Summer 1991

Integrated System Identification and Adaptive State Estimation for Control of Flexible Space Structures

Chung-Wen Chen
Old Dominion University

INTEGRATED SYSTEM IDENTIFICATION AND ADAPTIVE STATE ESTIMATION FOR CONTROL OF FLEXIBLE SPACE STRUCTURES

Chung-Wen Chen

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
in the Department of
Mechanical Engineering and Mechanics
Old Dominion University

August, 1991

Approved:

Jen-Kuang Huang (director)

Jer-Nan Juang (co-director)

Thomas Alberts

Gene Hou

Kyong Lim

ABSTRACT

INTEGRATED SYSTEM IDENTIFICATION AND ADAPTIVE STATE ESTIMATION FOR CONTROL OF FLEXIBLE SPACE STRUCTURES

Chung-Wen Chen

Department of Mechanical Engineering and Mechanics
Old Dominion University

Accurate state information is crucial for control of flexible space structures in which the state feedback strategy is used. The performance of a state estimator relies on accurate knowledge about both the system and its disturbances, which are represented by system model and noise covariances respectively. For flexible space structures, due to their great flexibility, obtaining good models from ground testing is not possible. In addition, the characteristics of the systems in operation may vary due to temperature gradient, reorientation, and deterioration of material, etc. Moreover, the disturbances during operation are usually not known. Therefore, adaptive methods for system identification and state estimation are desirable for control of flexible space structures. This dissertation solves the state estimation problem under three situations: having system model and noise covariances, having system model but no noise covariances, having neither system model nor noise covariances. Recursive least-squares techniques, which require no

initial knowledge of the system and noises, are used to identify a matrix polynomial model of the system, then a state space model and the corresponding optimal steady state Kalman filter gain are calculated from the coefficients of the identified matrix polynomial model. The derived methods are suitable for on-board adaptive applications. Experimental example is included to validate the derivations.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Jen-Kuang Huang, and co-advisor, Dr. Jer-Nan Juang, from whom I have received attentive instruction, care, motivation and guidance. The members of my doctoral committee, Dr. Thomas Alberts, Dr. Gene Hou, and Dr. Kyong Lim are appreciated for their constructive suggestions in revising the dissertation.

This dissertation will not be possible without my parents' love and support. Special acknowledgment goes to my wife, Gloria, for her constant love, patience and encouragement. It was not easy for her to take care of the family alone during the last few years of my Ph.D. program.

This work was supported by the Department of Mechanical Engineering and Mechanics of Old Dominion University, and NASA Langley Research Center under the Grant NAG1-830.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 1

INTRODUCTION

1.1 Background

During the last several decades, modern control theories have been widely established, and the applications of control theories in aerospace, military and civil industries have increased tremendously. This can be attributed to several reasons. Among them the most important is the advent of powerful and inexpensive modern digital computers. Without the powerful capabilities in calculation, data transmission and memory of computers, most modern control theories will be just too complicated to realize in practical applications. Without the manufacturing techniques which have substantially reduced the price of computers from their original astronomical figures, the applications would certainly be greatly restricted. Besides, the easy access to personal computers or work stations, which every university, research institute and laboratory can afford, has greatly helped researchers in developing, validating and testing control theories. Indeed, powerful, inexpensive digital computers have made some earlier engineers' dreams become possible.

The other major reason for the rapid development in control technology is due to necessity. Nowadays many physical plants constructed by human beings are very complicated and large-scaled and require high performance, such as huge chemical plants, jumbo airliners and large flexible space structures. To make these systems function satisfactorily, resorting to modern control techniques becomes inevitable.

The central problem of control is how to determine appropriate control forces so that the systems can accomplish the prescribed requirements. Mathematically, the solution of the problem is represented by a set of equations called "control laws". In practice, the controllers (computers) programmed with control laws play the role of determining control forces based on the situation of the systems. For controlling linear systems, state feedback strategy is a common technique, where state information, which represents the situation of the systems, is used for determining control forces. However, in general, state information cannot be measured directly and the quantities which can be measured are only some functions of states. Therefore, the technique of reconstructing state information from measurement data becomes an indispensable part of state feedback control.

In reconstructing state information from input/output data, one faces some inherent difficulties. First, measurement data are almost inevitably contaminated by noise due to imperfect instruments. Second, the number of output sensors is usually less than the number of the states of interest, which implies that the output measurement at a single moment alone is not sufficient to determine the state at that moment. Previous data, therefore, should somehow be utilized. But since systems are usually affected by unpredictable noises, uncertainty is introduced into the relation between previous data and the current state. Because the performance of feedback control relies on the accuracy of the reconstructed state, an effective state reconstruction method under all these difficulties plays a critical role in satisfactory control. State estimation is the technique of reconstructing state information from input and noise-corrupted output data, which is, therefore, very important in realizing feedback control theories to real world applications.

Generally speaking, state estimation belongs to a larger field call "filtering", which is basically a process of extracting signals of interest from some relevant data. For several decades researchers in the fields of economics, communication, guidance, navigation and control remain highly interested in the problem of filtering because accurate information and signals in useful form always play crucial roles in these fields. "Filter" in general is a generic term representing either a substantial hardware or an algorithm programmed in a computer to realize the signal extracting process. There are at least three different situations which require conducting filtering. First, the signals of interest are contaminated by noise; therefore, the sole objective of filtering is to clean up the noise. Second, the signal at hand is noise-free but not of interest by itself, and the objective is to deduce some other information from it. Third, the objective is to deduce information from the available signals which are contaminated by noise—a combination of the above two. State estimation problems belong to the third category.

The idea of estimating unknown quantities from observed data can be traced back to Gauss in the early 1800's. In his astronomical studies, in which planet and comet motion was studied using telescopic measurement data, he invented the famous least-square technique.[1] In more recent times, in order to solve the problem of controlling antiaircraft firing systems during World War II, Wiener developed the Wiener filter.[2] This filter is still widely acclaimed today as a cornerstone of modern estimation theory. Kolmogorov (1939) applied mean-square theory to discrete stochastic processes.[2] In 1960 Kalman published his famous method for sequential state estimation of discrete systems, known as the Kalman filter, using state space formulation.[3] This paper is a landmark in modern control history. Two years later, a version of the Kalman filter for continuous system was published.[4] After this a large number of papers appeared in the literature concerned with the properties, the modifications under some different assumptions, and the applications to some specific problems of the filter.[5]

The great body of literatures reveals the importance of the Kalman filter; however, at the same time it reveals the existence of some unsatisfactory features

as well. A well-known limitation in applying the conventional Kalman filter is its requirement of a priori knowledge about the system state space model and the covariances of process and measurement noises. This information, in practice, is either only partially known or totally unknown. Another limitation of the conventional Kalman filter is that it can neither adjust itself to trace a changing environment, nor can it correct the error caused by incorrect a priori information. In a sense, the conventional Kalman filter works as an open loop system, because the filter evolves according to preset formulas during operations and the estimation error never affects the filter itself. Moreover, after reaching its steady state, the filter "sleeps". That is, no matter how big the estimation error could be due to whatever reasons, the filter just remains unchanged. A phenomenon called filter divergence could happen.[6-9]

If a system model is known but its noise statistics are not, one should either use a set of guessed values of the noise covariances or conduct a systematic method to estimate the noise covariances or the Kalman filter gain before using the Kalman filter.[10-14] To enable the filter to acquire information during operation to improve upon the a priori assumptions that were made at the outset leads to the topic of adaptive filtering.[15-17] Adaptive Kalman filtering uses Kalman filter structure and modifies the philosophy of computing the filter gain so that the filter can monitor estimation error and feed back the information to improve its performance.[17,18] However, most existing adaptive Kalman filters and methods of estimating noise covariances or filter gain are complicated and not suitable for on-line application. Furthermore, most adaptive filtering methods are derived under the assumption that the system model is accurately known. Adaptive Kalman filtering for unknown or uncertain systems is seldom addressed.

The problem of adaptive Kalman filtering for unknown or uncertain systems is more complicated. Goodwin introduced some adaptive methods for state estimation under uncertain system models,[15] where the state vector is augmented to include uncertain system parameters. By this way the system parameters and

state can be estimated at the same time. However, nonlinear state estimation techniques, such as the extended Kalman filter, have to be used in the above approach because the system model becomes nonlinear due to state augmentation. For nonlinear estimation, the system is usually linearized at each estimated state, which is very time-consuming especially for a large order system. Moreover, the convergence of the estimate is not guaranteed.

To solve the problem of state estimation under unknown models and noise covariances, in general, system models should be identified before state estimation can be carried out. Therefore, it is a compound problem of system identification and state estimation. System identification, also called *"modeling"*, or *"time series analysis"*, is a technique of obtaining appropriate mathematical models for dynamical systems from their input/output data.[19–23] It is important in a diversity of fields such as engineering, economics, statistics, and physical science. In controls field, especially in adaptive controls where systems to be controlled are uncertain or time-varying, system identification is indispensible.

In conducting system identification, the format of a mathematical model is selected first, and then the parameters of the model are chosen to minimize a defined cost function which indicates the fitness of the model to the input/output data. In choosing a model format, the matrix polynomial model has the advantage of having a linear relation between the parameters and the input/output data; thus least-squares and its variations can be used to identify the model without requiring any a priori knowledge about the system. The matrix polynomial model is a system equation whose z-domain expression is a matrix polynomial equation. On the other hand, system identification using a state space model is a nonlinear optimization problem, which is more difficult to work with. However, a state space model is desirable for state estimation and control purpose, because most of the control theories are developed in state space representation.

1.2 Problem Description and Objective

For active control of large flexible space structures, accurate models and accurate state information of the systems are very important. However, because of its great flexibility and gravitational load, an accurate system model of a large flexible space structure can not be obtained from ground testing. In addition, the characteristics of a large flexible structure in space can vary due to such factors as temperature gradient induced by shadowing, reorientation of a large antenna or deterioration of material. The system model needs to be updated frequently. Moreover, space structures are working under unknown noises. Good performance of a state estimator relies both on an accurate system model and an accurate estimate of the noise statistics (or the optimal filter gain). Hence, for better control performance, strategies of on-line system identification and adaptive state estimation are required.

The objective of this dissertation is to develop effective *integrated* system identification and state estimation algorithm for on-line application in the control of linear systems, and in particular for flexible space structures.

1.3 Dissertation Outline

Aiming at the ultimate objective, this dissertation poses the problem of state estimation by dividing it into three stages according to the degree of complexity and investigates them progressively. These three stages are linear state estimation with both system model and noise covariances known, linear state estimation with system model known but noise covariances unknown, and linear estimation with both system model and noise covariances unknown.

Chapter 2 provides some background material. First, a brief review of three optimal estimation methods is given. These basic methods are cornerstones of many modern estimation methods. Next, a state space model of a linear system and the problem of state estimation are also described. Then, the Kalman filter is briefly introduced. The relations between parameter estimation and state estimation are also discussed.

