A NEW TECHNIQUE FOR THE OPTIMAL SMOOTHING
OF DATA

by

Donald C. Fraser

S. B. (1963), S. M. (1963)
Massachusetts Institute of Technology

Submitted in Partial Fulfillment
of the Requirements for the
Degree of Doctor of Science
at the
Massachusetts Institute of Technology
January 31, 1967

Signature of Author: ____________________________________________
Department of Aeronautics and Astronautics,
January 31, 1967

Certified by: ____________________________________________
Thesis Supervisor

______________________________________________________________
Thesis Supervisor

______________________________________________________________
Thesis Supervisor

Accepted by: __________________________________________________
Chairman, Departmental Graduate Committee
A NEW TECHNIQUE FOR THE OPTIMAL SMOOTHING OF DATA

by

Donald C. Fraser

Submitted to the Department of Aeronautics and Astronautics on January 31, 1967 in partial fulfillment of the requirements for the degree of Doctor of Science.

ABSTRACT

Optimum smoothing is a data reduction technique which uses in the best way possible all available information. Optimum smoothing differs from optimum filtering in that the optimum smoother can use information which is in the future with respect to some particular time of interest. The subject of this thesis is the solution of the optimum smoothing problem when the system and measurement functions are linear. The application of this solution to linear and nonlinear dynamic systems is described.
The optimal smoother is formulated as a combination of two optimum filters, one of which works forward over the data and the other of which works backward. A technique which was developed to process the filter which works backward over the data is extended to treat the general problem of optimal filtering when there is no prior estimate. The resulting technique constitutes a completely recursive method of starting a Kalman filter when no a priori information is available. A smoothability condition is derived. This condition enables the user to determine whether or not smoothing will yield results which cannot be obtained in a simpler manner from an optimum filter estimate. Several numerical problems are identified and solved.

Thesis Supervisor: Dr. Wallace E. Vander Velde
Title: Professor of Aeronautics and Astronautics

Thesis Supervisor: Dr. Yao Tzu Li
Title: Professor of Aeronautics and Astronautics

Thesis Supervisor: Dr. James E. Potter
Title: Assistant Professor of Aeronautics and Astronautics
ACKNOWLEDGEMENTS

I wish to express my appreciation to my thesis committee: Prof. W. E. Vander Velde, Prof. J. E. Potter, and Prof. Y. T. Li. To Prof. Vander Velde for assuming the major responsibility for the thesis supervision and for his careful reading of, and detailed comments on, the drafts of this document; to Prof. Potter for his numerous technical suggestions and invaluable mathematical advice throughout the research effort; and to Prof. Li for improving this presentation by improving the author's perspective on the problem through his stimulating discussions of the material.

The advice and encouragement I received from Prof. A. E. Bryson of Harvard University and Dr. F. Tung of the NASA Electronics Research Center during the early stages of this work was extremely helpful.

The comments and suggestions of my colleagues throughout the research effort is greatly appreciated. In particular, I would like to thank Dr. William S. Widnall for bringing to my attention Joseph's work on Kalman filters and Kalman's work on canonic forms.

I am grateful to Messrs. Edward M. Copps and Norman E. Sears for relieving me of other responsibilities to leave time for this research and for making available to me all of the facilities of the M. I. T. Instrumentation Laboratory.

The typing of the manuscript was ably undertaken by Mrs. Nancy Jordan and Miss Susan Gallagher. Their skill and perseverance are greatly appreciated.

I would also like to express my appreciation to my wife, Joanne, for proofreading the rough drafts and final manuscript of this document and for her patience and encouragement throughout the effort.

This report was prepared under DSR Project 55-23850, sponsored by the Manned Spacecraft Center of the National Aeronautics and Space Administration through Contract NAS 9-4065.

The publication of this report does not constitute approval by the Instrumentation Laboratory or the National Aeronautics and Space Administration of the findings or the conclusions contained herein. It is published only for the exchange and stimulation of ideas.
TABLE OF CONTENTS

Chapter I  INTRODUCTION  1

1.1  A Problem Description and Investigation Summary  1

1.2  Historical Background  7

Chapter II  THE BASIC PRINCIPLES OF OPTIMAL FILTERING AND SMOOTHING  11

2.1  A Numerical Example of the Kalman Filter  11

2.2  A Numerical Example of Smoothing  17

Chapter III  A DERIVATION OF THE OPTIMAL SMOOTHING EQUATIONS FOR LINEAR DYNAMIC SYSTEMS  25

3.1  Introduction  25

3.2  A Statement of the Smoothing Problem  25

3.3  Solution to the Calculus of Variations Problem  27

3.4  A Useful Relationship Between the Forward Filter and Smoother  30

3.5  The Smoother State Estimates  32

3.6  The Estimates of the Driving Force  32

3.7  Derivation of the Bryson-Frazier Smoothed Covariance Matrix Equations  33

3.8  Derivation of the Rauch-Tung-Striebel Smoothed Covariance Matrix Equations  40

3.9  The Bryson-Frazier Covariance of the Error in the Driving Force Estimate  44

3.10  The Rauch-Tung-Striebel Covariance of the Error in the Driving Force Estimate  45

3.11  The Discrete Smoother Equations  47
<table>
<thead>
<tr>
<th>Chapter IV</th>
<th>SOLUTIONS TO SOME PROBLEMS WHICH ARISE IN THE PRACTICAL APPLICATION OF THE FILTER AND SMOOTHER EQUATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>4.2</td>
<td>A Truncation Problem Arising in the Processing of the Filter Covariance Matrix</td>
</tr>
<tr>
<td>4.3</td>
<td>Schemes for Saving Computer Storage Space</td>
</tr>
<tr>
<td>4.4</td>
<td>A Solution to a Numerical Problem Arising During Smoothing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter V</th>
<th>A SMOOTHABILITY CONDITION FOR LINEAR SYSTEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>A Statement of Smoothability</td>
</tr>
<tr>
<td>5.2</td>
<td>A Derivation of the Smoothability Condition for the Continuous Formulation</td>
</tr>
<tr>
<td>5.3</td>
<td>An Alternate Derivation of the Smoothability Condition for the Discrete Formulation</td>
</tr>
<tr>
<td>5.4</td>
<td>Discussion of the Smoothability Condition</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter VI</th>
<th>THE OPTIMUM SMOOTHER AS A COMBINATION OF TWO KALMAN FILTERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>The Form of the New Smoother Equations</td>
</tr>
<tr>
<td>6.2</td>
<td>A Continuous Form Derivation of the New Smoother Covariance Matrix Equation</td>
</tr>
<tr>
<td>6.3</td>
<td>A Continuous Form Derivation of the New Smoother State Estimation Equation</td>
</tr>
<tr>
<td>6.4</td>
<td>A More Useful Form for the New Smoother Covariance Matrix Equation</td>
</tr>
<tr>
<td>6.5</td>
<td>A More Useful Form for the New Smoother State Estimation Equation</td>
</tr>
<tr>
<td>6.6</td>
<td>The Driving Force Estimates and Associated Covariances</td>
</tr>
</tbody>
</table>
Chapter VI  THE OPTIMUM SMOOTHER AS A COMBINATION OF TWO KALMAN FILTERS (Continued)

6.7  The Discrete Form of the Two Filter Smoother  93

6.8  An Electrical Network Described by the Smoother Equations  96

6.9  A Numerical Example of the Smoother Covariance Matrix Calculation  101

6.10 Application of the Forward-Backward Smoother Equations to Linearized Nonlinear Systems  103

6.11 Summary  106

Chapter VII  A METHOD OF FILTERING WHEN THERE IS INSUFFICIENT DATA TO COMPLETELY DETERMINE THE STATE  109

7.1  Introduction  109

7.2  Development of the Forward Cofilter  110

7.3  State Estimation with Insufficient Information  115

7.4  Uniqueness of the Solution  119

7.5  Relationship to Observability  124

7.6  Proof that the Estimator is Unbiased  127

7.7  Summary  129

Chapter VIII  APPLICATION OF SMOOTHING TECHNIQUES TO A LINEAR TWO DIMENSIONAL PROBLEM  131

8.1  General Comments  131

8.2  'The System Description  131

8.3  A Summary of the Results  133

Chapter IX  APPLICATION OF SMOOTHING TECHNIQUES TO A PARAMETER IDENTIFICATION PROBLEM  143

9.1  Introductory Remarks  143
Chapter IX  APPLICATION OF SMOOTHING TECHNIQUES TO A PARAMETER IDENTIFICATION PROBLEM (Continued)

9. 2  A Description of the System  145
9. 3  The Simulation  151
9. 4  Summary of Results  152

Chapter X  CONCLUSIONS AND RECOMMENDATIONS  175

10. 1  Summary  175
10. 2  Recommendations for Further Study  176

Appendix A  SOME USEFUL STATISTICAL PROPERTIES OF THE FILTERED ESTIMATE  177

A. 1  The Error in the Filtered Estimate  177
A. 2  Covariance Between the Filtered Error at Different Times  178
A. 3  Covariance Between the Filter Error and the Measurement Noise  179
A. 4  Covariance Between the Filter Error and the Noise Driving the State  179

Appendix B  DERIVATION OF THE KALMAN FILTER STATE ESTIMATION EQUATION  181

List of References  189

Biography  195
LIST OF SYMBOLS

General Notation

An underlined symbol indicates a column vector.

A single dot over a symbol indicates the first derivative with respect to current time, \( t \).

A caret over a symbol indicates that the quantity is an estimate.

A bar over a symbol or group of symbols indicates the expected value of what is beneath it.

Symbols

- \( a \) In Chapter II this is a constant used in the linearized model of a ship's motion which involves the drag parameters and speed of the ship. In Chapter IX it is the constant part of the slope of a time varying natural frequency.

- \( a \) A vector which selects certain dimensions or combinations of dimensions in the estimation scheme of Chapter VII.

- \( A_k \) A matrix used in the backward extrapolation or the forward filter covariance matrix.

- \( b \) Solution to Eq. (7.3.2)

- \( b_p \) A particular solution to Eq. (7.3.2)
\( B_k \)  
A matrix used in the backward extrapolation of the inverse of the forward filter covariance matrix

\( c \)  
An arbitrary scalar constant

\( C_k \)  
A weighting matrix used in the Rauch-Tung-Striebel smoother scheme

\( D \)  
The hydrodynamic drag on the ship in the ship's speed example of Chapter II

\( D_k \)  
A matrix used for clerical convenience in Chapter V. Defined by Eq. (5.3.5)

\( e \)  
The error in the forward filter state estimate

\( e_{-B} \)  
The error in the backward filter state estimate

\( e_T \)  
The error in the smoother state estimate

\( E \)  
The covariance matrix of forward filter state estimation errors (continuous case)

\( E_k \)  
The covariance matrix of forward filter state estimation errors at the \( k' \)th sample after incorporation of the \( k' \)th measurement statistics (discrete case)

\( E_k' \)  
The covariance matrix of forward filter state estimation errors at the \( k' \)th sample before incorporation of the \( k' \)th measurement statistics (discrete case)

\( f, f_k \)  
The functions which extrapolate the state forward for the nonlinear case
F

The fundamental matrix of the linear state equation

g, g_k

The functions which extrapolate the state backward for the nonlinear case

G, G_k

The gain matrix which premultiplies the driving force in the linear state equation

G'

The matrix which couples the deterministic part of the control to the derivative of the state (continuous case)

G_c

The matrix which couples the driving noise to the part of the state which is controllable by this noise

G_u

The matrix which couples the deterministic part of the control to the uncontrollable part of the state

h, h_k

The functions which relate the state to the measurement in the nonlinear case

H, H_k

The matrices which rotate the state into measurement coordinates

I

The identity matrix

I_R

Moment of inertia of spacecraft in spacecraft example

J, J^*

Cost functions for the optimum smoother problem

J_k

A weighting matrix used in the extrapolation of the backward filter. Defined by Eq. (6.7.4)

k

The measurement time (sample period) index
\( K_k \)

The Kalman gain matrix for the forward filter

\( \ell \)

Thrust lever arm in spacecraft example

\( m \)

Mass. In Chapter II, the mass of the ship and in Chapter IX, the mass of the spacecraft

\( m_1 \)

A driving vector introduced for clerical convenience. Defined by Eq. (B.5)

\( m_2 \)

A driving vector introduced for clerical convenience. Defined by Eq. (B.6)

\( M \)

A state transition matrix used in the derivation of the Kalman filter presented in Appendix B. Defined by Eq. (B.4)

\( n \)

The dimension of the state

\( n_k, n_k' \)

Null vectors of \( V_k \) and \( V_k' \)

\( N \)

An index which represents the total number of sample intervals (or measurements).

\( P \)

The covariance matrix of backward filter state estimation errors (continuous case)

\( P_k \)

The covariance matrix of backward filter state estimation errors at the \( k' \)th sample after incorporation of the \( k' \)th measurement statistics (discrete case)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_k'$</td>
<td>The covariance matrix of backward filter state estimation errors at the k'th sample before incorporation of the k'th measurement statistics (discrete case)</td>
</tr>
<tr>
<td>$P_{k/N}$</td>
<td>Covariance matrix of smoother state estimation errors at the k'th sample (discrete case)</td>
</tr>
<tr>
<td>$P_T$</td>
<td>Covariance matrix of smoother state estimation errors (continuous case)</td>
</tr>
<tr>
<td>$q$</td>
<td>The dimension of the driving force</td>
</tr>
<tr>
<td>$q_b$</td>
<td>Generalized bending coordinate in spacecraft example</td>
</tr>
<tr>
<td>$Q$</td>
<td>A priori covariance matrix of the driving noise (continuous case)</td>
</tr>
<tr>
<td>$Q_k$</td>
<td>A priori covariance matrix of the driving noise at the k'th sample (discrete case)</td>
</tr>
<tr>
<td>$Q_{k/N}$</td>
<td>Covariance matrix of smoother driving force estimation errors at the k'th sample (discrete case)</td>
</tr>
<tr>
<td>$Q_T$</td>
<td>Covariance matrix of smoother driving force estimation errors (continuous case)</td>
</tr>
<tr>
<td>$r$</td>
<td>The dimension of the measurement</td>
</tr>
<tr>
<td>$R, R_k$</td>
<td>A priori covariance matrix of the measurement noise</td>
</tr>
<tr>
<td>$s$</td>
<td>A dummy time variable</td>
</tr>
<tr>
<td>$S, S_k$</td>
<td>A matrix defined for convenience as $S = H^T R^{-1} H$</td>
</tr>
<tr>
<td>$t$</td>
<td>Current time</td>
</tr>
</tbody>
</table>
$t_0$ The time at which the a priori initial conditions are given. The beginning of the data interval.

$t_1$, $t_2$ Two arbitrary times

$T$ As a superscript this indicates the matrix transpose operation. As a time variable or subscript it indicates the time of the end of the data interval.

$T_e$ Engine thrust in spacecraft and ship's speed examples.

$u$ The actual driving force

$\bar{u}, \bar{u}_k$ The a priori mean of the driving force

$\hat{u}_k/N$ The smoothed estimate of the driving force at the $k^{th}$ sample (discrete case)

$\hat{u}_T$ The smoothed estimate of the driving force (continuous case)

$U$ The inverse of $P$

$U_k$ The inverse of $P_k$

$U'_k$ The inverse of $P'_k$

$v$ In Chapter II this is the deviation between the actual speed and the mean speed of the ship in the ship's speed example. Elsewhere it is a dummy time variable.

$\hat{v}$ A non-optimum estimate of the driving force
\( v_r \) The mean speed of the ship in the ship's speed example

\( V \) The inverse of \( E \)

\( V_k' \) The inverse of \( E_k' \)

\( w, w_k, w_k' \) Variables with the dimensions of the costate which characterize the forward cofilter

\( w_{B}, w_{B_k}, w_{B_k}' \) Variables with the dimensions of the costate which characterize the backward cofilter

\( W, W_k \) Weighting matrices used in the calculation of the smoother covariance matrix. Defined by Eqs. (6.4.4) and (6.7.3)

\( x, x_k \) The actual state

\( \hat{x} \) The forward filter state estimate (continuous case)

\( \hat{x}_B \) The backward filter state estimate (continuous case)

\( \hat{x}_{B_k} \) The backward filter estimate of the state at the \( k' \)th sample after incorporation of the \( k' \)th measurement (discrete case)

\( \hat{x}_{B_k} \) The backward filter estimate of the state at the \( k' \)th sample before incorporation of the \( k' \)th measurement (discrete case)
\( x_c \) The part of the state which is controllable by the driving noise

\( \hat{x}_k \) The forward filter estimate of the state at the \( k'\)th sample after incorporation of the \( k'\)th measurement (discrete case)

\( \hat{x}_k \) The forward filter estimate of the state at the \( k'\)th sample before incorporation of the \( k'\)th measurement (discrete case)

\( \hat{x}_{k/N} \) The smoother estimate of the state at the \( k'\)th sample (discrete case)

\( x_{R_k} \) In the nonlinear problem this is the reference solution state at the \( k'\)th sample

\( \hat{x}_T \) The smoother estimate of the state (continuous case)

\( x_u \) The part of the state which is not controllable by the driving noise

\( X \) A matrix introduced for clerical convenience in Appendix B. Defined by Eq. (B.15)

\( Y \) A non-optimum state estimate

\( Y, Y_k \) A matrix defined for convenience as \( Y \equiv GG^T \)

\( Y_c \) The part of \( Y \) corresponding to the part of the state which is controllable by the driving noise

\( z, z_k \) The actual measurement
The part of the costate for the smoother problem corresponding to that section of the state which is controllable by the driving noise

\( \lambda_c \)

\( \lambda_u \)

The part of the costate for the smoother problem corresponding to that section of the state which is not controllable by the driving noise

\[ \Lambda = \lambda \lambda^T \]

Defined as:

\( \Lambda, \lambda_k \)

The driving noise

\( \mu \)

A particular dimension or combination of dimensions of the state estimate

\( \xi \)

The mean squared error in the estimation of \( \xi \)

\( \sigma^2 \)

Normalized bending slope at inertial measurement unit location in spacecraft example

\( \sigma_a \)

Normalized bending slope at engine hinge location in spacecraft example

\( \tau \)

A dummy time variable

\( \phi_e \)

Normalized bending displacement at engine hinge location in spacecraft example

\( \Phi_{cc}, \Phi_{cu}, \Phi_{uu} \)

Elements of the state transition matrix when the state equation is expressed in canonic form
\( \alpha \) A zero vector used in the derivation of the smoother state estimates in Chapter III

\( \beta \) A vector defined for clerical convenience as \( H_{k \rightarrow k} \)

\( \gamma \) A vector defined for clerical convenience as \( V_{k \rightarrow k} \)

\( \delta \) Variational notation

\( \delta(t_1 - t_2) \)

\[
\begin{align*}
\delta(t_1 - t_2) & = 0 \quad t_1 \neq t_2 \\
& = 1 \quad t_1 = t_2
\end{align*}
\]

\( \delta_c \) Commanded engine angle in spacecraft example

\( \delta_e \) Actual engine angle in spacecraft example

\( \epsilon_{\nu} \) The error in the smoother driving force estimate

\( \epsilon_T \) In the nonlinear problem this is the smoother estimate of the deviation from the reference solution. It is the smoother error state (continuous case).

\( \epsilon_\xi \) The error in the estimate of \( \xi \)

\( \zeta \) Damping ratio of a second order mode

\( \eta \) The measurement noise

\( \theta \) Angle between the rigid vehicle centerline and the inertial reference line in spacecraft example

\( \lambda, \lambda_k \) The adjoint variable for the smoother problem
\( \Phi_k \) The state transition matrix which transfers the state from sample period \( k \) to sample period \( k + 1 \) (discrete case)

\( \psi(t_1, t_2) \) The state transition matrix associated with the smoother adjoint equation

\( \omega \) Natural frequency of a second order mode

\( \varpi \) An error vector in the cofilter. Defined by Eq. (7.6.1)

\( \Omega(t_1, t_2) \) The state transition matrix associated with the differential equation describing the error in the Rauch-Tung-Striebel smoother state estimate
CHAPTER I

INTRODUCTION

1.1 A Problem Description and Investigation Summary

The problem of interpreting measurements made in the presence of disturbances with imperfect instruments has been a concern of man since he first began collecting quantitative information about the world around him. The sailor had to determine his position on the sea from the deck of a rolling ship by sighting stars with a less than perfect sextant through an atmosphere capable of deflecting and obscuring light. The physicist, astronomer, and machinist have all worked for centuries with instruments which left uncertainties in their measurements. In recent years engineers have been confronted with such problems as identifying the meaningful information in a signal which is heavily corrupted with electrical noise. In these and many more cases the error in the measurement is often larger than is acceptable.

In many situations it is possible to reduce the effects of the disturbances by processing more measurements. Using an averaging process, one can determine the quantities of interest with more confidence. A simple example of this is the chemistry student who weighs a quantity of chemicals several times on his analytical balance before adding it to a mixture. By averaging the results obtained on each weighing he reduces the possibility of basing the weight of an ingredient on a single bad measurement.

When the quantity being measured changes from one measurement time to the next, the problem of interpreting multiple measurements becomes more complex. A technician measuring the voltage in a simple RC circuit while it is decaying could not describe the voltage by averaging a series of measurements and presenting a single number as
the result. If he knows the time constant of the exponential function which characterizes the decay, he can reflect each voltage measurement back to the initial time and average all the computed values to obtain an estimate of the initial voltage. This initial voltage together with the known exponential function time constant completely describe the voltage decay. When both the initial value and the decay rate are unknown, curve fitting techniques such as least squares* may be used to obtain estimates of initial voltage and time constant.

If the time variation of the quantities being measured is describable in terms of differential equations, these may be used in the interpretation of the measurements rather than curve fitting the data. There are two widely used techniques available for doing this. The first, usually called a Wiener filter, is valid for stationary processes and time invariant systems. It is a frequency domain technique which gives an estimate based upon minimizing the mean squared steady state error. The other formulation, which is much more pertinent to this thesis, is called a Kalman filter and is a time domain solution which is valid for nonstationary statistics and time varying systems. It gives the transient as well as the steady state solution. Another important advantage of Kalman's formulation is that it is easily programmed on a digital computer. This makes it a particularly attractive technique for large problems. It is important to realize, however, that this advantage of the Kalman filter over the Wiener filter is lost if a digital computer is not available. When Kalman did his work high speed digital computers were in widespread use; when Wiener obtained his solution practical digital computers had not yet been developed.

The Kalman filter provides a method for combining in an optimum fashion all the information available up to and including the time of the latest measurement to provide an estimate at that time. In addition to the measurements, information about the dynamics of the process, statistics of the disturbances involved, and a priori

*Hildebrand (1956) gives a general treatment of such numerical data reduction methods.
knowledge of the quantities of interest are included in the problem formulation. If the dynamics can be described by linear differential of difference equations and if the disturbances have Gaussian distributions, the resulting estimate is both a maximum likelihood and minimum variance estimate. As the name suggests, a maximum likelihood estimate has a higher probability of being correct than any other. Similarly, no estimator can give an estimate which has a smaller mean squared error than the minimum variance estimate.

The Kalman filter provides a method of combining redundant data by weighting appropriately each item of data as it is combined into the estimate. Poor measurements receive proportionately less weighting than good ones. Prior information is weighted into the estimate in the same way as the measurements. If the prior information is poor relative to the measurements it will have less effect upon the estimate than if the reverse is true. If there is noise driving the system between measurement times the filter will weight the extrapolated value of the old estimate less than if there were no noise. This is because noise introduces an uncertainty in the state of the system between measurements times. Consequently the estimate will depend less upon old estimates and more upon new measurements.

Another useful property of the Kalman filter is its recursive nature. The procedure begins by combining in an optimum fashion the a priori information, which in some cases must be extrapolated forward to the first measurement time, and the first measurement. The resulting estimate is then extrapolated forward to the second measurement time using the mathematical model of the dynamic process. The mean squared error at the first measurement time is also extrapolated to the time of the second measurement so that the proper weighting can be placed on the old information. The extrapolation of these statistics is accomplished using the model of the dynamic process and the mean squared value of the driving noise. A large driving noise variance will cause an increase in the mean squared error in the estimate when it is extrapolated from
one measurement time to the next. The estimate at the second time is obtained by combining optimally the extrapolated first estimate and the second measurement. This recursive procedure is repeated until all the measurements are incorporated. At each measurement time the new estimate is formed as the optimum combination of the old and new information.

As mentioned above, the Kalman filter provides an estimate at time T given all the information up to and including time T. This is to be distinguished from a predictor or a smoother. A predictor gives an estimate for some time \( t_1 \), greater than T, based upon all the measurements up to and including time T. A smoother gives an estimate at a time \( t_2 \), less than T, using all the measurements up to time T. Smoothing is the subject of this thesis.

Smoothing requires measurements which are in the future with respect to the time of interest, \( t_2 \). The only way to obtain such information is to wait until after time \( t_2 \) when the measurements for times later than \( t_2 \) have been made. Smoothing is thus a data reduction technique.

The practical applications of smoothing are numerous. It may be used in the form presented in this thesis as a data reduction technique in any situation where a linear or linearized model of the physical situation is reasonable. A typical example of an application where smoothing is useful is that of evaluating data which has been recorded during a test flight of an aircraft or spacecraft. The engineer can use the results of such a test to evaluate system performance and to obtain numerical values for the parameters of his system model. He can also use this information as an aid in designing subsequent systems. If linear or linearized models of the system are not applicable, solutions to the optimum smoothing problem can still be obtained if the user can solve a nonlinear two point boundary value problem.
Numerical techniques are available for solving such problems*. In this case, however, the statistics which describe the estimation error, if processed, must be computed using a linear or linearized system model.

Intuitively, one would expect that the optimum smoother using data which is both past and future with respect to a particular point in time should give a better estimate at that time than the optimum filter which uses only data up to and including that time. This is indeed true and it will be shown that statistically the smoother estimate can never be worse than the filter estimate. There are two basic physical reasons for this. First, the smoother has more measurements available and can use them to reduce the effect of the measurement noise. Second, the smoother can use future measurements of outputs as well as past and present ones, all with respect to time $t_2$, to estimate inputs which occurred prior to time $t_2$. The filter, not having future measurements available, cannot do as well. Using this additional information, the smoother can do a better job of attributing some of the estimation error to poor knowledge of the input. In this way it can improve accordingly its estimate of both the output and the input.

The smoothing solutions which have been presented in the literature and which are discussed in the next section obtain the smoothed estimate as a correction to the Kalman filter estimate. The Kalman filter is processed forward until all the measurements are incorporated. At the last measurement time the filter and smoother estimates are the same because they both use all the available information. These recursive smoothing schemes then step backward from the last measurement time and form the smoothed estimate by adding a correction term to the forward filter estimate.

*See, for example, Bellman and Dreyfus (1962), or Bryson and Denham (1961).
It is also possible to obtain the smoothed estimate non recursively by batch processing the data. This estimate will be identical to that provided by the optimum recursive smoother if the same information is used in both cases. In addition to the measurements, this includes information about the statistics of the driving and measurement noises, a mathematical model of the physical process, and an a priori estimate. In order to obtain the state estimate at each measurement time by batch processing the data it is necessary to invert a square matrix of order $Nn$ where $N$ is the number of measurements and $n$ is the dimension of the state. If there are only a few data points the recursive schemes offer no advantage over batch processing. When there are a large number of measurements, however, it becomes impractical to use the batch processing method because of the computation and storage requirements associated with the matrix inversion. This is the case for which the optimum recursive smoothing schemes are most useful.

The recursive smoothing solution which is presented in this thesis differs from these schemes in that it expresses the smoothed estimate as the optimum combination of two independent Kalman filter estimates. One of these filters is the conventional forward Kalman filter which starts from a priori initial conditions. The second filter starts at the last measurement time and works recursively backward over the data. The smoother estimate at a particular point within the data interval is obtained by properly combining the two filter estimates for the same point. All the information is reflected to the point of interest; the forward filter uses all the information up to the time of interest, $t_2$, to provide an estimate at this point and the backward filter estimate at this time is based upon all the information for time greater than $t_2$. It is assumed in obtaining this solution that the noises have white time distributions.

In addition to this new formulation of the smoothing problem a new solution is presented to the problem of filtering with information insufficient to determine completely all the quantities of interest. Part
of this solution constitutes a completely recursive method of starting a Kalman filter when there is no a priori information available and is used to process the backward filter which is involved in the new smoother formulation. This backward filter is identical to the conventional forward Kalman filter in all respects except two. First, the forward extrapolation procedure mentioned earlier is replaced by a backward one. (The estimate at the \( k + 1 \) measurement time is extrapolated back to the \( k \)'th measurement time so that the \( k \)'th measurement may be incorporated.) This means that the forward transition is replaced by its inverse. Second, the backward filter has no a priori information with which to start at the terminal time. This is reflected mathematically as an infinite mean squared error in the backward filter estimate at the terminal time. Physically, this is a consequence of the fact that there is no information available for any time later than the end of the data interval.

The concept of smoothability is introduced and a smoothability condition is derived. A quantity is considered smoothable if smoothing yields an estimate of that quantity which is better than that which can be obtained by extrapolating the final forward filter estimate backward in time. A number of numerical problems are identified and solved. These techniques are then applied to two numerical examples.

