January 01, 2011

Applied kalman filter theory

Yalcin Bulut

APPLIED KALMAN FILTER THEORY

A Dissertation Presented

by

Yalcin Bulut

to

The Department of Civil and Environmental Engineering

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the field of

Civil Engineering

Northeastern University
Boston, Massachusetts

August 2011

Abstract

The objective of this study is to examine three problems that arise in experimental mechanics where Kalman filter (KF) theory is used. The first is estimating the steady state KF gain from measurements in the absence of process and measurement noise statistics. In an off-line setting the estimation of noise covariance matrices, and the associated filter gain from measurements is theoretically feasible but lead to an ill-conditioned linear least square problem. In this work the merit of Tikhonov's regularization is examined in order to improve the poor estimates of the noise covariance matrices and steady state Kalman gain.

The second problem is on state estimation using a nominal model that represents the actual system. In this work the errors in the nominal model are approximated by fictitious noise and covariance of the fictitious noise is calculated using stored data on the premise that the norm of discrepancy between correlation functions of the measurements and their estimates from the nominal model is minimum. Additionally, the problem of state estimation using a nominal model in on-line operating conditions is addressed and feasibility of extended KF (EKF) based combined state and parameter estimation method is examined. This method takes the uncertain parameters as part of the state vector and a combined parameter and state estimation problem is solved as a nonlinear estimation using EKF.

The last problem is the issue of using the filter as a damage detector when the process and measurement noise statistics vary during the monitoring. The basic idea used to implement the filter as a detector is the fact that the innovation process is white. When the system changes due to damage the innovations are no longer white and correlation can be used to detect it. A difficulty arises, however, when the process and/or measurement noise covariance fluctuate because the filter detects these changes also and it becomes necessary to differentiate what comes from damage and what does not. In this work a modified whiteness test for innovations process is examined. The test uses correlation functions of the innovations evaluated at higher lags in order to increase the relative sensitivity of damage over noise fluctuations.

Acknowledgments

It is a pleasure to thank the many people who made this thesis possible.

First of all, I would like to thank to my parents for their life-long love and support. I thank them for enduring long periods of separation to help me in my betterment. To them I dedicate this thesis. I would also like to thank the rest of my family, my sisters and my brother for their inspiration. I would like to honor my grandfather who passed away during my studies, and ask for his mercy not being able to do the last task for him.

I would like to express my deep gratitude to the Civil and Environmental Engineering Department of Northeastern University for their generous funding throughout my graduate study.

I would like to express my sincere gratitude and appreciation to my advisor, Professor Dionisio Bernal. With his enthusiasm, his inspiration, and his great efforts to explain things clearly and simply, he helped to make mathematics fun for me. I would have been lost without him in a completely new area to me. I would also like to thank Professors Adams and Sznaier and Caracoglia for reading this dissertation and offering constructive comments.

I am indebted to my many colleagues for providing a stimulating and fun environment in which to learn and grow at the Center for Digital Signal Processing Lab at Northeastern University. I am grateful to Joan, Demetris, Jin, Necmiye, Dibo, Yiman, Harish, Anshuman, Vidyasagar, Srinivas, Osso, Bei, Parastoo, Rasoul, Yueqian, Yashar, Lang, Maytee and especially to Burak.

I would like to thank to my roommates during the years of my PhD studies; Murat, Nihal, Emrah, Omer, Serkan, Orcun, Onur and especially to Hasan for their continuous support.

Lastly, sincere thanks are extended to my other friends in Boston; Ece, Seda, Oguz, Mustafa, Anvi, Evrim, Emre, Sevket, Levent, Bilgehan, Akan, Alparslan, Volkan, Yalgin, Ahmet, Cihan and especially to Kate for helping me making my time in Boston the best it could possibly be.

Contents

Abstract **i**

1 Introduction **1**

1.1 Background and Motivation . . . 1

1.2 Objectives . . . 2

1.3 Outline . . . 4

2 Kalman Filter Theory For the User **5**

2.1 Introduction . . . 5

2.2 State Space Representations . . . 6

2.2.1 Continuous-Time State-Space Representation . . . 6

2.2.2 Discrete-Time State-Space Representation . . . 10

2.2.3 Controllability and Observability . . . 13

2.3 The Principles of Kalman Filtering . . . 14

2.3.1 The Dynamical System of Interest . . . 14

2.3.2 The State Estimator . . . 17

2.3.3 Selecting L_k . . . 22

2.3.4 Steady State Kalman Filter . . . 25

2.3.5 Innovations Form Kalman Filter . . . 25

2.3.6 Colored Process and Measurement Noise . . . 29

2.3.7 Correlated Process and Measurement Noise . . . 32

2.3.8 Biased Noise . . . 34

2.3.9 Properties of the Innovations Process . . . 35

2.3.10 Invariance Property of Kalman Gain . . . 41

2.4 Continuous Time Kalman Filter . . . 45

2.5 Extended Kalman Filter . . . 47

3 Steady State Kalman Gain Estimation **53**

3.1 Background and Motivation . . . 53

3.2 Indirect Noise Covariance Approaches . . . 59

3.2.1 The Output Correlations Approach . . . 59

3.2.2 The Innovations Correlations Approach . . . 62

3.3 Direct Kalman Gain Approaches . . . 65
3.3.1 The Output Correlations Approach . . . 65
3.3.2 The Innovations Correlations Approach . . . 68
3.4 Ill-conditioned Least Square Problem . . . 70
3.5 Numerical Experiments . . . 78
3.5.1 Experiment 1: Five-DOF Spring Mass System . . . 78
3.5.2 Experiment 2: Planar Truss Structure . . . 85
3.6 Summary . . . 89

4 Kalman Filter with Model Uncertainty 95
4.1 Background and Motivation . . . 95
4.2 Stochastic Modeling of Uncertainty . . . 98
4.2.1 Fictitious Noise Approach . . . 98
4.2.2 Equivalent Kalman Filter Approach . . . 102
4.3 Combined State and Parameter Estimation . . . 105
4.4 Numerical Experiment: Five-DOF Spring Mass System . . . 115
4.5 Summary . . . 120

5 Damage Detection using Kalman Filter 123
5.1 Background and Motivation . . . 123
5.2 Innovations Approach to Damage Detection . . . 128
5.2.1 Dependence of the Innovations Correlations on the Noise Covariances . . . 128
5.2.2 Effect of Damage On the Innovations Correlations . . . 130
5.2.3 Frequency Domain Interpretation . . . 132
5.3 A Modified Whiteness Test . . . 136
5.3.1 Hypothesis Testing . . . 137
5.3.2 The Test Statistics . . . 139
5.3.3 Modified Test Metric . . . 142
5.4 Numerical Experiment: Five-DOF Spring Mass System . . . 145
5.5 Summary . . . 150

6 Summary and Conclusions 153

A An Introduction to Random Signals and Noise 165
A.1 Random Variables . . . 165
A.2 Multivariate Random Variables . . . 168
A.3 Random Signals . . . 170

B Parametrization of Structural Matrices 172

List of Figures

3.1 Experimental PDFs of process noise covariance estimates. 59
3.2 Discrete Picard Condition 73
3.3 The generic form of the L-curve 77
3.4 Five-DOF spring mass system. 79
3.5 Change in the condition number of H matrix with respect to number of lags. 80
3.6 Discrete Picard Condition of five-DOF spring mass system 81
3.7 The L-curve for five-DOF spring mass system. 82
3.8 Histograms of Q and R estimates from 200 simulations in numerical testing using five-DOF spring mass system. 83
3.9 Five-DOF spring mass system estimated filter poles 84
3.10 Histograms of innovations covariance estimates from 200 simulations in numerical testing using five-DOF spring mass system. 84
3.11 Truss structure utilized in the numerical testing of correlations approaches. 85
3.12 Discrete Picard Condition for truss structure. 86
3.13 The L-curve for truss structure.. 87
3.14 Truss structure, estimates of filter poles. 88
3.15 Histograms of trace of state error covariance estimates for truss structure. 88

4.1 Estimate of second floor stiffness k_2 and error covariance. 114
4.2 The output correlation function of the five-DOF spring mass system. . . 117
4.3 Displacement estimate of the second mass 118
4.4 Histograms of filter cost from 200 simulations on state estimation using erroneous model 119
4.5 Spring stiffness estimates and error covariance for the five-DOF spring mass system 120

5.1 Frequency response of the transfer function from process noise to innovations. 136
5.2 PDF of ρ from healthy and a damage state. 138
5.3 Autocorrelation function of innovations process. 141

5.4 Trade-off between noise change and damage with respect to initial lag in 143
5.5 Largest eigenvalue of $(A - KC)^j$ in absolute value 145
5.6 Theoretical χ^2 CDF and PDF with 50 DOF in the numerical testing of the five-DOF spring mass system. 147
5.7 The largest eigenvalue of $(A - KC)^j$ in absolute value as the lag increases. 148
5.8 Auto-correlations of the innovations process for five-DOF sping mass system 149
5.9 Experimental χ^2 PDFs of ρ with 50 DOF from 200 simulations 150
5.10 Power of test, (P_T) at 5% Type-I error in the numerical testing in the numerical testing of five-DOF spring mass system. 150

A.1 Normal (Gaussian) distribution. 168
A.2 Uniform distribution. 169

List of Tables

2.1 Closed form discrete input to state matrices 11

4.1 The un-damped frequencies of five-DOF spring mass system. 116

5.1 Poles and zeros of the transfer functions in optimal case. 135
5.2 Poles and zeros of the transfer functions in damage case. 135
5.3 Chi square correlation test results for Type-I error probability, $\alpha = 0.05$. 142
5.4 Change in the first un-damped frequency (Hz) due to damage in five-DOF spring mass system 146

Chapter 1

Introduction

1.1 Background and Motivation

In many modern structural engineering applications such as structural health monitoring, control, model validation and damage detection, reconstructing the response of an existing structure subjected to partially unknown excitations is often required. If the structure is not instrumented, the best that can be done is to estimate the unknown loads and applying these together with the known loading to a model of the structure and obtaining the response. In this approach the accuracy of the estimations depends entirely on how accurate the model and predicted loading represent the physical system. Nevertheless, if the structure is instrumented and measurements of the response at some locations are available, the accuracy of the estimation can be improved by combining the measurements and the model in some consistent way.

The idea of response reconstruction for partially instrumented systems has been studied in the area of control and state estimation since the 1950s, and new contributions continue to appear. This work focuses on some applications of Kalman filter (KF) theory. The formulation of KF involves a blend of a model with measurements. The improvements in the predictions of the response, in comparison to the open loop, result from incorporating the measured data into the estimation process. The "open loop" is a system that does not contain any feedback terms from the outputs, as opposed to a "closed loop" system in which part of the excitation depends on the feedback from the response.

Kalman filter theory has received serious attention in many fields such as electrical engineering, robotics, navigation and economics since the 1960's. The problems and applications that originated the development of KF theory as it stands today are not problems related to structural engineering. However, the use of the KF theory may become more prominent because the types of problems that utilize it are more common than before, and there is significant merit and potential in incorporating KF theory into a variety of problems concerning existing structural systems. As new challenges arise regarding critical existing infrastructure, blending finite element models and measured data has a potential of becoming a more and more important part of modern structural engineering problems.

1.2 Objectives

The objective of this work is to approach three problems that arise in experimental mechanics where the Kalman filter is used:

The first problem consists of estimating the steady state Kalman filter gain from measurements. In the classical formulation of Kalman filter, one can calculate the filter gain given the information of a model and the covariance of unknown disturbances and measurement noise. If the noise covariance matrices are not available, it has been shown that the Kalman filter gain can be extracted from the data. Several methods have been proposed, however, all ultimately exploit the fact that the discrepancy between measured and predicted output signal, which is called innovations, is white when the filter gain is optimal. Examination of the literature shows that the problem is ill-conditioned. In this work we examine the merit of Tikhonov's regularization technique in the issue of calculating the Kalman gain from data.

The second problem is on state estimation using a nominal model that has uncertainty. In the classical Kalman filter theory, one of the key assumptions is that a priori knowledge of the system model, which represents the actual system, is known without error. In this work our objective is to examine the feasibility and merit of an approach that takes the effects of the uncertain parameters of the nominal model into account in state estimation using KF without estimating the uncertain parameters themselves. In this approach, the model errors are approximated by fictitious noise and the covariance of the fictitious noise is calculated on the premise that the norm of discrepancy between covariance functions of measurements and their estimates from the nominal model is minimum. Additionally, we aim to address the problem of state estimation using a nominal model in on-line operating conditions and examine the use of EKF-based combined state and parameter estimation method. This method takes the uncertain parameters as part of the state vector and a combined parameter and state estimation problem is solved as a nonlinear estimation using extended KF (EKF).

The last problem is related to the use of Kalman filter as a fault detector. It is well known that the innovations process of the Kalman filter is white. When the system changes due to damage the innovations are no longer white and correlations of the innovations can be used to detect damage. A difficulty arises, however, when the statistics of unknown excitations and/or measurement noise fluctuate because the filter detects these changes also and it becomes necessary to differentiate what comes from damage and what does not. In this work we investigate if the correlation functions of the innovations evaluated at higher lags can be used to increase the relative sensitivity of damage over noise fluctuations.

1.3 Outline

The chapters in the dissertation are organized as follows:

Chapter 2 presents background information that is frequently used throughout the dissertation on dynamic systems, state estimation and Kalman filter theory.

Chapter 3 is devoted to estimation of Kalman filter gain from measurements. An overview of correlation based procedures to estimate the Kalman filter gain from measurements is presented and the use of Tikhonov's regularization technique is examined to solve the poorly conditioned least square problem that involves the parameters of the noise covariance matrices and Kalman filter gain.

Chapter 4 addresses the issue of the KF with model uncertainty. The model errors are approximated by fictitious noise and the actual system is approximated with an equivalent Kalman filter model. Extended KF based combined state and parameter estimation method is examined.

Chapter 5 examines the use of Kalman filter as a fault detector. The fault detector is formulated based on the whiteness property of the Kalman filter innovations process. The use of the fault detector is examined to detect the changes in the system under changing noise covariances.

Chapter 6 presents a brief summary and conclusions.

Chapter 2

Kalman Filter Theory For the User

2.1 Introduction

The state of a dynamical system are variables that provide a complete representation of the internal condition or status of the system at a given instant of time. When the state is known, the evolution of the system can be predicted if the excitations are known. Another way to say the same is that the state consists of variables that prescribe the initial condition. When a model of the structure is available, its dynamic behavior can be estimated for a given input by solving the equations of motion. However if the structure is subjected to unknown disturbances and is partially instrumented, the response at the unmeasured degrees of freedom is obtained using state estimation.

The basic idea in state estimation theory is to obtain approximations of the state of a system which is not directly observable by using information from a model and from any available measurements. There are three estimation problems, namely:

- Estimate the current state, x_n given all the data including y_n; this is called *filtering.*

- Estimate some past value of x_k, $k < n$ given all the data including y_n; this is called *smoothing*.
- Estimate some future value of x_k, $k > n$ given all the data including y_n; this is called *prediction*.

State estimation can be viewed as a blend of information from the analytical model of the system and measurements to make predictions about present and future states of the system. Uncertainty in estimating the state of a dynamical system arises from three sources: (1) Stochastic disturbances, (2) Discrepancies between the real system and the model used to represent it (model uncertainty), (3) Unknown initial conditions.

The theory of state estimation originated with least squares method essentially established by the early 1800s with the work of Gauss [1]. The mathematical framework of modern theory of state estimation originated with the work of Wiener in the late 1940s, [2]. The field began to mature in the 1960's and 1970's after a milestone contribution was offered by R. E. Kalman in 1960 [3], which is very well-known as the *Kalman filter* (KF). The KF is a recursive data processing algorithm, which gives the optimal state estimate of the systems that are subjected stationary stochastic disturbances with known covariances.

2.2 State Space Representations

2.2.1 Continuous-Time State-Space Representation

A linear finite dimensional structural system subjected to a time varying excitation $u(t)$, can be described by the following ordinary linear differential equation

$$M\ddot{q}(t) + C_\xi \dot{q}(t) + Kq(t) = b_2 u(t) \tag{2.2.1}$$

where the dots represent differentiation with respect to time, $q \,\epsilon\, \mathbb{R}^{\eta x1}$ is the displacement vector at the degrees of freedom, M, C_ξ and K are the mass, damping and stiffness matrices respectively and $b_2 \,\epsilon\, \mathbb{R}^{nxr}$ is a vector describing the spatial distribution of the excitation $u(t) \,\epsilon\, \mathbb{R}^{rx1}$ and η and r are the number of degrees of freedom and input excitations respectively. Defining

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \overset{def}{=} \begin{bmatrix} q(t) \\ \dot{q}(t) \end{bmatrix} \tag{2.2.2}$$

and substituting Eq.2.2.2 into Eq.2.2.1 one gets

$$\begin{bmatrix} I & 0 \\ 0 & M \end{bmatrix} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -K & -C_\xi \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ b_2 \end{bmatrix} u(t) \tag{2.2.3}$$

In this dissertation, we limit examination to cases where the matrix M is non-singular. In this case, Eq.2.2.3 can be written as

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C_\xi \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ M^{-1}b_2 \end{bmatrix} u(t) \tag{2.2.4}$$

Taking $x(t) = \begin{bmatrix} x_1^T(t) & x_2^T(t) \end{bmatrix}^T \epsilon\, \mathbb{R}^{nx1}$ is the state vector and integer $n = 2\eta$ is the order of the system, it yields

$$\dot{x}(t) = A_c x(t) + B_c u(t) \tag{2.2.5}$$

where $A_c \, \epsilon \, \mathbb{R}^{nxn}$ is known as the continuous-time system matrix and $B_c \, \epsilon \, \mathbb{R}^{nxn}$ as the continuous-time state-to-input matrix. Assuming that the available measurements are linear combinations of the state and a direct transmission term one can write:

$$y(t) = C_c x(t) + D_c u(t) \tag{2.2.6}$$

where $C_c \, \epsilon \, \mathbb{R}^{mxn}$ is the state to output matrix and integer m is the number of measurements. The structure and entries of the matrix C_c depends on the specific quantity being measured. For displacement or velocity measurement;

$$C_{dis} = \begin{bmatrix} c_2 & 0 \end{bmatrix} \tag{2.2.7}$$

$$C_{vel} = \begin{bmatrix} 0 & c_2 \end{bmatrix} \tag{2.2.8}$$

$$\tag{2.2.9}$$

where c_2 is a matrix with each row having a one at the column corresponding to the degree of freedom being measured and zeros elsewhere. For acceleration measurements

$$C_{acc} = \begin{bmatrix} -M^{-1}K & -M^{-1}C_\xi \end{bmatrix} \tag{2.2.10}$$

The matrix $D_c \, \epsilon \, \mathbb{R}^{mxr}$ is constructed by taking rows of $M^{-1}b_2$ corresponding to the degree of freedom being load applied. D_c is zero, unless acceleration is measured at a collocated coordinate. The expressions given in Eqs.2.2.5 and 2.2.6 constitute the continuous-time state-space description of a linear time invariant system. The solution to Eq.2.2.5 at any time t is given by the well known convolution integral

$$x(t) = e^{A_c t}x(0) + \int_0^t e^{A_c t} B_c u(\tau) d\tau + D_c u(t) \tag{2.2.11}$$

where $x(0)$ is the initial state.

Transfer Matrix in Continuous-time Systems

The transfer matrix is a mathematical representation of the linear mapping between inputs and outputs of a linear time-invariant system. The transfer matrix of the continuous-time state-space system in Eq.2.2.5 can be obtained by taking a Laplace transform, as follows

$$sx(s) - x_0 = A_c x(s) + B_c u(s) \tag{2.2.12}$$

Solving for the state, one gets

$$x(s) = (I.s - A_c)^{-1} x_0 + (I.s - A_c)^{-1} B_c u(s) \tag{2.2.13}$$

Taking a Laplace transform of the output equation Eq.2.2.6 gives

$$y(s) = C_c x(s) + D_c u(s) \tag{2.2.14}$$

Combining Eq.2.2.13 and Eq.2.2.14, one obtains

$$y(s) = C_c(I.s - A_c)^{-1} x_0 + \left(C_c(I.s - A_c)^{-1} B_c + D_c\right) u(s) \tag{2.2.15}$$

The transfer function matrix is defined for zero initial condition (or after steady state is realized) as

$$H(s) = \frac{y(s)}{u(s)} = C(I.s - A_c)^{-1} B_c + D_c \tag{2.2.16}$$

2.2.2 Discrete-Time State-Space Representation

The experimental data is obtained in digital form in practice, so a discrete time solution of the Eq.2.2.11 is required. Defining a time step Δt, and assuming that the state at time $t = k\Delta t$ is known, the step $k+1$ is given by,

$$x_{k+1} = e^{A_c t} x_k + \int_{k\Delta t}^{(k+1)\Delta t} e^{A_c((k+1)\Delta t - \tau)} B_c u(\tau) d\tau \tag{2.2.17}$$

By defining $A = e^{A_c \Delta t}$, the Eq.2.2.17 can be written as

$$x_{k+1} = A x_k + \int_{k\Delta t}^{(k+1)\Delta t} e^{A_c((k+1)\Delta t - \tau)} B_c u(\tau) d\tau \tag{2.2.18}$$

The integration in Eq.2.2.18 depends on the inter-sample behavior of the input $u(t)$; this is often unknown in practice and an assumption needs to be made. A finite dimensional input is one that can be expressed in the inter-sample as [4]

$$u(\tau) = f_0(\tau) u_k + f_1(\tau) u_{k+1} \tag{2.2.19}$$

where f_0 and f_1 are signed-valued functions with the constraints of $f_0(0) = 1$, $f_0(\Delta t) = 0$ and $f_1(0) = 0$, $f_1(\Delta t) = 1$. These constrains are imposed in order to match the values of the u_k and u_{k+1}, which are known. Substituting Eq.2.2.19 into the integral in Eq.2.2.18 one can get a generic form of discrete-time state space equation

$$z_{k+1} = A z_k + B u_k \tag{2.2.20}$$

$$x_k = z_k + B_1 u_{k+1} \tag{2.2.21}$$

$$y_k = C z_k + D u_k \tag{2.2.22}$$

Table 2.1: Closed form discrete input to state matrices

Inter-sample Assumption	B_0	B_1
Zero Order Hold	$(A-I)A_c^{-1}B_c$	0
First Order Hold	$(A-I)A_c^{-1}B_c - B_1$	$(A - A_c\Delta t - I)A_c^{-2}B_c/\Delta t$

where

$$B = B_0 + AB_1 \tag{2.2.23}$$

$$C = C_c \tag{2.2.24}$$

$$D = C_cB_1 + D_c \tag{2.2.25}$$

The B_0 and B_1 depend on the inter-sample assumption and obtained from the following integrals

$$B_0 = \int_{k\Delta t}^{(k+1)\Delta t} e^{A_c((k+1)\Delta t-\tau)}B_cf_o(\tau)d\tau \tag{2.2.26}$$

$$B_1 = \int_{k\Delta t}^{(k+1)\Delta t} e^{A_c((k+1)\Delta t-\tau)}B_cf_1(\tau)d\tau \tag{2.2.27}$$

The most common assumptions for the inter-sample behavior are zero order hold (ZOH) and first order hold (FOH). The input inter-sample function for ZOH and FOH in the form of Eq.2.2.19 are $(f_0 = 1,\ f_1 = 0)$ and $(f_0 = 1 - \tau/\Delta t,\ f_1 = \tau/\Delta t)$, respectively. The closed form solutions of B_0 and B_1 for ZOH and FOH are presented in Table.2.1.

The reader is referred to [4] for other discrete to continuous transfer relationships such as half-step forward shift of zero-order-hold and band limited hold where the input is assumed to be sampled as modulated train of Dirac impulses.

Transfer Matrix in Discrete-time Systems

The transfer matrix in discrete-time systems is derived by using the z-transform, which converts a discrete time signal, into a complex frequency-domain representation. The z-transform, $X(z)$ of a sequence $x(n)$ is defined as

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) z^{-n} \tag{2.2.28}$$

where z is a complex variable. The connection between the z-transform and the discrete Fourier transform can be demonstrated by taking the complex variable z in Eq.2.2.28 as $z = re^{i\omega}$, which gives

$$X(re^{i\omega}) = \sum_{n=-\infty}^{\infty} x(n) r^{-n} e^{-i\omega n} \tag{2.2.29}$$

Eq.2.2.29 shows that the z-transform of $x(n)$ is equal to the discrete Fourier transform of $x(n)r^{-n}$. Therefore, for $|z| = 1$ (on the unit circle), the z-transform is identical to the discrete Fourier transform. The shift property of the z-transform is used in the derivation of the transfer matrix, namely

$$Z\left[y(n+n_0)\right] = z^{n_0} Y(z) \tag{2.2.30}$$

The transfer matrix is derived by taking the z-transform of the discrete time state space equations given in Eqs.2.2.20-2.2.22 as follows

$$zx(z) = Ax(z) + Bu(z) \tag{2.2.31}$$

$$y(z) = Cx(z) + Du(z) \tag{2.2.32}$$

Combining Eq.2.2.31 and Eq.2.2.32, one obtains the following relationship between the z transforms of the input and output

$$y(z) = C(I.z - A)^{-1}Bx_0 + \big(C(I.z - A)^{-1}B + D\big)\,u(z) \tag{2.2.33}$$

The transfer function matrix in discrete-time system is defined for zero initial condition (or after steady state is realized) as

$$H(z) = \frac{y(z)}{u(z)} = C(I.z - A)^{-1}B + D \tag{2.2.34}$$

2.2.3 Controllability and Observability

The concepts of controllability and observability are fundamental to state estimation theory, which were introduced by Kalman [5]. The following definitions of these concepts are adapted from the texts that cover the material [6].

A continuous-time system is *controllable* if for any initial state $x(0)$ and any final time $t > 0$ there exists a control force that transfers the state to any desired value at time t. A discrete-time system is controllable if for any initial state x_0 and some final time step k there exists a control force that transfers the state to any desired value at time k. A simple test of controllability for linear systems in the both the continuous-time and discrete time cases involves using the controllability matrix, namely

$$S = \begin{bmatrix} B & AB & A^2B & \cdots & A^{n-1}B \end{bmatrix} \tag{2.2.35}$$

where n is order of the state. The system is controllable if and only if rank of S is equal to the n.

A continuous-time system is *observable* if for any initial state $x(0)$ and any final time

$T > 0$ the initial state $x(0)$ can be uniquely determined by information of the input $u(t)$ and output $y(t)$ for all up to time $t \epsilon [0, T]$. A discrete-time system is observable if for any initial state x_0 and some final time step k the initial state x_0 can be uniquely determined by knowledge of the input u_j and output y_j for all $j \epsilon [0, k]$. A simple test of observability for linear systems in the both the continuous-time and discrete time cases involves using the observability matrix, namely

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{2.2.36}$$

The system is observable if and only if rank of O is equal to the n.

The concepts of *stabilizability* and *detectability* are also closely related to the controllability, observability and modes of the system. A system is *stabilizable* if its unstable modes are controllable. A system is *detectable* if its unstable modes are observable.

2.3 The Principles of Kalman Filtering

2.3.1 The Dynamical System of Interest

Consider a time invariant linear system where the model is known without uncertainty and it is subjected to deterministic input $u(t)$, unmeasured disturbances $w(t)$ and the available measurements $y(t)$ that are linearly related to the state vector $x(t)$.

Assuming the input in the system has the following description in sampled time:

$$x_{k+1} = Ax_k + Bu_k + Gw_k \tag{2.3.1}$$

$$y_k = Cx_k + v_k \tag{2.3.2}$$

where $A \,\epsilon\, \mathbb{R}^{nxn}$, $B \,\epsilon\, \mathbb{R}^{nxr}$ and $C \,\epsilon\, \mathbb{R}^{mxn}$ are the transition, input to state, and state to output matrices, $y_k \,\epsilon\, \mathbb{R}^{mx1}$ is the measurement vector and $x_k \,\epsilon\, \mathbb{R}^{nx1}$ is the state. The sequence $w_k \,\epsilon\, \mathbb{R}^{rx1}$ is the disturbance known as the process noise and $v_k \,\epsilon\, \mathbb{R}^{mx1}$ is the measurement noise. $G \,\epsilon\, \mathbb{R}^{nxs}$ is process noise to state matrix. In the treatment here, it is assumed that these are Gaussian stationary white noise sequences with zero mean and known covariance matrices, namely

$$E(w_k) = 0 \tag{2.3.3}$$

$$E(v_k) = 0 \tag{2.3.4}$$

and

$$E(w_k w_j^T) = Q\delta_{kj} \tag{2.3.5}$$

$$E(v_k v_j^T) = R\delta_{kj} \tag{2.3.6}$$

$$E(w_k v_j^T) = 0 \tag{2.3.7}$$

where δ_{kj} denotes the Kronecker delta function; that is, $\delta_{kj} = 1$ if $k = j$, and $\delta_{kj} = 0$ if $k \neq j$. $E(\cdot)$ denotes expectation. Q and R are covariance matrices of the process and measurement noise, respectively. We assume also that the state is uncorrelated with the process and measurement noise, namely

$$E(x_k v_k^T) = 0 \tag{2.3.8}$$

$$E(x_k w_k^T) = 0 \tag{2.3.9}$$

We note that the (A, C) pair is detectable, i.e. there is no unstable and unobservable eigenvalue in the system. Eq.2.3.7 expresses the stochastic independence of noises w_k and v_k. A formulation taking into account correlation between process noise and measurement noise is presented in Section 2.3.7. Since the model is linear and w_k and v_k are zero mean Gaussian white noise signals, the state x_k and the output y_k are also Gaussian signals and are therefore entirely characterized by their first and second moments as follows:

Mean state vector: Taking expectation of the Eq.2.3.1, one gets

$$\bar{x}_{k+1} = E(x_{k+1}) = A\bar{x}_k + Bu_k \tag{2.3.10}$$

State Covariance Matrix: By definition

$$\Sigma_k = E[(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T] \tag{2.3.11}$$

We can use Eq.2.3.1 and Eq.2.3.10 to obtain

$$x_k - \bar{x}_k = A(x_{k-1} - \bar{x}_{k-1}) + Gw_{k-1} \tag{2.3.12}$$

Post-multiplying Eq.2.3.12 with its transpose and taking expectations gives

$$\Sigma = A\Sigma A^T + GQG^T \tag{2.3.13}$$

which is a discrete Lyapunov equation. The steady-state covariance can be calculated from Eq.2.3.13. The conditions for uniqueness of solution of the Eq.2.3.13 briefly are; A is stable (must have all the eigenvalues inside the unit circle), and $GQG^T \geq 0$.

2.3.2 The State Estimator

We will drive the KF from the perspective of a general state estimator. Consider a linear state estimator of the form of

$$\hat{x}_{k+1} = H_k\hat{x}_k + Z_k u_k + L_k y_k \tag{2.3.14}$$

where $\hat{x}_k$ is the estimate of x_k and L_k is the observer gain. The objective is to select the matrices H_k, Z_k and L_k to maximize the accuracy of the estimate of the state. The error in the estimated state is

$$\varepsilon_k = x_k - \hat{x}_k \tag{2.3.15}$$

Substituting Eqs.2.3.1 and 2.3.14 in Eq.2.3.17, one obtains the error in the estimated state at station $k+1$ as follows

$$\varepsilon_{k+1} = Ax_k + Bu_k + Gw_k - H_k\hat{x}_k - Z_k u_k - L_k y_k \tag{2.3.16}$$

Substituting Eq.2.3.2 into Eq.2.3.16, it gives

$$\varepsilon_{k+1} = (A - L_kC)x_k + (B - Z_k)u_k + Gw_k - H_k\hat{x}_k - L_k v_k \tag{2.3.17}$$

We now express the true state as the estimate plus the error and get

$$\varepsilon_{k+1} = (A - L_k C)(\hat{x}_k + \varepsilon_k) + (B - Z_k)u_k + Gw_k - H_k\hat{x}_k - L_k v_k \tag{2.3.18}$$

which reduces to

$$\varepsilon_{k+1} = (A - L_k C)\varepsilon_k + (B - Z_k)u_k + H_k\hat{x}_k + Gw_k - L_k v_k \tag{2.3.19}$$

Taking the expected value of Eq.2.3.19 one has

$$E(\varepsilon_{k+1}) = (A - L_k C)E(\varepsilon_k) + (B - Z_k)E(u_k) + (A - L_k C - H_k)E(\hat{x}) \tag{2.3.20}$$

where the last two terms cancel due to assumptions that the process and the measurement noise are zero mean vectors. Now, if the estimation error is to be zero mean then the second and third terms on the right hand side of Eq.2.3.20 must vanish because the expectations of the state and the input are not necessarily zero. Specifically, we require that

$$Z_k = B \tag{2.3.21}$$

$$H_k = A - L_k C \tag{2.3.22}$$

Substituting Eqs.2.3.21 and 2.3.22 into Eq.2.3.19 gives

$$\varepsilon_{k+1} = (A - L_k C)\varepsilon_k + Gw_k - L_k v_k \tag{2.3.23}$$

which shows that the error in the estimated state at step $k + 1$ depends on the error at step k and on the process and measurement noise at step k also. As noises w_k and v_k are Gaussian and as the system is linear, one can state that ε_k is a Gaussian random signal. We note that, since w_k and v_k are assumed white, the error at step k is not correlated with the process and measurement noise at step k, namely

$$E(\varepsilon_k w_k) = 0 \tag{2.3.24}$$

$$E(\varepsilon_k v_k) = 0 \tag{2.3.25}$$

Taking S_k and H_k as indicated by Eqs.2.3.21 and 2.3.22, Eq.2.3.20 becomes;

$$E(\varepsilon_{k+1}) = (A - L_k C)E(\varepsilon_k) \tag{2.3.26}$$

Inspection of Eq.2.3.26 shows that the expected value of the error at step $k+1$ depends on the expected value at step k, so, if the expectation of the estimate is zero at the start ($k = 0$) then the expected value of the error will be zero throughout. Moreover even if the expected value of ε_0 is not zero the expected value will decay

$$lim_{k\to\infty} E(\varepsilon_{k+1}) = 0 \tag{2.3.27}$$

if the eigenvalues of the matrix $(A - L_k C)$ are inside the unit circle i.e. the complex plane. Taking Z_k and H_k as indicated by Eqs.2.3.21and 2.3.22, the observer model in Eq.2.3.14 becomes

$$\hat{x}_{k+1} = (A - L_k C)\hat{x}_k + Bu_k + L_k y_k \tag{2.3.28}$$

which can be re-organized as

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + L_k(y_k - C\hat{x}_k) \tag{2.3.29}$$

The state estimator in Eq.2.3.29 ensures that the estimation is unbiased whatever the matrices A, B, C of the system and the gain L_k provided that $(A - L_k C)$ is stable. A crucial result given in [7] shows that if pair(A, C) is observable then the eigenvalues of $(A - L_k C)$ can be arbitrarily placed by a suitable choice of L_k. The pair (A, C) is observable if the observability matrix

$$O = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{2.3.30}$$

is full rank. The proof can be found in many references e.g.[6]. We present an eigenvalue assignment algorithm for state observer design in the following.

