sequence tested as white noise, then no changes were made to the system model. However, if the innovation sequence was colored noise, then a Fast Fourier Transform was taken of the segment of the innovation sequence. Next, the Power Spectral Density of the innovation sequence was constructed using Equation 5.17. Then, the location of maximum power was identified, and the frequency of this power concentration was noted. The filter then proceeded to the next segment of data. If this sequence tested as non-white noise, then the PSD was again constructed. If the frequency of maximum power was within a given tolerance of the previously identified frequency, then a sine and cosine term at that frequency were added to the filter. This two step procedure ensures that the innovation sequence is colored due to periodic noise, not for another reason, such as a poor system model or a bad data point. Also, the next frequency must be identified within two estimation intervals of the first frequency. If this criterion is not met, then the frequency will not be added to the filter.

5.4 Filter Verification

Several tests were run with simulated data to verify the performance of this filter. First, periodic signals were added to a random process at both a single frequency and at multiple frequencies. Next, the periodic signals were given a frequency uncertainty of 0.1%, which is the typical uncertainty in the natural frequency of an inertial element in a micromechanical gyroscope*. For all of these tests, control tests were run to demonstrate that the filter only adds frequencies when they exist, not all the time.

5.4.1 Single and Multiple Frequency Noise

Table 5.1 presents the results from these simulated data test runs. In each of the tests, the goal was to estimate the mean of the data while eliminating unwanted periodic noise. The size of the FFT was chosen to be 8192 for all tests. The sampling rate for each of these data sets was 50 Hz. The estimation interval for each of these tests was 100 points. Also, the standard deviation of the measurement noise was 0.1 units for all tests. The Noise Frequency column in Table 5.1 gives the actual value of the frequency, before it

* A more detailed discussion of the micromechanical gyroscope is given in Chapter 7.

was aliased. However, when the FFT size was chosen, the aliased frequency of the noise could not be identified exactly, and usually a frequency was identified slightly above and slightly below the actual aliased frequency, as the results in Table 5.1 show. In Figures 5.2 through 5.5, the innovation sequence is shown for various tests. For Tests 1 through 5, the frequency tolerance was 1%, and the white noise tolerance was 5%. For a frequency tolerance of 5%, a beating pattern appeared in the innovation sequence due to the mismatching between the true aliased frequency and the filter estimate of the aliased frequency.

Table 5.1. Results of Periodic Noise Identification

Test Number	Mean	Noise Freq.	Freq. ID (Hz)	Mean Estimate
5A1	3.2	253 Hz	2.9968, 2.9785	3.1913 ± 0.0055
5A2	3.2	Not Added	None ID	3.2005 ± 0.0033
5A3	3.2	253 Hz	2.9846, 3.0151, 2.9541	3.2049 ± 0.0057
		1027 Hz	23.0225, 22.9980	N/A
5A4	3.2	253 Hz	3.0090, 2.9541	3.1986 ± 0.0044
5A5	3.2	253 Hz	2.94, 2.97, 2.86	3.1962 ± 0.0050
		1027 Hz	22.9919	N/A

Each of these tests is discussed below. Test 5A2, however, is not discussed, because it was a control test. This test demonstrated that the adaptive filter will not add frequency terms when there is no periodic signal present in the measurement data.

In Test 5A1, the aliased frequency of 253 Hz was 3 Hz. This periodic noise signal had unity amplitude. The first frequency identified was 2.9968 Hz, and it was identified at 4 seconds. The second frequency, 2.9785 Hz, was identified at 14 seconds.

In Test 5A3, the aliased frequencies of 253 Hz and 1027 Hz were 3 Hz and 23 Hz, respectively. Both of these periodic noise signals had unity amplitudes. The first frequency 23.0225 Hz, was identified at 5 seconds; the second, 2.9846 Hz, at 10 seconds; the third frequency, 22.9980 Hz at 16 seconds; 3.0151 Hz at 22 seconds; and finally, 2.9541 Hz at 34 seconds. These improvements are easily seen in Figure 5.3; at the times listed, the peak to peak variation of the innovation sequence is reduced by the addition of these terms.

For Test 5A4, the amplitude of the periodic noise was reduced from 1 to 0.1. In Figure 5.4, the removal of periodic noise is not noticeable as a change in amplitude. The first frequency, 3.009 Hz, was added at 6 seconds, and the second, 2.9541 Hz, at 30 seconds.

For Test 5A5, the amplitudes of the two frequencies were reduced from 1 to 0.1. The first frequency, 2.9358 Hz, was added at 8 seconds, 2.9663 Hz at 14 seconds, 22.9919 Hz at 18 seconds, and 2.8625 Hz at 30 seconds.

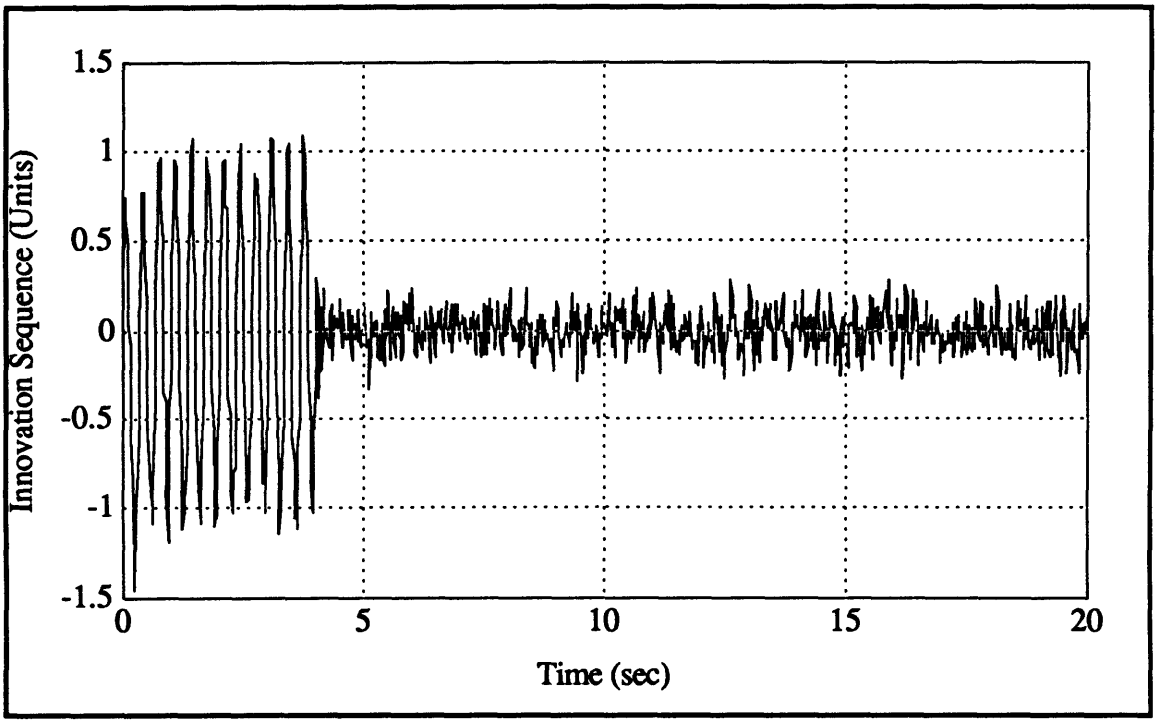

Figure 5.2. Innovation Sequence for Test 5A1

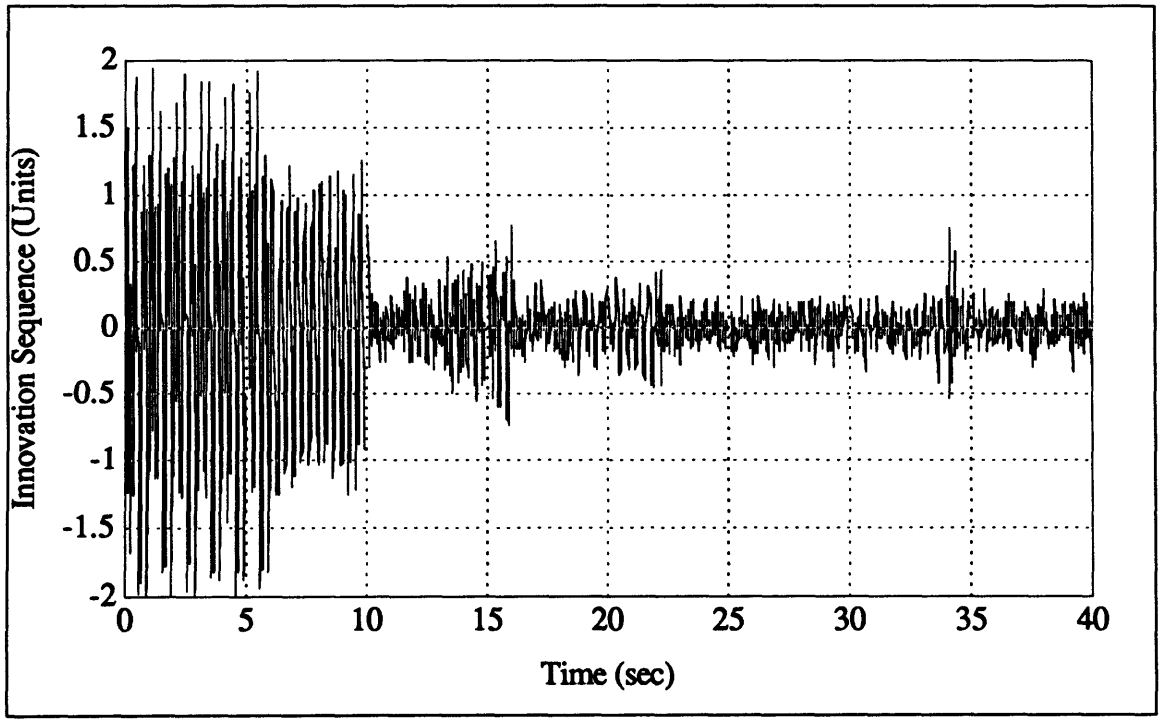

Figure 5.3. Innovation Sequence for Test 5A3

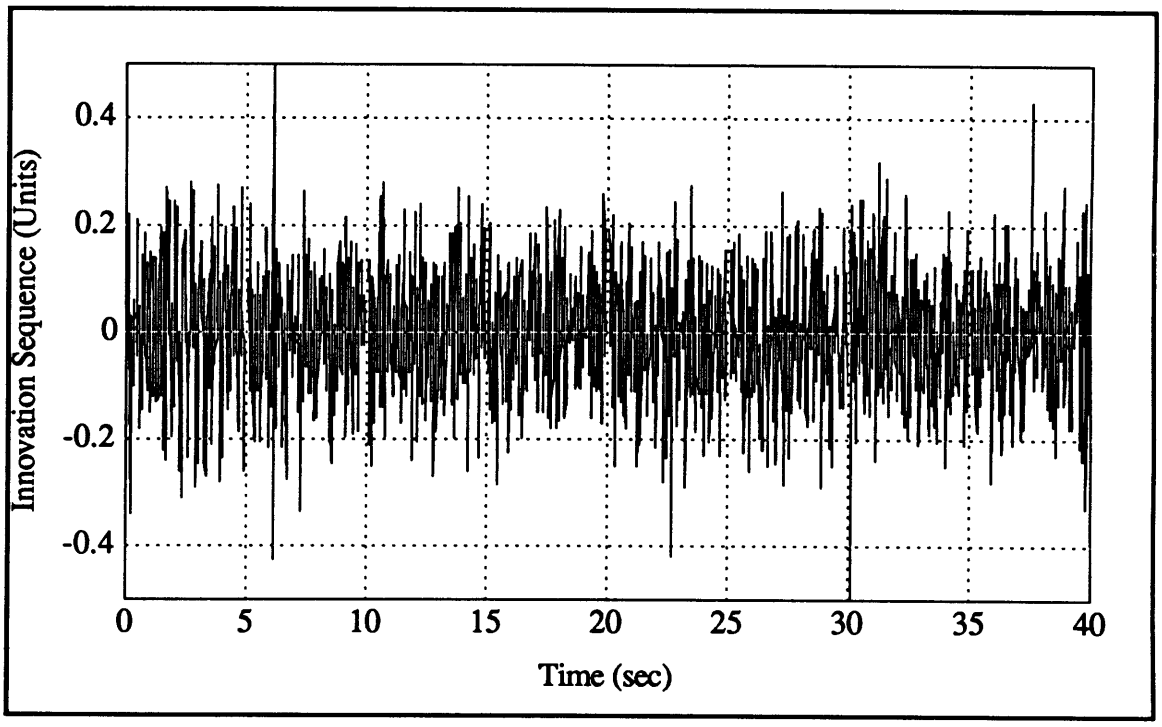

Figure 5.4. Innovation Sequence for Test 5A4

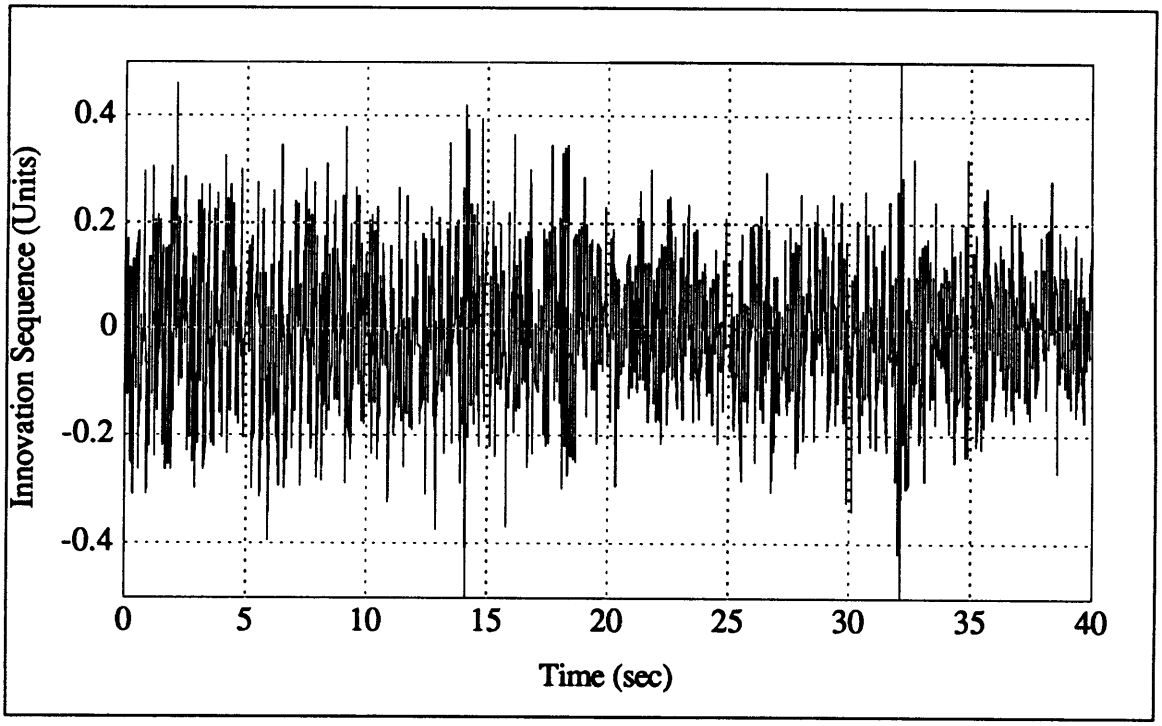

Figure 5.5. Innovation Sequence for Test 5A5

5.4.2 Distributed Frequency Noise

The purpose of this test was to determine the optimum sampling rate for identifying an aliased distributed high frequency noise signal. This signal may be present in the gyro output if a low pass filter does not adequately remove all high frequency components. This test demonstrated that a sampling rate of at least 1000 Hz will provide adequate information to remove an aliased high frequency signal of 4400 ± 4.4 Hz from the gyro output. In Chapter 8, real micromechanical gyroscope data is collected and analyzed using the conclusions of this simulation.

5.4.2.1 Motivation

One of the limiting factors in the performance of the micromechanical gyroscope is noise. Noise in the micromechanical gyros comes from measurement noise, random walk, amplification noise, and variation in the drive frequency. If any of these noise sources can be adequately modeled, then the characterization of the gyro performance will be enhanced.

In the micromechanical gyroscope*, the outer gimbal is driven at the resonant frequency of the inertial element (~2200 Hz). A rotation about the input axis will produce an output signal at this frequency. This output modulates a 100 kHz carrier signal, which, after amplification, is demodulated and passed through a low pass filter. The resulting signal at 2200 Hz is then demodulated at the resonant drive frequency. The two resulting frequencies are located at DC and at 4400 Hz. Only the DC signal passes through a baseband filter. This procedure is shown in Figure 7.2. However, if this low pass filter performs inadequately, then some of the 4400 Hz signal will be present in the gyro output.

If the resonant drive frequency was at exactly at 2200 Hz, then identification of any 4400 Hz noise could be accomplished at a low sampling rate because all the spectral power would be concentrated at 4400 Hz. However, the drive frequency may be mismatched with the resonant frequency, whose uncertainty is about 0.1%. Therefore, the sampling rate must be higher in order to recover enough of this information to determine the presence of poorly filtered high frequency noise. The spectral energy of the high frequency signal is distributed about 4400 Hz, not concentrated at 4400 Hz. As the signal is aliased, the distribution of the frequency becomes less recognizable, until the aliased signal has a completely random PSD.

Additionally, the filter designed for frequency location does not attempt to include a sinusoidal noise model unless the innovation sequence is colored noise. If the innovation

* A more detailed discussion of the micromechanical gyroscope is given in Chapter 7.

sequence tests as white noise, then no changes are made in the system model. Therefore, the sampling rate must be sufficiently high so that the innovation sequence will be colored noise if high frequency periodic noise is present.

These two requirements stipulate that the sampling rate must be sufficiently high so that the innovation sequence of the Kalman filter tests as colored noise and that the frequency detected by the FFT corresponds to the aliased value of the drive frequency.

5.4.2.2 Approach

A test plan was designed to determine the optimum sampling rate for high frequency noise detection. The frequency identification approach presented previously was implemented with a Kalman filter. The following test procedure was conducted for 1000 runs at each of 14 different sampling rates. First, a sinusoidal data set was generated using Matlab™ for the Macintosh [30]. The frequency for each point was determined by the *rand* function in a normal distribution mode. If the frequency exceeded the 0.1% tolerance, the frequency was recalculated using the *rand* function, until the value was within the 0.1% tolerance.

Once the data set was generated, it was run through a single state Kalman filter. When the filter was finished, a test was conducted to determine whether or not the innovation sequence was white noise. A count was made of the number of non-white sequences from each of the 1000 runs. Similarly, an FFT was taken of the innovation sequence, and the identified frequency from the analysis of the FFT was compared to the known value of 4400 Hz. If the identified frequency was within 2.5% of the expected value of 4400 Hz, the frequency was considered identified. A count was also made of the number of successful frequency identifications. While this approach is slightly different than that discussed in Section 5.4.1, the idea is the same: first, the data being evaluated must test as white noise; then the frequency must be identified.