Chapter 3 investigates the first stage of the problem, that is, state estimation under full a priori information. The projection filter is developed for this purpose based on parameter estimation techniques. This filter provides an alternative to the Kalman filter and the derivation process helps in understanding the characteristics of state estimation problems. The discussion of the relation between the projection filter and the Kalman filter also provides a better understanding of the Kalman filter. The relationship between the projection filter and the correlation canceler, which is frequently used in the field of signal processing,[24] is also discussed.

Chapters 4 and 5 investigate the state estimation problem under unknown noise covariances. Due to the lack of noise information the state estimation problem is solved under a deterministic framework, i.e., without requiring statistical information about input/output data and noise. Four methods are developed, one in Chapter 4 and three in Chapter 5. The approach in Chapter 4 is fundamentally different from all the others; therefore, it is separated as an independent chapter. In Chapter 4 the least-squares and recursive least-squares methods are used. The fading memory least-squares technique is also used to deal with the effect of unknown process noise. The relation between the least-squares filter and the Kalman filter is also discussed, which provides a deeper understanding of the Kalman filter as well as the estimation problem.

All three methods introduced in Chapter 5 aim at estimating optimal steady state Kalman filter gain directly. The relation between a state space model and a

matrix polynomial model is derived through Kalman filter formulation. The first method utilizes the equivalence of optimal linear output predictions made by different models and takes advantage of the inherent properties of models in different structures. The AutoRegressive with eXogeneous input (ARX) model[22] which is a matrix polynomial model, can yield linear prediction adaptively without requiring any initial knowledge about the system, but it can not offer state information. On the other hand, Kalman filter can provide state information, but it requires a priori information. This first method makes these two filters work together. The ARX model is used to generate one- to r-step-ahead linear predictions as references. Then a gain for the Kalman filter is chosen such that the Kalman filter will generate approximately the same linear output prediction. The second method uses the relation between the coefficients of the ARX model and the state space parameters and the optimal steady state Kalman filter gain. From the relationship the Kalman filter gain can be calculated from the estimated ARX coefficients, and thus the state estimation can be conducted. The third method utilizes the fact that the optimal Kalman filter residual is white and derives the optimal Kalman filter gain from a whitening filter.

Chapter 6 investigates the most challenging stage of state estimation problems, that is, state estimation under unknown system models and noise statistics. Under this situation, one faces a compound problem of system identification and state estimation. Two methods are developed in this chapter. The first method identifies a state space model and the corresponding Kalman filter gain simultaneously. The input/output data are first used to identify an ARX model of large order, taking advantage of its property of requiring no initial information. Then the eigensystem realization algorithm (ERA),[25] a system identification method, is used to decompose the coefficient of the ARX model into state space parameters and the corresponding Kalman filter gain. The second method is based on the projection filter theory developed in Chapter 3, where an ARX model of relatively smaller order is identified based on input/output data first. Using the coefficients of the identified ARX model, the system Markov parameters can be calculated.

Again, the ERA is used to decompose the Markov parameters into state space parameters. After having a system model, the corresponding optimal Kalman filter gain can be obtained using methods provided in Chapter 5.

Chapter 7 gives an experimental example to illustrate the feasibility of the methods derived in the previous chapters. A ten-bay structure located in NASA Langley Research Center is considered. System identification and state estimation are conducted and the reconstructed output is compared to the real output.

Finally, Chapter 8 provides conclusions and prospects for the extension of this research.

Chapter 2

OPTIMAL LINEAR ESTIMATION

2.1 Introduction

Estimation is a technique of deriving some interested quantities based on relevant data, usually conducted when the interested quantities cannot be measured either directly or correctly.[26] It is an information extraction process, extracting useful information out from available data. From another point of view, estimation theories also provide strategies of combining information obtained from several different sources to yield more accurate information.

Optimal estimation requires the estimates, the results of performing estimation, to be optimal under a certain optimality criterion. Therefore, based on different philosophies behind the optimality criteria there are different methods. Basically, there are major three different optimal estimation methods: least-squares, maximum likelihood and Bayesian estimations.[26–28] They are briefly introduced below.

Some terms which are often used in estimation literature are briefly explained here: an *unbiased* estimate is one in which expected value is the same as that of the quantity being estimated; a *minimum variance (unbiased)* estimate has the property that its error variance is less than or equal to that of any other unbiased estimate; a *consistent* estimate is one which converges to the true value as the number of measurements increases to infinity. In general, the desirable result of estimation is unbiased, consistent and has minimum variance.

2.2 Some Basic Optimal Estimation Methods

In this section, a general example is used to explain a few basic concepts in optimal estimation methods. Assume that the set of p measurements, Y, can be expressed as a linear combination of the columns of H plus a random, additive measurement error, w. That is, the measurement is modeled as

$$Y = Hx + w \tag{2.1}$$

where Y is a $p \times 1$ vector, x an $n \times 1$ vector, H a $p \times n$ matrix and w a $p \times 1$ vector. Assume $p > n$, i.e., the measurement set contains redundant information.

2.2.1 Least-Squares Estimation

In least-squares estimation, one chooses the value of estimate which minimizes the sum of squares of the deviations, $y_i - \hat{y}_i$.; i.e., it minimizes a scalar cost function, J, defined by

$$J = (Y - H\hat{x})^T(Y - H\hat{x}). \tag{2.2}$$

Hereafter, the symbol "$\wedge$" denotes the estimated value. The resulting least-squares

estimate found by setting $\delta J/\delta \hat{x} = 0$ is

$$\hat{x} = (H^T H)^{-1} H^T Y \tag{2.3}$$

If, instead, one seeks to minimize a weighted sum of squares of deviation,

$$J = (Y - H\hat{x})^T W (Y - H\hat{x}), \tag{2.4}$$

where W is a $p \times p$ symmetric, positive definite weighting matrix, the weighted-least-squares estimate becomes

$$\hat{x} = (H^T W H)^{-1} H^T W Y. \tag{2.5}$$

The results of least-squares have no direct probabilistic interpretation; they were derived through deterministic argument only. Consequently, the least-squares estimates may be preferred to other estimates when there is no basis for assigning probability density functions to x and Y.

2.2.2 Maximum Likelihood Estimation

The connotation of "maximum likelihood" is a setting in which nothing is known a priori about the unknown quantity, x, but there is prior information on the measurement Y itself. Thus, x is deterministic and Y is stochastic. The conditional probability density function (PDF) of Y given the unknown x, $f(Y \mid x)$, contains information about x. If it can be computed, x may be estimated according to the maximum-likelihood estimation criterion, which can be stated as follows. Given a measurement Y, the maximum-likelihood estimate $\hat{x}_{ML}$ is the value of x which maximizes $f(Y \mid x)$, the likelihood that x resulted in the observed Y.

In the above example, the conditional PDF of Y, conditioned on a given value of x, is the density for w centered around Hx. Assuming w is a zero-mean,

Gaussian distributed observation with covariance matrix R, we have

$$f(Y \mid x) = \frac{1}{(2\pi)^{\frac{p}{2}} \mid R \mid^{\frac{1}{2}}} exp\left[-\frac{1}{2}(Y - Hx)^T R^{-1}(Y - Hx)\right] \tag{2.6}$$

This is called a likelihood function. Since $ln(\cdot)$ is a monotonically increasing function, maximizing $f(Y \mid x)$ is equivalent to maximizing $ln(f(Y \mid x))$, which in turn is equivalent maximizing the exponent in the bracket of (2.6). This is equivalent to minimizing the cost function in (2.4), with weighting matrix replaced by R^{-1}. The result, of course, is as given in (2.5) with W replaced by R^{-1}. This approach provides a probabilistic basis for choosing the weighting matrix.

2.2.3 Baysian Estimation

In Bayesian estimation, statistic models are available for both parameter x and measurement Y, and one seeks the a posteriori conditional probability density function, $f(x \mid Y)$, since it contains all the statistic information of interest. In general, $f(x \mid Y)$ is evaluated as (Bayes' theorem)

$$f(x \mid Y) = \frac{f(Y \mid x)f(x)}{f(Y)} \tag{2.7}$$

where $f(x)$ is the a priori probability density function of x, and $f(Y)$ is the probability density function of the measurements Y. Depending on the criterion of optimality, one can compute estimate $\hat{x}$ from $f(x \mid Y)$. If the object is to find a generalized minimum variance Bayes' estimate, that is, to minimize the cost functional

$$J = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x - \hat{x})^T S(x - \hat{x})f(x \mid Y)dx_1 dx_2 \cdots dx_n \tag{2.8}$$

where S is an arbitrary, positive semidefinite matrix, we simply set $\delta J/\delta\hat{x} = 0$ to find that

$$\hat{x} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}xf(x \mid Y)dx_1 dx_2 \cdots dx_n = E[x \mid Y], \tag{2.9}$$

which is the conditional mean estimate, and is independent of S. The symbol E[·] denotes the expectation operation. Equation (2.8) has the characteristic structure

$$J = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(\tilde{x}) f(x \mid Y) dx_1 dx_2 \cdots dx_n, \tag{2.10}$$

where $\tilde{x} = x - \hat{x}$ is the estimation error and $L(\tilde{x})$ is a scalar loss function of $\tilde{x}$.

The result given in (2.9) holds for a wide variety of loss functions in addition to that used in (2.8). Equation (2.9) appears very simple, but the simplicity of its appearance belies the difficulty of its use. In fact, the computation of the conditional mean $E(x \mid Y)$ may be an intractable problem. Worse than this, to find $E(x \mid Y)$ we require $f(x \mid Y)$, which may not be known. In some cases, we may know only its first and second moments.

For the example as shown in (2.1), assuming gaussian distributions for x and w, the result of evaluating $E[x \mid Y]$ in (2.9) is

$$\hat{x} = (P_0^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} Y, \tag{2.11}$$

where P_0 is the a priori covariance matrix of x.

Comparing the various estimation just discussed, we note that if there is little or no a priori information, P_0^{-1} is very small and (2.11) becomes (2.5) with W replaced by R^{-1}. And if we argue that all measurement errors are uncorrelated (i.e., R is a diagonal matrix) and all errors have equal variance (i.e., $R = \sigma^2 \times I_p$), then (2.5) reduces to (2.3).

2.3 A State Space Model and State Estimation

In this section, the general linear state estimation problem of dynamical systems is described. Due to the growing importance of digital computers in practi-

cal applications of system identification, state estimation and control theories, the problem in this dissertation is confined to discrete-time linear systems.

A finite-dimensional, linear, discrete-time, time-invariant stochastic dynamic system can be represented by a state space model

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{2.12}$$

$$y_k = Cx_k + v_k, \tag{2.13}$$

where x is an $n \times 1$ state vector, u an $m \times 1$ input vector, and y a $p \times 1$ measurement or output vector. Matrices A, B, and C are the system matrix, input matrix and output matrix, respectively. The sequence $\{w_k\}$ is the process (input) noise, and the sequence $\{v_k\}$ is the measurement (output) noise. The integer k is the sample indicator. This model is used throughout the dissertation, except in some cases where the input term Bu_k is omitted for simplicity.