1.2 Historical Background

The history of optimum smoothing is brief in two senses. It has a short history when the number of years solutions have been available is considered and by comparison to other topics in modern control theory there has been relatively little mention of it in the literature. It will be possible in this section to enumerate virtually all the significant previous work on the subject.

As was mentioned above, some of the earliest work in the field of optimal estimation was done by Wiener (1949). This work was similar to some earlier work by Kolmogorov (1941). Since the publication of Wiener's results, other authors have published other solutions
to the problem. Parzen (1961) presents a good summary of these solutions. Until 1960 the Wiener filter was the most powerful tool available for solving the optimum estimation problem. In that year, Kalman (1960) published his work which not only permitted solutions to problems with time varying systems and non stationary statistics, but also was readily adaptable to mechanization on the high speed digital computers which had been developed since the period in which Wiener did his work. Using the method of orthogonal projections, Kalman obtained solutions to the optimum filtering and prediction problems. He defined the optimum smoothing problem but did not present its solution.

Bryson and Frazier (1962) were the first to extend the work of Kalman to the optimum smoothing problem. They derived a continuous form solution to the problem using the calculus of variations. Bryson later documented the discrete solution in lecture notes for his course at Harvard University (Bryson (1964)) and in the manuscript of a book he is currently writing with Ho (Bryson and Ho (1966)).

Cox, in his doctoral thesis (Cox (1963)), and later in a subsequent paper (Cox (1964)), provides a general treatment of the estimation of state variables of nonlinear systems in the presence of Gaussian driving and measurement noises. When he restricts his smoothing solution to linear systems he obtains the results of Bryson and Frazier.

A different form for the solution to the optimum linear smoothing problem was obtained by Rauch (1963) and Rauch, Tung, and Striebel (1965). Rauch (1963) simply presents results in a short paper while Rauch, Tung, and Striebel (1965) give a complete derivation of their form of the solution starting from maximum likelihood considerations. This form of the solution expresses the smoothed estimate recursively as a combination of the previous smoothed estimate with the filter estimates before and after the measurement. No direct use is made of the
measurements during smoothing. Bryson and Frazier express the smoothed estimate as a sum of the filter estimate plus a correction term which is based upon the adjoint variable from the calculus of variations approach to the smoothing problem. This solution directly uses the measurements but avoids a matrix inversion required by the solution of Rauch, Tung, and Striebel. All these works express the smoother estimate at the k'th sample period on the basis of N measurements by assuming N is fixed and k is varying. Rauch (1963) also documents a solution where k is fixed and N varies.

Alternate derivations of the results obtained by Rauch, Tung, and Striebel have been presented by Lee (1964) and Meditch (1966). Both derivations are for the discrete formulation. Lee used the calculus of variations to obtain his solution and Meditch used the method of orthogonal projections which Kalman (1960) used in obtaining his solution to the filtering and prediction problem.

These papers constitute the bulk of the published work on the theory of optimum smoothing. One more paper is work mentioning, however, because it represents an attempt to perform optimum smoothing when the process statistics are undefined. This is the work of Johansen (1965). Estimation with poorly defined or unknown statistics is important because it relieves the user of the necessity to specifically define the statistics of the noise involved and thus makes the solution less sensitive to errors in these definitions. In this paper Johansen obtains some very limited solutions to the smoothing problem by minimizing the maximum error which can be obtained when the driving noise is bounded. In this and another paper (Johansen (1966)) he treats the problem of filtering when the process statistics are undefined.

In all of the above literature the emphasis is on the theory rather than the application of optimal smoothing techniques. Some of the authors have included simple numerical illustrations of smoothing but these are mostly academic examples. To the knowledge of this author, the only work devoted specifically to the application of optimum
smoothing techniques to a real problem is that of Rauch (1965). In this paper he demonstrates the application of smoothing to the non-linear problem of estimating satellite trajectories in the presence of random fluctuations in drag.
CHAPTER II

THE BASIC PRINCIPLES OF OPTIMAL FILTERING AND SMOOTHING

2.1 A Numerical Example of the Kalman Filter

An understanding of the Kalman filtering technique is essential to the comprehension of the optimum smoothing methods considered in this thesis. We will in this section consider a scalar example of its use and explore the basic physical concepts involved. In the next section we will reconsider this example in the same way to study the improvement smoothing offers over filtering. Both here and in the next section we will consider only the variance equations. In practice, of course, we would also process the state estimate. We can compare the quality of the smoothed and filtered estimates by comparing the corresponding mean squared errors - something we could not do by examining the differences between particular smoother and filter estimates. For this reason we will emphasize the statistics. The physical intuition which we will apply to understanding the variance equations can also be used to understand the state estimation procedure.

We will consider the expected errors involved when a navigator attempts to measure the speed of his ship using the ship's velocity log. We will assume that he takes three measurements at equal intervals and that during the time these measurements are made the ship is being driven in a random manner by the action of waves. We will use a linearized model of the motion of the ship:

\[ \dot{v} = a v + u \]

where \( v \) is the difference between the actual speed of the ship and the
mean speed, \( u \) is the random acceleration cause by wave action, and \( a \) is a constant which results from linearizing the nonlinear equation of motion.\(^*\)

We will assume that the captain of the ship gives the navigator an estimate of the ship's speed based upon his experience. His uncertainty in that estimate may be expressed as a mean squared error of 3 knots\(^2\). After this, all speed information comes from the ship's velocity log. In addition, we will assume the variance of expected errors in measurements taken with the velocity log to be 1 knot\(^2\) and that the integrated effect over the constant time between measurements of the driving noise of the waves has a mean squared value of 1 knot\(^2\).

The scalar Kalman filter formula for the variance after a discrete measurement is incorporated is:

\[
E_k = \frac{E_k'}{E_k' + R} \quad (2.1.1)
\]

where \( E_k' \) and \( E_k \) are respectively the variance of the estimation error before and after incorporation of the \( k \)'th measurement and \( R \) is the variance of the measurement noise. This is also the maximum likelihood formula for combining the variances from two independent error sources.

If the navigator's initial estimate of speed was very good in comparison to the data from the ships velocity log (\( R \) much larger than \( E_k' \)) Eq. (2.1.1) would reduce to:

\[
T_e - D \quad (T_e \) is the thrust provided by the engine through the propellors, \( D \) is the hydrodynamic drag, and \( m \) is the mass of the ship) with respect to the speed, \( v \), evaluated at the mean speed, \( v_m \). The drag and thrust are functions of the speed. For details on these functional relationships see Hunsaker and Rightmire (1947).

\(^*\) The constant, \( a \), is the partial derivative of \( \frac{m}{v} \) (\( T_e \) is the thrust provided by the engine through the propellors, \( D \) is the hydrodynamic drag, and \( m \) is the mass of the ship) with respect to the speed, \( v \), evaluated at the mean speed, \( v_m \). The drag and thrust are functions of the speed. For details on these functional relationships see Hunsaker and Rightmire (1947).
\[ E_k' \approx E_k \]

which means that the variance before and after the measurement is approximately the same. In this case the measurement data has been rejected because it is so noisy as to be virtually useless. Since no new information has been added the variance remains the same. In the reverse case (\( R \) much smaller than \( E_k' \)) Eq. (2.1.1) reduces to:

\[ E_k' \approx R \]

which means that prior information is so poor compared to the measurement that it is rejected and the estimate is made solely on the basis of the measurement.

If \( E_k' = R \) the two variances are weighted equally and the resulting mean squared error is half of the two equal variances.

For the first measurement we use the mean squared value of the error in the a priori estimate (\( E' = 3 \)) and our value of \( R = 1 \). The result, using Eq. (2.1.1), is:

\[ E_1 = 0.75 \text{ Knot}^2 \]

Note that this is less than each of the variances which combine to determine it. This is because there is more information in the two sources (prior information and new measurement) than in each alone. Consequently, the resulting uncertainty in the estimate is smaller than that which we could obtain with each data source alone.

Before considering the extrapolation of this mean squared error to the second measurement time we will consider briefly the multidimensional equivalent of Eq. (2.1.1). The general Kalman filter update formula is:
\[ E_k = E_k' - E_k' H_k^T (H_k E_k' H_k^T + R_k)^{-1} H_k E_k' \]  \hspace{1cm} (2.1.2)

where all variables are matrices and \( H_k \) relates the state to the measurement. We can reduce Eq. (2.1.2) to our scalar formula by setting \( H_k = 1 \) and \( R_k = R \), and by considering all variables to be scalars:

\[ E_k = E_k' - \frac{E_k'^2}{E_k' + R} \]  \hspace{1cm} (2.1.3)

Collecting terms in Eq. (2.1.3) yields:

\[ E_k = \frac{E_k' R}{E_k' + R} \]

which is Eq. (2.1.1). The same physical intuition which we applied to Eq. (2.1.1) also is valid for interpreting the multidimensional case described by Eq. (2.1.2). In the multidimensional case, however, more than one measurement may be needed to obtain information about some of the quantities of interest. As an example of this, consider the fact that it takes two position measurements to make an estimate of both position and velocity.

In order to use Eq. (2.1.1) at the second measurement time we must first extrapolate the mean squared error in the estimate at the first measurement time forward to the time of the second measurement. The discrete scalar Kalman filter equation for extrapolating the variance from one measurement time to the next is:

\[ E_{k+1}' = \Phi^2 E_k + Q \]  \hspace{1cm} (2.1.4)
where $Q$ is the variance of the integrated effect of the driving noise, and:

$$\Phi = e^{at}$$

where $t$ is the time between measurements.

The solution to the equation:

$$\dot{v} = a v$$

is:

$$v(t_0 + t) = e^{at} v(t_0) = \Phi v(t_0)$$

where $t_0$ is the initial time. Thus $\Phi$ extrapolates the speed from one time to the next when the effects of the waves are ignored. In the multidimensional case $\Phi$ is called the state transition matrix.

Note that because variances are always positive $Q$ always causes $E_{k+1}$ to be larger than it would be if there were no driving noise. The meaning of this is that the driving noise introduces an uncertainty into how the estimate should propagate from one time to the next. In this example the waves change the speed of the ship between measurement times so we lose some of the confidence we had in our estimate at the previous measurement time.

The other term, $\Phi^2 E_k$, enables us to use our knowledge of the physical situation to decide how the variance should change from one time to the next. In our case $a$ is negative, hence $\Phi$ is less than unity. Consequently:

$$\Phi^2 E_k < E_k$$
The reason for this is that we have here a stable system in which errors die out in time. Indeed, if the waves were suddenly to disappear the ship would eventually return to its mean speed. After this happens we would have no error in our estimate at all. An unstable system would be characterized by a $\Phi$ which is larger than unity so the error would grow in time.

Equation (2.1.4) is very similar to the multidimensional covariance matrix extrapolation equation:

$$E'_{k+1} = \Phi_k E_k \Phi_k^T + Q_k$$

The intuitive reasoning which applied to Eq. (2.1.4) is also applicable to Eq. (2.1.5). Note that in the general multidimensional case (Eq. (2.1.5)) all variables are matrices and that the state transition matrix $\Phi_k$, and the covariance matrix of the driving noise, $Q_k$, can change from one sample period to the next.

Returning again to the scalar ship's speed example we may write, using Eq. (2.1.4) and $\Phi^2 = .5$:

$$E'_2 = 1.375 \text{ Knot}^2$$

Equations (2.1.1) and (2.1.4) describe a recursive formulation of the Kalman filter. To obtain $E_2$ we will use $E'_2$ and Eq. (2.1.1). Each piece of information is treated only once. The entire process is begun, as we did here, with the a priori information. The results of processing the recursive equations for the three measurements assumed here are summarized in Table 2.1. If we were to carefully examine the transient nature of the mean squared error we would discover that the variance decays approximately twice as fast as an exponential with time constant $1/a$ would. This is because $\Phi^2$ in Eq. (2.1.4) is $e^{-2a t}$.
The last entry in Table 2.1 gives the steady state variances. These are the mean squared errors we would obtain if we were to incorporate an infinite number of measurements. They are obtained by setting the variance at sample period \( k+1 \) to the corresponding variance at the \( k' \)th sample period.

<table>
<thead>
<tr>
<th>Measurement Time</th>
<th>Mean Squared Error Before the Measurement (Knots(^2))</th>
<th>Mean Squared Error After the Measurement (Knots(^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.0000</td>
<td>.7500</td>
</tr>
<tr>
<td>2</td>
<td>1.3750</td>
<td>.5789</td>
</tr>
<tr>
<td>3</td>
<td>1.2894</td>
<td>.5632</td>
</tr>
<tr>
<td>( \infty )</td>
<td>1.2807</td>
<td>.5615</td>
</tr>
</tbody>
</table>

2.2 A Numerical Example of Smoothing

In this section we will reconsider the ship's speed example of the previous section to demonstrate the improvement offered by smoothing.

We begin the smoothing process by noting that the boundary condition on the backward filter variance \( P_3' \) is:

\[
U_3' \Theta P_3'^{-1} = 0
\]

This infinite mean squared error is a consequence of the lack of any "prior" (future with respect to the end of the interval in which velocity measurements were taken) information about the speed at the last
measurement time. The meaning of this is that nothing is known about the speed for times later than the third measurement time. Note the use of the symbol \( U \) to denote \( P^{-1} \).

The formula for the mean squared error in the smoothed estimate at the last measurement time is:

\[
P_{3/3}^{-1} = E_{3}^{-1} + P_{3}^{-1} = \frac{1}{5632} + 0
\]

thus

\[
P_{3/3} = .5632 \text{ Knots}^2
\]

The symbol \( P_{3/3} \) denotes the mean squared error in the smoother estimate at the terminal time. The more general notation is \( P_{k/N} \) where \( k \) indicates which measurement time this variance is for and \( N \) indicates the total number of measurements. Note that \( P_{3/3} \) is the same as \( E_{3} \). The reason for this is that at the terminal time the backward filter has no information (hence infinite mean squared error) until after incorporation of the terminal measurement, if any. It therefore cannot offer any information not already contained in the forward filter estimate. Hence the smoother estimate is based entirely upon the forward filter estimate and the mean squared error in the smoother and forward filter estimates are thus identical.

The next step in the smoothing process is to update the inverse of the backward filter variance. The general update formula for this scalar example is:

\[
P_{k}^{-1} = P_{k}^{-1} + R^{-1}
\]  \hspace{1cm} (2.2.1)

Equation (2.2.1) is identical in form to Eq. (2.1.1). This fact is more obvious if we rewrite Eq. (2.2.1) in the form:
\[
P_k = \frac{P'_k R}{P'_k + R}
\]

Inasmuch as Eqs. (2.2.1) and (2.1.1) are identical in form, all the physical intuition we applied to Eq. (2.1.1) may also be used to understand Eq. (2.2.1). With \( R = R_k \) and all the variables matrices instead of scalars Eq. (2.2.1) is the multidimensional backward filter covariance matrix update formula. The same physical reasoning is also applicable in the multidimensional case.

Using Eq. (2.2.1) at the last measurement time we have:

\[
P_{3}^{-1} = P_3'^{-1} + R^{-1} = 0 + 1 = 1
\]

so:

\[
P_{3}^{-1} = U_3 = 1 \text{ Knot}^{-2}
\]

Before we can obtain the mean squared error in the smoothed estimate at the second measurement time we must extrapolate the backward filter variance to the time of the second measurement. The general scalar formula for doing this is:

\[
U_{k-1}' = \frac{\phi^2 U_k}{(1 + Q U_k)} \quad (2.2.2)
\]

Much of the same physical reasoning which we applied to the understanding of Eq. (2.1.4) can be used to understand Eq. (2.2.2). First note that as \( Q \) increases, \( U_{k-1}' \) decreases. The reason for this
is that the action of the waves changes the speed of the ship in an unknown manner between measurements and decreases the navigator's confidence in his old measurements of speed. This decrease in confidence is reflected in a larger \( P'_{k-1} \) hence smaller \( U'_{k-1} \).

For stable systems (stable in the direction of increasing time like the speed of our ship) \( \Phi \) is less than unity and will cause \( U'_{k-1} \) to be smaller than \( U_k \). The reason for this is that systems which are stable with time running forward are unstable with time running backward. Uncertainties which exist at one time in an unstable system will grow in magnitude in the direction in which the system is unstable. For the backward filter and \( \Phi \) less than one this means that \( P'_{k-1} \) will be larger than \( P_k \), hence \( U'_{k-1} \) will be smaller than \( U_k \). This is the case for our ships speed example.

Using Eq. (2.2.2) and our numerical values we obtain:

\[
U'_2 = \frac{\Phi^2 U_3}{(1 + Q U_3)} = \frac{(.5)1}{(1 + 1)} = .25 \text{ Knots}^{-2}
\]

The expression for the mean squared error in the smoother estimate at the \( k' \)th sample is:

\[
P'_{k/N} = E_k^{-1} + P_k^{-1} = E_k^{-1} + U_k'
\]  \hspace{1cm} (2.2.3)

Note that \( E_k \) is based upon all the error information available up to and including the mean squared error in the \( k' \)th measurement. This includes the error in the a priori estimate. \( P'_{k} \) is based upon all the information from the end of the data interval back to the \( k' \)th measurement time but not including the mean squared error in the \( k' \)th measurement. Together these two variances include exactly all the available error information, no more and no less. An entirely equivalent expression is:
\[
P^{-1}_{k/N} = E^{-1}_k + U_k
\]

In this expression the mean squared error in the \(k\)'th measurement is included in the backward filter variance instead of in the forward filter variance.

Next observe the similarity in form between Eqs. (2.2.1) and (2.2.3). The reason for this is that they both are the maximum likelihood formulas for combining the errors in two independent estimates. The mean squared errors in the two filter estimates combine to form the mean squared error in the smoother estimate in exactly the same way the mean squared errors in the measurement and prior estimate combined to form the mean squared error in the filter estimate. When the forward filter estimate is much better than the backward filter estimate (\(E_k\) much smaller than \(P_k'\)) the smoother estimate is based heavily upon the forward filter estimate and very little upon the backward filter estimate. The corresponding variance of errors in the smoother estimate is thus approximately equal to the variance of forward filter errors. The reverse is true when the backward filter estimate is much better than the forward filter estimate (\(P_k'\) much smaller than \(E_k\)). When the two filters have the same variances the mean squared error in the smoother is half that of the filters.

As a final observation about Eq. (2.2.3) note that the mean squared error in the smoother estimate is never larger than the mean squared error in either of the filter estimates. This is mathematically a consequence of the fact that both \(E_k\) and \(P_k'\) are greater than or equal to zero so \(P_{k/N}\) can never be larger than the smaller of the two. This is also a property of the optimality of the solution. Both filters and the smoother give the optimum estimate based upon the information available to them. The smoother, which has available to it the same information each of the filters uses, can not provide an estimate which has a larger mean squared error than either of the filters or its estimate would not be optimum.
Substitution of our numerical values into Eq. (2.2.3) yields:

\[ P_{2/3}^{-1} = E_2^{-1} + U_2 = \frac{1}{0.3789} + 0.25 = 1.977 \]

\[ P_{2/3} = 0.5057 \text{ Knots}^2 \]

Equations (2.2.1), (2.2.2) and (2.2.3) together with the results of forward filtering provide a recursive method of obtaining the smoothed estimate at each measurement time. Table 2.2 summarizes the results of applying this recursive scheme to the ship's speed example. Note that at all measurement times except the last the smoother gives a better estimate than either filter. At the last measurement time, the smoother and forward filter have identical errors. As was mentioned above, this is because they both use exactly the same information to form the estimate at this point.

If we had used a much larger number of measurements the backward filter would also have reached a steady state mean squared error. For this example this would have been 2.0 knots\(^2\) before the measurement. If both the forward and backward filters reached steady state in an overlapping region of sample periods the smoother would have a steady state mean squared error in this region. For this example the steady state smoother mean squared value would have been 0.4384 knot\(^2\).
<table>
<thead>
<tr>
<th>Measurement Time</th>
<th>Mean Squared Error in Forward Filter Estimate After Measurement (knots$^2$)</th>
<th>Mean Squared Error in Backward Filter Estimate Before Measurement (knots$^2$)</th>
<th>Mean Squared Error in Smoother Estimate(knots$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.7500</td>
<td>3.601</td>
<td>.6206</td>
</tr>
<tr>
<td>2</td>
<td>.5789</td>
<td>4.000</td>
<td>.5057</td>
</tr>
<tr>
<td>3</td>
<td>.5632</td>
<td>$\infty$</td>
<td>.5632</td>
</tr>
</tbody>
</table>
CHAPTER III

A DERIVATION OF THE OPTIMAL SMOOTHING EQUATIONS
FOR LINEAR DYNAMIC SYSTEMS

3.1 Introduction

The purpose of this chapter is to present a unified derivation of the optimal smoothing equations which form the basis for this investigation. The two best known solutions, that of Bryson and Frazier and that of Rauch, Tung, and Striebel, are derived from the common starting point of the least squares formulation.

The derivation presented here differs somewhat from those found in the literature primarily because the purpose here is to illustrate how both solutions follow from the same initial formulation. For alternate derivations of the method of Bryson and Frazier see Bryson and Frazier (1962) or Cox (1964). Rauch, Tung, and Striebel (1965), Lee (1964) and Meditch (1966) have all presented derivations of the solution of Rauch, Tung, and Striebel.

3.2 A Statement of the Smoothing Problem

Given:

1. A linear system described by the differential equation:

\[ \dot{x}(t) = F(t) x(t) + G(t) u(t) \]  (3.2.1)

2. A priori information about the initial statistics of the state estimate:

\[ \hat{x}(t_0) = \bar{x}(t_0) \]  (3.2.2)
\[ E(t_0) = \left[ \overline{x}(t_0) - x(t_0) \right] \left[ \overline{x}(t_0) - x(t_0) \right]^T \]  \tag{3. 2. 3}

3. A priori information about the statistics of the driving force \( \mu(t) \):

\[
\mu(t) = \overline{\mu}(t) + \mu(t) \quad \tag{3. 2. 4}
\]

\[
\mu(t) \mu^T(\tau) = Q(t) \delta(t - \tau) \quad \tag{3. 2. 5}
\]

Note that with this definition any random disturbances which are not part of a deterministic control effort are part of \( \mu(t) \). The corresponding dimensions of \( \overline{\mu}(t) \) would then be zero.

4. The measurement:

\[
\overline{z}(t) = H(t) x(t) + \eta(t) \quad \tag{3. 2. 6}
\]

where the measurement sequence \( \overline{z}(t) \) is known.

5. A priori information about the statistics of the measurement noise:

\[
\overline{\eta}(t) = 0
\]

\[
\eta(t) \eta^T(\tau) = R(t) \delta(t - \tau) \quad \tag{3. 2. 7}
\]

\[
\eta(t) \mu^T(\tau) = 0
\]

Find:

The best estimate of the state \( \hat{x}_T(t) \) in the least squares sense using the entire sequence of measurements \( z(t) \).
Note that if the driving disturbances have Gaussian distributions the best estimate in the least squares sense for this linear problem is equivalent to the maximum likelihood estimate and the minimum variance estimate.

3.3. Solution to the Calculus of Variations Problem

The optimal smoothed estimates in the least squares sense of the state and control \( \hat{X}_T(t) \) and \( \hat{U}_T(t) \) are those functions \( \chi(t) \) and \( \psi(t) \) which minimize the cost:

\[
J(t) = \frac{1}{2} \left[ \chi(t_0) - \overline{x}(t_0) \right]^T E^{-1}(t_0) \left[ \chi(t_0) - \overline{x}(t_0) \right] + \frac{1}{2} \int_{t_0}^T \left[ \left[ z(t) - H(t) \chi(t) \right]^T R^{-1}(t) \left[ z(t) - H(t) \chi(t) \right] \right. \\
\left. + \left[ \psi(t) - \overline{u}(t) \right]^T Q^{-1}(t) \left[ \psi(t) - \overline{u}(t) \right] \right] dt
\]

\[ (3.3.1) \]

subject to the constraint:

\[
\dot{\chi}(t) = F(t) \chi(t) + G(t) \psi(t)
\]

\[ (3.3.2) \]

Note the physical meaning of the individual terms of Eq. (3.3.1). Each term in square brackets is the difference between an estimate of a particular quantity and the best prior knowledge of that same quantity. For example, the first set of square brackets contains the difference between the estimate of the initial state and the a priori knowledge of the initial state. Each inner product is weighted with the corresponding confidence (inverse of a covariance matrix) in that difference. For the first term this confidence is the inverse of the a priori covariance matrix of state estimation errors. The resulting cost is thus normalized with respect to these confidences.
To solve this problem adjoin the constraint Eq. (3.3.2) to the cost Eq. (3.3.1). The result is a new cost:

\[ J'(t) = J(t) + \lambda^T(t) \left[ \dot{y}(t) - F(t) \chi(t) - G(t) \nu(t) \right] \tag{3.3.3} \]

Now consider the effect of variations of \( y(t) \) and \( \nu(t) \) around the optimum values \( \hat{\chi}_T(t) \) and \( \hat{\nu}_T(t) \). For convenience the time notation will be suppressed at this point on all but boundary conditions:

\[
\delta J' = \delta \chi^T(t_0) E^{-1}(t_0) \left[ \dot{y}(t_0) - \dot{\chi}(t_0) \right] + \int_{t_0}^{T} \left\{ \delta \chi^T \left[ -H^T R^{-1} z + H^T R^{-1} H \chi \right] \right. \\
+ \delta \nu^T \left[ Q^{-1} \nu - Q^{-1} \bar{\nu} \right] + \frac{\lambda^T}{\chi} \left[ \delta \dot{\chi} - F \delta \chi - G \delta \nu \right] \left\} \right\} dt
\]

Integrating the term \( \lambda^T \delta \dot{\chi} \) by parts and transposing the result yields:

\[
\delta J' = \delta \chi^T(t_0) E^{-1}(t_0) \left[ \dot{y}(t_0) - \dot{\chi}(t_0) \right] + \left[ \delta \chi^T \lambda \right]_{t_0}^{T} T \int_{t_0}^{T} \left\{ \delta \chi^T \left[ -H^T R^{-1} z \right. \\
+ H^T R^{-1} H \chi - F^T \lambda - \dot{\lambda} \right) + \delta \nu^T \left[ Q^{-1} (\nu - \bar{\nu}) - G^T \lambda \right] \right\} \right\} dt
\tag{3.3.4}
\]

We may set \( \delta J' = 0 \) by demanding the coefficient of each variation in Eq. (3.3.4) to be zero. The result is:

\[
\hat{\nu}_T = \bar{\nu} + Q G^T \lambda \tag{3.3.5}
\]

\[
\dot{\lambda} = -F^T \lambda + H^T R^{-1} H \hat{\chi}_T - H^T R^{-1} z \tag{3.3.6}
\]
\[ \hat{x}_T(t_0) = \bar{x}(t_0) + E(t_0) \lambda(t_0) \]  \hspace{1cm} (3.3.7)

\[ \lambda(T) = 0 \]  \hspace{1cm} (3.3.8)

where \( \hat{u}_T \) is the smoothed estimate of the driving force and \( \hat{x}_T \) is the smoothed estimate of the state.

Equations (3.3.5) to (3.3.8) together with the constraint equation:

\[ \dot{\hat{x}}_T = F \hat{x}_T + G \hat{u}_T \]  \hspace{1cm} (3.3.9)

define the solution to the smoothing problem in terms of a two point boundary value problem.

For this linear problem both boundary conditions may be obtained at the terminal time by realizing that the best estimate in the least squares sense at time \( T \) given the same information listed in Section 3.2 and all the measurements up to time \( T \) is the Kalman filter estimate at time \( T \). Thus the boundary conditions are:

\[ \hat{x}_T(T) = \check{x}(T) \]  \hspace{1cm} (3.3.10)

\[ \lambda(T) = 0 \]

where \( \check{x}(T) \) is the forward filter state estimate at time \( T \).

Substituting Eq. (3.3.5) into Eq. (3.3.9) we obtain a more convenient form for the smoothing equations:

\[ \frac{d}{dt} \begin{bmatrix} \hat{x}_T \\ \lambda \end{bmatrix} = \begin{bmatrix} F & Y \\ S & -F^T \end{bmatrix} \begin{bmatrix} \hat{x}_T \\ \lambda \end{bmatrix} + \begin{bmatrix} G \hat{u} \\ -H^T R^{-1} z \end{bmatrix} \]  \hspace{1cm} (3.3.11)

where \( S = H^T R^{-1} H \) and \( Y = G Q G^T \).
Equations (3.3.11) and (3.3.12) are to be integrated backward from the terminal conditions specified by Eq. (3.3.10).

Before we proceed to the derivation of the two major solutions to Eq. (3.3.11) we should note the content of Eq. (3.3.5). This equation, which resulted quite naturally from the solution to the calculus of variation problem, gives directly a method for estimating the driving force. The physical meaning of this is that the smoother is using the measurements of future outputs to infer what current and past inputs were. It is attributing some of the estimation error to poor knowledge of the input and is attempting to improve accordingly the estimate of the input.