5.4.2.3 Results

After completing 1000 runs at each of the 14 sampling rates, a plot (Figure 5.6) was generated to show, based on the simulation, the probability of detecting the correct frequency (solid line) and of identifying the innovation sequence as colored noise versus the sampling rate (dashed line). Also, Table 5.2 was created to show the same results. The figure clearly shows that a sampling rate of about 1000 Hz is needed in order to properly identify the aliased frequency noise. Figure 5.6 also shows that the probability of the innovation sequence testing as non-white noise was the driving factor in the choice of an optimal sampling rate. The values in Figure 5.6 are presented in Table 5.2.

Table 5.2. Probabilities for Aliased 4400 Hz Frequency Signal

Sampling Rate	P(Freq. ID)	P(colored noise)
280.74	0.2250	0.0250
328.33	0.3960	0.0350
383.99	0.6040	0.0700
449.09	0.4970	0.0970
525.22	0.8100	0.2290
614.26	0.7150	0.4620
718.38	0.7470	0.7340
840.17	0.9950	0.9530
982.59	1.0000	1.0000
1149.2	1.0000	1.0000
1344.0	1.0000	1.0000
1571.8	1.0000	1.0000
1838.3	1.0000	1.0000

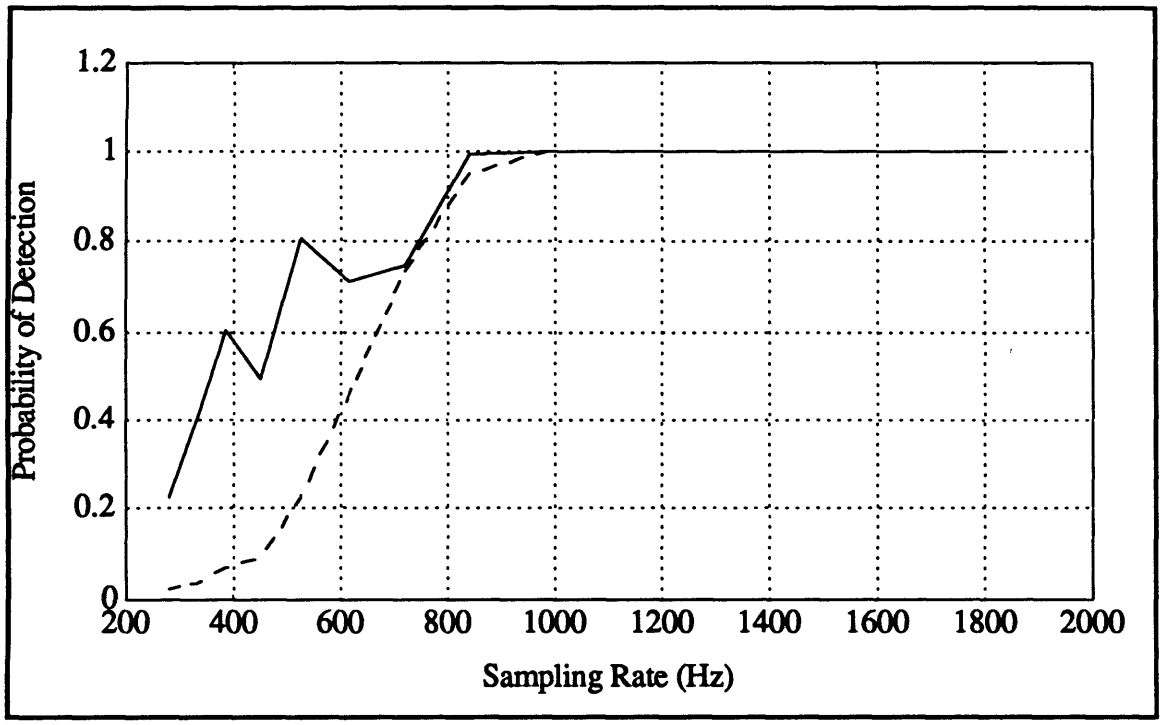

Figure 5.6. Probability of Detecting Frequency and Noise

Figure 5.7 shows the aliased frequency as a function of the sampling rate (solid line), and also shows the aliased frequencies that were identified in this simulation (dashed

line). The sawtooth pattern in the aliased frequency is the probable cause of the variation in frequency identification shown in Figure 5.6. A larger identification tolerance (~ 5%) would reduce this variation.

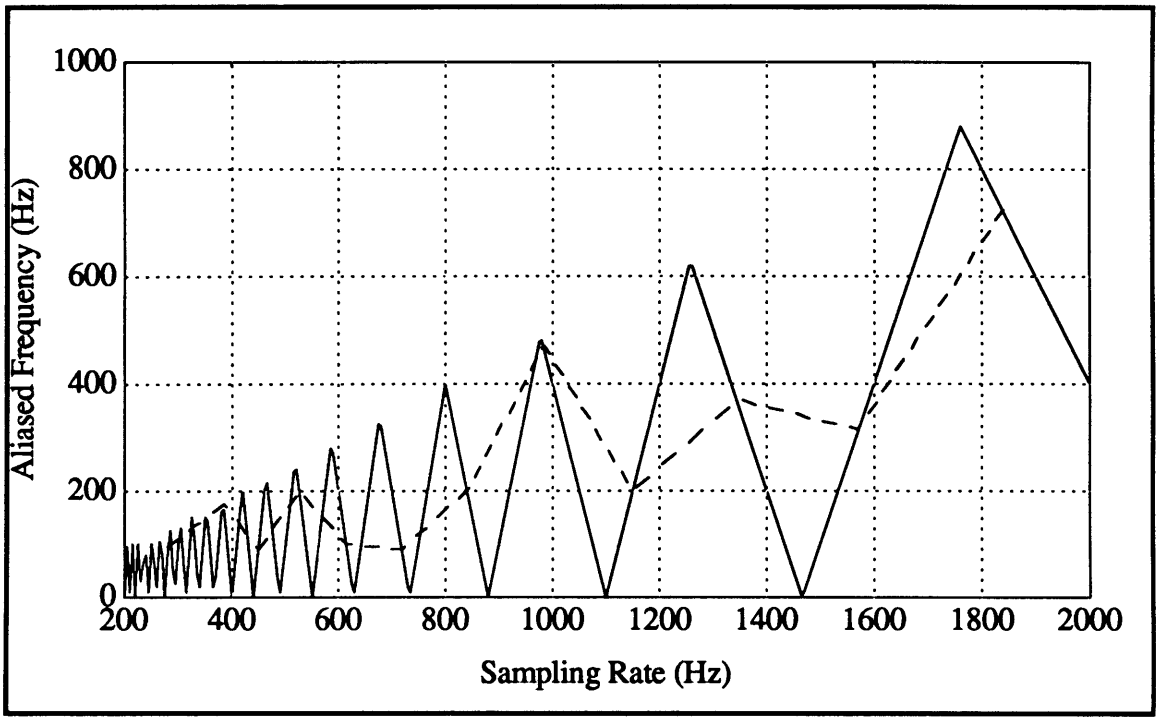

Figure 5.7. Aliased Frequency vs. Sampling Rate for 4400 Hz Signal

Additional tests were run for high frequency signals of 2200 Hz and 6600 Hz. The analysis methodology was identical to the 4400 Hz analysis. However, at each frequency, only 250 test runs were made. The results are shown below in Figures 5.8 and 5.9 and Tables 5.3 and 5.4. In each plot, the solid line represents the probability of correct frequency identification to within 2.5%, and the dotted line represents the probability of the innovation sequence testing as colored noise.

Table 5.3. Probabilities for Aliased 2200 Hz Frequency Signal

Sampling Rate	P(Freq. ID)	P(colored noise)
280.74	0.6560	0.3280
328.33	0.9280	0.6040
383.99	0.9920	0.8840
449.09	0.8840	0.9680
525.22	1.0000	1.0000
614.26	1.0000	1.0000
718.38	0.9200	1.0000
840.17	1.0000	1.0000
982.59	1.0000	1.0000
1149.2	1.0000	1.0000
1344.0	1.0000	1.0000
1571.8	1.0000	1.0000
1838.3	1.0000	1.0000

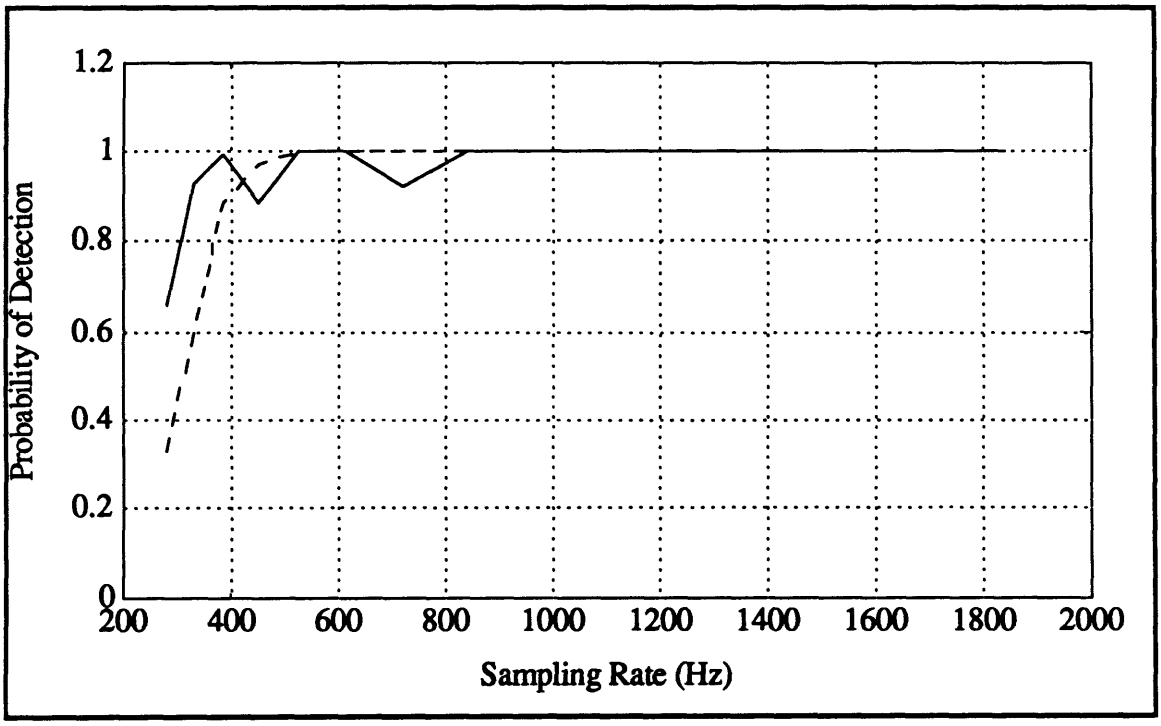

Figure 5.8. Probabilities for 2200 Hz Signal

Table 5.4. Probabilities for Aliased 6600 Hz Frequency Signal

Sampling Rate	P(Freq. ID)	P(colored noise)
280.74	0.1840	0.0200
328.33	0.0440	0.0040
383.99	0.1440	0.0160
449.09	0.3360	0.0200
525.22	0.4880	0.0440
614.26	0.5640	0.0840
718.38	0.6040	0.1800
840.17	0.6000	0.3400
982.59	0.9280	0.6400
1149.2	0.9840	0.8520
1344.0	0.7400	0.9760
1571.8	1.0000	1.0000
1838.3	1.0000	1.0000

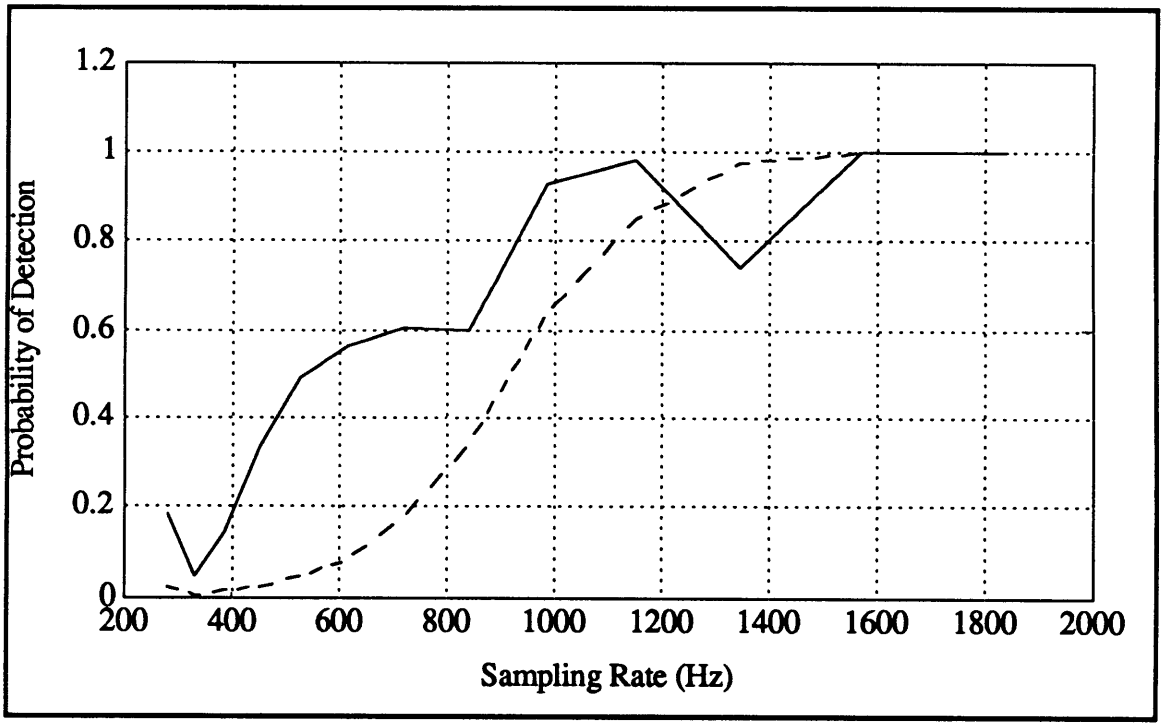

Figure 5.9. Probabilities for 6600 Hz Signal

These figures and tables show that the required sampling rate varies as the high frequency signal varies. Variations in the probabilities may also be linked to the estimation interval over which the white noise and frequency identification tests are applied.

5.5 Conclusions

The power spectral density approach is effective in identifying both concentrated and distributed frequency signals. This filter only considers adding periodic noise sources when the innovation sequence indicates that such model enhancement is necessary. Both the raw data reduction and reduced data analysis filters incorporate the PSD analysis.

Chapter 6

Adaptive Data Analysis Filter Development

6.1 Introduction

In this chapter, two adaptive filters are created: one for raw data reduction and one for reduced data analysis. Each of the three filter concepts developed and verified in Chapter 3 through Chapter 5 have been combined to produce these two filters. By implementing these two adaptive filters, micromechanical inertial instrument performance can be improved.

6.2 Raw Data Reduction Adaptive Filter

For raw data reduction, data points are averaged over a time interval, and the size of the data set is decimated. For example, a 100 Hz data set is averaged and decimated to a 1 Hz data set. Each decimated data point is a representation of data near that point, i.e., a mean or a weighted mean. Therefore, the purpose of a raw data reduction filter is to estimate a mean for a given set of data. By eliminating unwanted periodic noise from the data and by determining an accurate estimate of the unknown system parameters, the decimated data will be a more accurate representation of the raw data. The adaptive filter methods used for the raw data reduction adaptive filter are the maximum likelihood estimator and the PSD analysis. At the end of an estimation interval, the estimate of the mean is recorded as a decimated data point, and the state estimate of the mean and the covariance of this state estimate are reset to their initial values so that all decimated data points are independent.

6.2.1 Verification of Raw Data Reduction Filter

To verify this combination of filter concepts, simulations were run under various conditions. The simulated data sets were a combination of periodic noise signals and

unknown measurement noise standard deviations. The results for two runs, Test 6A1 and Test 6A2, are listed below in Table 6.1. For each of these tests, the simulated data set was over a 40 second time span at a sampling rate of 50 Hz, for a total of 2000 data points. In Test 6A1, a 3 Hz sine wave with an amplitude of 1 unit was added to a zero mean random process with a measurement noise σ of about 0.10 units. In Test 6A2, unity amplitude sine waves at both 3 Hz and 17 Hz were added to a zero mean random process with a measurement noise σ of about 0.10 units. The estimation interval for both tests was 2 seconds (100 points).

Table 6.1 shows that the filter was successful in identifying both the true value of the measurement noise standard deviation and the frequency of the periodic noise. As discussed in Chapter 5, the PSD analysis filter is not always able to exactly identify the frequency of the periodic noise because of the size of the Fast Fourier Transform. Table 6.1 lists all of the frequencies identified for each of the two tests.

Table 6.1. Results of Raw Data Reduction Filter Verification

Test Number	Value Estimated	Actual Value	Initial Estimate	Final Estimate
6A1	Measurement Noise	0.1001	1.0000	0.0969 ± 0.0069
	Frequency	3 Hz	None	2.991, 3.027 Hz
6A2	Measurement Noise	0.1019	1.0000	0.0931 ± 0.0073
	Frequency 1	3 Hz	None	2.997, 3.009 Hz
	Frequency 2	17 Hz	None	16.9983 Hz

Figures 6.1 and 6.2 show the innovation sequence for both tests. For Test 6A1, shown in Figure 6.1, the periodic noise was identified at the first possible opportunity, i.e., 4 seconds, because two sequential identifications of a given frequency are needed. Once identified, most of the periodic noise component was quickly removed from the data. Both this improvement in the innovation sequence and the proper estimate of the measurement noise covariance demonstrate the success of the raw data reduction adaptive filter.

For Test 6A2, shown in Figure 6.2, the first frequency component, 17 Hz, was also identified at 4 seconds. The second frequency component, 3 Hz, was first identified at 8 seconds. With a more accurate noise model, the improvement in the innovation sequence is quite significant, as Figure 6.2 shows. The data is reduced from a peak to peak variation of 4 units to a peak to peak variation of about 0.2 units, which is the expected variation due to the measurement noise.