Observer Design with Eigenvalue Assignment

A state observer is designed by calculating an observer gain matrix, L_k, with assigning eigenvalues of $(A - L_k C)$ in the predetermined locations. Several eigenvalue assignment approaches are presented in the literature. A polynomial approach to eigenvalue assignment commonly used in observer design, in which a symbolic characteristic polynomial of observer, $(A - L_k C)$, is formed and observer gain parameters are calculated to achieve the desired eigenvalues as roots of the characteristic polynomial, [8]. The polynomial approach is generally applicable to single-input single-output systems.

A general treatment of the eigenvalue assignment problem for multi-input multi-output systems using time invariant observer gain $(L_k = L)$ is presented in the following, initially proposed by Moore, [9]. We begin with noting that spectral decomposition of a matrix and and its transpose is same, namely,

$$(A - LC) = \Phi \Lambda \Phi^{-1} \tag{2.3.31}$$

$$(A^T - C^T L^T) = \Phi \Lambda \Phi^{-1} \tag{2.3.32}$$

where Λ is a diagonal matrix formed from the eigenvalues of $(A - LC)$, and the columns of Φ are the corresponding eigenvectors of $(A - LC)$. The classical eigenvalue problem is defined as the solution of the following equation

$$(A^T - C^T L^T)\phi_j - \lambda_j \phi_j = 0 \qquad j = 1, 2, \cdots, n \tag{2.3.33}$$

where ϕ_j is the corresponding eigenvector to the eigenvalue, λ_j. One can organize Eq.2.3.33 as follows

$$T_j q_j = 0 \tag{2.3.34}$$

where

$$T_j = \begin{bmatrix} (A^T - I\lambda_j) & -C^T \end{bmatrix} \tag{2.3.35}$$

and

$$q_j = \begin{bmatrix} \phi_j \\ L\phi_j \end{bmatrix} \tag{2.3.36}$$

After the location of the eigenvalue, λ_j has been decided, T_j can be calculated from Eq.2.3.35 and an arbitrary vector for q_j can be chosen from null space of T_j. One can collect all the q_j's in matrix $\mathbb{Q}$ and partition it as follows,

$$\mathbb{Q} = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = \begin{bmatrix} \mathbb{Q}_U \\ \mathbb{Q}_L \end{bmatrix} \tag{2.3.37}$$

where $\mathbb{Q}_U \,\epsilon\, \mathbb{C}^{nxn}$ and $\mathbb{Q}_L \,\epsilon\, \mathbb{C}^{mxn}$. Noting from Eq.2.3.36 that,

$$\mathbb{Q}_U = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_n \end{bmatrix} \tag{2.3.38}$$

and

$$\mathbb{Q}_L = L^T \mathbb{Q}_U \tag{2.3.39}$$

An expression for L^T is written from Eq.2.3.39 as follows,

$$L^T = \mathbb{Q}_L \mathbb{Q}_U^{-1} \tag{2.3.40}$$

It's important to note that this method allows us to place only m eigenvalues in the same location. This is due to the fact that $C \,\epsilon\, \mathbb{R}^{mxn}$ and $Null(T_j)$ has dimension of m and therefore one can only obtain m independent q_j vectors for same eigenvalue. The reader is referred to [10] and [11] for two other eigenvalue assignment approaches.

2.3.3 Selecting L_k

We recall that $\varepsilon_k = x_k - \hat{x}_k$ is a multivariate, centered (unbiased), Gaussian random signal. The Gaussian feature of this centered signal allows one to state that if the trace covariance of the estimation error ($P_k = E(\varepsilon_k \varepsilon_k^T)$) is minimized then, $\hat{x}_k$ is the best estimate of x_k. Therefore we seek L_k minimizing

$$J_k = trace(E(\varepsilon_k \varepsilon_k^T)) \tag{2.3.41}$$

Post-multiplying Eq.2.3.23 with transpose of itself and taking expectations on both sides, we find that the covariance of the error at step $k+1$ is

$$P_{k+1} = E(((A - L_k C)\varepsilon_{k+1} + Gw_k - L_k v_k)((A - L_k C)\varepsilon_{k+1} + Gw_k - L_k v_k)^T) \quad (2.3.42)$$

by introducing Eqs.2.3.5, 2.3.6, 2.3.24 and 2.3.25, Eq.2.3.42 reduces to

$$P_{k+1} = (A - L_k C)P_k(A - L_k C)^T + GQG^T + L_k R L_k^T \quad (2.3.43)$$

To minimize trace of (P_{k+1}), we have

$$\frac{\partial(trace(P_{k+1}))}{\partial L_k} = 0 \quad (2.3.44)$$

The derivative of the trace of the product of two matrices is given by

$$\frac{\partial(trace(XY))}{\partial X} = Y^T \quad (2.3.45)$$

and for a triple product we have

$$\frac{\partial(trace(XYX^T))}{\partial X} = XY^T + XY \quad (2.3.46)$$

Expanding the first term in Eq.2.3.43 and taking into account the fact that P_k is symmetric, from Eq.2.3.44 after some simple algebra one gets

$$K_k = AP_k C^T(CP_k C^T + R)^{-1} \quad (2.3.47)$$

where we denoted the solution of Eq.2.3.44 as $L_k = K_k$ that is the Kalman gain matrix of the minimal error variance state observer namely Kalman filter. We update the observer model in Eq.2.3.29 using the Kalman gain matrix K_k as observer gain, namely

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + K_k(y_k - C\hat{x}_k) \tag{2.3.48}$$

Substituting Eq.2.3.47 into 2.3.43 leads to following version of the error covariance recurrence

$$P_{k+1} = AP_kA^T - AP_kC^T(CP_kC^T + R)^{-1}CP_kA^T + GQG^T \tag{2.3.49}$$

Equations 2.3.47, 2.3.48 and 2.3.49 constitute the discrete-time Kalman filter. The Kalman filtering estimation of the state is as follows:

1. At $k = 0$ the covariance of the initial state (P_0) is (assumed) known so Eq.2.3.47 can be used to compute the Kalman gain at $k = 0$.
2. Use Eq.2.3.48 to compute the estimate of the state at $k + 1$ (in the first step the expected value of the initial state ($\hat{x}_0$) is used).
3. Use Eq.2.3.49 to update the covariance of the error.
4. Use Eq.2.3.48 to update the Kalman gain.
5. Go back to step 2.

Inspection of the Eqs.2.3.47 and 2.3.49 shows that the computation of K_k and P_k does not depend on the measurements but depends only on the system (A, C, G) and noise covariance matrices (Q and R). That means that the K_k and P_k can be calculated off-line before the filter is actually implemented. Then when operating the system in real time, only the state estimation in Eq.2.3.48 need to be implemented.

2.3.4 Steady State Kalman Filter

The underlying system of interest and the process and measurement-noise covariances in Eqs.2.3.1-2.3.2 are time-invariant (in our limited scope). Therefore once the transient response due to error in the initial state estimate $\hat{x}_0$ and covariance of the error in the initial state estimate P_0 is finished, the state estimation error ε_k becomes a stationary random signal and the covariance of the error P_k reaches a steady state value. When P_k converges to a steady state value then $P_k = P_{k+1}$ for large k. We will denote this steady state value as P, which means that from Eq.2.3.49 we can write

$$P = APA^T - APC^T(CPC^T + R)^{-1}CPA^T + GQG^T \tag{2.3.50}$$

which is called a discrete algebraic Riccati equation. The reader will find methods to compute the solution of a Riccati equation in [12, 13]. The conditions for uniqueness of solution of the Eq.2.3.50 briefly are; A is stable (must have all the eigenvalues inside the unit circle), the pair A, C is observable, the pair A, GQG^T is controllable, $R > 0$ and $GQG^T \geq 0$. From a practical point of view, such solvers are available in most software like Matlab or Scilab. Once we have P, we can substitute it to Eq.2.3.47 in order obtain the steady-state Kalman gain K, that is

$$K = APC^T(CPC^T + R)^{-1} \tag{2.3.51}$$

The steady state Kalman filter often performs nearly as well as the time-varying Kalman filter provided that Q is stationary.

2.3.5 Innovations Form Kalman Filter

The filter derived in the previous section is called one step predicting form Kalman filter, which is based on the assumption that the best estimate of the state vector at time

k is a function of the output measurements up to $k-1$ but not of the measurements at k. In this section we drive the innovations form Kalman filter in which the measurements at time k are used to improve the estimate of the state. We start with postulating the updated estimate of the state in two steps as follows

$$\hat{x}_k^- = A\hat{x}_{k-1}^+ + Bu_{k-1} \tag{2.3.52}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-) \tag{2.3.53}$$

where is $\hat{x}_k^+$ the estimate after the information from the measurement at time k is taken into consideration and $\hat{x}_k^-$ is the estimate before. In first step presented in Eq.2.3.52, the open loop prediction of the state is calculated, which is called *prediction step* of the Kalman filter. In the second step presented in Eq.2.3.53, the open loop prediction is corrected (updated) with the difference between the measurement y_k and the prediction of the measurement $\hat{y}_k = C\hat{x}_k^-$ through the Kalman gain K_k, which is called *update step*. We define the error in the state before and after the update as

$$\varepsilon_k^- = x_k - \hat{x}_k^- \tag{2.3.54}$$

$$\varepsilon_k^+ = x_k - \hat{x}_k^+ \tag{2.3.55}$$

We define the covariance of the estimation error before and after the update as

$$P_k^- = E((x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T) \tag{2.3.56}$$

$$P_k^+ = E((x_k - \hat{x}_k^+)(x_k - \hat{x}_k^+)^T) \tag{2.3.57}$$

We follow a similar approach presented in the previous section to derive the Kalman filter gain as a function of the covariance prior to the update P_k^- and after the update P_k^+. The estimation error before the update is obtained by substituting Eqs.2.3.1 and 2.3.52 into Eq.2.3.54 as follows

$$\varepsilon_k^- = Ax_{k-1} + Bu_{k-1} + Gw_{k-1} - A\hat{x}_{k-1}^+ + Bu_{k-1} \tag{2.3.58}$$

which reduces to

$$\varepsilon_k^- = A\varepsilon_{k-1}^+ + Gw_{k-1} \tag{2.3.59}$$

Post-multiplying Eq.2.3.59 with its transpose and taking expectations on both sides, we find that the covariance of the estimation error before the update is given by

$$P_k^- = AP_{k-1}^+ A^T + GQG^T \tag{2.3.60}$$

Substituting Eq.2.3.2 into Eq.2.3.53 gives

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(Cx_k + v_k - C\hat{x}_k^-) \tag{2.3.61}$$

Substituting Eqs.2.3.54 and 2.3.55 into Eq.2.3.61 gives

$$x_k - \varepsilon_k^+ = x_k - \varepsilon_k^- + K_k(Cx_k + v_k - C\hat{x}_k^-) \tag{2.3.62}$$

and it can be organized as

$$\varepsilon_k^+ = \varepsilon_k^- - K_k(C\varepsilon_k^- + v_k) \tag{2.3.63}$$

Post-multiplying Eq.2.3.63 with its transpose and taking expectations on both sides, we find that the covariance of the estimation error after the update is given by

$$P_k^+ = E(\varepsilon_k^+ \varepsilon_k^{+T}) = E((\varepsilon_k^- - K_k(C\varepsilon_k^- + v_k))(\varepsilon_k^- - K_k(C\varepsilon_k^- + v_k)^T) \tag{2.3.64}$$

by introducing Eqs.2.3.56 and Eq.2.3.6, Eq.2.3.64 reduces to

$$P_k^+ = P_k^- - P_k^- C^T K_k + K_k R K_k^T - K_k C P_k^- + K_k C P_k^- C^T K_k^T \tag{2.3.65}$$

We seek K_k that minimizes the trace of P_k^+. Therefore, we take the derivative of the trace of Eq.2.3.65 and equate it to zero, which gives

$$K_k = P_k^- C^T (C P_k^- C^T + R)^{-1} \tag{2.3.66}$$

Substituting Eq.2.3.66 into 2.3.65, we can find that the estimation error covariance after the update simplifies to

$$P_k^+ = (I - K_k C) P_k^- \tag{2.3.67}$$

Equations 2.3.52, 2.3.53, 2.3.60, 2.3.66 and 2.3.67 constitute the innovations form Kalman filter framework and estimation of the state is as follows

1. At $k = 0$ the error covariance of the initial state (P_0^-) is (assumed) known so Eq.2.3.47 can be used to compute the Kalman gain at $k = 0$.
2. Use Eq.2.3.52 to compute the priori estimate of the state at k.
3. Use Eq.2.3.60 to calculate the covariance of the error before the update at k.
4. Use Eq.2.3.66 to update the Kalman gain.
5. Use Eq.2.3.53 to compute updated estimate of the state at k (in the first step the expected value of the initial state $\hat{x}_0^- = E(\hat{x}_0)$ is used).
6. Use Eq.2.3.67 to calculate the covariance of the error after the update at k.
7. Go back to step 2.

The steady state state error covariance in the innovations form Kalman filter is the solution of the Riccati equation

$$P = APA^T - APC^T(CPC + R)^{-1}CPA^T + GQG^T \tag{2.3.68}$$

and the steady state Kalman gain, K is

$$K = PC^T(CPC^T + R)^{-1} \tag{2.3.69}$$

2.3.6 Colored Process and Measurement Noise

Derivation of the Kalman filter in previous sections assumed that the process and measurement noise were both white. In this section, we illustrate how to deal with colored process and measurement noise. We assume that spectrum of the colored noise is the spectrum of the response of a linear system subjected to white noise and that this spectrum is known. Therefore, in the time domain, we assume that a state-space stochastic representation for the colored process and measurement noise are given as follows

$$w_{k+1} = \Phi w_k + \Psi \xi_k \tag{2.3.70}$$

$$v_{k+1} = \Delta v_k + \Upsilon \rho_k \tag{2.3.71}$$

where ξ_k and ρ_k are zero-mean white noise signals with covariance matrices

$$E(\xi_k \xi_k^T) = Q \tag{2.3.72}$$

$$E(\rho_k \rho_k^T) = R \tag{2.3.73}$$

$$E(\xi_k \rho_k^T) = 0 \tag{2.3.74}$$

We assume that ξ_k and ρ_k are uncorrelated with w_k and v_k, respectively. The system matrices Φ, Ψ, Δ Υ and noise covariance matrices Q and R are assumed to be known. Post-multiplying Eq.2.3.70 with w_k^T and taking expectations on both sides, we find that

$$E(w_{k+1} w_k^T) = \Phi E(w_k w_k^T) \tag{2.3.75}$$

$$E(v_{k+1} v_k^T) = \Delta E(v_k v_k^T) \tag{2.3.76}$$

The key for the colored noise case is to comprise a new state space model by combining Eqs.2.3.1, 2.3.70 and 2.3.71 as follows

$$z_{k+1} = \bar{A}z_k + \bar{B}u_k + \bar{G}\bar{w}_k \tag{2.3.77}$$

$$y_k = \bar{C}z_k \tag{2.3.78}$$

where

$$\bar{A} = \begin{bmatrix} A & G & 0 \\ 0 & \Phi & 0 \\ 0 & 0 & \Delta \end{bmatrix} \tag{2.3.79}$$

$$\bar{B} = \begin{bmatrix} B & 0 & 0 \end{bmatrix}^T \tag{2.3.80}$$

$$\bar{G} = \begin{bmatrix} 0 & 0 \\ \Psi & 0 \\ 0 & \Upsilon \end{bmatrix} \tag{2.3.81}$$

$$\bar{C} = \begin{bmatrix} C & 0 & I \end{bmatrix} \tag{2.3.82}$$

and

$$z_k = \begin{bmatrix} x_k & w_k & v_k \end{bmatrix}^T \tag{2.3.83}$$

$$\bar{w}_k = \begin{bmatrix} \xi_k \\ \rho_k \end{bmatrix} \tag{2.3.84}$$

This is an augmented state space model with a new state z_k, new system matrices $\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{G}$, and a new process noise vector $\bar{w}_k$ whose covariance is given as follows

$$\bar{Q} = \begin{bmatrix} E(\xi_k \xi_k^T) & 0 \\ 0 & E(\rho_k \rho_k^T) \end{bmatrix} = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \tag{2.3.85}$$

We note that, the measurement noise in the augmented state space model in Eq.2.3.77 and 2.3.78 is zero. The augmented model satisfies all the assumptions of the system given in Eqs.2.3.1 and 2.3.2, therefore, the Kalman filter can be applied as presented in the previous sections.

2.3.7 Correlated Process and Measurement Noise

Our derivation of the Kalman filter in previous sections assumed that the process and measurement noise were uncorrelated. In this section, we will present recurrent equations of the Kalman filter when the w_k and v_k are mutually correlated. We start with recalling noise covariance matrices as follows

$$E\left(\begin{bmatrix} w_k \\ v_k \end{bmatrix} \begin{bmatrix} w_k^T & v_k^T \end{bmatrix} \right) = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta_{kj} \tag{2.3.86}$$

where S is the cross-covariance between process and measurement noise. We consider one step predicting form Kalman filter first and follow from the state error recurrence in Eq.2.3.23. We re-derive the state error covariance recurrence by introducing Eqs.2.3.24, 2.3.25 and 2.3.86 into Eq.2.3.42 as follows

$$P_{k+1} = (A - K_k C) P_k (A - K_k C)^T + GQG^T + K_k R L_k^T - GSK_k^T - K_k S G^T \tag{2.3.87}$$

Kalman filter gain K_k that minimizes the trace of P_k can be obtained by taking the derivative of the trace of Eq.2.3.87 and equating it to zero, which gives

$$K_k = (AP_kC^T + GS)(CPC^T + R)^{-1} \tag{2.3.88}$$

Substituting Eq.2.3.88 into Eq,2.3.87 leads to following version of the error covariance recurrence

$$P_{k+1} = AP_kA^T - (AP_kC^T + GS)(CP_kC^T + R)^{-1}(APC^T + GS)^T + GQG^T \tag{2.3.89}$$

The steady state state error covariance in the one step prediction form Kalman filter is the solution of the Riccati equation

$$P = APA^T - (APC^T + GS)(CPC^T + R)^{-1}(APC^T + GS)^T + GQG^T \tag{2.3.90}$$

and the steady state Kalman gain, K is

$$K = (APC^T + GS)(CPC^T + R)^{-1} \tag{2.3.91}$$

The innovation form steady state Kalman filter when w_k and v_k are mutually correlated is

$$\hat{x}_k^- = A\hat{x}_{k-1}^+ + Bu_{k-1} + G\hat{z}_{k-1} \tag{2.3.92}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K^x(y_k - C\hat{x}_k^-) \tag{2.3.93}$$

$$\hat{z}_k = K^z(y_k - C\hat{x}_k^-) \tag{2.3.94}$$

where $\hat{z}_k$ is an auxiliary variable. Eq. 2.3.92 is the prediction step of the filter and Eqs.2.3.93 and 2.3.94 constitute update step. The filter gains K^x and K^z are calculated from

$$K^x = PC^T(CPC^T + R)^{-1} \quad (2.3.95)$$

$$K^z = S(CPC^T + R)^{-1} \quad (2.3.96)$$

where P, the steady state covariance of the state error, is the solution of the Riccati equation in Eq.2.3.90.

2.3.8 Biased Noise

The mean of a random signal, which is also called a bias, is considered to be deterministic. As seen in Eqs.2.3.3-2.3.4, derivation of the Kalman filter in previous sections assumed that the process and measurement noise were unbiased (centered). In this section, we present how to deal with biased process and measurement noise and we assume that $E(w_k)$ and $E(v_k)$ are known. The key is adding the terms $GE(w_k) - GE(w_k)$ and $E(v_k) - E(v_k)$ to the right hand side of the Eqs.2.3.1 and 2.3.2 respectively and re-organizing them as follows

$$x_{k+1} = Ax_k + \begin{bmatrix} B & G \end{bmatrix} \begin{bmatrix} u_k \\ E(w_k) \end{bmatrix} + G\bar{w}_k \quad (2.3.97)$$

$$y_k = Cx_k + E(v_k) + \bar{v}_k \quad (2.3.98)$$

where $\bar{w}_k$ and $\bar{v}_k$ are new centered process noise and measurement noise, namely

$$\bar{w}_k = w_k - E(w_k) \tag{2.3.99}$$

$$\bar{v}_k = v_k - E(v_k) \tag{2.3.100}$$

The state space model in Eqs. 2.3.97-2.3.98 satisfies all the assumptions of the system given in Eqs.2.3.1 and 2.3.2, therefore, the Kalman filter can be applied as presented in the previous sections.

2.3.9 Properties of the Innovations Process

The difference between measured and estimated output is called the innovation, namely

$$e_k = y_k - C\hat{x}_k \tag{2.3.101}$$

This is the part of the measurement that contains new information about the state, hence it's called innovations and they are uncorrelated with the previous measurements, namely

$$E(e_k y_{k-j}^T) = 0, \qquad j \geqq 1 \tag{2.3.102}$$

Substituting Eq.2.3.2 into Eq.2.3.101 and using the state error definition in Eq.2.3.15, we rewrite the innovation sequence as follows

$$e_k = C\varepsilon_k + v_k \tag{2.3.103}$$

Post-multiplying Eq.2.3.103 with its transpose and taking expectations on both sides, we find that the covariance of the innovations in steady state is given by

$$E(e_k e_k^T) = F = CPC^T + R \tag{2.3.104}$$

The properties of the sequence of e_k was first discussed by Kailath [14]. The most important characteristic of the sequence is that during optimal filtering, the innovation sequence will be white, namely

$$E(e_k e_{k+j}^T) = F, \qquad j = 0 \tag{2.3.105}$$

$$E(e_k e_{k+j}^T) = 0, \qquad j \neq 0 \tag{2.3.106}$$

In fact, Kailath derived the Kalman filter as a whitening filter which whitens measurements and generates a white innovations process; hence it extracts the maximum possible amount of information from the measurements. Consider one step predicting form steady Kalman filter model in Eq.2.3.48 is re-organized as follows.

$$\hat{x}_{k+1} = (A - KC)\hat{x}_k + Bu_k + Ky_k \tag{2.3.107}$$

$$e_k = -C\hat{x}_k + y_k \tag{2.3.108}$$

In this form, the measurement y_k is the input to the filter and the innovations process e_k is the output of the filter. The filter works as “whitening filter” for the output that the white noise innovations process is generated. The Kalman filter innovations process is a Gaussian white noise random signal which has the characteristic properties in the following:

1. The innovations process is white.
2. The innovations process a is stationary signal with a constant covariance of F.
3. The innovations process is zero mean.

The covariance of innovations process, F, has significant importance for validation of the optimality of the Kalman gain, K. The consistency between theoretical covariance of innovations process, F, and its experimental estimate demonstrates that the filter gain is optimal.

Auto-correlation Functions of Innovations

In this subsection, we demonstrate that the Kalman filter gain makes the innovations process uncorrelated. We suppose that an arbitrary filter is used to generate innovations process, which is correlated. Here we use auto-correlations functions of the correlated innovations process and show that when the arbitrary filter gain equal to the Kalman filter gain, the correlations functions are equal to zero, namely innovations are uncorrelated.

The derivations of auto-correlation functions the innovations generated from an arbitrary filter are described in the following. These derivations yield explicit expressions that relate the system and noise covariance matrices to the sample auto-correlation functions of innovations. The derivation of auto-correlation functions of innovations from an arbitrary filter is a result initially presented by Mehra, [15] and we re-derive it for mutually correlated process and measurement noise. We consider the stochastic dynamical system presented in Eqs.2.3.1-2.3.2 and one step predicting form Kalman filter model in Eq.2.3.48 with an arbitrary stable gain K_0, whose state estimates we denote as $\overline{x}_k$ and the innovations process as $\overline{e}_k = y_k - C\overline{x}_k$. The arbitrary stable filter gain K_0, which can be obtained by placing the eigenvalues of $A - K_0C$ inside the unit circle. Another option to calculate K_0 might be using classical Kalman filter formulations for some Q, R and S from whatever a priori knowledge there may be. The correlation function of the innovations from the presented filter model is,

$$\mathcal{L}_j \overset{def}{=} E(\overline{e}_k \overline{e}_{k-j}) \tag{2.3.109}$$

We recall the estimation error definition,

$$\overline{\varepsilon}_k \overset{def}{=} x_k - \overline{x}_k \tag{2.3.110}$$

and the innovation sequence recurrence from Eq.2.3.103 as function of estimation error and measurement noise as follows

$$\overline{e}_k = C\overline{\varepsilon}_k + v_k \tag{2.3.111}$$

Post-multiplying Eq.2.3.111 with its transpose and taking expectations on both sides and introducing the definition in Eq.2.3.109, we write the auto-correlation function of the innovations as follows

$$\mathcal{L}_j = CE(\overline{\varepsilon}_k \overline{\varepsilon}_{k-j}^T)C^T + CE(\overline{\varepsilon}_k v_{k-j}^T) \tag{2.3.112}$$

To obtain the terms $E(\overline{\varepsilon}_k \overline{\varepsilon}_{k-j}^T)$ and $E(\overline{\varepsilon}_k v_{k-j}^T)$ in Eq.2.3.112, we recall the state error recurrence from Eq.2.3.25 and introduce to the arbitrary gain K_0 into it

$$\overline{\varepsilon}_k = (A - K_0C)\overline{\varepsilon}_{k-1} - K_0 v_{k-1} + Gw_{k-1} \tag{2.3.113}$$

An equation for the state error covariance, $\bar{P}$ follows from Eq.2.3.113 and the assumption of stationarity. Post-multiplying Eq.2.3.113 by its transpose and taking expectations one gets

$$\bar{P} = \bar{A}\bar{P}\bar{A} + K_0 R K_0^T + GQG^T - K_0 S G^T - GS^T K_0^T \tag{2.3.114}$$

where

$$\bar{A} = A - K_0 C \tag{2.3.115}$$

To get $E(\bar{\varepsilon}_k \bar{\varepsilon}_{k-j}^T)$ we follow the recurrence in Eq.2.3.113 and carrying it j steps back one finds that

$$\bar{\varepsilon}_k = \bar{A}^j \bar{\varepsilon}_{k-j} - \sum_{t=1}^{j} \bar{A}^{j-1} K_0 v_{k-t} + \sum_{t=1}^{j} \bar{A}^{j-1} B w_{k-t} \tag{2.3.116}$$

Post-multiplying Eq.2.3.116 by $\bar{\varepsilon}_{k-j}^T$ and taking expectation

$$E(\bar{\varepsilon}_k \bar{\varepsilon}_{k-j}^T) = \bar{A}^j \bar{P} \tag{2.3.117}$$

where $\bar{P}$ is the steady state error covariance which is the solution of the Riccati equation given in Eq.2.3.114. Post-multiplying Eq.2.3.116 by v_{k-j}^T and taking expectation

$$E(\bar{\varepsilon}_k v_{k-j}^T) = \bar{A}^{j-1} GS - \bar{A}^{j-1} K_0 R \tag{2.3.118}$$

Substituting Eqs.2.3.117 and 2.3.118 into 2.3.112 one finally finds

$$\mathcal{L}_j = C\bar{A}^j \bar{P} C^T + C\bar{A}^{j-1} GS - C\bar{A}^{j-1} K_0 R \qquad j > 0 \tag{2.3.119}$$

Post-multiplying Eq.2.3.111 by its transpose and taking expectations gives zero lag correlation, namely

$$\mathcal{L}_j = C\bar{P}C^T + R \qquad j = 0 \tag{2.3.120}$$

Eqs.2.3.119 and 2.3.120 constitute the correlations of the innovations from an arbitrary gain, K_0. Inspection of 2.3.120 show that the correlations decrease with lag if the eigenvalues of the matrix $\bar{A}$ are inside the unit circle. As previously noted, Kalman filter generates uncorrelated innovations and this property can be shown by equating Eq.2.3.119 to zero and solving the gain that leads to zero covariance at all lags (other than zero), which gives

$$K_0 = (APC^T + GS)(CPC^T + R)^{-1} \tag{2.3.121}$$

which, of course, is the steady-state Kalman gain as given in Eq.2.3.91.

Another version of auto-correlation functions of innovations from an arbitrary filter can be derived using error definition between optimal and suboptimal estimates, namely, [16]

$$\tilde{\varepsilon}_k \stackrel{def}{=} \hat{x}_k - \overline{x}_k \tag{2.3.122}$$

where $\hat{x}_k$ is the state estimate from optimal filter gain, K and $\overline{x}_k$ is the estimate from suboptimal filter gain K_0. Following similar steps as shown previously, the functions can be obtained as

$$\mathcal{L}_j = C\tilde{P}C^T + F \qquad j = 0 \tag{2.3.123}$$

$$\mathcal{L}_j = C\bar{A}^j\tilde{P}C^T + \bar{A}^{j-1}KF - \bar{A}^{j-1}K_0F \qquad j > 0 \tag{2.3.124}$$

where F is the covariance of innovations from optimal gain K. $\tilde{P}$ is the state error covariance in line with the state error definition given in Eq.2.3.122, namely

$$\tilde{P} = E\left[(\hat{x}_k - \overline{x}_k)(\hat{x}_k - \overline{x}_k)^T\right] \tag{2.3.125}$$

which is the solution of following Lyapunov equation,

$$\tilde{P} = \bar{A}\tilde{P}\bar{A} + (K - K_0)F(K - K_0)^T \tag{2.3.126}$$

2.3.10 Invariance Property of Kalman Gain

We provide some insights on uniqueness of the Kalman gain, K and covariance of innovation process F with respect to noise covariance matrices, which is initially introduced by Son and Anderson, [17]. Consider the output correlation function of the stochastic system in Eq2.3.1-2.3.2 presented system is given

$$\Lambda_j \overset{def}{=} E(y_k y_{k-j}^T) \tag{2.3.127}$$

We suppose that there is no deterministic input acting on the system, namely $u_k = 0$. Substituting Eq.2.3.2 into Eq.2.3.127 one gets

$$\Lambda_j = CE(x_k x_{k-j}^T)C^T + E(x_k v_{k-j}^T)C^T \tag{2.3.128}$$

From Eq.2.3.1 one can show that

$$x_k = A^j x_{k-j} + A^{j-1}Bw_{k-j} + A^{j-2}Bw_{k-j+1} + ... + Bw_{k-1} \tag{2.3.129}$$

Post-multiplying x_k^T and taking expectations on both sides of Eq.2.3.129 gives

$$E(x_k x_{k-j}^T) = A^j\Sigma \tag{2.3.130}$$

where Σ is the state covariance in steady state, namely

$$\Sigma \overset{def}{=} E(x_k x_k^T) \tag{2.3.131}$$

One gets a Lyapunov equation for Σ by post-multiplying Eq.**??** with transpose of itself and taking expectations on both sides which gives

$$\Sigma = A\Sigma A^T + GQG^T \tag{2.3.132}$$

Post-multiplying Eq.2.3.129 with v_k^T and taking expectations on both sides gives

$$E(x_k v_{k-j}^T) = A^{j-1}GS \tag{2.3.133}$$

Substituting Eqs.2.3.131 and 2.3.133 into Eq.2.3.128 gives the auto-correlation functions of output as follows

$$\Lambda_j = C\Sigma C^T + R \qquad j = 0 \tag{2.3.134}$$

$$\Lambda_j = CA^j\Sigma C^T + CA^{j-1}GS \qquad j > 0 \tag{2.3.135}$$

Defining,

$$M \overset{def}{=} E(x_k y_{k-1}) \tag{2.3.136}$$

substituting Eqs.2.3.1-2.3.2 into Eq.2.3.136 and imposing the assumptions in Eqs.2.3.8-2.3.9 one gets

$$M = A\Sigma C^T + GS \tag{2.3.137}$$

The function of output correlations in Eq.2.3.135 can be reorganized as and using the definition of Eq.2.3.137 it yields,

$$\Lambda_j = CA^{j-1}M \qquad j > 0 \tag{2.3.138}$$

The main observation from Eq.2.3.138 is that the output correlations are function of A, C and M and although it involves the Q and S implicitly, different noise covariance matrices can lead to same output correlations. To illustrate, consider a stochastic state space model, in which the system matrices are given as follows,

$$A = \begin{bmatrix} 0.1 & 0.1 \\ 0 & 0.2 \end{bmatrix},\ G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},\ C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Assuming process and measurement noise are uncorrelated, $S = 0$, let us define two scenarios for noise statistics as follows,

$$Q_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad R_1 = 1$$
$$Q_2 = \begin{bmatrix} 1.0103 & 0.0206 \\ 0.0206 & 0.0110 \end{bmatrix} \quad R_2 = 1$$

For these two scenarios of noise statistics, one can obtain the state error covariance, P, Kalman gain, K and covariance of innovations process, F using Eqs.2.3.50, 2.3.51 and 2.3.104, respectively, which are

$$P_1 = \begin{bmatrix} 1.0157 & 0.0210 \\ 0.0210 & 1.0417 \end{bmatrix} \quad K_1 = \begin{bmatrix} 0.0514 \\ 0.0021 \end{bmatrix} \quad F_1 = 2.0157$$

$$P_2 = \begin{bmatrix} 1.0157 & 0.0210 \\ 0.0210 & 0.0114 \end{bmatrix} \quad K_2 = \begin{bmatrix} 0.5014 \\ 0.0021 \end{bmatrix} \quad F_2 = 2.0157$$

As seen, different Q matrices lead the same K and F even though the state error covariance matrices, P_1 and P_2 are different. In fact there are many Q for the given stochastic model that have same K, and there is one common property of these noise covariance matrices that is they give different realization of output with same output statistics. Consequently, it is not the Q but the K that is uniquely related to the output statistics.