When the time argument of Eqs. (3.3.11) and (3.3.12) is the terminal time T, we have an expression for the forward filter estimate (see Eq. (3.3.10)). By considering T to be variable and using the boundary conditions given by Eqs. (3.3.7) and (3.3.8) we may derive the equations which describe the forward filter estimate from Eqs. (3.3.11) and (3.3.12) - the same equations which will yield the smoother equations. This is done in Appendix B.

3.4 A Useful Relationship Between the Forward Filter and Smoother

In this section we will derive an expression which will simplify the reduction of Eqs. (3.3.11) and (3.3.12) to the more familiar forms of the smoother equations. We will show here that:

\[ E(t) \lambda(t) = \hat{x}_T(t) - \hat{x}(t) \]  \hspace{1cm} (3.4.1)

where \( \hat{x}(t) \) is the forward filter state estimate and \( E(t) \) is the covariance matrix of forward filter state estimation errors.

Equation (3.4.1) is true at the terminal time because:

\[ \lambda(T) = 0 \]

\[ \hat{x}_T(T) = \hat{x}(T) \]
To show that Eq. (3.4.1) is true at all other times define the variable:

$$\alpha = \hat{\alpha} \left( \hat{x}_T - \hat{x} - E \lambda \right)$$  \hspace{1cm} (3.4.2)

Because Eq. (3.4.1) is true at the terminal time $\alpha(T) = 0$.

Differentiating Eq. (3.4.2) yields:

$$\dot{\alpha} = \hat{\alpha} - \hat{x} - \hat{E} \lambda - E \dot{\lambda} .$$  \hspace{1cm} (3.4.3)

Expressions for $\dot{E}$ and $\dot{\lambda}$ have been derived by many authors. See for example Bryson and Frazier (1962). They are:

$$\dot{E} = FE + EF^T + Y - ESE$$  \hspace{1cm} (3.4.4)

$$\dot{\lambda} = F \hat{\lambda} + G \dot{u} + E H^T R^{-1} (z - H \hat{\lambda})$$  \hspace{1cm} (3.4.5)

Substituting Eqs. (3.3.11), (3.3.12), (3.4.4) and (3.4.5) into Eq. (3.4.3) and cancelling terms yields:

$$\dot{\alpha} = F \hat{x}_T - F \hat{x} - FE \lambda - E S \hat{x}_T + ES \hat{\lambda} + E S E \lambda$$  \hspace{1cm} (3.4.6)

Substituting Eq. (3.4.2) into Eq. (3.4.6) leaves:

$$\dot{\alpha} = (F - E S) \alpha$$  \hspace{1cm} (3.4.7)

Equation (3.4.7) shows that $\alpha$ satisfies an undriven linear differential equation. Since $\alpha(T) = 0$, $\alpha(t)$ must be zero everywhere. Thus Eq. (3.4.1) is true for all values of time.
3.5 The Smoother State Estimates

With the result of the previous section the smoother state estimates are very easily obtained. The Bryson-Frazier estimate is the result of rearranging Eq. (3.4.1):

\[ \hat{x}_T = \hat{x} + E \lambda \]

(3.5.1)

Substituting Eq. (3.5.1) into Eq. (3.3.12) yields the corresponding equation for the adjoint variable:

\[ \dot{\lambda} = -(F - ES)^T \lambda - H^T R^{-1} (z - H \hat{x}) \]

(3.5.2)

Equations (3.5.1) and (3.5.2) are the Bryson-Frazier formulation of the continuous smoother equations. Equation (3.5.2) is integrated backward from the terminal condition \( \lambda(T) = 0 \).

To obtain the Rauch-Tung-Striebel state estimate first express Eq. (3.4.1) in the form:

\[ \lambda = E^{-1} (\hat{x}_T - \hat{x}) \]

(3.5.3)

Substitution of Eq. (3.5.3) into Eq. (3.3.11) yields the Rauch-Tung-Striebel continuous smoothed state estimation equation:

\[ \hat{x}_T = F \hat{x}_T + G \hat{u} + Y E^{-1} (\hat{x}_T - \hat{x}) \]

(3.5.4)

Equation (3.5.4) is integrated backward from the terminal condition \( \hat{x}_T(T) = \hat{x}(T) \).

3.6 The Estimates of the Driving Force

The estimate of the driving force for the method of Bryson and
Frazier is given by Eq. (3.3.5). The adjoint variable is calculated from Eq. (3.5.2).

To obtain the driving force estimate in the formulation of Rauch, Tung, and Striebel substitute Eq. (3.5.3) into Eq. (3.3.5). The result is:

\[
\hat{u}_T = \overline{u} + QG^TE^{-1}(\hat{x}_T - \hat{x})
\]  

(3.6.1)

3.7 Derivation of the Bryson-Frazier Smoothed Covariance Matrix Equations

These equations may be derived directly from the state estimation formulas presented in Section 3.5.

Begin by making the definitions:

\[
e_T \triangleq x - \hat{x}_T
\]  

(3.7.1)

\[
e \triangleq x - \hat{x}
\]  

(3.7.2)

\[
P_T \triangleq e_T e_T^T
\]  

(3.7.3)

\[
E \triangleq e e^T
\]  

(3.7.4)

From Eqs. (3.7.1), (3.7.2) and (3.5.1) we have

\[
e_T = e - E \lambda
\]  

(3.7.5)

Applying Eq. (3.7.3) and Eq. (3.7.4) to this result gives:
\[ P_T = E - e e^T E - \lambda e \lambda^T E - E \lambda e e^T + E \lambda \lambda^T E \]  
\[ \text{(3.7.6)} \]

Substituting \( e \) from Eq. (3.7.5) into the second and third terms of Eq. (3.7.6) yields:

\[ P_T = E - E \lambda \lambda^T E - E \lambda e e^T - E \lambda e e^T \]  
\[ \text{(3.7.7)} \]

Now consider one of the cross terms, \( e e^T \lambda^T \):

Combining Eqs. (3.5.1) and (3.3.11) gives:

\[ \dot{x}_T = F \dot{x}_T + G \bar{u} + Y E^{-1} (\dot{x}_T - \hat{x}) \]  
\[ \text{(3.7.8)} \]

From Eq. (3.2.1) and Eq. (3.2.4) we have:

\[ \dot{x} = F x + G \bar{u} + G \mu \]  
\[ \text{(3.7.9)} \]

Combining the derivative of Eq. (3.7.1) with Eq. (3.7.8) and Eq. (3.7.9) gives:

\[ \dot{e}_T = F e_T + G \mu - Y E^{-1} (e - e_T) \]  
\[ \text{or} \]

\[ \dot{e}_T = (F + Y E^{-1}) e_T + G \mu - Y E^{-1} e \]  
\[ \text{(3.7.10)} \]

*Equation (3.7.8) is the Rauch-Tung-Striebel state estimate equation. To be consistent with the scheme of deriving each set of equations separately the Bryson-Frazier result of (3.5.1) is used to generate (3.7.8) here.
Define $\Omega(t_1, t_2)$ as the solution to:

$$\frac{d}{dt} \begin{bmatrix} \Omega(t_1, t_2) \end{bmatrix} = \begin{bmatrix} F(t_1) + Y(t_1) E^{-1}(t_1) \end{bmatrix} \Omega(t_1, t_2)$$

(3.7.11)

Then the solution to Eq. (3.7.10) is:

$$e_T(t) = \Omega(t, T) e(T) - \int_t^T \Omega(t, s) \begin{bmatrix} G(s) \mu(s) - Y(s) E^{-1}(s) e(s) \end{bmatrix} ds$$

(3.7.12)

In writing Eq. (3.7.12) we have used the fact that the filter and smoother errors are identical at the terminal time:

$$e_T(T) = e(T)$$

Substituting Eqs. (3.2.6) and (3.7.2) into Eq. (3.5.2) gives:

$$\dot{\lambda} = - (F^T - S E) \lambda - H^T R^{-1} (H e + \eta)$$

(3.7.13)

If we define $\Psi(t_1, t_2)$ as the solution to:

$$\frac{d}{dt} \begin{bmatrix} \Psi(t_1, t_2) \end{bmatrix} = - \begin{bmatrix} F(t_1) - S(t_1) E(t_1) \end{bmatrix} \Psi(t_1, t_2)$$

(3.7.14)

then the solution to (3.7.13) is:
\[ \lambda(t) = \psi(t, T) \lambda(T) + \int_t^T \psi(t, s) \left[ S(s) \varepsilon(s) + H^T(s) R^{-1}(s) \eta(s) \right] ds \]

Applying the terminal condition on \( \lambda \) leaves:

\[ \lambda(t) = \int_t^T \psi(t, s) \left[ S(s) \varepsilon(s) + H^T(s) R^{-1}(s) \eta(s) \right] ds \]

(3.7.15)

From Eqs. (3.7.12), (3.7.15) and (3.2.7) we may write:

\[ e_T(t) \lambda_T(t) = \Omega(t, T) \int_t^T \left[ e(T) e_T(s) S(s) + e(T) \eta_T(s) R^{-1}(s) H(s) \right] \psi^T(t, s) ds \]

\[ - \int_t^T ds \Omega(t, s) \int_v^v \left[ G(s) \mu(s) e^T(v) S(v) - Y(s) E^{-1}(s) e(s) e^T(v) S(v) \right. \]

\[ \left. - Y(s) E^{-1}(s) e(s) \eta^T(v) R^{-1}(v) H(v) \right] \psi^T(t, v) \]

(3.7.16)

Substituting the various filter covariance results from Appendix A into Eq. (3.7.16) gives:
\[ e_{T}(t) \lambda^{T}(t) = \Omega(t, T) \int_{t}^{T} \left[ \psi^{T}(s, T) E(s) S(s) - \psi^{T}(s, T) E(s) S(s) \right] \psi^{T}(t, s) \, ds \]

\[ - \int_{t}^{T} ds \Omega(t, s) G(s) \int_{s}^{T} dv Q(s) G^{T}(s) \psi(s, v) S(v) \psi^{T}(t, v) \]

\[ + \int_{t}^{T} ds \Omega(t, s) Y(s) E^{-1}(s) \left\{ \int_{t}^{s} dv \psi^{T}(v, s) E(v) S(v) + \int_{s}^{T} dv E(s) \psi(s, v) S(v) \right\} \psi^{T}(t, v) \]

\[ - \int_{t}^{T} ds \Omega(t, s) Y(s) E^{-1}(s) \int_{t}^{s} dv \psi^{T}(v, s) E(v) S(v) \psi^{T}(t, v) \]

Adding, we obtain the result

\[ e_{T}(t) \lambda^{T}(t) = 0 \] (3.7.17)

since

\[ Y = G Q G^{T} \]

This means that the smoother adjoint variable is uncorrelated with the error in the smoothed estimate.

With this result Eq. (3.7.7) reduces to:

\[ P_{T} = E - E \lambda \lambda^{T} E \] (3.7.18)

Begin the evaluation of \( \lambda \lambda^{T} \) by using Eq. (3.7.15) to form the indicated product;
\[ \lambda(t) \lambda^T(t) = \int_t^T ds \psi(t, s) \int_t^T dv \left[ S(s) e(s) e^T(v) S(v) + S(s) e(s) \eta^T(v) R^{-1}(v) H(v) \right. \\
+ H^T(s) \eta(s) e^T(v) S(v) + H^T(s) \eta(s) \eta^T(v) R^{-1}(v) H(v) \left. \right] \psi^T(t, v) \]

(3.7.19)

Applying the filter results given in Appendix A to Eq. (3.7.19) we obtain:

\[ \lambda(t) \lambda^T(t) = \int_t^T ds \psi(t, s) S(s) \left\{ \int_t^s dv \psi^T(v, s) E(v) S(v) \psi^T(t, v) + \int_s^T dv E(s) \psi(s, v) S(v) \psi^T(t, v) \right\} \]

\[ - \int_t^T \psi(t, s) H^T(s) R^{-1}(s) ds \int_s^T dv H(s) E(s) \psi(s, v) S(v) \psi^T(t, v) dv \]

\[ - \int_t^T ds \psi(t, s) S(s) \int_t^s dv \psi^T(v, s) E(v) S(v) \psi^T(t, v) \]

\[ + \int_t^T \psi(t, s) S(s) \psi^T(t, s) ds \]

Recalling that \( S = H^T R^{-1} H \) and adding gives the result:

\[ \lambda(t) \lambda^T(t) = \int_t^T \psi(t, s) S(s) \psi^T(t, s) ds \]

(3.7.20)
Define:

\[ \wedge(t) \triangleq \frac{\lambda(t)}{\lambda(t)} \lambda^T(t) = \int_t^T \psi(t, s) S(s) \psi^T(t, s) \, ds \quad (3.7.21) \]

Because \( \wedge(T) \) is zero we may rewrite Eq. (3.7.21) in the form:

\[ \wedge(t) = \psi(t, T) \wedge(T) \psi^T(t, T) + \int_t^T \psi(t, s) S(s) \psi^T(t, s) \, ds \quad (3.7.22) \]

Differentiating Eq. (3.7.22) using Eq. (3.7.14) we obtain:

\[ \dot{\wedge}(t) = - \left[ F^T(t) - S(t) E(t) \right] \psi(t, T) \wedge(T) \psi^T(t, T) - \psi(t, T) \wedge(T) \psi^T(t, T) \left[ F(t) - E(t) S(t) \right] \]

\[ - \int_t^T \left[ F^T(t) - S(t) E(t) \right] \psi(t, s) S(s) \psi^T(t, s) \, ds - \int_t^T \psi(t, s) S(s) \psi^T(t, s) \left[ F(t) - E(t) S(t) \right] \, ds \]

\[ - S(t) \quad (3.7.23) \]

Substitution of (3.7.22) into (3.7.23) gives:

\[ \dot{\wedge}(t) = - \left[ F(t) - E(t) S(t) \right]^T \wedge(t) - \wedge(t) \left[ F(t) - E(t) S(t) \right] - S(t) \quad (3.7.24) \]

Equation (3.7.24) together with:

\[ P_T(t) = E(t) - E(t) \wedge(t) E(t) \quad (3.7.25) \]
represent the Bryson-Frazier method of calculating the smoothed covariance matrix. Equation (3.7.24) must be integrated backward from the terminal condition \( \Lambda(T) = 0 \).

3.8 Derivation of the Rauch-Tung-Striebel Smoothed Covariance Matrix Equations

These covariance matrix equations may be derived directly from the state estimation equations presented in Section 3.5. The derivation is begun by developing a differential equation for the error in the smoothed estimate starting from the Rauch-Tung-Striebel state estimate equation. This is presented in the sequence of Eqs. (3.7.8) to (3.7.12) of the previous section and will not be repeated here.

Using Eq. (3.7.12) we may define the smoothed covariance matrix:

\[
P_T = e_T(t) e_T^T(t) = \Omega(t, T) e(T) e^T(T) \Omega^T(t, T) - \Omega(t, T) \int\limits_t^T e(T) \mu(T) e^T(s) G^T(s) e(T) e^T(s) E^{-1}(s) Y(s) \Omega^T(t, s) ds - \int\limits_t^T \Omega(t, s) \left[ G(s) \mu(s) e^T(T) - Y(s) E^{-1}(s) e(s) e^T(T) \right] ds \Omega^T(t, T)
\]

\[
+ \int\limits_t^T \Omega(t, s) \left[ G(s) \mu(s) - Y(s) E^{-1}(s) e(s) \right] ds \int\limits_t^T \left[ \mu^T(v) G^T(v) - e^T(v) E^{-1}(v) Y(v) \right] \Omega^T(t, v) dv
\]

(3.8.1)

Consider first the evaluation of the second term in Eq. (3.8.1).
Using the results given in Appendix A the second term of Eq. (3.8.1) becomes:

$$2\text{nd~term} = \Omega(t, T) \int_t^T \left[ \psi_T(s, T) G(s) Q(s) G^T(s) - \psi_T(s, T) E(s) E^{-1}(s) Y(s) \right] \Omega^T(t, s) \, ds$$

Since \( GQG^T = Y \) this second term is zero. The third term of Eq. (3.8.1) is also zero because it is the transpose of the second term.

Now consider the product of integrals in the last term of Eq. (3.8.1):

$$\text{last~term} = \int_t^T \Omega(t, s) G(s) \int_t^T dv \mu(s) \mu^T(v) G^T(v) \Omega^T(t, v)$$

$$- \int_t^T \Omega(t, s) G(s) \int_t^T dv \mu(s) e^T(v) E^{-1}(v) Y(v) \Omega^T(t, v)$$

$$- \int_t^T \Omega(t, s) Y(s) E^{-1}(s) \int_t^T dv \mu(s) \mu^T(v) G^T(v) \Omega^T(t, v)$$

$$+ \int_t^T \Omega(t, s) Y(s) E^{-1}(s) \int_t^T dv \mu(s) e^T(v) E^{-1}(v) Y(v) \Omega^T(t, v)$$

(3.8.2)

Applying the results of Appendix A to Eq. (3.8.2) gives:
\[ \text{last term} = \int_t^T \Omega(t, s) \ Y(s) \ \Omega^T(t, s) \ ds \]

\[ - \int_t^T ds \ \Omega(t, s) G(s) \int_s^T dv \ Q(s) \ G^T(s) \ \Psi(s, v) \ E^{-1}(v) \ Y(v) \ \Omega^T(t, v) \]

\[ - \int_t^T ds \ \Omega(t, s) Y(s) E^{-1}(s) \int_t^s dv \ \Psi^T(v, s) G(v) Q(v) G^T(v) \ \Omega^T(t, v) \]

\[ + \int_t^T ds \ \Omega(t, s) Y(s) E^{-1}(s) \int_t^s dv \ E(s) \ \Psi(s, v) \ E^{-1}(v) \ Y(v) \ \Omega^T(t, v) \]

Adding leaves:

\[ \text{last term} = \int_t^T \Omega(t, s) \ Y(s) \ \Omega^T(t, s) \ ds \]

Collecting these results for the second, third, and fourth terms of Eq. (3.8.1) and substituting them into Eq. (3.8.1), we obtain:

\[ P_T(t) = \Omega(t, T) \ P_T(T) \ \Omega^T(t, T) + \int_t^T \Omega(t, s) \ Y(s) \ \Omega^T(t, s) \ ds \]

(3.8.3)
Differentiating Eq. (3.8.3) using Eq. (3.7.11) yields:

\[
\dot{P}_T(t) = \left[ F(t) + Y(t) E^{-1}(t) \right] \Omega(t, T) P_T(T) \Omega^T(t, T) \\
+ \Omega(t, T) P_T(T) \Omega^T(t, T) \left[ F(t) + Y(t) E^{-1}(t) \right]^T \\
+ \int_t^T \left[ F(t) + Y(t) E^{-1}(t) \right] \Omega(t, s) Y(s) \Omega^T(t, s) \, ds \\
+ \int_t^T \Omega(t, s) Y(s) \Omega^T(t, s) \left[ F(t) + Y(t) E^{-1}(t) \right]^T \, ds - Y(t)
\]

(3.8.4)

Substitution of Eq. (3.8.3) into Eq. (3.8.4) gives the Rauch-Tung-Striebel formula for the smoother covariance matrix:

\[
\dot{P}_T(t) = \left[ F(t) + Y(t) E^{-1}(t) \right] P_T(t) + P_T(t) \left[ F(t) + Y(t) E^{-1}(t) \right]^T - Y(t)
\]

(3.8.5)

This is integrated backward from the terminal condition

\[
P_T(T) = E(T)
\]
3.9 The Bryson-Frazier Covariance of the Error in the Driving Force Estimate

If we define

\[ \epsilon_u = u - \hat{u}_T \]  
(3.9.1)

and

\[ Q_T = \epsilon_u \epsilon_u^T \]  
(3.9.2)

we may write using Eqs. (3.2.4) and (3.3.5):

\[ \epsilon_u = \mu - Q G^T \lambda \]  
(3.9.3)

Substitution of Eq. (3.9.3) into (3.9.2) yields:

\[ Q_T = \mu \mu^T - \mu \lambda^T G Q - Q G^T \lambda \mu^T + Q G^T \lambda \lambda^T G Q \]  
(3.9.4)

Consider the expression \( \lambda \mu^T \) in the third term. Using Eq. (3.7.15) we may write this term as:

\[ \lambda(t) \mu^T(t) = \int_t^T \psi(t, s) \left[ S(s) \epsilon(s) \mu^T(t) + H^T(s) R^{-1}(s) \eta(s) \mu^T(t) \right] ds \]  
(3.9.5)

Substitution of Eqs. (3.2.7) and (A.4.3) into Eq. (3.9.5) results in:

\[ \lambda(t) \mu^T(t) = \int_t^T \psi(t, s) S(s) \psi^T(t, s) G(t) Q(t) ds \]  
(3.9.6)
By comparing Eqs. (3.9.6) and (3.7.20) we may deduce:

\[ \lambda(t) \mu^T(t) = \lambda(t) \lambda^T(t) G(t) Q(t) \]  \hspace{1cm} (3.9.7)

Recalling that:

\[ \wedge(t) = \lambda(t) \lambda^T(t) \]  \hspace{1cm} (3.9.8)

we may combine Eqs. (3.9.4), (3.2.5), (3.9.7) and (3.9.8) to obtain the desired result:

\[ Q_T(t) = Q(t) \delta(0) - Q(t) G^T(t) \wedge(t) G(t) Q(t) \]  \hspace{1cm} (3.9.9)

Because \( \mu(t) \) is white noise the first term in Eq. (3.9.9) is infinite. For the continuous case, then, we make a finite correction to an infinite quantity when computing the smoother covariance of the driving noise. In the discrete case, however, we have:

\[ \mu_i \mu_i^T = Q_k \quad i = k \]

\[ = 0 \quad i \neq k \]

Thus in the discrete case all quantities are finite.

3.1.0 The Rauch-Tung-Striebel Covariance of the Error in the Driving Force Estimate

The derivation of the Rauch-Tung-Striebel equivalent of Eq. (3.9.9)
is begun in the same way as the derivation of the previous section.

We begin with Eq. (3.3.5) as before and continue the derivation in the same manner as in Section 3.9 until we obtain:

\[ \mathcal{Q}_T(t) = Q(t) \delta(0) - Q(t) G^T(t) \Lambda(t) \Lambda^T(t) G(t) Q(t) \quad (3.10.1) \]

To evaluate the expected value in the second term use Eq. (3.4.1) and the definition of the smoother and filter errors. The result of these substitutions is:

\[ \Lambda(t) \Lambda^T(t) = E^{-1} \left[ e e^T - e e^T - e e^T + e e^T \right] E^{-1} \quad (3.10.2) \]

Again applying Eq. (3.4.1) enables us to write:

\[ e e^T = (e e^T + E \Lambda) e^T = P_T + E \Lambda e^T \quad (3.10.3) \]

In Section 3.7 using the Rauch-Tung-Striebel form of the error in the smoothed estimate we learned that the error in the smoothed estimate is uncorrelated with the costate for the smoother problem. Using this fact in Eq. (3.10.3) we have:

\[ e e^T = P_T \quad (3.10.4) \]

which means that the covariance of the error in the smoothed estimate with itself and with the error in the filtered estimate are equal.

Collecting Eqs. (3.10.1) through (3.10.4) we obtain the desired result:
\[ Q_T(t) = Q(t) \delta(0) - Q(t) G^T(t) E^{-1}(t) \left[ E(t) - P_T(t) \right] E^{-1}(t) G(t) Q(t) \]

(3.10.5)

Note that here again we make a finite correction to an infinite quantity. In the discrete case all quantities are finite.

3.11 The Discrete Smoother Equations

All the derivations presented in this chapter for the continuous formulation of the smoother equations can be derived in a completely analogous fashion for the discrete case. The discrete state equation is:

\[ x_{k+1} = \Phi_k x_k + G_k u_k \]

where \( \Phi_k \) is the state transition matrix.

The discrete equivalent of the noise statistics of Section 3.2 are:

\[ u_k = \bar{u}_k + \mu_k \]

\[ \mu_k \mu_i^T = Q_k \quad k = i \]

\[ = 0 \quad k \neq i \]

\[ \eta_k \eta_i^T = R_k \quad k = i \]

\[ = 0 \quad k \neq i \]

\[ \eta_k \mu_i^T = 0 \]

The discrete formulation of the Bryson-Frazier smoother equations are:
\[ \hat{x}_{k/N} = \hat{x}_k + E_k \phi_k^T \lambda_k \]  
\( (3.11.1) \)

\[ \lambda_{k-1} = (I - E_k S_k) \phi \left( \hat{x}_k + H_k R_k^{-1} (z_k - H_k \hat{z}_k) \right) \]  
\( (3.11.2) \)

\[ \hat{u}_{k/N} = \hat{u}_k + Q_k G_k^T \lambda_k \]  
\( (3.11.3) \)

\[ P_{k/N} = E_k - E_k \phi_k^T \lambda_k \phi_k E_k \]  
\( (3.11.4) \)

\[ \wedge_{k-1} = (I - E_k S_k) \phi \phi_k \left( I - E_k S_k \right) + S_k - S_k P_k S_k \]  
\( (3.11.5) \)

\[ Q_{k/N} = Q_k - Q_k G_k^T \lambda_k G_k Q_k \]  
\( (3.11.6) \)

In these equations the subscript \( k/N \) indicates the smoother estimate at sample period \( k \) given a total of \( N \) samples.

The results for the method of Rauch, Tung, and Striebel are:

\[ \hat{x}_{k/N} = \hat{x}_{k+1} + E_k \phi_k^T E_{k+1}^{-1} \left( \hat{x}_{k+1/N} - \hat{x}_{k+1} \right) \]  
\( (3.11.7) \)

\[ \hat{u}_{k/N} = \hat{u}_{k+1} + Q_k G_k^T E_{k+1}^{-1} \left( \hat{x}_{k+1/N} - \hat{x}_{k+1} \right) \]  
\( (3.11.8) \)

\[ P_{k/N} = E_k - E_k \phi_k^T E_{k+1}^{-1} \left[ E_{k+1} - P_{k+1/N} \right] E_{k+1}^{-1} \phi_k E_k \]  
\( (3.11.9) \)

\[ Q_{k/N} = Q_k - Q_k G_k^T E_{k+1}^{-1} \left[ E_{k+1} - P_{k+1/N} \right] E_{k+1}^{-1} G_k Q_k \]  
\( (3.11.10) \)
The results presented in this chapter summarize the two major formulations of the smoother equations. They will provide a basis for the work presented in the next few chapters.
CHAPTER IV

SOLUTIONS TO SOME PROBLEMS WHICH ARISE
IN THE PRACTICAL APPLICATION OF THE FILTER AND
SMOOTHER EQUATIONS

4.1 Introduction

There are four major elements involved in the practical application of the smoothing equations using a general purpose digital computer. They are:

1. Compatibility of the raw data with the computer input configuration.
2. Computer time necessary to manipulate the equations.
3. Information storage capacity needed during the computation process.

The problems involved with the first of these will vary with each application and with each computation center so they cannot be considered here.

The basic nature of smoothing requires measurements both past and future with respect to any particular point within the time interval of interest. The requirement of availability of future information immediately prohibits the use of this smoothing scheme as a real time technique. For this reason we are considering here a data reduction problem. This situation does not have the stringent speed requirements of the real time application because the urgency of providing output information before or at fixed times is removed. Computational speed thus affects only cost and not ability to use the technique.
There are at least two more reasons for hesitating to place great emphasis on manipulating the smoothing equations to obtain greater speed. First, the increase in speed is sometimes bought by sacrificing accuracy; second, efficient and clever programming of the smoothing scheme can yield savings in computer time comparable to those afforded by using equation forms which require less arithmetic. For all these reasons no special emphasis will be placed upon finding forms for the smoothing equations which require less arithmetic.

Lack of sufficient storage capacity and loss of accuracy in the computation are handicaps which can prevent the meaningful use of the smoothing equations. Some of these problems are identified and solved in this chapter for the two smoothing techniques derived in Chapter 3. Because of the dependence of the smoother upon the filter, some attention is given to processing of the filter equations.

4.2 A Truncation Problem Arising in the Processing of the Filter Covariance Matrix

Three popular update equations for the filter covariance matrix of state estimation errors are:

\[ E_k = \left( I - K_k H_k \right) E'_k \]  \hspace{1cm} (4.2.1)

\[ E_k = E'_k - E'_k H_k \left( H_k E'_k H_k^T + R_k \right)^{-1} H_k E'_k \]  \hspace{1cm} (4.2.2)

\[ E_k = \left( I - K_k H_k \right) E'_k \left( I - K_k H_k \right)^T + K_k R_k K_k^T \]  \hspace{1cm} (4.2.3)
where $K_k$, the optimal gain matrix, is given by:

$$K_k = E_k^T H_k \left( H_k E_k^T H_k + R_k \right)^{-1} \quad (4.2.4)$$

Equations (4.2.1), (4.2.2), and (4.2.3) are algebraically equivalent; however Eqs. (4.2.1) and (4.2.2) require less arithmetic than Eq. (4.2.3). The purpose of this section is to demonstrate that Eq. (4.2.3) is less sensitive to arithmetic truncation than the simpler, faster forms of Eqs. (4.2.1) and (4.2.2).