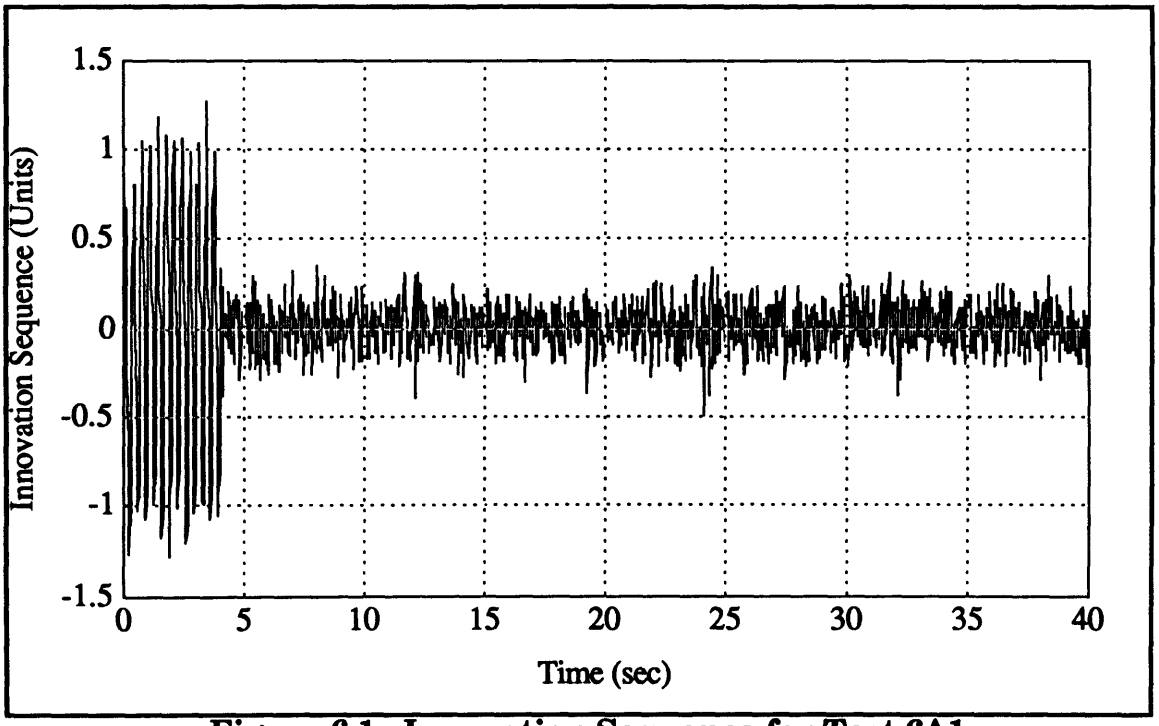

Figure 6.1. Innovation Sequence for Test 6A1

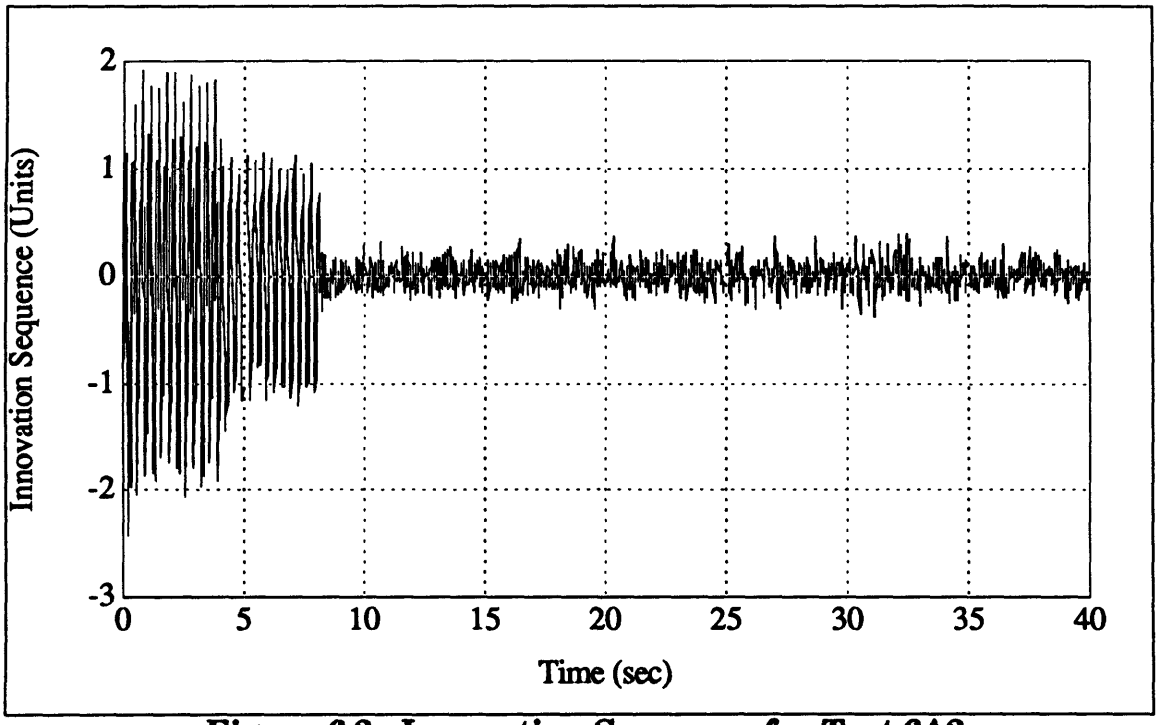

Figure 6.2. Innovation Sequence for Test 6A2

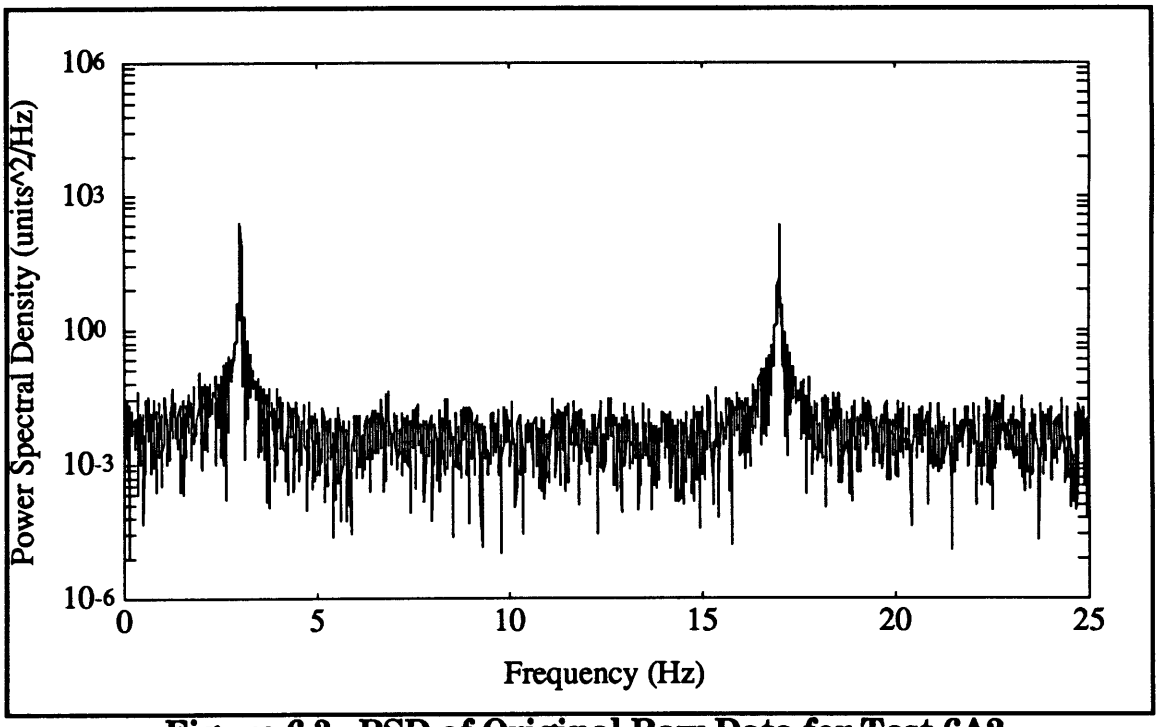

Figure 6.3. PSD of Original Raw Data for Test 6A2

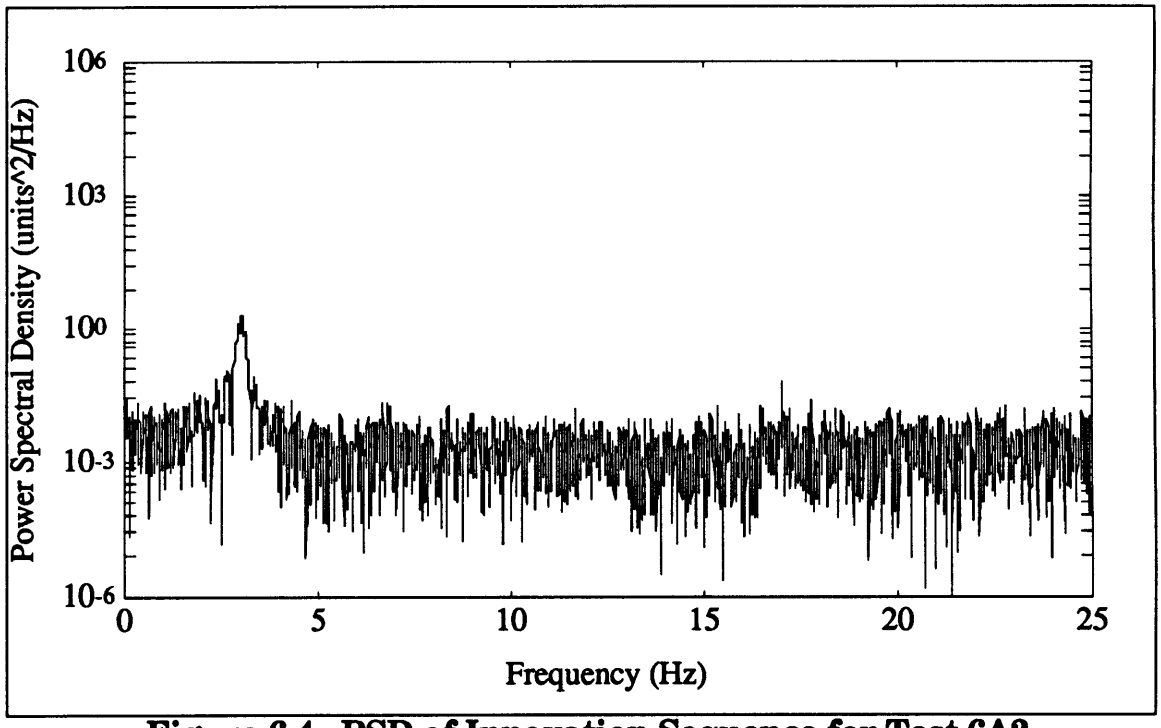

Figure 6.4. PSD of Innovation Sequence for Test 6A2

Figure 6.3 and 6.4 are a comparison of the power spectral density of the original data to the power spectral density of the innovation sequence for Test 6A2. Figure 6.4 represents the PSD of the innovation sequence after both frequency terms have been

removed, i.e. the innovation sequence from 8 seconds to 40 seconds. The frequency concentrations at 3 Hz and at 17 Hz in Figure 6.3 do not exist in Figure 6.4 because the raw data reduction filter identified and removed these frequency components. In this portion of the innovation sequence, small concentrations at both 3 Hz and 17 Hz exist, but these amounts are three orders of magnitude less than that in the original data.

6.2.2 Raw Data Reduction Adaptive Filter vs. Triangular Filter

Once these tests were run to verify the operation of the filter, simulation tests were run to determine any improvement in data decimation using this raw data reduction adaptive filter over a triangular filter. Simulated 100 Hz data sets were generated over 100 seconds for various cases. These data sets were then analyzed using both the raw data reduction adaptive filter and a triangular filter. In data decimation, the purpose is to provide estimates of the mean of a segment of data over a given interval. Therefore, the raw data reduction adaptive filter must "re-estimate" the mean over each interval. This requires that, after the estimate of the mean is recorded, the values of the state vector and its covariance are reset to their initial values and allowed to converge again over the next estimation interval. However, the noise estimates, as well as the frequency estimates, will be allowed to converge over the entire data set. This is permissible because it is assumed that the noise sources, such as the frequency of the periodic noise, are nearly constant throughout the data acquisition.

The comparison of these filters was based on equal bandwidths. By comparing bandwidths, each filter makes estimates based on the same number of data points, and an accurate picture of the behavior of each filter can be drawn. In a triangular filter, a weighted average of data is taken over a certain number of points [22]. For equal bandwidth, the full size of the triangular filter must be equal to the estimation interval of the raw data reduction adaptive filter. The frequency of the bandwidth is equal to one-half the sampling frequency divided by the number of points used in the interval.

For the first test, Test 6B1, a periodic signal of frequency 1206.25 Hz with an amplitude of 0.5 units was added to a zero mean white noise source with a standard deviation of 0.10 units. Figure 6.5 is a semilog plot of the standard deviation of the decimated data versus the bandwidth for both the triangular and adaptive filters. The solid line represents the log of the standard deviation of the raw data reduction filter, and the dashed line represents the log of the standard deviation for the triangular filter. As Figure 6.5 shows, the raw data reduction filter performs better than the triangular filter. At the higher bandwidths, the improvement in the data is visibly obvious, about 33 percent at a 2 Hz bandwidth. At the lower bandwidths, the triangular filter performance

approaches that of the raw data reduction filter because of the attenuation of the periodic noise by the triangular filter.

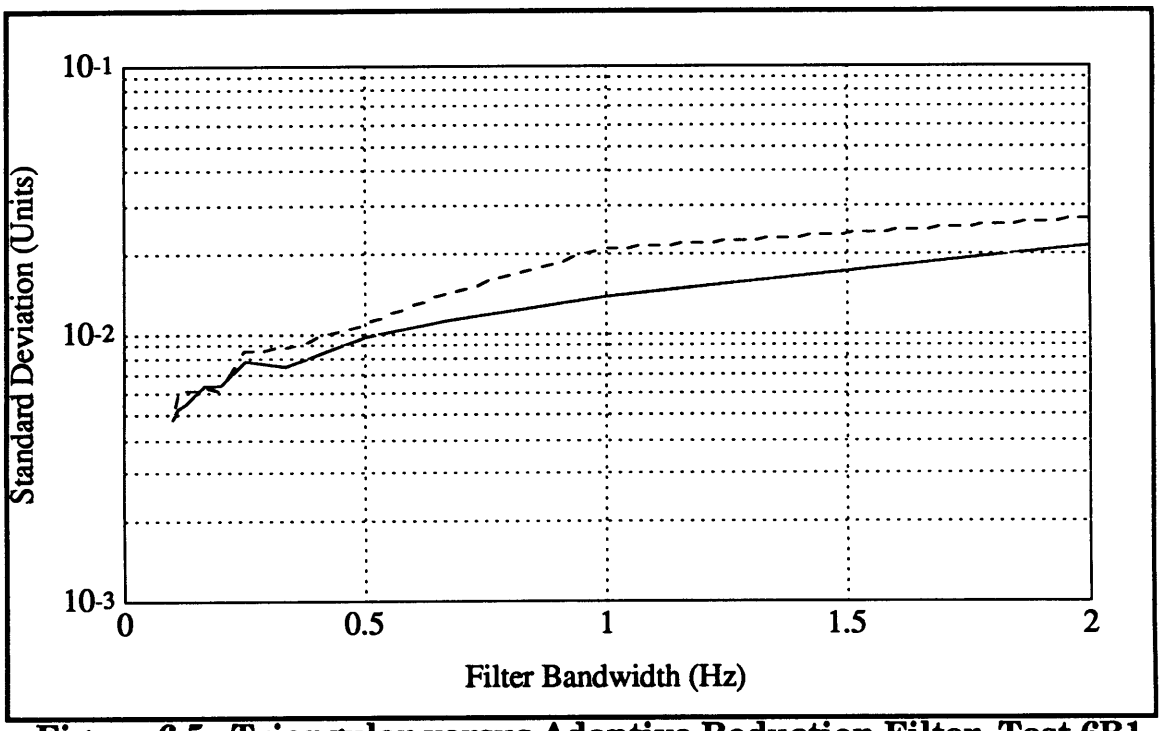

Figure 6.5. Triangular versus Adaptive Reduction Filter, Test 6B1

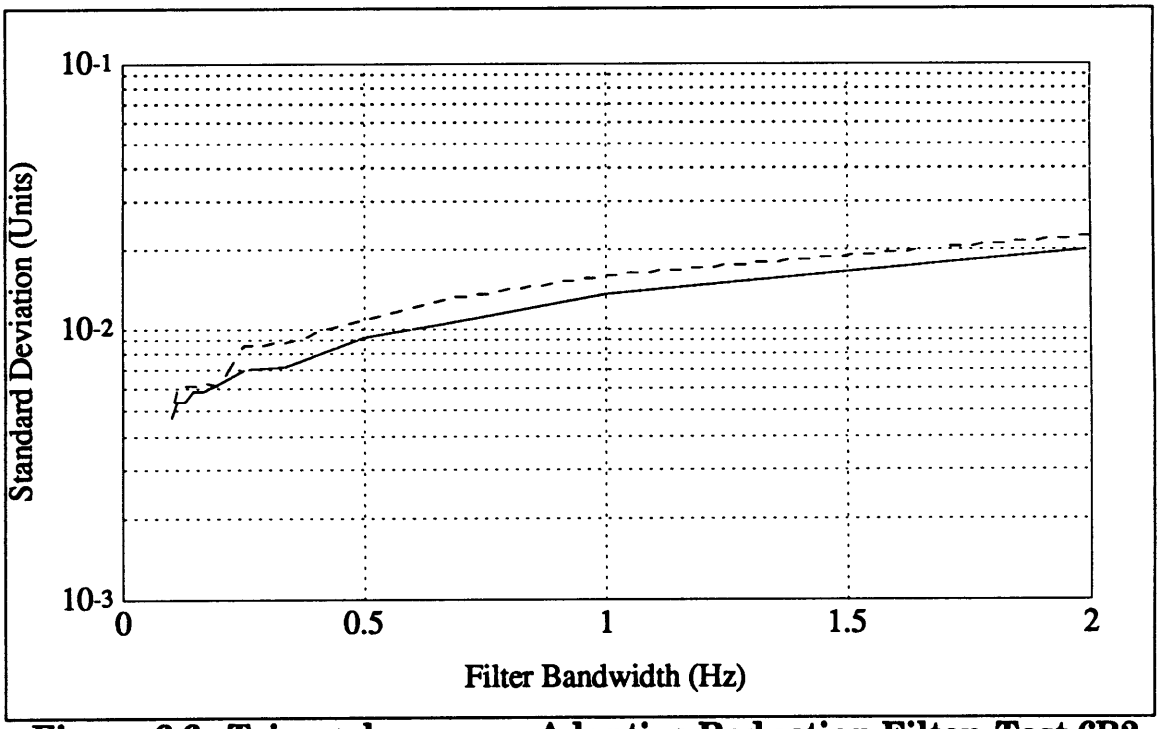

Figure 6.6. Triangular versus Adaptive Reduction Filter, Test 6B2

Next, a data set without periodic noise was generated for a white noise source with a standard deviation of 0.10 units and a mean of 1.42 units, Test 6B2. A comparison of the standard deviations for this test are shown in Figure 6.6. Once again, the solid line represents the log of the standard deviation of the raw data reduction filter, and the dashed line represents the log of the standard deviation for the triangular filter. In this test, the raw data reduction adaptive filter gave a lower standard deviation of the decimated data than the triangular filter did, about 10 percent at 2 Hz bandwidth.