There is one particular Q and R duplet among the all possible Q and R solution set that has an important relation with the K. Consider the innovation form of the Kalman filter model given in Eq.2.3.48, which is

$$\hat{x}_{k+1} = A\hat{x}_k + Ke_k \tag{2.3.139}$$

$$y_k = C\hat{x}_k + e_k \tag{2.3.140}$$

where e_k is the innovation sequence as defined in Eq.2.3.101. In this form, the Ke_k can be considered as an input to the filter and e_k is a direct noise added to the measurements. One can easily realize the relationship between system and filter models given in Eqs.2.3.1-2.3.2 and Eqs.2.3.139-2.3.140, respectively. The covariance of the input (Ke_k) and innovations process (e_k) in the filter model can be calculated from $E(Ke_ke_k^TK^T)$ and $E(e_ke_k^T)$, respectively. Moreover, these covariance matrices lead to same output statistics as the others, therefore, they also are in the set of all possible Q and R. However, it is important to note that in this case process and measurement noise are mutually correlated and cross-covariance, S, can be calculated from $E(Ke_ke_k^T)$. The noise covariance matrices from this analogy are calculated as

$$Q = \begin{bmatrix} 0.0053 & 0.0002 \\ 0.0002 & 0.000008 \end{bmatrix} \; R = 2.0157 \quad S = \begin{bmatrix} 0.1037 \\ 0.0042 \end{bmatrix}$$

which lead to a Kalman gain that is equal to K_1 and innovations covariance that is equal to F_1.

The invariance property of K and F with respect to noise covariance matrices constitutes the fundamental idea of direct estimation of K and F from measurements in the absence of noise covariance matrices, which is explored in Chapter 3.

2.4 Continuous Time Kalman Filter

Kalman and Bucy presented continuous-time version of the Kalman filter [18] one year after Kalman's work on the optimal filtering. For this reason, the continuous-time filter is sometimes called the Kalman-Bucy filter. The Kalman filter applications are implemented in digital computers, therefore, the continuous time Kalman filter has been used more for theory than practice. Consider a linear system in which the state $x(t)$ and measurements $y(t)$ satisfy

$$\dot{x}(t) = A_c x(t) + B_c u(t) + G_c w(t) \tag{2.4.1}$$

$$y(t) = C_c x(t) + v(t) \tag{2.4.2}$$

with notations defined in Section 2.3, subscript c denotes continuous-time. $\dot{x}(t)$ denotes the derivative of the state $x(t)$. We assume that process noise $w(t)$ and measurement noise $v(t)$ are uncorrelated Gaussian stationary white noise with zero mean, namely

$$E(w(t)) = 0 \tag{2.4.3}$$

$$E(v(t)) = 0 \tag{2.4.4}$$

and

$$E(w(t)w(\tau)^T) = Q_c\delta(t-\tau) \tag{2.4.5}$$

$$E(v(t)v(\tau)^T) = R_c\delta(t-\tau) \tag{2.4.6}$$

$\delta(t-\tau)$ is the delta dirac function, which has a value of ∞ at $t = \tau$, a value of 0 everywhere else. We note that, discrete-time white noise with covariance Q in a system with a sample period of $\triangle t$, is equivalent to continuous-time white noise with covariance $Q_c = Q/\triangle t$, [12]. The continuous-time Kalman filter has the form:

$$\dot{\hat{x}}(t) = A_c\hat{x}(t) + B_c u(t) + K_c(t)(y(t) - C_c\hat{x}(t)) \tag{2.4.7}$$

where the Kalman gain $K_c(t)$ is

$$K_c(t) = P_c(t)C_c^T R_c^{-1} \tag{2.4.8}$$

and the state error covariance matrix $P_c(t)$ satisfies

$$\dot{P}_c(t) = A_c P_c(t) + P_c(t) A_c^T - P_c(t) C_c^T R_c^{-1} C_c P_c(t) + G_c Q_c G_c^T \qquad (2.4.9)$$

which is called a differential algebraic Riccati equation. By letting $t \to \infty$ such that $\dot{P}_c(t) = 0$, a steady state solution for $P_c(t)$, which we denote as P_c, is obtained from

$$0 = A_c P_c + P_c A_c^T - P_c C_c^T R_c^{-1} C_c P_c + G_c Q_c G_c^T \qquad (2.4.10)$$

The conditions for uniqueness of solution of the equation Eq.2.4.10 are; A_c is stable (must have all the eigenvalues on the left half complex plane), the pair $(A_c,\, C_c)$ is observable, the pair $(A_c,\, G_c Q_c G_c^T)$ is controllable, $R_c > 0$ and $G_c Q_c G_c^T \geq 0$.

The expressions given in Eqs.2.4.7, 2.4.8 and 2.4.9 constitute the continuous-time Kalman filter. The distinction between the prediction and update steps of discrete-time Kalman filtering does not exist in continuous time and the covariance of the innovation process ($e(t) = y(t) - C_c \hat{x}(t)$) is equal to the covariance of measurement noise R_c, namely

$$E(e(t)e(\tau)^T) = R_c \delta(t - \tau) \qquad (2.4.11)$$

Integrating Eq.2.4.9 and 2.4.7 numerically, the solution is advanced one time step to obtain $\hat{x}(t+1)$. $P_c(t+1)$.

2.5 Extended Kalman Filter

The extended Kalman filter (EKF) is an extension of the Kalman filter to the case of nonlinear dynamical systems. Consider a nonlinear system that it is subjected to deterministic input $u(t)$, unmeasured zero mean white Gaussian disturbances $w(t)$ with a covariance Q_c described by

$$\dot{x}(t) = f(x,\, u,\, w,\, t) \tag{2.5.1}$$

where $f \,\epsilon\, \mathbb{R}^{nxn}$ is an arbitrary vector valued function. Measurements are assumed to be linearly related to the state vector $x(t)$ and available at discrete time steps and contaminated by a realization of white noise v_k with a covariance R, namely

$$y_k = Cx_k + v_k \tag{2.5.2}$$

In the treatment here, we assumed that unmeasured stochastic disturbance and measurement noise are mutually uncorrelated. Given the nonlinear equations of motion and measurement data, the EKF is used to calculate the minimum variance estimate of $x(t)$. The main idea of the extended Kalman filter is the linearization of the nonlinear equations of motion using Taylor series expansion around the Kalman filter estimate, and calculation of the Kalman filter estimate based on the linearized system. The recursive algorithm for the EKF is in essence the same as in linear Kalman filter, and consists of two distinct steps: Prediction and Update.

Prediction Step:

In the prediction step, the state is propagated using a linear approximation of nonlinear system. Consider expanding the vector valued function f in Eq.2.5.1 in a Taylor series about a nominal state x_0, namely

$$\dot{x}(t) = f(x_0,\ u_0,\ w_0) + \left.\frac{\partial f}{\partial x}\right|_{x_0} (x(t) - x_0) + \left.\frac{\partial f}{\partial u}\right|_{u_0} (u(t) - u_0) +$$
$$+ \left.\frac{\partial f}{\partial w}\right|_{w_0} (w(t) - w_0) + H.O.T. \tag{2.5.3}$$

We suppose that the mean state at time t, $\hat{x}(t)$ is known, and we want to propagate the state $x(t)$ at nominal state $x_0 = \hat{x}(t)$, $u_0(t) = u(t)$ and $w_0(t) = 0$. Introducing these into Eq.2.5.4 and ignoring high order terms (H.O.T.) gives

$$\dot{x}(t) = f(\hat{x}(t),\ u(t)) + \left.\frac{\partial f}{\partial x}\right|_{\hat{x}} (x(t) - \hat{x}(t)) + \left.\frac{\partial f}{\partial w}\right|_{w_0} w(t) \tag{2.5.4}$$

Taking the expectation of both sides gives

$$\dot{\hat{x}}(t) = \hat{f}(\hat{x}(t),\ u(t)) \tag{2.5.5}$$

Thus a first order Taylor series approximation of the evolution equations of the state and the state estimate are obtained. A differential equation for the estimation error covariance matrix

$$P(t) = E((x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T) \tag{2.5.6}$$

is derived by using Eqs.2.5.1 and 2.5.5

$$\dot{P}(t) = E(x(t)f^T) - \hat{x}(t)\hat{f}^T + E(fx(t)^T) - \hat{x}(t)\hat{f}^T \tag{2.5.7}$$

Although the Eq.2.5.7 is exact, an approximation is needed to obtain the terms $E(x(t)f^T)$ and $E(fx(t)^T)$ since they cannot be computed in general [19]. Substituting the expansion of the state evolution in Eq.2.5.4 and carrying out the indicated expectation operations gives

$$\dot{P}(t) = \Delta(t)P(t) + P(t)\Delta(t)^T + G_c Q G_c^T \tag{2.5.8}$$

where $\Delta(t)$ and G_c are given by

$$\Delta(t) = \left.\frac{\partial f(x,\, u,\, w,\, t)}{\partial x}\right|_{x(t)=\hat{x}(t)} \tag{2.5.9}$$

$$G_c = \left.\frac{\partial f(x,\, u,\, w,\, t)}{\partial w}\right|_{w(t)=0} \tag{2.5.10}$$

$\Delta(t)$ is the Jacobian of the nonlinear function $f(x, u,\, w,\, t)$ around current state $\hat{x}(t)$. The Eqs.2.5.5 and 2.5.8 constitute the prediction step of the EKF. Integrating numerically these equations, the solution for state estimate and state error covariance are advanced one time step to obtain $\hat{x}(t+1) = \hat{x}_{k+1}^-$ and $P(t+1) = P_{k+1}^-$, respectively.

Update Step:

In the update step of the EKF, we suppose that the measurement at step $k+1$ has just been processed. The linearized system in Eq.2.5.4 changes based upon the predicted state estimate but remains constant during the measurement step, therefore, the state estimate is updated as in the linear Kalman filter. In this step, the updated estimate of the state $\hat{x}^+_{k+1}$ is obtained from

$$\hat{x}^+_{k+1} = \hat{x}^-_{k+1} + K_{k+1}(y_{k+1} - C\hat{x}^-_{k+1}) \tag{2.5.11}$$

The Kalman gain K_{k+1} and the updated state error covariance P^+_{k+1} are calculated from

$$K_{k+1} = P^-_{k+1}C^T(CP^-_{k+1}C^T + R)^{-1} \tag{2.5.12}$$

$$P^+_{k+1} = (I - K_{k+1}C)P^-_{k+1}(I - K_{k+1}C)^T + K_{k+1}RK^T_{k+1} \tag{2.5.13}$$

where R is the covariance of the measurement noise as discrete process. The filter is initialized by using initial state estimate and state error covariance, namely

$$\hat{x}_0 = E(\hat{x}_0) \tag{2.5.14}$$

$$P_0 = E((x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T) \tag{2.5.15}$$

For the linear Kalman filter, P is equal to the covariance of the estimation error. This is not true in the extended Kalman filter since only first order terms are used in linearization of Eq.2.5.1. However, if the linearization errors are small then P should be approximately equal to the covariance of the estimation error. The main difference between EKF and linear Kalman filter in the implementation is that the state error covariance and the Kalman gain cannot be computed as closed form in the EKF. The difficulty in EKF approach is need to solve Eq. 2.5.8 at each time step since the Jacobian of the nonlinear function $\Delta(t)$ is updated using current state. For long-duration simulations and large size models this is computationally demanding.

Chapter 3

Steady State Kalman Gain Estimation

3.1 Background and Motivation

In the classical presentation of the Kalman filter, the filter gain, K, is computed given the state space model parameters (A, G, C) and noise covariance matrices (Q, R and S, which are referred to as covariance of unknown excitations, covariance of measurement noise, and cross-covariance between these two, respectively). Q, R and S are seldom known a priori and work to determine how to estimate these matrices, and K from the measured data began soon after introduction of the filter. Methods to estimate the Kalman gain from measurements can be classified into two groups: (1) Direct Kalman gain approaches, (2) Indirect noise covariance approaches.

The direct approaches identify the K directly from measured data. The indirect approaches estimate the Q, R and S first, and then use them to obtain K. The state error covariance matrix, P, provides an estimate of the uncertainty in estimates of the state. Therefore one can require P to evaluate the performance of the state estimation as well as K. If all that is of interest is K, one can take the direct approach, i.e., from the data to K, or the indirect one that estimates Q, R and S first. Nevertheless, if in addition to estimating the state one is interested in the covariance of the state error, then the computation of Q, R and S becomes necessary since there is no way to go from the data to the state error covariance, P.

Another classification of the methods to estimate Kalman filter and noise covariance matrices from data, which is initially presented by Mehra [20], has four categories : (1) Bayesian estimation, (2) Maximum likelihood estimation, (3) Covariance matching and (4) Correlation approaches. In the first two categories the problem is posed as parameter estimation. The Bayesian estimation appraoch involves numerical integrations over a candidate parameter space, which is reported being time consuming and impractical in applications, [21]. In the maximum likelihood estimation, the parameters of noise covariance matrices or the Kalman gain are calculated by solving a nonlinear optimization problem in which the convergence is not guaranteed, [22]. In the covariance matching technique, noise covariance matrices are calculated on the premise that the sample covariance of the innovations is consistent with its theoretical value. The convergence of the covariance matching technique has not been proved [23]. Among the methods that have been developed to estimate Kalman gain and noise covariance matrices from data, the correlation methods have received most attention. Since the output statistics have the information of Q, R, S and K, the correlations are used to extract this information from the measurements or the innovations from an arbitrary filter gain. The correlation methods are mainly applicable to time invariant systems and estimate the K and noise covariance matrices in an off-line setting.

In the innovations approach, one begins by postulating an arbitrary stable filter and calculates the optimal gain, the K, from analysis of the resulting innovations. The basic idea is to derive a set of equations relating the system matrices to the sample autocorrelation functions of innovations or measurements, and these equations are solved simultaneously for the noise covariance matrices or the steady state Kalman gain. The output correlation approach can be viewed as a special case of the innovations where the initial gain is taken as zero. In practice, however, the output correlation alternative has been discussed as a separate approach (probably) because the mathematical arrangements that are possible in this special case cannot be easily extended to the case where $K \neq 0$. The correlation approach requires the system to be stable and the output to be stationary. The innovations correlation approach, however, is applicable to unstable systems. Both methods require the system to be completely controllable and observable. Output and innovations correlation approaches are considered for both direct and indirect Kalman gain estimation.

The study in this chapter examines the correlation approaches. A fundamental contribution on the innovations correlations approach is the paper by Heffes [24], who derived an expression for the covariance of the state error of any suboptimal filter as a function of Q, R, assuming process and measurement noise are mutually uncorrelated, namely $S = 0$. The use of correlations of measurements and innovations from an arbitrary filter gain to estimate the noise covariance matrices first was considered by Mehra [20, 15], who built on Heffes's expression. Given Mehra's algorithm, it may appear that the Q and R problem had been solved, but interest lingered because the solution is sensitive to the inevitable errors in the estimation of the covariance of the innovations. Some modifications to Mehra's algorithm that could lead to improved performance were noted by Neethling and Young [25], namely: 1) solving a single least square problem by combining the parameters of Q and R in a vector, 2) enforcement of semi-definitiveness in the covariance matrices and 3) formulation of the problem as an over-determined weighted least squares one.

In the innovations approach results used for Q and R depend on the initial estimation of noise covariance matrices to obtain a suboptimal filter gain and a question that arises is whether the answer is sensitive to this selection or not. In this regard, Mehra [15] suggested that the estimation of Q and R could be repeated by starting with the gain obtained in the first attempt, but this expectation was challenged by Carew and Belanger [16], who noted that if the exact gain is used as the initial guess the approximations in the approach are such that the correctness will not be confirmed. Recently, some other contributions to the Mehra's approach on the estimation of noise covariance matrices are presented. Odelson, Rajamani, and Rawlings applied the suggestions of Neethling and Young's on Mehra's approach and used the vector operator solution for state error covariance Riccatti equation of suboptimal filter, [26]. Akesson et al. extended their work for mutually correlated process and measurement noise case i.e., $S \neq 0$, [27]. Bulut, Vines-Cavanaugh and Bernal compared the performance of the output and innovations correlations approaches to estimate noise covariance matrices, [28]. Dunik, Simandl, and Straka compared the method presented by Odelson, Rajamani, and Rawlings to a combined state and parameter estimation approach [29].

The most widely quoted strategies to carry out the direct estimation of K are due to Mehra [15] and the subsequent paper by Carew and Belanger [16]. The techniques are both iterative in nature and use correlations of innovations from an arbitrary gain. Mehra's approach has disadvantage of solving another Lyapunov equation in each iteration. Moreover, examination show that the accuracy of the estimates highly depends on the initial gain. The method requires the initial gain being close to the correct one, otherwise, the Kalman gain estimates are converging to wrong values even though exact correlation functions of innovations are used. Carew and Bellanger's strategy to calculate K is based on the estimation error that is defined as discrepancy between optimal state estimates obtained from Kalman gain, and state estimates obtained form an arbitrary gain. The main advantages of this method are solving Lyapunov equation for state error covariance and updating the estimate of K in the same recursion as opposed to Mehra's approach. Carew and Bellanger's technique is robust to initial gain. Output correlation approach for direct estimation of Kalman gain is presented by Mehra in 1972 [20] and one year later the same approach has been shown by Son and Anderson in more detail, [17]. The most important difference of the output correlation approach compared to the innovations correlation approach is its being non-recursive. The drawback of the output correlations approach is that poor estimates of sample output correlations functions can lead to an ill-conditioned Riccati equation.

In an off-line setting the estimation of noise statistics and the associated optimal gain lead to a problem of the form;

$$HX = L \tag{3.1.1}$$

where H is a tall matrix formed by using system matrices and the arbitrary filter gain, X is a column vector contains the entries in noise covariance matrices as unknowns and L is a vector valued function of correlation functions calculated from

$$\Lambda_j \overset{def}{=} E(y_k y_{k-j}^T) = \frac{1}{N-j} \sum_{k=1}^{N-j} y_k y_{k-j}^T \tag{3.1.2}$$

$$\mathcal{L}_j \overset{def}{=} E(\bar{e}_k \bar{e}_{k-j}^T) = \frac{1}{N-j} \sum_{k=1}^{N-j} \bar{e}_k \bar{e}_{k-j}^T \tag{3.1.3}$$

where Λ_j is the correlation functions of measurements y_k, $\mathcal{L}_j$ is the correlation functions of innovations $\bar{e}_k$ from an arbitrary filter, and N is the number of time steps. In theory, when $N \to \infty$, exact values of Λ_j and $\mathcal{L}_j$ can be calculated from Eqs.3.1.2-3.1.3, from which an exact L is formed. When H is full rank and L is exact, a unique and exact solution for X is obtained from Eq.3.1.1. However in practical applications, only an estimate of Λ_j and $\mathcal{L}_j$ can be calculated since a finite duration of observed data is available, namely $N \ll \infty$. Therefore, the error in L is an inevitable problem in the correlations approaches. Let $\hat{L}$ be an estimate of L with error ε, namely $L = \hat{L} + \varepsilon$. Substituting it in Eq.3.1.1 gives

$$HX = \hat{L} + \varepsilon \tag{3.1.4}$$

A minimum error norm approximation of X, which we denote as $\hat{X}$ can be calculated from Eq.3.1.4, however, the estimates rely heavily on the accuracy of the $\hat{L}$, which, in general, requires a large amount of data. An illustration to this point by quantifying the size of the variance in inflation associated with correlation methods is depicted in Fig.3.1, in which a three-degree-of-freedom shear frame subjected to unmeasured Gaussian disturbance with a unit covariance at the first story is considered. Covariance of unmeasured disturbance is estimated from correlation of innovation process generated an arbitrary gain, The results are obtained from 100 simulations for three cases with durations of measurements as {100, 200, 500} seconds. The probability distribution functions of estimates of Q are estimated from 100 simulations by fitting a generalized extreme value density function. As can be seen, the increase in the durations of measurements is resulting in better estimates of Q. However, although 500 seconds of data (sampled at 200Hz) is used, the estimates of Q has a standard deviation of 0.465, and in some some cases the estimates are negative which is unphysical.

The approximate solution of noise covariance matrices and the K calculated from Eq.3.1.4 is very sensitive to the changes in the right hand side due to the fact that coefficient matrix, H is in general ill-conditioned. This implies that the possibility exists for the small error in L which leads to very large changes on the calculated solution of X. Therefore, correlation approaches lead to very poor estimates of covariance matrices and

Kalman gain, not to mention that in some cases they are simply wrong. The study in this chapter is aimed to investigate the merit of Tikhonov's regularization technique in the estimation of the noise covariance matrices and optimal gain. Part of the objective of the chapter is to offer a detailed review of the classical correlation based approaches that may prove useful in structural engineering.

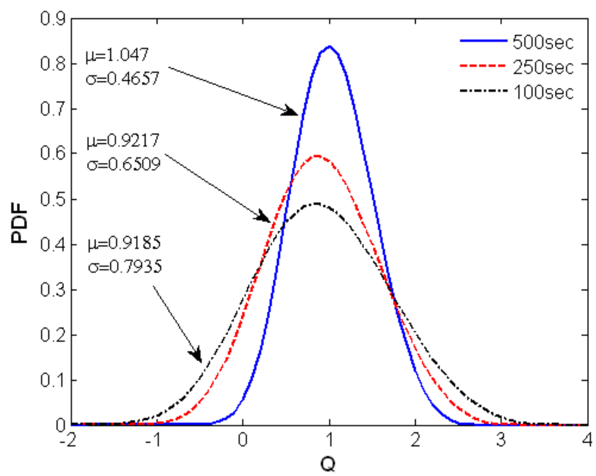

Figure 3.1: Experimental PDFs of process noise covariance estimates from 100 simulations.

3.2 Indirect Noise Covariance Approaches

3.2.1 The Output Correlations Approach

Output correlations approach to estimate noise covariance matrices has two main steps. In first step estimates of Q and S are obtained using sample correlation functions of measurement. In the second step R is calculated by using sample zero lag correlation and state covariance, Σ.

Estimation of Q and S

Applying the *vec* operator to Eqs.2.3.135 and 2.3.132 one obtains

$$vec(\Lambda_j) = (C \otimes CA^j)vec(\Sigma) + (I \otimes CA^{j-1}G)vec(S) \tag{3.2.1}$$

and

$$vec(\Sigma) = (I - (A \otimes A))^{-1}(G \otimes G)vec(Q) \tag{3.2.2}$$

where $\otimes$ denotes the Kronecker product. Substituting Eq.3.2.2 into Eq.3.2.1 and organizing one gets

$$l_j = \begin{bmatrix} h_j^Q & h_j^S \end{bmatrix} \begin{bmatrix} vec(Q) \\ vec(S) \end{bmatrix} \tag{3.2.3}$$

where

$$l_j = vec(\Lambda_j) \tag{3.2.4}$$

$$h_j^Q = \left[C \otimes CA^j)(I - (A \otimes A))^{-1}(G \otimes G)\right] \tag{3.2.5}$$

$$h_j^S = \left[(I \otimes CA^{j-1}G)\right] \tag{3.2.6}$$

Listing explicitly the correlation functions in Eq.3.2.3 for lags $j = 1,\ 2\,,..p$, and writing in matrix form one has

$$HX = L \tag{3.2.7}$$

where

$$H = \begin{bmatrix} h_1^Q & h_1^S \\ h_2^Q & h_2^S \\ \vdots & \vdots \\ h_p^Q & h_p^S \end{bmatrix}, \quad L = \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_p \end{bmatrix}, \quad X = \begin{bmatrix} vec(Q) \\ vec(S) \end{bmatrix} \tag{3.2.8}$$

Estimates of Q and S can be obtained from Eq.3.2.7. Examination of Eq.3.2.7 shows that H has dimensions $m^2px(r^2 + mr)$ where we recall that m and r represent the numbers of outputs and independent disturbances, respectively.

The sufficient condition for the uniqueness of the solution of Eq.3.2.7 is defined as follows in the general case: the number of unknown parameters in Q and S have to be smaller than the product of number of measurements and the state. The error in solving Eq.3.2.7 for X is entirely connected to the fact that the L is approximate since it is constructed from sample correlation functions of the output which are estimated from finite duration signals. Substituting $\hat{L}$ as the estimate of L, the solution of Eq.3.2.7 can be presented as in the following.

Case $\#1 \; mn \geqq (r^2 + mr)$

In this case H is full rank and there exists a unique minimum norm solution for a weighting matrix I given in the following,

$$\hat{X} = (H^T H)^{-1} H^T \hat{L} \tag{3.2.9}$$

Case $\#2 \; mn < (r^2 + mr)$

In this case the matrix is rank deficient, and the size of null space of H can be calculated from $t = r^2 - mn$. The solution is written as follows,

$$\hat{X} = \hat{X}_0 + null(H)Y \tag{3.2.10}$$

where $\hat{X}_0$ is the minimum norm solution given in Eq.3.2.9 and $Y \epsilon R^{tx1}$ is an arbitrary vector. Therefore, we conclude Eq.3.2.7 has infinite solution when $mn < (r^2 + mr)$.

Estimation of R

Estimation of R is quite simple after one has Q and S. It first requires solving the Lyapunov equation in Eq.2.3.132 for Σ, then using the zero lag correlation function in Eq.2.3.134, estimate of R can be obtained as follows,

$$\hat{R} = \hat{\Lambda}_0 - C\hat{\Sigma}C^T \tag{3.2.11}$$

Although a unique solution for Q and S does not exist when $mn < (r^2 + mr)$, any of the solution for Q, R and S from Eqs.3.2.10 and 3.2.11, still gives the optimal Kalman gain. K can be calculated using classical approach which involves solving Riccati equation given in Eq.2.3.50 for error covariance, P and obtaining K from Eq.2.3.51. In fact in this case although resulting K and covariance of innovations F are correct, P is not correct and it cannot be calculated without knowing the correct noise covariance matrices.

3.2.2 The Innovations Correlations Approach

As opposed to output correlations approach, innovations correlations approach gives estimates of noise covariance matrices from a single step procedure. We start applying *vec* operator to both sides of the auto-correlation function of the innovations in Eqs.2.3.119-2.3.120, namely

$$vec(\mathcal{L}_j) = (C \otimes C)vec(\bar{P}) + vec(R) \qquad j = 0 \tag{3.2.12}$$

$$\begin{aligned} vec(\mathcal{L}_j) = (C\bar{A}^j \otimes C)vec(\bar{P}) + (G^T \otimes C\bar{A}^{j-1})vec(S)- \\ -(I \otimes C\bar{A}^{j-1}K_0)vect(R) \qquad j > 0 \end{aligned} \tag{3.2.13}$$

and applying vec operator to error covariance equation in Eq.2.3.114, one has

$$\begin{aligned} vec(\bar{P}) = [I - (\bar{A} \otimes \bar{A})]^{-1}[(K_0 \otimes K_0)vec(R) + G \otimes Gvec(Q)- \\ - (G \otimes K_0)vec(S) - (K_0 \otimes G)vec(S^T)] \end{aligned} \tag{3.2.14}$$

Substituting Eq.3.2.14 into Eqs.3.2.12 and 3.2.13 , and adding the terms related to S^T to the terms related to S and canceling S^T , one finds

$$vec(\mathcal{L}_j) = \begin{bmatrix} h_j^Q & h_j^S & h_j^R \end{bmatrix} \begin{bmatrix} vec(Q) \\ vec(S) \\ vec(R) \end{bmatrix} \tag{3.2.15}$$

where

$$h_j^Q = (C \otimes C)[I - (\bar{A} \otimes \bar{A})]^{-1}(G \otimes G) \qquad j = 0 \tag{3.2.16}$$

$$h_j^Q = (C \otimes C\bar{A}^j)[I - (\bar{A} \otimes \bar{A})]^{-1}(G \otimes G) \qquad j > 0 \tag{3.2.17}$$

$$h_j^S = -2I(C \otimes C)[I - (\bar{A} \otimes \bar{A})]^{-1}(G \otimes K_0) \qquad j = 0 \tag{3.2.18}$$

$$h_j^S = (B^T \otimes C\bar{A}^{j-1}) - 2I[(C \otimes C\bar{A}^j)[I - (\bar{A} \otimes \bar{A})]^{-1}(G \otimes K_0)] \;\; j > 0 \tag{3.2.19}$$

$$h_j^R = (C \otimes C)[I - (\bar{A} \otimes \bar{A})]^{-1}(K_0 \otimes K_0) + I \qquad j = 0 \tag{3.2.20}$$

$$h_j^R = (C \otimes C\bar{A}^j)[I - (\bar{A} \otimes \bar{A})]^{-1}(K_0 \otimes K_0) - (I \otimes C\bar{A}^{j-1}K_0) \quad j > 0 \tag{3.2.21}$$

Listing explicitly the correlation functions in Eq.3.2.15 for lags $j = 1,\, 2\,, ..p$ and writing in matrix form one has

$$HX = L \tag{3.2.22}$$

where

$$H = \begin{bmatrix} h_0^Q & h_0^S & h_0^R \\ h_1^Q & h_1^S & h_1^R \\ h_2^Q & h_2^S & h_2^R \\ \vdots & \vdots & \vdots \\ h_p^Q & h_p^S & h_p^R \end{bmatrix},\; L = \begin{bmatrix} vec(\mathcal{L}_0) \\ vec(\mathcal{L}_1) \\ vec(\mathcal{L}_2) \\ \vdots \\ vec(\mathcal{L}_p) \end{bmatrix},\; X = \begin{bmatrix} vec(Q) \\ vec(S) \\ vec(R) \end{bmatrix}$$

Estimates of Q, S and R can be obtained from Eq.3.2.22. From its inspection, one finds that H has dimensions $m^2px(r^2 + m^2 + mr)$. The sufficient condition for the uniqueness of the solution of Eq.3.2.22 is identical with the uniqueness of the solution of Eq.3.2.7. The observations presented in section 3.2.1 on the uniqueness of the Kalman gain, K , the error covariance, P, and covariance of the innovations, F, hold in this case as well.

3.3 Direct Kalman Gain Approaches

3.3.1 The Output Correlations Approach

Output correlations approach for direct estimation of Kalman gain has three steps. In the first step, an estimate of $A\Sigma C^T + GS$ is obtained from sample correlation functions of measurement solving a least square problem. In the second step, a Riccati equation is solved for state error estimate covariance using the estimate of $A\Sigma C^T + GS$ and zero lag correlation. In the last step, the Kalman gain K is obtained using the state estimate covariance Σ and zero lag correlation of output. We start with listing explicitly the correlation functions in Eq.2.3.135 for lags $j = 1,\ 2\ ,..p$, namely

$$\Lambda_1 = C(A\Sigma C^T + GS) \tag{3.3.1}$$

$$\Lambda_2 = CA(A\Sigma C^T + GS) \tag{3.3.2}$$

$$\Lambda_2 = CA^3(A\Sigma C^T + GS) \tag{3.3.3}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\Lambda_p = CA^{p-1}(A\Sigma C^T + GS) \tag{3.3.4}$$

from where one can write

$$\Lambda = \mathcal{T}(A\Sigma C^T + GS) \tag{3.3.5}$$

where

$$\Lambda = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \\ \vdots \\ \Lambda_p \end{bmatrix} \tag{3.3.6}$$

and

$$\mathcal{T} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{p-1} \end{bmatrix} \tag{3.3.7}$$

The estimate of $(A\Sigma C^T + GS)$ can be obtained from minimum norm solution of Eq.3.3.5 using sample correlations as follows

$$A\Sigma C^T + GS = \mathcal{T}^{\dagger}\hat{\Lambda} \tag{3.3.8}$$

where $\mathcal{T}^{\dagger}$ is the pseudo-inverse of $\mathcal{T}$. On the assumption that the system is stable and observable one concludes that $\mathcal{T}$ attains full column rank when p is no larger than the order of the system, n. The error in solving Eq.3.3.8 for $A\Sigma C^T + GS$ is entirely connected to the fact that the $\hat{\Lambda}$ is approximate since it is constructed from sample correlation functions of the output which are estimated from finite duration signal.

One can write the relation between state error covariance and state covariance as follows, [30]

$$P = \Sigma - \Gamma \tag{3.3.9}$$

where Γ is the state estimate covariance, namely

$$\Gamma \overset{def}{=} E(\hat{x}_k \hat{x}_k^T) \tag{3.3.10}$$

Substituting Eq.3.3.9 into Eq.2.3.91, one gets

$$K = (A\Sigma C^T - A\Gamma C^T + GS)(C\Sigma C^T - C\Gamma C^T + R)^{-1} \tag{3.3.11}$$

Defining

$$M \overset{def}{=} A\Sigma C^T + GS \tag{3.3.12}$$

and using zero lag correlation of output in Eq.2.3.134, yields

$$K = (M - A\Gamma C^T)(\Lambda_0 - C\Gamma C^T)^{-1} \tag{3.3.13}$$

One requires state estimate covariance, Γ to obtain an estimate of K from 3.3.13. Substituting Σ from Eq.2.3.132 and P from Eq.2.3.90 into Eq.3.3.9 one can obtain the following Riccati equation for Γ,

$$\Gamma = A\Gamma A^T - (A\Sigma C - A\Gamma C + GS)(C\Sigma C^T - C\Gamma C^T + R)^{-1}(A\Sigma C - A\Gamma C + GS)^T \tag{3.3.14}$$

Using the definition in Eq.3.3.12 and zero lag correlation in Eq.2.3.134, yields

$$\Gamma = A\Gamma A^T - (M - A\Gamma C)(\Lambda_0 - C\Gamma C^T + R)(M - A\Gamma C)^T \tag{3.3.15}$$

Substituting the estimates of M and zero lag correlation Λ_0 into Eq.3.3.13, one gets

$$\hat{\Gamma} = A\hat{\Gamma}A^T - (\hat{M} - A\hat{\Gamma}C)(\hat{\Lambda}_0 - C\hat{\Gamma}C)^{-1}(\hat{M} - A\hat{\Gamma}C)^T \tag{3.3.16}$$

Finally, estimate of K is obtained by solving the Riccati Eq.3.3.16 for $\hat{\Gamma}$, and substituting it into Eq.3.3.13, gives

$$\hat{K} = (\widehat{M} - A\hat{\Gamma}C^T)(\hat{\Lambda}_0 - C\hat{\Gamma}C^T)^{-1} \tag{3.3.17}$$

We note that an estimate of covariance of innovations process from optimal filter can be obtained from

$$\hat{F} = \hat{\Lambda}_0 - C\hat{\Gamma}C^T \tag{3.3.18}$$

Output correlations approach to estimate K is summarized in three steps in the following:

(1) Obtain estimate of M from Eq.3.3.8.

(2) Solve the Riccati equation in Eq.3.3.16 for $\hat{\Gamma}$ using the estimates of M and Λ_0.