The superiority of Eq. (4.2.3) is most noticeable in those cases where the difference between the first and second terms in Eq. (4.3.2) or the difference between $I$ and $K_k H_k$ in Eq. (4.3.1) contains a small error due to a bad subtraction. In these two equations the error appears in $E_k$ to first order. In Eq. (4.2.3) the term containing the error is of second order. The additive first order term does not contain the error, thus $E_k$ calculated in this fashion will contain only second order errors.

This error is especially severe in those cases where the measurement greatly increases the confidence in some or all of the dimensions of the state estimate. When this occurs the elements of $E_k$ associated with these dimensions of the state will be much smaller than the corresponding elements of $E_k'$. Examination of Eq. (4.2.1) reveals that in these situations some elements of $K_k H_k$ will be very nearly equal to corresponding elements of $I$. Similarly some elements of the second term in Eq. (4.2.2) will be very nearly equal to corresponding elements of $E_k'$. The resulting difference of almost equal quantities can have many less significant digits than either of the original arrays.
Joseph (1964) gives an alternate explanation for the numerical superiority of Eq. (4.2.3) which is based upon the assumption that there is a small numerical error in the calculation of $K_k$. In many instances, however, an error in the computation of $E_k$ using Eq. (4.2.1) occurs due to loss of significance in the floating point subtraction even with a perfect value of $K_k$.

As a demonstration of this problem consider a one dimensional update with:

\[
\begin{align*}
    E_k' &= 1 \\
    R_k &= 3 \times 10^{-5} \\
    H_k &= 1
\end{align*}
\]  

\( (4.2.5) \)

This set of numbers is typical of the first update (or for the multidimensional case, the first few updates) when the measurements are very good and the a priori state estimate is poor in comparison to the measurements.

Direct application of Eqs. (4.2.4) and (4.2.1) with $H_k = 1$ for this scalar example gives:

\[
E_k = \frac{R_k}{1 + \frac{R_k}{E_k'}}
\]  

\( (4.2.6) \)

which shows that the variance of the state estimation error after the update
must be smaller than the variance of the measurement noise.

Table 4.1 summarizes the results of machine computation using Eqs. (4.2.1), (4.2.2), and (4.2.3) with the numerical values given in Eq. (4.2.5). A comparison is provided between these results and the correct values.

A truncation problem occurs in this example even though a ten digit machine has been used. It could be worse with a shorter word length. Note also in Table 4.1 that because $K_k$ is very nearly unity there are very few significant digits remaining in the difference $(1-K_k)$. When using Eq. (4.2.1) to compute $E_k$, the error in $(1-K_k)$ causes $E_k$ to be larger than is theoretically possible according to Eq. (4.2.6). Similarly, the subtraction error resulting when the difference is taken between the two terms of Eq. (4.2.2) causes the same incorrect answer. When using Eq. (4.2.3) however, we experience no such problems.

In this example of a single update we advanced from a variance with no numerical errors to one which contained, when using Eq. (4.2.1) or Eq. (4.2.2), an error in the sixth significant digit. As the filtering process continues these arithmetic errors can propagate causing the subtraction problem to become even more severe.

An additional advantage of Eq. (4.2.3) over Eq. (4.2.1) is that it is a symmetric form. This improves the symmetry of the resulting $E_k$ as is readily apparent from Table 4.2.

Table 4.2 uses a two dimensional example to demonstrate the improved symmetry of the resulting covariance matrix provided by Eq. (4.2.3). Selected resulting matrices from a typical run are tabulated with the difference between the off diagonal terms entered as an error in the last digit. A difference of 2 in the 10th digit is listed as 2; a difference of 2 in the 9th digit is listed as 20. The underlined digits in the off diagonal terms are those which agree with the corresponding digits of the opposite term.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Correct Value</th>
<th>Machine Calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Using Eq. (4.2.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.99997000009</td>
</tr>
<tr>
<td>$K_k$</td>
<td>.99997000009</td>
<td>.99997000009</td>
</tr>
<tr>
<td>$E_k'$ $H_k$ $^T$ $\left( H_k E_k' H_k^T + R_k \right)^{-1} H_k E_k'$</td>
<td>.99997000009</td>
<td></td>
</tr>
<tr>
<td>$1 - K_k$</td>
<td>$2.999910000 \times 10^{-5}$</td>
<td>$3.000000000 \times 10^{-5}$</td>
</tr>
<tr>
<td>$E_k$</td>
<td>$2.999910000 \times 10^{-5}$</td>
<td>$3.000000000 \times 10^{-5}$</td>
</tr>
<tr>
<td>Sample Period</td>
<td>Covariance Matrix Calculated From Eq. (4.2.1)</td>
<td>Difference Between off Diagonal Elements in tenth digit</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>10</td>
<td>8.953810562 -08 -1.033097424 -07 -1.033266666 -07 2.358726160 -06</td>
<td>109,242</td>
</tr>
<tr>
<td>15</td>
<td>3.123562330 -08 -2.811619382 -08 -2.811962505 -08 8.049500531 -07</td>
<td>343,123</td>
</tr>
<tr>
<td>20</td>
<td>1.148567499 -08 -7.224615955 -09 -7.225845402 -09 2.986065359 -07</td>
<td>1,230,447</td>
</tr>
<tr>
<td>25</td>
<td>4.258445363 -09 -1.575923306 -09 -1.576385709 -09 1.158539276 -07</td>
<td>462,403</td>
</tr>
<tr>
<td>30</td>
<td>1.570403877 -09 -2.404440200 -10 -2.406208678 -10 4.627255216 -08</td>
<td>1,768,478</td>
</tr>
</tbody>
</table>
4.3 Schemes for Saving Computer Storage Space

Both formulations of the smoothing equations presented in the previous chapter require availability of information from filtering during smoothing. One quick way of providing this information is to store it during the filtering process and recover it during smoothing. For a large problem this technique can become infeasible. Consider for example the application of the Bryson - Frazier discrete smoothing equations to a problem of state dimension $n$, driving force dimension $q$, and measurement dimension $r$ with $N$ measurement points. The storage required at each point is as follows:

- Filter measurement residual: $r$
- Measurement noise covariance matrix: $\frac{r(r + 1)}{2}$
- A priori driving force: $q$
- Driving noise covariance matrix: $\frac{q(q + 1)}{2}$
- Filter state estimate: $n$
- State transition matrix: $n^2$
- Filter covariance matrix: $\frac{n(n + 1)}{2}$
- Measurement matrix: $rn$
- Total per sample: $\frac{1}{2} (r^2 + 3r + q^2 + 3q + 3n^2 + 3n + 2rn)$
- Total for all samples: $\frac{N}{2} (r^2 + 3r + q^2 + 3q + 3n^2 + 3n + 2rn)$
In this summary we have reduced the storage requirement for the covariance matrices by realizing that since they are symmetric they may be stored as triangular matrices.

As a numerical example of the storage requirement consider the results from a hypothetical flight test of a navigational system with:

\[ n = 6 \]
\[ r = 3 \]
\[ q = 3 \]
\[ N = 10,000 \]

This number of data points would be typical of a two orbit test flight of a spacecraft with telemetry information received once per second. For this example the total storage required could reach 3.5 million words.

To avoid such a sizeable storage requirement in large problems we must recalculate some of the filter results during smoothing. This reduces the storage requirements at the expense of computation time.

Inasmuch as the smoother equations are solved by calculating backward from the terminal conditions we must recalculate the filter results in the same fashion - backward from their terminal values. In the continuous case this can be accomplished by simply integrating the forward filter equations backward from the terminal values reached when integrating forward.

The equations for stepping the discrete forward filter results backwards from their terminal values are:
\[ \hat{X}_k^\dagger = \left( I - E_k S_k \right)^{-1} \left( \hat{X}_k - E_k H_k T \right)^{-1} \]

(4.3.1)

\[ \hat{X}_k = \Phi_k^{-1} \left[ \frac{\hat{X}_k^\dagger}{\hat{X}_k+1} - G_k \hat{u}_k \right] \]

(4.3.2)

\[ A_k = E_k H_k T \left( H_k E_k H_k T - R_k \right)^{-1} \]

(4.3.3)

\[ E_k^\dagger = \left( I - A_k H_k \right) E_k \left( I - A_k H_k \right)^{-1} A_k T - A_k R_k A_k T \]

(4.3.4)

\[ E_k = \Phi_k^{-1} \left( E_k^\dagger + Y_k \right) \Phi_k^{-1} \]

(4.3.5)

If the terminal value of the forward filter covariance matrix is numerically inverted the inverse at prior times may be obtained recursively without inverting the covariance matrix again. The recursion formulas are:

\[ E_k^{-1} = E_k^{-1} - H_k T R_k^{-1} H_k \]

(4.3.6)

\[ B_k = E_k^\dagger + G_k \left( G_k T E_k^\dagger - G_k - Q_k^{-1} \right)^{-1} \]

(4.3.7)
\[ E_k^{-1} = \phi_k^T \left[ \left( I - B_k G_k^T \right) E_{k+1}^{-1} \left( I - B_k G_k^T \right)^T - B_k Q_k^{-1} B_k^T \right] \phi_k \]

Equations (4.3.4) and (4.3.8) use the expanded computational form which is less sensitive to numerical errors than the equivalent shorter form. This makes the results insensitive to loss of significance in the subtractions \((I - A_k H_k)\) and \((I - B_k G_k^T)\) for reasons identical to those discussed for the subtraction problem of the previous section.

There does exist, however, another subtraction problem which does not arise with the forward recursion formulas. This occurs in Eq. (4.3.3) when elements of \(H_k E_k H_k^T\) are almost equal to corresponding elements of \(R_k\) and similarly in Eq. (4.3.7) with respect to \(G_k E_{k+1}^{-1} G_k^T\) and \(Q_k^{-1}\). When the filtering is dominated by good measurements the covariance matrix of state estimation errors will pass through points where elements of \(H_k E_k H_k^T\) nearly equal corresponding elements of \(R_k\). When the filtering is dominated by the noise driving the state a similar relationship exists between elements of \(G_k E_{k+1}^{-1} G_k^T\) and \(Q_k^{-1}\). Thus there are real applications where this subtraction problem will arise.

If these situations arise a loss of numerical significance in the result can occur. To understand this consider the scalar analog:

\[ y = x^{-1} \]

\[ dy = \frac{1}{x^2} dx. \]
When $x$ is very small a small error, $dx$, in $x$ can cause a very large error, $dy$, in $y$. A similar sensitivity is found in Eqs. (4.3.3) and (4.3.7).

As an example of this numerical problem consider the application of equations (4.3.3) to (4.3.5) to a system characterized by:

$$
\Phi_k = \text{constant} = \begin{bmatrix} \cdot836 & \cdot086 \\ \cdot3.09 & \cdot673 \end{bmatrix}
$$

$$
Y_k = \text{constant} = 0
$$

$$
R_k = \text{constant scalar} = 10^{-6}
$$

$$
H_k = \text{constant} = (1, 0)
$$

$$
E_N = \begin{bmatrix} 1.570 \times 10^{-9} & -2.458 \times 10^{-10} \\ -2.458 \times 10^{-10} & 4.545 \times 10^{-8} \end{bmatrix}
$$

$$
N = 30
$$

Table 4.3 shows the effect of the numerical error on the one-one element of the covariance matrix both at the sample period in which it occurred and the one following it. A ten decimal digit word length machine was used to obtain the machine calculated values. The correct values were generated using double precision arithmetic on the same machine.
### Table 4.3

Display of Numerical Error in Backward Stepping of Forward Filter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Correct Value</th>
<th>Machine Calculated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>((E_k)_{11})</td>
<td>(9.999985947 \times 10^{-7})</td>
<td>(9.999964460 \times 10^{-7})</td>
</tr>
<tr>
<td>(H_k E_k H_k^T - R_k)</td>
<td>(1.405300000 \times 10^{-12})</td>
<td>(3.554000000 \times 10^{-12})</td>
</tr>
<tr>
<td>((E'<em>k)</em>{11})</td>
<td>(0.7116380364)</td>
<td>(0.281370000)</td>
</tr>
<tr>
<td>((E_{k-1})_{11})</td>
<td>(9.999999858 \times 10^{-7})</td>
<td>(-1.584652080 \times 10^{-6})</td>
</tr>
<tr>
<td>((E'<em>{k-1})</em>{11})</td>
<td>(70.59539234)</td>
<td>(-6.131007310 \times 10^{-7})</td>
</tr>
</tbody>
</table>

Note that \(H_k E_k H_k^T\) is almost equal to \(R_k\). When the two are subtracted the result is heavily dependent upon the low order digits of \((E_k)_{11}\) which are not correct due to accumulative arithmetic errors. The resulting loss of significance in the difference is sufficient to cause gross errors in subsequent calculations. In this example the error was sufficient to cause the covariance matrix calculated using equations (4, 3.3) through (4, 3.5) to become non positive definite.

One way of avoiding this problem is to store the forward filter results as they are generated and recover them during smoothing. If insufficient storage is available a partial storage technique can be employed. This consists of making the decision about possible subtraction problems during forward filtering (by forming \(H_k E_k H_k^T\) and comparing terms to \(R_k\)) and storing the filter values only when a bad subtraction may occur. The storage requirements consist of space for a binary flag (to specify whether or not the filter covariance is stored) at each sample.
period plus the covariance matrix only at the troublesome points. During smoothing the filter covariance matrix and/or its inverse are recomputed everywhere except at the points the flag indicates they should be taken from storage.

4.4 A Solution to a Numerical Problem Arising During Smoothing

A loss of numerical significance due to subtraction of two very nearly equal quantities can arise when computing the smoother covariance matrix by either method presented in Chapter 3. The problem occurs in applying Eq. (3.11.4) or Eq. (3.11.9) when elements in the first term are very nearly equal to the corresponding elements of the second term. The resulting difference is very sensitive to errors in the low order digits of each term.

Note that in both Eq. (3.11.4) and (3.11.9) the smoother covariance matrix is generated by subtracting a correction term from the forward filter covariance matrix. Consequently this numerical problem is likely to arise in those cases where the confidence in the smoothed estimate of one or more dimensions of the state is much higher than the confidence in the forward filter estimate of the same dimensions at the same sample period. Unfortunately these situations are the ones in which smoothing is most valuable.

Equation (3.11.9) involves a matrix inversion. This makes it possible to manipulate the Rauch-Tung-Striebel smoother covariance matrix formula into a form which replaces the difference of two positive definite matrices by a sum. This manipulation follows:

If we define:

\[ C_k = E_k \Phi_k T_{k+1}^{-1} \]

(4.4.1)
we may rewrite Eq. (3.11.9) in the form:

\[
P_{k/N} = E_k - C_k \left[ E'_{k+1} - P_{k+1/N} \right] C_k^T
\]

\[
= E_k - C_k E'_{k+1} C_k^T - C_k E'_{k+1} C_k^T + C_k E'_{k+1} C_k^T + C_k P_{k+1/N} C_k^T
\]

Substituting for \( C_k \) in the second term and \( C_k^T \) in the third gives:

\[
P_{k/N} = E_k - E_k \phi_k T C_k T - C_k \phi_k E_k + C_k \left[ \phi_k E_k \phi_k^T + Y_k \right] C_k^T + C_k P_{k+1/N} C_k^T
\]

\[
= E_k - E_k \phi_k T C_k T - C_k \phi_k E_k + C_k \phi_k E_k \phi_k^T C_k T + C_k Y_k C_k^T
\]

\[
+ C_k P_{k+1/N} C_k^T
\]

Thus

\[
P_{k/N} = \left( I - C_k \phi_k \right) E_k \left( I - C_k \phi_k \right)^T + C_k \left[ P_{k+1/N} + Y_k \right] C_k^T
\]  \hspace{1cm} (4.4.2)
Equations (4.4.1) and (4.4.2) thus give a way of computing the Rauch-Tung-Striebel smoother covariance matrix which is not as sensitive to numerical problems as Eq. (3.11.9). A similar form for the Bryson-Frazier formulation which does not involve a matrix inversion has not been found.

As a demonstration of the cause of the numerical problem and of the superiority of Eq. (4.4.2) over those presented in Chapter 3 consider the system described in section 4.3 with an initial a priori covariance matrix of:

$$E_0 = \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}$$

Tables 4.3 and 4.4 summarize the development of the error in Eqs. (3.11.4) and (3.11.9). Table 4.5 provides a comparison between the results of computation using Eq. (3.11.9) and (4.4.2). The correct values were calculated using double precision arithmetic. Inspection of these tables reveals that the numerical problem can cause large errors even when a ten decimal digit word length machine is used to perform the computation.

Note that although the bad subtraction is avoided when using Eq. (4.4.2), errors in $C_k$ appear in $P_{k/N}$ to first order. In the symmetric forms for the filter equations this was not true with respect to errors in the gain matrices $K_k$, $A_k$, and $B_k$. If we apply Joseph's argument to Eq. (4.4.2) with respect to errors in $C_k$ we obtain:

$$\delta P_{k/N} = C_k P_{k+1/N} \delta C_k^T + \delta C_k P_{k+1/N} C_k^T$$

(4.4.3)
<table>
<thead>
<tr>
<th>Quantity</th>
<th>Computed Value</th>
<th>Correct Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.571517980 -06 9.687057404 01</td>
<td>-3.571517980 -06 9.687057404 01</td>
</tr>
<tr>
<td>$C_k \left[ E_{k+1}' - P_{k+1}/N \right] C_k^T$</td>
<td>6.829907246 -07 -3.000000000 -06</td>
<td>6.828124412 -07 -3.003650185 -06</td>
</tr>
<tr>
<td></td>
<td>-3.000420200 -06 9.684700935 01</td>
<td>-3.003650185 -06 9.687055929 01</td>
</tr>
<tr>
<td>$P_{k/N}$</td>
<td>3.170092624 -07 -5.710977800 -07</td>
<td>3.171875458 -07 -5.678677951 -07</td>
</tr>
<tr>
<td></td>
<td>-5.715179800 -07 2.356469000 -02</td>
<td>-5.678677951 -07 1.474934419 -05</td>
</tr>
<tr>
<td>$E_{k-1}$</td>
<td>1.000000000 02 0.000000000 00</td>
<td>1.000000000 02 0.000000000 00</td>
</tr>
<tr>
<td></td>
<td>0.000000000 00 1.000000000 02</td>
<td>0.000000000 00 1.000000000 02</td>
</tr>
<tr>
<td>$C_{k-1} \left[ E_k' - E_k/N \right] C_{k-1}^T$</td>
<td>9.999974663 01 2.468018000 -03</td>
<td>9.999999995 01 8.294412254 -07</td>
</tr>
<tr>
<td></td>
<td>2.468100000 -03 9.997592666 01</td>
<td>8.294412254 -07 9.999984800 01</td>
</tr>
<tr>
<td>$P_{k-1/N}$</td>
<td>2.534000000 -04 -2.468100000 -03</td>
<td>4.643005134 -07 -8.294412540 -07</td>
</tr>
<tr>
<td></td>
<td>-2.468018000 -03 2.407340000 -02</td>
<td>-8.294412254 -07 1.520015809 -05</td>
</tr>
<tr>
<td>Quantity</td>
<td>Computed Value</td>
<td>Correct Value</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td></td>
<td>-3.571517980 -06 9.687057404 01</td>
<td>-3.571517980 -06 9.687057404 01</td>
</tr>
<tr>
<td>$E_k^T \Phi_k \wedge_k \Phi_k E_k$</td>
<td>6.282902441 -07 -3.006688431 -06</td>
<td>6.828124412 -07 -3.003650185 -06</td>
</tr>
<tr>
<td></td>
<td>-3.005691163 -06 9.684303410 01</td>
<td>-3.003650185 -06 9.687055929 01</td>
</tr>
<tr>
<td>$P_{k/N}$</td>
<td>3.170975455 -07 -5.658268170 -07</td>
<td>3.171875458 -07 -5.678677951 -07</td>
</tr>
<tr>
<td></td>
<td>-5.648295490 -07 2.753994000 -02</td>
<td>-5.678677951 -07 1.474934419 -05</td>
</tr>
<tr>
<td>$E_{k-1}$</td>
<td>1.000000000 02 0.000000000 00</td>
<td>1.000000000 02 0.000000000 00</td>
</tr>
<tr>
<td></td>
<td>0.000000000 00 1.000000000 02</td>
<td>0.000000000 00 1.000000000 02</td>
</tr>
<tr>
<td>$E_{k-1}^T \Phi_{k-1} \wedge_{k-1} \Phi_{k-1} E_{k-1}$</td>
<td>9.186147196 01 -8.314161000 -01</td>
<td>9.999999995 01 8.294412254 -07</td>
</tr>
<tr>
<td></td>
<td>-8.529041240 -01 9.988177495 01</td>
<td>8.294412254 -07 9.999998480 01</td>
</tr>
<tr>
<td>$P_{k-1/N}$</td>
<td>8.138528100 00 8.529041240 -01</td>
<td>4.643005134 -07 -8.294412540 -07</td>
</tr>
<tr>
<td></td>
<td>8.314161000 -01 1.182251000 -01</td>
<td>-8.294412254 -07 1.520015809 -05</td>
</tr>
<tr>
<td>Quantity</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Correct $P_{k/N}$</td>
<td>$3.171875458 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-5.678677951 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td>$P_{k/N}$ by Eq. (3.11.9)</td>
<td>$3.170092624 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-5.710977800 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td>$P_{k/N}$ by Eq. (4.4.2)</td>
<td>$3.173207111 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-5.682210880 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td>Correct $P_{k-1/N}$</td>
<td>$4.643005134 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-8.294412540 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td>$P_{k-1/N}$ by Eq. (3.11.9)</td>
<td>$2.534000000 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-2.468100000 \times 10^{-3}$</td>
<td></td>
</tr>
<tr>
<td>$P_{k-1/N}$ by Eq. (4.4.2)</td>
<td>$4.797986952 \times 10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$-9.788842160 \times 10^{-7}$</td>
<td></td>
</tr>
</tbody>
</table>
We will see in Chapter VI that there exists another formulation of the smoothing equations which eliminates this type of first order error. An improvement over the results when using Eq. (4.4, 2) is obtained.

This chapter provides means of increasing the accuracy and decreasing the storage requirements for the two major smoothing formulations. In the next two chapters we will consider these problems again - both for these formulations and a new one which is presented in Chapter VI.
CHAPTER V

A SMOOTHABILITY CONDITION FOR LINEAR SYSTEMS

5.1  A Statement of Smoothability

The purpose of this chapter is to develop a criterion to judge whether or not smoothing will provide an estimate of the state which is superior to that which can be obtained by simpler means. This criterion will be called the "smoothability condition". Specifically:

A state will be considered smoothable if smoothing of the data yields an estimate of the state which is different from that which would be obtained by extrapolating the final forward filter state estimate backward in time.

With this definition we will find that the smoothability condition is:

Only those dimensions of the state which are controllable by the noise driving the state are smoothable.

5.2  A Derivation of the Smoothability Condition for the Continuous Formulation

The solution to Eq. (3.3.11) in terms of the smoother costate is:

\[ \hat{x}_T(t) = \Phi(t, T) \hat{x}(T) - \int_t^T \Phi(t, s) G(s) \bar{u}(s) \, ds - \int_t^T \Phi(t, s) Y(s) \lambda(s) \, ds \]

\[ (5.2.1) \]
The first two terms of Eq. (5.2.1) extrapolate the final forward filter state estimate backward in time. Since it is the third term which causes the smoothed estimate to differ from backward extrapolation it is this term which is of interest for smoothability. Those dimensions which have a third term contribution of zero are not smoothable.

Kalman (Kalman (1962) and Kalman (1963)) has shown that a linear system as described in Section 3.2 may be reduced to a canonical form in which the state is partitioned into its controllable and uncontrollable parts. We will use this canonical form to reduce our system to a form in which the state is partitioned into two parts: one of which is controllable by the driving noise \( \mu(t) \) and the other of which is not.

Using the subscripts \( c \) and \( u \) to designate respectively the parts of the state which are and are not controllable by the driving noise we have:

\[
\begin{bmatrix}
  x_c \\
  x_u
\end{bmatrix}
= \begin{bmatrix}
  \lambda_c \\
  \lambda_u
\end{bmatrix}
\begin{bmatrix}
  \phi_{cc} & \phi_{cu} \\
  0 & \phi_{uu}
\end{bmatrix}
\begin{bmatrix}
  x_c \\
  x_u
\end{bmatrix}
\]

With this canonical form the driving term in the state equation is:

\[
\text{driving term} = G' \, u + \begin{bmatrix}
  G_c \\
  0
\end{bmatrix} \mu.
\]

Due to the assumed canonical nature of the system equations \( G_c \) contains no zero elements.

The coefficient of \( \mu \) in the driving term is used to form \( Y \) in the new coordinate system. Using the definition of \( Y \) we have:
\[ Y = \begin{bmatrix} G_c \\ 0 \end{bmatrix} Q \begin{bmatrix} G_c^T & 0 \end{bmatrix} = \begin{bmatrix} Y_c \\ 0 \end{bmatrix} \]

Substitution of these canonic forms into the third term of Eq. (5.2.1) yields:

\[ 3\text{rd term} = -\int_t^T \begin{bmatrix} \Phi_{cc}(t, s) Y_c(s) & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_c(s) \\ \lambda_u(s) \end{bmatrix} \, ds \]

(5.2.2)

Inspection of Eq. (5.2.2) reveals that the third term of Eq. (5.2.1) contributes only to those dimensions of the state which are controllable by the driving noise. Due to the assumed canonic form of the system this third term contributes to all the dimensions which are controllable by the driving noise. Since it is only the third term which causes the smoother estimate to differ from a backward extrapolation of the final forward filter estimate we may conclude that only those dimensions of the state which are controllable by the noise driving the state are smoothable.

5.3 An Alternate Derivation of the Smoothability Condition for the Discrete Formulation

The purpose of this section is to provide an alternate derivation of the smoothability condition for the discrete case. We will use the Rauch-Tung-Striebel formulation this time and will again express the state equation in the canonic form in which the state is partitioned into two sections: one of which is controllable by the driving noise and one of which is not.

In this canonical form the state equation is:
\[
\begin{bmatrix}
    x_c \\
    x_u
\end{bmatrix}_{k+1} =
\begin{bmatrix}
    \Phi_{cc} & \Phi_{cu} \\
    0 & \Phi_{uu}
\end{bmatrix}
\begin{bmatrix}
    x_c \\
    x_u
\end{bmatrix}_k +
\begin{bmatrix}
    G_c \\
    G_u
\end{bmatrix}_k \begin{bmatrix}
    u_k \\
    0
\end{bmatrix}_k +
\begin{bmatrix}
    G_c
\end{bmatrix}_k \mu_k
\]
(5.3.1)

Note that \( G_k \) in Section 3.11 and in the covariance matrix expressions of Chapter IV is now the coefficient of \( \mu_k \) in Eq. (5.3.1).

In terms of the notation introduced in Chapter IV the Rauch-Tung-Striebel state estimate (Eq. (3.11.7)) may be written:

\[
\hat{x}_{k/N} = \hat{x}_k + C_k \left[ \hat{x}_{k+1/N} - \hat{x}_{k+1} \right]
\]
(5.3.2)

where \( C_k \) is defined in Eq. (4.4.1).