Given a set of input and output data recorded from the beginning to the current moment, the state estimation problem involves finding the "best" estimate of the current state under some pre-defined optimality criterion using all the knowledge available about the system and noises.

2.4 Optimal Kalman Filter for State Estimation

The Kalman filter is a natural extension of the estimation methods discussed above. This point will be clearly shown in the later chapters. The Kalman filter for discrete systems with stationary, white process and measurement noises which are not correlated to each other can be summarized as follows:[28]

a. Initial Conditions:

$$E[x_0] = \hat{x}_0, \qquad E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] = P_0$$

b. Prediction (Extrapolation):

$$\hat{x}_k^- = A\hat{x}_{k-1}^+ + Bu_{k-1} \tag{2.14}$$

$$P_k^- = AP_{k-1}^+A^T + Q \tag{2.15}$$

c. Measurement Update:

$$\hat{x}_k^+ = \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-) = (I_n - K_kC)\hat{x}_k^- + K_ky_k \tag{2.16}$$

$$P_k^+ = (I_n - K_kC)P_k^- \tag{2.17}$$

$$K_k = P_k^-C^T(CP_k^-C^T + R)^{-1} \tag{2.18}$$

where Q and R are the covariances of process and measurement noises respectively, $\hat{x}$ the estimated state vector, P the corresponding estimation error covariance matrix, I_n the n-dimensional identity matrix, K_k the Kalman filter gain and the superscripts $-$ and $+$ distinguish the estimates before and after taking account of the current measurement data respectively.

The inner operation of Kalman filtering can be explained as follows. Given the state, x_{k-1}, at time $k-1$ and its corresponding error covariance, P_{k-1}^+, the Kalman filter propagates the state and the error covariance to the next moment k ((2.14) and (2.15)) using the system model, and the results are x_k^- and P_k^- respectively. This procedure is called *"prediction"* or *"extrapolation"*, because the current state is calculated based on previous data. Upon the arrival of the measurement y_k at time k, there are two sources of information about the state at time k: the propagated state with its error covariance and the new measurement with measurement noise covariance. The measurement is related to the state through measurement equation (2.13). Using a minimum-mean-square estimation error criterion, the Kalman filter provides a method of combining these two sources of information into an optimal estimate of state x_k. This is done by adding a modifying term to the predicted value, where the modifying term is computed by pre-multiplying the output prediction error (the difference between the real and the predicted measurements) with a weighting matrix. This weighting matrix is

called the optimal Kalman filter gain, and is given by (2.18). This procedure is called *"measurement update"*. After measurement update, the next prediction can be made, and so on. By this way the Kalman filter can use data recursively to yield optimal estimated state. There is no need to keep the record of previous data.

2.5 State Estimation and Parameter Estimation

In a sense, state estimation is also a kind of parameter estimation problem because both of them estimate unknown quantities based on relevant data. However, state estimation is different from general parameter estimation problems in two aspects. First, in general parameter estimation the objective is to estimate some unknown parameters, which is constant or, in some special cases, slowly time-varying. However, for a state estimation problem, the objective is to estimate the state vector, which evolves rapidly through time. Second, in general parameter estimation, redundant data directly related to the parameters under estimating are used to determine the values of the parameters. However for state estimation problem, only one measurement directly related to the state under estimating is available. Due to these differences state estimation is essentially a more difficult problem than the general parameter estimation problems. The difficulty is many-fold, which can be briefly stated as follows:

Because of sensor hardware limitations, the number of measurements is usually less than the number of state of interest. Mathematically, this means that the dimension of the measurement vector is smaller than that of the state vector. Hence, for each time step, the measurement vector at that moment alone is not sufficient to determine the corresponding state vector uniquely because the number of unknowns is more than the number of equations. Consequently, previous

measurements are utilized to determine the current state. However, previous data are not directly related to the current state. To relate them the system model should be utilized. However, due to the effect of process noise, previous data are less reliable when compared to the current one in terms of bearing the information about the current state. Therefore, any effective state estimation method should properly weigh the previous and current data based on the knowledge of process and measurement noises.

Chapter 3

LINEAR STATE ESTIMATION WITH SUFFICIENT A PRIORI INFORMATION

3.1 Introduction

In this chapter state estimation of linear time-invariant dynamical systems with sufficient a priori information (i.e., with the knowledge of system model, noise covariances and initial conditions) is investigated. This is the first stage of the state estimation problem and is the simplest case among all the cases this dissertation aims to solve. With the above assumptions, the conventional Kalman filter is already an optimal solution for state estimation. Although it appears that there is no need to investigate this solved problem further, the above problem is analyzed from a different point of view and provides new insight.

In this chapter the state estimation problem is solved from a view point different from the conventional Kalman filter, and, as a result, an alternative method called *the projection filter* is developed. When compared to the Kalman filter, this alternative method has both advantages and disadvantages. The derivation of the

projection filter provides good insight regarding the relation between the classical estimation theory and the Kalman filter. The studies also provide the background knowledge for the subsequent research; therefore, the effort is worthy.

The state space model of the system studied in this chapter is

$$x_{k+1} = Ax_k + w_k \tag{3.1}$$

$$y_k = Cx_k + v_k, \tag{3.2}$$

where all the symbols are defined in section 2.4. Both the process noise sequence, $\{w_k\}$, and the measurement noise sequence, $\{v_k\}$, are assumed to be Gaussian, zero-mean, white and stationary with covariance matrices Q and R, respectively. These two sequences are also assumed statistically uncorrelated with each other, i.e., $E[w_i v_j^T] = 0$ for any i and j.

Compared with the model described in (2.12), (3.1) does not have the input term, Bu_k. Because the system parameter set $[A,\ B,\ C]$ and the input force $\{u_k\}$ are known, the effect of the term Bu_k on the output is also a known deterministic quantity. If the term Bu_k is included in the system model, one can always subtract its influence out from the output. Therefore, it makes no difference conceptually if the input term is omitted in the beginning.

Some extreme cases of the Kalman filter are briefly discussed in Section 3.2, providing a background for comparison with the projection filter derived later. In Section 3.3 the projection filters for systems under various conditions are derived based on parameter estimation theories. Section 3.4 discusses the relation between the projection filter and the correlation canceler, an often used processor in signal processing. Section 3.5 addresses the relation between the projection filter and the Kalman filter. Numerical examples are provided in Section 3.6 to validate the derivation of the projection filter and to compare the projection filter with the Kalman filter under various situation.

3.2 Special Cases of Kalman Filter

This section discusses three special cases of Kalman filtering. Although these special cases are either trivial or unlikely to exist in practice, they are of theoretical interest and hence are included for completeness.

3.2.1 Kalman Filter for Noise Free Systems

If the system is noise free ($Q = 0$ and $R = 0$) with non-zero initial error covariance ($P_0 \neq 0$), the Kalman filter gain (see (2.18)) becomes

$$K_k = P_k^- C^T (C P_k^- C^T)^{-1}. \tag{3.3}$$

Note that $CK_k = I_p$; therefore, K_k is actually a weighted pseudo-inverse of C and P_k^- is the weighting matrix. Premultiplying (2.16) by C gives

$$C\hat{x}_k^+ = C(I_n - K_k C)\hat{x}_k^- + CK_k y_k = y_k, \tag{3.4}$$

which implies that the estimate satisfies the measurement equation exactly for each time step. Define the a posteriori estimation error e_k^+ by

$$e_k^+ = x_k - \hat{x}_k^+. \tag{3.5}$$

By system model (3.1), (3.2) and filter equation (2.16), it easily follows that

$$e_{k+1}^+ = (I_n - K_k C) A e_k^+. \tag{3.6}$$

Matrix $(I_n - K_k C)$ has p zero eigenvalues and $n - p$ unit eigenvalues if C has rank p, hence, $(I_n - K_k C)A$ is a stable matrix if A is stable. Thus, based on the dynamics in (3.6), e_k^+ will converge to zero asymptotically. Equivalently, based on (2.15) and (2.17) where A represents a stable system and $Q = 0$, it can be seen

that the Kalman filter state error covariance will converge to zero in steady state ($P_\infty^+ = P_\infty^- = 0$). When P_k^- is approaching zero the matrix inversion in (3.3) is near singular; therefore, to prevent numerical difficulty a threshold value of P_k^- should be set so that when P_k^- is smaller than the threshold, the filter gain is set to be zero thereafter.

3.2.2 Kalman Filter for Systems without Process Noise

If the system has no process noise but has measurement noise ($Q = 0$ and $R \neq 0$), again based on (2.15) and (2.17) the Kalman filter has a zero steady state error covariance ($P_\infty^+ = P_\infty^- = 0$) and a zero steady state Kalman gain ($K_\infty = 0$). In this case the threshold value of P_k^- is not required since the existence of R prevents the singularity in the matrix inversion. The zero error covariance implies that the estimate is perfect in the steady state. The optimal steady state estimate becomes

$$\hat{x}_k^+ = \hat{x}_k^- = A\hat{x}_{k-1}^-,$$

which indicates that in steady state the Kalman filter ignores the noise-corrupted measurement and relies solely on system model to predict state information. In this case the prediction is the same as the estimation in steady state.

3.2.3 Kalman Filter for Systems without Measurement Noise

If the system has process noise but has no measurement noise ($Q \neq 0$ and $R = 0$), the Kalman filter is calculated based on (3.3), (2.15) and (2.17) but does not vanish in the steady state. Hence the filter gain and the error covariance have non-zero steady state values. In other words, the estimation can never be perfect except in a trivial case when matrix C is square and nonsingular.

As discussed in Section 2.4, the optimal Kalman estimation of the current state is formed by adding a term to modify the predicted state (see (2.16)). The modifying term is the residual premultiplied by the Kalman filter gain, and the predicted state is obtained by propagating the optimal estimate of the last time step through system model. Kalman filter gain, in a sense, indicates the weight of the residual in the filter. It provides a way to fuse the information about the current state comeing from two sources: prediction and measurement. If the process noise is strong but the measurement noise is weak, the measurement should be more reliable than the prediction which is made by ignoring process noise. In this case the residual bears more significant information which can not be ignored, and the filter gain will be relatively large. On the contrary, if the measurement noise is strong but the process noise is weak, the residual can be regarded as mainly caused by the measurement noise and thus is less important. The filter gain will be smaller. In the extreme case of no process noise as described in Section 3.2.2, the residual in steady state is nothing but measurement noise and should be totally ignored. Therefore the filter gain is zero.

3.3 Projection Filters

Projection filter is a linear operator which projects (or transforms, or maps) measurement vector from a finite measurement space to a state space such that the image of the projection is an optimal estimate of the current state under some optimality criterion. In other words, by defining a measurement vector $Y_{q,k}$ as

$$Y_{q,k} = \begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+1} \end{bmatrix}, \tag{3.7}$$

the projection filter is the matrix F_q such that

$$\hat{x}_k = F_q Y_{q,k}, \tag{3.8}$$

which provides an optimal estimate $\hat{x}_k$ of the current state x_k. In the above equations q denotes the number of successive previous measurements, including the current one, contained in the measurement vector. The transformation matrix F_q is called the *projection filter* of *order* q. Here the generic term "filter" is used to represent the data processing procedure which receives measurement as input and produces the information interested as output. In the following subsections, projection filters for systems under various conditions are discussed.