(3) Obtain estimate of K from 3.3.17 using the estimates of M , Λ_0 and Γ.

3.3.2 The Innovations Correlations Approach

Here we describe an innovations correlations approach for direct estimation of K which initially presented by Carew and Bellanger [16]. We start rewriting innovations correlation functions by combining Eqs.2.3.123 and 2.3.124 as follows

$$\mathcal{L}_j = C\bar{A}^{j-1}(A\tilde{P}C^T + KF - K_0\mathcal{L}_0) \qquad j > 0 \tag{3.3.19}$$

Listing explicitly the correlation functions in Eq.3.3.19 for lags $j = 1,\, 2\,, ..p$ one has

$$\mathcal{L}_1 = C(A\tilde{P}C^T + KF - K_0\mathcal{L}_0) \tag{3.3.20}$$

$$\mathcal{L}_2 = C\bar{A}(A\tilde{P}C^T + KF - K_0\mathcal{L}_0) \tag{3.3.21}$$

$$\mathcal{L}_3 = C\bar{A}^2(A\tilde{P}C^T + KF - K_0\mathcal{L}_0) \tag{3.3.22}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\mathcal{L}_p = C\bar{A}^{p-1}(A\tilde{P}C^T + KF - K_0\mathcal{L}_0) \tag{3.3.23}$$

from where one can write

$$\mathcal{L} = \mathcal{O}(A\tilde{P}C^T + KF - K_0\mathcal{L}_0) \tag{3.3.24}$$

where

$$\mathcal{L} = \begin{bmatrix} \mathcal{L}_1 \\ \mathcal{L}_2 \\ \mathcal{L}_3 \\ \vdots \\ \mathcal{L}_p \end{bmatrix}, \ \mathcal{O} = \begin{bmatrix} C \\ C\bar{A} \\ C\bar{A}^2 \\ \vdots \\ C\bar{A}^{p-1} \end{bmatrix}$$

As can be seen, matrix $\mathcal{O}$ is the observability block of an observer whose gain is K_0. On the assumption that the closed loop is stable and observable one concludes that $\mathcal{O}$ attains full column rank when p is no larger than the order of the system, n. Defining

$$Z = A\tilde{P}C^T + KF \tag{3.3.25}$$

and accepting that the matrix, $\mathcal{O}$ is full rank and using the estimates of $\mathcal{L}_0$ and $\mathcal{L}$ one finds that the unique least square solution to Eq.3.3.24 is

$$\hat{\mathcal{Z}} = \mathcal{O}^{\dagger}\hat{\mathcal{L}} + K_0\hat{\mathcal{L}}_0 \tag{3.3.26}$$

where $\mathcal{O}^{\dagger}$ is the pseudo-inverse of $\mathcal{O}$. The error in solving Eq.3.3.26 for $\mathcal{Z}$ is entirely connected to the fact that the $\hat{\mathcal{L}}_0$ and $\hat{\mathcal{L}}$ matrices are approximate since they constructed from sample correlation functions of the innovations which are estimated from finite duration signal.

Solving Lyapunov equation in Eq.2.3.126 for $\tilde{P}$ is essential to calculate K from $\mathcal{Z}$. This requires a recursive algorithm which is shown in the following,

$$F_k = \hat{\mathcal{L}}_0 - C\tilde{P}_kC^T \tag{3.3.27}$$

$$K_k = (\hat{\mathcal{Z}} - A\tilde{P}_kC^T)F_k^{-1} \tag{3.3.28}$$

$$\tilde{P}_{k+1} = \bar{A}\tilde{P}_k\bar{A}^T + (K_k - K_0)F_k(K - K_0)^T \tag{3.3.29}$$

Algorithm can be started assuming $\tilde{P}_0$ is a null matrix. Reader is referred to [16] for convergence analysis of the algorithm.

3.4 Ill-conditioned Least Square Problem

In this section we address the ill-conditioned least square problem that arise in correlation approaches and describe Tikhonov's regularization technique to approach the matter. As shown in the sections 3.2-3.3, the correlation methods to estimate the Kalman gain from measurements lead to a problem of the form (refer to Eqs.3.2.7-3.3.8 and Eqs.3.2.22-3.3.26 for output and innovations correlations approaches, respectively),

$$HX = L \tag{3.4.1}$$

where $H \epsilon \mathbb{R}^{axb}$ is a tall matrix formed by using system matrices and the arbitrary filter gain, $X \epsilon \mathbb{R}^{bx1}$ is a column vector contains the entries in noise covariance matrices as unknowns and $L \epsilon \mathbb{R}^{ax1}$ is a vector valued function that contains correlation functions of the output or innovations correlations. In the limiting case where the output sequences is infinite an exact L can be calculated. When H is full rank and L is exact, a unique and exact least squares minimum norm solution of Eq.3.4.1 is calculated from

$$X = \underset{X}{min} \left(\|HX - L\|_2 \right) \tag{3.4.2}$$

where $\left\|\cdot\right\|_2$ denotes the L_2 vector norm. A closed form solution is

$$X = H^{\dagger} L \tag{3.4.3}$$

where $\dagger$ denotes pseudo inverse, namely

$$H^{\dagger} = (H^T H)^{-1} H^T \tag{3.4.4}$$

Another form of the solution can be given by using singular value decomposition (SVD) of H, which is

$$H = U \Sigma V^T = \sum_{i=1}^{b} u_i \sigma_i v_i^T \tag{3.4.5}$$

where $U = (u_1, \ldots, u_b)$ and $V = (v_1, \ldots, v_b)$ are matrices with orthonormal columns, $U^T U = V^T V = I$, and $\Sigma = diag(\sigma_1, \ldots, \sigma_b)$ has non-negative diagonal elements appearing in decreasing order such that

$$\sigma_1 \geqslant \ldots \sigma_b \geqslant 0 \tag{3.4.6}$$

Replacing H in Eq.3.4.3 by its SVD yields

$$X = V\Sigma^{\dagger}U^{T}L \tag{3.4.7}$$

$$X = \sum_{i=1}^{b} \frac{u_i^T L}{\sigma_i} v_i \tag{3.4.8}$$

As can be seen from Eq.3.4.8 that the solution is controlled by the singular values σ_i. A condition that insures the solution stability is that the term $u_i^T L$ on the average has to decay to zero faster than the corresponding values of σ_i, which is known as Discrete Picard Condition (DPC).

The issue in the correlations approaches is that, in practical applications, only an estimate of output or innovations correlations can be calculated since a finite duration of observed data is available. Therefore, the uncertainty in the estimate of L is an inevitable problem. Let $\hat{L}$ be an estimate of L with error ε, namely $\hat{L} = L + \varepsilon$. Substituting it in Eq.3.1.1 gives

$$HX = L + \varepsilon \tag{3.4.9}$$

A least squares approximation of X, which we denote as $\hat{X}$, is calculated from

$$\hat{X} = \min_{X} \left(\left\| HX - \hat{L} \right\|_2 \right) \tag{3.4.10}$$

and SVD form of the solution is

$$\hat{X} = \sum_{i=1}^{b} \frac{u_i^T L}{\sigma_i} v_i + \sum_{i=1}^{b} \frac{u_i^T \varepsilon}{\sigma_i} v_i \tag{3.4.11}$$

Thus $\hat{X}$ consists of two terms, one is the actual solution of the error-free problem, and the other is the contribution from the error. In correlations approaches, the accuracy of $\hat{X}$ rely heavily on the second term which does not satisfy the DPC. The term $u_i^T \varepsilon$ corresponding to the smaller σ_i do not decay as fast as the σ_i, therefore the solution governed by the terms in the sum corresponding to the smallest σ_i . An illustration of this issue is depicted in Fig.3.2. As can be seen most of the $u_i^T L$ satisfied the DPC however for large i both the $u_i^T L$ and the σ_i become dominated by rounding errors. For the same problem with $\hat{L}$, we see that the $u_i^T \hat{L}$ become dominated by the error so small values of σ_i are causing the ratios $\frac{u_i^T \hat{L}}{\sigma_i}$ blow up.

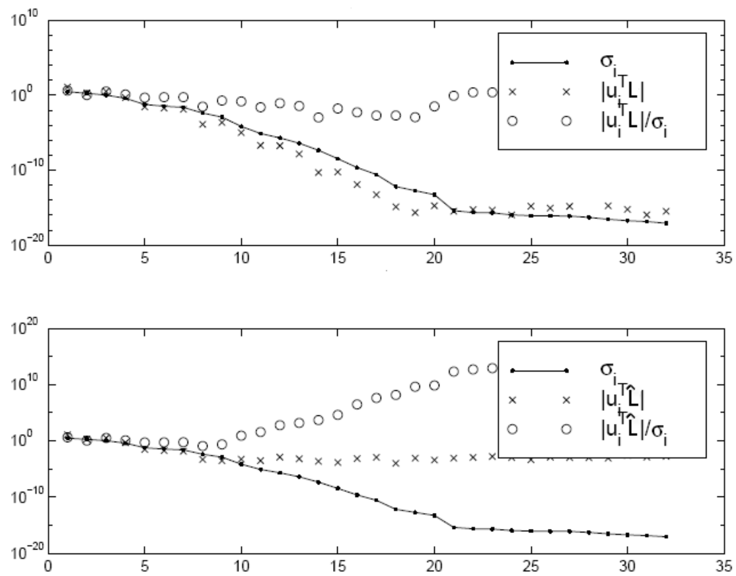

Figure 3.2: Discrete Picard Condition (DPC) [31], Top: DPC for L and, Bottom: DPC for $\hat{L}$.

In practical applications, the condition number of H is used to describe how sensitive the solution is to changes in $\hat{L}$, which is defined as the ratio of the largest singular values of the H to the smallest, namely

$$\kappa_H = \frac{\sigma_1}{\sigma_b} \tag{3.4.12}$$

The big condition number implies that the possibility exists for small error in L lead to very large changes on the calculated solution of X. A matrix with a low condition number is said to be well-conditioned, while a matrix with a high condition number is said to be ill-conditioned. The ill-conditioning of H matrix in Eq.3.4.1 is a concern only when one must operates with finite length sequences which is the case in real situations. In the limiting case where the output sequences are infinite in principle the results from output correlations and innovations correlations converge to the exact solution independently of the conditioning of the H matrix. The theory for ill-conditioned problems is well developed in the literature. A review can be found in [32].

Tikhonov's Regularization

Numerical regularization methods aim to calculate an approximate solution for ill-conditioned problem by formulating a related well-conditioned problem that applies some constraint to the desired solution. In general, the constraint is chosen such that the size of the solution is controlled, namely

$$\|X\|_2 < \delta \tag{3.4.13}$$

The constraint on the solution leads to a multi-objective problem where one tries to balance the perturbations and the constraint on the solution. Therefore regularization methods has two aspects: (1) Computation of the regularization parameter and (2) Computation of the regularized solution, [33]. Among the regularization methods the Tikhonov's method is most commonly used and well-known form of regularization, because of the existence of a closed-form solution. A general form of the Tikhonov's regularization, which is basically adding a side constraint to the objective function, is

$$\hat{X} = \min_{X} \left(\left\| HX - \hat{L} \right\|_2^2 + \lambda^2 \left\| (X - \hat{X}_0) \right\|_2^2 \right) \tag{3.4.14}$$

where is λ regularization parameter chosen by the user which introduces a suitable bias-variance trade-off between the residual norm, $\left\| HX - \hat{L} \right\|_2$ and solution norm, $\left\| (X - \hat{X}_0) \right\|_2$ and $\hat{X}_0$ is initial guess for the solution. The closed form of the Tikhonov's regularization in Eq.3.4.14 can be written as

$$\hat{X} = \hat{X}_0 + (H^T H + \lambda^2 I)^{-1} H^T (\hat{L} - H\hat{X}_0) \tag{3.4.15}$$

where I is the identity matrix. λ controls the sensitivity of the regularized solution $\hat{X}$ to the norm of the solution and perturbations in L. A large λ (equivalent to a large amount of regularization) favors a small solution norm at the cost of a large residual norm, while a small λ has the opposite effect.

An alternative form of the Tikhonov's regularization is given as follows

$$\hat{X} = \min_{X} \left(\left\| \begin{bmatrix} H \\ \lambda I \end{bmatrix} X - \begin{bmatrix} \hat{L} \\ 0 \end{bmatrix} \right\|_2 \right) \tag{3.4.16}$$

which is also called *damped least square problem*. Applying SVD of the H matrix given in Eq.3.4.5 yields

$$\hat{X} = V(\Sigma^T \Sigma + \lambda^2 I)^{-1} \Sigma^T U^T \hat{L} \tag{3.4.17}$$

$$\hat{X} = \sum_{i=1}^{b} \varphi_i \frac{u_i^T \hat{L}}{\sigma_i} v_i \tag{3.4.18}$$

where φ_i is called Tikhonov's regularization filter factors

$$\varphi_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \simeq \begin{cases} 1 & \sigma_i \gg \lambda \\ \left(\frac{\sigma_i}{\lambda}\right)^2 & \sigma_i \ll \lambda \end{cases} \tag{3.4.19}$$

As can be seen from Eq.3.4.18 that Tikhonov's regularization performs as a filter where the contribution of the singular values, that make the system ill-conditioned, is damped out. Thus, the regularization parameter, λ is an important quantity which controls the effect of singular values that close zero on the solution, [31].

Many numerical methods are proposed to compute λ in the literature, in which L-curve and generalized cross-validation (GCV) methods are the most commonly used and well-known. GCV technique is based on the strategy that if an arbitrary element of the L is left out, then the corresponding regularized solution should predict this element of L with good approximation [34, 31]. In the L-curve method, one plots the norm of the regularized solution $||X - X_0||_2$ versus the corresponding residual norm, $\left\|H\hat{X} - \hat{L}\right\|_2$ for all possible regularization parameters. The plot demonstrates the trade off between minimization of these two quantities, and when log-log scale is used for plotting, almost always has a characteristic L-shaped appearance with a corner separating the vertical and the horizontal parts of the curve. The λ used in Tikhonov's regularization is the one that gives the corner of the L curve, [35]. The generic form of an L-curve is depicted in Fig.3.3. The vertical part of the L-curve corresponds to solutions where $||X - X_0||_2$ is very sensitive to changes in λ because of the error in $\hat{L}$. The horizontal part of the L-curve corresponds to solutions where it is the residual norm $\left\|HX - \hat{L}\right\|_2$ that is most sensitive to the regularization parameter because the solution is dominated by the regularization error.

There are many other regularization methods besides Tikhonov's regularization with properties that make them better suited to certain problems. For other methods, the

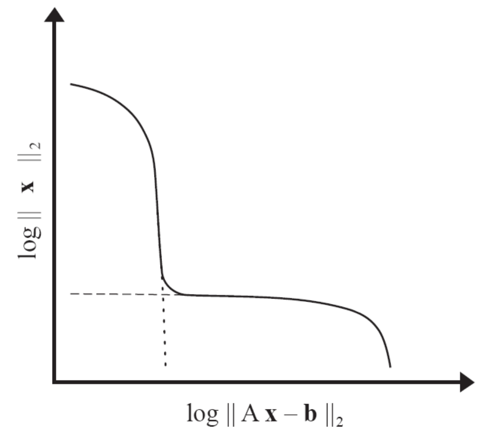

Figure 3.3: The generic form of the L-curve [31].

reader is referred to [32, 31].

Enforcing Positive Semi-definitiveness

By definition, the noise covariance matrix constructed from Q, R and S is positive semi-definite (i.e., all their eigenvalues are ≥ 0), namely

$$\triangle \stackrel{def}{=} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \geqslant 0 \tag{3.4.20}$$

However, due to error in sample correlation functions, the least squares solution of from Eq.3.4.10 may not satisfy this requirement. In the general case one can satisfy positive semi-definitiveness by recasting the problem as an optimization with constraints which is minimizing the $||HX - \hat{L}||_2$ subject to constraint that all eigenvalues of $\triangle \geqslant 0$ namely

$$\hat{X} = \min_{\hat{\triangle} \geqslant 0} \left(\left\| HX - \hat{L} \right\|_2 \right) \tag{3.4.21}$$

The positive semi-definitiveness constraint on the regularized solution can be applied by recasting the problem as

$$\hat{X} = \min_{\hat{\triangle}\geqslant 0}\left(\left\|HX - \hat{L}\right\|_2^2 + \lambda^2\left\|(X - \hat{X}_0)\right\|_2^2\right) \tag{3.4.22}$$

Eqs.3.4.21-3.4.22 are called are least square problem with semi-definite constraint and they can be solved by optimization, [36, 27]. As described in the sections 3.2-3.3, the expressions in innovations correlations approach allows to enforce the positive semi-definiteness of the solution when solving for Q, R and the S. However in output correlations approach the X in Eq.3.2.7 involves only the unknowns of Q and S. R has to be calculated from Eq.3.2.11, which requires the estimates of Q and S. Therefore enforcing the positive semi-definiteness of the solution for X is not possible in the output correlations approach.

3.5 Numerical Experiments

3.5.1 Experiment 1: Five-DOF Spring Mass System

In this experiment, we present an application of innovations correlations approaches to estimate steady Kalman filter gain using the five-DOF spring mass system depicted in Fig.3.4. The un-damped natural frequencies (in Hz) are {2.45, 6.75, 12.21, 13.32, 16.12, 16.5}.

We obtain results for a single output sensor located at coordinate three, which is measuring the velocity data at 100Hz sampling. The stationary stochastic disturbances are assumed to act at all masses. The measurement noise is prescribed to have a root-mean-square (RMS) equal to approximately 10% of the RMS of the response measured. Unmeasured excitations and measurement noise are assumed to be mutually uncorrelated, namely $S = 0$. Two cases of noise statistics are considered which are given in the following,

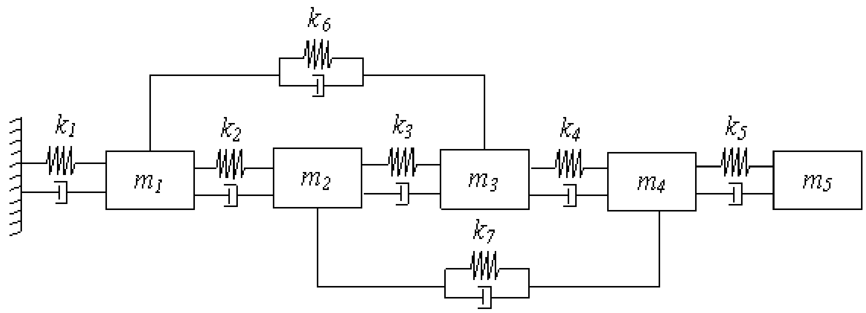

Figure 3.4: Five-DOF spring mass system, $m_1 = m_3 = m_4 = m_5 = 0.05$, $m_2 = 0.10$, $k_1 = k_3 = k_5 = k_7 = 100$, $k_2 = k_4 = k_6 = 120$, (in consistent units). Damping is 2% in all modes.

Case I: Uncorrelated Process Noise;

$$Q = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix} \qquad R = 0.030 \qquad S = 0$$

Case II: Correlated Process Noise;

$$Q = \begin{bmatrix} 19 & 6 & 11 & 0 & -4 \\ 6 & 6 & 5 & 1 & -5 \\ 11 & 5 & 9 & -2 & -4 \\ 0 & 1 & -2 & 5 & 1 \\ -4 & -5 & -4 & 1 & 6 \end{bmatrix} \qquad R = 0.125 \qquad S = 0$$

Innovations correlations approaches are applied to estimate the steady state Kalman gain K in line with sections 3.2-3.3. Innovations process is generated using an arbitrary gain, K_0, that is chosen such that eigenvalues of the matrix $A - K_0C$ are assumed to have the same phase as those of A but with a 20% smaller radius. In the indirect noise covariance approach, the construction of H matrix from Eq.2.5.9 requires only A, C and K_0.

In general case, where one does not know the spatial distribution of the noise and correlation terms in the covariance, full noise covariance matrices with no zero terms can be considered. In Case II, by taking the symmetry into account, the number of unknowns in Q is 15 and in R is 1, which results to an H matrix with 16 columns. However, in Case I only diagonal terms of Q are of interest, so H is constructed by taking only related columns of the full H matrix. The change in the condition number of H matrix for two noise cases, from a range of $p = \{6 - 60\}$ is depicted in Fig.3.5, and 50 lags of correlation functions of innovations process is taken into consideration for further calculations.

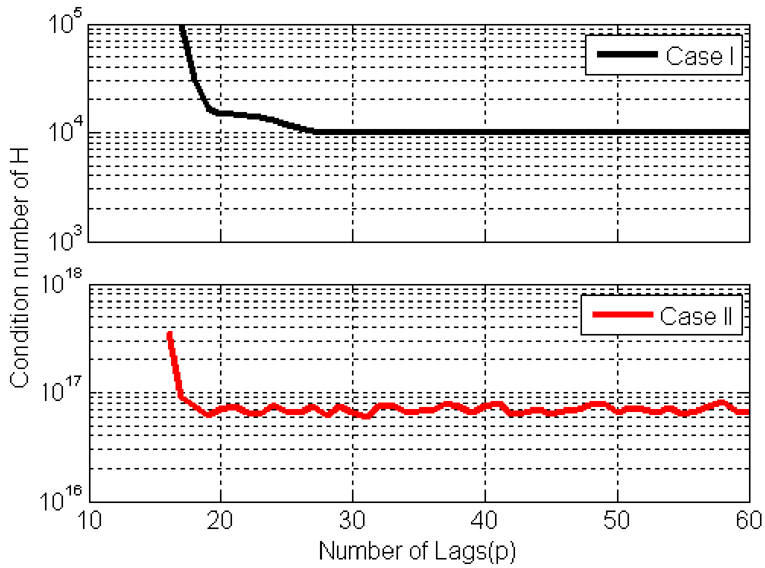

Figure 3.5: Change in the condition number of H matrix with respect to number of lags.

The condition number of H matrix in Eq.3.2.22 for $p = 50$ is $1.0043x10^4$ and $6.53x10^{16}$ for Case I and Case II, respectively. In Case I, the number unknown parameters in Q is smaller than mxn, namely

$$mxn = 1x10 = 10 > 5$$

Therefore, uniqueness condition of noise covariance matrices is satisfied and Q and R matrices are estimated uniquely. In Case II, the number of unknown parameters in Q is bigger than mxn, namely

$$mxn = 1x10 = 10 < 15$$

Therefore the H matrix is rank deficient and solution for noise covariance matrices is not unique. Although a unique solution does not exist in this case, the Q and R estimates are used to calculate the K from classical formulations of the KF. The sample innovation correlations functions are calculated using 200 seconds of data.

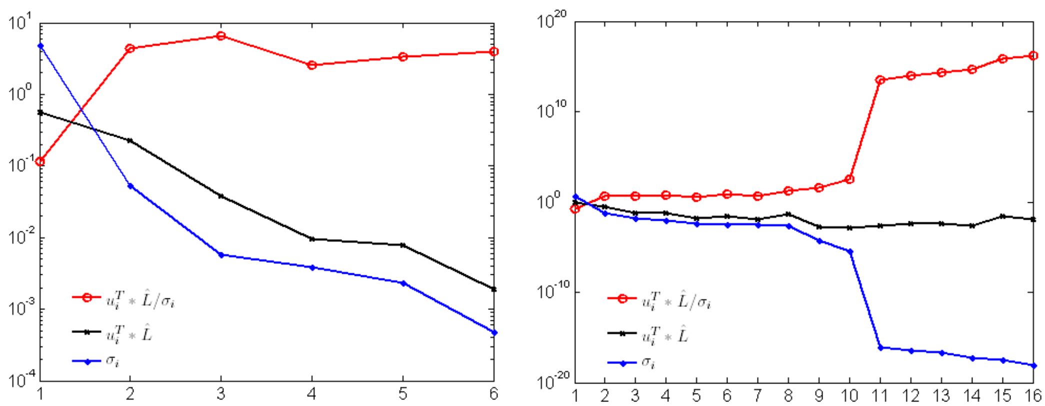

Figure 3.6: Discrete Picard Condition of five-DOF spring mass system; Left: Case I, Right: Case II.

Stability of the least squares solution is examined using Discrete Picard Condition, which is depicted in Fig.3.6. As can be seen poor conditioned H matrix and insufficient accuracy in the estimates of innovation correlations lead to and ill-conditioned least square problem. Particularly, the estimates of K and noise covariance matrices from the ill-conditioned least square problem in Case II are simply wrong. The H matrix in Case II is more poorly conditioned than the H matrix in Case I. That is due to the fact that the number of unknowns in Case II much more than the Case I. The Tikhonov's regularization with enforcing positive semi-definitiveness of the Q and R on the solution is applied in accordance with section 3.4. Regularization parameter is calculated $\lambda = 0.00028$ using L-curve approach. An illustration of L-curve from one simulation is presented in Fig.3.7.

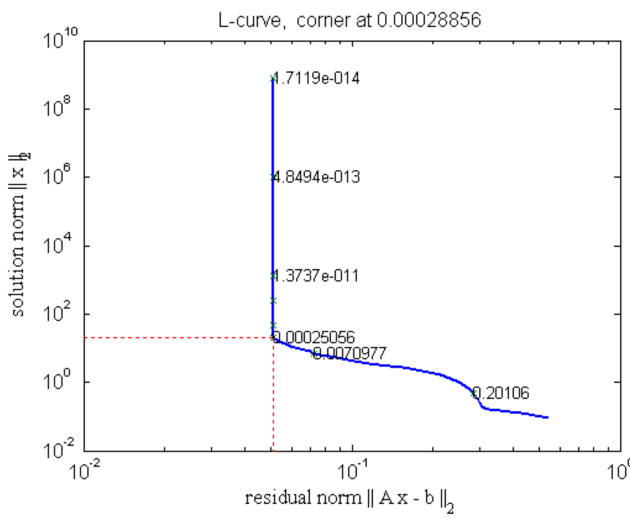

Figure 3.7: The L-curve for five-DOF spring mass system (Case II).

Histograms of Q and R estimates from 200 simulations for Case I are depicted in Fig.3.8. As can be seen, the estimates of R and 2^{nd}, 3^{rd}, and 5^{th} diagonals of Q are quite successful with ratio of $\sigma/\mu < 0.1$. The estimates of 1^{st} and 4^{th} diagonals of Q are poor, they have values of 0.53 and 0.61 for ratio of σ/μ, respectively.

We present the poles of the estimated Kalman filter calculated from indirect noise covariance and direct filter gain approaches in Fig.3.9. As can be seen, the estimated filter poles are very close to correct values.

We check optimality of the estimated Kalman gain by comparing theoretical covariance of innovations process, F, and with its experimental estimate. Theoretical values of

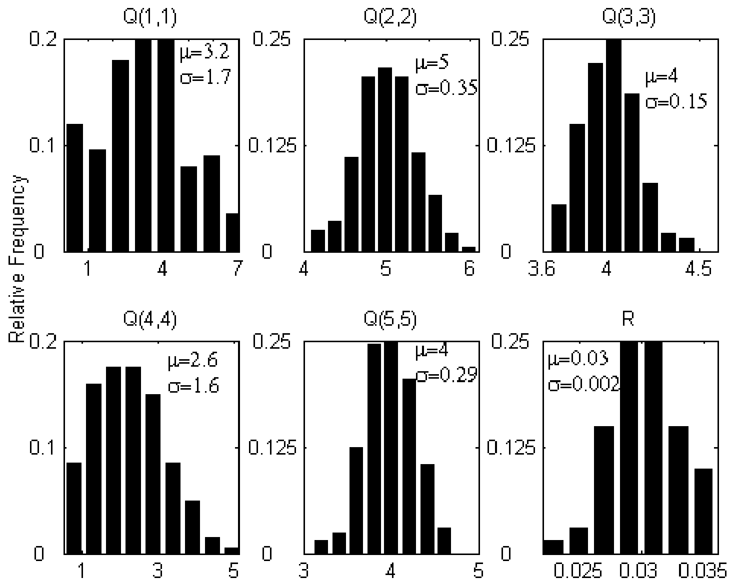

Figure 3.8: Histograms of Q and R estimates from 200 simulations for Case I in numerical testing using five-DOF spring mass system.

covariance of the optimal innovations are 0.33 and 0.55 in Case 1 and in Case 2, respectively. Histograms of innovations covariance estimates from 200 simulations are depicted in Fig.3.10. As can be seen variances of the innovations covariances estimates from 200 simulations are very small and the mean values are identical with the theoretical values. This results shows that the estimated K are nearly optimal.

The estimation of K from measurements is successfully exemplified on a five-DOF spring-mass model, which demonstrated that the Tikhonov's regularization is a useful tool in order to obtain the estimates of K from finite data. However the challenges in structural engineering applications remains to be checked since the size of the model can be an issue, which is examined in the following numerical experiment using a truss structure.

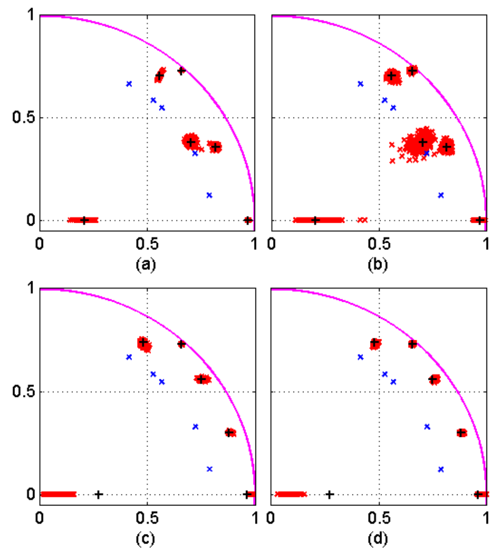

Figure 3.9: Five-DOF spring mass system estimated filter poles, a) Case I - Indirect Approach b) Case I - Direct Approach c) Case II - Indirect Approach d) Case II - Direct Approach, (blue: Estimated gain poles, Red: Initial gain poles, Black: Optimal gain poles).

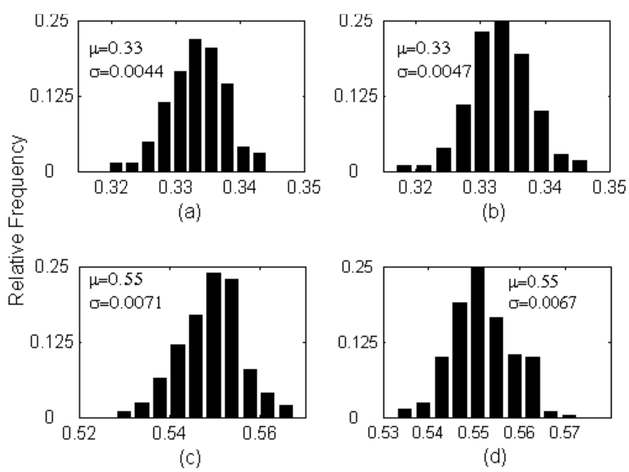

Figure 3.10: Histograms of innovations covariance estimates from 200 simulations in numerical testing using five-DOF spring mass system, First Row: Case I, Second Row: Case II, a-c) Indirect noise covariance approach, b-d) Direct Kalman gain approach

3.5.2 Experiment 2: Planar Truss Structure

This simulation experiment demonstrates the application of innovations correlations approaches to estimate steady state Kalman gain for a truss structure. The planar truss structure considered is depicted in Fig.3.11. It has 44 bars and a total of 39 DOF. All the bars are made of steel (with E = 200 GPa) with an area of 64.5 cm^2. Mass of 1.75*10^5 kg at each coordinate. Damping is 2% in all modes. The first five un-damped natural frequencies (in Hz) are {0.649, 1.202, 1.554, 2.454, 3.301} and the largest one is 16.584 Hz. The system is statically indeterminate both externally and internally.

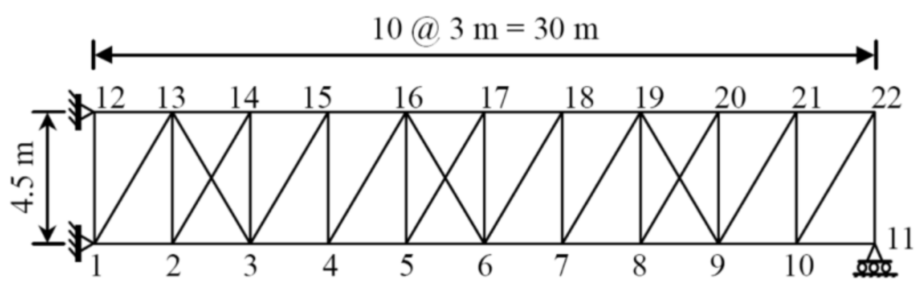

Figure 3.11: Truss structure utilized in the numerical testing of correlations approaches.

Three sensors recording motions in the vertical and horizontal directions are located at the joints of {3, 6, 9}, which are measuring the velocity data at 50Hz sampling. The unmeasured stationary and mutually correlated excitations are assumed to act at all joints on truss. The measurement noise in each simulation is prescribed to have RMS equal to approximately 10% of the RMS of the response measured. Unmeasured excitations and measurement noise are assumed to be mutually uncorrelated, namely $S = 0$.

The number of unknown parameters in Q is bigger than mxn, namely

$$mxn = 6x78 = 468 < 780$$

Therefore the H matrix is rank deficient and solution for noise covariance matrices is not unique. Although a unique solution does not exist in this case, the Q and R estimates are used to calculate the K from classical formulations of the KF.

Innovations process obtained from an arbitrary gain K_0 that is chosen such that eigenvalues of the matrix $A - K_0C$ are assumed to have the same phase as those of A but with a 15% smaller radius. 150 lags of correlation functions are considered. The sample innovation correlations functions are calculated using 600 seconds of data. In this case the condition number of H matrix in Eq.3.2.22 is calculated as $1.45x10^{21}$. Stability of the least squares solution is examined using Discrete Picard Condition, which is depicted in Fig.3.12. As can be seen poor conditioned H matrix and insufficient accuracy in the estimates of innovation correlations lead to an ill-conditioned least square problem, in which the solution is blowing up due to contribution of singular values that are close to zero.