Figures 6.5 and 6.6 show that the raw data reduction filter is more effective than the triangular filter, whether or not a periodic signal exists. These two tests demonstrate that the raw data reduction filter is more effective than a triangular filter for data decimation.

6.3 Reduced Data Analysis Adaptive Filter

Once the data has been prepared for analysis, e.g. data decimation, instrument performance must be characterized according to a system model. Traditional analysis has used a Kalman filter with a prescribed system model. Because of the rigid system model, this method cannot account for an inaccurate description of the noise, unmodeled system behavior, and unknown periodic noise. By combining the three adaptive filter concepts discussed in Chapters 3 through 5, a more powerful adaptive Kalman filter has been developed. All three concepts are included so that, if the raw data reduction adaptive filter is not needed, the reduced data analysis adaptive filter can determine the system parameters as well as identify periodic noise.

One modification made in the combination of these filters is addition of a fault tolerant algorithm. A faulty commanded rate test encouraged the development of this fault tolerant algorithm. This commanded rate test is discussed in Section 8.3, and the fault tolerant algorithm is presented in this chapter.

6.3.1 Fault Tolerant Algorithm

A fault tolerant algorithm can detect unexpected jumps in the gyro output by examining the innovation sequence. The variable ε is defined in Equation 6.1 as

$$\varepsilon = \sum_{j=1}^{N} \mathbf{r}_j^{\mathrm{T}} \mathbf{S}_j^{-1} \mathbf{r}_j \tag{6.1}$$

where N is the length of the estimation interval, $\mathbf{r}$ is the innovation sequence and $\mathbf{S}$ is the covariance of the innovation sequence. This variable ε is a chi-squared distribution random variable with $m \times N$ degrees of freedom, where m is the dimension of the

innovation vector **r**. For order greater than 30, the chi-squared variable can be approximated by a normal distribution.

For each estimation interval, the variable ε is computed, and the mean, which is equal to the degrees of freedom, is subtracted from the value. This value is then compared to a confidence limit, which is equal to some number times the standard deviation. For this variable, the standard deviation is equal to

$$\sigma_\varepsilon = \sqrt{2mN} \tag{6.1}$$

If the value of ε exceeds this confidence limit, then the covariance matrix **P** is reset to its initial values, and the state vector **x** is allowed to converge to new values.

6.3.2 Verification of Reduced Data Analysis Adaptive Filter

Before this adaptive Kalman filter can be compared with a traditional Kalman filter, the proper behavior of this filter must be verified. Once again, simulated data sets were generated for a variety of conditions. Table 6.2 shows the results from these filter verification tests. For Tests 6C1 through 6C4, the data was generated at 100 Hz over a 45 second span. The estimation interval for each test was 100 points, or 1 second. In all of these tests, the maximum likelihood estimator determined the measurement noise standard deviation. Random walk was not studied in these tests. Also, the data analyzed here is not decimated; there is no difference in the performance of the filter whether the data set is decimated or not.

In the first test, Test 6C1, the unmodeled term was a rate squared term, using a commanded rate profile similar to that discussed in Section 4.4 for the rate squared test. In Test 6C2, trend was the unmodeled term. For Test 6C3, both trend and a periodic noise component were added to the data set. In Test 6C4, only the measurement noise standard deviation, bias and scale factor were estimated. Although these data files were not decimated before analysis, no differences in operation would occur in decimated data analysis.

In Test 6C1, the rate squared term was successfully added to the system model. Once added, the filter correctly determined the true values of the bias, scale factor, and rate squared terms, as well as the measurement noise σ. In Test 6C2, the trend term was added, and again all noise parameters were determined accurately. In Test 6C3, the filter identified the trend, as well as the periodic noise component. The filter located the correct aliased frequency of 6.25 Hz, and determined that most of the periodic noise was at this frequency. Therefore, only the sine and cosine estimates at 6.25 Hz are presented in Table 6.2. The estimates of the other identified frequency, 6.2378 Hz, were of the

order 10^{-3}, and the covariances of these estimates were also of this magnitude. For Test 6C4, the filter accurately estimated the proper values for the bias, scale factor and measurement noise standard deviation. Plots of these tests are not included, but would appear similar to those presented in Chapter 4.

Table 6.2. Results of Reduced Data Analysis Filter Verification

Test	True Model	Actual Values	Initial Values	Final Estimates
6C1	Bias	1.1000	0.0000	1.0975 ± 1.96e-3
	Scale Factor	0.0150	0.0000	0.0151 ± 4.36e-5
	Rate Squared	1.200e-4	Not Modeled	1.205e-4 ± 6.9e-7
	Measurement Noise	0.09957	1.0000	0.1006 ± 0.0065
6C2	Bias	1.1000	0.0000	1.1082 ± 0.0067
	Scale Factor	0.0150	0.0000	0.0150 ± 6.20e-5
	Trend	0.0200	Not Modeled	0.0197 ± 2.43e-4
	Measurement Noise	0.09957	1.000	0.0999 ± 0.0065
6C3	Bias	1.1000	0.0000	1.1201 ± 0.0154
	Scale Factor	0.0150	0.0000	0.0150 ± 6.5e-5
	Trend	0.0200	Not Modeled	0.0193 ± 4.49e-4
	Sine - 1206.25 Hz	0.2500	Not Modeled	0.2201 ± 0.0079
	Cosine - 1206.25 Hz	Not Modeled	Not Modeled	-0.0923 ± 0.0079
	Measurement Noise	0.09957	1.0000	0.09907 ± 0.0065
6C4	Bias	1.1000	0.0000	1.0984 ± 0.0012
	Scale Factor	0.0150	0.0000	0.0150 ± 2.81e-5
	Measurement Noise	0.07468	1.0000	0.0754 ± 0.0049

6.3.3 Comparison of Reduced Data Analysis Filter with Kalman Filter

Once the behavior of the reduced data analysis filter was verified, a comparison was made with the performance of a traditional Kalman filter. Each of the data sets in Table 6.2 was also analyzed by a Kalman filter, and a comparison of the performances for each filter are presented in Table 6.3. The four tests examined the performance of each filter for unmodeled trend terms, unmodeled periodic noise, unmodeled rate squared terms, and an inadequate measurement noise model.

For Test 6C1, the reduced data analysis filter performed significantly better than the Kalman filter. With the adaptive filter, the bias estimate is closer to the true value, and the innovation sequence RMS for the adaptive filter is half the RMS of the Kalman filter.

Also, the rate squared term was identified in the adaptive filter; it was never modeled in the Kalman filter. If the *a posteriori* residual was calculated, the adaptive filter RMS would be significantly lower than that of the Kalman filter. The *a posteriori* residual is the difference between the measured and expected values at each data point, based on the final state vector estimates.

For Test 6C2, the trend term is unmodeled in the Kalman filter. Again, the innovation sequence RMS is higher for the Kalman filter. The uncertainty of the bias estimate in the Kalman filter is about 0.0015. This uncertainty means that the estimate cannot converge more rapidly to the correct estimate than this value. In fact, the plot of the bias estimate in Figure 6.7 shows that the bias estimate was accounting for the trend term. However, in Figure 6.8, the adaptive filter bias estimate is unaffected by the trend in the data, once the final system model had been determined. The adaptive filter is a better filter not only because it removes the time dependence of the estimates by modeling for trend, but also because it determines the true model of the system.

In Test 6C3, neither the trend nor the periodic noise are modeled by the Kalman filter. This mismodeling of the state vector in the Kalman filter results in poor estimates of both the bias and the scale factor. While the adaptive filter estimates for all the state parameters are accurate to within a standard deviation, the Kalman filter estimates are several standard deviations away from the true system values. Also, for the Kalman filter, the innovation sequence RMS is twice as large as the RMS for the adaptive filter.

Inadequate information about the measurement noise standard deviation was examined in Test 6C4. In this test, the Kalman filter performed as well as the reduced data analysis filter. This test demonstrated that the reduced data analysis filter provides performance at least equal to that of a Kalman filter.

Table 6.3 Kalman Filter versus Reduced Data Analysis Filter

Test	Actual Model		Kalman Filter		Reduced Data Analysis Filter	
#	Terms	Values	Initial Values	Estimates	Initial Values	Estimates
6C1	Bias	1.1000	0.0000	1.3495 ± 1.491e-3	0.0000	1.0975 ± 1.96e-3
	Scale Factor	0.0150	0.0000	0.0150 ± 3.26e-5	0.0000	0.0151 ± 4.36e-5
	Rate Squared	1.200e-4	Not Modeled	Not Modeled	Not Modeled	1.205e-4 ± 6.9e-7
	Msmt. Noise	0.09957	0.1000*	Not Estimated	1.0000	0.1006 ± 0.0065
	Innov. Seq. RMS	—	—	0.4142	—	0.2110
6C2	Bias	1.1000	0.0000	1.5484 ± 1.491e-3	0.0000	1.1082 ± 0.0067
	Scale Factor	0.0150	0.0000	0.0109 ± 3.25e-5	0.0000	0.0150 ± 6.20e-5
	Trend	0.0200	Not Modeled	Not Modeled	Not Modeled	0.0197 ± 2.43e-4
	Msmt. Noise	0.09957	0.1000	Not Estimated	1.000	0.0999 ± 0.0065
	Innov. Seq. RMS	—	—	0.2073	—	0.1068
6C3	Bias	1.1000	0.0000	1.5486 ± 1.49e-3	0.0000	1.1201 ± 0.0154
	Scale Factor	0.0150	0.0000	0.0109 ± 3.25e-5	0.0000	0.0150 ± 6.5e-5
	Trend	0.0200	Not Modeled	Not Modeled	Not Modeled	0.0193 ± 4.49e-4
	Sine - 1206.25 Hz	0.2500	Not Modeled	Not Modeled	Not Modeled	0.2201 ± 0.0079
	Msmt. Noise	0.09957	0.1000	Not Estimated	1.0000	0.0993 ± 0.0065
	Innov. Seq. RMS	—	—	0.2734	—	0.1224
6C4	Bias	1.1000	0.0000	1.0989 ± 0.0015	0.0000	1.0984 ± 0.0012
	Scale Factor	0.0150	0.0000	0.0150 ± 3.25e-5	0.0000	0.0150 ± 2.81e-5
	Msmt. Noise	0.07468	0.1000	Not Estimated	1.0000	0.0754 ± 0.0049
	Innov. Seq. RMS	—	—	0.0800	—	0.0801

*This is a reasonable estimate of the noise of the instrument, based on previous analysis.

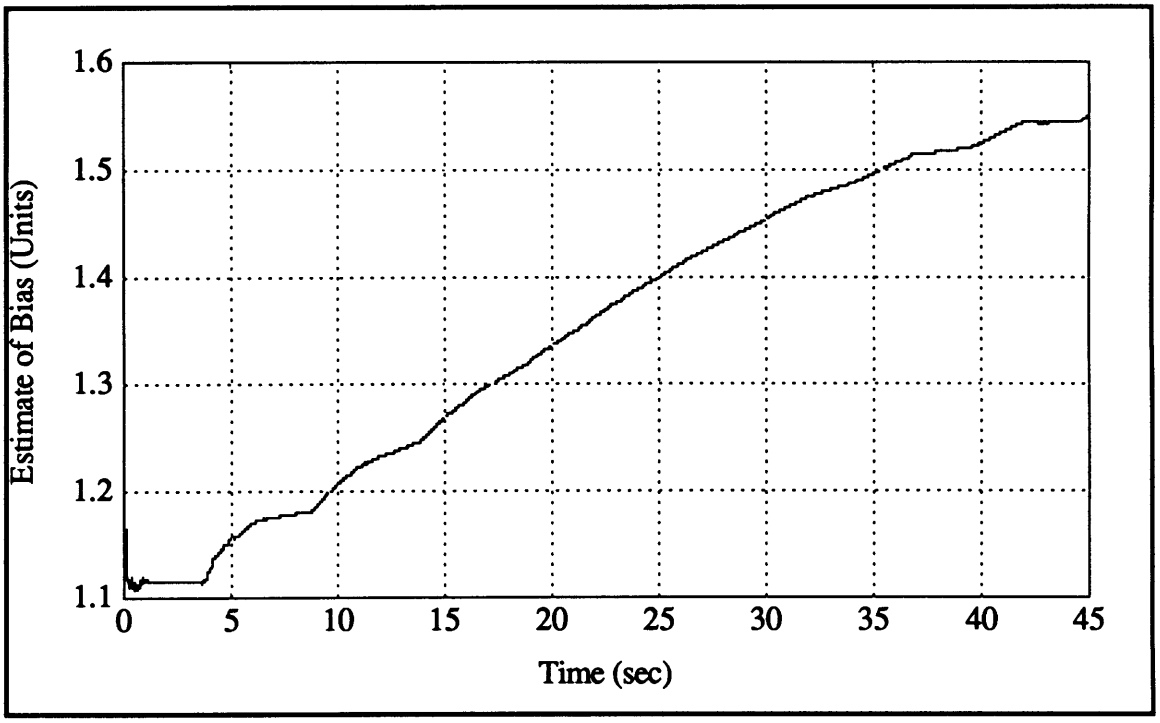

Figure 6.7. Bias Estimate versus Time for Kalman Filter, Test 6C2

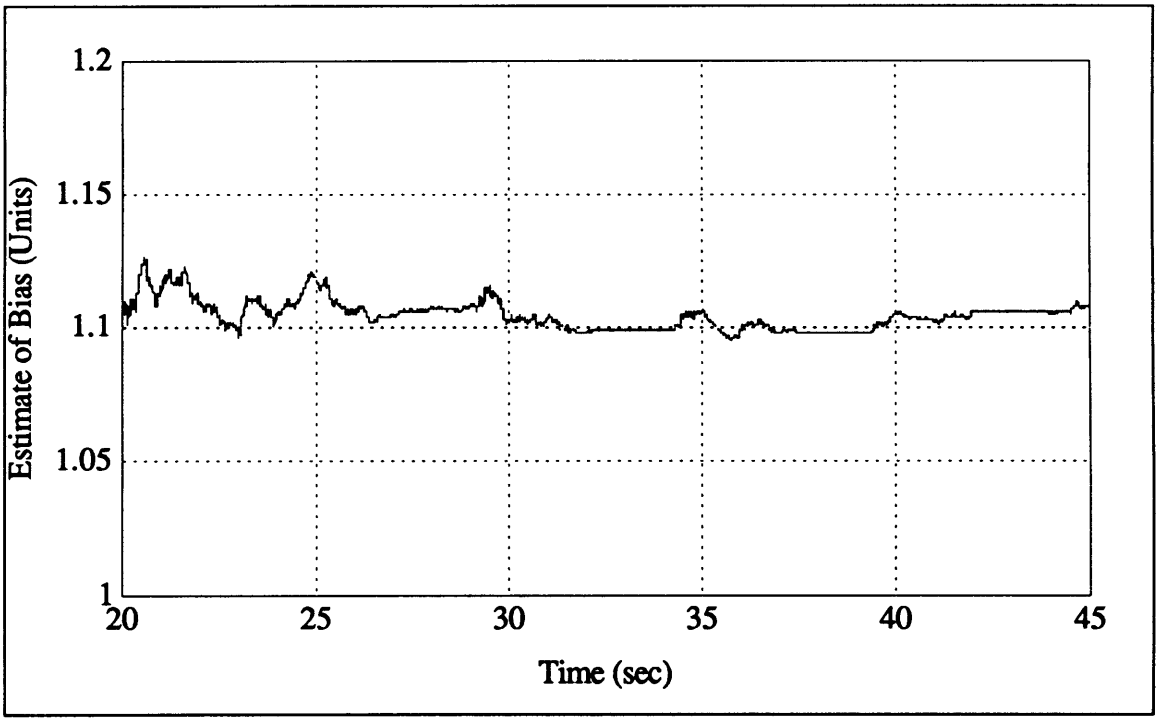

Figure 6.8. Bias Estimate versus Time for Adaptive Filter, Test 6C2

6.4 Conclusions on Adaptive Filters

These simulated data tests have demonstrated that both the raw data reduction and reduced data analysis adaptive filters can improve instrument performance over the traditional filters used for these tasks. The raw data reduction adaptive filter reduces the standard deviation of the decimated data whether or not periodic noise exists. The reduced data analysis adaptive filter is better than the Kalman filter for two main reasons. When no improvement to the system model is made, the adaptive filter still determines the true value of the measurement noise and process noise, instead of using a fixed value. If model adaptation is necessary, the adaptive filter performs substantially better. This improvement alone is important for determining instrument performance. In Chapter 8, these adaptive filters will be applied to micromechanical instrument data in order to improve the characterization of instrument performance.

Chapter 7

Micromechanical Inertial Instruments

7.1 Introduction

Micromechanical inertial instruments are inertial instruments such as accelerometers and gyroscopes that are fabricated on a silicon wafer. Because the fabrication techniques are similar to those used for computer chip production, the required electronics can be placed on the same chip as the inertial instrument, minimizing off-chip processing. These instruments can be mass produced, resulting in uniformity, low cost, small size, low readout noise and high production rates. Also, because these instruments are solid state, they can withstand large g-forces without failing. However, these instruments, because of their minute size, have a wide range of noise sources. Brownian motion, as well as pressure variations, have a significant effect on these instruments.

7.2 Micromechanical Gyroscopes

Presently, several forms of micromechanical gyroscopes are being developed. Each is a unique attempt to improve upon the performance limits of the previous gyroscope. The vibratory micromechanical gyroscope demonstrated the concept of micromechanical gyroscopes, but the performance was not adequate. The inverted gyroscope design improved on the vibratory gyro with subtle changes in operation. The tuning fork gyroscope is a fundamentally different approach to the same problem.

7.2.1 Vibratory Gyroscope

The basic vibratory micromechanical gyroscope is fabricated from a single silicon block, with gold electroplate as an inertial element, as shown in Figure 7.1. Two torsional flexures connect the outer gimbal to the case. Similarly, two flexures connect

the inner gimbal to the outer gimbal. These two orthogonal sets of flexures are weak in torsion, but stiff in all other directions [12,13].