3.3.1 Projection Filter for Noise-Free Systems

Define an observability-type matrix H_q, called measurement matrix, as

$$H_q = \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+1} \end{bmatrix} \tag{3.9}$$

with dimension $pq \times n$; here p is the dimension of output and, again, q denotes the number of measurement used. It is obvious that

$$Y_{q,k} = H_q x_k. \tag{3.10}$$

This linear equation suggests a least-square solution of state x_k, that is,

$$\hat{x}_k = H_q^{\dagger} Y_{q,k}, \tag{3.11}$$

where $H_q^{\dagger}$ is the pseudo-inverse of H_q. In this noise-free case, (3.10) is an exact equation, hence the estimation is perfect. The pseudo-inverse matrix $H_q^{\dagger}$ in this case is the projection filter for the system, i.e.,

$$F_q \triangleq H_q^{\dagger}. \tag{3.12}$$

Note that H_q and $H_q^{\dagger}$ are not functions of k. If the rank of H_q, denoted by n, is smaller than pq (i.e. H_q is a long and full-column-ranked matrix), then

$$F_q = H_q^{\dagger} = (H_q^T H_q)^{-1} H_q^T. \tag{3.13}$$

If H_q is a nonsingular square matrix of dimension n, then

$$F_q = H_q^{-1}. \tag{3.14}$$

Matrix H_q requires that q equals or exceeds a minimum number, denoted by q_{min}, to make itself full column-ranked. In other words, the projection filter requires at least q_{min} measurements to operate. If more measurements than necessary are used ($q > q_{min}$), the computational load will increase without improvement on the estimation, which suggests the optimal number of measurement used is q_{min}.

Compared to the Kalman filter for the same noise-free case, the projection filter does not need the initial values of state and its error covariance to initiate the estimation, while the Kalman filter does. The projection filter can also achieve perfect estimation once the required minimum number of measurements becomes available, while the Kalman filter takes more steps to converge to zero-error estimation if the initial values are poorly estimated. However, the projection filter can not produce estimation until the matrix H_q becomes full-column-ranked, while the Kalman filter can.

3.3.2 Projection Filter for Systems without Process Noise

If a system has measurement noise but has no process noise, from (3.1) and (3.2) the following equations can be derived:

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+1} \end{bmatrix} = \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+1} \end{bmatrix} x_k + \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-q+1} \end{bmatrix}, \tag{3.15}$$

or in short,

$$Y_{q,k} = H_q x_k + V_{q,k}, \tag{3.16}.$$

where the noise vector $V_{q,k}$ is Gaussain, zero-mean and the entries are not correlated with each other.

Because the system has no process noise, the state x_k is a deterministic quantity; therefore, a classical deterministic parameter estimation technique [26,27] can be applied in estimating the state x_k from data vector $Y_{q,k}$. The technique is briefly introduced as follows.

3.3.2.1 Deterministic Parameter Estimation

The assumptions and constraints of a deterministic parameter estimation problem are stated first. Given an observation model

$$y = Hx + v, \tag{3.17}$$

where y is a $p \times 1$ measurement vector, H a $p \times n$ known constant matrix with $p \geq n$, x an $n \times 1$ deterministic unknown vector and v a $p \times 1$ Gaussian, zero-mean noise vector with covariance matrix R_v, the problem is to find an estimate $\hat{x}$ of x such that the mean square estimation error, $E[(x-\hat{x})^T(x-\hat{x})]$, is minimized. We impose two constraints: (a) the estimate should be unbiased, and (b) $\hat{x}$ should be a linear transformation of the measurement vector (i.e., $\hat{x} = Fy$).

Using the above two constraints we obtain

$$x = E[\hat{x}] = E[Fy] = E[FHx + Fv] = FHx, \tag{3.18}$$

where the assumption $E[v] = 0$ is used. The mean square estimation error is

$$\begin{aligned} E[(x-\hat{x})^T(x-\hat{x})] &= E[(x-Fy)^T(x-Fy)] \\ &= E[(x-FHx-Fv)^T(x-FHx-Fv)] \\ &= E[(x-x-Fv)^T(x-x-Fv)] \\ &= E[\text{trace}(v^TF^TFv)] = E[\text{trace}(Fvv^TF^T)] \\ &= \text{trace}(FR_vF). \end{aligned} \tag{3.19}$$

The problem reduces to finding a matrix F which minimizes trace(FR_vF) subjected to the requirement (3.18). The solution is [27]

$$F = (H^T R_v^{-1} H)^{-1} H^T R_v^{-1}, \tag{3.20}$$

and

$$\hat{x} = (H^T R_v^{-1} H)^{-1} H^T R_v^{-1} y. \tag{3.21}$$

The corresponding estimation error covariance P is

$$\begin{aligned} P &= E[(x - \hat{x})(x - \hat{x})^T] = F R_v F^T \\ &= (H^T R_v^{-1} H)^{-1} H^T R_v^{-1} R_v R_v^{-1} H (H^T R_v^{-1} H)^{-1} \\ &= (H^T R_v^{-1} H)^{-1}. \end{aligned} \tag{3.22}$$

Equation (3.20) shows that F is actually a weighted pseudo-inverse of H with weighting matrix R_v, which is identical to the least-squares estimation introduced in Section 2.2.1 except that the weighting matrix has a physical meaning. In addition, (3.22) gives the error covariance of the estimate while the least-squares method doesn't.

3.3.2.2 Projection Filter Using Deterministic Parameter Estimation Technique

Since (3.16) takes the same form as (3.17), the technique of deterministic parameter estimation just shown can be applied directly to the state estimation problem. Consequently, the optimal estimate of x_k is

$$\hat{x}_k = (H_q^T \bar{R}^{-1} H_q)^{-1} H_q^T \bar{R}^{-1} Y_{q,k} \tag{3.23}$$

where $\bar{R} = E[V_{q,k} V_{q,k}^T] = R \otimes I_q$, and "$\otimes$" denotes the Kronecker product. The corresponding state error covariance is

$$P_k = (H_q^T \bar{R}^{-1} H_q)^{-1}, \tag{3.24}$$

which is a constant.

If the noise level of every element of the measurement vector is the same and can be characterized by its variance σ^2, in other words, the covariance matrix of the measurement noise can be written as $R = \sigma^2 \times I_p$, then (3.23) becomes

$$\hat{x}_k = (H_q^T H_q)^{-1} H_q^T Y_{q,k} = H_q^{\dagger} Y_{q,k}, \tag{3.25}$$

which is the same as (3.11) for the noise-free case. The corresponding error covariance becomes

$$P_k = \sigma^2 (H_q^T H_q)^{-1}. \tag{3.26}$$

If the covariance matrix of the measurement noise is not of the form $\sigma^2 \times I_p$, the projection filter is then a weighted pseudo-inverse of the observability matrix as shown in (3.23), with $\bar{R}^{-1}$ as its weighting matrix.

Compared to the Kalman filter under the same situation of no process noise, the projection filter with a small order q is less accurate because P_k in (3.26) is not zero for a small q, while the Kalman filter can achieve perfect estimation ($P_k^+ = 0$). It can approach perfect estimation if the order q becomes sufficiently large. However, the computational load increases as the order increases. On the other hand, the projection filter does not need the initial values of state and its error covariance to start the estimation.

3.3.3 Projection Filter for Systems with both Process and Measurement Noises

For a linear system with both process and measurement noises, from (3.1) and (3.2), the following equation can be derived:

$$x_{k-1} = A^{-1} x_k - A^{-1} w_{k-1},$$
$$y_{k-1} = CA^{-1} x_k - CA^{-1} w_{k-1} + v_{k-1},$$

$$\vdots$$

$$y_{k-q+1} = CA^{-q+1}x_k - \sum_{i=1}^{q-1} CA^{-q+i}w_{k-i} + v_{k-q+1}. \tag{3.27}$$

The above equations can be expressed in a matrix form as

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+2} \\ y_{k-q+1} \end{bmatrix} = \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+2} \\ CA^{-q+1} \end{bmatrix} x_k$$

$$- \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & CA^{-1} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & CA^{-q+2} & \cdots & CA^{-1} & 0 \\ 0 & CA^{-q+1} & \cdots & CA^{-2} & CA^{-1} \end{bmatrix} \begin{bmatrix} w_k \\ w_{k-1} \\ \vdots \\ w_{k-q+2} \\ w_{k-q+1} \end{bmatrix} + \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-q+2} \\ v_{k-q+1} \end{bmatrix}, \tag{3.28}$$

or compactly,

$$Y_{q,k} = H_q x_k - M_q W_{q,k} + V_{q,k} \tag{3.29}$$

where

$$W_{q,k} = [w_k^T, \cdots, w_{k-q+1}^T]^T$$
$$V_{q,k} = [v_k^T, \cdots, v_{k-q+1}^T]^T$$

and M_q is the coefficient matrix of $W_{q,k}$. Defining $\xi_{q,k} \equiv= -M_q W_{q,k} + V_{q,k}$, Equation (3.29) can be further reduced to

$$Y_{q,k} = H_q x_k + \xi_{q,k}. \tag{3.30}$$

Because the noise vector $\xi_{q,k}$ is a linear transformation of some independent Gaussian, zero-mean random vectors, it remains Gaussian and zero-mean. The covariance of $\xi_{q,k}$ is

$$\Sigma_k = Cov[\xi_{q,k}] = M_q \bar{Q} M_q^T + \bar{R} \tag{3.31}$$

where $\bar{Q} = Cov[W_{q,k}] = Q \otimes I_q$, and $\bar{R} = Cov[V_{q,k}] = R \otimes I_q$; $Cov[\,\cdot\,]$ denotes the covariance operation.

From (3.30), one can use the technique of random parameter estimation for linear equation to solve it. The random parameter estimation technique is therefore briefly introduced next.

3.3.3.1 Random Parameter Estimation

Consider a measurement model defined as

$$y = Hx + v, \tag{3.32}$$

where y is a $p \times 1$ measurement vector, H a $p \times n$ known constant matrix, x an $n \times 1$ unknown random vector with known autocorrelation matrix $E[xx^T] = \Phi_x$, and v a $p \times 1$ Gaussian, zero-mean noise vector with covariance matrix R_v. The cross-correlation of x and v, $E[xv^T] = \Phi_{xv}$, is assumed known. The problem is to find an optimal estimate $\hat{x}$ by using the criterion of least mean square error, subjected to the requirement of being a linear and unbiased estimation.