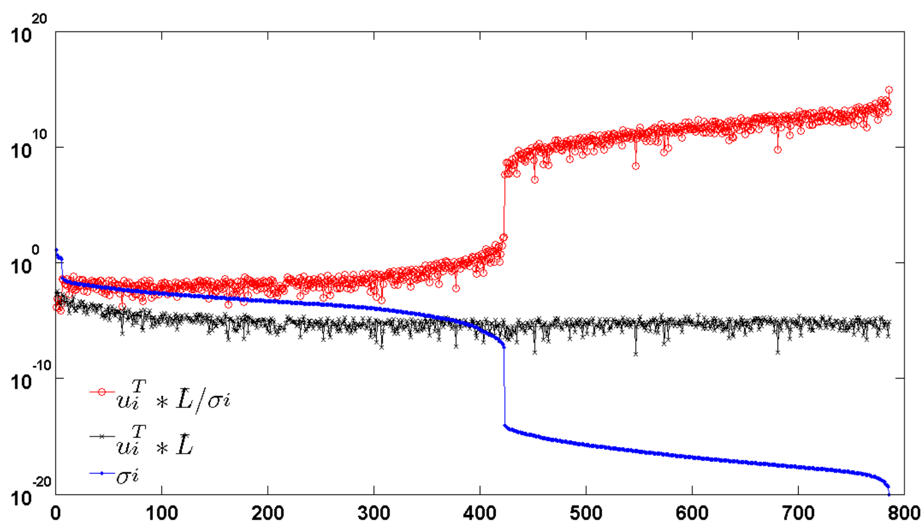

Figure 3.12: Discrete Picard Condition for truss structure.

The Tikhonov's regularization with enforcing positive semi-definitiveness of the Q and R on the solution is applied in accordance with section 3.4. Regularization parameter is calculated as $\lambda = 0.0011$ using L-curve approach. An illustration of L-curve from one simulation is depicted in 3.13.

The estimated filter poles from direct and indirect innovations correlations approaches with Tikhonov's regularization are depicted in Fig.3.14. As can be the seen the filter poles are estimated with a good approximation with the use of Tikhonov's regularization although the size of the problem is relatively large with 468 unknowns parameter of in the K matrix. Indirect noise covariance approach has better performance compared to direct Kalman gain approach in this experiment. This can be due to the fact that the indirect noise covariance approach allows enforcing positive semi-definitiveness of the noise covariances while this is not the case in the direct Kalman gain approach.

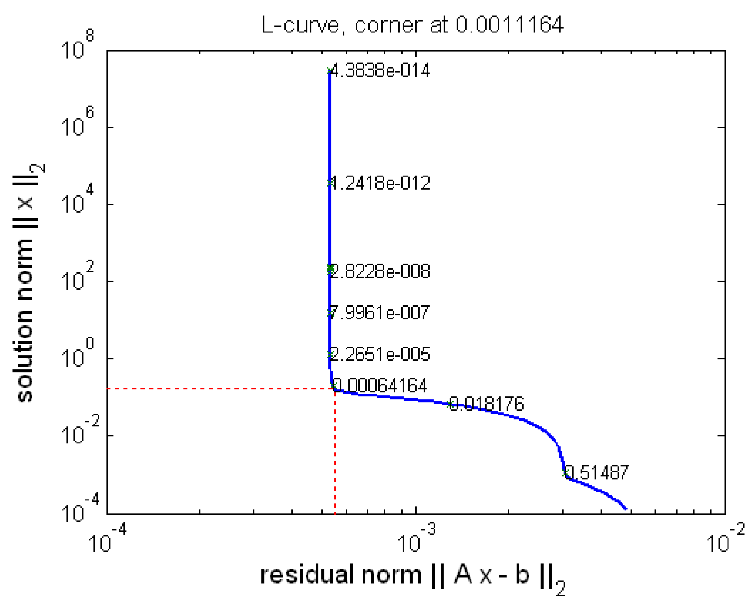

Figure 3.13: The L-curve for truss structure.

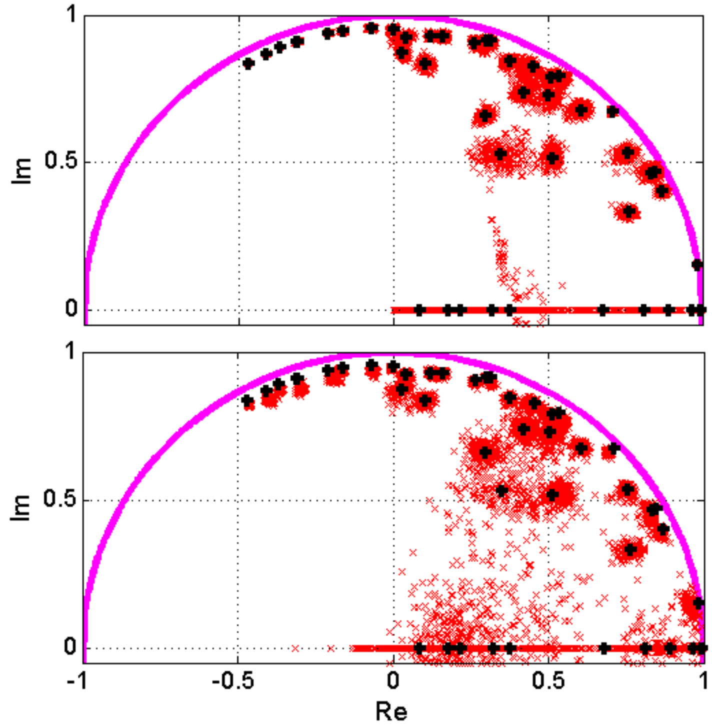

Figure 3.14: Truss structure, estimates of filter poles for 200 simulations; Top: Indirect noise covariance approach, Bottom: Direct Kalman gain Approach (Red: Estimated gain poles, Blue: Optimal gain poles)

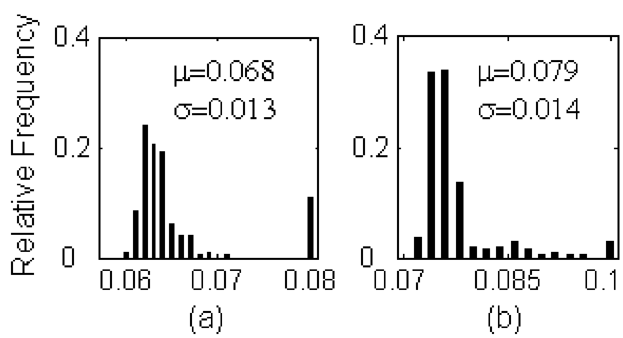

Figure 3.15: Histograms of trace of state error covariance estimates from 200 simulations - truss structure, a) Indirect noise covariance approach b) Direct Kalman gain approach

The performance of the estimated filter gain is evaluated using experimental state error covariance for each simulation, which is calculated from,

$$\hat{P} = \frac{1}{N}\sum_{k=1}^{N}\left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T\right] \tag{3.5.1}$$

where x_k is correct state, $\hat{x}_k$ is state estimate obtained from calculated Kalman gain, N is the number of time steps. Theoretical value of trace of state error covariance P for optimal gain is 0.060. Histograms of trace of experimental state error covariance from 200 simulations are presented in Fig.3.15. As can be seen indirect noise covariance approach performs better compared to direct Kalman gain approach, which is a result in line with Fig.3.14. Mean value of the trace of error covariance estimates from both methods, is larger than 0.060 therefore we conclude that filter gain estimates from correlations approaches are suboptimal in this numerical experiment.

3.6 Summary

This chapter studies estimation of steady state Kalman gain K for time invariant stochastic systems. The operating assumptions are that the system is linear and subjected to unmeasured Gaussian stationary disturbances and measurement noise, which are (in general) correlated. In classical Kalman filter theory, the noise covariance matrices (Q, R and S) are assumed known. Here, we assumed that system matrices (A, C) are known without model error however Q, R and S are not known. The chapter presented a complete description of the classical correlations approaches to estimate the K as well as Q, R and S . The correlations approaches examined use the measurements obtained from a data collection session, so the results are restricted to problems where the estimation is done off-line. The procedures has to be carried out off-line, but in many applications in structural engineering this is not an issue. There are two strategies to calculate K from correlations of measurements or innovations of an arbitrary filter:

1) Indirect noise covariance approach. 2) Direct Kalman gain approach. The direct approach identifies the K directly from measured data. The indirect noise covariance approach estimates the Q, R and S first, and then use them to calculate K from classical Kalman filter formulations.

In theory, K and corresponding covariance of innovations, F can be computed from measurements or innovations of an arbitrary filter because correlation functions of measurements or the innovations sequence can be related to K and F. However, state error covariance matrix, P corresponding to optimal Kalman gain cannot be computed without the information of noise covariance matrices. In theory, Q, R and S can be computed from measurements or innovations of an arbitrary filter because correlations functions of measurements or the innovations of any arbitrary filter can be related to the Q, R and S linearly. In an off-line setting, the estimation of K and noise covariance matrices lead to a problem of the form;

$$HX = L \tag{3.6.1}$$

where L are the correlation functions of measurements or innovations from an arbitrary filter and X contains the entries in noise covariance matrices as unknowns. H is calculated using system matrices and arbitrary gain K_0, and is known without any error. From the results presented in the previous sections, we can identify the following conclusions:

- Computing noise covariance matrices from Eq.3.6.1 may have infinite solutions. A unique solution for noise covariance matrices exists only if the number of unknown parameters in Q and S is smaller than the product of the number of measurements and the number of state. However, when uniqueness condition of solving for the noise covariance matrices is not satisfied, the optimal Kalman gain, K and covariance of the optimal innovations, F can still be computed from any of the solutions for noise covariance matrices. Note that, in this case although any of the solution for noise covariance matrices is resulting to correct K and F , the resulting covariance of state error, P is not the correct one, and it cannot be calculated without getting the unique solution for noise covariance matrices.

- The innovations correlations approach leads to expressions that are more complex than the output correlations scheme, but the differences are not important when it comes down to computer implementation. Since the innovations are less correlated than the output, the innovations approach is more efficient and gives more accurate estimates with short data compare to output correlations approaches.

- The expressions in the innovations correlations approach allows to enforce the positive semi-definiteness when solving for Q, R and the S. However in output correlations approach, the unknown vector of least square problem involves only the unknowns of Q and S. R has to be calculated from another equation which requires the Q and S estimates. Therefore enforcing the positive semi-definitiveness of the solution is not possible in the output correlations approach.

- In general, the least square problem of estimating the K and noise covariance matrices from correlation approaches has an ill-conditioned coefficient matrix. The examinations show that, in the indirect noise covariance approaches, the condition number of the coefficient matrix increases with an increase in the number of unknown parameters of noise covariance matrices.

- In real applications, the right hand side of the Eq.3.6.1 has some uncertainty since it is constructed from sample correlation functions of innovations process calculated using finite data. The accuracy of the sample correlation functions of innovations process is improved by using long data, however, due to fact that coeffecient matrix is ill-conditioned, the stability of the solution being sensitive to the errors in correlations functions is examined using Discrete Picard Condition (DPC). Numerical examinations show that the correlations approaches do not satisfy the DPC, therefore, the estimates obtained from the classical least square solution are simply wrong. In this study we examined the merit of using Tikhonov's regularization to approach the ill-conditioned problems of correlations approaches. Numerical examinations show that the estimates can be significantly improved by applying Tikhonov's regularization to ill-conditioned problems of correlations approaches. This is shown for a simulated five DOF spring mass system and a truss structure.

Given the results of the examinations, we conclude that the direct Kalman gain approach examined in this study is recursive and convergence of the solution is heavily related to accuracy in the sample correlation functions of innovations and it is not guaranteed. We recommend the use of indirect noise covariance approach to estimate the steady state Kalman gain from measurements. To apply the indirect noise covariance approach to estimate the steady state Kalman gain from available measurements, the reader can follow the instructions below:

1. Construct the coefficient matrix H in Eq.3.2.22 for given set of (A, C, G) and choosing an arbitrary stable gain K_0.

2. Use available measurements and obtain the innovations process of the filter with gain K_0.

3. Construct the L matrix in Eq.3.2.22 using sample correlations of the innovations process that are calculated from Eq.3.1.3.

4. Check the stability of the H being sensitive to the errors in L using DPC.

5. If #4 shows that the problem is ill-conditioned, calculate regularization parameter from L-shape approach and apply Tikhonov's regularization in Eq.3.4.15 to obtain estimates of noise covariance matrices.

6. Check the noise covariance estimates from #5. If they are not positive semi-definitive matrices, enforce the positive semi-definitiveness using optimization algorithms.

7. Use the classical Kalman filter formulations presented in Chapter 2 and calculate the K using system parameters (A, C, G) and noise covariance estimates obtained from #6.

Chapter 4

Kalman Filter with Model Uncertainty

4.1 Background and Motivation

In the classical Kalman filter theory, one of the key assumptions is that a priori knowledge of the system model, which represents the actual system, is known without uncertainty. In reality, due to the complexity in the systems, it is often impractical (and sometimes impossible) to model them exactly. Therefore, there is considerable uncertainty about the system model and the error-free model assumption of classical Kalman filtering is not realistic in applications. Methods for addressing the Kalman filtering with model uncertainty can be classified into two groups: (1) Robust Kalman Filtering (RKF), (2) Adaptive Kalman Filtering.

The key idea in RKF is to design a filter such that a range of model parameters are taken into account. In this case, the filter gain is calculated by minimizing a bound on the trace of the state error covariance, not the trace itself. One of the fundamental contribution on the RKF is the work by Xie, Soh and Souza, who considered to design of Kalman filter for linear discrete-time systems with norm-bounded parameter uncertainty in the state and output matrices, [37]. They calculated the filter gain on the premise that the covariance of the state estimation error is guaranteed to be within a certain bound for all admissible parameter uncertainties. They showed that a steady-state solution to the robust Kalman filtering problem is related to two algebraic Riccati equations. The formulation of RKF is computationally intensive and solving two Riccati equations in systems of large model size may be impracticable. We refer to work by Petersen and Savkin [38] for a general treatment and a review of RKF algorithms.

The adaptive Kalman filtering can be categorized into two approaches. One is simultaneous estimation of the parameters and the state, which is applicable in two ways: (1) The bootstrap approach, (2) The combined state and parameter estimation approach. In the bootstrap approach, the estimation is carried out in two steps. In the first step the states are estimated with the assumed nominal values of the parameters. In the second step the parameters are calculated using the recent estimates of the state from step one in addition to measurements, [39, 40]. Probably, the first bootstrap solution for parameter and state estimation problem is proposed by Cox [41] who obtained the estimates based on the maximization of the likelihood function of the measurement constrained by the nominal model of the system. El Sherief and Sinha [42] have also proposed an other bootstrap method to obtain estimates of the parameters of an Kalman filter model as well as the state.

In the combined state estimation approach, the unknown parameters are augmented to the state vector for their online identification. This idea was initially introduced by Kopp and Orford [43], who derived a recursive relationship for the updated estimates of the parameters and state as a function of measurement. Since the problem posed as nonlinear, nonlinear filtering techniques such as particle filter, extended Kalman filter

(EKF) and unscented Kalman filter (UKF) are used to obtain the combined estimates of parameters and [44, 45, 46]. In literature the problem is called with various names such as dual estimation [47, 48], combined state estimation [39, 49, 50], augmented state estimation [51, 52] and joint state estimation, [53]. This chapter examines the use of EKF for on-line state and parameter estimation. A fundamental contribution on the theory of the EKF as a parameter estimator for linear systems is the work of Ljung [54], who presented asymptotic behavior of the filter. Panuska extended the work to the systems which are subjected to the correlated noise and presented another form of the filter, where the state consists only of the parameters to be estimated [55, 56]. Recently, Wu and Smyth [57] are compared the performance of the EKF as an parameter estimator against that of the UKF.

The other approach in adaptive Kalman filtering, instead of estimating the uncertain parameters themselves, includes the effect of the uncertain parameters in state estimation, [19, 58]. In this approach, the model errors are approximated by fictitious noise and the covariance of the noise is tuned based on an analytical criteria. To the best of the writer's knowledge, this idea is first applied by Jazwinski [59] who determined the covariance fictitious noise so as to produce consistency between Kalman filter innovations and their statistics.

The objective in this chapter is to address the uncertainty issue in model that is used in Kalman filtering. We examine the feasibility and merit of an approach that takes the effects of the uncertain parameters of the nominal model into account in state estimation. In this approach, the system is approximated with a stochastic model and the problem is addressed in off-line conditions. The model errors are approximated by fictitious noise and the covariance of the fictitious noise is calculated on the premise that the norm of discrepancy between covariance functions of measurements and their estimates from the nominal model is minimum. Additionally, the problem in considered in on-line operating conditions, and the EKF-based combined parameter and state estimation method is examined.

4.2 Stochastic Modeling of Uncertainty

In this section, for the situation where the uncertainty in the state estimate, in addition to the disturbances, derives from error in the matrices of the state space model. Specifically, we consider the situation given by

$$x_{k+1} = (A_n + \triangle A)x_k + (G_n + \triangle G)w_k \tag{4.2.1}$$

$$y_k = Cx_k + v_k \tag{4.2.2}$$

where A_n and G_n are nominal model matrices; $\triangle A$ and $\triangle G$ error matrices. Suppose that the noise covariance and error matrices are unknown. The objective is to obtain an estimate of the state x_k using the information of nominal model matrices and stored data of measurement sequence y_k.

4.2.1 Fictitious Noise Approach

An approximation of the state sequence of the system in Eq.4.2.1-4.2.2 is obtained from an equivalent stochastic model, namely

$$\bar{x}_{k+1} = A_n\bar{x}_k + \bar{w}_k \tag{4.2.3}$$

$$\bar{y}_k = C\bar{x}_k + \bar{v}_k \tag{4.2.4}$$

Suppose that $\bar{w}_k$ are $\bar{v}_k$ are white noise sequences, with covariance matrices $\bar{Q}$ and $\bar{R}$, respectively The equivalent disturbance $\bar{w}_k$ obtained by comparing Eqs.4.2.1 and 4.2.3 is

$$\bar{w}_k = \triangle A\bar{x}_k + Gw_k \tag{4.2.5}$$

If the $\bar{Q}$ and $\bar{R}$ are known, then the KF can be applied to the equivalent stochastic model in Eqs.4.2.3-4.2.4 to obtain an estimate of the state. Since the actual system and equivalent model are stochastic systems, the outputs y_k and $\bar{y}_k$ can be characterized with the covariance functions. The main idea explored here is that the covariance of $\bar{w}_k$ and $\bar{v}_k$ are calculated on the premise that the norm of discrepancy between correlation functions of y_k and $\bar{y}_k$ is minimum, namely minimizing the cost function

$$J = \|corr(y) - corr(\bar{y})\|$$

The solution of a similar problem is presented in Chapter 3, in which the noise covariance matrices of a model error free stochastic system are calculated. The fundamental steps of the solution involves: (1) Theoretical correlation functions of $\bar{y}_k$ is derived as a linear function of A_n, C and noise covariance matrices $\bar{Q}$ and $\bar{R}$. (2) Using available stored data, an estimate of the correlations function of y_k, which we denote as Λ_j, is calculated from

$$\Lambda_j \overset{def}{=} E(y_k y_{k-j}^T) = \frac{1}{N-j}\sum_{k=1}^{N-j} y_k y_{k-j}^T \tag{4.2.6}$$

where N is the number of time steps. (3) A linear least square problem is formed considering a number of lags of correlations and solved for equivalent noise covariance matrices. Since the state sequence x_k is not a white process and the white noise approximation of $\bar{w}_k$ in Eq.4.2.5 is theoretically not correct. However, since our aim is to obtain an estimate of the state, we examine the merit of using some noise covariance matrices that make the output correlations of the actual system and equivalent model approximately equal. The correlations is a function lag, therefore, the solution will be dependent on the number of lags considered. Given the fact output correlations of a stochastic system approach to zero as seen in Eq.2.3.135, a solution that gives a better approximation of output correlations of the actual system requires taking a range of lags starting from zero.

Since the $\bar{w}_k$ and $\bar{v}_k$ are fictitious, the uniqueness of the solution of the least square problem is not a concern, as long as the positive definitiveness of the covariance matrices are provided. Therefore, the information of covariance matrices of w_k and v_k does not apply any condition to the solution. For instance, one can force the equivalent disturbances $\bar{w}_k$ and measurement noise $\bar{v}_k$ noise to be mutually correlated, namely, $\bar{S} \neq 0$, although the actual system has mutually uncorrelated w_k and v_k.

As noted in Chapter 3, the drawback of output correlations approach is that the calculations of the $\bar{Q}$ and $\bar{R}$ matrices are performed in two steps and it does not allow to force positive definitiveness of the solution for these matrices. Moreover, the output correlations approach requires very long data to obtain accurate estimates of noise covariance matrices since the measurements are generally overly correlated.

Another approach to this problem uses the correlations of innovations process. In this approach, the available measurements are filtered with an arbitrary gain and the correlations of resulting innovations are used. Suppose that the measurements $y_k$is filtered thorough an arbitrary filter, in which we denote the gain as K_0, namely

$$\hat{x}_{k+1} = (A_n - K_0 C)\hat{x}_k + K_0 y_k \tag{4.2.7}$$

$$e_k = y_k - C\hat{x}_k \tag{4.2.8}$$

and $\bar{y}_k$ is filtered with the same filter, namely

$$\hat{x}_{k+1} = (A_n - K_0 C)\hat{x}_k + K_0 \bar{y}_k \tag{4.2.9}$$

$$\bar{e}_k = \bar{y}_k - C\hat{x}_k \tag{4.2.10}$$

In this case, the covariance of $\bar{w}_k$ and $\bar{v}_k$ are calculated on the premise that the norm of discrepancy between correlation functions of e_k and $\bar{e}_k$ is minimum, namely minimizing the cost function

$$J = \|corr(e) - corr(\bar{e})\|$$

The solution involves three fundamental steps similar to the output correlations: (1) Theoretical correlation functions of $\bar{e}_k$ is derived as a linear function of A_n, C, K_0 and noise covariance matrices $\bar{Q}$ and $\bar{R}$. (2) An estimate of the correlations function of e_k, which we denote as $\mathcal{L}_j$, is calculated from

$$\mathcal{L}_j \overset{def}{=} E(e_k e_{k-j}^T) = \frac{1}{N-j}\sum_{k=1}^{N-j} e_k e_{k-j}^T \tag{4.2.11}$$

where N is the number of time steps. (3) A linear least square problem is formed considering a number of lags of innovations correlations and solved for equivalent noise covariance matrices. Since the innovations are less correlated, the innovations approach is more efficient and gives more accurate estimates with short data compare to output correlations approaches. The innovations correlations approach allows to enforce the positive semi-definiteness when solving for $\bar{Q}$, $\bar{R}$ and $\bar{S}$. The reader is referred to the section 3.2 for a detailed review of innovations and output correlations approaches to calculate noise covariance matrices.

4.2.2 Equivalent Kalman Filter Approach

An approximation of the state sequence of the system in Eq.4.2.1-4.2.2 can be calculated using a Kalman filter model that is constructed from the nominal model and available measurements. Suppose that operating condition is off-line and consider an output form Kalman filter is given, namely

$$\hat{x}_{k+1} = (A_n - KC)\hat{x}_k + Ky_k \tag{4.2.12}$$

$$\hat{y}_k = C\hat{x}_k \tag{4.2.13}$$

where $\hat{y}_k$ is the measurement predictions of the filter. The main idea here, which is initially described by Juang, Chen and Phan [60], is to calculate the filter gain K by minimizing the norm of the discrepancy between available measurement y_k and its estimate $\hat{y}_k$ from the filter model is minimum, namely, minimizing the cost function

$$J = \|y - \hat{y}\|$$

Assuming the initial state is zero and $(A_n - KC)$ is asymptotically stable, one can write auto-regressive (AR) model of the output form Kalman filter model as follows,

$$\hat{y}_k = \sum_{j=0}^{p} Y_j y_{k-j} \tag{4.2.14}$$

where Y_j is the markov parameters of the output form Kalman filter, namely

$$Y_j = C\bar{A}_n^{j-1}K \tag{4.2.15}$$

where

$$\bar{A}_n = A_n - KC \tag{4.2.16}$$

It is assumed that for a sufficiently large value of p in Eq.4.2.14,

$$\bar{A}_n^k \approx 0 \qquad k > p \tag{4.2.17}$$

To obtain markov parameters of AR model from observations one can write Eq.4.2.14 in matrix form for a given set of data as follows,

$$e = y - YV \tag{4.2.18}$$

where

$$V = \begin{bmatrix} y_0 & y_1 & \cdots & y_{n-1} \\ 0 & y_0 & \cdots & y_{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & y_{n-p-2} \end{bmatrix} \tag{4.2.19}$$

and

$$Y = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_p \end{bmatrix}, \ y = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_n^T \end{bmatrix}^T, \quad e = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_n^T \end{bmatrix}^T \tag{4.2.20}$$

n is the data length. Assuming the innovations are minimal and uncorrelated Y can be computed by least square solution as follows,

$$Y = yV^{\dagger} \tag{4.2.21}$$

where † denotes the pseudo inverse. The Y matrix contains the markov parameters of moving average (MA) model of the filter, namely

$$Y_k^0 = CA_n^{j-1}K \tag{4.2.22}$$

One can obtain these parameters from the following recursion equation

$$Y_k^0 = Y_k + \sum_{j=0}^{k-1} Y_{k-j}^0 Y_j \qquad k = 1,\ 2\,,3 \cdots p \tag{4.2.23}$$

Finally, the filter gain can be solved from the markov parameters of MA model of the filter as follows,

$$K = \mathcal{O}^{\dagger} Y^0 \tag{4.2.24}$$

where † denotes the pseudo inverse and $\mathcal{O}$ is the observability block of A_n and C, Y^0 is the matrix formed by Y_k^0's, namely,

$$Y^0 = \begin{bmatrix} Y_1^0 \\ Y_2^0 \\ \vdots \\ Y_p^0 \end{bmatrix} = \begin{bmatrix} CK \\ CA_nK \\ \vdots \\ CA_n^{p-1}K \end{bmatrix} \tag{4.2.25}$$

and

$$\mathcal{O} = \begin{bmatrix} C \\ CA_n \\ \vdots \\ CA_n^{p-1} \end{bmatrix} \tag{4.2.26}$$

The algorithm to estimate the filter gain can be summarized in three steps:

1. Obtain markov parameters of the AR model, solving the least square problem in Eq.4.2.21.

2. Obtain markov parameters of the MA model, from Eq.4.2.23.

3. Obtain the filter gain, K from Eq.4.2.24.

4.3 Combined State and Parameter Estimation

In this section we outline the extended Kalman filter approach to the parameter estimation problem in the case where the system is linear and the non-linearity arises from the augmentation of the state vector with unknown parameters. Let the system be described by

$$\dot{x}(t) = A_c(\theta)x(t) + B_c(\theta)u(t) + G_c w(t) \tag{4.3.1}$$

with notations defined in section 2.4. θ is finite dimensional parameter vector which denotes unknown parameters of the system. The available measurements has the following description in sampled time:

$$y_k = Cx_k + v_k \tag{4.3.2}$$

The $w(t)$ and v_k are uncorrelated Gaussian stationary white noise sequences with zero mean and covariance of Q_c and R respectively, with notations defined in section 2.5. Additionally, it is also assumed that $w(t)$ and v_k are independent of θ. One begins by augmenting the state with the parameter vector, namely

$$z(t) \stackrel{def}{=} \begin{bmatrix} x(t) \\ \theta(t) \end{bmatrix} \tag{4.3.3}$$

We suppose that the parameters are constant, namely

$$\dot{\theta}(t) = 0 \tag{4.3.4}$$

The second step involves comprising a new state space model for the augmented state by combining Eqs.4.3.1 and 4.3.4, namely

$$\dot{z}(t) = \bar{A}(\theta)z(t) + \bar{B}(\theta)u(t) + \bar{G}\bar{w}(t) \tag{4.3.5}$$

$$y_k = \bar{C}z_k + v_k \tag{4.3.6}$$

where $\bar{w}(t)$ is the process noise of the combined model described by

$$\bar{w}(t) = \begin{bmatrix} w(t) \\ w_p(t) \end{bmatrix} \tag{4.3.7}$$

$w_p(t)$ is a pseudo noise with a covariance q introduced to drive the filter to change the estimate of θ. Augmented state model system matrices are formed as

$$\bar{A} = \begin{bmatrix} A(\theta) & 0 \\ 0 & 0 \end{bmatrix} \tag{4.3.8}$$

$$\bar{B} = \begin{bmatrix} B^T(\theta) & 0 \end{bmatrix}^T \tag{4.3.9}$$

$$\bar{C} = \begin{bmatrix} C & 0 \end{bmatrix} \tag{4.3.10}$$

$$\bar{G} = \begin{bmatrix} G & 0 \\ 0 & I \end{bmatrix} \tag{4.3.11}$$

The augmented state space model in Eq.4.3.5 has the unknown parameters as additional state of the system. It's important to note that due to the coupling of the state with the parameters, the estimation problem becomes nonlinear although the system given in Eq.4.3.1 is linear. Consequently, nonlinear techniques have to be used to perform state estimation for this model, where we utilize from EKF. The last step of the combined state and parameter estimation involves formulating the EKF for the nonlinear model in Eq.4.3.5 in accordance with the section 2.5. The prediction and update steps of the EKF are presented in the following.

Prediction Step:

Since the disturbances are not known the derivative of the augmented state is obtained as

$$\dot{\hat{z}}(t) = \bar{A}(\theta)\hat{z}(t) + \bar{B}(\theta)u(t) \tag{4.3.12}$$

The a priori state error covariance of the augmented model is described by

$$\bar{P}(t) = \begin{bmatrix} P_x(t) & [0] \\ [0] & P_\theta(t) \end{bmatrix} \tag{4.3.13}$$

where

$$P_x(t) = E\left[(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T\right] \tag{4.3.14}$$

$$P_\theta(t) = E\left[(\theta - \hat{\theta}(t))(\theta - \hat{\theta}(t))^T\right] \tag{4.3.15}$$

and $\bar{P}(t)$ satisfies

$$\dot{\bar{P}}(t) = \Delta(t)\bar{P}(t) + \bar{P}(t)\Delta^T(t) + \bar{G}\bar{Q}\bar{G}^T \tag{4.3.16}$$

where

$$\bar{Q} = \begin{bmatrix} Q & [0] \\ [0] & q \end{bmatrix} \tag{4.3.17}$$

$\Delta(t)$ is the Jacobian of the nonlinear model around current state $\hat{z}(t)$, which is calculated from

$$\Delta(t) = \left.\frac{\partial \dot{z}(t)}{\partial z}\right|_{z=\hat{z}(t)} = \begin{bmatrix} A(\hat{\theta}(t)) & D(t) \\ 0 & 0 \end{bmatrix} \tag{4.3.18}$$

$$D(t) = \left.\frac{\partial A(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} \widehat{x}(t) + \left.\frac{\partial B(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} u(t) \tag{4.3.19}$$

Integrating Eqs.4.3.12 and 4.3.16 numerically, the solution for state estimate and state error covariance are advanced one time step to obtain $\hat{z}(t+1) = \hat{z}^-_{k+1}$ and $\bar{P}(t+1) = \bar{P}^-_{k+1}$, respectively.

Update Step:

Upon arrival of the measurement the posterior estimate of the state is computed from

$$\hat{z}^+_{k+1} = \hat{z}^-_{k+1} + K_{k+1}(y_{k+1} - \bar{C}\hat{z}^-_{k+1}) \tag{4.3.20}$$

The Kalman gain K_{k+1} and the a posteriori error covariance $\bar{P}^+_{k+1}$ are calculated from

$$K_{k+1} = \bar{P}^-_{k+1}\bar{C}^T(\bar{C}\bar{P}^-_{k+1}\bar{C}^T + R)^{-1} \tag{4.3.21}$$

$$\bar{P}^+_{k+1} = (I - K_{k+1}\bar{C})\bar{P}^-_{k+1}(I - K_{k+1}\bar{C})^T + K_{k+1}RK^T_{k+1} \tag{4.3.22}$$

The filter is initialized by using initial state estimate and state error covariance, namely

$$\hat{z}_0 = \begin{bmatrix} E\left[\hat{x}_0\right] \\ E\left[\hat{\theta}_0\right] \end{bmatrix} \tag{4.3.23}$$

$$\bar{P}_0 = \begin{bmatrix} P_{x_0}(t) & [0] \\ [0] & P_{\theta_0}(t) \end{bmatrix} \tag{4.3.24}$$

Convergence of the augmented filter model requires

$$\left.\frac{\partial K_x(\theta)}{\partial \theta}\right|_{x=\hat{x},\,\theta=\hat{\theta}} \neq 0 \tag{4.3.25}$$

where K_x is the partition of the Kalman gain, corresponding to the un-augmented state, namely,

$$K_k \stackrel{def}{=} \begin{bmatrix} K_x \\ K_\theta \end{bmatrix}$$

The lack of coupling between K_x and θ in the filter may lead to divergence of the estimates, [54]. An illustration that involves the steps of the EKF-based parameter estimation algorithm is presented in the following example.