We may obtain an expression for \( E_{k+1}^{-1} \) (which is involved in \( C_k \)) either by solving Eqs. (4.3.7) and (4.3.8) for \( E_{k+1}^{-1} \) or by inverting the forward filter covariance matrix extrapolation equation. This extrapolation equation together with a matrix inversion identity useful for inverting it may be found in Rauch, Tung, Striebel (1965). The result is:

\[
E_{k+1}^{-1} = \Phi_k T^{-1} \left[ E_k^{-1} - E_k^{-1} \Phi_k^{-1} G_k (G_k \Phi_k T^{-1} E_k^{-1} \Phi_k^{-1} G_k + Q_k^{-1})^{-1} \right]
\]

\[
\begin{bmatrix}
    G_k T & E_k^{-1} \Phi_k^{-1} G_k + Q_k^{-1}
\end{bmatrix}\phi_k^{-1}
\]
(5.3.3)

Substitution of Eq. (5.3.3) into Eq. (4.4.1) gives

\[
C_k = \begin{bmatrix}
    I - \Phi_k^{-1} G_k (G_k T E_k^{-1} \Phi_k^{-1} G_k + Q_k^{-1})^{-1} C_k T & E_k^{-1}
\end{bmatrix}\phi_k^{-1}
\]
(5.3.4)
For convenience define the $q \times q$ matrix:

$$D_k \triangleq (G_k^T \Phi_k^{-1} E_k^{-1} \Phi_k^{-1} G_k + Q_k^{-1})^{-1}.$$  \hfill (5.3.5)

Combining Eqs. (5.3.4) and (5.3.5) we have:

$$C_k = \left[ I - \Phi_k^{-1} G_k D_k G_k^T \Phi_k^{-1} E_k^{-1} \right] \Phi_k^{-1}.$$  \hfill (5.3.6)

Inverting the state transition matrix of Eq. (5.3.1) gives:

$$\Phi_k^{-1} = \begin{bmatrix} \Phi_{cc}^{-1} & -\Phi_{cc}^{-1} \Phi_{cu} \Phi_{uu}^{-1} \\ \Phi_{cc}^{-1} \Phi_{cu} & \Phi_{uu}^{-1} \end{bmatrix}.$$  \hfill (5.3.7)

Recalling that $G_k$ in Eq. (5.3.6) is the coefficient of $\mu_k$ in Eq. (5.3.1) we may write:

$$\Phi_k^{-1} G_k = \begin{bmatrix} \Phi_{cc}^{-1} G_c \\ 0 \end{bmatrix}.$$  \hfill (5.3.8)

If we partition $E_k^{-1}$ as above and define for convenience its matrix elements to be:

$$E_k^{-1} \triangleq \begin{bmatrix} V_{cc} & V_{cu} \\ V_{cu} & V_{uu} \end{bmatrix}_k.$$  \hfill (5.3.9)
we may combine Eqs. (5.3.6), (5.3.8), and (5.3.9) to give:

\[
C_k = \begin{bmatrix}
(I - \Phi^{-1}_{cc} G_c D G_c^T \Phi^{-1}_{cc} V_{cc}) & \Phi^{-1}_{cc} G_c D G_c^T \Phi^{-1}_{cc} V_{cu} \\
0 & I
\end{bmatrix} \Phi^{-1}_k
\]  
(5.3.10)

Combining Eq. (5.3.7) and Eq. (5.3.10) we may conclude that \(C_k\) is of the form:

\[
C_k = \begin{bmatrix}
C_{cc} & C_{cu} \\
0 & \Phi^{-1}_{uu}
\end{bmatrix} \Phi^{-1}_k
\]  
(5.3.11)

Rearranging Eq. (5.3.2) and writing it in partitioned form we have:

\[
\begin{bmatrix}
\hat{\mathbf{x}}_c \\
\hat{\mathbf{x}}_u
\end{bmatrix}^{k/N} = (I - C_k \Phi_k) \hat{\mathbf{x}}_k + C_k \hat{\mathbf{x}}^{k+1/N} - C_k \begin{bmatrix}
G_c \\
G_u
\end{bmatrix} \bar{u}_k
\]  
(5.3.12)

Substitution of \(C_k\) from Eq. (5.3.11) and \(\Phi_k\) from Eq. (5.3.1) into Eq. (5.3.12) leaves:

\[
\begin{bmatrix}
\hat{\mathbf{x}}_c \\
\hat{\mathbf{x}}_u
\end{bmatrix}^{k/N} = \begin{bmatrix}
(I - C_{cc} \Phi_{cc}) & -(C_{cc} \Phi_{cu} + C_{cu} \Phi^{-1}_{uu}) \\
0 & 0
\end{bmatrix} \begin{bmatrix}
\hat{\mathbf{x}}_k \\
\bar{u}_k
\end{bmatrix}
\]  
(5.3.13)
From this result we have:

\[ \left( \widehat{\chi}_u \right)_{k/N} = \Phi^{-1}_{uu_k} \left[ \left( \widehat{\chi}_u \right)_{k+1/N} - G_{u_k} \overline{u}_k \right] \]  

(5.3.14)

Equation (5.3.14) represents the extrapolation of the smoother estimate from sample period k+1 to sample period k. The extrapolation process is begun with \( \left( \widehat{\chi}_u \right)_{N/N} \) which is the final forward filter estimate. The smoother estimate of this part of the state is thus a backward extrapolation of the final forward filter estimate. Consequently the part of the state which is uncontrollable by the driving noise is non-smoothable.

Due to the assumed canonic form of the system all the elements of the first row of the coefficient of \( \Delta_k \) in Eq. (5.3.13) cannot be zero. Thus the smoother estimate of \( x_c \) differs from a backward extrapolations of the final forward filter estimate. Hence, we may conclude that the only part of the state which is smoothable is that which is controllable by the noise driving the state.

5.4 Discussion of the Smoothability Condition

The smoothability condition derived in this chapter may be employed to guide the user in the application of the smoothing technique to real problems. The most important application of this condition is probably its use in organizing the computation in a more efficient and accurate manner.

First of all it is quite clear that one should not consider smoothing in those linear cases where there is no driving noise. In these applications the entire state is non-smoothable and backward extrapolation of the state would theoretically provide the same result as processing the smoothing equations. Inasmuch as the smoothing equations are more complex than those which describe the backward extrapolation, a saving in computation time may be obtained by using
extrapolation. The resulting reduction in arithmetic should reduce the effect of truncation errors causing the estimate obtained by extrapolation to be numerically more accurate than that which would be obtained by processing the smoothing equations.

In this case what is true for the whole is true for the parts. In those applications where only part of the state is smoothable a saving in computation time and increased accuracy may be obtained by partitioning the state into its smoothable and non-smoothable parts. The estimate of the non-smoothable part can then be obtained by backward extrapolation while the estimate of the smoothable part is obtained by processing the smoothing equations.

Insight into smoothability for some nonlinear systems can be obtained from the derivation presented here even though this derivation is only applicable to linear systems. These nonlinear systems may be classified as those which may be treated by a linearization process.

If the nonlinear equations are linearized about a given reference solution the smoothability condition is applicable to the linear equations which describe the deviation from the reference solution. The smoothed estimate for the nonlinear problem is obtained by adding the smoothed estimate of the linear deviation to the given reference solution.

If, however, the reference solution is updated by incorporating the deviation into the nominal the smoothability condition is no longer applicable. This is most easily understood by considering the derivation for the discrete case. During the calculation of the forward filter estimate both the state estimate and the covariance matrix are extrapolated using a state transition matrix which is based upon a linearization about the filtered estimate of the state. The simplification of the smoothing equations which led to the smoothability condition will no longer be possible. An example of one place this occurs is in the substitution of Eq. (5.3.3) into Eq. (4.4.1) to obtain Eq. (5.3.4). The transition matrix in Eq. (5.3.3) is based upon the filtered estimate while that of Eq. (4.4.1) is based upon the smoothed estimate. Because of this we no longer have:
\[ \Phi_k^{-1} \Phi_k^T = I \]

and the simple form of Eq. (5.3.4) and subsequent equations is lost.

If this process converges\(^*\) an iterative technique consisting of using the smoother results to begin the filtering over again may be employed. After several iterations the reference should not change much and the above comments for the case of a given reference solution become applicable. In particular the smoothability condition will be approximately valid.

\(^*\) Ho and Lee (1964) offer a proof that it should converge for the filtering process when the nonlinear problem is an identification problem organized by linearization.
CHAPTER VI

THE OPTIMUM SMOOTHER AS A COMBINATION OF TWO KALMAN FILTERS

6.1 The Form of the New Smoother Equations

The Kalman filter equations when processed from the beginning of the data interval to some time t within the interval yield the best estimate* of the state at time t based upon all the measurements from the beginning of the interval to time t. If, however, the Kalman filter equations are processed from the end of the data interval backward to time t they will yield the best estimate of the state at time t based upon all the measurements from time t to the end of the time interval. Together these two filters use all the measurements available - both past and future with respect to time t. We will demonstrate in this chapter that we can obtain the smoother estimate by properly combining these two filter estimates. For linear systems with Gaussian noises the smoother estimate is the maximum likelihood combination of the two independent filter estimates. Specifically we will find that the smoother state estimate is a weighted sum of the two filter estimates. The weighting factors are proportional to the reciprocal of the covariance of errors in each estimate; thus they are normalized confidences in each estimate. The inverse of the smoother covariance matrix of state estimation errors is obtained by simply adding the inverses of the corresponding filter covariance matrices.

6.2 A Continuous Form Derivation of the New Smoother Covariance Matrix Equation

In this section we will demonstrate that the smoother covariance matrix is given by the maximum likelihood formula for combining the variances of two independent estimates:

* Best in the sense given in Section 3.2. For a linear system with Gaussian noises this is in the maximum likelihood, minimum variance, and least squares senses.
\[ P_t^{-1}(t) = E^{-1}(t) + P^{-1}(t) \]  \hspace{1cm} (6.2.1)

where \( P(t) \) is the covariance matrix of state estimation errors for the Kalman filter which works backward from the end of the data. We will hereafter refer to this filter as the backward filter and to the other as the forward filter.

For convenience we will use the definitions:

\[ V = E^{-1} \]  \hspace{1cm} (6.2.2)

\[ U = P^{-1} \]

With this convention Eq. (6.2.1) may be rewritten as:

\[ P_t^{-1} = V + U. \]  \hspace{1cm} (6.2.3)

From Chapter III we have:

\[ P_T(T) = E(T) \]

Thus:

\[ P_T^{-1}(T) = V(T) \]

Hence, if we demand:

\[ U(T) = 0 \]  \hspace{1cm} (6.2.4)

Eq. (6.2.3) will be true at the terminal time. The meaning of this boundary condition is that the backward filter begins with no prior information; there is no confidence in the a priori state estimate because it does not exist, thus, \( U(T) \) must be zero. A Kalman filter
which operates with this boundary condition is called an unbiased filter. For a discussion of such filters see Potter and Stern (1964).

We have established that Eq. (6.2.3) is true at the terminal time. To show that it is true at all other times we will demonstrate that the derivatives are always equal, or that:

\[ P^{-1} = \dot{V} + \dot{U}. \]  
\[ (6.2.5) \]

We may easily obtain an expression for \( V \) by differentiating the identity:

\[ V E = I \]

The result is:

\[ \dot{V} = - V \dot{E} V \]  
\[ (6.2.6) \]

Substitution of Eq. (3.4.4) into Eq. (6.2.6) yields:

\[ \dot{V} = - V F - F^T V - V Y V + S \]  
\[ (6.2.7) \]

To obtain an expression for \( \dot{U} \) begin by defining the backward time variable:

\[ \tau = T - t \]

Then

\[ \frac{dt}{d\tau} = -1 \]  
\[ (6.2.8) \]

Using Eq. (6.2.8) and the chain rule for derivatives we may rewrite Eq. (A.1.3) as:
\[
\frac{d}{d\tau} (e_B) = - \left[ F(\tau) + P(\tau) S(\tau) \right] e_B(\tau) - G(\tau) \mu(\tau) + P(\tau) H(\tau) R^{-1}(\tau) \eta(\tau)
\]

(6.2.9)

where \( e_B \) is the error in the backward filter state estimate.

We may derive an expression for the corresponding covariance matrix by methods analogous to those used in Section 3.8 to obtain the smoother covariance matrix. The result is:

\[
\frac{dP}{d\tau} = - F(\tau) P(\tau) - P(\tau) F^T(\tau) + Y(\tau) - P(\tau) S(\tau) P(\tau)
\]

(6.2.10)

Another argument which leads to Eq. (6.2.10) begins by noting that Eq. (6.2.9) may be derived from Eq. (A.1.3) by changing \( t \) to \( \tau \), \( F \) to \( -F \), \( G \) to \( -G \), and \( H \) to \( -H \). Applying this observation to Eq. (3.4.4) yields Eq. (6.2.10).

Using Eq. (6.2.8) and the chain rule for derivatives we may rewrite Eq. (6.2.10) with time running forward. The result is:

\[
\frac{dP}{dt} = F(t) P(t) + P(t) F^T(t) - Y(t) + P(t) S(t) P(t)
\]

(6.2.11)

Using a result for \( P \) and \( U \) which is completely analogous to Eq. (6.2.6) we may write, employing Eq. (6.2.11):

\[
\hat{U} = - U F - F^T U + U Y U - S
\]

(6.2.12)

Using Eqs. (6.2.7) and (6.2.12) we may form the right side of Eq. (6.2.5):

\[
\hat{V} + \hat{U} = - (V + U) F - F^T (V + U) - V Y V + U Y U
\]

(6.2.13)

Substitution of Eq. (6.2.3) into Eq. (6.2.13) yields:

\[
\hat{V} + \hat{U} = - P^{-1}_T F - F^T P^{-1}_T - V Y V + (P^{-1}_T - V) Y (P^{-1}_T - V).
\]

Expanding and collecting terms leaves:

\[
\hat{V} + \hat{U} = - P^{-1}_T (F + Y V) - (F + Y V)^T P^{-1}_T + P^{-1}_T Y P^{-1}_T
\]

(6.2.14)
Using a result for \((V + U)\) and its inverse which is completely analogous to Eq. (6.2.6) we may invert Eq. (6.2.14):

\[
\frac{d}{dt} (V + U)^{-1} = (F + Y V) P^T + P_T (F + Y V)^T - Y \quad (6.2.15)
\]

The right side of Eq. (6.2.15) is the right side of the Rauch-Tung-Striebel state equation (Eq. (3.8.5)). Thus, we may write:

\[
P_T = \frac{d}{dt} (V + U)^{-1} \quad (6.2.16)
\]

Equation (6.2.16) is the derivative of Eq. (6.2.3) when it is written in the form:

\[
P_T = (V + U)^{-1}
\]

Since the derivatives are equal for all time and Eq. (6.2.3) is true at the terminal time, Eqs. (6.2.3) and (6.2.1) are true for all time.

6.3 A Continuous Form Derivation of the New Smoother State Estimation Equation

In this section we will demonstrate that the smoother state estimate is given by the maximum likelihood formula for the combination of two independent estimates:

\[
P_T (E^{-1} \hat{x} + P^{-1} \hat{x}_B)
\]

(6.3.1)

where \(\hat{x}_B\) is the state estimate of the backward filter. To show this we
will derive the Rauch-Tung-Striebel smoothed state estimation formula from Eq. (6.3.1).

Using the definitions of Eq. (6.2.2) and differentiating Eq. (6.3.1) we obtain:

\[ \dot{\hat{\chi}}_T = \dot{P}_T \left[ V \dot{\hat{x}} + U \hat{x}_B \right] + P_T \left[ \dot{V} \dot{\hat{x}} + \dot{U} \hat{x}_B + U \dot{\hat{x}}_B \right] \]  \hspace{1cm} (6.3.2)

We already have obtained expressions for most of the derivatives on the right side of Eq. (6.3.2). \( \dot{P}_T \) is defined by Eq. (3.8.5), \( \dot{V} \) by Eq. (6.2.7), \( \dot{U} \) by Eq. (6.2.12), and \( \dot{\hat{x}} \) by Eq. (3.4.5). The equation for the backward filter state estimate is similar to Eq. (3.4.5). With time running backward it is:

\[ \frac{d}{d\tau} (\hat{x}_B) = - F(\tau) \hat{x}_B(\tau) - G(\tau) \bar{\nu}(\tau) + P(\tau) H^T(\tau) R^{-1}(\tau) \left[ z(\tau) - H(\tau) \hat{\chi}_B(\tau) \right] \]  \hspace{1cm} (6.3.3)

To obtain the proper form for substitution into Eq. (6.3.2), we must have time running forward. This is accomplished by applying Eq. (6.2.8) to Eq. (6.3.3). The result is:

\[ \dot{\hat{x}}_B = F \hat{x}_B + G \bar{\nu} - P H^T R^{-1} \left[ z - H \hat{\chi}_B \right] \]  \hspace{1cm} (6.3.4)

Equation (6.3.1) may be rewritten in the form:

\[ V \dot{\hat{x}} + U \dot{\hat{x}}_B = P_T^{-1} \dot{\hat{\chi}}_T \]  \hspace{1cm} (6.3.5)
Substitution of Eqs. (3.8.5), (6.2.7), (6.2.12), (3.4.5), (6.3.4) and (6.3.5) into Eq. (6.3.2) yields, after collecting and cancelling terms:

\[
\dot{\hat{x}}_T = F \hat{x}_T + Y V (\hat{x}_T - \hat{x}) + P_T (U + V) G \bar{u} \\
+ (P_T V Y U - Y U + P_T U Y U) \hat{x}_B
\]  

(6.3.6)

Substituting Eq. (6.2.3) into the third and fourth terms on the right side of Eq. (6.3.6) gives:

\[
\dot{\hat{x}}_T = F \hat{x}_T + Y V (\hat{x}_T - \hat{x}) + G \bar{u} + (Y U - P_T U Y U) \\
- Y U + P_T U Y U) \hat{x}_B
\]

Cancelling and rearranging terms yields:

\[
\dot{\hat{x}}_T = F \hat{x}_T + G \bar{u} + Y E^{-1} (\hat{x}_T - \hat{x})
\]  

(6.3.7)

Equation (6.3.7) is identical to Eq. (3.5.4), the Rauch-Tung-Striebel smoother state estimation formula. If Eq. (6.3.1) gives the correct smoothed state estimate at any time we may conclude that it is correct for all times because its derivative is equal to the derivative of the smoothed state estimate.

Employing the boundary condition on U given by Eq. (6.2.4) we may write the smoother estimate (according to Eq. (6.3.1)) at the terminal time as:

\[
\hat{x}_T(T) = P_T(T) E^{-1}(T) \hat{x}(T)
\]  

(6.3.8)
Since:

\[ P_T(T) = E(T) \]
\[ \hat{x}_T(T) = \hat{x}(T) \]

Equation (6.3.8) is an identity. Thus, Eq. (6.3.1) is a valid expression for the smoother state estimate.

6.4 A More Useful Form for the New Smoother Covariance Matrix Equation

Equation (6.2.1) is not very convenient for calculation of the smoother covariance matrix for two reasons:

1. \( P(t) \) is difficult to compute because it is infinite at the terminal time.

2. Three inversions of matrices which are of the dimension of the state are necessary.

The problem with the infinite \( P(t) \) can be avoided by working with the matrix \( U(t) \), which is the inverse of \( P(t) \). An expression can then be derived which does not involve any infinite quantities and which requires only one \( n \times n \) matrix inversion. This derivation is begun by rewriting Eq. (6.2.1) in the form:

\[ P^{-1}_T = E^{-1} + U \]

then

\[ E P^{-1}_T = I + E U \]

\[ P_T = (I + E U)^{-1} E \quad (6.4.1) \]
A more symmetric form may be obtained by premultiplying the right side of Eq. (6.4.1) by \((I + E U - E U)\).

\[
P_T = \left[ (I + E U) - E U \right] (I + E U)^{-1} E
\]

\[
P_T = E - E U (I + E U)^{-1} E \tag{6.4.2}
\]

This in turn may be expanded into a symmetric form which is the sum of two positive definite matrices. The result is:

\[
P_T = (I - W U) E (I - W U)^T + W U W^T \tag{6.4.3}
\]

where:

\[
W = E (I + E U)^{-1}^T \tag{6.4.4}
\]

To establish this expand Eq. (6.4.3):

\[
P_T = E - E U W^T + W U \left[ (I + E U) W^T - E \right]
\]

Using Eq. (6.4.4) to substitute for \(W\) gives:

\[
P_T = E - E U (I + E U)^{-1} E + W U \left[ (I + E U) (I + E U)^{-1} E - E \right]
\]

\[
P_T = E - E U (I + E U)^{-1} E
\]

which is Eq. (6.4.2).

Equations (6.4.3) and (6.4.4) are equivalent to Eq. (6.2.1). These expanded equations are free from singularity problems, require
fewer matrix inversions than Eq. (6.2.1), avoid forming the result as a difference of two positive definite matrices, and are of a symmetric form.

6.5 A More Useful Form for the New Smoother State Estimation Equation

Calculation of the backward filter state estimate using Eq. (6.3.4) involves the use of the backward filter covariance matrix, P(t), and the terminal condition, \( \dot{\hat{x}}_B(T) \). Both of these are undesirable. The boundary condition on U(T) specified by Eq. (6.2.4) causes P(T) to be infinite and we have not obtained a boundary condition to specify \( \dot{\hat{x}}_B(T) \). These difficulties can be avoided by introducing the new variable:

\[
\dot{w}_B \triangleq U \dot{\hat{x}}_B
\]  \hspace{1cm} (6.5.1)

which has the terminal value:

\[
\dot{w}_B(T) = 0
\]  \hspace{1cm} (6.5.2)

because U(T) = 0.

Differentiating both sides of Eq. (6.5.1) gives:

\[
\dot{\dot{w}}_B = \dot{U} \dot{\hat{x}}_B + U \ddot{\hat{x}}_B
\]  \hspace{1cm} (6.5.3)

Substituting Eqs. (6.2.12) and (6.3.4) into Eq. (6.5.3) and cancelling terms leaves:

\[
\dot{\dot{w}}_B = -F^T U \dot{\hat{x}}_B + U Y U \dot{\hat{x}}_B + U G \bar{u} - H^T R^{-1} z
\]  \hspace{1cm} (6.5.4)

Using Eq. (6.5.1) in Eq. (6.5.4) we obtain:
\[ \hat{w}_B = -(F - Y U)^T w_B + U G \hat{u} - H^T R^{-1} z \quad (6.5.5) \]

Equation (6.5.5) must be integrated backward from the terminal condition specified by Eq. (6.5.2). The resulting value of \( w_B \) can then be used in the calculation of the smoothed state estimate via the equation:

\[ \hat{x}_T = P_T (E^{-1} \hat{x} + w_B) \quad (6.5.6) \]

In this manner we avoid the problems of singularity associated with Eq. (6.3.4).

For stable systems integrating Eq. (6.5.5) backward instead of Eq. (6.3.4) offers an additional numerical advantage. For such stable systems Eq. (6.3.4) is unstable when integrated backward; thus, small errors which arise due to imperfect numerical integration will tend to grow in magnitude as the backward integration proceeds. Near the terminal time (and for some applications for much longer periods) when \( U(t) \) is small Eq. (6.5.5) is stable when integrated backward.

Equation (6.5.6) involves \( E^{-1} \), a matrix inversion not yet required. This can be avoided by replacing it with the matrix inversion needed in the computation of the smoother covariance matrix according to Eqs. (6.4.3) and (6.4.4). To accomplish this first write Eq. (6.5.6) in the form:

\[ \hat{x}_T = P_T E^{-1} \hat{x} + P_T w_B \quad (6.5.7) \]

then postmultiply both sides of Eq. (6.4.1) by \( E^{-1} \).

\[ P_T E^{-1} = (I + E U)^{-1} \quad (6.5.8) \]
Substitution of Eq. (6.5.8) into Eq. (6.5.7) yields the desired result:

\[ \hat{x}_T = (I + E U)^{-1} \hat{x} + P_T w_B \]  \hspace{1cm} (6.5.9)

Equation (6.5.9) enables us to use in the smoothed state estimation the matrix inversion we already needed in the computation of the smoother covariance matrix.

6.6  The Driving Force Estimates and Associated Covariances

An expression for the estimate of the driving force can be easily obtained by comparing the Bryson-Frazier state estimation equation to a rearranged form of Eq. (6.5.6).

Adding and subtracting \( P_T U \hat{x} \) to Eq. (6.5.6) gives:

\[ \hat{x}_T = (P_T E^{-1} + P_T U) \hat{x} + P_T w_B - P_T U \hat{x} \]  \hspace{1cm} (6.6.1)

Premultiplying both sides of Eq. (6.2.1) by \( P_T \) we obtain:

\[ I = P_T E^{-1} + P_T U \]

Using this identity in the first term of Eq. (6.6.1) results in:

\[ \hat{x}_T = \hat{x} + P_T (w_B - U \hat{x}) \]  \hspace{1cm} (6.6.2)

Comparison of Eqs. (3.5.1) and (6.6.2) reveals the relationship:

\[ \lambda = E^{-1} P_T (w_B - U \hat{x}) \]  \hspace{1cm} (6.6.3)
Substitution of Eq. (6.6.3) into Eq. (3.3.5) yields the estimate of the driving force in this new smoother formulation:

\[ \hat{u}_T = \bar{u} + Q G^T E^{-1} P_T (w_B - U \hat{x}) \]  

(6.6.4)

We can avoid inverting the forward filter covariance matrix by applying the transpose of Eq. (6.5.8) to Eq. (6.6.4). The result is:

\[ \hat{u}_T = \bar{u} + Q G^T (I + E U)^{-1} T \left[ w_B - U \hat{x} \right] \]  

(6.6.5)

Equation (6.6.5) gives the smoother estimate of the driving force in the new formulation using the same matrix inversion required for the smoother state estimate and covariance matrix.

The covariance of the errors in the smoothed estimate of the driving noise expressed in this new formulation is easily obtained by comparing Eq. (6.4.2) to the Bryson-Frazier result of Eq. (3.7.25). This comparison reveals the relationship:

\[ \hat{\eta} = U (I + E U)^{-1} \]  

(6.6.6)

Substitution of Eq. (6.6.6) into Eq. (3.9.9) gives the desired result:

\[ Q_T(t) = Q(t) \delta(0) - Q(t) G^T(t) U(t) \left[ I + E(t) U(t) \right]^{-1} G(t) Q(t) \]  

(6.6.7)

6.7 The Discrete Form of the Two Filter Smoother

The derivations presented in the previous five sections for the continuous case can be repeated in a completely analogous fashion for the discrete formulation.
The resulting smoother covariance matrix relationship is:

\[ P_{k/N}^{-1} = E_k^{-1} + P_k'^{-1} \]  \hspace{1cm} (6.7.1)

where \( P_k' \) is the backward filter covariance matrix at the \( k' \)th sample before incorporation of the statistics of the \( k' \)th measurement. Note in Eq. (6.7.1) that there is no overlap of measurements. The forward filter covariance contains the statistics of the \( k' \)th measurement but the backward filter covariance matrix does not. An equivalent formulation which will not be presented here uses \( E_k' \) and \( P_k \) instead of \( E_k \) and \( P_k' \).

A more useful form for computation of \( P_{k/N} \) is:

\[ P_{k/N} = (I - W_k U'_k) E_k (I - W_k U'_k)^T + W_k U'_k W_k^T \]  \hspace{1cm} (6.7.2)

where:

\[ W_k = E_k (I + E_k U'_k)^{-1}^T \]  \hspace{1cm} (6.7.3)

The inverse of the backward filter covariance matrix is processed from the terminal condition \( U_N = 0 \) using the relationships:

\[ J_k = U_k G_{k-1} \begin{bmatrix} G_{k-1}^T U_k G_{k-1} + Q^{-1}_{k-1} \end{bmatrix}^{-1} \]  \hspace{1cm} (6.7.4)

\[ U'_{k-1} = \Phi_{k-1}^T \begin{bmatrix} (I - J_k G_{k-1}^T) U_k (I - J_k G_{k-1}^T)^T \end{bmatrix} + J_k Q^{-1}_{k-1} J_k^T \Phi_{k-1} \]  \hspace{1cm} (6.7.5)

* All sample periods are counted from the beginning of the time interval of interest. At the \( k' \)th sample the backward filter equations have been processed through \( N-k \) cycles.
\[ U_k = U_k' + H_k^T R_k^{-1} H_k = U_k' + S_k \] (6.7.6)

The associated state estimation equation is:

\[ \hat{x}_k/N = P_k/N \left[ E_k^{-1} \hat{x}_k + P_k^{1-1} \hat{z}_{B_k} \right] \] (6.7.7)

For reasons entirely analogous to those cited in Section 6.5, it is desirable to introduce a new variable to describe the backward filter. The discrete equivalent of Eq. (6.5.9) then becomes:

\[ \hat{x}_k/N = (I + E_k U_k')^{-1} \hat{x}_k + P_k/N \tilde{w}_{B_k}' \] (6.7.8)

where the prime indicates the value of \( \tilde{w}_{B_k} \) before incorporation of the \( k \)'th measurement.

The recursion formulas for \( \tilde{w}_{B_k} \) are obtained in a manner which is the discrete equivalent of the derivation of Section 6.5. The results are:

\[ \tilde{w}_{B_k}' = \Phi_k^T \left[ I - U_k G_{k-1} G_{k-1}^T \left( G_{k-1}^T U_k G_{k-1} + Q_{k-1}^{-1} \right)^{-1} G_k^T \right] \]

\[ \left[ \tilde{w}_{B_k}' - U_k G_{k-1} \tilde{u}_{k-1} \right] \]

or

\[ \tilde{w}_{B_k}' = \Phi_k^T \left[ I - \tilde{J}_k G_k^T \right] \left[ \tilde{w}_{B_k}' - U_k G_{k-1} \tilde{u}_{k-1} \right] \] (6.7.9)

\[ \tilde{w}_{B_k}' = \tilde{w}_{B_k} + H_k^T R_k^{-1} \tilde{z}_k \] (6.7.10)
The discrete form of the smoothed driving force estimate is:

\[
\hat{u}_k/N = \bar{u}_k + Q_k G_k^T \Phi_k^{-1} \left[ I + E_k U_k' \right]^{-1} \left[ w_k' - B_k \bar{u}_k + \Delta_k \right]
\] (6.7.11)

and the associated covariance matrix equation is:

\[
Q_k/N = Q_k - Q_k G_k^T \Phi_k^{-1} U_k' \left( I + E_k U_k' \right)^{-1} \Phi_k^{-1} G_k Q_k
\] (6.7.12)

6.8 An Electrical Network Described by the Smoother Equations

In the scalar case the discrete smoother equations describe a passive electrical network. The interpretation of the smoother as a weighted combination of two filters is illustrated by the network which is described by the smoother equations. This network is formed by combining two other networks, one of which is described by the discrete forward filter equations and the other of which is characterized by the backward filter equations.