(1) Random Parameter Estimation for Zero-Mean Parameters

Suppose x is zero-mean, then y is also zero-mean. The correlation matrices Φ_x and Φ_{xv} become covariances matrices Ω_x and Ω_{xv}, respectively. The filtering form is assumed to be

$$\hat{x} = Fy, \tag{3.33}$$

where F is the desired filter. The estimation error e is defined as

$$e \triangleq x - \hat{x} = x - Fy. \tag{3.34}$$

To obtain an "optimal" filter, a scalar cost function $J(F)$ is defined first, which is the 2-norm of the error vector:

$$\begin{aligned} J(F) &= E[e^T e] = \text{trace}E[(x - Fy)(x - Fy)^T] \\ &= \text{trace}E[xx^T - xy^T F^T - Fyx^T + Fyy^T F^T]. \end{aligned} \tag{3.35}$$

The function is then minimized with respect to matrix F, which requires that

$$\frac{dJ}{dF} = 0.$$

Taking derivative of (3.35) with respect to F, we obtain

$$\begin{aligned} \frac{dJ}{dF} &= -2E[xy^T] + 2FE[yy^T] = 0, \\ FE[yy^T] &= E[xy^T], \\ F\Omega_y &= \Omega_{xy}, \end{aligned} \tag{3.36}$$

where Ω_y and Ω_{xy} are the auto- and cross-correlation matrices of y and between x and y, respectively. Assuming Ω_y is invertible, then the filter F becomes

$$\begin{aligned} F &= \Omega_{xy}\Omega_y^{-1} \\ &= E[x(x^TH^T + v^T)]E[(Hx + v)(x^TH^T + v^T)]^{-1} \\ &= (\Omega_x H^T + \Omega_{xv})(H\Omega_x H^T + H\Omega_{xv} + \Omega_{xv}^T H^T + R_v)^{-1}. \end{aligned} \tag{3.37}$$

The corresponding error covariance of the estimation is

$$\begin{aligned} P &= E[ee^T] \\ &= E[xx^T - xy^TF^T - Fyx^T + Fyy^TF^T] \\ &= E[xx^T] - E[xy^T]F^T - FE[yx^T] + E[xy^T]E[yy^T]^{-1}E[yy^T]F^T \\ &= \Omega_x - F\Omega_{yx} \\ &= \Omega_x - \Omega_{xy}\Omega_y^{-1}\Omega_{yx}. \end{aligned} \tag{3.38}$$

Consider a special case when $\Omega_{xv} = 0$, that is, x is not correlated with the noise v. The optimal estimate and its corresponding error covariance then become

$$\hat{x} = \Omega_x H^T(H\Omega_x H^T + R_v)^{-1}y, \tag{3.39}$$

$$P = \Omega_x - \Omega_x H^T(H\Omega_x H^T + R_v)^{-1}H\Omega_x. \tag{3.40}$$

By using matrix inversion lemma[26], the above equations can be further simplified as

$$\begin{aligned}\hat{x} &= \Omega_x H^T \left(R_v^{-1} - R_v^{-1}H(\Omega_x^{-1} + H^T R_v^{-1} H)^{-1} H^T R_v^{-1}\right) y \\ &= \Omega_x H^T R_v^{-1} - \Omega_x H^T R_v^{-1} H(\Omega_x^{-1} + H^T R_v^{-1} H)^{-1} H^T R_v^{-1} y \\ &= \Omega_x H^T R_v^{-1} - \left(\Omega_x(\Omega_x^{-1} + H^T R_v^{-1} H) - I_n\right) \\ &\quad \times (\Omega_x^{-1} + H^T R_v^{-1} H)^{-1} H^T R_v^{-1} y \\ &= (H^T R_v^{-1} H + \Omega_x^{-1})^{-1} H^T R_v^{-1} y, \end{aligned} \tag{3.41}$$

$$\begin{aligned} P &= \Omega_x - (H^T R_v^{-1} H + \Omega_x^{-1})^{-1} H^T R_v^{-1} H \Omega_x \\ &= (H^T R_v^{-1} H + \Omega_x^{-1})^{-1} (H^T R_v^{-1} H + \Omega_x^{-1} - H^T R_v^{-1} H)\Omega_x \\ &= (H^T R_v^{-1} H + \Omega_x^{-1})^{-1}. \end{aligned} \tag{3.42}$$

From these two equations, it can be clearly seen that when $\Omega_x^{-1} = 0$, that is, no a priori information about x is available, (3.41) and (3.42) reduce to the form of deterministic parameter estimation as shown in (3.21) and (3.22).

The formulations derived above assumed the filtering form as shown in (3.33). Next we check the bias of the estimation. For an unbiased estimate

$$\begin{aligned} E[e] &= E[x - \hat{x}] = E[x - Fy] \\ &= \bar{x} - FE[Hx + v] = (I_n - FH)\bar{x} = 0, \end{aligned} \tag{3.43}$$

where $\bar{x} \triangleq E[x]$. To satisfy (3.43), either FH should be an identity matrix, or matrix $I_n - FH$ should be singular with $\bar{x}$ in its null space, or $\bar{x}$ should be zero. In the case $\bar{x} = 0$ the requirement of unbiased estimation is satisfied. If x is not zero-mean, then y is not zero-mean also; hence, covariance matrices Ω_x and Ω_{xv} become correlation matrices Φ_x and Φ_{xv}, respectively. Writing FH as

$$FH = (\Phi_x H^T + \Phi_{xv})(H\Phi_x H^T + H\Phi_{xv} + \Phi_{xv}^T H^T + R_v)^{-1} H, \tag{3.44}$$

it is clear that in general FH is not an identity matrix (if $\Phi_{xv} = 0$ and $\Phi_x^{-1} = 0$ then $FH = I_n$), and it is unlikely that $\bar{x}$ is always in the null space of matrix

$I_n - FH$. This implies the estimation is biased even though it satisfies the least-mean-square criterion. This is because we have imposed the filtering form in (3.33). To obtain an unbiased estimation when $\bar{x} \neq 0$, the filter form should be modified.

(2) Random Parameter Estimation for Non-Zero-Mean Parameters

The optimal linear estimation is the mean of the conditional density function of the parameter x given data y. [6] The conditional mean for linear Gaussian measurement (3.39) is of the form [26,27]

$$E[x \mid y] = \bar{x} + \Omega_{xy}\Omega_y^{-1}(y - \bar{y}), \tag{3.45}$$

where $\bar{y} = E[y]$, Ω_{xy} and Ω_y are the cross-covariance of x and y and auto-covariance of y, respectively. Therefore,

$$\hat{x} = E[x \mid y] = \bar{x} + F(y - \bar{y}) \tag{3.46}$$

where

$$\begin{aligned} F &= \Omega_{xy}\Omega_y^{-1} \\ &= E[(x - \bar{x})(H(x - \bar{x}) + v)^T]E[(H(x - \bar{x}) + v)(H(x - \bar{x}) + v)^T]^{-1} \\ &= (\Omega_x H^T + \Omega_{xv})(H\Omega_x H^T + H\Omega_{xv} + \Omega_{xv}^T H^T + R_v)^{-1}, \end{aligned} \tag{3.47}$$

and $\bar{y} = H\bar{x}$. The corresponding error covariance is

$$\begin{aligned} P &= Cov[(x - \bar{x})] = Cov[(x - \bar{x}) - F(y - \bar{y})] \\ &= \Omega_x - F\Omega_{yx} - \Omega_{xy}F^T + F\Omega_y F^T \\ &= \Omega_x - F\Omega_{yx}. \end{aligned} \tag{3.48}$$

Note that (3.47) and (3.48) are identical with (3.36) and (3.37).

Taking expectation of (3.46) yields $E[\hat{x}] = \bar{x}$, which indicates the estimate is unbiased. Therefore, (3.46) and (3.48) provide an unbiased optimal estimate of the random parameter x and its corresponding error covariance P.

If parameter x and noise v are uncorrelated ($i.e., \Omega_{xv} = 0$), (3.46), (3.47) and (3.48) reduce to

$$\hat{x} = \bar{x} + \Omega_x H^T (H\Omega_x H^T + R)^{-1}(y - H\bar{x}) \tag{3.49}$$

$$P = \Omega_x - \Omega_x H^T (H\Omega_x H^T + R_v)^{-1} H\Omega_x \tag{3.50}$$

or alternatively, based on the same matrix inversion lemma used in the previous section we have

$$\hat{x} = \bar{x} + (H^T R_v^{-1} H + \Omega_x^{-1})^{-1} H^T R_v^{-1}(y - H\bar{x}) \tag{3.51}$$

$$P = (H^T R_v^{-1} H + \Omega_x^{-1})^{-1}. \tag{3.52}$$

If $\Omega_x^{-1} = 0$, i.e., no a priori information about x is available, then $\bar{x}$ and $\bar{y}$ vanish and (3.51) and (3.52) reduce to (3.21) and (3.22), respectively, of the deterministic parameter estimation case.

3.3.3.2 Projection Filter Using Random Parameter Estimation Technique

Since (3.30) has the same form as (3.32), the technique of random parameter estimation can be applied in the state estimation problem. For this purpose, the auto-covariance of x_k and the cross-covariance of x_k and $\xi_{q,k}$ should be calculated first. The noise $\xi_{q,k}$ is correlated with the state x_k in this case due to the existence of process noise. The covariance matrices can be derived as follows. Defining $x_{k-1}^* \triangleq x_{k-1} - \bar{x}_{k-1}$, then

$$\begin{aligned}\Omega_{x_k} &= Cov[x_k - \bar{x}_k] = Cov[Ax_{k-1}^* + w_{k-1}] \\ &= A\Omega_{x_{k-1}} A^T + Q_{k-1}\end{aligned} \tag{3.53}$$

$$\begin{aligned}\Omega_{x_k \xi_k} &= E[x_k^* \xi_{q,k}^T] = E[x_k^*(-M_q W_{q,k} + V_{q,k})^T] \\ &= -E[x_k^* W_{q,k}^T] M_q^T \\ &= -\left[E[x_k^* w_k^T], \cdots, E[x_k^* w_{k-q+1}^T]\right] M_q^T \\ &= -\left[0_n,\ Q_{k-1}, \cdots, A^{q-2} Q_{k-q+1}\right] M_q^T\end{aligned} \tag{3.54}$$

where $0_n = 0 \times I_n$, and the following relation is used,

$$E[x_k^* w_{k-p}^T] = E\left[(A^p x_{k-p}^* + \sum_{i=1}^{p} A^{i-1} w_{k-i}) w_{k-p}^T\right] = A^{p-1} Q_{k-p},$$

$p = 0, 1, \cdots, q-1$. Note that Ω_{x_k} can be computed recursively using (3.53) from the given initial Ω_{x_0} and will converge to a steady state value if the system is stable.