Example:

The following example presents the EKF applied to parameter estimation. Consider an un-damped two-degree-of-freedom shear frame structure whose story stiffnesses and story masses are given in consistent units as $\{100,\ 100\}$ and $\{1.0,\ 1.0\}$, respectively. The un-damped frequencies of the structure are $\{0.98,\ 2.57\}$ in Hz. The unmeasured excitation is assumed to act at the 1st floor, which has a unit variance ($Q_c = 1$) in discrete data sampled at 100Hz. There is no deterministic excitation acting on structure, namely

$u(t) = 0$. We obtain results for output sensor at the second floor, which is recording displacement data at 100Hz sampling. The measurement noise is assumed to have a standard deviation that is equal 10% standard deviation of the response; the variance is calculated as $R = 1.25x10^{-5}$. The the state $x(t)$ and measurements y_k satisfy the following state space model given by

$$\dot{x}(t) = A(\theta)x(t) + Gw(t) \tag{4.3.26}$$

$$y_k = Cx_k + v_k \tag{4.3.27}$$

where

$$A(\theta) = \begin{bmatrix} [0] & [I] \\ -M^{-1}K(\theta) & [0] \end{bmatrix} \tag{4.3.28}$$

M and $K(\theta)$ are the mass and stiffness matrices of the structure, respectively, which are obtained from

$$K = \begin{bmatrix} \theta + k_1 & -k_1 \\ -k_1 & -k_1 \end{bmatrix} \tag{4.3.29}$$

$$M = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \tag{4.3.30}$$

Input to state matrix G and state to output matrix C are obtained from,

$$G = \begin{bmatrix} 0 & 0 & 1/m_1 & 0 \end{bmatrix}^T \tag{4.3.31}$$

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix} \tag{4.3.32}$$

It's apparent from Eq.4.3.29 that k_1 denotes the stiffness of the first floor and the θ denotes the stiffness of the second floor, namely

$$\theta \overset{def}{=} [k_2] \tag{4.3.33}$$

We suppose that the k_2 is constant, namely

$$\dot{\theta}(t) = 0 \tag{4.3.34}$$

We form a new state space model by combining Eqs.4.3.26 and 4.3.34 as follows

$$\dot{z}(t) = \bar{A}(\theta)z(t) + \bar{G}\bar{w}(t) \tag{4.3.35}$$

$$y_k = \bar{C}z_k + v_k \tag{4.3.36}$$

where $z(t)$ is the new state is obtained by augmenting the state with the parameter vector

$$z(t) \overset{def}{=} \begin{bmatrix} x(t) & \theta(t) \end{bmatrix}^T = \begin{bmatrix} x_1(t) & x_2(t) & \dot{x}_1(t) & \dot{x}_2(t) & \theta(t) \end{bmatrix}^T \tag{4.3.37}$$

The $\bar{w}(t)$ is the process noise of the combined model, which is formed by

$$\bar{w}(t) = \begin{bmatrix} w(t) \\ w_p(t) \end{bmatrix} \tag{4.3.38}$$

where $w_p(t)$ is a pseudo noise with a variance q introduced to drive the filter to change the estimate of θ. Augmented state model system matrices are formed as

$$\bar{A}(\theta) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -(\theta + k_1)/m_1 & k_1/m_1 & 0 & 0 & 0 \\ k_1/m_2 & -k_1/m_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.3.39}$$

$$\bar{C} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{4.3.40}$$

$$\bar{G} = \begin{bmatrix} 0 & 0 & 1/m_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T \tag{4.3.41}$$

The state space model in Eq.4.3.35 is nonlinear due to fact that the augmented state is coupled with system parameters. We use the EKF to estimate the state of this combined state model. We suppose that the variance of the pseudo noise q in the estimate of θ is fixed as zero, and the filter is initialized with

$$\hat{z}(0) = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}^T \tag{4.3.42}$$

$$P(0) = I \tag{4.3.43}$$

The prediction step of the EKF is adapted by using the Eqs.2.5.5 and 2.5.8, where the

Jacobian of the combined state model $\Delta(t)$ around $\hat{z}(t)$ is

$$\Delta(t) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -(\hat{\theta}+k_1)/m_1 & k_1/m_1 & 0 & 0 & -\hat{x}_1/m_1 \\ k_1/m_2 & -k_1/m_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.3.44}$$

The $\bar{A}(\theta)$ and $\Delta(t)$ are updated using current estimate of the parameter θ at every time step before prediction step calculations are performed. The noise covariance matrix that is used for calculation of state error covariance is given by

$$\bar{Q} = \begin{bmatrix} Q_c & [0] \\ [0] & q \end{bmatrix} \tag{4.3.45}$$

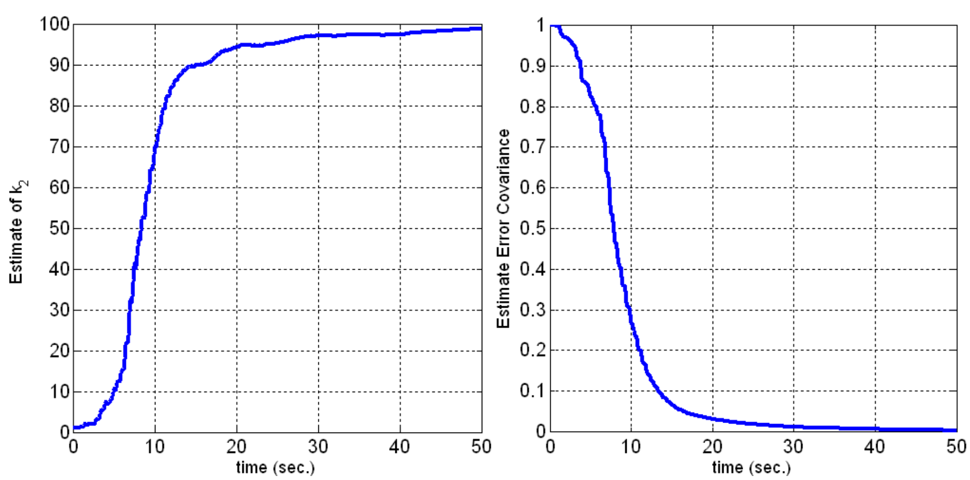

Figure 4.1: Estimate of second floor stiffness k_2 and error covariance.

Fourth-order Runge–Kutta method is used in order to integrate the first order differential functions in the prediction step of the EKF numerically. The estimate of k_2

and the state error covariance are presented in Fig.4.1. As can be seen the estimate of k_2 is converging to true value of 100 at 50 seconds, and the state error covariance is converging to zero as expected.

EKF with Large Size Models

The EKF approach for combined state estimation algorithm requires that at each time station one write the state space formulation explicitly as a function of the state and parameters in order to calculate the Jacobian. This is easily done when treating small models as shown in the example, but can be impractical when the parameter vector is large. Here we use a parametrization, described in the Appendix B, that simplifies implementation of the EKF-based parameter estimation algorithm efficiently regardless of the size of the model and parameter vector.

4.4 Numerical Experiment: Five-DOF Spring Mass System

In this numerical experiment we use the five-DOF spring mass system depicted in Fig.3.4 in order to examine the uncertainty modeling methods for Kalman filtering. We suppose that true stiffness and mass values are given in consistent units as $k_i = 100$ and $m_i = 0.05$, respectively. We assumed that the spring stiffness values of the model has uncertainty and the nominal model (A_n) is constructed based on the stiffness values are $\{80, 110, 90, 85, 110, 110, \text{and } 105\}$. The un-damped frequencies of the system and the model used in Kalman filtering are depicted in Table 4.1.

Damping is classical with 2% in each mode. We obtain results for output sensors at the third masses, which are recording velocity data at 100Hz sampling. The measurement

Table 4.1: The un-damped frequencies of the spring mass system and erroneous model.

Frequency No.	System	Model	%Change
1	0.582	0.545	6.349
2	1.591	1.594	0.184
3	2.851	2.883	1.104
4	3.183	3.119	2.008
5	3.434	3.470	1.073

noise is prescribed to have a root-mean-square (RMS) equal to approximately 10% of the RMS of the response measured. Unmeasured excitations and measurement noise are assumed to be mutually uncorrelated, with the covariance matrices,

$$Q = \begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix} \qquad R = 0.030 \qquad S = 0$$

Stochastic Modeling of Uncertainty

Fictitious noise approach is applied in line with section 4.2.1 and using correlations of innovations process. The arbitrary filter gain K_0, that is chosen such that eigenvalues of the matrix $(A_n - K_0C)$ are assumed to have the same phase as those of A_n but with a 20% smaller radius. 80 lags of correlation functions of innovations process is taken into consideration and the sample innovation correlations functions are calculated using 600 seconds of data. The noise covariance matrices of the equivalent stochastic model is calculated as,

$$\bar{Q} = \begin{bmatrix} 11.86 & 5.64 & 4.33 & -7.23 & -0.015 \\ 5.64 & 3.58 & 0.90 & -1.80 & 1.33 \\ 4.33 & 0.90 & 3.27 & -4.99 & -1.94 \\ -7.23 & -1.80 & -4.99 & 7.69 & 2.719 \\ -0.015 & 1.33 & -1.94 & 2.719 & 2.244 \end{bmatrix} \qquad \bar{R} = 0.0067 \qquad \bar{S} = \begin{bmatrix} -0.218 \\ -0.058 \\ -0.147 \\ 0.227 \\ 0.077 \end{bmatrix}$$

The output correlation function of the actual system and equivalent stochastic model are depicted in Fig.4.2. The $\bar{Q}$, $\bar{R}$ and $\bar{S}$ are used to calculate a filter gain from the classical formulations of the Kalman filter and the state estimates are obtained from this filter.

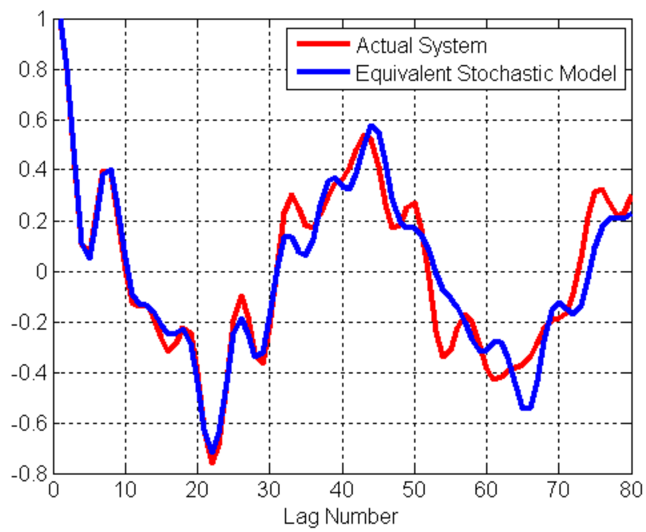

Figure 4.2: The output correlation function of the five-DOF spring mass system.

The equivalent Kalman Filter approach is applied in line with section 4.2.2 and the state estimates are obtained using the calculated filter gain. Since the optimal estimate of the state can only be calculated using a Kalman filter that is constructed error free model of the actual system and the true noise covariance matrices, the methods examined in this study are suboptimal. For the comparison of the methods, we take the discrepancy between state estimate from optimal Kalman filter and suboptimal state estimate, namely

$$\varepsilon = \hat{x}_{optimal} - \hat{x}_{suboptimal} \tag{4.4.1}$$

and define the filter cost as

$$J = trace(E(\varepsilon\varepsilon^T)) \tag{4.4.2}$$

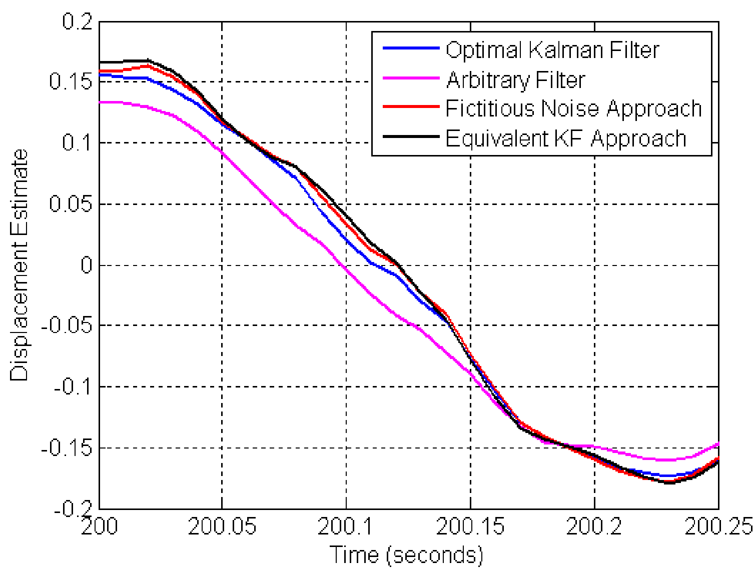

Figure 4.3: Displacement estimate of the second mass

The displacement estimate of the second mass from the fictitious noise and equivalent Kalman filter approaches are depicted in Fig.4.3. As can be seen examined methods give better estimates compare to the arbitrary filter. Histograms of filter cost from 200 simulations for the examined methods and the arbitrary filter are depicted in Fig.4.4. As can be seen, the estimate from the arbitrary filter is the worst with a mean of the filter cost $\mu = 11.84$. The fictitious noise approach performs better compare to the equivalent Kalman filter approach. The mean of the filter cost from 200 simulations are 1.60 and 2.65, respectively.

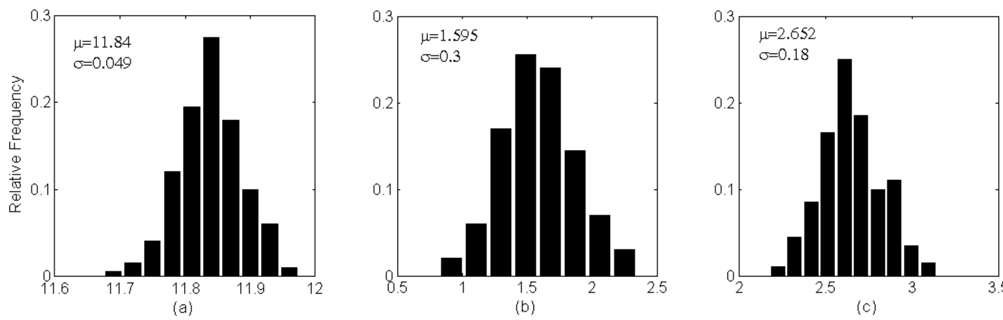

Figure 4.4: Histograms of filter cost from 200 simulations, a) Arbitrary filter b) Fictitious Noise Approach c) Equivalent Kalman Filter Approach

Combined State and Parameter Estimation

The initial error covariance for each spring is taken as 200. The combined state and parameter estimation is applied using parametrization scheme described in the Appendix B. The results from a single simulation of 120 seconds are presented in Fig.4.5. As can be seen, the stiffness values are converging to 100 and error covariance values are converging to zero. The estimated spring stiffness values at the end of the simulation are {102.5, 97.9, 99.1, 100.1, 98.02, 104.3 and 102.9}.

Although the EKF-based parameter estimation method leads to satisfying results in this numerical experiment, examinations show that the method is not robust to large uncertainties in initial parameter estimate and error covariance matrix when unknown parameter vector is large. For instance, in this experiment an initial stiffness parameter vector of {25, 150, 50, 170, 10, 190, 250} leads to divergence in some of the parameter estimates. The dependence of estimation results on the initial estimation error covari-

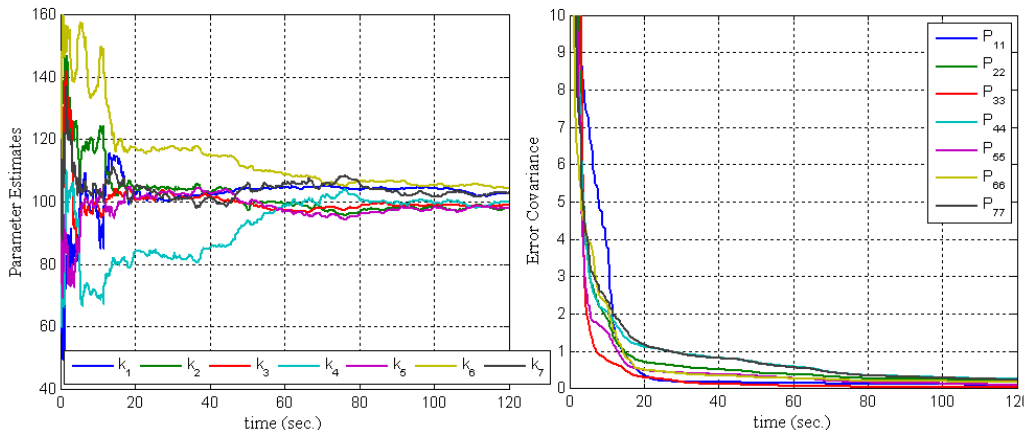

Figure 4.5: Spring stiffness estimates and error covariance for the five-DOF spring mass system

ance matrix and the initial value of parameters to be estimated is also reported by Liu, Escamilla-Ambrosio and Lieven [46]. The authors examined the merit of using a multiple model adaptive estimator consisting of a bank of EKF, in which a different error covariance for each filter in the bank is chosen. Due to the fact that the algorithm has high computational cost; implementation is difficult.

4.5 Summary

This chapter studies state estimation of time invariant stochastic systems that analytical model of the system has uncertainty. The operating assumptions are that the system is linear and subjected to unmeasured Gaussian stationary disturbances and measurement noise, which are correlated. In classical Kalman filter theory, the model that represents the actual system is assumed known. Here, we assumed that state transition matrices (A) has error and noise covariance matrices Q, R and S are not known. The chapter is addressed the uncertainty issue in model that is used in Kalman filtering and examined the feasibility and merit of an approach that takes the effects of the uncertain parameters of the nominal model into account in state estimation. In this approach, the system is approximated with a stochastic model and the problem is addressed in off-line conditions. The model errors are approximated by fictitious noise and covariance of the fictitious noise is calculated using the data on the premise that the norm of discrepancy between correlation functions of the measurements and their estimates from the nominal model is minimum. Another approach examined approximates the system with an equivalent Kalman filter model and the filter gain is calculated using the data on the premise that the norm of measurement error of the filter is minimum. Additionally, the problem is addressed in on-line operating conditions, and the EKF-based combined parameter and state estimation method is examined. In this method, the uncertain parameters are taken as part of the state vector and the combined parameter and state estimation problem is solved as a nonlinear estimation via EKF. From the results presented in the previous sections, we can identify the following conclusions:

- The fictitious noise and equivalent Kalman filter approaches are applicable in off-line conditions where stored measurement data is available. The fictitious noise approach leads to expressions that are more complex than the equivalent Kalman filter approach scheme, but the differences are not important when it comes down to computer implementation. Examinations show that although the state estimates from these two approaches are suboptimal, they both perform better than an arbitrary filter.

- In general, the fictitious noise approach performs better than the equivalent Kalman filter approach. However, the performance of the fictitious noise approach is depended on the length of the data. This is due to the fact that the output correlations are calculated from a finite length sequences and the accuracy of the sample correlations increase with longer data. Therefore, when the sample correlations are not calculated with a good approximation, e.g. they are calculated from short data, the equivalent Kalman filter approach gives better state estimates for the same data.

- Online estimation of the uncertain model parameters using EKF-based combined parameter and state estimation strategy is simple in theory but is not trivial in applications when the model is large and the uncertain model parameters are too many. The EKF requires the computation of Jacobian of the augmented model. That is easy for systems with small model that state and measurement equations can be written explicitly. However, writing explicit state and measurement equations are impractical for systems with large models. A parametrization scheme for structural matrices (M, C, K) is presented that simplifies implementation of the EKF-based combined state estimation algorithm regardless of the size of the model and parameter vector. The performance of the EKF-based parameter estimation for linear systems is shown for a simulated five-DOF spring mass system.

- Examinations show that the EKF-based combined parameter estimation approach is not robust to large uncertainties in initial parameter estimate and error covariance matrix when unknown parameter vector is large.

Chapter 5

Damage Detection using Kalman Filter

5.1 Background and Motivation

Damage is caused by factors such as aging, fatigue, earthquakes and blast-loading. Implementing a damage identification strategy for civil engineering structures involves observing the structures over time using periodically taken measurements, extracting damage-sensitive features from these measurements and performing a statistical analysis of these features to determine if the system has changed notably from the reference state, [61]. Two major components comprise a typical damage identification system: a network of sensors for observations; and data process algorithms for structural assessment of physical condition, [62]. In structural engineering, one has to deal with four classes of changes in monitored structures, as described below:

(1) Structural changes: Structural changes refer to changes in the dynamical system itself and boundary conditions. Examples of structural changes include concrete

degradation and steel corrosion in structural members and connections or settlements in foundations, etc.

(2) Changes due to environmental conditions: One of the most important issues in damage detection in civil structures is changing environmental conditions. In laboratory tests, environmental conditions usually do not play any role. However, in reality, civil structures are subjected to temperature differences and those differences have an influence on the dynamic characteristics of structures. Environmental conditions include wind, temperature and humidity.

(3) Operational condition changes: Civil engineering structures are constantly excited by many unknown disturbances such as traffic, building occupants, operating machinery, etc. Operational condition changes include operational speed and mass loading such as traffic loading and speed.

(4) Malfunctioning instruments.

In past decades, a broad spectrum of methods has been proposed for damage identification. These methods are classified into local and global. Local methods often are known as Non-Destructive Evaluation. Common local damage detection methods include thermography, magnetic-field analysis, ultrasonic inspection, radiography and edyy-current methods. These methods can successfully detect damage if the possible damage regions are known and if the regions of the structure needing inspection are accessible, [63]. In global damage identification, damage is defined as change in system parameters that are identified from measurement data. Global methods often are developed using measured vibration data on the premise that commonly measured dynamic quantities and global modal characteristics (frequencies, mode shapes, and modal damping) are functions of the physical properties of the structure (mass, damping, and stiffness). Consequently, it is assumed that changes in the physical properties, such as reductions in stiffness resulting from damage, will cause detectable changes in the modal properties of the structures. A review of the literature on global damage identification methods in structural engineering published before 1996 is presented by Doebling et.al. [64]; more recent advances are presented by Sohn et.al. [65]. Global damage identification includes modal-based techniques, model-updating techniques, and damage detection filter methods.

The work in this chapter falls within the category of the detection filter methods and the use of Kalman filter as a damage detector is examined. Our focus is on the detection rather than localization of the damage. To the best of the writer's knowledge, the Kalman filter based damage detection is first discussed by Mehra and Peshon, [66]. They used the Kalman filter for residual generation under the premise that an accurate model for the system is available and that the covariance matrices of the disturbance and the measurement noise are known. The seminal work of damage detection filters is done by Beard [67] and Jones [68]. The main idea is that filter gain is chosen such that particular damaged scenario can be seen in the residuals. Kranock extended Beard's work and proposed the use of a structural connectivity matrix and treated the forces resulting from damage as inputs to the system. His method attempts to detect the damage in real time. Another extension of the damage detection filters is proposed by Liberatore, Speyer and Hsu [69]. They used the damage direction vectors obtained from pre-defined damage locations in the structure which are used as the basis for identification of each of the possible damage locations.

The use of bank of parallel detection filters have had quite attention in the literature. The idea is that different damage scenarios are defined for each of the filter models in the bank, and multiple innovations tests respond to one of them, indicating damages included. The bank of parallel filters strategy is first applied by Montgomery and Caglayan, [70]. The filter based damage detection in structural engineering first applied by Fritzen, Seibold and Buchen [71], who used a bank of Kalman filters that each of the filter models represents a special type of damage. The filter residuals are used to localize the damage. Seibold and Weinert applied the same idea for the localization of cracks on a rotor dynamic system, [72]. Fritzen and Mengelkamp [73] and Yan et al. [74] used the Kalman filter model obtained from observations to detect damage by checking covariance of the innovation process.

There are research efforts that focus on identification of system changes due to changing environmental conditions. In this regard, Peeters and Roeck [75] attempted to develop models from observations for environmental changes. They identified a stochastic model using temperature data as input and estimated frequencies of the structure from vibration data as output. The model than used to predict natural frequencies and discrepancy between predicted and measured frequency is used to separate the influences of temperature from damage on dynamical modal parameters. Kim, Yun and Yi described a set of empirical frequency-correction formulas using the relationship between temperature and natural frequencies, [76]. They adjusted the measured frequencies by the frequency-correction formulas and fed them into the damage-detection scheme. It's known that varying environmental conditions are changing the modal parameters of structures slowly. Hence, one can update the analytical model to take into account the changes in the structure due to varying temperature.

The damage detection strategy examined in this chapter attempts to take operational condition changes into consideration. On this subject, Zhang, Fan and Yuan reported traffic-induced variability in the dynamic properties of a cable-stayed bridge, [77]. They showed that the natural frequencies of the bridge can exhibit as much as 1% variation within a day under a steady wind and temperature environment. Ruotolo and Surace investigated a singular value decomposition based damage detection approach that compares the data obtained from the system before and after the damage, [78]. They focused on alterations during normal operating conditions, such as changes in mass, and tried to distinguish between changes in the working conditions and the onset of damage. Variations of this appraoch were presented by Vanlanduit et al. to detect damage using both numerical and experimental data. They used an aluminum beam with different damage scenarios and performed damage detection under several conditions including different varying operating conditions, [79]. Shon et al. used a combination of time series analysis and statistical pattern recognition techniques to detect damages on a roller coaster, in which data were subsequently acquired during a test operation with varying speeds, mass loading and damage conditions. The reader is referred to [80] for other research efforts focused on damage detection under varying operational conditions in civil structures.

In this work, we suppose that the unmeasured disturbances acting on the structures are characterized with stochastic process with a stationary Gaussian distribution. Using available measured data and an analytical model of the structure, we use the Kalman filter to generate the innovations process. Our focus is on the operational conditions where covariance of unmeasured disturbances is subject to change between data collection sessions and we aim to detect damages using Kalman filter innovations. The classical Kalman filter approach to damage detection uses the whiteness property of the innovations process and applies a correlations based whiteness test to the innovations for deciding whether there is change in the system or not, [?]. The classical correlations based whiteness test used, in which the criteria is that the correlations of innovations inside the confidence interval indicate that the innovations are uncorrelated. Changes of statistical characteristics of the innovation process are caused by variability in covariance of disturbance and measurement noise and change in dynamical system parameters due to damage. The objective is to make the detection insensitive to disturbance changes and, at the same time, sensitive to damages.

5.2 Innovations Approach to Damage Detection

The innovations approach to damage detection involves the following components: 1) Kalman filter that is used to generate innovations process using available measurements, 2) Metric derived from the innovations, 3) Criteria for damage detection based on the metric. The properties of the Kalman filter innovations process are presented in section 2.3.9. Here we explore the effect of variability in noise covariance statistics and damage on the correlations functions of innovations process.

5.2.1 Dependence of the Innovations Correlations on the Noise Covariances

We consider the discrete dynamic system described in Eqs.2.3.1-2.3.2, and refer to it as reference system. We suppose that the innovations process is generated using a Kalman filter in the form of Eq.2.3.48 that is formulated using the reference system. Since any change in the the reference system parameters and noise statistics make the Kalman filter suboptimal, and we denote the filter gain as K_0. As shown in section 2.3.9, the correlations function of innovations process $\mathcal{L}_j$ can be expressed as a function of K_0 and noise covariance matrices, Q, R and S. Applying the vec operator to both sides of Eqs.2.3.119 and 2.3.114 one has

$$\begin{aligned} vec(\mathcal{L}_j) = (C\bar{A}^j \otimes C)vec(\bar{P}) + (G^T \otimes C\bar{A}^{j-1})vec(S) - \\ -(I \otimes C\bar{A}^{j-1}K_0)vect(R) \qquad j \neq 0 \end{aligned} \tag{5.2.1}$$

and

$$vec(\bar{P}) = [I - (\bar{A} \otimes \bar{A})]^{-1}[(K_0 \otimes K_0)vec(R) + G \otimes Gvec(Q) - \\ - (G \otimes K_0)vec(S) - (K_0 \otimes G)vec(S^T)] \quad (5.2.2)$$

Substituting Eq.5.2.2 into Eq.5.2.1, and adding the terms related to S^T to the terms related to S and canceling S^T, after organizing one finds

$$vec(\mathcal{L}_j) = \begin{bmatrix} h_j^q & h_j^s & h_j^r \end{bmatrix} \begin{bmatrix} vec(Q) \\ vec(S) \\ vec(R) \end{bmatrix} \quad (5.2.3)$$

where

$$h_j^q = (C \otimes C\bar{A}^j)[I - (\bar{A} \otimes \bar{A})]^{-1}(G \otimes G) \qquad j \neq 0 \quad (5.2.4)$$

$$h_j^s = (G^T \otimes C\bar{A}^{j-1}) - 2I[(C \otimes C\bar{A}^j)[I - (\bar{A} \otimes \bar{A})]^{-1}(G \otimes K_0)] \;\; j \neq 0 \quad (5.2.5)$$

$$h_j^r = (C \otimes C\bar{A}^j)[I - (\bar{A} \otimes \bar{A})]^{-1}(K_0 \otimes K_0) - (I \otimes C\bar{A}^{j-1}K_0) \quad j \neq 0 \quad (5.2.6)$$

In ideal conditions, when K_0 is the Kalman filter gain, the expected value of correlation functions of innovations is zero. When noise covariance matrices are changed, the reference filter becomes suboptimal and the innovations process become correlated. It is apparent from Eq.5.2.3 that the correlations of the innovations process is linearly related to the noise covariance matrices, Q, R and S. The variations in noise covariance matrices have the effect in the correlations is determined by the coefficient matrices h^q, h^s and h^r, and these matrices have a norm that decays with lags. In particular, changes that have strong projections in the direction of the vectors associated with the higher singular values of h^q, h^s and h^r introduce large change in the correlations while those that project on vectors associated with the smaller singular values have small effect.

5.2.2 Effect of Damage On the Innovations Correlations

The effect of the change in dynamical system parameters on the innovations process is explored in this subsection.. We start with introducing parameter changes due to damage in the system dynamics equation. Consider state space model of a dynamical system after damage occurred is given in the following,

$$x_{k+1} = Ax_k + \triangle Ax_k + Gw_k \tag{5.2.7}$$

$$y_k = Cx_k + v_k \tag{5.2.8}$$

where $\triangle A \, \epsilon \, \mathbb{R}^{nxn}$ represents the change in the system parameters due to damage. We refer to the original system for the assumptions on the system and noise characteristics given in Eqs.2.3.1 and 2.3.2. While system parameters change due to damage, the Kalman filter model remains the same as given in Eq.2.3.48. Thus, the filter is not optimal anymore since the observations, y_k from the damaged system used in the filtering process. Therefore we denote the filter gain as K_0 and innovations process as $\bar{e}_k$, and start recalling the correlation function of innovations from Eq.2.3.112, which is

$$E(\bar{e}_k\bar{e}_{k-j}) = CE(\bar{\varepsilon}_k\bar{\varepsilon}_{k-j}^T)C^T + CE(\bar{\varepsilon}_k v_{k-j}^T) \tag{5.2.9}$$

To obtain the terms $E(\bar{\varepsilon}_k\bar{\varepsilon}_{k-j})$ and $E(\bar{\varepsilon}_k v_{k-j}^T)$ in Eq.5.2.9, we use similar steps to obtain them as previously followed in section 2.3.9. The recurrence for the state error in the damaged system case is presented in the following,

$$\bar{\varepsilon}_k = A(I - K_0 C)\bar{\varepsilon}_{k-1}^T - K_0 v_{k-1} + G w_{k-1} + \triangle A x_{k-1} \tag{5.2.10}$$

Carrying Eq.5.2.10 j steps back one find that,

$$\bar{\varepsilon}_k = \bar{A}^j \bar{\varepsilon}_{k-j} - \sum_{t=1}^{j} \bar{A}^{t-1} K_0 v_{k-t} + \sum_{t=1}^{j} \bar{A}^{t-1} G w_{k-t} + \triangle A \sum_{t=1}^{j} \bar{A}^{t-1} x_{k-t} \tag{5.2.11}$$

Post-multiplying Eq.5.2.11 by $\bar{\varepsilon}_{k-j}$ and taking expectation, it follows that

$$E(\bar{\varepsilon}_k \bar{\varepsilon}_{k-j}^T) = \bar{A}^j E(\bar{\varepsilon}_{k-j} \bar{\varepsilon}_{k-j}^T) + \triangle A \sum_{t=1}^{j} \bar{A}^{j-1} E(x_{k-t} \bar{\varepsilon}_{k-j}^T) \tag{5.2.12}$$

Post-multiplying Eq.5.2.11 by v_{k-j}^T and taking expectation,

$$E(\bar{\varepsilon}_k v_{k-j}^T) = \bar{A}^{j-1} G S - \bar{A}^{j-1} A K_0 R + \triangle A \sum_{t=1}^{j} \bar{A}^{j-1} E(x_{k-t} v_{k-j}^T) \tag{5.2.13}$$

Substituting Eqs.5.2.12 and 5.2.13 into 5.2.9 one finds

$$\begin{aligned} E(\bar{e}_k \bar{e}_{k-j}) = {} & C \bar{A}^j E(\bar{\varepsilon}_{k-j} \bar{\varepsilon}_{k-j}^T) C^T + C \bar{A}^{j-1} G S - C \bar{A}^{j-1} A K_0 R + \\ & C \triangle A \left[\sum_{t=1}^{j} \bar{A}^{j-1} E(x_{k-t} \bar{\varepsilon}_{k-j}^T) \right] C^T + C \triangle A \left[\sum_{t=1}^{j} \bar{A}^{j-1} E(x_{k-t} v_{k-j}^T) \right] \end{aligned} \tag{5.2.14}$$

The main observation from Eq.5.2.14 is that after damage occurs in the dynamical system. The expected value of correlations of filter innovations do not vanish as the lag approaches large values. If the eigenvalues of the matrix $\bar{A}$ are inside the unit circle (stability condition of the discrete filter model), then the first three terms in the Eq.5.2.14 vanishes as the lag j increases and the other summation terms are governed by the few lower powers.

We note that, the change in the system parameters, due to discretization of the system, always lead to changes in input to state matrix (G) as well. However, the change in G is trivial in our examination since it does not make any change in the assumptions on the w_k and in the conclusions from the Eq.5.2.14.

5.2.3 Frequency Domain Interpretation

An examination of the transfer function from process noise to innovations helps to develop insight into the mechanism through which damage affects the correlation of the Kalman filter innovations. The Kalman filter model in output form has the following description:

$$\dot{\hat{x}}(t) = (A_c - KC_c)\hat{x}(t) + Ky(t) \tag{5.2.15}$$

$$e(t) = y(t) - C_c\hat{x}(t) \tag{5.2.16}$$

Subscript c is used to refer to continuous system matrices. One can show that the relation between between the process noise and the output in the s-domain is

$$y(s) = C_c(sI - A_c)^{-1}G_c w(s) + v(s) \tag{5.2.17}$$

and from Eqs.5.2.15-5.2.16 one has

$$e(s) = \left(-C_c[sI - (A_c - KC]^{-1}K + I\right) y(s) \tag{5.2.18}$$

Combining Eq.5.2.17 and Eq.5.2.18 one gets that

$$e(s) = \begin{bmatrix} T_{w/e}(s) & T_{v/e}(s) \end{bmatrix} \begin{bmatrix} w(s) \\ v(s) \end{bmatrix} \tag{5.2.19}$$

where

$$T_{w/e}(s) = \left(-C_c[sI - (A_c - KC_c]^{-1}K + I\right)\left(C_c(sI - A_c)^{-1}B_c\right) \tag{5.2.20}$$

and

$$T_{v/e}(s) = \left(-C_c[sI - (A_c - KC_c]^{-1}K + I\right) \tag{5.2.21}$$

$T_{w/e}(s)$ and $T_{w/e}(s)$ are transfer functions from the process and measurement noise to the innovations respectively. When the Kalman filter is optimal, the innovation process is white signal and the parameters of the coefficient matrix in Eq.5.2.19 prove to be constants. To analyze this in a simpler format, we neglect measurement noise and consider the single input single output case. In this case, coefficient matrix in Eq.5.2.19 involves only transfer function from process noise to innovations, namely $T_{w/e}(s)$. It is convenient to factor the $T_{w/e}(s)$ in the numerator and denominator, and to write it in terms of those factors as follows

$$T_{w/e}(s) = g_1 g_2 \frac{\prod s - z_i}{\prod s - p_i} \frac{\prod s - z_j}{\prod s - p_j} \tag{5.2.22}$$

where the definitions of subscripts i and j are clear from Eq.5.2.20 that they refer to transfer matrices from output to innovations (closed loop) and from process noise to output (open loop) respectively, g_1 and g_2 are the gains of the corresponding transfer functions. The values for z_i and z_j where the numerator of the transfer functions equal zero are the complex frequencies that make transfer function zero, and they are called zeros of the transfer function. The values for p_i and p_j, where the denominator of the transfer function equal zero, are the complex frequencies that make the transfer function infinite. When the filter gain is optimal, the poles p_i are equal to the zeros $z_j$and the zeros z_i are equal to the poles p_j so the transfer is just the product of the gains, g_1 and g_2. The system changes can be observed in $T_{w/e}(s)$ as follows:

When damage takes place the transfer function from output to innovations remains the same and so p_i and z_i stay unchanged however the p_j and z_j change. Therefore, in damage case, the innovations process may have large contributions from sinusoids for which the new values of p_j as well as p_i which are identical to the zeros of the open loop before the system get damaged, namely z_j. In some cases the poles of the closed loop, p_i might be far from the imaginary, and the fact that the zeros z_j may vary in position as a result of the damage does not tend to have much influence on what happens along the imaginary line so the innovations process may have large contributions only from sinusoids due to the new values of p_j.