Figure 6.1 illustrates the network which is described by the discrete forward filter equations and Fig. 6.2 gives a circuit which is characterized by the discrete backward filter equations. In these networks we have either:

\[
H_k = G_k = 1
\]

or:

\[
\bar{u}_k = G_k \bar{u}_k
\]

\[
z_k = z_k/H_k
\]

\[
R_k = R_k/H_k^2
\]

\[
Q_k = Q_k/G_k^2
\]

The circles in Figs. 6.1, 6.2, and 6.3 are a.c. voltage sources and the square in Fig. 6.4 is an a.c. current generator. All circuit elements are assumed to be ideal.
Figure 6.1 Discrete Forward Filter Network Equivalent

Figure 6.2 Discrete Backward Filter Network Equivalent
Figure 6.3 Discrete Smoother Network Equivalent
Figure 6.4 Norton Equivalent Circuit of the Discrete Backward Filter
The state estimates are the voltages across the voltage source-resistor combinations which characterize the measurements. The variances associated with the state estimates are the impedances at the same points. The state transitions are represented by the ideal transformers.

The resistors represent the driving and measurement noise variances. Consider the effect of a typical measurement noise resistor, $R_k'$. If $R_k$ is zero there is no measurement noise and the $k$'th state voltage consists of only the measurement voltage - the state estimate is based only upon the perfect $k$'th measurement. In the opposite case, when $R_k$ is infinite (open circuit) the $k$'th state voltage is not affected by $z_k$ - the state estimate is based solely upon the extrapolated prior estimate and ignores entirely the infinitely bad measurement.

The same is true with respect to the driving noise. A large driving noise (large resistance)$Q_k$ causes the $k$'th state estimate to be less dependent upon the extrapolated estimate (voltages coming from sources other than the $k$'th measurement voltage). A small driving noise resistance causes the opposite effect.

The measurement and driving noise variances (resistors) thus control the relative weighting of the measurement (measurement voltage, $z_k$) and extrapolated prior estimate (in these filters, the effect of all other sources measured at the state estimate position) in forming the state estimate. It is important that these values describe the actual noises involved.

Figure 6.3 represents the smoother network. It is the combination of the filter networks illustrated in Figs. 6.1 and 6.2. Thus the smoother is identical to the filters - the difference is that the estimate (voltage) and variance (impedence) are measured at an intermediate point instead of at the endpoints. Note the infinite impedance (open circuit) at the right end of the networks of Figs. 6.2 and 6.3. This reflects the terminal condition $\mathbf{U}_N = 0$ on the backward filter.
For the reader with a background in electronics these circuits may also provide some physical intuition about the variable \( w^1_{Bk} \) used in the new smoother equations. If the backward filter circuit of Fig. 6.2 is reduced to its Norton equivalent the result is the simple circuit of Fig. 6.4. The current generator in the Norton equivalent circuit generates a current of magnitude \( w^1_{Bk} \).

6.9 A Numerical Example of the Smoother Covariance Matrix Calculation

Table 6.1 is an expanded version of Table 4.5. The results of computing the smoother covariance matrix using Eqs. (6.7.2) and (6.7.3) have been added to the values already presented in Table 4.5. The system used is the same as that described in Sections 4.3 and 4.4.

For this example, the forward-backward smoother equations give much more accurate results than any of the other formulations, better even than the results obtained using Eq. (4.4.2). Equations (4.4.2) and (6.7.2) both express the smoother covariance matrix as a sum of positive definite matrices; neither base the result on the difference of almost equal quantities. The explanation of the superiority of Eq. (6.7.2) over Eq. (4.4.2) thus probably lies in an argument similar to that offered by Joseph (1964) for the forward filter. When Eqs. (4.4.1) and (4.4.2) are used to evaluate the smoother covariance matrix, errors in the weighting factor \( C_k \) contribute to errors in \( P_k/N \) to first order. This is demonstrated by Eq. (4.4.3). When we apply the same argument to Eqs. (6.7.2) and (6.7.3) we discover that errors in the weighting factor \( W_k \) contribute to the error in \( P_k/N \) only to second order, that is:

\[
\delta P_k/N = 0.5 \delta W_k
\]

By this argument, Eq. (6.7.2) should be numerically superior to Eq. (4.4.2). The results given in Table 6.1 do show this superiority. Note however, that this argument is valid only if there is an error in \( W_k \).
<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct $P_{k/N}$</td>
<td>3.171875458 -07</td>
</tr>
<tr>
<td></td>
<td>-5.678677951 -07</td>
</tr>
<tr>
<td>$P_{k/N}$ by Eq. (3.11.9)</td>
<td>3.170092624 -07</td>
</tr>
<tr>
<td></td>
<td>-5.715179800 -02</td>
</tr>
<tr>
<td>$P_{k/N}$ by Eq. (4.4.2)</td>
<td>3.173207111 -07</td>
</tr>
<tr>
<td></td>
<td>-5.682210880 -05</td>
</tr>
<tr>
<td>$P_{k/N}$ by Eq. (6.7.2)</td>
<td>3.172123206 -07</td>
</tr>
<tr>
<td></td>
<td>-5.679121439 -07</td>
</tr>
<tr>
<td>Correct $P_{k-1/N}$</td>
<td>4.643005134 -07</td>
</tr>
<tr>
<td></td>
<td>-8.294412540 -07</td>
</tr>
<tr>
<td>$P_{k-1/N}$ by Eq. (3.11.9)</td>
<td>2.534000000 -04</td>
</tr>
<tr>
<td></td>
<td>-2.468100000 -03</td>
</tr>
<tr>
<td>$P_{k-1/N}$ by Eq. (4.4.2)</td>
<td>4.797986952 -07</td>
</tr>
<tr>
<td></td>
<td>-9.788842110 -07</td>
</tr>
<tr>
<td>$P_{k-1/N}$ by Eq. (6.7.2)</td>
<td>4.643252560 -07</td>
</tr>
<tr>
<td></td>
<td>-8.293936553 -07</td>
</tr>
</tbody>
</table>
6.10 Application of the Forward-Backward Smoother Equations to Linearized Nonlinear Systems

When linearizing a set of nonlinear equations in order to apply these linear filtering and smoothing techniques we can either keep the reference solution fixed or update it using the filter and smoother state estimates. If we do not change the reference solution using the filter and smoother state estimates we can directly apply the linear filter and smoother theory to the resulting linear equations which describe the deviation from the reference solution. To obtain the state estimate at any time we simply add the best estimate of the deviation at that time to the reference solution.

If we desire to update the reference solution at each sample period (discrete case) on the basis of the best estimate of the deviation we must add the deviation to the reference and appropriately redefine the deviation. In the case of the forward filter the estimate of the deviation is set to zero after it is incorporated into the reference solution. In practice we never actually have to assign storage to the deviation; it appears as an additive term in the state update equation. The forward filter equations are:

\[
\begin{align*}
\hat{x}_k^t &= \hat{x}_{k-1}^t + \Phi_{k-1} \bar{u}_k - Y_{k-1} \\
E_k^t &= \Phi_{k-1} E_{k-1} \Phi_{k-1}^T + Y_{k-1} \\
K_k &= E_k^t H_k^T (H_k E_k^t H_k^T + R_k)^{-1} \\
\hat{x}_k &= \hat{x}_k^t + K_k \left[ z_k - h (\hat{x}_k, k) \right] \\
E_k &= (I - K_k H_k) E_k^t (I - K_k H_k)^T + K_k R_k K_k^T
\end{align*}
\]

(6.10.1) (6.10.2) (6.10.3) (6.10.4) (6.10.5)

where:

\[
z_k = h (x_k, k)
\]

and where \( f \), the nonlinear function which describes the state transition,
is evaluated using the most recent state estimate. The state transition matrix for the linearized system, $\Phi_k$, is evaluated using

$$\frac{\partial f_k}{\partial \hat{x}_k} \bigg|_{\hat{x}_k = \hat{x}_k}.$$ 

Note that the forward filter estimate of the difference between the actual state state and the reference state is the second term on the right side of Eq. (6.10.4). No storage space was required for this deviation vector.

The application of the forward-backward smoothing technique to the linear deviation equations when the reference solution is updated on the basis of the smoothed state estimate requires more bookkeeping than the application of the forward filter equations. To understand this, we will first introduce the linearized discrete equivalent of Eq. (6.5.1):

$$w_{B_k} = U_k (\hat{x}_{B_k} - \hat{x}_{R_k}) \quad (6.10.6)$$

where $\hat{x}_{R_k}$ is the reference solution at the k'th sample period. If we change the reference solution at this k'th sample by an amount $\delta \hat{x}_{R_k}$ the corresponding change in $w_{B_k}$ is given by:

$$\delta w_{B_k} = -U_k \delta \hat{x}_{R_k} \quad (6.10.7)$$

This means that $w_{B_k}$ does not in general become zero when the reference solution is updated as was the case for the deviation when we were considering the forward filter solution. Because of this, we must process the equations which extrapolate $w_{B_k}$ backward
to \( \mathbf{w}_{\mathbf{B}_{k+1}}' \) an extrapolation procedure which was not necessary for the forward filter deviation because it was set to zero after a reference update. The computational procedure is thus summarized by the following sets of equations:

1. Backward extrapolation:

\[
J_k = U_k G_{k-1}^T (G_{k-1} U_k G_{k-1} + Q_{k-1}^{-1})^{-1}
\]  
(6.10.8)

\[
U'_{k-1} = \Phi_{k-1} \left[ (I - J_k G_{k-1}^T) U_k (I - J_k G_{k-1}^T)^T + J_k Q_{k-1}^{-1} J_k^T \right] \Phi_{k-1}
\]  
(6.10.9)

\[
\mathbf{w}'_{\mathbf{B}_{k-1}} = \Phi_{k-1} \left[ I - J_k G_{k-1}^T \right] \left[ \mathbf{w}_{\mathbf{B}_k} - U_k G_{k-1} \bar{x}_{k-1} \right]
\]  
(6.10.10)

\[
\mathbf{x}_{\mathbf{R}_{k-1}} = g(x_{\mathbf{R}_k}, u_{k-1}, k-1)
\]  
(6.10.11)

where \( g \) is the nonlinear function which describes the backward state extrapolation. The state transition matrix \( \Phi_k \) is computed using

\[
\frac{\partial f_k}{\partial \mathbf{x}_k} \bigg|_{\mathbf{x}_k = \mathbf{x}_{\mathbf{R}_k}}
\]

Both \( f \) and \( g \) are evaluated using \( \mathbf{x}_{\mathbf{R}_k} \).

2. Update:

\[
W_k = E_k (I + E_k U_k')^{-1}^T
\]  
(6.10.12)

\[
P_{k/N} = (I - W_k U_k') E_k (I - W_k U_k')^T + W_k U_k' W_k^T
\]  
(6.10.13)
\[ \hat{x}_k / N = x_{R_k} + (1 + E_k U_k')^{-1} (\hat{x}_k - x_{R_k}) + P_{k/N} w'_{k} \quad (6.10.14) \]

\[ w'_{B_k} = w_{B_k} - U_k' (\hat{x}_k/N - x_{R_k}) \quad (6.10.15) \]

\[ x_{R_k} = \hat{x}_k / N \quad (6.10.16) \]

\[ w_{B_k} = w_{B_k} + H_k^T R_k^{-1} \left[ z_k - h (x_{R_k}, k) \right] \quad (6.10.17) \]

\[ U_k = U_k + S_k \quad (6.10.18) \]

Equations (6.10.15) and (6.10.16) describe the effect of updating the reference solution. If the reference solution is not to be updated at every sample period, these two equations should be processed at those samples for which no reference update is desired.

In the continuous case simply apply the continuous smoother techniques to the linear deviation equations until a reference solution update is desired. At these times, add the smoothed deviation to the reference and redefine the deviation to be zero. Before setting the deviation to zero, correct \( w_B \):

\[ w_B = w_B + U \epsilon_T \quad (6.10.19) \]

where \( \epsilon_T \) is the smoother error state.

6.11 Summary

In this chapter a new formulation of the smoother equations has been presented. The smoother estimate is expressed as a weighted combination of two Kalman filter estimates, one of which is the estimate of a filter running forward over the data and the other of which is the estimate of a filter which proceeds backwards over the data. The weighting factors are proportional to the confidences in each estimate.
Not only is this formulation appealing because it readily lends itself to physical understanding, it also is less sensitive to numerical problems than the older formulations. An example of this superiority appears in Section 6.9.

The key equations for the continuous formulation are Eqs. (6.4.3), (6.4.4), (6.5.5) and (6.5.6). The corresponding equations for the discrete case are Eqs. (6.7.2), (6.7.3), (6.7.8), (6.7.9) and (6.7.10).
CHAPTER VII

A METHOD OF FILTERING WHEN THERE IS INSUFFICIENT DATA TO COMPLETELY DETERMINE THE STATE

7.1 Introduction

In this chapter we will explore in more detail the techniques introduced in Sections 6.4 and 6.5. Sufficient material was presented in Sections 6.4, 6.5, 6.7, and 6.10 to enable the user to apply these techniques to smoothing without reading this chapter. The purpose here is to demonstrate that these methods can be used during filtering to obtain parts of the state even when there is not enough information available to determine the entire state.

This situation of being able to determine some but not all of the dimensions of the state arises in practice when there is no a priori state information and the measurements are discrete with dimension less than that of the state. The backward filter used in smoothing can be of this nature. As an example of this, consider the problem of estimating a state consisting of position, velocity, and acceleration from discrete measurements of position when no initial state information is available. With one measurement we can estimate position but not velocity or acceleration. With two measurements we can obtain position and difference to obtain velocity but we cannot estimate acceleration. With three or more measurements we can completely determine the state. Another example of a practical application where a priori information is often not available is the orbit determination problem. Gauss (1809) determined the orbit of the asteroid Ceres without any a priori information. He used his method of least squares to reduce the astronomical observations which were sent to him.

The usual method of treating such cases is to batch process
enough measurements to uniquely determine the state and then begin the filtering process. In those applications where there is no driving noise the generalized inverse of a matrix* may be used to obtain the minimum norm estimate of the state when there are insufficient measurements to completely determine the state. Both of these techniques are discussed by Lee (1964). The technique to be presented here enables the user to determine which dimensions can be obtained from the available measurements, provides a method for estimating these dimensions, is directly applicable even when there is driving noise, and is completely recursive from the beginning of the problem. This author is not aware of any previous development of this technique in the literature.

7.2 Development of the Forward Cofilter

In this section, we will derive the forward equivalent of Eq. (6.5.5) from a special case of the smoother cost function given by Eq. (3.3.1). The resulting equations and their discrete equivalents are the key parts of the technique considered in this chapter and in Sections 6.4 and 6.5; consequently we will find it as useful to refer to them by name as we have in referring to Eqs. (3.4.4) and (3.4.5) and their discrete equivalents as a forward filter. We will refer to Eqs. (6.2.12) and (6.5.5) and their discrete equivalents as a backward cofilter and to the forward versions of these equations as a forward cofilter. The main reasons for the choice of the name cofilter are that \( \hat{w}_B \) (in the forward case, \( w \)) has the same dimensions as the costate and bears a relationship to \( \hat{x}_B \) (in the forward case \( \hat{x} \)) which is similar to that between the smoother state and costate. The variables \( w \) and \( \hat{w}_B \) are not, however, the costates for the filters - these are zero. This may be seen by recalling that the filter costate can be obtained by considering the terminal time, \( T \), in the smoother equations.

*For details of the generalized inverse and its use, see Penrose (1955); Penrose (1956); and Rust, Burrus, and Schneeberger (1966).
to be the current time. The boundary condition on \( \lambda(T) \) given by Eq. (3.3.10) thus causes the filter costate to be zero.

We are considering here the problem of state estimation when there is no prior estimate. This is equivalent to having any finite a priori state estimate with an associated infinite covariance. To modify the cost function accordingly, we must set \( E^{-1}(t_0) \) in Eq. (3.3.1) to zero. If we then follow the same procedure as in Section 3.3 we obtain the same differential equations (Eqs. (3.3.11) and (3.3.12)) with a different set of boundary conditions. The new set of boundary conditions is:

\[
\begin{align*}
\lambda(t_0) &= 0 \\
\lambda(T) &= 0
\end{align*}
\]  

(7.2.1)

We will now proceed as in Appendix B by converting the smoother equations (in particular, Eqs. (3.3.11) and (3.3.12)) into filter equations by considering the terminal time, \( T \), to be the variable current time, \( t \). In this way we eliminate all information which is in the future with respect to \( t \). Since the filter and smoother are equivalent at the final time (which is now the current time) the smoothing equations describe the filter when they are evaluated at the final time. Using the state transition matrix defined by Eq. (B.4) and the boundary conditions provided by Eq. (7.2.1) we may write:

\[
M_{11} \hat{x}(t_0) + m_1 = \hat{x}(t)
\]  

(7.2.2)

\[
M_{21} \hat{x}(t_0) + m_2 = 0
\]  

(7.2.3)

Solving Eq. (7.2.2) for \( \hat{x}(t_0) \) and substituting the result into Eq. (7.2.3) yields:

\[
M_{21} M_{11}^{-1} \hat{x}(t) = M_{21} M_{11}^{-1} m_1 - m_2
\]  

(7.2.4)
Next, we will show that for this problem:

\[ V = M_{21}^{-1} M_{11}^{-1} \]  \hspace{1cm} (7.2.5)

or, in words, that \( M_{21}^{-1} M_{11}^{-1} \) is the inverse of the forward filter covariance matrix when this covariance matrix is infinite at time \( t_0 \). Equation (7.2.5) is valid at time \( t_0 \) because \( M_{21} \) is zero initially; consequently:

\[ V(t_0) = 0 \]  \hspace{1cm} (7.2.6)

To show that Eq. (7.2.5) is true at all times, we will demonstrate that the derivatives of both sides are identical. Differentiating the right side of Eq. (7.2.5) gives:

\[ \frac{d}{dt} (M_{21}^{-1} M_{11}^{-1}) = \dot{M}_{21}^{-1} M_{11}^{-1} - M_{21}^{-1} \dot{M}_{11}^{-1} \]  \hspace{1cm} (7.2.7)

Substituting for \( \dot{M}_{21} \) and \( \dot{M}_{11} \) from Eq. (B.4) into Eq. (7.2.7) we obtain:

\[ \frac{d}{dt} (M_{21}^{-1} M_{11}^{-1}) = S - F^T M_{21}^{-1} M_{11}^{-1} - M_{21}^{-1} \dot{M}_{11}^{-1} F - M_{21}^{-1} Y M_{21}^{-1} \]  \hspace{1cm} (7.2.8)

Using Eq. (7.2.5) in Eq. (7.2.8) there results:

\[ \frac{d}{dt} (M_{21}^{-1} M_{11}^{-1}) = S - F^T V - V F - V Y V \]  \hspace{1cm} (7.2.9)

Comparing Eq. (7.2.9) with Eq. (6.2.7) we may conclude:
\[ \dot{V} = \frac{d}{dt} (M_{21} M_{11}^{-1}) \]

thus, Eq. (7.2.5) is true for all time.

If we next define:

\[ w \triangleq \dot{V} \hat{x} \quad (7.2.10) \]

we may then combine Eqs. (7.2.4), (7.2.5), and (7.2.10) to yield:

\[ w = V m_1 - \dot{m}_2 \quad (7.2.11) \]

We thus have:

\[ w(t_0) = 0 \quad (7.2.12) \]

because as a consequence of their definitions (Eqs. (B.5) and (B.6)) \( m_1(t_0) \) and \( \dot{m}_2(t_0) \) are both zero.

Differentiating Eq. (7.2.11) we obtain:

\[ \dot{w} = \dot{V} \dot{m}_1 + V \dot{m}_1 - \dot{m}_2 \]

Substituting for the derivatives on the right side from Eqs. (6.2.7), (B.19), and (B.20) and cancelling terms yields:

\[ \dot{w} = -F^T (V m_1 + m_2) + H^T R^{-1} z - V Y (V m_1 + m_2) + V G \bar{u} \quad (7.2.13) \]

Substitution of Eq. (7.2.11) into Eq. (7.2.13) yields the desired result:
\[
\dot{\bar{w}} = -F^T \bar{w} + H^T R^{-1} \bar{z} - V Y \bar{w} + V \bar{G} \bar{u}
\]  

(7.2.14)

Equations (6.2.7), (7.2.6), (7.2.12), and (7.2.14) describe the continuous form of the forward cofilter.

An analogous derivation may be performed for the discrete case. The results are:

\[
V_0 = 0
\]  

(7.2.15)

\[
\bar{w}_0 = 0
\]  

(7.2.16)

\[
V_k = V_k' + S_k
\]  

(7.2.17)

\[
\bar{w}_k = w_k' + H_k^T R_k^{-1} z_k
\]  

(7.2.18)

\[
J_k = V_k \Phi_k^{-1} G_k \left( G_k^T \Phi_k^{-1} V_k \Phi_k^{-1} G_k + Q_k^{-1} \right)^{-1}
\]  

(7.2.19)

\[
V_{k+1}' = \Phi_k^{-1} \left[ \left( I - J_k G_k^T \Phi_k^{-1} \right) V_k \left( I - J_k G_k^T \Phi_k^{-1} \right)^T + J_k Q_k^{-1} J_k^T \right] \Phi_k
\]  

(7.2.20)

\[
w_{k+1}' = \Phi_k^{-1} \left[ I - J_k G_k^T \Phi_k^{-1} \right] \left[ \bar{w}_k + V_k \Phi_k^{-1} G_k \bar{u}_k \right]
\]  

(7.2.21)

where

\[
\bar{w}_k = V_k \hat{x}_k
\]  

(7.2.22)

\[
w_k' = V_k' \hat{x}'_k
\]  

(7.2.23)

If there is no driving noise Eqs. (7.2.19) through (7.2.21) are replaced with:

\[
V_{k+1}' = \Phi_k^{-1} V_k \Phi_k^{-1}
\]  

(7.2.24)
\[ w_{k+1}^i = \Phi_k^{-1} \left[ w_k + V_k \Phi_k^{-1} G_k \bar{u}_k \right] \]  \hspace{1cm} (7.2.25)

Note that when \( V_k \) is nonsingular we can obtain an estimate of the entire state:

\[ \hat{x}_k = V_k^{-1} w_k \]  \hspace{1cm} (7.2.26)

This condition exists only after enough measurements have been taken to determine all the dimensions of the state and we can then always obtain the state estimate using Eq. (7.2.26). In practice, however, we can avoid repeating the matrix inversion needed in Eq. (7.2.24) by using Eq. (7.2.26) once at the first point all the dimensions of the state can be obtained. After this, we would process the forward filter equations instead of the forward cofilter equations.

7.3 State Estimation with Insufficient Information

When \( V_k \) is singular we cannot completely determine the state because of the matrix inversion required in Eq. (7.2.26). In the case under consideration (no a priori information available) \( V_k \) is singular because inability to estimate some or all of the dimensions of the state with the information available makes \( E_k \) infinite. Although we cannot completely determine the state when \( V_k \) is singular we may be able to estimate certain dimensions or combinations of dimensions of the state. In this section we will present a method for doing this.

If we wish to estimate a quantity \( \xi \) which is related to the state estimate according to:

\[ \xi = a^T \hat{x}_k \]  \hspace{1cm} (7.3.1)

we can obtain \( \xi \) if we can find a vector \( b \) which satisfies the equation:

\[ V_k b = a \]  \hspace{1cm} (7.3.2)
To show that we can obtain \( \xi \) if a solution to Eq. (7.3.2) exists first transpose Eq. (7.3.2) and substitute the result for \( a^T \) in Eq. (7.3.1). The result is:

\[
\xi = b^T V_k \hat{x}_k. \tag{7.3.3}
\]

Substitution of Eq. (7.2.22) into Eq. (7.3.3) gives:

\[
\xi = b^T w_k \tag{7.3.4}
\]

Thus, if we can find a vector \( b \) to satisfy Eq. (7.3.2) we can obtain \( \xi \) using the cofilter and Eq. (7.3.4). Note that this procedure is not necessary when smoothing because of the availability of the forward filter estimate during the period when the backward filter has insufficient information to completely determine the state.

Solutions to Eq. (7.3.2) exist only if the rank of the rectangular matrix formed by augmenting \( a \) to \( V_k \) (adding \( a \) to \( V_k \) as another column) has the same rank as \( V_k \) itself*. We can thus determine whether or not we can find a particular \( \xi \) by augmenting the corresponding \( a \) to \( V_k \) and comparing the rank of the augmented matrix to the rank of \( V_k \).

The associated mean squared error in the estimate, \( \sigma^2 \), is given by:

\[
\sigma^2 = a^T b \tag{7.3.5}
\]

This result can be derived starting with the relationship:

\[
\epsilon_\xi = a^T \epsilon_k
\]

where \( \epsilon_\xi \) is the error in \( \xi \) and \( \epsilon_k \) is the discrete equivalent of the forward filter error defined by Eq. (3.7.2).

* For a definition of these terms and a discussion of solutions to such linear equations see Hildebrand (1952), pp. 21-29.

116
With this definition we have:

\[ \sigma^2 = \bar{e}^2_\xi = \bar{a}^T \mathbf{e}_k \mathbf{e}_k^T \bar{a} \]

\[ = \bar{a}^T \mathbf{E} \bar{a} \]

\[ = \bar{a}^T \mathbf{V} \mathbf{b} \]

\[ \sigma^2 = \bar{a}^T \mathbf{b} \]

which is Eq. (7.3.5).

As a numerical example of this technique, consider the first measurement in a two dimensional system with:

\[ H_1 = [1, 0] \quad z_1 = 2 \quad R_1 = 1 \]

Using Eqs. (7.2.15) through (7.2.18) we obtain:

\[ V_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \]

\[ w_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \]

To determine the first dimension of the state we form \( \xi \) with:

\[ \xi = [1, 0] \hat{\omega} \]

The augmented matrix is:
\[
\begin{bmatrix}
1 & 0 & 1 \\
0 & 0 & 0
\end{bmatrix}
\]

which has rank one as does \( V_1 \). Thus, we can obtain an estimate of
the first dimension. To estimate the second dimension we would define
\[
\mathbf{a} = \begin{bmatrix}
0 \\
1
\end{bmatrix}
\]

The corresponding augmented matrix is:
\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

which has a rank of two. Thus, we cannot estimate dimension two of
the state with only this first measurement.

The solution to Eq. (7.3.2) for the case of:
\[
\mathbf{a} = \begin{bmatrix}
1 \\
0
\end{bmatrix}
\]

is:
\[
\mathbf{b} = \begin{bmatrix}
1 \\
c
\end{bmatrix}
\]

where \( c \) is an arbitrary constant.

Using Eq. (7.3.4) we obtain the estimate of the first dimension
of the state as:
\[ \xi = 2 \]

The associated mean square error in this estimate computed using Eq. (7.3.5) is:

\[ \sigma^2 = 1 \]

### 7.4 Uniqueness of the Solution

Note in the example of the previous section that the constant \( c \) did not appear in the solution for the estimate of the first dimension of the state. This constant is the parameter that characterized the one parameter family of solutions to Eq. (7.3.2) for that example. In general there will be as many arbitrary parameters as the difference between the order and rank of \( V_k \). We will demonstrate in this section that the solutions obtained using Eq. (7.3.4) are independent of these arbitrary parameters; thus the solutions are unique.

The solution to Eq. (7.3.2) is not unique. To any particular solution, \( b_p \) of Eq. (7.3.2) we may add a null vector of \( V_k \) and still have a valid solution:

\[ b = b_p + n_k \]

where the null vector \( n_k \) is a solution to:

\[ V_k n_k = 0 \]

We may think of these null vectors as scaled unit vectors. These unit vectors define directions in state space along which we have no information and there are as many independent unit vectors as the difference between the dimension and rank of \( V_k \). The arbitrary parameters referred to above are the constants which scale these unit vectors. There is thus the same number of arbitrary parameters as there are unit null vectors.
We will now demonstrate that these parameters do not appear in $\xi$ because the null vectors are orthogonal to the $w_k$. We will employ an inductive reasoning process which will be based upon showing that if this is true during an arbitrary extrapolation it is true during an arbitrary update and vice versa. By demonstrating that it is true at any point we can then conclude that it is always true.

We will first consider an arbitrary update by premultiplying Eqs. (7.2.17) and (7.2.18) by $n_k^T$, a null vector of $V_k$:

$$n_k^T V_k = 0 = n_k^T V_k' + n_k^T S_k$$  \hspace{1cm} (7.4.1)

$$n_k^T w_k = n_k^T w_k' + n_k^T H_k^T R_k^{-1} z_k$$  \hspace{1cm} (7.4.2)

Next postmultiply Eq. (7.4.1) by $n_k$:

$$n_k^T V_k' n_k + n_k^T S_k n_k = 0$$  \hspace{1cm} (7.4.3)

Since $V_k'$ and $S_k$ are at least positive semidefinite, each term on the left side of Eq. (7.4.3) must be zero:

$$n_k^T V_k' n_k = 0$$  \hspace{1cm} (7.4.4)

$$n_k^T S_k n_k = n_k^T H_k^T R_k^{-1} H_k n_k = 0$$  \hspace{1cm} (7.4.5)

Define:

$$\gamma = V_k' n_k$$  \hspace{1cm} (7.4.6)

120
Then write the quadratic form:

\[(n_k + c \gamma)^T V_k' (n_k + c \gamma) \geq 0\]

where \(c\) is an arbitrary constant.