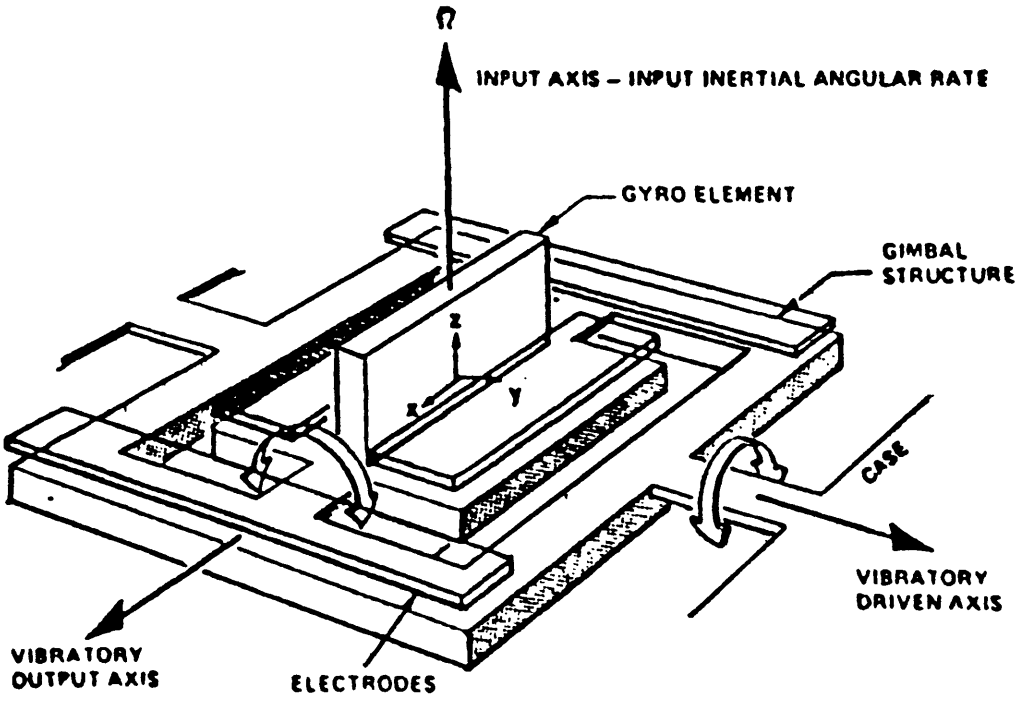

Figure 7.1. Vibratory Micromechanical Gyroscope

Torquing electrodes embedded in the chip vibrate the outer gimbal, creating an oscillatory angular momentum vector, **H**, about the Vibratory Driven Axis, shown in Figure 7.1. In order to achieve maximum resolution, the outer gimbal is vibrated at the resonant frequency of the inner gimbal. If there is an angular rate about the input axis, the inner gimbal will oscillate about its flexures at a magnitude proportional to the angular rate about the input axis. Readout electrodes, mounted above the inner gimbal, detect a change in capacitance as the inner gimbal oscillates. By sensing this change in capacitance, the angular rate about the input axis can be determined using Equation 7.1 [12,13].

$$\theta = \left[\frac{I_x + I_y - I_z}{I_x}\right]\frac{\varphi_0 \Omega Q}{\omega_n} \tag{7.1}$$

where,

θ = the output angle of the inner gimbal

I_j = moment of inertia about the j axis

φ_0 = the drive angle of the outer gimbal

Q = the mechanical resonant quality factor

Ω = the angular input rate to the gyroscope

and,

ω_n = the natural frequency of the inner gimbal.

Closed loop operation is achieved by torquing the inner gimbal back to a null position. This gyro has been designed, built and tested successfully. One major drawback to this design is the presence of noise.

Open loop operation of a typical micromechanical gyroscope is shown in Figure 7.2.* The inertial element is driven at its natural frequency, and a 100 kHz carrier signal is applied to the output axis. Motion from mass imbalance or rate input causes a modulation of this 100 kHz signal at the resonant frequency of the gyro element. If the resonant frequency of the element is 30 kHz, then the modulated frequencies are 70 kHz and 130 kHz. These carrier signals then pass through the preamp, postamp and various filters. At the carrier demodulation, the 100 kHz signal is removed, and the resulting frequency components are 30 kHz, 170 kHz and 230 kHz. The low pass filter removes the 170 kHz and 230 kHz signals. The remaining 30 kHz signal is then demodulated at the natural frequency, resulting in a D.C. signal and a 60 kHz signal. Next, a low pass filter allows only the D.C. signal to pass. This D.C. signal is the rate sensed by the gyroscope. For closed loop operation, this signal would be remodulated and applied to rebalance the capacitor plates.

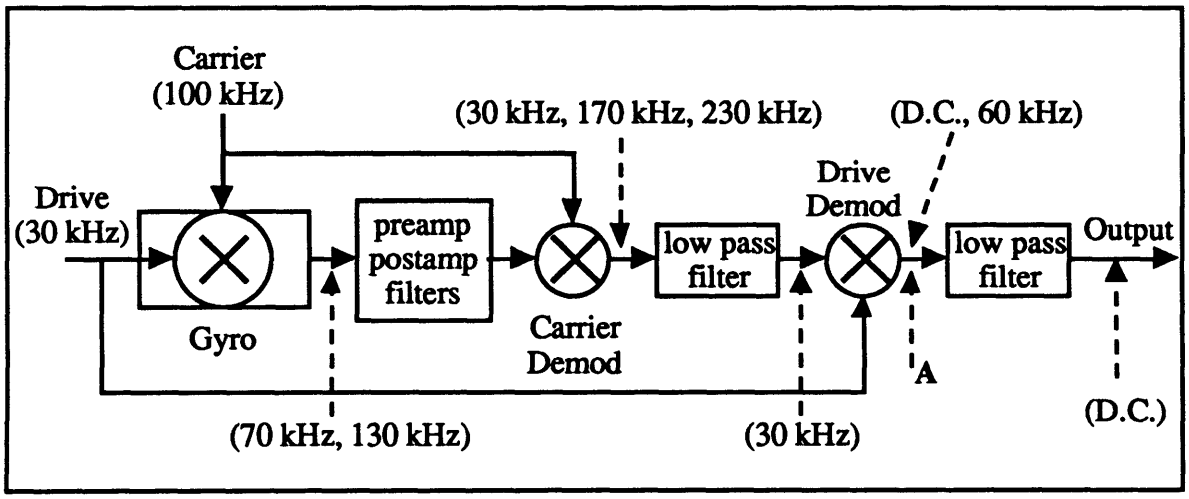

Figure 7.2. Schematic of Micromechanical Gyroscope Operation

* Schematic provided by Anthony Kourepenis.

7.2.2 Noise Sources

One major source of error is the stability of the scale factor that converts the gyro output from volts to deg/hr. Presently, the scale factor is so large (~10^6 deg/hr/volt) that the output voltage reading must be accurate to μvolts in order to achieve an accuracy of less than 10 deg/hr. A preamplifier was added to the circuit to reduce this problem. However, more improvement is necessary in this area before the gyroscope performance is satisfactory.

Because of the minute size of this instrument, Brownian motion and damping are forces that significantly limit gyroscope performance [3,40]. A method of isolating the gyroscope in a vacuum is needed in order to reduce the impact of these effects. Once this problem has been solved, the performance of the gyroscope will greatly improve because the Q (Equation 7.1) will increase, raising the magnitude of the inner gimbal oscillation for a given input rate. The data analysis adaptive filter can adjust for noise sources such as Brownian motion and pressure sensitivity. Other contributing noise sources are preamplifier noise and feedback capacitance. Feedback capacitance is a significant problem because the capacitance change sensed by the electrodes is on the order of tens of attofarads (10^{-18} Farads). However, thermal variations have a very small effect on these instruments.

Bandwidth is an important component of the gyroscope performance [3,40]. The readout noise is proportional to the bandwidth, and the mechanical noise of the instrument is proportional to the square root of the bandwidth. In order to improve instrument performance, the noise due to bandwidth must be reduced. Improvement at higher bandwidths has been demonstrated in Chapter 6, using the raw data reduction adaptive filter.

Another significant noise source is random walk [3,40]. Using the maximum likelihood estimator, the random walk parameter can be adequately determined, and more accurate estimates of the system model may be obtained.

7.2.3 Inverted Gyroscope

The inverted gyroscope design is an attempt to improve the vibratory gyroscope performance. This design is very similar to the basic vibratory gyroscope, but the method of operation is different. In this gyroscope, the inner gimbal is vibrated at the natural frequency of the outer gimbal. When an inertial rate is applied about the input axis of the gyro, the outer gimbal will oscillate with a magnitude proportional to the input rate. Because the outer gimbal oscillates to generate output, the oscillations are relative to a fixed case, providing a more stable and more accurate output than the vibratory

gyroscope. The scale factor of the inverted gyro is four times less than that of the original micromechanical gyroscope (360,000 deg/hr/volt vs. 1,200,000 deg/hr/volt). The inverted gyroscope design is affected by the same noise sources as the original design.

7.2.4 Tuning Fork Gyroscope

The third design, expected to be the most accurate, is the tuning fork gyroscope. This design exploits the Coriolis force in order to detect input rates. In this design, a tuning fork is constructed on a silicon wafer, as shown in Figure 7.3. The prongs of the fork are vibrated 180° out of phase at a given frequency, creating an oscillating velocity vector at the end of each prong.

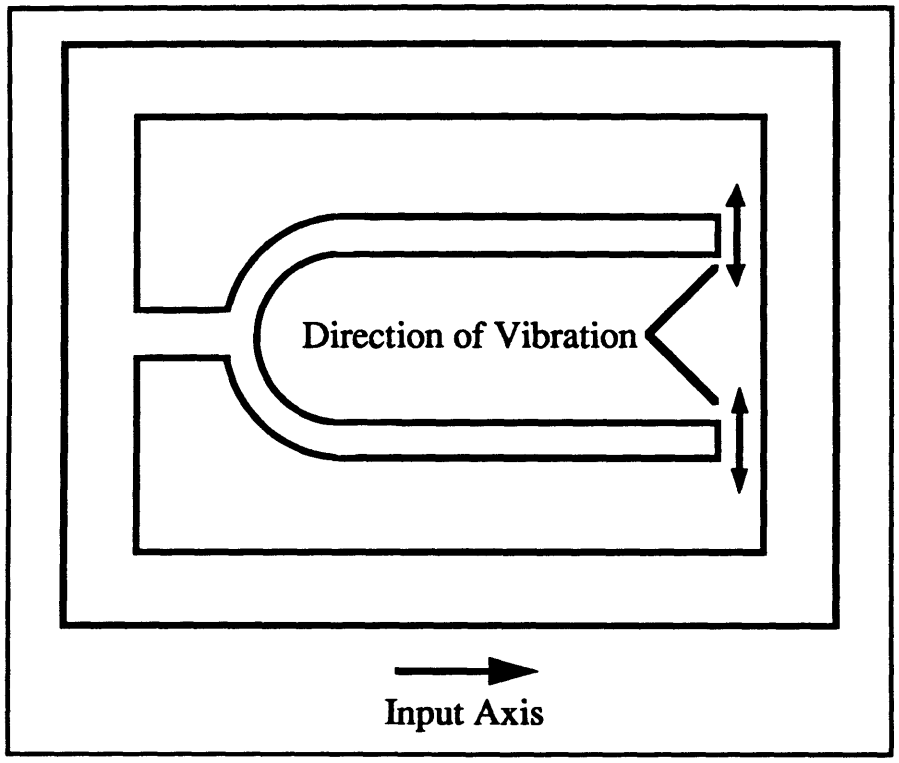

Figure 7.3. Schematic of Tuning Fork Gyroscope

If an angular rate is applied along the axis of the tuning fork, the Coriolis force bends the prongs of the fork in the direction perpendicular to the plane of the gyroscope as shown in Figure 7.4. Using electrodes above and below the tuning fork to measure capacitance, the deflection of the prongs can be measured and the input rate to the gyro can be calculated.

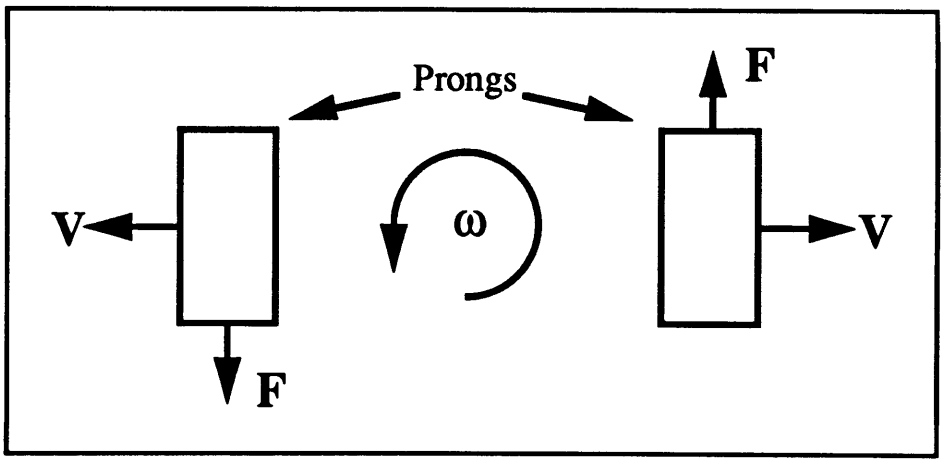

Figure 7.4. Tuning Fork Gyro Operation

The Coriolis force is proportional to the angular rate and velocity at a given time, as shown in Equation 7.2.

$$\mathbf{F} = 2\boldsymbol{\omega} \times \mathbf{v} \tag{7.2}$$

where,

$\mathbf{F}$ = the Coriolis force about the input axis

$\boldsymbol{\omega}$ = the input rate

and,

$\mathbf{v}$ = velocity of the prong.

Using beam theory, the deflection measured at the end of the tuning fork can be converted into a force applied at the end of the beam. With a known velocity vector, the angular rate about the input axis of the tuning fork gyro can be determined.

7.3 Conclusions

Many of the noise sources that affect micromechanical gyroscopes can be modeled for use in an adaptive Kalman filter. The maximum likelihood estimator can determine the random walk of the instrument, as well as the measurement noise σ. Pressure sensitivities may also be included in the system model with the correlation method. Bandwidth of the instrument may be increased by using the raw data reduction adaptive filter, which has shown improvement at higher bandwidths over the triangular filter. Brownian motion may be modeled for using the maximum likelihood estimator, as discussed by Matthew Skeen in his thesis, *Maximum Likelihood Estimation of Fractional Brownian Motion and Markov Noise Parameters* [43].

Chapter 8

Micromechanical Gyroscope Data Analysis

8.1 Approach

Data from different micromechanical inertial instruments and from various test procedures were analyzed to demonstrate that both of the adaptive filters improve the estimates of instrument performance. These analyses were compared with those of the traditional triangular and Kalman filters. Data was collected from both the original micromechanical design and from the inverted gyro design. The tests that were conducted include stationary drift tests, commanded rate tests and pressure variation tests; sampling rates were also varied throughout these tests. For all of these tests, six channels of data were taken: rate input, gyro output, fixture temperature, quadrature, pump pressure, and gyro temperature.

8.2 Stationary Drift Test with Original Gyro Design, Test 8A

For Test 8A, data was collected at 3 Hz for 16 hours in a stationary drift test with IA perpendicular to the earth's axis using the original vibratory gyroscope design. The gyro output was analyzed using both raw data reduction and reduced data analysis.

8.2.1 Raw Data Reduction of 3 Hz Stationary Drift Test

A PSD of the gyro output was constructed for Test 8A. As shown in Figure 8.1, no significant frequency component exists in this output. The gyro output was decimated at various bandwidths, ranging from 0.002 Hz to 0.033 Hz, using both the raw data reduction adaptive filter and the triangular filter (Test 8A1). Figure 8.2 presents a comparison of the standard deviations of the decimated data at various bandwidths (in deg/hr, using a conversion factor of 1.23887×10^6 deg/hr/volt). The dashed line represents the triangular filter standard deviations, and the solid line represents the raw

data reduction adaptive filter standard deviations. The data reduction adaptive filter performed, on average, 10 percent better than the triangular filter. The greatest improvement was 403 deg/hr at a bandwidth of 0.033 Hz.

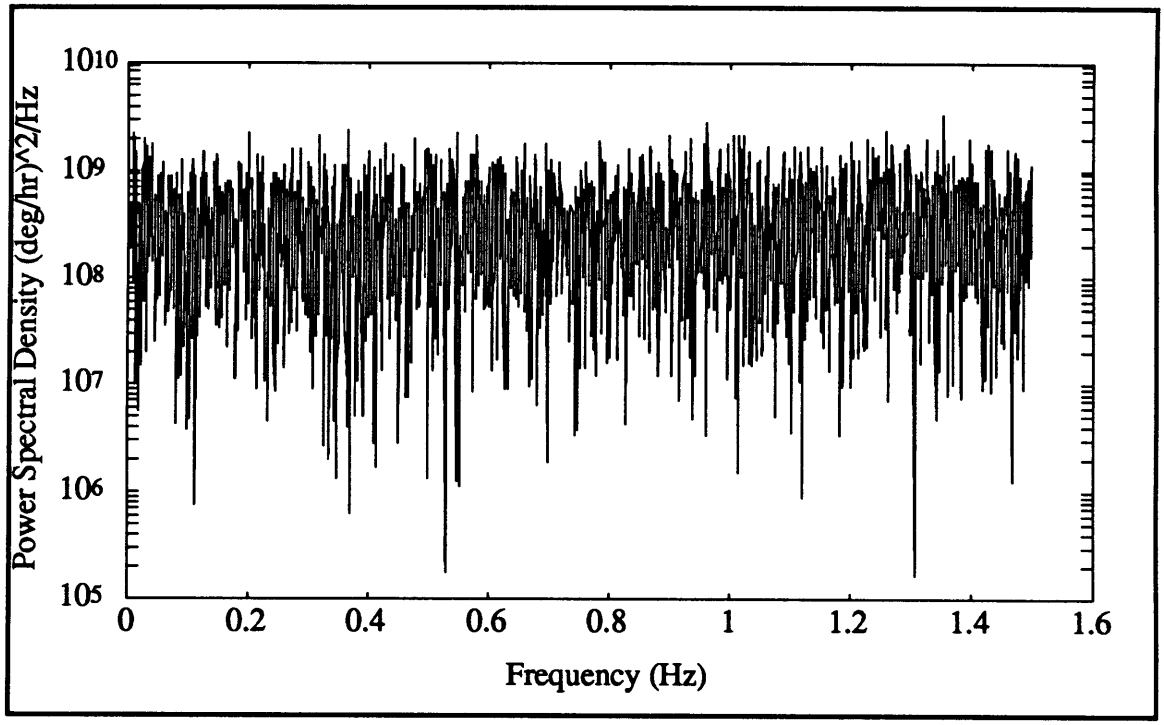

Figure 8.1. Power Spectral Density for Test 8A

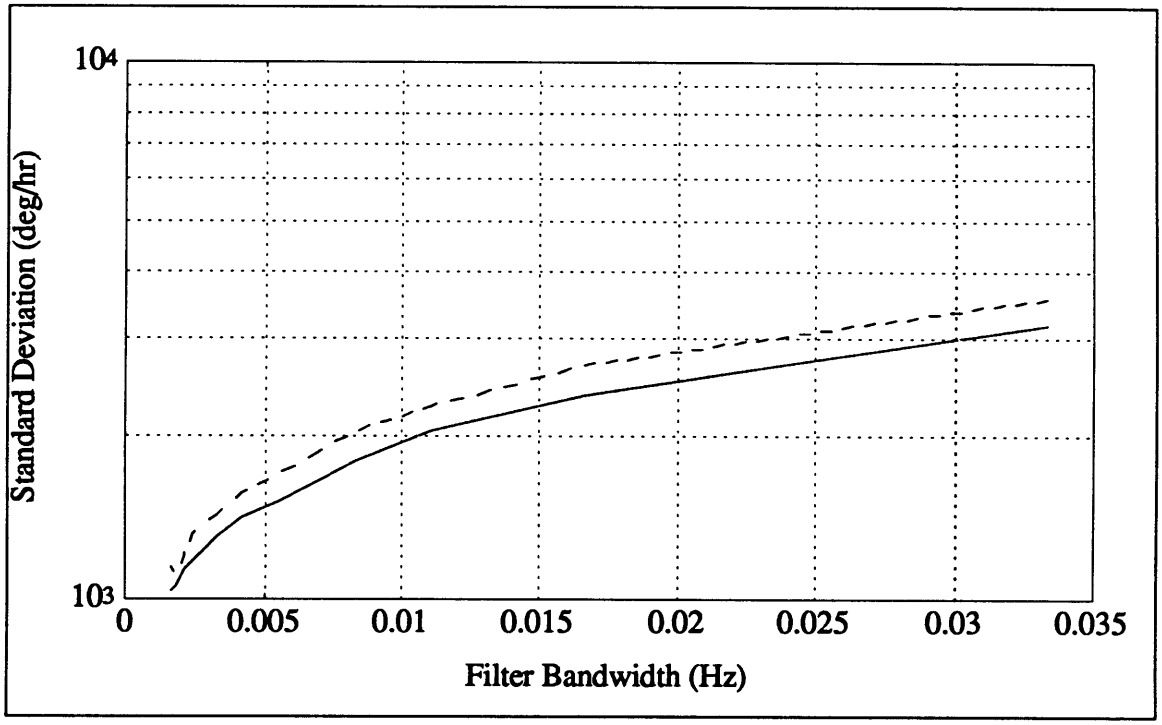

Figure 8.2. Standard Deviation vs. Bandwidth for Test 8A1

8.2.2 Reduced Data Analysis of 3 Hz Stationary Drift Test

After the data decimation analysis, the gyro output was then analyzed using the reduced data analysis adaptive filter with an estimation interval of 1000 points. Because this is a stationary drift test, the bias term cannot be distinguished from the rate dependent terms. Therefore, in Test 8A2, only the bias and the measurement noise standard deviation were estimated; a random walk parameter was included in a second analysis of this data, Test 8A3. This data was then analyzed using a Kalman filter, Test 8A4, with a given measurement noise standard deviation of 123,887 deg/hr. In Tables 8.1 through 8.3, these results are compared, and the innovation sequence RMS is presented. This data has been converted to deg/hr using a conversion factor of 1.23887×10^6 deg/hr/volt.