Because x_k and y_k are not necessarily zero-mean processes, based on the theory of random parameter estimation (3.46), the optimal estimated state for (3.30) should be

$$\begin{aligned}\hat{x}_k &= \bar{x}_k + F_{q,k}(Y_{q,k} - \bar{Y}_{q,k}) \\ &= \bar{x}_k + \Big[(\Omega_{x_k} H_q^T + \Omega_{x_k \xi_k})(H_q \Omega_{x_k} H_q^T \\ &\qquad + H_q \Omega_{x_k \xi_k} + \Omega_{x_k \xi_k}^T H_q^T + \Sigma_k)^{-1}\Big](Y_{q,k} - \bar{Y}_{q,k}) \end{aligned} \tag{3.55}$$

where $\bar{x}_k$ can be obtained by propagating the initial state

$$\bar{x}_k = A^k \bar{x}_0.$$

The corresponding error covariance is

$$P_k = \Omega_{x_k} - F_{q,k}(\Omega_{x_k} H_q^T + \Omega_{x_k \xi_k})^T. \tag{3.56}$$

If both the process and measurement noises are stationary with covariance matrices Q and R, respectively, then the noise covariance Σ_k and the cross-covariance matrix $\Omega_{x_k \xi_k}$ become time-invariant. Furthermore, if the system is stable, the state auto-covariance matrix Ω_{x_k} will converge to a steady state value Ω_x; then (3.55) and (3.56) become

$$\hat{x}_k = \bar{x}_k + (\Omega_x H_q^T + \Omega_{x\xi})(H_q \Omega_x H_q^T + H_q \Omega_{x\xi} + \Omega_{x\xi}^T H_q^T + \Sigma)^{-1}(Y_{q,k} - \bar{Y}_{q,k}), \tag{3.57}$$

$$P_k = \Omega_x - F_q(\Omega_x H_q^T + \Omega_{x\xi})^T, \tag{3.58}$$

where

$$\Omega_{x\xi} = E[x_k^* \xi_k^T] = -\left[0_n,\ Q,\ AQ, \cdots, A^{q-2}Q\right] M_q^T,$$
$$\Sigma = M_q \bar{Q} M_q + \bar{R}.$$

The projection filter F_q in (3.57) becomes time-invariant and is a function of q only.

3.4 The Relation Between Projection Filter and the Correlation Canceler

The concept of correlation cancelation plays a central role in the development of many optimum signal processing algorithms because a correlation canceler is also the best linear processor for estimating one signal from another. The concept is well-known in the information and data processing disciplines.[24]

Consider two zero-mean random vectors x and y of dimensions n and p, respectively. If x and y are correlated with each other in the sense that $\Omega_{xy} = E[xy^T] \neq 0$, then we would like to remove such correlations by means of a linear transformation of the form

$$e = x - Fy \tag{3.59}$$

where the $n \times p$ matrix F must be suitably chosen so that the new pair of vector (e, y) are no longer correlated with each other; that is, we require

$$\Omega_{ey} = E[ey^T] = 0 \tag{3.60}$$

Using (3.59), we obtain

$$\begin{aligned} \Omega_{ey} &= E[(x - Fy)y^T] = E[xy^T] - FE[yy^T] \\ &= \Omega_{xy} - F\Omega_y \end{aligned} \tag{3.61}$$

Then the condition $\Omega_{ey} = 0$ immediately implies that

$$F = \Omega_{xy}\Omega_y^{-1} = E[xy^T]E[yy^T]^{-1}. \tag{3.62}$$

Using $\Omega_{ey} = 0$, the covariance matrix of the resulting vector e is easily found to be

$$\begin{aligned}\Omega_e &= E[ee^T] = E[e(x^T - y^T F^T)] = \Omega_{ex} - \Omega_{ey}F^T \\ &= \Omega_{ex} = E[(x - Fy)x^T] = \Omega_x - F\Omega_{yx} = \Omega_x - \Omega_{xy}\Omega_y^{-1}\Omega_{yx}.\end{aligned} \tag{3.63}$$

The vector

$$\hat{x} = Fy = \Omega_{xy}\Omega_y^{-1}y, \tag{3.64}$$

obtained by linear processing the vector y through the matrix F is called the linear regression, or orthogonal projection, of x on the vector y. In a sense, $\hat{x}$ represents the best "copy", or estimate, of x that can be made on the basis of the vector y. Thus, the vector $e = x - Fy = x - \hat{x}$ may be treated as the estimation error. Actually, it is better to treat $\hat{x}$ not as an estimation of x but rather as an estimation of the part of x which is correlated with y. To elaborate this point, suppose that x consists of two parts,

$$x = x_1 + x_2,$$

such that x_1 is correlated with y, but x_2 is not. Then,

$$\Omega_{xy} = E[(x_1 + x_2)y^T] = \Omega_{x_1 y}$$

and therefore,

$$\hat{x} = \Omega_{xy}\Omega_y^{-1}y = \Omega_{x_1 y}\Omega_y^{-1}y = \hat{x}_1.$$

The vector $e = x - \hat{x} = (x_1 - \hat{x}_1) + x_2$ consists of the estimation error of the x_1-part plus the x_2-part. Both of these two terms are uncorrelated with y. The corelation cancellation may be summarized as follows: if x has a part x_1 which is correlated with y, then this part can be canceled as much as possible by using

a linear processor F to convert y into the best copy $\hat{x}_1$ of x_1 and subtracting it from x. The remainder is no longer correlated with y. The part x_2 of x which is uncorrelated with y remains entirely unaffected. It cannot be estimated in terms of y.

At this stage one can find (3.62), (3.63) and (3.64) are the same as the projection filter, (3.53), (3.52) and (3.51), respectively, if x and y are zero-mean. In other words, *the projection filter is also an optimal correlation canceler.* It can extract as much as possible the information about the current state from data. However, the projection filter additionally provides a method to deal with the signals which are not zero-mean.

3.5 A Relation Between Projection Filter and Kalman Filter

The projection filter and the Kalman filter are closely related. In fact, a projection filter of order q ($q \geq 2$) can be transformed to have a Kalman filter structure, and the recursive projection filter of order one is identical to the Kalman filter. The identity of the two filters can be proved by re-deriving the Kalman filter through the recursive projection filter of order one and will be proved later. But first, the transformation of the projection filter to the Kalman filter structure is derived.

The transformation is based on the concept of correlation cancellation introduced in the last section. Taking expectation of both sides of (3.30) one has

$$\bar{Y}_{q,k} = H_q \bar{x}_k, \tag{3.65}$$

because $\xi_{q,k}$ is zero-mean. Subtracting (3.65) from (3.30) on both sides yields

$$Y_{q,k} - \bar{Y}_{k,q} = H x_k^* + \xi_{q,k}, \tag{3.66}$$

where $x_k^* \triangleq x_k - \bar{x}_k$. The noise term $\xi_{q,k}$ is correlated with x_k^*, and the covariance is shown in (3.54). According to the correlation cancellation theory, $\xi_{q,k}$ can be divided into correlated and uncorrelated parts, namely,

$$\xi_{q,k} = H' x_k^* + \zeta_k, \tag{3.67}$$

where $H' x_k^*$ is the correlated part. The orthogonal projection, denoted by H', is

$$H' = \Omega_{\xi_k x_k} \Omega_{x_k}^{-1} \tag{3.68}$$

according to (3.62), where $\Omega_{\xi_k x_k} = E[\xi_{q,k} x_k^{*T}]$ and $\Omega_{x_k} = Cov[x_k^*]$. Introducing (3.67) and (3.68) into (3.66) yields

$$\begin{aligned} \tilde{Y}_{q,k} &\triangleq Y_{q,k} - \bar{Y}_{q,k} \\ &= (H_q + H') x_k^* + \zeta_k, \\ &= \tilde{H}_q x_k^* + \zeta_k. \end{aligned} \tag{3.69}$$

From (3.68) one can derive

$$\Omega_{\xi_k x_k} = H' \Omega_{x_k} \tag{3.70}$$

and

$$\Omega_{x_k \xi_k} = \Omega_{x_k} (H')^T. \tag{3.71}$$

The noise ζ_k in (3.69) is uncorrelated with x_k^*, and its covariance is (see (3.63))

$$\begin{aligned} \tilde{\Sigma}_k &\triangleq Cov[\zeta_k] \\ &= \Omega_{\xi_k} - H' \Omega_{x_k \xi_k} \\ &= \Sigma_k - H' \Omega_{x_k} (H')^T, \end{aligned} \tag{3.72}$$

where (3.31) and (3.71) are used. According to (3.55) the projection filter is

$$\begin{aligned} \hat{x}_k = \bar{x}_k + \Big[& (\Omega_{x_k} H_q^T + \Omega_{x_k \xi_k})(H_q \Omega_{x_k} H_q^T \\ & + H_q \Omega_{x_k \xi_k} + \Omega_{x_k \xi_k}^T H_q^T + \Sigma_k)^{-1} \Big] (Y_{q,k} - \bar{Y}_{q,k}). \end{aligned}$$

and by substituting (3.70) to (3.72) into the corresponding terms in (3.55), one obtain

$$\begin{aligned}\hat{x}_k &= \bar{x}_k + \Big[(\Omega_{x_k} H_q^T + \Omega_{x_k}(H')^T)(H_q \Omega_{x_k} H_q^T \\ &\qquad + H_q \Omega_{x_k}(H')^T + H' \Omega_{x_k} H_q^T + H' \Omega_{x_k}(H')^T + \tilde{\Sigma}_k)^{-1}\Big](Y_{q,k} - \bar{Y}_{q,k}) \\ &= \bar{x}_k + \Omega_{x_k} \tilde{H}_q^T \left(\tilde{H}_q \Omega_{x_k} \tilde{H}_q + \tilde{\Sigma}_k\right)^{-1} \tilde{Y}_{q,k}. \end{aligned} \tag{3.73}$$

Similarly, the corresponding error covariance (3.56) becomes

$$\begin{aligned} P_k &= \Omega_{x_k} - F_{q,k}(\Omega_{x_k} H_q^T + \Omega_{x_k \xi_k})^T \\ &= \Omega_{x_k} - F_{q,k}(H_q \Omega_{x_k} + H' \Omega_{x_k}) \\ &= (I - F_{q,k} \tilde{H}_q)\Omega_{x_k}. \end{aligned} \tag{3.74}$$

A close examination of (3.73) and (3.74) shows that they have a Kalman filter format, where $F_{q,k}$ is equivalent to the Kalman filter gain K_k, and $\tilde{H}_q$, $\tilde{\Sigma}_k$ and Ω_{x_k} are equivalent to C, R and P_k^-, respectively. Equations (3.73) and (3.74) can also be derived from (3.69) using (3.39) and (3.40) directly, because in (3.69) all the variables are zero-mean and the noise ζ_k is uncorrelated with x_k^*. In the above fashion, x_k^* is estimated first, and the final estimate is given by $\hat{x}_k = \bar{x}_k + x_k^*$.

Based on the above derivation, we can see that despite the apparent difference, the information extraction philosophies of projection and Kalman filters are actually the same. Yet there are still some differences.

The a priori values of the estimates ($\bar{x}_k$) and their corresponding error covariances (Ω_{x_k}) of the projection filter are either obtained by propagating from their initial values or set to be the steady state values. However, the counterparts in the Kalman filter ($\hat{x}_k^-$ and P_k^-) are conditional means and covariances, conditioned on all previous data. Using conditional means as the a priori estimates allows the Kalman filter to utilize all the data (from the beginning till the current moment) recursively in estimating the current state. On the contrary, a projection filter of order q uses only q's most recent data to do the same task. As a result, the Kalman

filter in general is more accurate than a projection filter of small order. Besides, the Kalman filter treats only one measurement at each step, while the projection filter of order q needs to treat a batch of q data. Therefore, computationally the Kalman filter is more efficient.