Expanding using Eqs. (7.4.4) and (7.4.6) to eliminate terms we obtain:

\[2c |\gamma|^2 + c^2 \gamma^T V_k' \gamma \geq 0\]  \hspace{1cm} (7.4.7)

If:

\[\gamma^T V_k' \gamma = 0\]

the only way to satisfy Eq. (7.4.7) is:

\[\gamma = 0.\]

If:

\[\gamma^T V_k' \gamma > 0\]

we can violate Eq. (7.4.7) by choosing \(c\) such that:

\[c = -\frac{|\gamma|^2}{\gamma^T V_k' \gamma}\]

Since \(\gamma^T V_k' \gamma\) can never be less than zero the only way Eq. (7.4.7) can be true is if:
\[ \gamma = V'_k n_k = 0 \]  (7.4.8)

Next define

\[ \beta = H_k n_k \]

then Eq. (7.4.5) becomes:

\[ \beta^T R_k^{-1} \beta = 0 \]  (7.4.9)

Since \( R_k^{-1} \) is positive definite the only way to satisfy Eq. (7.4.9) is to demand that:

\[ \beta = H_k n_k = 0 \]  (7.4.10)

Using Eq. (7.4.10) we can deduce that the second term on the right side of Eq. (7.4.2) is zero.

Equation (7.4.8) shows that \( n_k \) is a null vector of \( V'_k \). If we assume for the moment that null vectors of \( V'_k \) are orthogonal to \( w'_k \) (we will show this next) the first term on the right side of Eq. (7.4.2) is also zero. Thus

\[ n_k^T w_k = 0 \]

or null vectors of \( V_k \) are orthogonal to \( w_k \).

We will begin the consideration of an arbitrary extrapolation by premultiplying Eq. (7.2.21) by \( n'_{k+1} \), a null vector of \( V'_{k+1} \):

\[ n'_{k+1}^T w'_{k+1} = n'_{k+1}^T \phi'_{k+1} \left[ I - J_k G_k T \phi'_{k+1} T^{-1} \right] \left[ w_k + V_k \phi'_{k+1} G_k u_k \right] \]  (7.4.11)
Next we will simplify Eq. (7.4.11) by demonstrating that:

\[ n_{k+1}^T \Phi_k^{-1} V_k = 0 \]  
(7.4.12)

To do this, use Eqs. (4.3.7) and (4.3.8) to express \( V_k \) in terms of \( V_{k+1}^i \). The result is:

\[ V_k = \Phi_k^T \left[ V_{k+1}^i - V_{k+1}^i G_k \left( G_k^T V_{k+1}^i G_k - Q_k^{-1} \right)^{-1} G_k^T V_{k+1}^i \right] \Phi_k \]  
(7.4.13)

Substituting Eq. (7.4.13) into Eq. (7.4.12) we obtain:

\[ n_{k+1}^T V_{k+1}^i \left[ I - G_k \left( G_k^T V_{k+1}^i G_k - Q_k^{-1} \right)^{-1} G_k^T V_{k+1}^i \right] \Phi_k = 0 \]  
(7.4.14)

Since \( n_{k+1}^i \) is a null vector of \( V_{k+1}^i \) Eq. (7.4.14) is true and thus Eq. (7.4.12) is also true. Using Eq. (7.4.12) together with Eq. (7.2.19) we may reduce Eq. (7.4.11) to:

\[ n_{k+1}^T w_{k+1}^i = n_{k+1}^T \Phi_k^{-1} \]

But according to Eq. (7.4.12) \( n_{k+1}^T \Phi_k^{-1} \) is a null vector of \( V_k \). If null vectors of \( V_k \) are indeed orthogonal to \( w_k \) we may conclude:

\[ n_{k+1}^T w_{k+1}^i = 0 \]

or null vectors of \( V_{k+1}^i \) are null vectors of \( w_{k+1}^i \).

We have thus far demonstrated that the orthogonality property remains valid during an arbitrary update if it remains true during an
arbitrary extrapolation and vice versa. To complete this proof by induction we must demonstrate that it is true at some point in the iterative procedure. If it is true at any point it will then be true at all times.

The boundary condition provided by Eq. (7.2.16) enables us to write:

\[ n_0^T \mathbf{w}_0 = 0 \]

Thus we may make the general conclusion that null vectors of \( V_k \) and \( V'_k \) are orthogonal to the corresponding \( \mathbf{w}_k \) and \( \mathbf{w}'_k \).

Since all the arbitrary parameters involved in the solution to Eq. (7.3.2) are contained in the null vectors these parameters do not appear in \( \xi \) when using Eq. (7.3.4). Thus the solutions obtained with Eq. (7.3.4) are unique.

7.5 **Relationship to Observability**

When we are able to estimate some dimensions of the state and not others what we are doing is estimating those dimensions which are observable with the number of measurements available. Those dimensions of the state which are unobservable with these measurements are the ones we cannot estimate. As outlined in Section 7.3 those dimensions or combinations of dimensions which we can estimate are those for which we can obtain a solution to Eq. (7.3.2). An equivalent statement is that we can obtain those \( \xi \)'s (see Eq. (7.3.1)) for which the corresponding \( \mathbf{a} \)'s are orthogonal to the null vectors of \( V_k \). This is most easily understood by referring to Hildebrand (1952) who shows that the equation:

\[ V_k \mathbf{b} = \mathbf{a} \]

possesses a solution if and only if \( \mathbf{a} \) is orthogonal to the null vectors
of $V_k^T$. Since $V_k$ is symmetric the $a$'s must also be orthogonal to the
null vectors of $V_k$. In this section we will show that the rows of the
observability matrix* at the k'th sample period are orthogonal to the
null vectors of $V_k$. Thus the rows of the observability matrix can be
used to select dimensions and combinations of dimensions which we
can estimate. This is one more way of stating the fact that we can
estimate those dimensions of the state which are observable with the
measurements available.

If we neglect the effects of noise and the driving force we can
relate the state at the k'th sample period to the first k measurements
in the following fashion:

\[
\begin{align*}
  z_k & = H_k x_k \\
  z_{k-1} & = H_{k-1} \Phi_{k-1}^{-1} x_k \\
  z_{k-2} & = H_{k-2} \Phi_{k-2}^{-1} \Phi_{k-1}^{-1} x_k \\
  \vdots & \\
  z_1 & = H_1 \Phi_1^{-1} \Phi_2^{-1} \cdots \Phi_{k-1}^{-1} x_k
\end{align*}
\]

where we have assumed scalar measurements. The matrix $H_k$ is
thus a row vector. Rewriting these equations we have:

\[
\begin{bmatrix}
  z_k \\
  z_{k-1} \\
  z_{k-2} \\
  \vdots \\
  z_1
\end{bmatrix}
= \begin{bmatrix}
  H_k \\
  H_{k-1} \Phi_{k-1}^{-1} \\
  H_{k-2} \Phi_{k-2}^{-1} \Phi_{k-1}^{-1} \\
  \vdots \\
  H_1 \Phi_1^{-1} \Phi_2^{-1} \cdots \Phi_{k-1}^{-1}
\end{bmatrix}
\begin{bmatrix}
  x_k \\
  A x_k
\end{bmatrix}
\]

* We are interested here in the incomplete observability matrix which
is formed for the first k measurements where k < n. See Lee (1964),
pg. 83, or Tou (1964), pp. 154-155,for alternate presentations of the
observability matrix.
where the matrix $A$ is the observability matrix when $k$ equals the
dimension of the state. In this discussion we are interested in the
case where $k < n$. We will now demonstrate that the rows of $A$ are
orthogonal to the null vectors of $V_k$ by using some of the results pre-
sented in the previous section.

First, from Eq. (7.4.10) we know:

$$H_k n_k = 0$$

(7.5.1)

Next we must show that:

$$H_{k-1} \Phi_{k-1}^{-1} n_k = 0$$

(7.5.2)

From Eq. (7.5.1) we know that null vectors of $V_{k-1}$ are orthogonal to
$h_{k-1}$. If we can demonstrate that $\Phi_{k-1}^{-1} n_k$ is a null vector of $V_{k-1}$ we
may then conclude that Eq. (7.5.2) is true. Thus, to demonstrate the
validity of Eq. (7.5.2) we must show that:

$$V_{k-1} \Phi_{k-1}^{-1} n_k = 0$$

(7.5.3)

Substitution of Eq. (7.4.13) into Eq. (7.5.3) yields:

$$\Phi_{k-1}^T \left[ I - V_k G_{k-1} \left( G_{k-1}^T V_k G_{k-1} - Q_{k-1}^{-1} \right)^{-1} G_{k-1}^T \right] V_k n_k = 0$$

(7.5.4)

Applying Eq. (7.4.8) in Eq. (7.5.4) we obtain an identity.

Hence, Eq. (7.5.3) is true and thus Eq. (7.5.2) is also true.

To show that:
\[ H_{k-2} \Phi_{k-2}^{-1} \Phi_{k-1}^{-1} n_k = 0 \]  \hspace{1cm} (7.5.5)

we will use the fact that if \( \Phi_{k-2}^{-1} \Phi_{k-1}^{-1} n_k \) is a null vector of \( V_{k-2} \) Eq. (7.5.5) is true by virtue of Eq. (7.5.1). Thus we must show that

\[ V_{k-2} \Phi_{k-2}^{-1} \Phi_{k-1}^{-1} n_k = 0 \]  \hspace{1cm} (7.5.6)

Using Eq. (7.4.13) in Eq. (7.5.6) we obtain:

\[
\Phi_{k-2}^T \left[ I - V_{k-1}' G_{k-2} \left( G_{k-2}' V_{k-1} G_{k-2} - Q_{k-2}^{-1} \right)^{-1} G_{k-2}' \right] V_{k-1}' \Phi_{k-1}^{-1} n_k = 0
\]

(7.5.7)

From Eq. (7.5.3) we know that \( \Phi_{k-1}^{-1} n_k \) is a null vector of \( V_{k-1} \) and from Eq. (7.4.8) we know that null vectors of \( V_{k-1}' \) are also null vectors of \( V_{k-1}' \). Thus:

\[ V_{k-1}' \Phi_{k-1}^{-1} n_k = 0 \]

consequently Eq. (7.5.7) is an identity. Hence Eq. (7.5.6) is true thus Eq. (7.5.5) is also valid.

Repeating this procedure for the other rows of the matrix \( A \) reveals the fact that all the rows of the observability matrix formed for the first \( k \) measurements are orthogonal to the null vectors of \( V_k' \). When these rows or linear combinations of these rows are used as in Eq. (7.3.1) solutions for the corresponding \( \xi \) will exist.

7.6 Proof that the Estimator is Unbiased

In this section we will show that if there is no driving noise and all the measurements are perfect the estimated state vector
coincides with the actual state after enough information is available to completely determine the state. Estimators with this characteristic are called unbiased estimators*.

We begin by defining:

$$\omega \hat{=} V e$$  \hspace{1cm} (7.6.1)

where $e$ is the forward filter error defined by Eq. (3.7.2). Note that:

$$\omega(0) = 0$$  \hspace{1cm} (7.6.2)

because:

$$V(0) = 0$$

Differentiating Eq. (7.6.1) yields:

$$\dot{\omega} = \dot{V} e + V \dot{e}$$

Substituting for $\dot{V}$ and $\dot{e}$ from Eqs. (6.2.7) and (A.1.3) and cancelling we obtain:

$$\dot{\omega} = -F^T V e - V Y V e + V G \mu - H^T R^{-1} \eta$$  \hspace{1cm} (7.6.3)

We are interested in the case where there is no driving or measurement noise. Setting $\mu$ and $\eta$ to zero and using Eq. (7.6.1) we can reduce Eq. (7.6.3) to:

$$\dot{\omega} = -F^T \omega - V Y \omega$$

* They also have the property that for the same statistics $E_{\text{biased}} < E_{\text{unbiased}}$. Potter and Stern (1964) discuss these and other properties of unbiased filters.
which is an undriven differential equation in \( \omega \). From Eq. (7.6.2) we know that \( \omega \) is zero initially; since \( \omega \) is undriven \( \omega \) is always zero.

After enough information is available to completely determine the state \( V \) will be non singular. In the continuous case this will usually be immediately after the measurement process begins. When \( V \) is non-singular we can obtain the estimation error by:

\[
e = V^{-1} \omega
\]

which will be zero when there is no noise. Thus when there is no driving noise and the measurements are perfect this estimation scheme yields an estimate which coincides with the actual state as soon as there is enough information to completely determine the state.

7.7 Summary

When there is no a priori state estimate, situations can arise where part but not all of the state is observable. This chapter contains a method for estimating these dimensions which is based upon the cofilter that is developed here and in Chapter VI. No batch processing of data is necessary. The cofilter is completely recursive in the discrete case and is described by differential equations in the continuous case. The technique can be used to start a Kalman filter for which no a priori information has been provided.

The important equations are Eqs. (6.2.7), (7.2.6), (7.2.12), (7.2.14), (7.2.15) through (7.2.25), (7.3.1), (7.3.2) (7.3.4), and (7.3.5).
CHAPTER VIII

APPLICATION OF SMOOTHING TECHNIQUES TO A LINEAR TWO DIMENSIONAL PROBLEM

8.1 General Comments

The purpose of this chapter is to illustrate the application of the smoothing equations to a very simple problem. A two dimensional example has been chosen because it is the lowest dimension which retains the complexities of matrix algebra. Only the statistics are processed because the state estimate is too dependent upon the actual sequence of random numbers which form the measurements and driving force. One processing of the statistics summarizes the errors in a large number of Monte Carlo runs of the state estimation procedure. A study is made of smoother performance as compared to filter performance.

8.2 The System Description

We will consider the application of the discrete smoother covariance matrix equations to a lightly damped, constant coefficient, second order oscillator. Two examples of such a system are a mechanical mass, spring, dashpot arrangement (for higher damping ratios this would approximate an automobile suspension system) and an electrical RLC circuit.

A fundamental matrix for the state space representation of the oscillator is:

\[
F = \begin{bmatrix}
0 & 1 \\
-\omega^2 & -2\zeta\omega
\end{bmatrix}
\]

We will obtain the constant state transition matrix by integrating the equation:
\[
\frac{d}{dt} \left[ \Phi(t_1, t_0) \right] = F(t_1) \Phi(t_1, t_0)
\]

from the initial condition at time \( t_0 \):

\[
\Phi(t_0, t_0) = I
\]

to one sample period later.

A direct measurement of the first dimension of the state is assumed, thus:

\[
H_k = H = [1, 0]
\]

The integrated effect of the driving noise will be considered a random variable. This term will be a discrete scalar. We will vary the nature of the gain matrix which couples this force to the state but will keep it fixed during any one run:

\[
G_k = G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}
\]

The scalars \( G_1 \) and \( G_2 \) will take on values of one or zero.

The noise variances \( R_k \) and \( Q_k \) will be assumed scalar constants for any given run. We will explore the smoother performance as \( R \) and \( Q \) vary from run to run.
8.3 A Summary of the Results

A MAC* language computer program was written to study the application of the discrete smoother covariance matrix equations of Chapter VI to the system described in the previous section. The program was run on a Honeywell 1800 computer. The calculations were done with floating point arithmetic using a 10 decimal digit mantissa.

Due to space limitations only a few representative results are graphically presented here. The remainder of the results are summarized in the text of this section.

Figures 8.1 and 8.2 illustrate the diagonal elements of the filter and smoother covariance matrices. In this typical run the following numerical values were used:

\[
F_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad R = 10^{-4} \quad Q = 10^{-2}
\]

\[
G = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{Sample Period} = .1
\]

\[
\omega = 6 \quad \zeta = .16
\]

*MAC is an algebraic computer language developed at the M.I.T. Instrumentation Laboratory for engineering applications. It has a basic three line format but otherwise is similar to the more generally used Fortran.
Figure 8.1 1-1 Element of Covariance Matrices, $R = 10^{-4}$, $Q = 10^{-2}$
Figure 8.2 2-2 Element of Covariance Matrices, $R = 10^{-4}$, $Q = 10^{-2}$
In all figures of this chapter the logs of the diagonal elements of the forward filter covariance matrix are indicated by squares. The corresponding quantities for the smoother are identified with circles. These squares and circles are for identification purposes only - they do not indicate sample periods. These marks occur at a frequency of once every five sample periods.

Figures 8.1 and 8.2 illustrate some results that occurred in many of the runs. Note that the improvement of smoothing over filtering is noticeably more significant in dimension 2 than dimension 1. This is evidenced by the fact that the smoother variance is smaller than the filter variance by a greater amount in dimension 2 than in dimension 1. As the ratio $Q/R^*$ increases this improvement becomes larger. Figure 8.3 illustrates the results for dimension 2 with all numerical values the same as above except that $Q = 10$. In this run the separation between smoother and filter variances is larger than the previous run which had a smaller driving noise. As the ratio $Q/R$ becomes very large the separation between the filter and smoother variances becomes proportional to $Q/R$. Figure 8.4 illustrates the other extreme. In this run there was no driving noise ($Q = 0$). Note how there is virtually no separation between the smoother and filter variances.

An explanation for this trend lies in the smoothability condition presented in Chapter V. For the case of no driving noise the state is non-smoothable and the smoother result is merely the backward extrapolation of the final forward filter covariance. For this example such an extrapolation is characterized by the two variances being almost the same at each sample periods as in Fig. 8.4. As the driving noise increases, the state becomes smoothable and the curves separate. Smoothing improves the estimate of the second dimension more than the first because the second dimension is more strongly smoothable. The reason for

*In the more general case this would be $X/S$. 

136
Figure 8.5  2-2 Element of Covariance Matrices, $R = 10$, $Q = 10^{-4}$
this is that dimension 2 is controlled directly by the driving noise while dimension 1 is separated from the driving noise by an integration. This makes dimension 1 more weakly controllable by the driving noise (hence more weakly smoothable) than dimension 2. When the gain matrix was changed such that $G_1 = 1$ the first dimension became more strongly controllable by the driving noise. The difference in the improvement brought by smoothing in the two dimensions was not as noticeable. This is because the good quality direct measurement of dimension one was able to contain the effect of the driving noise.

In those cases where smoothing significantly improved the filter estimate (examples of this may be seen in Figs. 8.2 and 8.3) both the filter and smoother variances very quickly reached steady state values. In these cases we used a large $Q/R$ ratio. The problem was thus dominated by a large driving noise and, by comparison, good measurements. This caused the filter to quickly reach steady state because the driving noise reduced the confidence in prior estimates (including the initial a priori estimate) due to the uncertainty involved in propagating the state from one sample period to the next. The result was that the estimate was heavily dependent upon each new measurement and not closely related to prior estimates. This repeated pattern plus the fact that the noise statistics were stationary caused the filter to quickly obtain the statistical steady state condition.

All the remarks of the preceding paragraph were directed at the forward filter - the diagonal elements of which are displayed in the illustrations of this chapter. The same remarks are also true for the backward filter. The smoother is a weighted combination of these two filters, each of which reaches a steady state condition shortly after the beginning of its operation. Thus the smoother response should have a transient at each end with a steady state region between the transients.
Note in Fig. 8.2 and 8.3 that there are indeed two transients during smoothing. This is characteristic of all cases where smoothing improves the estimate which can be obtained by filtering. This is most easily understood in terms of the forward-backward smoother formulation. The transient in the beginning is the effect of the transient in the forward filter and the transient at the end is the effect of the transient in the backward filter.

In these same cases (those in which smoothing significantly improves the filter estimate) the steady state values obtained by the filter and smoother are independent of the initial forward filter covariance matrix. The reason the forward filter is essentially independent of this initial covariance matrix over most of the interval of interest is the same as that given for the brief transient. The problem is dominated by the driving noise and good measurements. The backward filter is by definition independent of the initial forward filter covariance matrix. The smoother must then also be virtually independent of this initial condition because it depends only upon these two filters.

As an example of a case where there is a long transient consider Fig. 8.5. In this run all parameters are the same as above except \( Q \) and \( R \). Figure 8.5 shows what happens to the estimation of dimension 2 when \( R = 10 \) and \( Q = 10^{-4} \). Note the long transient and lack of improvement during smoothing. The oscillations which are evident in Fig. 8.5 occur at a frequency of \( 2\omega \). This is characteristic of this type problem.

The preceding remarks have been directed toward the covariance matrix of state estimation errors. The same general conclusions apply to the driving force estimate and its covariance. The smoother is able to learn about the driving force by using outputs which occur after the force is applied. The forward filter, which does not have this additional information available, cannot provide an
Figure 8.4  2-2 Element of Covariance Matrices, \( R = 10^{-4} \), \( Q = 0 \)
estimate which is better than the a priori information. The smoother also cannot improve this estimate if there is no driving noise. When there is no driving noise the driving force is completely deterministic and no amount of estimation can improve this already perfect knowledge. Smoothing thus improves the estimate of those driving force dimensions which are corrupted with additive noise. This is the smoothability condition applied to the estimate of the driving force.

As a final observation in this simple example note that the smoother estimate is never worse than the forward filter estimate at the same time. This is really a consequence of the optimality of the smoothing solution. The smoothing solution could not be the optimum solution if there existed a better estimate. Thus the smoother estimate can be no worse (in the statistical sense) than the forward filter estimate. This is also a consequence of Eq. (6.2.1) which is repeated here:

$$P_T^{-1}(t) = E^{-1}(t) + P^{-1}(t) \quad (6.2.1)$$

The matrices on the right are either positive definite or positive semidefinite. The inverse of the smoother covariance matrix is the sum of these two matrices. Since $P(t)$ is at least semidefinite $P_T(t)$ must be less than $E(t)$. 

142
CHAPTER IX

APPLICATION OF SMOOTHING TECHNIQUES
TO A PARAMETER IDENTIFICATION PROBLEM

9.1 Introductory Remarks

The purpose of this chapter is to demonstrate that the smoothing techniques discussed in the previous chapters are useful even when applied to a large dimensional, nonlinear problem. Such complex problems are typical of the interesting applications the practicing engineer is likely to encounter. We will use as an example the problem of determining an unknown time varying parameter in a fourth order system. This application is representative of a large number of practical situations where the engineer is uncertain about some of the parameters which he has used in modeling a dynamic process. By properly reducing data taken during a performance test of the system he can obtain information about the uncertain parameters. If the data reduction can be done after the test is completed, filtering followed by smoothing is a possible candidate for processing the recorded measurements.

The identification problem is basically nonlinear. This is perhaps most clearly demonstrated by considering the simple first order system which is characterized by:

\[
x = ax
\]

where \(a\) is an unknown constant. If we consider \(x\) and \(a\) to be state variables we have the equations:
\dot{x} = a x

(9.1.1)

\dot{a} = 0

The product of state variables in Eq. (9.1.1) causes the problem to be nonlinear.

The technique to be employed here is one which has been found to be useful by other workers.* This is the method of augmenting the unknown parameters to the state and linearizing the resulting nonlinear equations about a reference solution which is based upon the best estimate of the augmented state. We will then apply the optimum filtering and smoothing techniques to the linear equations which describe the deviation from the reference solution.

In the linear example of the previous chapter the covariance matrix computation was entirely independent of the state estimation procedure. For this reason we were able to process only the statistics. In the nonlinear problem under consideration in this chapter such a separation is not valid. The reason for this is that the reference solution is updated using the state estimate. Since the extrapolation of the covariance matrix depends upon this reference solution the statistics are coupled to the state estimation procedure. We thus lose the ability to describe an ensemble of runs with one processing of the statistics as we did in Chapter VIII. We must instead resort to Monte Carlo simulations to study the performance of the smoothing techniques when applied to such nonlinear problems. If, however, we do not update the reference solution using the state estimate the covariance matrix computation would be uncoupled from the state estimation procedure. If the deviations from the reference solution remain small

*See for example Ho and Lee (1964) or Kopp and Orford (1963).
the statistics will closely approximate the ensemble average of the estimation errors and we could replace the Monte Carlo simulations with one processing of the covariance matrix computation.

9.2 A Description of the System

We will consider the application of the smoothing techniques presented in Section 6.10 to the identification of a time varying natural frequency which is involved in the description of a fourth order linear system. Examples of such a system include a double spring-mass arrangement with varying mass, an electrical network with adjustable components, and a flexible spacecraft with time varying bending characteristics. The spacecraft example is one of current interest so we will use it to develop and motivate the system equations.

We will include in our simplified planar dynamic model of the spacecraft only the effects of rigid body rotation and the first bending oscillation. We will assume that the vehicle is driven by a gimballed engine at its base. With these assumptions the equation which describes rigid body rotation is:

\[ T_e \left[ \ell \sigma_e + \phi_e \right] q_b - T_e \ell \dot{\delta}_e + I_R \ddot{\theta} = 0 \]  \hspace{1cm} (9.2.1)

The length \( \ell \) and the angles \( \delta_e \) and \( \theta \) are illustrated in Fig. 9.1. The thrust provided by the engine is denoted as \( T_e \), \( q_b \) is the generalized bending coordinate, \( I_R \) is the moment of inertia of the vehicle, \( \sigma_e \) is the normalized bending slope at the engine hinge point and \( \phi_e \) is the normalized bending displacement at the hinge point.

The equation which describes body bending is:
Fig. 9.1 Coordinate System for Spacecraft Example
\[ T_e \phi_e \delta_e + m \ddot{q}_b + 2m \xi \omega \dot{q}_b + m \omega^2 q_b = 0 \quad (9.2.2) \]

where \( m \) is the mass of the bending body and \( \xi \) and \( \omega \) are respectively the damping ratio and natural frequency of the bending mode.

The engine deflection, \( \delta_e \), in Eqs. (9.2.1) and (9.2.2) is the sum of a commanded value, \( \delta_c \), and a random error, \( \mu_1 \):

\[ \delta_e = \delta_c + \mu_1 \quad (9.2.3) \]

We will model the natural frequency as a random walk with a preferred direction:

\[ \dot{\omega} = a + \mu_2 \quad (9.2.4) \]

where \( a \) is an unknown constant and \( \mu_2 \) is additive white noise. This means that we expect the natural frequency to change due to changes in mass and rigidity as the flight progresses but we are not sure of its exact time variation. For this reason we approximate the time variation as roughly linear (but with unknown slope) and add the driving noise term to account for the expected deviations from linearity. In these filter and smoother applications it is often useful to make up for inadequacies in the model by adding noise terms in the appropriate places as we did here for the natural frequency. For this example the unknown constant, \( a \), would probably be positive because the natural frequency should increase as the mass decreases due to fuel consumption.
We will assume that we measure only vehicle attitude using an inertial measurement unit. The measurement is then:

\[ z = \theta + \sigma_a q_b + \eta \tag{9.2.5} \]

where \( \sigma_a \) is the normalized bending slope at the IMU location and \( \eta \) is the error (assumed white noise) involved in making the measurement.

If we define the six dimensional state:

\[ \begin{bmatrix} \theta \\ \dot{\theta} \\ q \\ \dot{q} \\ \omega \\ a \end{bmatrix} \tag{9.2.6} \]

Eqs. (9.2.1) to (9.2.4) may be written in the form

\[ \dot{x} = f(x, u) \tag{9.2.7} \]

where

\[ u = \begin{bmatrix} \delta c \\ 0 \end{bmatrix} + \mu \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{9.2.8} \]
and

\[
\begin{bmatrix}
  x_2 \\
  \frac{T_e}{I} \left\{ l(\delta_c + \mu_1) - (l\sigma_e + \phi_e) x_3 \right\} \\
  x_4 \\
  \left\{ -2\xi x_5 x_4 - x_5^2 x_3 - \frac{T_e \phi_e}{m} (\delta_c + \mu_1) \right\} \\
  x_6 + \mu_2 \\
  0
\end{bmatrix}
\]

(9.2.9)

\[
f =
\]

If \( x_R \) denotes the reference solution, the deviation from this reference is:

\[
\delta x = \dot{x} - x_R
\]

and the linear equation describing this deviation is:

\[
\frac{d}{dt} (\delta x) = F \delta x + Gu
\]

(9.2.10)
where:

\[
F = \left[ \frac{\partial f}{\partial x} \right]_{x = x_R} = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{T_e}{I} (\ell \sigma_e + 4e) & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & -x_{R3}^2 & -2\xi x_{R5} - (2\xi x_{R4} + 2x_{R3} x_{R5}) & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

(9.2.11)

and:

\[
G = \begin{bmatrix}
0 & 0 \\
\frac{T_e \ell}{I} & 0 \\
0 & 0 \\
\frac{T_e \phi_e}{m} & 0 \\
0 & 1 \\
0 & 0 \\
\end{bmatrix}
\]

(9.2.12)

Equations (9.2.7), (9.2.8) and (9.2.9) describe the nonlinear system. Equations (9.2.10), (9.2.11), and (9.2.12) characterize the linearized version of this nonlinear system.
9.3 The Simulation

A MAC language computer program was written to study the application of the discrete smoothing technique presented in Section 6.10 to the nonlinear example described in the previous section. This program was arranged as a simulation and run on a Honeywell 1800 computer.