Table 8.1. Data Analysis on Stationary Drift Data, Test 8A2

Term (deg/hr)	Adaptive Kalman Filter	
	Initial	Final
Mean	0.0000	-449,164 ± 48
$\sigma_{measurement}$	1,238,870	19,377 ± 438
Random Walk (°/hr/√s)	Not Modeled	Not Estimated
Innov. Seq. RMS	—	19,447

Table 8.2. Data Analysis on Stationary Drift Data, Test 8A3

Term (deg/hr)	Adaptive Kalman Filter	
	Initial	Final
Mean	0.0000	-448,455 ± 307
$\sigma_{measurement}$	1,238,870	19,323 ± 437
Random Walk (°/hr/√s)	12,388.7	451 ± 726
Innov. Seq. RMS	—	19,451

Table 8.3. Data Analysis on Stationary Drift Data, Test 8A4

Term (deg/hr)	Traditional Kalman Filter	
	Initial	Final
Mean	0.0000	-449,093 ± 303
$\sigma_{measurement}$	123,887	123,887
Random Walk (°/hr/√s)	Not Modeled	Not Estimated
Innov. Seq. RMS	—	19,452

The adaptive filter improved the standard deviation of the estimate of the mean by 84 percent over the traditional Kalman filter. Table 8.2 suggests that there is no random walk in the data over the 16 hour time span, because the standard deviation of the random walk parameter is almost twice as large as the estimate itself.

8.3 Scale Factor Test with Inverted Gyroscope Design, Test 8B

The inverted gyroscope design was used for the second test, Test 8B. Data was collected at 1 Hz over 95 minutes for a commanded rate test, with a commanded rate profile of 100 deg/sec to -100 deg/sec in 10 deg/sec increments, as shown in Figure 8.3. Figure 8.4 shows the difficulties encountered during data acquisition. The bias term of the gyroscope jumped dramatically (around 1.5 volts) near 3900 seconds. This unexpected shift in bias encouraged the development of the fault tolerant algorithm that was discussed in Chapter 6. This data set was used to successfully demonstrate that the fault tolerant algorithm is capable of detecting and accounting for radical jumps in system behavior.

Raw data reduction was not done for this data because the comparison used in this thesis, the standard deviation, loses it meaning in a commanded rate test. In order to properly determine the standard deviation of the reduced data, each segment must be examined individually, and this procedure is not necessary for demonstrating the effectiveness of the raw data reduction adaptive filter.

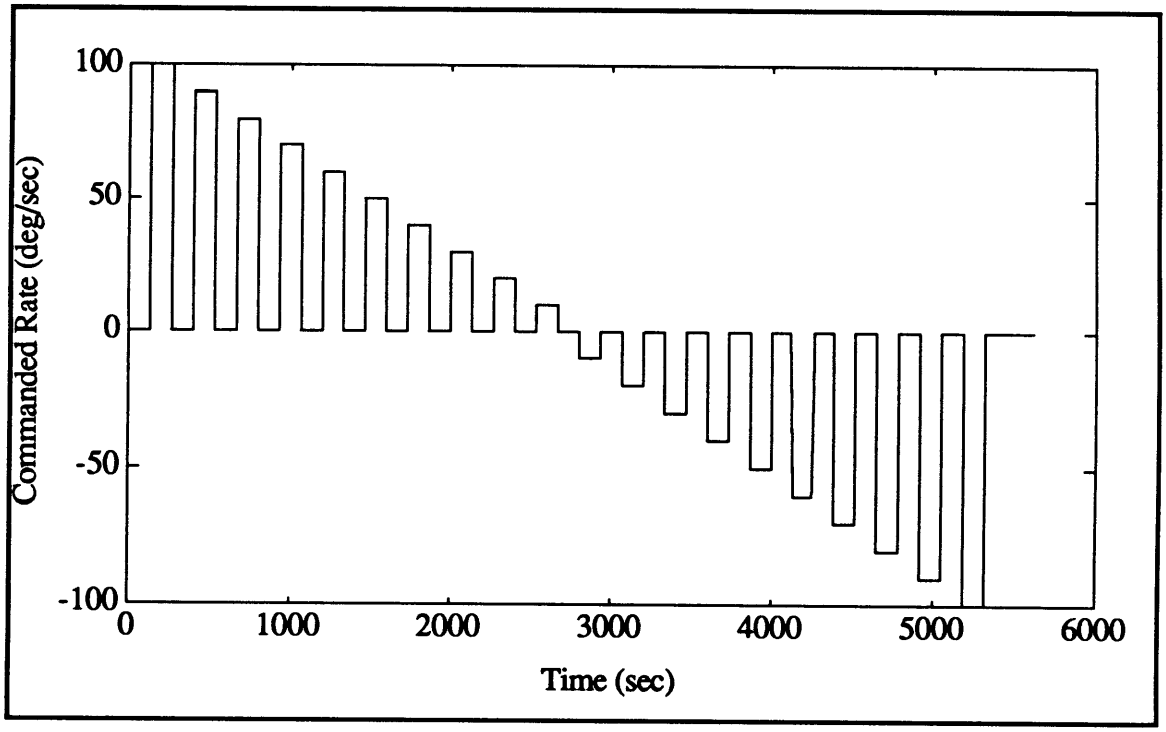

Figure 8.3. Commanded Rate Profile for Test 8B

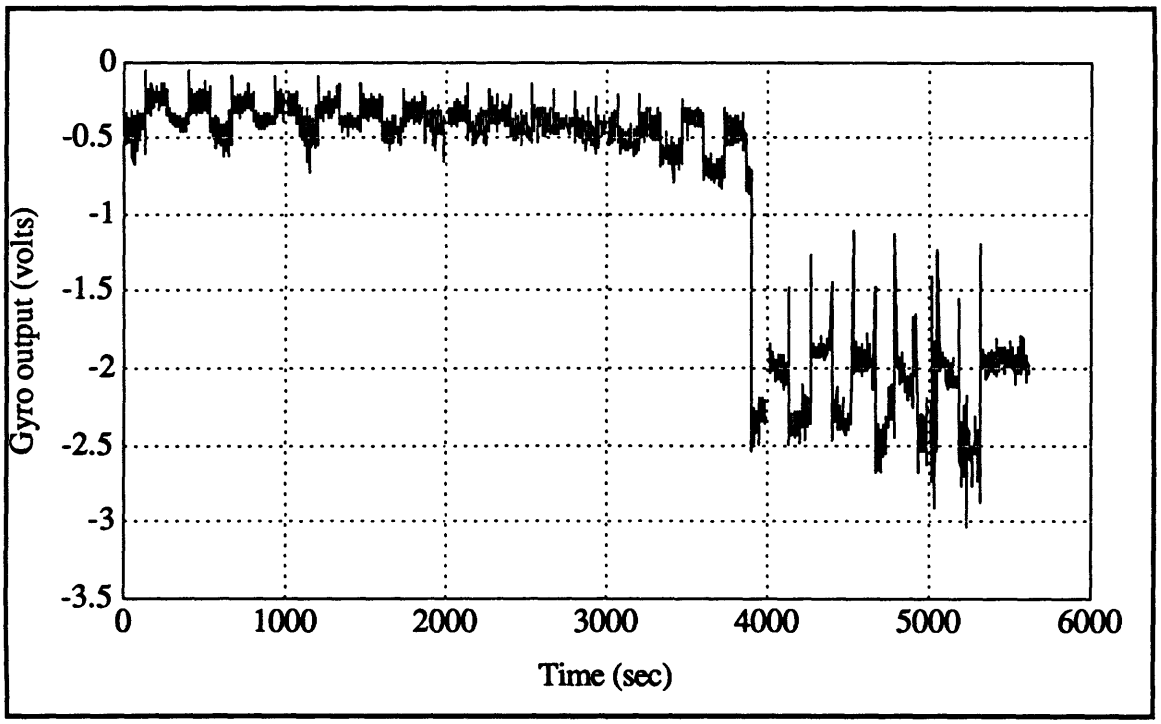

Figure 8.4. Inverted Gyro Output for Test 8B

8.3.1 Reduced Data Analysis of Commanded Rate Test

When this data was analyzed using the reduced data analysis filter, the fault tolerant algorithm shifted the bias estimate at 3900 seconds, as expected. A comparison between the reduced data analysis adaptive filter and Kalman filter is shown in Table 8.4. All of the results in this table are given in volts. No attempt is made to convert the voltage output into deg/hr because of the noise in the test and the large bias shift. In later tests with better data, the results will be presented in deg/hr.

Table 8.4. Results for Commanded Rate Test, Test 8B

Parameter	Kalman Filter, Test 8B1		Adaptive Filter, Test 8B2	
(Volts)	Initial Value	Final Estimate	Initial Value	Final Estimate
Bias	0.0000	-0.9340 ± 1.3e-3	0.0000	-1.9412 ± 7.3e-3
Scale Factor	0.0000	0.0124 ± 3.1e-5	0.0000	6.23e-3 ± 1.7e-4
Msmt Noise	0.1000	0.1000	1.0000	0.0656 ± 0.0052
Innov. RMS	—	0.6246	—	0.1553

Figure 8.5 shows the innovation sequence for the traditional Kalman filter. This filter is unable to account for the jump in the bias term, and the innovation sequence reflects the poor bias estimate. Figure 8.6 shows the innovation sequence versus time for the

reduced data analysis adaptive filter. With this filter, the bias estimate quickly adjusts for the jump in the data, and the innovation sequence quickly returns to a white noise sequence.