One might ask if the projection filter can somehow be modified to have the capability of utilizing all the data available so that it may produce the same result as the Kalman filter. The answer is yes. Because both filters are optimal linear filters, based on the same given conditions, they should be equivalent.

For a projection filter with order q, a number of q's most recent measurements should be kept in record at each step. The estimate made at each step does not take advantage of previous estimates. In other words, the estimation is totally based on the finite data in the current record. In order to use all measurements, one may increase the order as the time step increases. By thus doing, however, the computational load will soon become too heavy to bear in practice. Hence, from a computational standpoint, a recursive type of projection filter is preferable. The recursive projection filter is derived as follows.

Based on (3.2) and the assumption that state and measurement noise are uncorrelated, the projection filter of order one and its corresponding estimation error covariance, according to (3.57) and (3.58), are

$$\hat{x}_k = \bar{x}_k + \Omega_{x_k} C^T (C\Omega_{x_k} C^T + R)^{-1} (y_k - C\bar{x}_k), \tag{3.75}$$

$$P_k^+ = \Omega_{x_k} - \Omega_{x_k} C^T (C\Omega_{x_k} C^T + R)^{-1} C\Omega_{x_k}, \tag{3.76}$$

where $\bar{x}_k$ is the unconditional mean of x_k. In order to take advantage of previous estimations, conditional mean of state and conditional state covariance should be used. Suppose the optimal estimate of the state $\hat{x}_{k-1}$ at time $k-1$ and its corresponding error covariance P_{k-1} have been obtained using the projection filter of order $k-1$ based on all the data from y_1 to y_{k-1}. Because the projection filter also calculates the conditional mean of the parameters under estimating,

conditional on all the data used, [26] $\hat{x}_{k-1}$ can be written as

$$\hat{x}_{k-1} = E[x_{k-1} \mid Y_{k-1}]$$

where $Y_{k-1} = \{y_1, \cdots, y_{k-1}\}$.

Using the estimate at $k-1$ and the system model, a prediction of the state at k and its corresponding error covariance can be made. That is,

$$\begin{aligned} \bar{x}_k = E[x_k \mid Y_{k-1}] &= E[Ax_{k-1} + w_{k-1} \mid Y_{k-1}] \\ &= E[Ax_{k-1} \mid Y_{k-1}] \\ &= A\hat{x}_{k-1}. \end{aligned} \tag{3.77}$$

Since $\bar{x}_k$ is the conditional mean, conditioned on Y_{k-1}, it is also the a priori estimate of x_k. Similarly, the conditional state covariance can be calculated as

$$\begin{aligned} \Omega_{x_k} = \Omega_{x_k|Y_{k-1}} &= Cov[x_k - \bar{x}_{k|Y_{k-1}}] \\ &= Cov[Ax_{k-1} + w_{k-1} - A\hat{x}_{k-1}] \\ &= Cov[Ae_{k-1} + w_{k-1}] \\ &= AP_{k-1}A^T + Q_{k-1} \\ &\triangleq P_k^-. \end{aligned} \tag{3.78}$$

Therefore, the recursive projection filter and its corresponding posterior error covariance become

$$\begin{aligned} \hat{x}_k &= A\hat{x}_{k-1} + P_k^- C^T (CP_k^- C^T + R)^{-1}(y_k - CA\hat{x}_{k-1}) \\ &= A\hat{x}_{k-1} + K_k(y_k - CA\hat{x}_{k-1}) \end{aligned} \tag{3.79}$$

where

$$K_k = P_k^- C^T (CP_k^- C^T + R)^{-1}, \tag{3.80}$$

and

$$\begin{aligned} P_k^+ &= P_k^- - P_k^- C^T (CP_k^- C^T + R)^{-1} CP_k^- \\ &= (I_n - K_k C)P_k^-. \end{aligned} \tag{3.81}$$

Equations (3.77) to (3.81) are exactly the same as the Kalman filter formulations. This proves that *the recursive projection filter of order one is identical to the Kalman filter.*

Enlightened by the equivalence just derived, immediately we can find a modified Kalman filter for the situation in which measurement noise is correlated with state. For this situation the conventional Kalman filter ceases to be optimal if applied directly. The structure of the modified filter is the same as the conventional Kalman filter, but the filter gain becomes

$$K_k = (P_k^- C^T + P_{xv})(CP_k^- C^T + CP_{xv} + P_{xv}^T C^T + R)^{-1}, \tag{3.82}$$

where P_{xv} is the covariance of state error and measurement noise. This modified filter is a benefit of deriving the Kalman filter from the projection filter.

For recursive projection filters of orders greater than one in which the recursive feature is obtained by using conditional a priori mean and state covariance, the formulations are the same as (3.77) to (3.81), except C and R should be replaced by $\tilde{H}_q$ and $\tilde{\Sigma}$, respectively, and I_n in (3.81) replaced by the identity matrix of proper dimension. In this case, some measurements are used more than one time in estimating one single state, that is, some measurements are used both in calculating the a priori estimate and in calculating the filter part, or the modifying part. However, this does not help in improving the results. Since the projection filter seeks the conditional mean of the state, it makes no difference whether a measurement is conditioned once, twice, or more. Consequently, recursive projection filter of an order greater than one is computationally inefficient. Though there is no benefit in computation, the concept of recursive projection filter is still valuable. The property of equivalence between recursive projection filter and Kalman filter helps in the development of an effective system identification method which is introduced in Chapter 6.

3.6 Numerical Examples

Numerical examples are provided to verify the projection filter formulations and to compare the projection filter with the Kalman filter. A lumped-mass beam-like dynamical system as shown in Fig. 3.1 is simulated. The system has three modes (six states). The modal frequency and the damping ratio of each mode are listed as follows:

Mode	Frequency (rad/sec)	Damping (%)
1	1.6369	0.63
2	4.4719	1.01
3	6.1085	1.30

The sampling frequency is 10 Hz. The state space parameters A and C are

$$A = diag\left\{\begin{bmatrix} 0.9856 & 0.1628 \\ -0.1628 & 0.9856 \end{bmatrix} \begin{bmatrix} 0.8976 & 0.4305 \\ -0.4305 & 0.8976 \end{bmatrix} \begin{bmatrix} 0.8127 & 0.5690 \\ -0.5690 & 0.8127 \end{bmatrix}\right\}$$

$$C = \begin{bmatrix} 1.5119 & 0. & 2.0000 & 0. & 1.5119 & 0. \\ 1.3093 & 0. & 0. & 0. & -1.3093 & 0. \end{bmatrix}$$

where A is a block diagonal matrix.

The projection filters for noise-free systems and systems without process noise are trivial, and hence are exempted from numerical example. Only the case of systems with both process and measurement noise is illustrated. The system starts from an initial condition $x_0 = 0$, and is excited by Gaussian white noise with covariance $Q = 0.005 \times I_6$. The two output measurements are contaminated by additive Gaussian white noise with covariances $R = 0.3 \times I_2$. The noise-to-signal ratio in variances is about 10%.

The optimal Kalman and the projection filters of order 2, 10 and 20 are used to estimate the state using the same initial conditions ($\hat{x}_0 = 0$ and $P_0 = 4.0 \times I_6$) and noise covariances. The results are shown in Figs. 3.2 and 3.3. Figure 3.2

shows the estimations of the first state, while Fig. 3.3 shows the sixth state, which is of higher frequency. In both figures, the results of the optimal Kalman filter are compared with the results of the projection filters with different order. The solid lines represent the true states, and the dashed lines represent the estimated values. Figure 3.4 shows the error variances of state 1 and state 6 of the projection filter vs. filter order. The dashed line represents the error variance of the optimal Kalman filter estimation which is shown for comparison. As the order of the projection filter increases the results are improved and approach those of the Kalman filter. The error variances shown are the values calculated by averaging 200 samples, discarding the first 50 samples (not obtained from theory). Similarly, Figs. 3.5, 3.6 and 3.7 show the comparison of the optimal Kalman filter and the constant projection filters which use steady state covariance instead of the propagating time-varying state covariance. Note that in this case the constant projection filter is even better than the time-varying one. This is due to the poor estimate of the initial state error covariance for the time-varying case.

3.7 Concluding Remarks

The results reported in this chapter are summarized as follows:

(1) State estimation of linear systems can be realized using projection filters, which are based on optimal parameter estimation theories.

(2) The projection filter for noise-free systems is a least-squares filter, which is simply the pseudo-inverse of the measurement matrix. It does not need initial values of state nor its error covariance, but cannot yield estimates until it has accumulated a certain number of data. The estimation is perfect.

(3) The projection filter for systems without process noise is a weighted least-squares filter, in which the weighting matrix is determined by the covariance of measurement noise. It also does not need initial values to initiate the

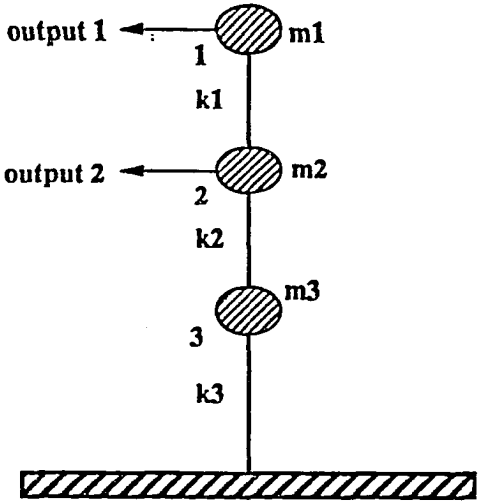
estimation but cannot yield estimates until it has accumulated a certain number of data. The estimation is less accurate when compared to the Kalman filter.

(4) The projection filter for systems with both process and measurement noises is a Baysian estimation, which compute the conditional mean of the state based on all the measurements used. It needs the same amount of a priori information of system and noises as the Kalman filter does. The projection filter of small order is less accurate and requires more computation than the Kalman filter.

(5) The projection filter is also an optimal correlation canceler which extracts all correlated information associated with the state estimated and the measurement vector.

(6) The recursive projection filter is equivalent to the Kalman filter. The recursive projection filter of order one has exactly the same formulations as the Kalman filter does.

(7) The recursive projection filter provides a viable method for state estimation when the measurement noise is correlated with state.



m1	0.5
m2	1.0
m3	1.0
k1	10.0
k2	10.0
k3	10.0

Fig. 3.1 A simulated lumped-mass beam-like system.