The following example is intended to provide more insight on this point and to exemplify the frequency domain interpretation of damage detection using Kalman filter innovations. Consider a two-DOF shear frame whose story stiffnesses and story masses are given in consistent units as {1000, 1000} and {1.25, 1.10} respectively. The undamped frequencies are {2.91,7.42} in Hz. Damping is classical with 2% in each mode. We obtain transfer functions for a single input and single output sensor arrangement at first floor. Measurement sensor is recording velocity data without noise and the deterministic excitation is a white noise with a unit variance. The damage scenario examined is 10% loss of stiffness in each floor.

Poles and zeros of the transfer functions from output to innovations (closed loop) and process noise to output (open loop) in optimal case is depicted in Table 5.1. As can been seen, zeros of the closed loop are equal poles of the open loop. And the poles of the closed loop are equal with zeros of the open loop with two additional zeros. Therefore transfer function from output to innovations has only two zeros which are on the real axis.

Table 5.1: Poles and zeros of the transfer functions in optimal case.

Open Loop		Closed Loop	
zeros	poles	zeros	poles
-0.54±30.15i	-0.093±46.62i	-0.093±46.62i	-0.54±30.15i
	-0.37±18.28i	-0.37±18.28i	-0.01
			-7.99x10^4

Poles and zeros of the closed loop and open loop in damage case are depicted in Table 5.2. As can been seen, poles and the zeros of the closed loop stay unchanged since the filter model remains the same. However poles and zeros of the open loop are shifted due to damage; therefore transfer function from process noise to innovations in this case has three zeros and five poles, in which two of them are on the real axis.

Table 5.2: Poles and zeros of the transfer functions in damage case.

Open Loop		Closed Loop	
zeros	poles	zeros	poles
-0.54±28.59i	-0.093±44.23i	-0.093±46.62i	-0.54±30.15i
	-0.36±17.34i	-0.37±18.28i	-0.01
			-7.99x10^4

Frequency response of the transfer function from process noise to innovations are depicted in Fig.5.1. Two of the resonant peaks of the frequency response plot in damage case at 2.47Hz and 7.04Hz are the natural frequencies of the damaged system, and the third one at 4.8Hz is the zero of the healthy system. Two of the anti-resonant peaks of the frequency response plot in damage case at 2.91Hz and 7.42Hz are the natural frequencies of the healthy system, and the third one is the zero of the damaged system.

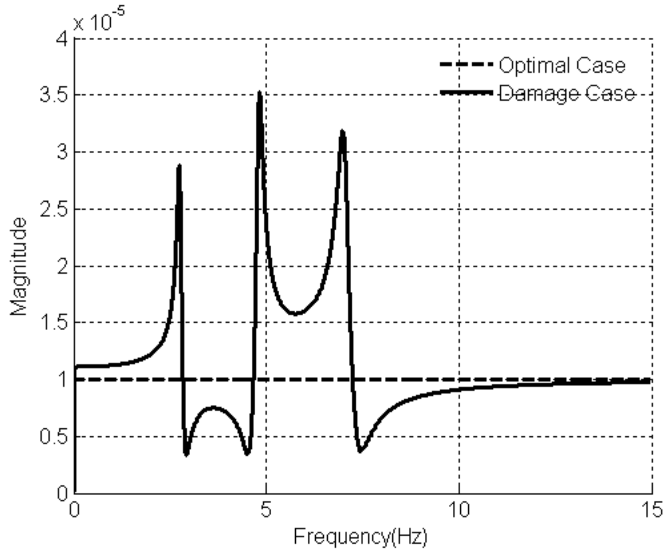

Figure 5.1: Frequency response of the transfer function from process noise to innovations.

5.3 A Modified Whiteness Test

In statistics, two correlation tests for time series data are commonly used: (1) Independence test, (2) Whiteness test. According to the independence test criteria, the optimal Kalman filter innovations are uncorrelated with past inputs. This method is applicable only to the systems in which the input is deterministic and known. According to the whiteness test criteria, the correlations of innovations inside the confidence interval indicate that the innovations are uncorrelated and decisions on the whiteness of the signal are made using a statistical hypothesis test, [81].

5.3.1 Hypothesis Testing

The damage detection approach based on correlations of the Kalman filter innovations makes the decision on two hypotheses, which are

- $\mathcal{H}_0 : \ \rho \leq \rho_0 \quad$ {'structure is undamaged'}
- $\mathcal{H}_1 : \ \rho > \rho_0 \quad$ {'structure is damaged'}

$\mathcal{H}_0$ is the null hypothesis, $\mathcal{H}_1$ is the alternative hypothesis and ρ is a metric used for classification between damage and no damage whose value is equal to a certain value e.g. it is smaller than cut-ff ρ_0 under $\mathcal{H}_0$. The crucial part of the hypothesis testing is the set of outcomes which, if they occur, will lead us to decide that there is damage. That is, cause the null hypothesis to be rejected in favor of the alternative hypothesis. The cutt-off as to which hypothesis to accept is selected based on a set of observations whose PDF conditioned on $\mathcal{H}_0$ and $\mathcal{H}_1$

Two types of errors can be observed in the hypothesis testing approach for damage detection. If $\rho > \rho_0$ and one incorrectly claims that the structure is damaged, this constitutes Type-I error. The opposite situation, in which $\rho < \rho_0$ and one incorrectly claims that the structure is healthy, is called Type-II error. Probabilities for these errors are defined in the following,

- Type-I error probability: $\alpha = P\{\mathcal{H}_1 | \mathcal{H}_0\} = P\{accept\ \mathcal{H}_1 | \mathcal{H}_0\ true\}$
- Type-II error probability: $\beta = P\{\mathcal{H}_0 | \mathcal{H}_1\} = P\{accept\ \mathcal{H}_0 | \mathcal{H}_1\ true\}$

An illustration of possible distributions of ρ from healthy and a damage state is depicted in Fig.5.2. The probabilities of Type-I error and Type-II error for a given cut-off ρ_0 above which damage is to be announced are illustrated in the figure. Power of the test (P_T) , also known as the probability of detection, is defined as one minus the probability of Type-II error, namely

$$P_T = P\{\mathcal{H}_1 | \mathcal{H}_1\} = 1 - \beta \tag{5.3.1}$$

and it measures the performance of the test capability to detect $\mathcal{H}_1$ when it is true.

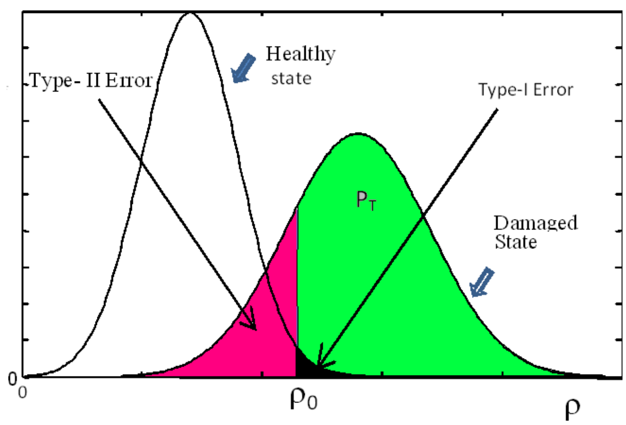

Figure 5.2: PDF of ρ from healthy and a damage state.

The hypothesis testing is conditional on the fact that the system is undamaged and probability distribution of the metric in healty state is known. The operating assumption on the damaged state is that probability distribution of metric for the all possible damage scenarios of interest is shifted to the right relative to the reference. The test is performed using a cut-off ρ_0 that is selected from probability distribution of the metric in healty state for a given Type-I error propability, α.

The examinations show the performance of the examined appraoch in this work is depended on the size of the damage introduced to the system. That is due to the fact that when the probability distribution of the metric for a damage scenario is not shifted to the right relative to the reference, i.e when damage produces very small change in dynamics charateristic of the system, the power of test is very low.

5.3.2 The Test Statistics

A test statistic that quantifies the "whiteness" of a signal is defined using auto-correlations of the signal. We use the sum of the auto-correlations of the innovations for a preselected number of lags. We begin with obtaining a unit variance normalized innovation sequence. To do this, the sample covariance matrix of the innovations

$$\hat{C}_0 = \frac{1}{N}\sum_{k=1}^{N}(e_k - \overline{e})(e_k - \overline{e})^T \tag{5.3.2}$$

is computed, where e_k is the innovations process, N is the length of the sequence and $\overline{e}$ is the mean. The normalized innovations are obtained from,

$$\tilde{e}_k = \frac{e_k}{\sqrt{\hat{C}_0}} \tag{5.3.3}$$

An un-biased estimate of auto-correlation function of innovations is computed from

$$\hat{l}_j = \frac{1}{N-j}\sum_{k=1}^{N-j}(\tilde{e}_k)(\tilde{e}_{k-j})^T \quad for\; j = 1, 2, ...p \tag{5.3.4}$$

where j is the number of lags, [82]. The auto-correlation, $\hat{l}_j$ is equal to 1 at zero lag and remains between -1 and +1. On the premise that $N \gg j$ all $\hat{l}_j$, under under $\mathcal{H}_0$, are identically distributed random variables with a variance [83]

$$Var(\hat{l}_j) = \frac{1}{N} \tag{5.3.5}$$

To have a unit variance for each lag of correlations we normalize the correlation function, namely

$$\bar{l}_j = \hat{l}_j \sqrt{N} \tag{5.3.6}$$

And finally, we define the following metric for whiteness test,

$$\rho = \sum_{j=1}^{s} \bar{l}_j^2 \tag{5.3.7}$$

which follows a χ^2 distribution (under $\mathcal{H}_0$) with s degrees of freedom (DOF), [81]. The probability that the value from Eq.5.3.7 in any given realization is larger than any given number is obtained from the cumulative distribution function (CDF) of the χ^2 distribution for the appropriate number of DOF. A threshold for the metric, ρ_0 is obtained from χ^2 CDF for s DOF and a preselected Type-I error probability, α. The null hypothesis $\mathcal{H}_0$ is accepted if the test statistic is smaller than the selected threshold ρ_0 and rejected otherwise. The test requires a single measurement channel, for multi-output cases, one can treat each available channel as a detector and announced damage if the metric for any one exceeds the selected threshold.

Example:

Consider a three-degree-of-freedom shear frame whose story stiffnesses and story masses are given in consistent units as {100, 100, 100} and {0.120, 0.110, 0.100} respectively. The first un-damped frequency is 2.18Hz. Damping is classical with 2% in each mode. It's assumed there is a velocity sensor located at the first story and there is an unmeasured Gaussian disturbance with a covariance of $Q = 10$ at the third story. The exact response is computed at 100Hz sampling and Gaussian noise added to measurement with a covariance of $R = 0.01$ (consistent with a noise that has 10% of the standard deviation of the measurement). A single simulation is carried out with a duration of

300 seconds. To present behavior of innovations process due to change in the dynamical system, two additional cases are defined as follows:

Case #1: Unmeasured disturbance at the third story is scaled by 2 ($Q = 40$), and R is consistent with a noise that has 10% of the RMS of the measurement.

Case #2: 5% stiffness loss in the second floor. The first un-damped frequency after the stiffness change is introduced is 2.16Hz.

The Kalman filter innovations are generated for using the measurements from the healthy system and changed systems given in two cases. Sample autocorrelation functions of 50 lags for two cases are presented in Fig.5.3. The optimal case in Fig.5.3 refers to original dynamical system without any change. The expected value of correlation of a random noise signal with infinite duration is zero for any non-zero lag. However it's important to note that, due to finiteness of data, the autocorrelations are always significantly different from zero, as seen in Fig.5.3.

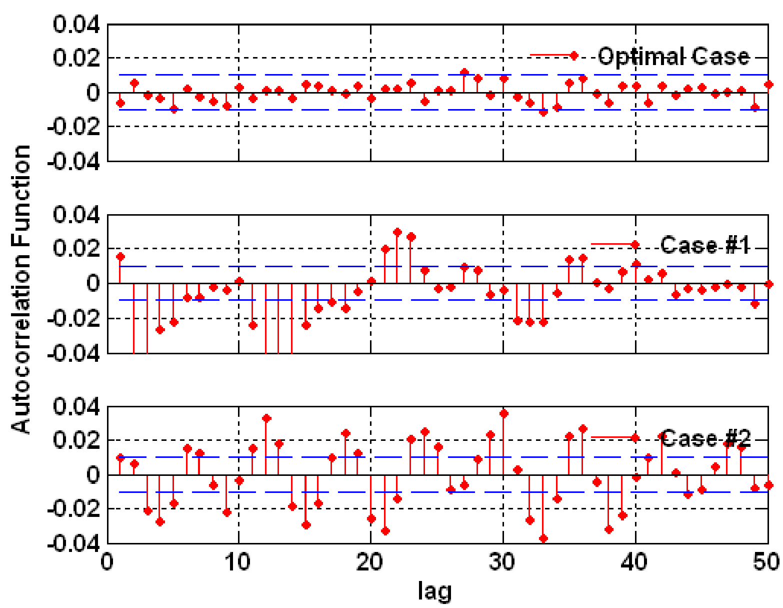

Figure 5.3: Autocorrelation function of innovations process. Dash line represents the 95% confidence interval.

For the example considered, the Chi-square whiteness test is carried out using 50 lags of correlations of innovations process with a Type-I error probability $\alpha = 5\%$ and the results are depicted in Table 5.3. It's clear from the table that Kalman filter is able to detect changes in the dynamical system and it shows that the innovations from the cases #1 and #2 are correlated. Consequently, the innovations are sensitive to change in disturbances as well as system changes and it is necessary to differentiate what comes from damage and what does not, which is addressed in the following section.

Table 5.3: Chi square correlation test results for Type-I error probability, $\alpha = 0.05$.

Optimal Case	$\rho_{50} = 49.87 < \chi^2_\alpha(50) = 67.5$	$\surd$
Case #1	$\rho_{50} = 848.3 > \chi^2_\alpha(50) = 67.5$	X
Case #2	$\rho_{50} = 718.12 > \chi^2_\alpha(50) = 67.5$	X

5.3.3 Modified Test Metric

The dependence of the innovations correlations on the noise covariances matrices is explored in section 5.2. Inspection of Eq.5.2.3 shows that the matrices h^q, h^s and h^r, involve the matrix $(A - KC)$ raised to powers that increase with the lag. Since this matrix has all eigenvalues in the unit circle (i.e., the filter is stable) the entries decrease as the lags increase and one concludes that, for sufficiently large lags the changes in the disturbances will have no effect on the correlations function. Using large lags of the correlations, a metric based on the modification of Eq.5.3.7 can be given as follows,

$$\rho = \sum_{j=d_1}^{d_2} \bar{l}_j^2 \tag{5.3.8}$$

where the first lag is taken as d_1 instead of one, and the number of lags $s = d_2 - d_1 + 1$. The modified metric will have a distribution that is essentially independent of the variations in the statistics of Q, R and S while correlations from damage are retained, provided that d_1 is large enough.

The range of lags that are used in the test is critical. If the correlation introduced by damage persisted at all lags d_1 could be selected arbitrarily (provided it is small compared to N) but examinations shows that this is not the case. The sensitivity to damage of the metric of Eq.5.3.8 also decreases with lags, although it has a different rate than that due to variations in the noise covariance matrices. This issue is illustrated using the example presented in the previous section. We calculated the metric ρ for 200 simulations and obtain an experimental PDF of ρ by fitting a generalized extreme value (GEV) density function for the system changes considered. Changes from one simulation to the next come from randomness in the unmeasured excitations and the measurement noise. We obtained the modified test results for three different number of lag values, namely $s = 25$, 50, 100 and 30 initial lags (d_1), which are chosen by starting from the lag #1 and by shifting at every 10th up to the lag #301. Power of test (P_T) results at 5% Type-I error for each case are depicted in Fig.5.4.

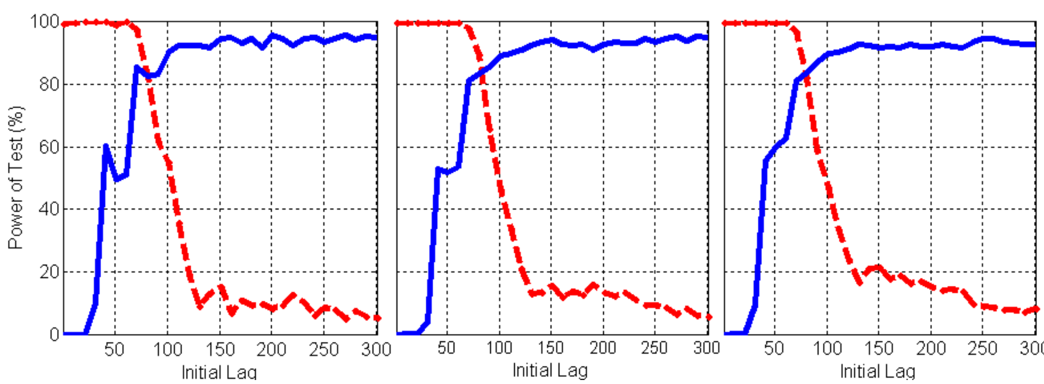

Figure 5.4: Trade-off between noise change and damage with respect to initial lag. Dash-Line: Damage Case, Solid Line: Noise Change, Left: $s = 25$, Middle: $s = 50$, Right: $s = 100$.

As can be seen from Fig.5.4, when d_1 is chosen in the first 25 lags, the test cannot differentiate the change in the noise covariance so the test fails almost all the time. The advantage of modified test becomes clear when d_1 is larger than 60. In this case the power of test for the noise covariance change case is increasing up to 90%.

It's also obvious from Fig.5.4 that there is a trade-off between noise change and damage cases in terms of location of initial lag (d_1), with respect to the power of whiteness test. Using higher lag bands gives better results for noise covariance change while lower lag bands lead high power of test in the damage case. In the noise covariance change case, after lag 100 the power of test fluctuates around 90%; however, in the damage case it drops drastically after the lag 60. Therefore, using a lag range starting from $d_1 = 75$ would give the maximum power of test which is around 85% for described system changes in the considered example.

An approach to choose d_1 is by inspecting the eigenvalues of the matrix in $(A-KC)^j$ and selecting a value such that the largest eigenvalue in absolute value, raised to d_1, is smaller than some pre-selected number. The behavior of the of largest eigenvalue of $(A-KC)^j$ in absolute value is depicted in Fig.5.5. As can be seen the maximum eigenvalue of $(A-KC)^j$ is decreasing to 0.1 after the lag 200, which show that the change in noise covariance matrices in the correlations functions of innovations will have no perceptible effect after the lag 200. In this experiment, however, the sensitivity to damage of the metric of Eq.5.3.8 also decreases quickly after the lag 75 with a corresponding largest eigenvalue in absolute value is 0.4.

As can be seen in Fig.5.4, using $s = 25$, 50 or 100 doesn't make much difference in this experiment. After damage appeared in the system, the innovations process involves oscillations with the same frequency content as the damaged system. Assuming the damage produces small shift in the un-damped natural frequencies of the healthy system, a heuristic criteria on selecting the number of lags used in the modified whiteness test, s, can be introduced as follows

$$s > \frac{T}{\Delta t} \tag{5.3.9}$$

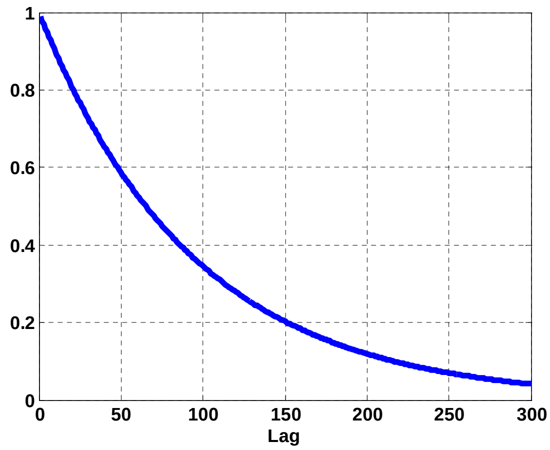

Figure 5.5: Largest eigenvalue of $(A - KC)^j$ in absolute value.

where Δt is the sampling interval of the discrete system and T is the fundamental period of the healthy system. The idea here is to cover all the lags of correlation functions that are in one period of the system's oscillation.

5.4 Numerical Experiment: Five-DOF Spring Mass System

In this experiment, we present an application of the innovations based damage detection technique and perform a Monte Carlo simulation using the five-DOF spring mass system depicted in Fig.3.4. We obtain results for a single input and single output sensor arrangement at coordinate #5. Measurement sensor is recording velocity data at 100Hz sampling. The deterministic excitation is a white noise with a unit variance. The unmeasured excitations were assumed to act at all masses and to have an RMS that, for each signal, is 10% of the RMS of the deterministic excitation, namely $Q = 0.01 * I$. Measurement noise has 10% of the standard deviation of the output and the variance is calculated as $R = 0.0016$. Unmeasured excitations and measurement noise are assumed to be mutually uncorrelated, namely $S = 0$. The Kalman filter is designed with this information of noise covariance matrices from the reference model and used to generate innovations process from the system subjected to the following changes.

- Change in noise statistics: Each entry of the diagonals in Q is allowed to vary independently between 0.25 and 4 times the value from the model. R is consistent with a noise that has 10% of the standard deviation of the output.

- Damage in the system: The damage scenarios examined are loss of stiffness in each one of the seven springs (one at a time) at three levels of severity: 2.5%, 5% and 10%.

200 simulations are performed with a duration of 400 seconds in each simulation. Change from one simulation to the next is coming from randomness in the unmeasured excitations and the measurement noise.

Table 5.4: Change in the first un-damped frequency (Hz) due to three damage cases in tfive-DOF spring mass system (as percent of healthy system frequency).

	Springs						
Damage (%)	k_1	k_2	k_3	k_4	k_5	k_6	k_7
2.5	0.8798	0.1548	0.0008	0.0391	0.0387	0.1309	0.0229
5	1.7822	0.3141	0.0017	0.0796	0.0797	0.2656	0.0466
10	3.6585	0.6471	0.0035	0.1650	0.1690	0.5475	0.0964

Test Parameters:

The whiteness test parameters, the location of first lag (s) and the number lags (d) are chosen in the line of Section 5.3. The number of lags being examined is calculated using Eq.5.3.9, namely

$$s > \frac{T}{\Delta t} = \frac{0.408}{0.01} = 40.8 \tag{5.4.1}$$

from where s is chosen as 50. Type-I error probability, P_{E_I} is assumed as 5% and the threshold for whiteness test is calculated as $\rho_0 = \chi^2_{0.05}(50) = 67.50$. Theoretical χ^2 CDF and PDF with 50 DOF and the threshold, ρ_0 is depicted in Fig.5.6

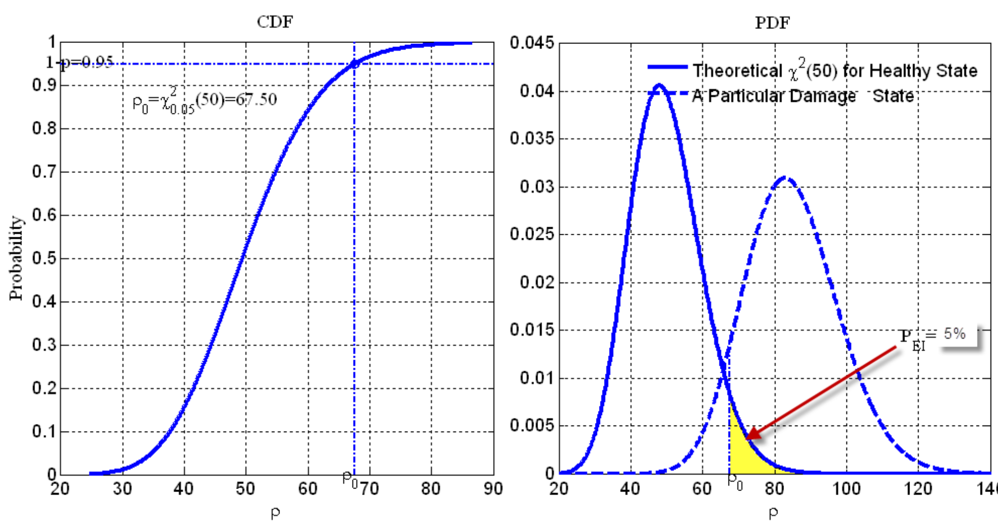

Figure 5.6: Theoretical χ^2 CDF and PDF with 50 DOF in in the numerical testing of the five-DOF spring mass system.

For the selection of the first lag, d_1 , the study in Fig.5.7 is performed to see how the largest eigenvalue of $(A - KC)^j$ decay as the lag increases. The location of the first lag, $d_1 = 60$ is chosen such that the largest eigenvalue of $(A - KC)^j$ in absolute value has decreased to 0.2.

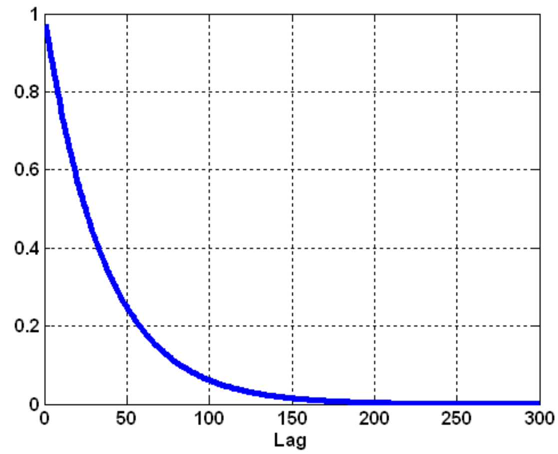

Figure 5.7: The largest eigenvalue of $(A - KC)^j$ in absolute value as the lag increases in numerical testing of five-DOF spring-mass system.

Results:

Simulation results are depicted in Figs.5.8-5.10. In Fig.5.8, autocorrelations of the innovations process are presented from a single simulation for two particular system changes. These particular changes are chosen from predefined system change scenarios, which are: (1) Disturbance at mass coordinate #2 is scaled by 2 and R is taken consistent with a measurement noise that has 10% of the RMS of the response. (2) 5% stiffness loss in the first spring.

As can be seen from Fig.5.8, the rate of decay of the correlations induced by changes in the noise statistics is much faster than the one for system changes. The correlations from the case of changes in the noise statistics fluctuate in the 95% confidence interval after lag 50.

We compare the behavior of the correlations for two range of lags, namely {1 to 50} and {61 to 110}. The experimental PDFs of ρ are estimated from 200 simulations by fitting a generalized extreme value (GEV) density function for the system changes considered. Fig.5.9 presents experimental PDFs of ρ with 50 DOF for both range of

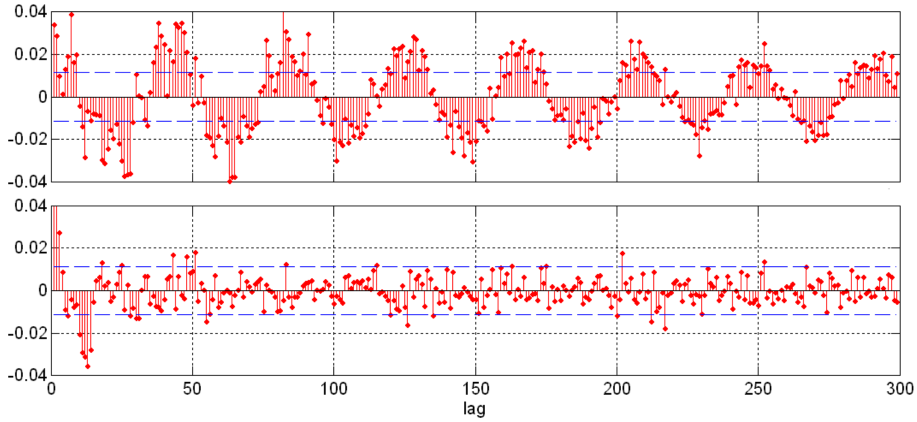

Figure 5.8: Auto-correlations of the innovations process from a single simulation, Bottom: Noise Change Case, disturbance at mass coordinate #2 is scaled by 2, Top: Damage Case with a 5% stiffness loss in the second spring. Dash line represents the 95% confidence interval.

lags from the case of change in the noise statistics. As can be seen, for the high lags band, the estimated PDF of $\chi^2(50)$ is very close to the theoretical one. In this case the discrepancy between experimental and theoretical PDFs might partially stem from duration of the data used in simulations. However, for the low lags band, experimental PDF is shifted significantly away from the theoretical one.

Power of test, P_T is calculated for each of the 21 damage cases using experimental PDFs estimated from 200 simulations. Figure 5.10 presents the power of test result with an 5% Type-I error for low and high range of lags. As can be seen, comparison between two range of lags band shows that low and high lags band leads to almost the same performance in the cases of 5% and 10% damage, with a 100% P_T for all springs. However, in the 2.5% damage case, the 3^{rd} and 7^{th} springs are poorly detectable even when low lags band is used, in which power of test is 48%. The resolution of the low lags band is superior to high lags band at 2.5% stiffness loss for the 2^{nd} and 4^{th} springs.

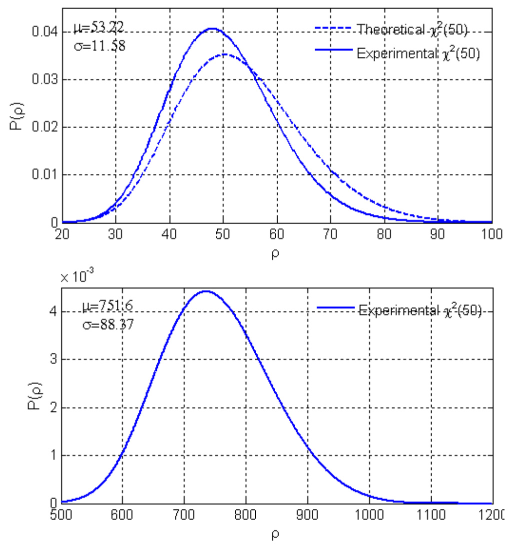

Figure 5.9: Experimental χ^2 PDFs of ρ with 50 DOF from 200 simulations for change in noise statistics, Range of Lags: Top= $61 \, to \, 110$, Bottom $= 1 \, to \, 50$.

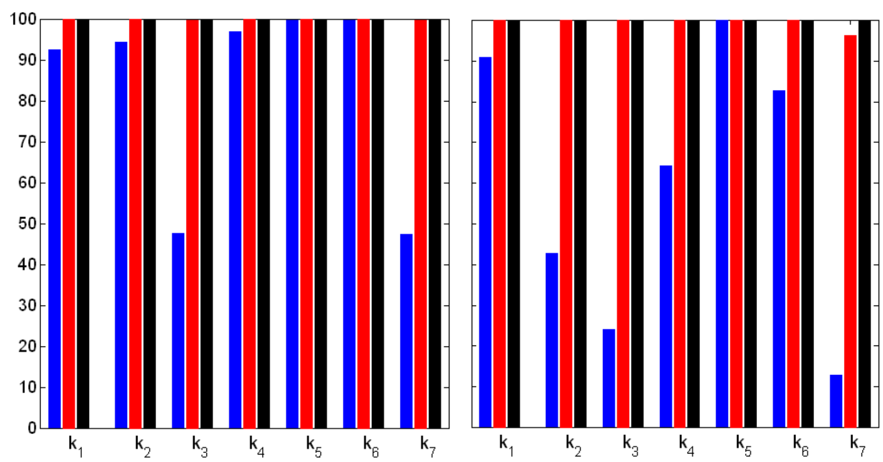

Figure 5.10: Power of test, (P_T) at 5% Type-I error in the numerical testing of five-DOF spring mass system. Damage Levels: Blue={2.5%}, Red={5%}, Black={10%}, Range of Lags: Left= {1 to 50}, Right = {61 to 110}.

5.5 Summary

The objective of the study in this chapter is to examine a damage detection technique based on whiteness property of Kalman filter innovations process. The system considered has time invariant discrete-time dynamics and is subjected to unmeasured stationary disturbances. The measurements are corrupted by white noise and available in discrete-time. It is assumed that the disturbance and measurement noise covariance fluctuates between data collection sections, and so the standard whiteness test for innovations process generated by reference Kalman filter model becomes ineffective. From the results presented in the previous sections, we can identify the following conclusions:

- Any change in the reference system parameters and noise statistics make the Kalman filter suboptimal. Theoretical derivations show that the correlations of the innovations from an arbitrary stable filter gain decrease with lag and asymptotically approach zero. However when the system changes, filter innovations do not vanish and asymptotically approach a value.

- A modified whiteness test is introduced. The test is insensitive to changes in the statistics of the disturbances and the measurement noise. The proposed whiteness test can be successfully applied to the damaged detection problem in structural systems that an analytical model is available without uncertainty. This is shown for a simulated five-DOF spring mass system.