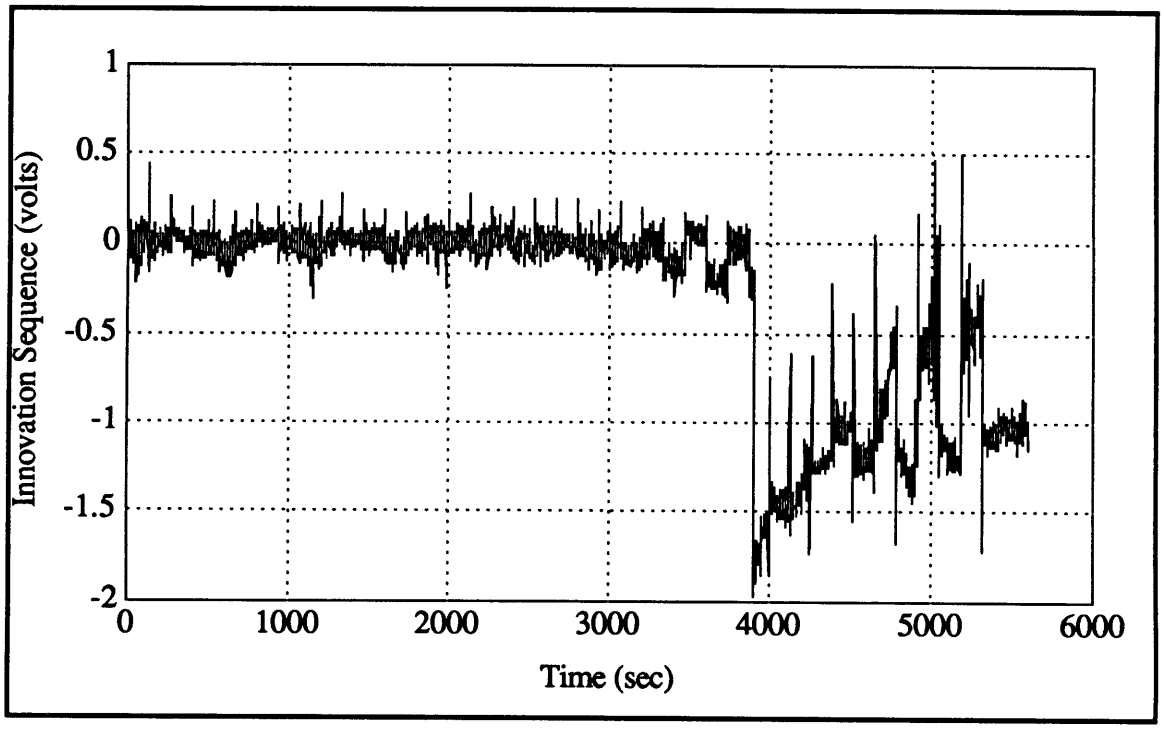

Figure 8.5. Innovation Sequence of Kalman Filter for Test 8B1

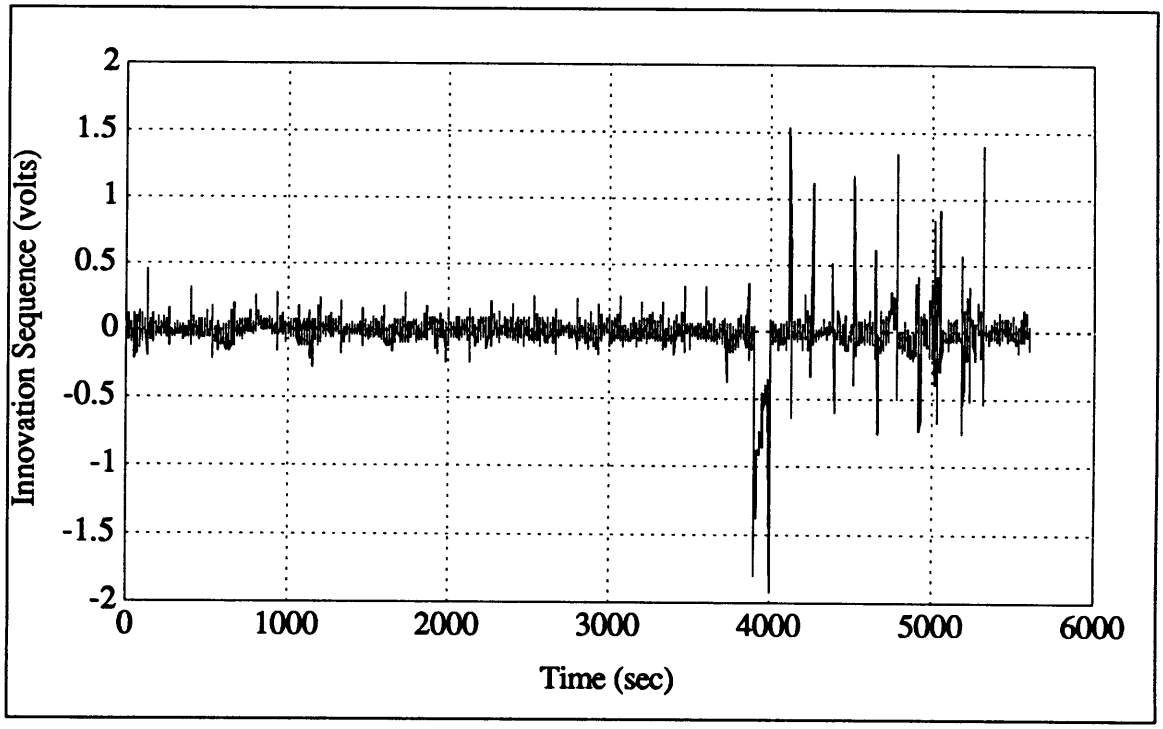

Figure 8.6. Innovation Sequence of Adaptive Filter for Test 8B2

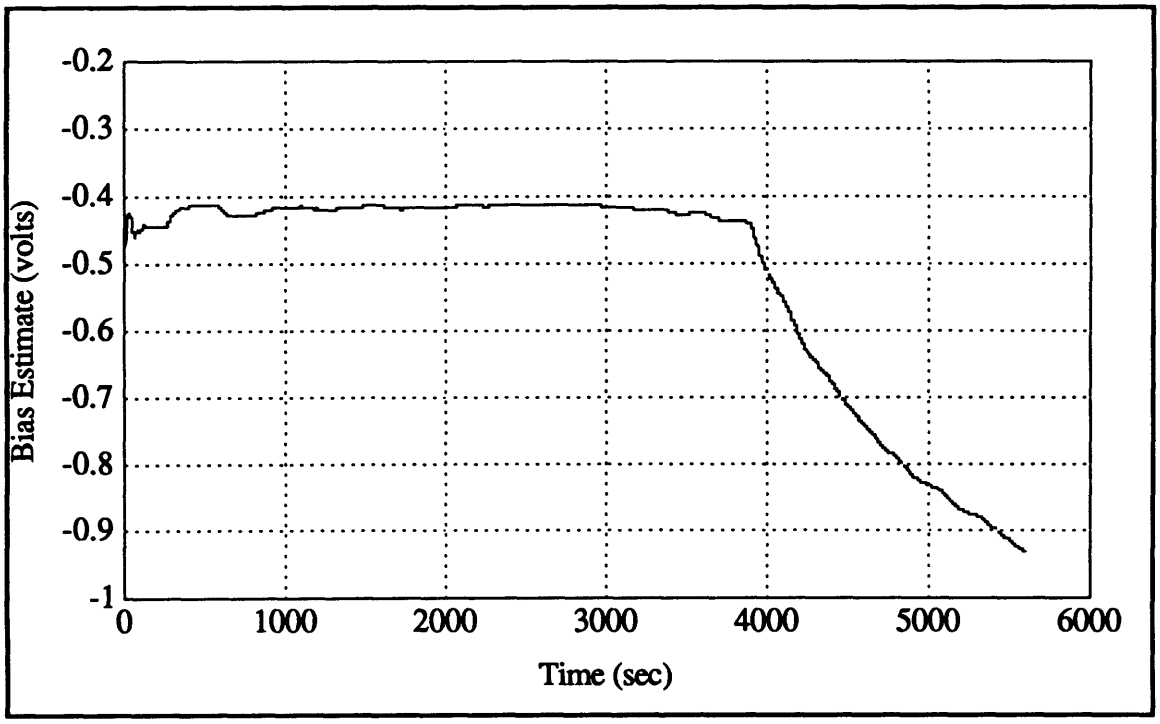

Figure 8.7. Bias Estimate of Kalman Filter for Test 8B1

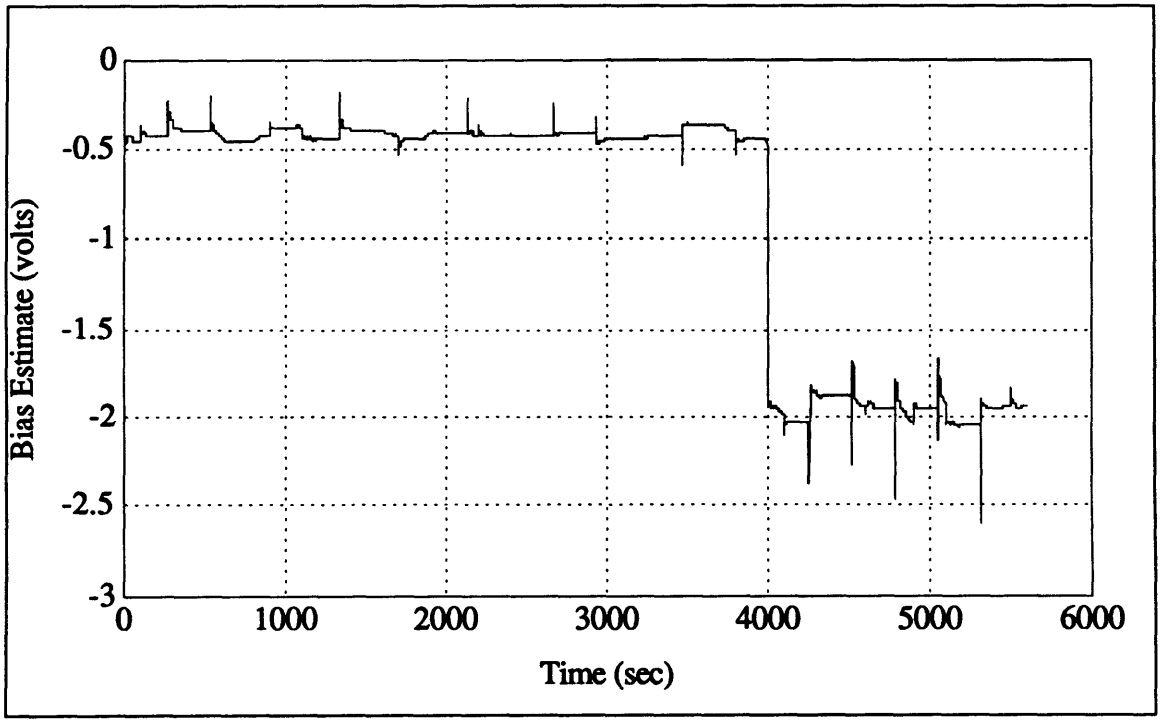

Figure 8.8. Bias Estimate of Adaptive Filter for Test 8B2

As shown in Figures 8.5 and 8.6, the adaptive filter was able to compensate for the jump in the gyro output. While the innovation sequence of the Kalman filter in Figure 8.5 is erratic after the bias jump, the adaptive filter innovation sequence, in Figure 8.6, has adjusted for this bias shift. Figures 8.7 and 8.8 show that the adaptive filter bias estimate (Figure 8.8) successfully shifted to the new value of the bias, while the bias estimate of the Kalman filter (Figure 8.7) gradually shifted towards the new bias estimate.

8.4 Stationary Drift Test with Inverted Design, Test 8C

For the third test, Test 8C, the motivation was to examine the performance of the raw data reduction filter to identify and model periodic noise. This concept was explored in detail in Section 5.4.2. To obtain data with periodic noise, the data acquisition system was connected to point A in Figure 7.2. Since the data is taken after the drive frequency demodulation, but before the baseband filter, there will be a high frequency noise signal equal to twice the drive frequency of the inertial element of the gyroscope. For this inverted gyroscope, the drive frequency is 2500 Hz, so that the high frequency noise is at 5000 Hz. The data was sampled at 1500 Hz, which causes the 5000 Hz signal to be aliased to a 500 Hz signal. This stationary drift test was conducted for 6 seconds. Figure 8.9 shows the PSD of the original signal from this test. This data set was analyzed using both the raw data reduction adaptive filter and triangular filter.

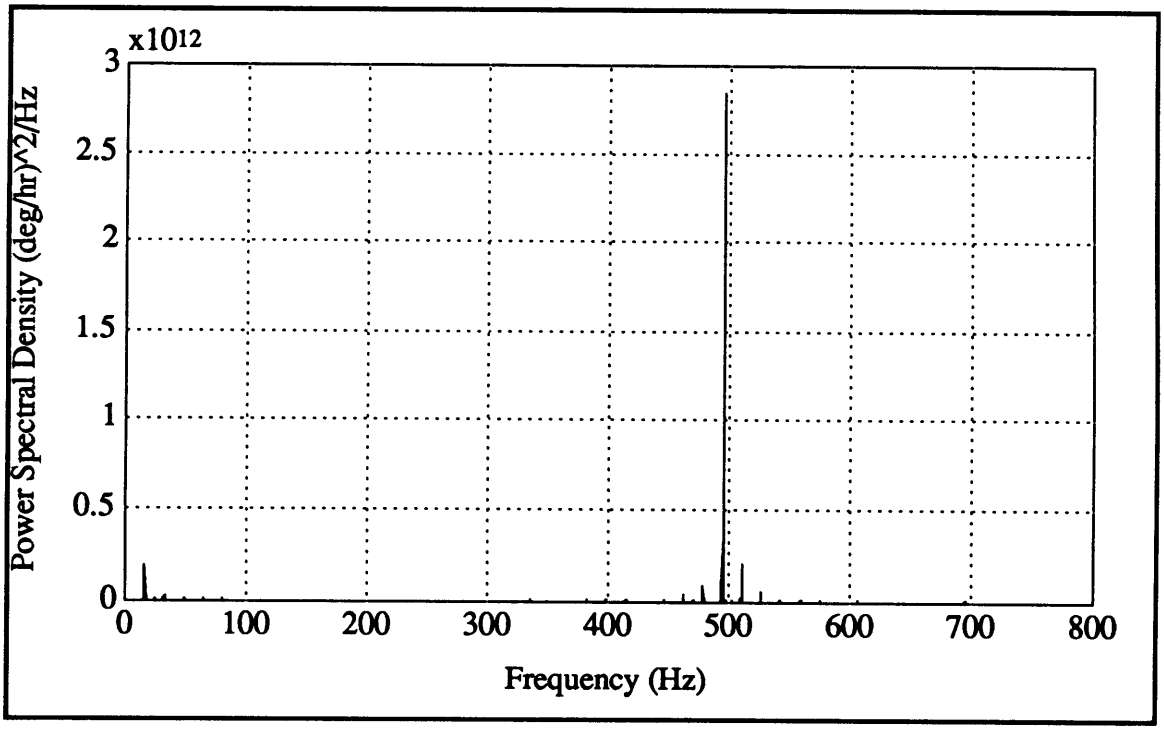

Figure 8.9. PSD of Gyro Output for Stationary Drift Test, Test 8C

8.4.1 Raw Data Reduction of 1500 Hz Stationary Drift Test

For Test 8C1, the data was decimated using both the raw data reduction adaptive filter and the triangular filter at various bandwidths. Figure 8.10 is a plot of the log of the standard deviations in deg/hr for both filters as a function of bandwidth. The solid line represents the adaptive filter; the dashed line represents the triangular filter. As the bandwidth increases, the adaptive filter has a lower standard deviation that achieves a 51.38 percent improvement at 10.71 Hz bandwidth, which, using a conversion factor of 360,000 deg/hr/volt, gives an improvement of about 3620 deg/hr. Larger bandwidths are not possible with the adaptive filter because there are too few data points (less than 70 points) in an estimation interval, and the filter cannot converge to the true estimate of the mean in such a short interval.

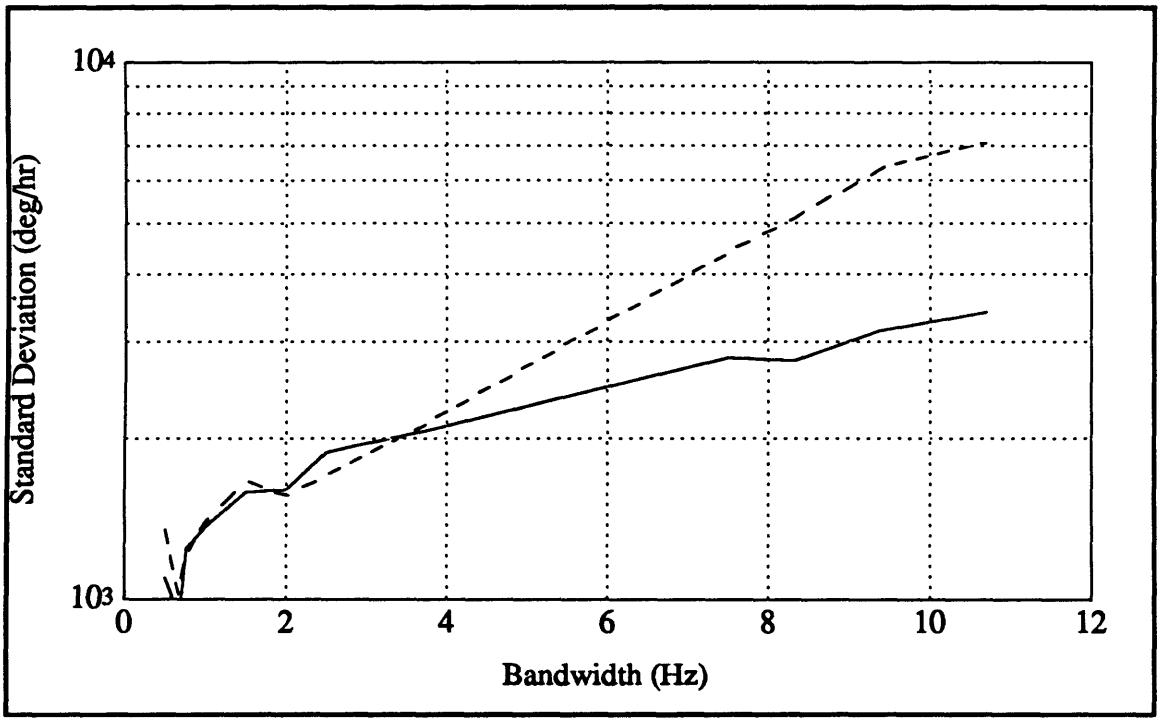

Figure 8.10. Standard Deviation vs. Bandwidth for Test 8C1

8.4.2 Reduced Data Analysis of 1500 Hz Stationary Drift Test

After completing the raw data reduction, the data set was analyzed using two state vector estimators: the reduced data analysis adaptive filter and the Kalman filter. Because this was a stationary drift test, only the mean of the data could be estimated. Two tests were run using the reduced data analysis adaptive filter with an estimation interval of 100 points: one estimating measurement noise σ (Test 8C2), and one estimating both the noise σ and the random walk parameter (Test 8C3). Table 8.5 through Table 8.7 present the results of these analyses. In the first two tests, frequency

components were removed at multiple frequencies. The reason for the wide range of frequency components most likely has to do with frequency beating that results from the inability of the filter to identify a frequency component exactly. In Test 8C4, a Kalman filter was used, with a given measurement noise standard deviation of 36,000 deg/hr.

Table 8.5. Data Analysis on High Frequency Drift Data, Test 8C2

Term (deg/hr)	Reduced Data Analysis Filter	
	Initial	Final
Mean	0.0000	37,404 ± 1008
$\sigma_{measurement}$	360,000	26,388 ± 1836
Random Walk	Not Modeled	Not Estimated
Freq. ID (Hz)	15.93, 478.6, 494.8, 510.3, 526.6	
Innov. Seq. RMS	—	31,428

Table 8.6. Data Analysis on High Frequency Drift Data, Test 8C3

Term (deg/hr)	Reduced Data Analysis Filter	
	Initial	Final
Mean	0.0000	38,268 ± 2664
$\sigma_{measurement}$	360,000	27,180 ± 1944
Random Walk	3600	11,628 ± 20,988
Freq. ID (Hz)	16.11, 478.64, 494.75, 510.31	
Innov. Seq. RMS	—	31,752

Table 8.7. Data Analysis on High Frequency Drift Data, Test 8C4

Term (deg/hr)	Traditional Kalman Filter	
	Initial	Final
Mean	0.0000	36,432 ± 396
$\sigma_{measurement}$	36,000	36,000
Random Walk	Not Modeled	Not Estimated
Freq. ID (Hz)	Not Estimated	
Innov. Seq. RMS	—	53,496

These three tables show that the traditional Kalman filter performs non-optimally when unmodeled noises, such as periodic noise, exist in the signal. The periodic noise sources could be removed during raw data reduction, but if a periodic signal is not identified and removed during this phase of data analysis, the Kalman filter will give non-

optimal results. The RMS of the innovation sequence in Test 8C4 is almost twice as large as those in Tests 8C2 and 8C3. Test 8C2 and Test 8C3 have nearly identical performance, suggesting that, over this time span of 6 seconds, random walk is not identifiable. The large uncertainty in the random walk parameter estimate also suggests that random walk does not exist.

8.5 Pressure Variation Test for Inverted Design

For Test 8D, the pressure around the gyroscope varied because of a leak. The test itself involved commanded rates at 100 deg/sec, 90 deg/sec and 80 deg/sec with 0 deg/sec rates between each of the nonzero commanded rates. Data was taken at 1 Hz for 12.5 minutes, as shown in Figure 8.11.

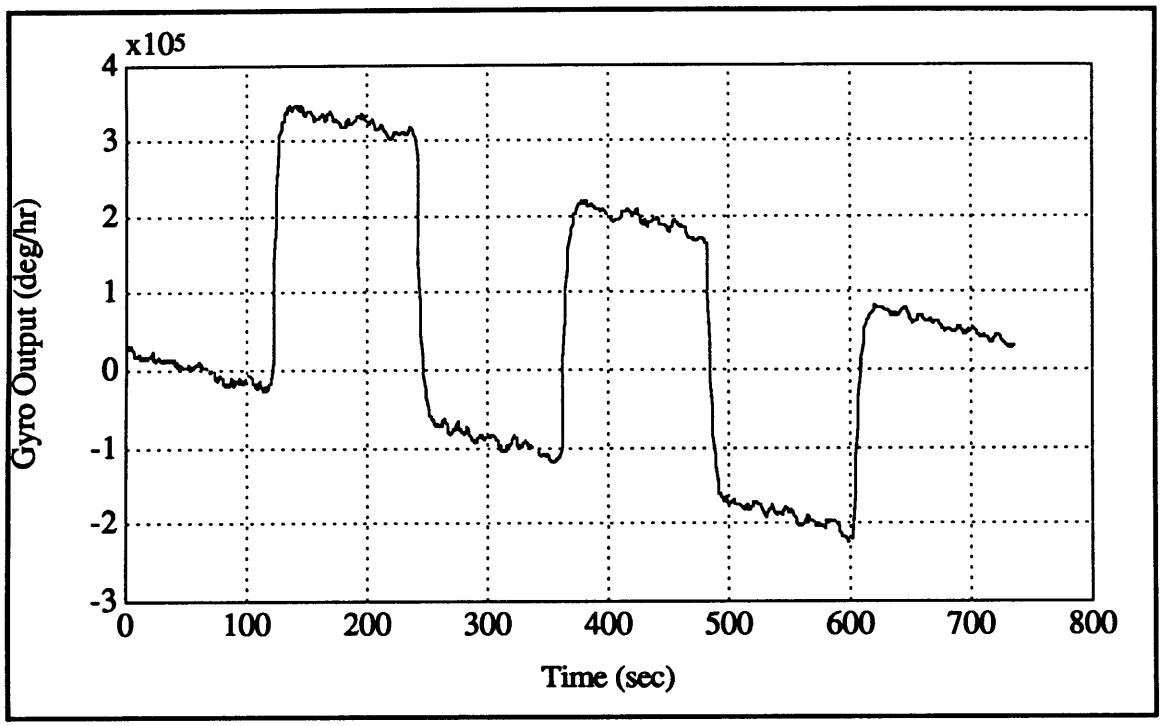

Figure 8.11. Gyroscope Output for Test 8D

Because of the current setup of the gyroscope, however, pressure readings near the gyroscope cannot be taken accurately; the pressure is on the order of millitorr, much smaller than the pressure reading accuracy. Therefore, the pressure variation was modeled as linearly dependent on time. The slope for this approximation is unknown because of the limits of the measuring instruments. With this linear model, time was used to test whether an additional pressure-dependent term could be added to the system model. The gyro output was analyzed using both the reduced data analysis adaptive filter and the traditional Kalman filter. The reduced data analysis adaptive filter successfully

identified the time/pressure dependence of the instrument, and adjusted the system model accordingly. Also, the adaptive filter determined the correct value for the measurement noise, using an estimation interval of 100 points. The results from these two test runs are compared in Table 8.8, using a conversion factor of 360,000 deg/hr/volt.

Table 8.8. Results for Pressure Variation Test, Test 8D

Parameter	Kalman Filter, Test 8D1		Adaptive Filter, Test 8D2	
(deg/hr)	Initial Value	Final Estimate	Initial Value	Final Estimate
Bias	0.0000	-89,450 ± 18900	0.0000	42,118 ± 9166
Scale Factor	0.0000	3174 ± 30	0.0000	3431 ± 60
Press. (°/hr/s)	Not Modeled	Not Modeled	Not Modeled	-405 ± 17
Msmt Noise	36,000	36,000	1.0000	55,388 ± 1783
Innov. RMS	—	92,469	—	52,145

As Table 8.8 shows, the bias estimate for the Kalman filter is very poor, and the Kalman filter innovation sequence RMS is almost twice as large as that of the reduced data analysis adaptive filter. The adaptive filter added the pressure sensitivity/trend term at the first interval, when the inclusion of the term improved the RMS of the innovation sequence by 60.2 percent. Figure 8.12 shows the innovation sequence before (dashed line) and after (solid line) the pressure sensitivity/trend term was added to the model.