The measurements required by the filter and smoother were generated in the program by numerically integrating Eq. (9.2.7) and using Eq. (9.2.5) to form the measurements. The measurement and driving noises were taken from random number generators with Gaussian distributions. These measurements were then processed using the filter and smoother equations to determine an estimate of the state. An evaluation of the usefulness of these techniques was then made by comparing the state estimates to the actual state which was generated in the simulation.

Except for in a few cases which we will note and discuss in the next section the reference solution was updated at each measurement time using the best available estimate of the deviation of the state from the reference solution. The forward extrapolation indicated by Eq. (6.10.1) was accomplished by numerically integrating Eq. (9.2.7) with \( \tilde{x} \) substituted for \( x \) and \( \tilde{u} \) substituted for \( u \). The backward extrapolation indicated by Eq. (6.10.11) was done by numerically integrating Eq. (9.2.7) backward with \( \varphi_{R} \) substituted for \( x \) and \( \tilde{u} \) for \( u \). All other computations were performed as outlined in Section 6.10 for the case where the reference solution is changed every sample period.

In the simulation results to be considered in the next section the following numerical values were used in all runs:

\[
\begin{align*}
T_e &= 1 \\
L &= .5 \\
I &= 1 \\
m &= 1 \\
\xi &= .5 \\
\phi_e &= -.5 \\
\sigma_e &= 1.2 \\
\sigma_A &= 0
\end{align*}
\]
Note that the choice $\sigma_A = 0$ gives us a direct measurement of only the first dimension of the state.

All runs were processed with a sample period of .5 and a maximum time of 15. The initial conditions and statistics were varied from run to run.

9.4 Summary of Results

In this section we will consider some of the results of the simulation discussed in the previous sections. Space limitations prohibit presenting all the results so we will consider representative illustrations of the important dimensions of the problem. Since we organized this example as an identification problem we will denote most of our attention to the last two dimensions of the state.

Figures 9.2 through 9.5 illustrate results obtained with the numerical values:

$$E_0 = \text{diagonal (100, 100, 100, 100, 5, 5)}$$

$$R = 10^{-6} \quad Q = 10^{-2} I$$

The random numbers actually used in the simulation were taken from distributions with the same variances as $R$ and $Q$.

Figures 9.2 and 9.3 illustrate respectively the fifth ($\omega$) and sixth ($a$) dimensions of the state. The actual state, which was generated by the simulation, represents the correct answer and is marked with diamonds. The forward filter estimate of these dimensions is identified with squares and the corresponding smoother estimate is
Figure 9.2 Comparison of Estimates and Actual State for Dimension 5
Figure 9.4 5-5 Element of Covariance Matrices
marked with circles. These marks are for identification purposes only; they occur at a frequency of once every four sample periods. Figures 9.4 and 9.5 are coded in the same way and illustrate respectively the log of the fifth and sixth diagonal elements of the filter and smoother covariance matrices.

Note in Fig. 9.5 that the actual value of the constant, $x_6$, was chosen negative. This departure from the expected positive value for $x_6$ in the spacecraft example was for clerical convenience only. The numbers used do not represent any particular spacecraft and could just as easily describe parameters in a model of an electrical circuit or mass-spring-dashpot arrangement. By choosing $x_6$ negative and by selecting the numerical integration time step properly initially we can be sure that we will never encounter instabilities in the numerical integration due to too large a time step. In this way the programming which would be necessary to detect and remedy such numerical problems was eliminated.

These figures illustrate results which were typical of those obtained in many runs. The variances illustrated in Figs. 9.4 and 9.5 indicate that the smoother estimate of dimensions five and six should be significantly better than the corresponding forward filter estimate. Figures 9.2 and 9.3 represent one member of the ensemble which is approximately* summarized by the variances of Figs. 9.4 and 9.5. In these figures the smoother does indeed give a noticeably better state estimate than the forward filter over most of the time period involved.

*Because the extrapolation of the statistics depends upon the reference solution which in turn depends upon a state estimate which is a function of the actual driving and measurement noises, different variances will result for different noise sequences. This coupling is second order, however, so the covariances approximately summarize the ensemble of state estimation errors.
Figure 9.2 also provides some insight into the choice of the word "smoothing" to describe this estimation procedure. Note how the curve representing the smoother estimate "smooths" out the oscillations in the filter estimate to provide a closer approximation to the actual state.

The smoothability condition derived in Chapter V was applicable to linear systems only. The fact that it does not in general apply to nonlinear systems is visibly displayed by Fig. 9.3. The sixth dimension of the state (the slope, a) is not controllable by the driving noise. If the smoothability condition were applicable to this system we would expect the smoother estimate of dimension six to be a backward extrapolation of the final forward filter estimate. The backward extrapolation of this constant would be a straight line. Note that the smoother estimate in Fig. 9.3 is definitely not a straight line.

In Section 5.4 it was stated that if the reference solution does not change very much when updated on the basis of the state estimates then the smoothability condition will be approximately valid. Figure 9.6 supports this argument. In this run there was less driving noise and the a priori covariance of the driving noise was reset accordingly:

\[ Q = 10^{-4} I \]

All other parameters remained the same. The resulting estimates did not significantly change the reference solution over most of the time period. Note how the smoother estimate of dimension six in this case was approximately a straight line (backward extrapolation of the terminal estimate) for most of the time period. A very slight slope in the line develops at about time = 7 (returning from time = 15) when the forward filter begins oscillating severely and changing the forward filter reference
Figure 9.6 Comparison of Estimates and Actual States for Dimension 6
solution. The smoother does not exhibit the large transient found in the filter response at time = 2 because at this time there is much more confidence in the backward filter estimate than in the forward filter estimate. Consequently the smoother depends much more at this point on the backward filter estimate than on the corresponding forward filter estimate.

In this case the results for dimension 5 also show that when the reference solution is not heavily dependent upon the state estimates the smoothability condition is approximately valid. For dimension 5 the backward extrapolation of the final forward filter estimate is a straight line with a slope given by the value of the constant dimension six. Figure 9.7 illustrates that in this case smoothing does indeed produce such a straight line estimate.

A lesson to be learned from results such as these is that in those cases where the reference solution does not change much when the estimate is incorporated there is no real need to update the reference solution. This perhaps obvious conclusion can be used to simplify the mechanization of the filter and smoother equations with a resulting saving in computation time. If the entire filter-smoother technique is iterated a few times the reference solution should become less and less sensitive to the state estimate and it will no longer be necessary to update it. If however this situation does not arise as it did not in Figs. 9.2 and 9.3, then the reference solution should be updated.

As an example of the results obtained when the reference solution is not updated using the state estimates consider Figs. 9.8 through 9.13. These plots illustrate the diagonal elements of the covariance matrices for a run with all numerical values the same as above except:

\[ Q = 10^{-6} I \]
Figure 9.7 Comparison of Estimates and Actual State for Dimension 5
Figure 9.8 1-1 Element of Covariance Matrices for Unupdated Reference Solution
Figure 9.9 2-2 Element of Covariance Matrices for Unupdated Reference Solution
Figure 9.10  3-3 Element of Covariance Matrices for Unupdated Reference Solution
Figure 9.11 4-4 Element of Covariance Matrices for Unupdated Reference Solution
Figure 9.12 5-5 Element of Covariance Matrices for Unupdated Reference Solution
Figure 9.13 6-6 Element of Covariance Matrices for Unupdated Reference Solution
The covariance matrix calculations are uncoupled from the state estimation procedure in this example because the reference solution is not updated using the estimates. Note that of the six dimensions of the state, dimension one is improved the least by smoothing. This trend was also noticeable in those cases where the reference solution was updated. The explanation for this is the same as for the smaller dimension example of the previous chapter and is twofold. First, dimension one is the most weakly controllable by the driving noise of all the dimensions; consequently smoothing should not be as effective for this dimension as for the others. Second, the measurement is of the first dimension only. This direct measurement of dimension one enables the filter to regain at the measurement times any information which was lost due to the driving noise between these times. Since the filter already does such a good job of containing the driving noise, the task the smoother is useful for, the smoother cannot give much improvement in the estimate of the first dimension.

Figures 9.8 through 9.13 illustrate variances which do not differ significantly from those obtained for this case when the reference solution is updated. As an example of this compare Fig. 9.14 to Fig. 9.12. Figure 9.14 illustrates results from a typical run with the same parameters used to generate Figs. 9.8 through 9.13; in this case, however, the reference solution was updated using the state estimate. The reason for the small difference between the results for the two cases is that the deviations from the reference solution remained small. The resulting second order differences in the covariance matrices are thus very small.

Two additional observations about Figs. 9.8 through 9.13 are worth mentioning. First, in most of these illustrations an oscillation is evident in the curves. These oscillations occur at approximately the natural frequency and decrease in frequency as the natural frequency decreases. Second, the smoother variance in Fig. 9.13 is a straight line. The reason for this is that when the reference solution is not
updated dimension six is not smoothable. The variance is thus the backward extrapolation of the final forward filter variance. Since dimension six is a constant, this backward extrapolation is a straight line.

As a final example of the results obtained for this nonlinear problem we will consider a case which should serve as a note of caution to any prospective user of these techniques. Figure 9.15 illustrates the results of the estimation of $\omega$ for a case with all numerical values as above except:

$$Q = I$$

The actual driving noises were taken from distributions with unit variance. In this case the forward filter estimate diverged from the correct answer and the smoother estimate was almost as poor. The reason for this was that the filter covariance matrix indicated a small error in the estimate when there indeed existed a large error. This reduced the Kalman gains too much with the result than new measurements were not weighted heavily enough to enable the new information to improve the estimate. This was able to happen because when the estimate was incorrect the linearization occurred around this incorrect estimate. The result was that the estimate did not improve in the nonlinear case as fast as it should according to linear theory. Since the covariance calculation is based upon the linear theory the covariance matrix became too small to soon.

As a demonstration that such problems do not render these techniques useless consider Fig. 9.16. This plot illustrates the results obtained for a run where everything (including the actual state which is scaled differently in Figs. 9.15 and 9.16) remained the same as before except the a priori covariance of the driving noise which was artificially raised to:
Figure 9.16 Comparison of Estimates and Actual State for Dimension 5
\[ Q = 100 I \]

in order to offset the effects of the nonlinearity. This increased covariance of the driving noise caused the forward filter to rely less upon extrapolated estimates and more upon new measurements than before by appropriately increasing the Kalman gains. The result was that the forward filter estimate no longer diverged from the correct value. A similar increased dependence upon new measurements was obtained in the backward filter; consequently the smoother estimate was also greatly improved. More improvement could probably have been obtained by further adjustment of critical parameters in the filter. Our purpose here, however, is not to show that we can improve the results of a particular example. The purpose is to demonstrate that in such nonlinear applications these problems can arise and that good judgement as to why they occur can lead to practical solutions for them. This we have already achieved so we will close this example with the recommendation that the user always simulate the application of these techniques to his nonlinear problem in as realistic a manner as possible before processing actual test data. In this way he will be able to determine under what circumstances these techniques work well and how to adjust them in those cases where the results are not acceptable.

In summary of the results obtained in this nonlinear study we may conclude that the smoothing technique presented in Section 6.10 is a useful one for application to state estimation in those nonlinear systems which can be piecewise approximated by linearization about a reference solution. There do exist situations in which the estimate may diverge from the actual state but these can be found by simulation and made to converge by appropriate adjustment of filter parameters. These situations arise when the linearized model does not closely approximate the errors between the actual and reference state. The Kalman gains are no longer near optimum and large estimation errors
can arise. These nonlinear effects can usually be offset by artificially increasing the appropriate noise covariances. For state type nonlinearities this would be the driving noise and for measurement nonlinearities it would be the measurement noise. Even in the linear case it is usually wise to use driving noise covariances which are a little larger than expected to offset the effect of nonlinearities in the calculation. An example of such a nonlinearity is truncation of numbers in the computer.
CHAPTER X

CONCLUSIONS AND RECOMMENDATIONS

10.1 Summary

A new method for the optimal smoothing of data is presented. The results of numerical studies show that this new recursive scheme is superior to the older methods in those cases when smoothing gives an estimate which is significantly better than that which can be obtained using optimum filtering techniques. In addition to this numerical superiority, the new smoothing scheme is more easily understood in terms of physical reasoning than the earlier methods.

A smoothability condition is derived which enables the user to determine under what circumstances smoothing yields an estimate which differs from a backward extrapolation of the final forward filter estimate. If he determines that part or all of the state is non-smoothable, the user can realize a saving in computer time by using backward extrapolation of the final forward filter state estimate to obtain the smoothed estimate of these quantities.

A recursive method for starting a Kalman filter when no a priori information is available is presented. The recursion formulas are called a cofilter and are used in the new smoother formulation. A method of performing partial state estimation when there is insufficient information to completely determine the state is developed. This technique is also based upon the cofilter.

The new smoothing technique was applied to a nonlinear parameter identification problem. The results obtained from this study, which was organized as a simulation, show that optimum smoothing is a useful technique in those nonlinear situations where linearization about a reference solution is valid.
10.2 Recommendations for Further Study

The work presented herein is for the case where the user has all the data ahead of time and desires to obtain the smoothed estimate for each point within the data interval. Another problem of more limited interest is that of obtaining the smoothed estimate at a particular point within the interval as more data is received. Rauch (1963) considered this problem. It would be interesting to extend the interpretation of the smoother as two Kalman filters to this situation and yet retain the recursive nature of the solution.

For every filtering problem there is a corresponding smoothing problem. With this in mind we can say that it would be useful to attempt to extend to smoothing any and all new techniques which are developed to solve the filtering problem. A few examples of this follow.

Deyst (1964), and Bryson and Johanson (1965) have solved the optimum linear filtering problem when the measurements are corrupted by noise which does not have a white time distribution. A useful extension of this work would be to apply it to smoothing. It should be possible in this case to express the smoother estimate as the optimum combination of two correlated Kalman filter estimates.

The optimum smoother, just like the optimum filter, is only optimum if the correct noise statistics are used. Any techniques which are developed to filter in the presence of uncertain statistics should be extended to smoothing. Johansen (Johansen (1965) and Johansen (1966)) offers one approach to this problem. Another approach is to attempt to estimate the statistics using the measurements.

Jazwinski (1966) and others have presented approximate nonlinear minimum variance filtering techniques. Any useful results which come from this and other investigations into nonlinear estimation schemes should also be extended to smoothing.
APPENDIX A

SOME USEFUL STATISTICAL PROPERTIES OF
THE FILTERED ESTIMATE

In this appendix the statistical properties of the error in the filtered estimate which were found useful in Chapter 3 are derived. The notation of that chapter is used here to provide continuity with the derivations in that chapter.

A.1 The Error in the Filtered Estimate

The equation describing the filter estimate is:

\[ \dot{\hat{x}} = FX + G\bar{u} + EH^T R^{-1}(z - H\hat{x}), \]

(A.1.1)

The actual state is described by:

\[ \dot{x} = FX + G\bar{u} + Gu. \]

(A.1.2)

Subtracting (A.1.1) from (A.1.2) and using (3.2.6) and (3.7.2) gives an expression describing the filter error:

\[ \dot{e} = (F-ES)e + Gu - EH^T R^{-1}n. \]

(A.1.3)

By comparing Eq. (A.1.3) to Eq. (3.7.13) and (3.7.14) we may deduce that the state transition matrix associated with (A.1.3) is the inverse of the transpose of that associated with (3.7.13) or:

*This equation has been derived by many workers. See for example Appendix B or Bryson and Frazier (1962).
\[ \psi^T(t_2, t_1). \]

Thus the solution to Eq. (A.1.3) is:

\[ e(t) = \psi^T(\tau, t)e(\tau) + \int_\tau^t \psi^T(s, t) \left[ G(s)\mu(s) - E(s)H^T(s)R^{-1}(s)\eta(s) \right] ds. \]  

(A.1.4)

A.2 Covariance Between the Filter Error at Different Times

Consider the autocorrelation \( e(t)e^T(\tau) \) for \( t > \tau \).

From Eq. (A.1.4) we have:

\[ e(t) = \psi^T(\tau, t)e(\tau) + \int_\tau^t \psi^T(s, t) \left[ G(s)\mu(s) - E(s)H^T(s)R^{-1}(s)\eta(s) \right] ds \]

\[ e(t)e^T(\tau) = \psi^T(\tau, t)e(\tau)e^T(\tau) + \int_\tau^t \psi^T(s, t) \left[ G(s)\mu(s)e^T(\tau) \right. \]

\[ - E(s)H^T(s)R^{-1}(s)\eta(s)e^T(\tau) \right] ds. \]  

(A.2.1)

Now \( \mu(s)e^T(\tau) \) and \( \eta(s)e^T(\tau) \) are both zero because \( s > \tau \) and the driving terms \( \mu(s) \) and \( \eta(s) \) cannot influence \( e(t) \) until \( t > s \).

Thus Eq. (A.2.1) reduces to:

\[ e(t)e^T(\tau) = \psi^T(\tau, t)E(\tau) \quad t > \tau. \]  

(A.2.2)

Had we instead elected to evaluate \( e(\tau)e^T(t) \) for \( t > \tau \) we would have obtained the result:

\[ e(\tau)e^T(t) = E(\tau)\psi(\tau, t) \quad t > \tau. \]  

(A.2.3)
A. 3 Covariance Between the Filter Error and the Measurement Noise

In this section we will consider the covariance \( \mathbf{e}(t) \mathbf{\eta}^T(\tau) \). Inasmuch as the noise on the measurement cannot influence the error until after it actually occurs we immediately deduce:

\[
\mathbf{e}(t) \mathbf{\eta}^T(\tau) = 0 \quad t \leq \tau.
\] (A. 3.1)

For \( \tau < t \) we may write, using (A.1.4):

\[
\mathbf{e}(t) \mathbf{\eta}^T(\tau) = \mathbf{\psi}^T(t_0, t) \mathbf{e}(t_0) \mathbf{\eta}^T(\tau) + \int_{t_0}^{t} \mathbf{\psi}^T(s, t) \left[ \mathbf{G}(s) \mathbf{\mu}(s) \mathbf{\eta}^T(\tau) - \mathbf{E}(s) \mathbf{H}^T(s) \mathbf{R}^{-1}(s) \mathbf{\eta}(s) \mathbf{\eta}^T(\tau) \right] ds.
\] (A. 3.2)

Now the initial error in the filter estimate is uncorrelated with the measurement noise and the noise driving the state is uncorrelated with the measurement noise. Using these facts and substituting Eq. (3.2.7) into (A. 3.2) we obtain:

\[
\mathbf{e}(t) \mathbf{\eta}^T(\tau) = -\mathbf{\psi}^T(\tau, t) \mathbf{E}(\tau) \mathbf{H}^T(\tau) \quad t > \tau.
\] (A. 3.3)

A. 4 Covariance Between the Filter Error and the Noise Driving the State

The covariance \( \mathbf{e}(t) \mathbf{\mu}^T(\tau) \) is considered in this section. Just as in the case of the measurement noise, the noise driving the state cannot influence the filter error until after the noise actually occurs. Thus:
\[
e(t)u^T(\tau) = 0 \quad t \leq \tau. \quad (A.4.1)
\]

Using Eq. (A.1.4) we may write the following equation for \( t > \tau \):

\[
e(t)u^T(\tau) = \psi^T(t_0, t)e(t_0)u^T(\tau) + \int_{t_0}^{t} \psi^T(s, t) \left[ G(s)\mu(\mu)^T(\tau) 
- \bar{E}(s)H^T(\mu)R^{-1}(\mu)\bar{\eta}(\mu)^T(\tau) \right] ds. \quad (A.4.2)
\]

Invoking Eqs. (3.2.5), (3.2.7), and the fact that the initial error in the filter is uncorrelated with the driving noise, we may reduce Eq. (A.4.2) to:

\[
e(t)u^T(\tau) = \psi^T(\tau, t)G(\tau)Q(\tau) \quad t > \tau. \quad (A.4.3)
\]
APPENDIX B
DERIVATION OF THE KALMAN FILTER
STATE ESTIMATION EQUATION

The purpose of this appendix is to demonstrate that the filter equations may be derived from the smoother problem formulation. To show this we will derive the forward filter state estimation equation from Eqs. (3.3.11) and (3.3.12). These two equations were the result of the calculus of variations approach to minimizing the cost function which characterized the smoother problem formulation.

We begin by noting as we did in Chapter III that the smoother and filter are equivalent at the terminal time, \( T \), because both operate using all the available information. If we now consider the terminal time to be a variable we have the relationship:

\[
\frac{d \hat{x}(T)}{dT} = \left[ \frac{d \hat{x}(t)}{dt} \right]_{t = T} \quad (B.1)
\]

where:

\[
\hat{x}(T) = \hat{x}_T(T)
\]

\[
E(T) = P_T(T)
\]

By restricting ourselves to the case where the terminal time, \( T \), is the current time, \( t \), we can transform the smoother equations into a set of
equations which describe the forward filter. Using Eq. (B.1) we may rewrite Eqs. (3.3.11) and (3.3.12) with \( t = T \) as:

\[
\begin{align*}
\frac{d}{dT} \begin{bmatrix} \hat{X}_T(T) \\ \lambda(T) \end{bmatrix} &= \begin{bmatrix} F(T) & Y(T) \\ S(T) & -F^T(T) \end{bmatrix} \begin{bmatrix} \hat{X}_T(T) \\ \lambda(T) \end{bmatrix} \quad + \quad \begin{bmatrix} G(T)\bar{u}(T) \\ -H^T(T)R^{-1}(T)\bar{z}(T) \end{bmatrix} \\
\end{align*}
\]

(B.2)

Using the fact that the filter and smoother estimates are identical at \( t = T \) we may write the solution to Eq. (B.2) as:

\[
\begin{bmatrix} \hat{X}(T) \\ \lambda(T) \end{bmatrix} = M(T, t_0) \begin{bmatrix} \hat{X}(t_0) \\ \lambda(t_0) \end{bmatrix} + \int_{t_0}^{T} M(T, s) \begin{bmatrix} G(s)\bar{u}(s) \\ -H^T(s)R^{-1}(s)\bar{z}(s) \end{bmatrix} \, ds
\]

(B.3)

where:

\[
\frac{d}{dT} \begin{bmatrix} M(T, t_0) \end{bmatrix} = \begin{bmatrix} F(T) & Y(T) \\ S(T) & -F^T(T) \end{bmatrix} M(T, t_0)
\]

(B.4)
For clerical convenience define:

$$m_1 = \int_{t_0}^{T} \left[ M_{11}(T, s)G(s)\mu(s) - M_{12}(T, s)H^T(s)R^{-1}(s)\mu(s) \right] ds$$  \hspace{1cm} (B. 5)

$$m_2 = \int_{t_0}^{T} \left[ M_{21}(T, s)G(s)\mu(s) - M_{22}(T, s)H^T(s)R^{-1}(s)\mu(s) \right] ds$$  \hspace{1cm} (B. 6)

then Eq. (B. 3) may be written in the form:

$$\begin{bmatrix} \hat{x}(T) \\ \hat{\lambda}(T) \end{bmatrix} = M(T, t_0) \begin{bmatrix} \hat{x}(t_0) \\ \hat{\lambda}(t_0) \end{bmatrix} + \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$$  \hspace{1cm} (B. 7)

Applying the boundary condition provided by Eq. (3. 3. 8) to $\lambda(T)$ in Eq. (B. 7) yields:

$$M_{21}(T, t_0)\hat{x}(t_0) + M_{22}(T, t_0)\hat{\lambda}(t_0) + m_2 = 0$$  \hspace{1cm} (B. 8)
We have been interpreting the final time, $T$, to be the current time, $t$. At this point we will change notation such that $T$ is replaced by the move familiar current time variable, $t$. Furthermore we will suppress the explicit time notation on all but initial conditions. When $M$ and its elements appear without time arguments this will mean $M(t, t_0)$. Using this simpler notation we may write the result of substituting the boundary condition provided by Eq. (3.3.7) into Eq. (B.8) as:

$$M_{21} \begin{bmatrix} x(t_0) + E(t_0)\lambda(t_0) \end{bmatrix} + M_{22} \lambda(t_0) + m_2 = 0$$

or

$$\lambda(t_0) = \left[ M_{21}E(t_0) + M_{22} \right]^{-1} \begin{bmatrix} M_{21}x(t_0) + m_2 \end{bmatrix}$$  \hspace{1cm} (B.9)$$

Applying Eq. (3.3.7) to the $\hat{x}(T)$ part of Eq. (B.7) leaves, upon substituting $t$ for $T$:

$$\hat{x}(t) = M_{11}x(t_0) + \left[ M_{11}E(t_0) + M_{12} \right] \lambda(t_0) + m_1$$  \hspace{1cm} (B.10)$$

Substitution of Eq. (B.9) into Eq. (B.10) yields:
\[ \hat{\Delta}(t) = M_{11} \bar{x}(t_0) - \left[ M_{11} E(t_0) + M_{12} \right] \left[ M_{21} E(t_0) + M_{22} \right]^{-1} \left[ M_{21} \bar{x}(t_0) + m_2 \right] + m_1 \]  

(B.11)

Now define:

\[ E(t) = \left[ M_{11} E(t_0) + M_{12} \right] \left[ M_{21} E(t_0) + M_{22} \right]^{-1} \]  

(B.12)

Note that this is an identity for \( t = t_0 \).

Differentiating Eq. (B.12), substituting the appropriate derivatives for \( M \) from Eq. (B.4), canceling terms, and using Eq. (B.12), yields:

\[ \dot{E} = FE + EF^T + Y - ESE \]  

(B.13)

Substitution of Eq. (B.12) into Eq. (B.11) gives:

\[ \hat{\Delta}(t) = (M_{11} - EM_{21}) \bar{x}(t_0) - Em_2 + m_1 \]  

(B.14)

Defining:
\[
X = M_{11} - EM_{21}
\]  
(B. 15)

we may write Eq. (B.14) as:

\[
\hat{X}(t) = X(t_0) - EM_2 + m_1
\]  
(B. 16)

Differentiating Eq. (B.16) we obtain:

\[
\dot{\hat{X}}(t) = \dot{X}(t_0) - \dot{EM}_2 - \dot{EM}_2 + \dot{m}_1
\]  
(B. 17)

Using Eqs. (B.15), (B.4) and (B.13) we can show that:

\[
\dot{X} = (F - ES) X
\]  
(B. 18)

An expression for \( \dot{m}_1 \) can be obtained using Eqs. (B.6), (B.4), and (B.7). The result is:

\[
\dot{m}_1 = Fm_1 + Ym_2 + GU
\]  
(B. 19)
A similar result may be obtained for \( \dot{m}_2 \) using the same equations. The result is:

\[
\dot{m}_2 = S m_1 - F^T m_2 - H^T R^{-1}z
\]  \hspace{1cm} (B.20)

Substituting Eqs. (B.13), (B.18), (B.19), and (B.20) into Eq. (B.17) and canceling terms yields:

\[
\dot{\hat{x}} = (F - E S) \left[ X(t_0) - E m_2 + m_1 \right] + E H^T R^{-1}z + G \bar{u}
\]  \hspace{1cm} (B.21)

Application of Eq. (B.16) to Eq. (B.21) yields the desired result:

\[
\dot{\hat{x}} = F \hat{x} + G \bar{u} + E H^T R^{-1} (z - H \hat{x})
\]  \hspace{1cm} (B.22)

Equation (B.22) is the forward filter state estimation formula. We could derive the corresponding covariance matrix relationship in a manner analogous to that used in Chapter III to obtain the smoother covariance matrices. In doing this we would learn that \( E \) as defined by Eq. (B.12) is this covariance matrix and that its derivative is indeed given by Eq. (B.13). This will not be done here because we have already accomplished out objective of demonstrating that the Kalman filter equations can be derived from the smoother problem formulation and because other workers have previously presented such a derivation.
LIST OF REFERENCES


Bryson, A. E., (1964), Notes from course Eng. 218, Harvard University, Spring Term.


191


BIOGRAPHY

Donald C. Fraser was born April 20, 1941 in New York City. He attended public schools in New York City; Tappan, New York; Dumont, New Jersey; and graduated from the Northern Valley Regional High School in Demarest, New Jersey in June, 1958.

Mr. Fraser entered M. I. T. in September, 1958 and received the degrees of Bachelor of Science and Master of Science in Aeronautics and Astronautics from M. I. T. in June, 1963. As an undergraduate he held various scholarships and was a member of the Honors Program of the Department of Aeronautics and Astronautics.

While a graduate student, Mr. Fraser held the Bendix and Sperry Rand Fellowships. During the summer of 1962 he was employed by the Autonetics Division of North American Aviation Corporation as a research engineer. In September, 1962 he joined the M. I. T. Instrumentation Laboratory as a research assistant under the supervision of Mr. John R. McNeil. In this capacity he worked on the analysis of "strapped down" inertial guidance systems until joining the Space Guidance Analysis Group under the general supervision of Dr. Richard H. Battin in May, 1964. In this group he worked on control system analysis for the Apollo spacecraft under the supervision of Mr. Edward M. Copps.

Mr. Fraser is married to the former Joanne Murray of Port Chester, New York. He is a Member of Tau Beta Pi, Sigma Xi, and Sigma Gamma Tau honorary fraternities.