- A special care has to be taken in order to choose range of lags used in the modified whiteness test. Using higher lag bands gives better results for noise change case while lower lag bands lead high power of test in the damage case. Therefore this trade-off has to be considered when the location of initial lag is decided. This is shown for a simulated three-DOF shear frame structural system.

Chapter 6

Summary and Conclusions

The studies described in this dissertation have attempted to approach three problems that arise in experimental mechanics where Kalman filter (KF) theory is used. The operating assumptions are that dynamical system of interest is linear time invariant and subjected to unmeasured Gaussian stationary disturbances. An analytical model that represents the system is assumed to be known and measurements are corrupted by white noise and available in discrete-time. From the results presented in the previous chapters, we can summarize the problems examined and identify conclusions in the following:

- The first problem is estimating the steady state KF gain from measurements in the absence of process and measurement noise statistics and we examined merit of correlations based methods to approach to the problem. In an off-line setting the estimation of noise covariance matrices, and the associated filter gain from correlations of measurements or innovations process from an arbitrary filter is theoretically feasible but lead to an ill-conditioned linear least square problem. In real applications, the right hand side of the least square problem has some uncertainty since it is constructed from sample correlation functions of the innovations process or measurements calculated using finite data. The accuracy of the sample correlation functions of innovations process is improved by using long data, however, due to fact that coefficient matrix is ill-conditioned, the stability of the solution being sensitive to the errors in correlations functions is examined using Discrete Picard Condition (DPC). Examinations showed that the correlations approaches do not satisfy the DPC, therefore, the estimates obtained from the classical least square solution are simply wrong. In this study we examined the merit of using Tikhonov's regularization to approach the ill-conditioned problems of correlations approaches. Numerical examinations showed that the noise covariance and the optimal filter gain estimates can be significantly improved by applying Tikhonov's regularization to ill-conditioned problems of correlations approaches.

- The second problem is on state estimation using a nominal model that represents the actual system. We examined an approach that takes the effects of uncertain parameters of the nominal model into account in state estimation using KF. In this approach the errors in the nominal model are approximated by fictitious noise and covariance of the fictitious noise is calculated using the stored data on the premise that the norm of discrepancy between correlation functions of the measurements and their estimates from the nominal model is minimum. Another approach examined approximates the system with an equivalent Kalman filter model. In this approach the filter gain is calculated using the stored data on the premise that the norm of measurement error of the filter is minimum. The fictitious noise and equivalent Kalman filter approaches are applicable in off-line conditions where stored measurement data is available. The fictitious noise approach leads to expressions that are more complex than the equivalent Kalman filter approach scheme, but the differences are not important when it comes down to computer implementation. Examinations showed that the state estimates from these two approaches are suboptimal, however, they both perform better than an arbitrary filter. The performance of the fictitious noise approach is depended on the length of the data since it requires the sample output correlations that are calculated from a finite length sequences. Therefore, when the sample correlations are not calculated with a good approximation, e.g. they are calculated from short data, the equivalent Kalman filter approach gives better state estimates for the same data.

- Additionally, the problem of state estimation using a nominal model is addressed in on-line operating conditions using EKF-based combined state and parameter estimation method. This method takes the uncertain parameters as part of the state vector and a combined parameter and state estimation problem is solved as a nonlinear estimation using extended KF (EKF). This strategy is simple in theory but is not trivial in applications when the model is large and the uncertain model parameters are too many. The EKF requires the computation of Jacobian of the augmented model and writing explicit state and measurement equations, which are impractical for systems with large models. A parametrization scheme for structural matrices (M, C, K) is presented that simplifies implementation of the EKF-based combined state estimation algorithm regardless of the size of the model and parameter vector. Examinations showed that the EKF-based combined parameter estimation approach is not robust to large uncertainties in initial parameter estimate and error covariance matrix when unknown parameter vector is large.

- The last problem is related to the use of Kalman filter as a fault detector. It is well known that the innovations process of the Kalman filter is white. When the system changes due to damage the innovations are no longer white and correlations of the innovations can be used to detect damage. A difficulty arises, however, when the statistics of unknown excitations and/or measurement noise fluctuate because the filter detects these changes also and it becomes necessary to differentiate what comes from damage and what does not. In this work, we showed that, theoretically, the correlations of the innovations from an arbitrary stable filter gain decrease with lag and asymptotically approach zero. However when the system changes, filter innovations do not vanish and asymptotically approach a value. We investigated if the correlation functions of the innovations evaluated at higher lags can be used to increase the relative sensitivity of damage over noise fluctuations. A modified whiteness test is introduced, that is insensitive to changes in the statistics of the disturbances and the measurement noise. The test is successfully applied to a damaged detection problem that is numerically simulated.

Bibliography

[1] C. F. Gauss. Theoria motus corporum coelestium. 1809.

[2] N. Wiener. *Extrapolation, Interpolation,and Smoothing of Stationary Time Series.* The M.I.T. Press,, Cambridge, MA, 1949.

[3] R. E. Kalman. A new approach to linear filtering and prediction problems. *ASME Journal of Basic Engineering*, 82:35–45, 1960.

[4] D. Bernal. Optimal discrete to continuous transfer for band limited inputs. *Journal of Engineering Mechanics, ASCE*, 133-12:1370–1377, 2007.

[5] R. Kalman, Y. Ho, and K. Narendra. Controllability of linear dynamical systems. *Contributions to Differential Equations*, 2:189–213, 1963.

[6] D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* Wiley-Interscience, 2006.

[7] D. G. Luenberger. An introduction to observer. *IEEE Trans. on Automatic Control*, 16:596–602, 1971.

[8] A. G. O. Mutambara. *Design and Analysis of Control Systems.* CRC Press, NY, 1999.

[9] B. Moore. On the flexibilty offered by state feedback in multivariable control beyond closed loop eigenvalue assignment. *IEEE Trans. Autom. Control*, 21:689–692, 1976.

[10] J. Kautsky, N. K. Hurvich, and P. V. Dooren. Robust pole assignment in linear state feedback. *International Journal of Control*, 41:1129–1155, 1985.

[11] W. Wonham. On pole assignment in multi-input controllable linear systems. *IEEE Transactions on Automatic Control*, 12(6):660 – 665, dec 1967.

[12] D. Simon. *Optimal state estimation: Kalman, H-infinity and nonlinear approaches.* John Wiley and Sons, Inc., Hoboken, New Jersey, 2006.

[13] P. Lancaster and L. Rodman. *Algebraic Riccati Equations.* Oxford University Press, 1995.

[14] T. Kailath. Innnovations approach to least-squares estimation - part i: Linear filtering in additive white noise. *IEEE Transactions on Automatic Control*, 13-6:646–655, 1968.

[15] R. K. Mehra. On the identification of variance and adaptive kalman filtering. *IEEE Transactions on Automatic Control*, 15:175–184, 1970.

[16] B. Carew and P. R. Belanger. Identification of optimum filter steady-state gain for systems with unknown noise covariances. *IEEE Transactions on Automatic Control*, 18:582–587, 1974.

[17] L. H. Son and B. D. O. Anderson. Design of kalman filters using signal model output statistics. *Proc.IEE*, 120-2:312–318, 1973.

[18] R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *ASME Journal of Basic Engineering*, 83:95–108, 1961.

[19] A. H. Jazwinski. *Stochastic process and filtrering theory*. Academic Press, New York, 1970.

[20] R. K. Mehra. Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 83:95–108, 1972.

[21] C. G. Hilborn and D. G. Lainiotis. Optimal estimation in the presence of unknown parameters. *IEEE Trans. Svst. Sci. Cybern*, SSC-5:38–43, 1969.

[22] R. L. Kashyap. Maximumlikelihood identification of stochastic linear systems. *IEEE Trans on AC.*, 15(1):25–34, 1970.

[23] K. A. Myers and B. D. Tapley. Adaptive sequential estimation with unknown noise statistics. *IEEE Trans. Auto. Cont.*, 21:520–523, 1976.

[24] H. Heffes. The effect of erroneous models on the kalman filter response. *IEEE Transactions on Automatic Control*, 11:541–543, 1966.

[25] C. Neethling and P. Young. Comments on identification of optimum filter steady-state gain for systems with unknown noise covariances. *IEEE Transactions on Automatic Control*, 19:623–625, 1974.

[26] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings. A new autocovariance least-squares method for estimating noise covariances. *Automatica*, 42(2):303–308, February 2006.

[27] B. M. Akesson, J. B. Jùrgensen, N. K. Poulsen, and S. B. Jùrgensen. A generalized autoco-variance least-squares method for kalman filter tuning. *Journal of Process Control*, 42(2), June 2007.

[28] Y. Bulut, D. Vines-Cavanaugh, and D. Bernal. Process and measurement noise estimation for kalman filtering. *IMAC XXVIII, A Conference and Exposition on Structural Dynamics*, February, 2010.

[29] J. Dunik, M. Simandl, and O. Straka. Methods for estimating state and measurement noise covariance matrices: Aspects and comparison. *15th IFAC Symposium on System Identification, 2009*, 15-1, June 2009.

[30] B. D. Anderson and J. B. Moore. *Optimal Filtering.* Prentice-Hall, Englewood Cliffs, NJ, 1979.

[31] P. C. Hansen. *Regularization Tools-Report.* Numerical Algorithms 46, 2007.

[32] A. Neumaier. Solving ill-conditioned and singular linear systems: A tutorial on regularization. *SIAM Review*, 40-3:636–666, 1998.

[33] L. M. Rojas. *Regularization of Large Scale Ill-Conditioned Least Square Problems.* Center for Research on Parallel Computation, Rice University, 1996.

[34] C. R. Vogel. *Solving ill-conditioned linear systems using the conjugate gradient method.* Report, Dept. of Mathematical Sciences, Montana State University, 1987.

[35] P. C. Hansen. The l-curve and its use in the numerical treatment of inverse problems. *Computational Inverse Problems in Electrocardiology*, 2000.

[36] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, pages 49–95, 1996.

[37] L. Xie, Y. C. Soh, and C. E. de Souza. Robust kalman filtering for uncertain discrete-time systems. *Automatic Control, IEEE Transactions on*, 39(6):1310 –1314, june 1994.

[38] I. R. Petersen and A. V. Savkin. Robust kalman filtering for signals and systems with large uncertainties. *Birkhausen*, 1999.

[39] N. K. Sinha and B. Kuszta. *Modeling and identification of dynamic systems.* Van Nostrand Reinhold Co., New York, 1983.

[40] J. N. Yang, H. Huang, and S. Lin. Sequential non-linear least-square estimation for damage identification of structures. *International Journal of Non-Linear Mechanics*, 41(1):124 – 140, 2006.

[41] H. Cox. On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Transactions on Automatic Control,*, 9:5 – 12, 1964.

[42] H. El Sherief and N. Sinha. Bootstrap estimation of parameters and states of linear multivariable systems. *IEEE Transactions on Automatic Control*, 24(2):340–343, Feb. 1979.

[43] R. E. Kopp and R. J. Orford. Linear regression applied to system identification for adaptive control systems. *AIAA Journal*, 1-10:2300–2306, 1963.

[44] K. K. Kim, J. T. Lee, D. K. Yu, and Y. S. Park. Parameter estimation of noisy passive telemetry sensor system using unscented kalman filter. *Future Generation Communication and Networking*, 2:433–438, 2007.

[45] T. Chen, J. Morris, and E. Martin. Particle filters for state and parameter estimation in batch processes. *Journal of Process Control*, 15:665–673, 2005.

[46] X. Liu, P. J. Escamilla-Ambrosio, and N. A. J. Lieven. Extended kalman filtering for the detection of damage in linear mechanical structures. *Journal of Sound and Vibration*, 325(4-5):1023 – 1046, 2009.

[47] E. W. Ericwan, R. V. D. Merwe, and A. T. Nelson. Dual estimation and the unscented transformation. In *Neural Information Processing Systems*, pages 666–672. MIT Press, 2000.

[48] E. A. Wan and A. T. Nelson. Dual kalman filtering methods for nonlinear prediction, smoothing, and estimation. In *Advances in Neural Information Processing Systems 9*, 1997.

[49] P. S. Maybeck. *Combined state and parameter estimation for on-line applications.* PhD thesis, Massachusetts Institute of Technology. Dept. of Aeronautics and Astronautics, 1972.

[50] S. Zhong and A. Abur. Combined state estimation and measurement calibration. *IEEE Transactions on Power Systems*, 20(1):458–465, Feb. 2005.

[51] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation.* John Wiley and Sons, Inc., Hoboken, New Jersey, 2001.

[52] J. Aßfalg and F. Allgöwer. Fault diagnosis of nonlinear systems using structured augmented state models. *International Journal of Automation and Computing*, 4:141–148, 2007.

[53] P. Moireau, D. Chapelle, and P. Le Tallec. Joint state and parameter estimation for distributed mechanical systems. *Computer Methods in Applied Mechanics and Engineering*, 197:659–677, 2008.

[54] L. Ljung. Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control,*, 24(1):36 – 50, 1979.

[55] V. Panuska. A new form of the extended kalman filter for parameter estimation in linear systems. volume 18, pages 927 –932, dec. 1979.

[56] V. Panuska. A new form of the extended kalman filter for parameter estimation in linear systems with correlated noise. *Automatic Control, IEEE Transactions on*, 25(2):229 – 235, apr. 1980.

[57] M. Wu and A. W. Smyth. Application of the unscented kalman filter for real-time nonlinear structural system identification. *Struct. Control Hlth.,*, 14:971–990, 2007.

[58] S. F. Schmidt. Application of state-space methods to navigation problems. *Advances in Control Systems*, 3:293–340, 1966.

[59] A. H. Jazwinski. Adaptive filtering. *Automatica*, 3:475–485, 1969.

[60] J. Juang, C. Chen, and M. Phan. Estimation of kalman filter gain from output residuals. *Journal of Guidance, Control, and Dynamics*, 16(5):903–908, 1993.

[61] C. R. Farrar and K. Worden. An introduction to structural health monitoring. *Royal Society of London Transactions Series A*, 365-1851:303–315, 2007.

[62] A. Hera. *Instantaneous Modal Parameters and Their Applications to Structural Health Monitoring*. PhD thesis, Worcester Institute of Technology, MA, 2005.

[63] C. J. Hellier. *Handbook of Nondestructive Evaluation*. McGraw-Hill, New York, 2001.

[64] S. W. Doebling, C. R. Farrar, and M. B. Prime and D. W. Shevitz. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review. *Technical Report, Los Alamos National Lab*, May 1996.

[65] H. Sohn, C. R. Farrar, F. M. Hemez, D. D. Shunk, D.W. Stinemates, and B. R. Nadler. A review of structural health monitoring literature: 1996-2001. *Los Alamos National Laboratory, NM.*, Report: LA-13976-MS, 2003.

[66] R. K. Mehra and I. Peshon. An innovations approach to fault detection and diagnosis in dynamic systems. *Automatica*, 7:637–640, 1971.

[67] R. V. Beard. Failure accommodation in linear system through self reorganization. *Ph.D. thesis, Massachussets Institute of Technologya*, 1971.

[68] H. L. Jones. Failure detection in linear systems. *Ph.D. thesis, Massachussets Institute of Technology*, 1973.

[69] S. Liberatore, J. L. Speyer, and A. C. Hsu. Application of a fault detection filter to structural health monitoring. *Automatica*, pages 1199–1209, 2006.

[70] R. C. Montgomery and A. K. Caglayan. Application of a fault detection filter to structural health monitoring. *AIAA 12th Aerospace Sciences Meeting*, 12th Aerospace Sciences Meeting, Washington, DC:1199–1209, 1974.

[71] C. P. Fritzen, S. Seibold, and D. Buchen. Application of filter techniques for damage identification in linear and nonlinear mechanical structures. *Proc. of the 13th International Modal Analysis Conference*, pages 1874–1881, 1995.

[72] S. Seibold and K. Weinert. A time domain method for the localization of cracks in rotors journal of sound and vibration. *Journal of Sound and Vibration*, 195-1:57–73, 1995.

[73] C. P. Fritzen and G. Mengelkamp. A kalman filter approach to the detection of structural damage. *Proc. of the 4th Int. Workshop on Structural H. Mon.*, pages 1275–1284, 2003.

[74] A. M. Yan, P. De Boe, and J.C. Golinval. Structural damage diagnosis by kalman model based on stochastic subspace identification. *Structural Health Monitoring*, 3-2:103–119, 2004.

[75] B. Peeters and G. D. Roeck. One-year monitoring of the z24-bridge: environmental effects versus damage events. *Earthquake Engineering - Structural Dynamics*, 30-2:149–171, 2000.

[76] J. T. Kim, C. B. Yun, and J. H. Yi. Temperature effects on frequency-based damage detection in plate-girder bridges. *KSCE Journal of Civil Engineering*, 7-6:725–733, 2008.

[77] Q. W. Zhang, L. C. Fan, and W. C. Yuan. Traffic-induced variability in dynamic properties of cable-stayed bridge. *Earthquake Eng. Struct. Dyn.*, 31:2015–2021, 2002.

[78] R. Ruotolo and C. Surace. Using svd to detect damage in structures with different operational conditions. *Journal of Sound and Vibration*, 226-3:425–439, 1999.

[79] S. Vanlanduit, E. Parloo, B. Cauberghe, P. Guillaume, and P. Verboven. A robust singular value decomposition for damage detection under changing operating conditions and structural uncertainties. *Journal of Sound and Vibration*, 284:1033–1050, 2005.

[80] H. Sohn. Effects of environmental and operational variability on structural health monitoring. *Philosophical Transactions of the Royal Society*, 365:539–560, 2007.

[81] L. Ljung. *System Identification Theory for the User.* Prentice Hall PTR, NJ, 1999.

[82] G. M. Jenkins and D. G. Watts. *Spectral analysis and its applications.* Holden-Day, London, 1969.

[83] M. H. Clifford, C. M. Hurvich, J. S. Simonoff, and S. L. Zeger. Variance estimation for sample autocovariances direct and resampling approaches. *Australian and New Zealand Journal of Statistics*, 231:23–43, 2008.

[84] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* Mc-Graw Hill, 1984.

[85] P. Peebles. *Probability Random variables and Random Signal Principles.* Mc-Graw Hill, 2000.

Appendix A

An Introduction to Random Signals and Noise

In this appendix we list some concepts from probability theory that are adapted from the original texts that cover the material [84, 85].

A.1 Random Variables

A random variable is a number X assigned to every outcome ζ of an experiment Γ. Let X be a random variable defined on the real space $\mathbb{R}$. The (cumulative) probability distribution function (CDF) $F(x)$ associates, to each real value x, the probability of the occurrence $X \leq x$, namely

$$F : \mathbb{R} \rightarrow \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{A.1.1}$$

$$F(x) = P\left\{X \leq x\right\} \tag{A.1.2}$$

$F(x)$ is a monotonous, increasing function, and can be continuous or discrete depending on X has continuous or discrete values, respectively. The resulting random variable X of the experiment Γ must satisfy the following conditions,

- The set $(X \le x)$ is an event for every x.
- $\underset{x \to +\infty}{lim} F(x) = 1$; $\underset{x \to -\infty}{lim} F(x) = 0$;

The derivative

$$f(x) = \frac{dF(x)}{dx} \tag{A.1.3}$$

$$f(x)dx = P\left\{x \le X \le x + dx\right\} \tag{A.1.4}$$

of $F(x)$ is called probability density function(PDF) of the random variable X.

To characterize a random variable X, one can use the moments of this variable. The first moment is called mean value or expected value. The second central moment is called variance and is denoted $var(X) = \sigma_x^2$ where is σ_x is the standard deviation, namely.

$$Expected\,Value: \quad E(X) = \int_{-\infty}^{+\infty} x f(x)dx \tag{A.1.5}$$

$$k^{th}\ moment: \quad E(X^k) = \int_{-\infty}^{+\infty} x^k f(x)dx \tag{A.1.6}$$

$$k^{th}\ central\ moment: \quad E\left((X - E(X))^k\right) = \int_{-\infty}^{+\infty} (x - E(x))^k f(x)dx \tag{A.1.7}$$

Full description of a random variable requires the characterization of all its moments. But from a practical point of view third and higher moments are not used because they cannot be computed or derived easily. If the X is of discrete type taking the values x_i with probabilities p_i then

$$f(x) = \sum_i p_i \delta(x - x_i) \tag{A.1.8}$$

where $\delta(x)$ is delta dirac function and $p_i = P\{x = x_i\}$. The definition of moments involves a discrete sum :

$$E(X^k) = \sum_i x_i^k p_i \delta(x - x_i) \tag{A.1.9}$$

Gaussian Random Variable:

A random variable X is called normal or Gaussian if its probability density is the shifted or/and scaled Gaussian function, namely

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma_x^2} \tag{A.1.10}$$

where μ and σ_x denote mean and standard deviation of the random variable X, respectively. This is bell-shape curve, symmetrical about the line $x = \mu$ and the corresponding cumulative distribution function (CDF) is given by

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \tag{A.1.11}$$

The normal (Gaussian) random variables are entirely defined by the first and second moments. The distribution functions of Gaussian random variables for a set of $\{\mu,\, \sigma_x\}$

are depicted in Fig.A.1.

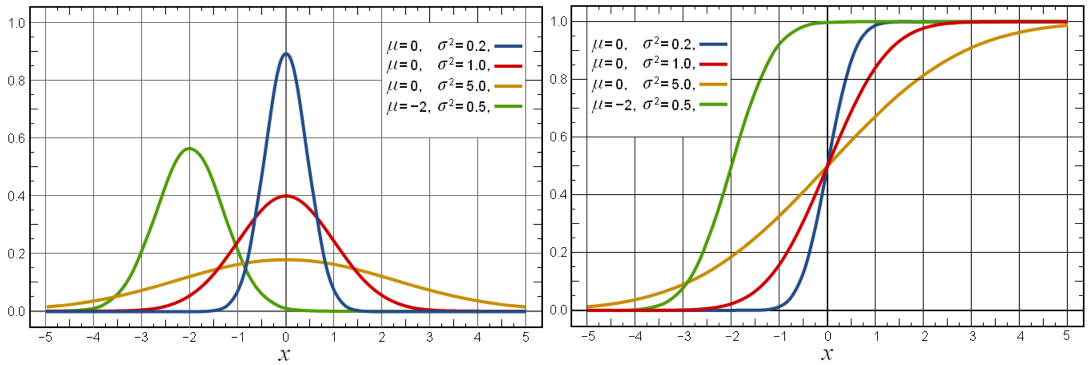

Figure A.1: Normal (Gaussian) distribution, Left: Probability density function, Right: Cumulative distribution function

Uniform Distributed Random Variable:

A random variable X is called uniform between x_1 and x_2 if its probability density is constant in the interval $(x_1,\ x_2)$ and zero elsewhere, namely

$$f(x) = \begin{cases} \dfrac{1}{x_2 - x_1} & x_1 \leq x \leq x_2 \\ 0 & otherwise \end{cases} \tag{A.1.12}$$

Typical distribution functions of uniform distributed random variables are depicted in Fig.A.2.

A.2 Multivariate Random Variables

A multivariate random variable is a vector $X = [X_1, \ldots, X_q]^T$ whose components are random variables on the same probability space. Let X be defined on the real space

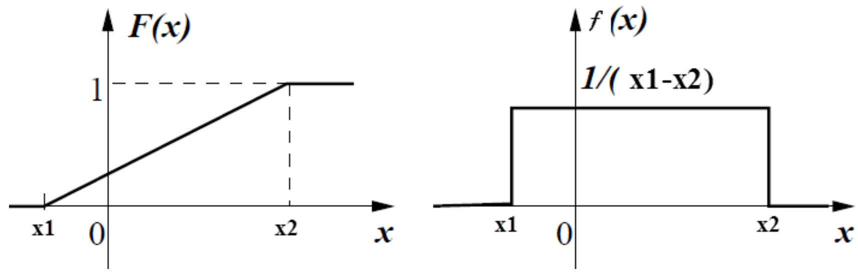

Figure A.2: Uniform distribution, Right: Probability density function, Left: Cumulative distribution function.

$\mathbb{R}^q$.

Probability distribution function:

$$F(x_1, \ldots, x_q) = P\{X_1 < x_1 \; and \; X_2 < x_1 \; and \cdots and \; X_q < x_q\} \tag{A.2.1}$$

Probability density function:

$$f(x_1, \ldots, x_q) = \frac{\partial^q F(x_1, \ldots, x_q)}{\partial x_1 \ldots \partial x_q} \tag{A.2.2}$$

Moments:

Only the first moment (that is the mean vector) and the second central moment (that is the covariance matrix) are presented in the following:

$$Expected\,Value: \quad E(X) = [\; E(X_1), \;\; \cdots, \;\; E(X_q) \;]^T \tag{A.2.3}$$

$$Covariance: \quad Cov_x = E\left[(X - E(X)(X - E(X))^T\right] \tag{A.2.4}$$

The component $Cov_x(i,j)$ at row i and column j of this covariance matrix verifies:

$$Cov_x(i,j) = \int_{\mathbb{R}^2} (x_i - E(X_i)(x_j - E(X_j))^T dF(x_i, x_j) \tag{A.2.5}$$

The covariance matrix is definite, positive and symmetric.

Independence:

Two random variables X_1 and X_2 are independent if and only if:

$$F(x_1, x_2) = F(x_1)F(x_2) \tag{A.2.6}$$

A.3 Random Signals

Given a random variable X, the random signal $x(t)$ is a function of time t such that for each given t, $x(t)$ corresponds to a value (a sample) of X. Let $w(t)$ be a random signal, then:

Moments for random signal:

$$First\,Moment: \qquad m(t) = E[w(t)] \tag{A.3.1}$$

$$Second\,Moment: \qquad \psi_{ww}(t,\tau) = E[w(t)w(t+\tau)^T] \tag{A.3.2}$$

The second moment of a random signal is called the auto-correlation function.

Stationarity:

A random signal is defined to be stationary if its mean is constant $m(t) = m$ and if its auto-correlation function depends only on τ, namely $\psi_{ww}(t,\tau) = \psi_{ww}(\tau)$.

Power Spectral Density:

Stationary random signals can also be characterized by their frequency domain representation called Power Spectral Density (PSD). The spectrum in the s-plane of a stationary random signal is the Laplace transform of the auto-correlation function, namely

$$\Phi_{ww}(s) = \mathcal{L}(\psi_{ww}(\tau)) = \int_{-\infty}^{\infty} \psi_{ww}(\tau)e^{-\tau s}d\tau \tag{A.3.3}$$

White Noise:

White noise is a random signal with with a flat power spectral density. In other words, the signal contains equal power within a fixed bandwidth at any center frequency. In statistical sense, a time series is characterized as white noise if it is a sequence of serially uncorrelated random variables; whose auto-correlation function is proportional to a Dirac function. From a practical point of view, it is not possible to simulate, on a numerical computer, a perfect continuous-time white noise characterized by a finite PSD. However an approximation is often used, which consists in holding band limited frequency range over a sample data. This approximation corresponds to the band-limited white-noise.

Appendix B

Parametrization of Structural Matrices

The calculation of the Jacobian of the parametric state space models makes use of the EKF for combined state and parameter estimation impractical for large size models. Here we introduce a parameterization scheme based on dynamical system matrices in order to calculate the Jacobian. Let the equations of equilibrium of a linear dynamical system be written as

$$M(\theta^m)\ddot{y}(t) + C_s(\theta^c)\dot{y}(t) + K(\theta^k)y(t) = b_2 u(t) \tag{B.1}$$

where M, K and C_s are global mass, stiffness and damping matrices and θ^m, $\theta^k$and θ^c are finite dimensional parameter vectors which contain parameters related to the mass, stiffness and damping properties of the elements of the structure model, respectively. We define global parameter vector as

$$\theta = \begin{bmatrix} \theta^m & \theta^k & \theta^c \end{bmatrix}^T \tag{B.2}$$

and write the structural matrices which consist of f elements as follows

$$M(\theta^m) = \sum_{j=1}^{f} \theta_j^m M_j = \theta_1^m M_1 + \theta_1^m M_2 + \cdots + \theta_p^m M_p + \cdots + \theta_f^m M_f \tag{B.3}$$

$$K(\theta^k) = \sum_{j=1}^{f} \theta_j^k K_j = \theta_1^k K_1 + \theta_2^k K_2 + \cdots + \theta_p^k K_p + \cdots + \theta_f^k K_f \tag{B.4}$$

$$C_s(\theta^c) = \sum_{j=1}^{f} \theta_j^c (C_s)_j = \theta_1^c (C_s)_1 + \cdots + \theta_p^c (C_s)_p + \cdots + \theta_j^c (C_s)_f \tag{B.5}$$

where M_j, K_j and $(C)_j$ denote the stiffness, mass and damping matrices for the jth element in the global coordinate system and p is the number of the unknown parameters in the structural system matrices. The parameters are in order such that 1 to p refer to unknown parameters of structural matrices and $p+1$ to n refer to known parameters. One can partition these matrices as follows

$$M(\theta^m) = M_P + M_R \tag{B.6}$$

$$K(\theta^k) = K_P + K_R \tag{B.7}$$

$$C_s(\theta^c) = C_P + C_R \tag{B.8}$$

where subscript P and R denote the partition due to unknown and known parameters in the structural system matrices, respectively, namely

$$M_P = \sum_{j=1}^{p} \theta_j M_j \qquad M_R = \sum_{j=p+1}^{f} \theta_j M_j \tag{B.9}$$

$$K_P = \sum_{j=1}^{p} \theta_j^k K_j \qquad K_R = \sum_{j=p+1}^{f} \theta_j^k K_j \tag{B.10}$$

$$C_P = \sum_{j=1}^{p} \theta_j^c (C_s)_j \qquad C_R = \sum_{j=p+1}^{f} \theta_j^c (C_s)_j \tag{B.11}$$

We extend the parameterization of structural matrices to the state space form system matrices in the following. We recall the the state space matrices $A(\theta)$ and $B(\theta)$, which are given in the following form,

$$A(\theta) = \begin{bmatrix} 0 & I \\ -M(\theta^m)^{-1} K(\theta^k) & -M(\theta^m)^{-1} C_s(\theta^c) \end{bmatrix} \tag{B.12}$$

$$B(\theta) = \begin{bmatrix} 0 \\ M(\theta^m)^{-1} b_2 \end{bmatrix} \tag{B.13}$$

Using the definitions in Eqs.B.6-B.8, one can partition the state transition matrix $A(\theta)$ in Eq.B.12 due to unknown and known parameters as follows,

$$A(\theta) = A_R + A_P \tag{B.14}$$

where

$$A_R = \begin{bmatrix} 0 & I \\ -M_R^{-1}K_R & -M_R^{-1}(C_s)_R \end{bmatrix} \tag{B.15}$$

$$A_P = \begin{bmatrix} 0 & 0 \\ -\frac{1}{\theta_1^m}\theta_1^k M_1^{-1}K_1 & -\frac{1}{\theta_1^m}\theta_1^k M_1^{-1}(C_s)_1 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & 0 \\ -\frac{1}{\theta_p^m}\theta_1^k M_1^{-1}K_1 & -\frac{1}{\theta_1^m}\theta_p^k M_p^{-1}(C_s)_p \end{bmatrix} \tag{B.16}$$

and state to input matrix, $B(\theta)$ is partitioned as follows,

$$B(\theta) = B_R + B_P \tag{B.17}$$

where

$$B_R = \begin{bmatrix} 0 \\ -M_R^{-1}b_2 \end{bmatrix} \tag{B.18}$$

$$B_P = \begin{bmatrix} 0 \\ -\frac{1}{\theta_p^m}M_1^{-1}b_2 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ -\frac{1}{\theta_p^m}M_p^{-1}b_2 \end{bmatrix} \tag{B.19}$$

The Jacobian of the augmented state space model in Eq.4.3.18 is presented in a closed form as follows;

$$\Delta(\hat{z}(t)) = \left.\frac{\partial(\dot{z}(t))}{\partial\theta}\right|_{z=\hat{z}(t)} = \begin{bmatrix} A(\hat{\theta}(t)) & D^m(\hat{\theta}(t)) & D^k(\hat{\theta}(t)) & D^c(\hat{\theta}(t)) \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{B.20}$$

where

$$D^m(\hat{\theta}(t)) \stackrel{def}{=} \left.\frac{\partial(A(\theta))\hat{x}}{\partial\theta^m}\right|_{\theta=\hat{\theta}} + \left.\frac{\partial(B(\theta))u(t)}{\partial\theta^m}\right|_{\theta=\hat{\theta}} = \begin{bmatrix} d_1^m & \ldots & d_p^m \end{bmatrix} \tag{B.21}$$

$$D^k(\hat{\theta}(t)) \stackrel{def}{=} \left.\frac{\partial(A(\theta))\hat{x}}{\partial\theta^k}\right|_{\theta=\hat{\theta}} = \begin{bmatrix} d_1^k & \ldots & d_p^k \end{bmatrix} \tag{B.22}$$

$$D^c(\hat{\theta}(t)) \stackrel{def}{=} \left.\frac{\partial(A(\theta))\hat{x}}{\partial\theta^c}\right|_{\theta=\hat{\theta}} = \begin{bmatrix} d_1^c & \ldots & d_p^c \end{bmatrix} \tag{B.23}$$

and

$$d_j^m = \begin{bmatrix} 0 & 0 \\ \frac{1}{(\hat{\theta}_j^m)^2}\hat{\theta}_j^k M_j^{-1} K_j & \frac{1}{(\hat{\theta}_j^m)^2}\hat{\theta}_j^c M_j^{-1} (C_s)_j \end{bmatrix} \hat{x} + \begin{bmatrix} 0 \\ \frac{1}{(\hat{\theta}_j^m)^2} M_j^{-1} b_2 \end{bmatrix} u(t) \tag{B.24}$$

$$d_j^k = \begin{bmatrix} 0 & 0 \\ -\frac{1}{\hat{\theta}_j^m} M_j^{-1} K_j & 0 \end{bmatrix} \hat{x} \tag{B.25}$$

$$d_j^c = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{\hat{\theta}_j^m} M_j^{-1} (C_s)_j \end{bmatrix} \hat{x} \tag{B.26}$$

d_j is a column vector with the size $nx1$ where we recall that n is the order of the system. The size of Jacobian matrix $F(\hat{z}(t))$ is $(n+3p)x(n+3p)$. It's apparent that the calculation of d_j requires only global structural matrices of the elements that the parameters are calculated, namely M_j, K_j and $(C)_j$. After one have the a priori estimate of the augmented state, the d_j's for each parameter can be calculated, then Jacobian matrix $\Delta(\hat{z}(t))$ is constructed from B.20.