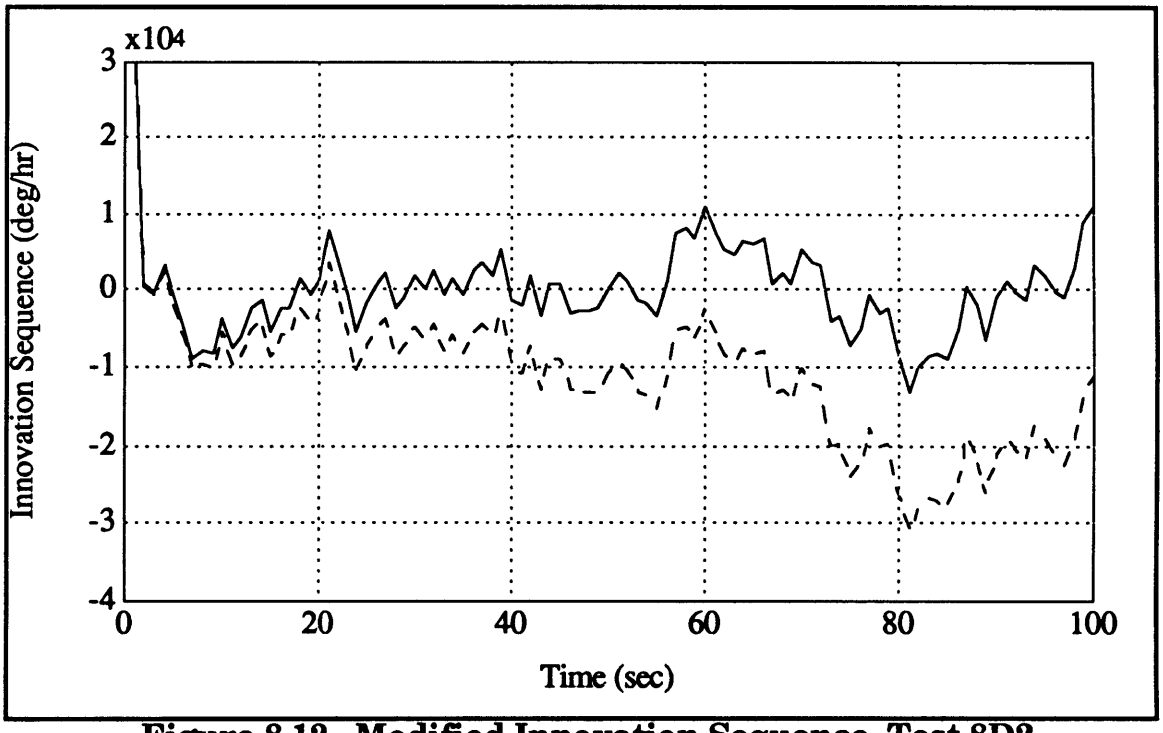

Figure 8.12. Modified Innovation Sequence, Test 8D2

Although the pressure contributes significantly to the performance of the gyroscope, the final design of the gyroscope will be sealed in a vacuum, and therefore unaffected by variations in pressure.

8.6 Commanded Rate Test for Original Design

For this test, Test 8E, the original micromechanical gyro design was given commanded rates from +100 deg/sec to -100 deg/sec in 10 deg/sec increments four times. This data was sampled at 1 Hz over a 16 hour period. The first data analysis test done, Test 8E1, used a traditional Kalman filter to determine the bias and scale factor. Test 8E2 was conducted with the reduced data analysis adaptive filter, testing for a rate squared term and the measurement noise standard deviation using an estimation interval of 200 points. Table 8.9 shows a comparison between Test 8E1 and Test 8E2. The results have been converted to deg/hr using a conversion factor of 1.23887×10^6 deg/hr/volt. The addition of the rate squared term improved the innovation sequence by 5.73 percent. This improvement is very close to the tolerance given, but the final estimate of the rate squared term suggests that it may exist.

Table 8.9. Results for Commanded Rate Test, Test 8E

Parameter	Kalman Filter, Test 8E1		Adaptive Filter, Test 8E2	
(deg/hr)	Initial Value	Final Estimate	Initial Value	Final Estimate
Bias	0.0000	-447,346 ± 516	0.0000	-447,778 ± 98
Scale Factor	0.0000	3587 ± 11.7	0.0000	3588 ± 2.23
Rate Squared	Not Modeled	Not Modeled	Not Modeled	0.2360 ± 0.0337
Msmt Noise	123,887	123,887	1,238,870	20,934 ± 1671
Innov. RMS	—	21,416	—	21,577
Residual RMS	—	21,294	—	21,286

The Residual RMS mentioned in Table 8.9 is the *a posteriori* residual that was discussed in Section 6.3.3. Table 8.9 demonstrates that the reduced data analysis adaptive filter can detect terms that have a very slight affect on the innovation sequence. At the highest rate, 100 deg/sec, the rate squared term contributes only 2360 deg/hr to the instrument, which has a noise of about 21,000 deg/hr. This slight improvement, however, is significant to the instrument's performance.

If the given measurement noise was an order of magnitude smaller, then the Kalman filter's performance would approach that of the adaptive filter, except for the rate squared

term. The high given measurement σ does demonstrate that the Kalman filter is limited by its given parameters, whereas the adaptive filter can adjust to the true values.

For Test 8E3, the temperature was used as the potential mismodeled term. The filter did not identify this term as significant. This result is not surprising since the performance of the micromechanical instruments has not been shown to be dependent on temperature. The results for this test are not shown.

8.7 Conclusions

Characterization of micromechanical inertial instruments can be improved using these adaptive filters. For raw data reduction, improvements from 10 to 50 percent over the triangular filter have been achieved. With the reduced data analysis adaptive filter, both a pressure dependent and a rate squared dependent term have been identified. Also, the measurement noise of the instrument can be determined, even if it is initially unknown. The reduced data analysis filter provides information as least as accurate as the traditional Kalman filter, and it should be used for data analysis.

Chapter 9

Conclusions

9.1 Results

Although the Kalman filter is an efficient estimator, it has certain limitations to its performance. If the given state vector is inadequate, then suboptimal results will be obtained. Also, the system parameters are constant in a Kalman filter. To overcome the limitations of the Kalman filter, adaptive filter concepts have been used to derive and develop two adaptive filters.

A maximum likelihood estimator has been developed for parameter estimation. This estimator is capable of accurate estimates of both **Q** and **R**, as well as estimating time varying parameters. A state vector determination filter has been developed using correlation methods. This filter is effective in online location and inclusion in the state vector of previously unmodeled terms. Trend, pressure and acceleration dependent terms have been identified and included in the state vector. Power spectral density analysis is used for periodic noise identification. This filter has successfully identified unmodeled frequency components in both simulated and real data.

These three concepts have been combined into two adaptive filters; one for raw data reduction, the other for reduced data analysis. These filters improve the analysis over traditional methods, including triangular filters and Kalman filters. The raw data reduction adaptive filter has improved data decimation by as much as 50 percent in real micromechanical gyroscope output. The reduced data analysis filter is quite effective in identifying unmodeled terms. This adaptive filter has identified both a pressure sensitivity of the micromechanical gyroscope, as well as a rate squared dependence. This filter has also accurately estimated the measurement noise standard deviation of various instruments. Random walk was found to not be a significant error source in any of the

instruments analyzed. Because of their versatility and performance accuracy, both of these filters should be considered for use in future data analysis.

9.2 Implementation of Adaptive Filters

Both of these filters can be used online. For the 3 Hertz drift data, the filters could easily process this data online; it took about 5 hours for each filter to analysis 16 hours of data on a Macintosh IIfx. However, for the 1500 Hz data, online implementation is not possible. Future research is necessary to determine the capabilities of these filters for real time applications.

9.3 Recommendations for Future Work

There is plenty of additional work that may be done to improve both of these adaptive filters. The Matlab™ *xcorr* function that was used to generate the correlation sequences can probably be improved. For large data sets, computing the correlations was very burdensome; a more efficient algorithm would greatly improve the online capabilities of this filter. This algorithm could be implemented with minimum effort. Also, amplitude tolerances could be used for the PSD analysis.

Also, by using parallel processors, this filter could be implemented online. By using three processors, one filter could perform the Kalman filter and maximum likelihood estimator equations to generate the innovation sequence, another filter could determine the PSD of this innovation sequence, and the third processor could test additional terms for inclusion in the system model. With parallel processing, the filter outputs would not be delayed by either the PSD or correlation methods; real time output would be generated, and the model could be adapted as the secondary processors determined a new system model.

The fault tolerant algorithm introduced in Chapter 6 should be more fully developed. This algorithm, though effective, is very simple, and a fault tolerant algorithm similar to that presented by Alan Willsky and Harold Jones in "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems" [51] should be included in this filter.

References

[1] Alspach, D.L. "A Parallel Filtering Algorithm for Linear Systems with Unknown Time Varying Noise Statistics." *IEEE Transactions on Automatic Control.* Vol. AC-19, No. 5, October 1974. pp. 552-556.

[2] Boozer, Drayton D. and Willie L. McDaniel, Jr. "On Innovation Sequence Testing of the Kalman Filter." *IEEE Transactions on Automatic Control.* Vol. AC-17, No. 1, February 1972. pp. 158-160.

[3] Boxenhorn, B. and J Connally. "Micromechanical Gyro Noise Analysis." CSDL Memo EAC-90-396. November 12, 1990.

[4] Brown, Robert Grover and Patrick Y.C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering.* New York: John Wiley & Sons, Inc., 1992.

[5] Carew, Burian and Bélanger, Pierre R. "Identification of Optimum Filter Steady-State Gain for Systems with Unknown Noise Covariances." *IEEE Transactions on Automatic Control.* Vol. AC-18, No. 6 December 1973. pp. 582-587.

[6] Chung, Richard C. and Pierre R. Bélanger. "Minimum-Sensitivity Filter for Linear Time-Invariant Stochastic Systems with Uncertain Parameters." *IEEE Transactions on Automatic Control.* Vol. AC-21, No. 1, February 1976. pp. 98-100.

[7] Elwell, John and Capt. Tim Poth. "Micromechanical Inertial Guidance System." C.S. Draper Laboratory Memo, 1990.

[8] Gelb, A., A. Dushman and H.J. Sanberg. "A Means for Optimum Signal Identification." Northeast Electronics Research and Engineering Meeting, 1963. Volume 5, 1963. pp. 80-81.

[9] Gelb, Arthur. *Applied Optimal Estimation.* Cambridge, MA: The M.I.T. Press, 1974.

[10] Godbole, S.S. "Kalman Filtering with No *A Priori* Information About Noise–White Noise Case: Identification of Covariances." *IEEE Transactions on Automatic Control.* Vol. AC-19, No. 5, October 1974. pp. 561-563.

[11] Graupe, Daniel. *Time Series Analysis, Identification and Adaptive Filtering.* Malabar, FA: Robert E. Krieger Publishing Company, 1984.

[12] Greiff, P. and B. Boxenhorn, *et al.* "Silicon Monolithic Micromechanical Gyroscope." C.S. Draper Laboratory Memo, 1991.

[13] Greiff, P. and B. Boxenhorn. "A Vibratory Micromechanical Gyroscope." C.S. Draper Laboratory Memo, 1988.

[14] Hawkes, R.M. and J.B. Moore. "Adaptive Estimation via Sequential Processing." *IEEE Transactions on Automatic Control*. Vol. AC-20, No. 1, February 1975. pp. 137-138.

[15] Jazwinski, Andrew H. "Adaptive Filtering." *Automatica*. Vol. 5, July 1969. pp. 475-485. Reprinted in [44].

[16] Jazwinski, Andrew H. *Stochastic Processes and Filtering Theory*. San Diego, CA: Academic Press, Inc., 1970.

[17] Jora, B. "A Useful Transformation in the Synthesis of Adaptive Observers." *IEEE Transactions on Automatic Control*. Vol. AC-21, No. 3, June 1976. pp. 417-419.

[18] Kailath, Thomas. "An Innovations Approach to Least-Squares Estimation Part I: Linear Filtering in Additive White Noise." *IEEE Transactions on Automatic Control*. Vol. AC-13, December 1968. pp. 646-655. Reprinted in [44].

[19] Kalman, R.E. "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering* (ASME). Vol. 82D, March 1960. pp. 35-45.

[20] Kalman, R.E. and Bucy, R.S. "New Results in Linear Filtering and Prediction." *Journal of Basic Engineering* (ASME). Vol. 83D, March 1961. pp. 95-108.

[21] Kaufman, Howard and Daniel Beaulier. "Adaptive Parameter Identification." *IEEE Transactions on Automatic Control*. Vol. AC-17, No. 5, October 1972. pp. 729-731.

[22] Koenigsberg, W.D. and R.A. Harris. *Quantization Noise, Digital Filtering and Cross Spectral Analysis*. R-869. Charles Stark Draper Laboratory, Cambridge, MA. December 1974.

[23] Landau, I.D. "Unbiased Recursive Identification Using Model Reference Adaptive Techniques." *IEEE Transactions on Automatic Control*. Vol. AC-21, No. 2, April 1976. pp. 194-202.

[24] Leathrum, James F. "On Sequential Estimation of State Noise Variances." *IEEE Transactions on Automatic Control*. Vol. AC-26, No. 3, June 1981. pp. 745-746.

[25] Leondes, C.T. (Ed). *Theory and Applications of Kalman Filtering*. AGARDograph No. 139, 1970.

[26] Ljung, Lennart and Torsten Söderström. *Theory and Practice of Recursive Identification*. Cambridge, MA: The MIT Press, 1985.

[27] Lüders, Gerd and Kumpati S. Narendra. "A New Canonical Form for an Adaptive Observer." *IEEE Transactions on Automatic Control*. Vol. AC-19, April 1974. pp. 117-119.

[28] Lüders, Gerd and Kumpati S. Narendra. "Stable Adaptive Schemes for State Estimation and Identification of Linear Systems." *IEEE Transactions on Automatic Control*. Vol. AC-19, No. 6, December 1974. pp. 841-847.

[29] Martin, W.C. and A.R. Stubberud. "An Additional Requirement for Innovations Testing in System Identification." *IEEE Transactions on Automatic Control.* Vol. AC-19, No. 5, October 1974. pp. 583-584.

[30] *Matlab for Macintosh Computers: User's Guide.* The Mathworks, Inc. 1989.

[31] Maybeck, Peter S. *Stochastic Models, Estimation and Control Volume 2.* New York, Academic Press, 1982.

[32] Mehra, Raman K. "Approaches to Adaptive Filtering." *IEEE Transactions on Automatic Control.* Vol. AC-17, October 1972. pp. 693-698.

[33] Mehra, Raman K. "On the Identification of Variances and Adaptive Kalman Filtering." *IEEE Transactions on Automatic Control.* Vol. AC-15, No. 2, April 1970. pp. 175-184.

[34] Mehra, Raman K. "On-line Identification of Linear Dynamic Systems with Applications to Kalman Filtering." *IEEE Transactions on Automatic Control.* Vol. AC-16, No. 1, February 1971. pp. 12-21.

[35] Mendel, Jerry M. "Postflight Data Analysis by Means of Adaptive, Iterated, Extended Kalman Filtering." *IEEE Transactions on Automatic Control.* Vol. AC-19, No. 5, October 1974. pp. 467-474.

[36] Miller, Kenneth S. and Leskiw, Donald M. *An Introduction to Kalman Filtering with Applications.* Malabar, FA: Robert E. Krieger Publishing Company, 1987.

[37] Myers, Kenneth A. and Byron D. Tapley. "Adaptive Sequential Estimation with Unknown Noise Statistics." *IEEE Transactions on Automatic Control.* Vol. AC-21, No. 4, August 1976. pp. 520-523.

[38] Nahi, Nasser E. and Brian M. Schaefer. "Decision-Directed Adaptive Recursive Estimators: Divergence Prevention." *IEEE Transactions on Automatic Control.* Vol. AC-17, No. 1, February 1972. pp. 61-68.

[39] Pandya, Rajendra N. "A Class of Bootstrap Estimators and Their Relationship to the Generalized Two Stage Least Squares Estimators." *IEEE Transactions on Automatic Control.* Vol. AC-19, No. 6 December 1974. pp. 831-835.

[40] Petranic, T. and C. Pu. "μGyro Noise Analysis". CSDL Memo TDP:μGyro-90-01. 2 January 1990.

[41] Sandel, Nils R. Jr and Khaled I. Yared. *Maximum Likelihood Identification of State Space Models for Linear Dynamic Systems.* ESL-R-814. MIT Electronic Systems Laboratory, Cambridge, MA. April 1978.

[42] *Signal Processing Toolbox for use with Matlab™.* The Mathworks, Inc. 1988.

[43] Skeen, Matthew E. *Maximum Likelihood Estimation of Fractional Brownian Motion and Markov Noise Parameters.* CSDL-T-1109. C. S. Draper Laboratory, Cambridge, MA. December 1991.

[44] Sorenson, Harold W. (Ed). *Kalman Filtering: Theory and Application.* New York: IEEE Press, 1985.

[45] Sorenson, Harold W. "Least Squares Estimation: from Gauss to Kalman." *IEEE Spectrum*. Vol. 7, July 1970. pp. 63-68.

[46] Tajima, Koji. "Estimation of Steady-State Kalman Filter Gain." *IEEE Transactions on Automatic Control*. Vol. AC-23, No. 5, October, 1978. pp. 944-945.

[47] Taylor, Robert M. *Dynamic System Identification Based on Propagation of Probability Densities and Their Partial Derivatives*. CSDL-P-1929. C. S. Draper Laboratory, Cambridge, MA. 1 July 1984.

[48] Upadhyay, Triveni N. and Demetrios G. Lainiotis. "Joint Adaptive Plant and Measurement Control of Linear Stochastic Systems." *IEEE Transactions on Automatic Control*. Vol. AC-19, No. 5, October 1974. pp. 567-571.

[49] Vander Velde, Wallace E. "Class Notes: Statistical Problems in Automatic Control." M.I.T. Fall 1992.

[50] Weinberg, M. "Notes on Micromechanical Gyro Performance." CSDL Memo EEN-91-813. June 4, 1991.

[51] Willsky, Alan S. and Harold L. Jones. "A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems." *IEEE Transactions on Automatic Control*. Vol. AC-21, No. 1, February 1976. pp. 108-112.

[52] Wittenmark, Björn. "A Self-Tuning Predictor." *IEEE Transactions on Automatic Control*. Vol. AC-19, No. 6, December 1974. pp. 848-851.