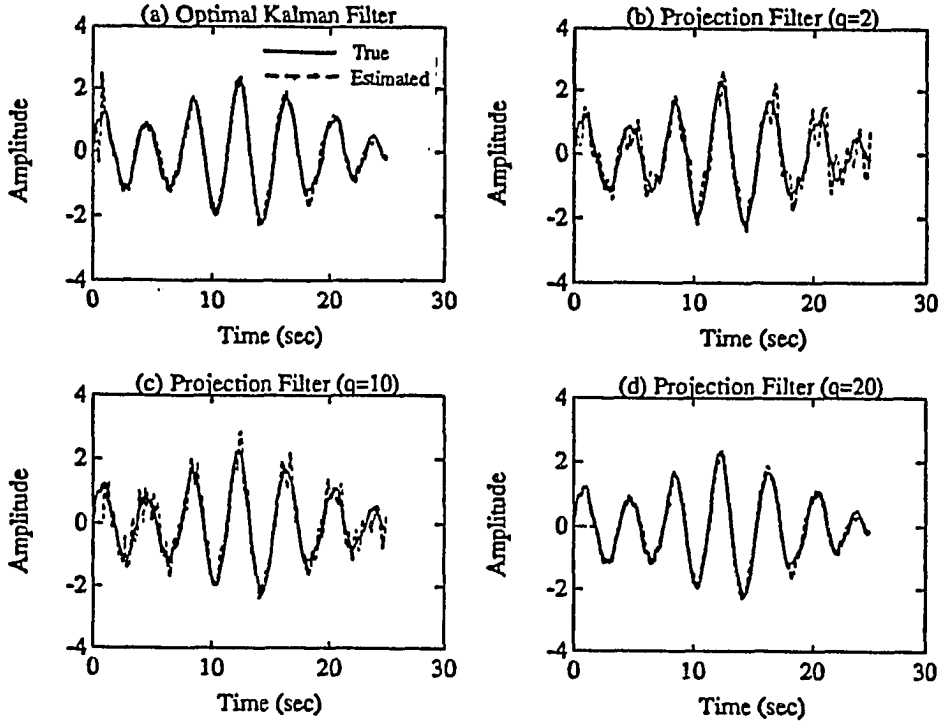


Fig. 3.2 Estimation of the first state by optimal Kalman filter and by time-varying projection filters with different orders.

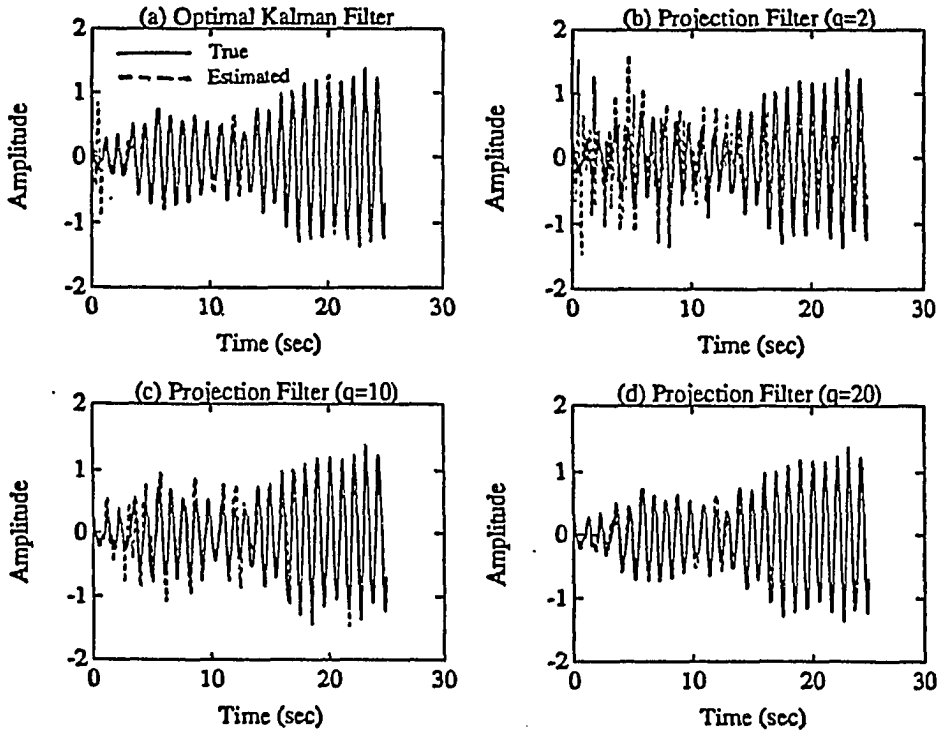


Fig. 3.3 Estimation of the sixth state by optimal Kalman filter and by time-varying projection filters with different orders.

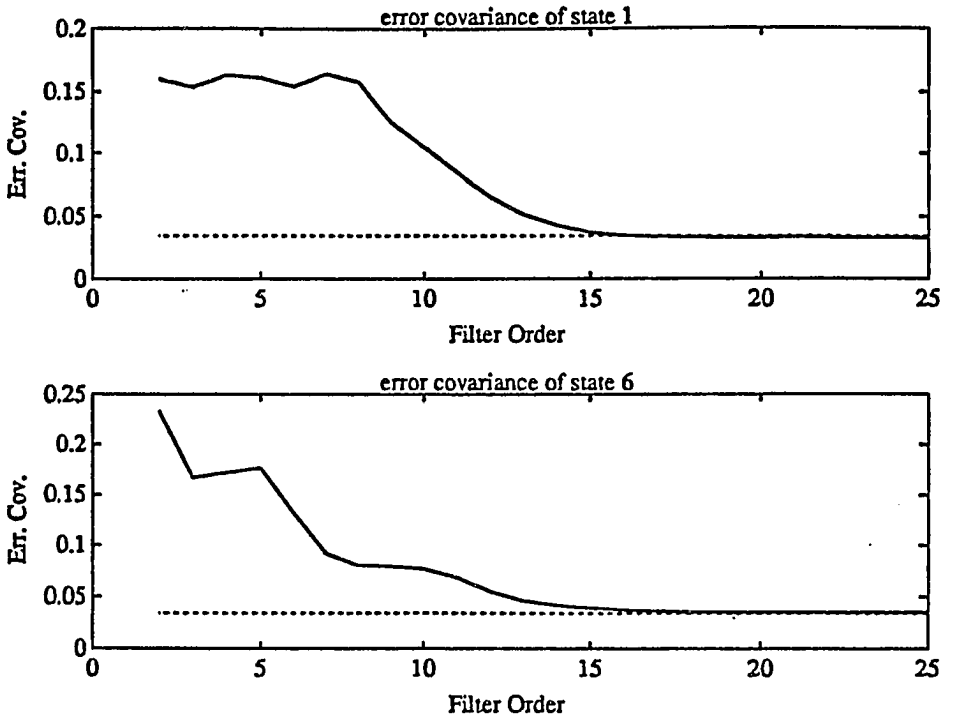


Fig. 3.4 State error variance by time-varying projection filter as function of order.

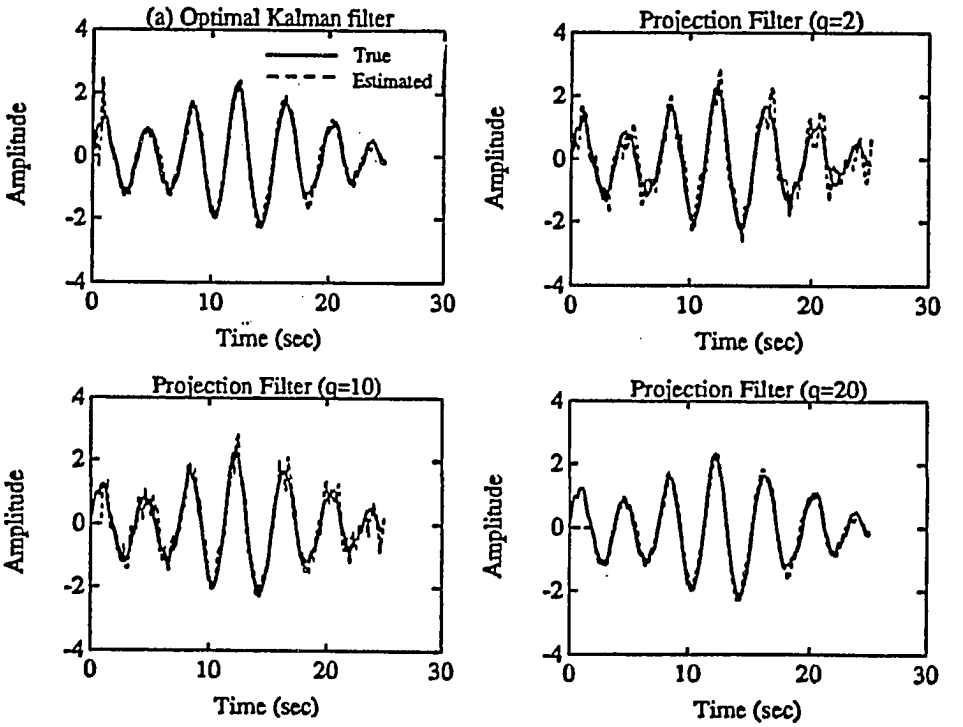


Fig. 3.5 Estimation of the first state by optimal Kalman filter and by constant projection filters with different orders.

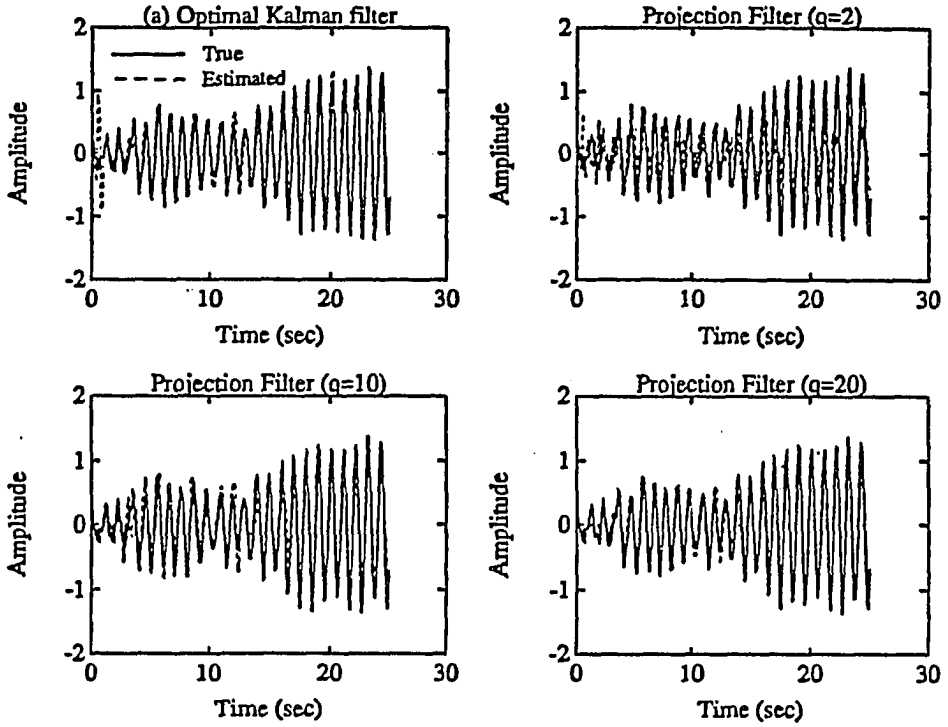


Fig. 3.6 Estimation of the sixth state by optimal Kalman filter and by constant projection filters with different orders.

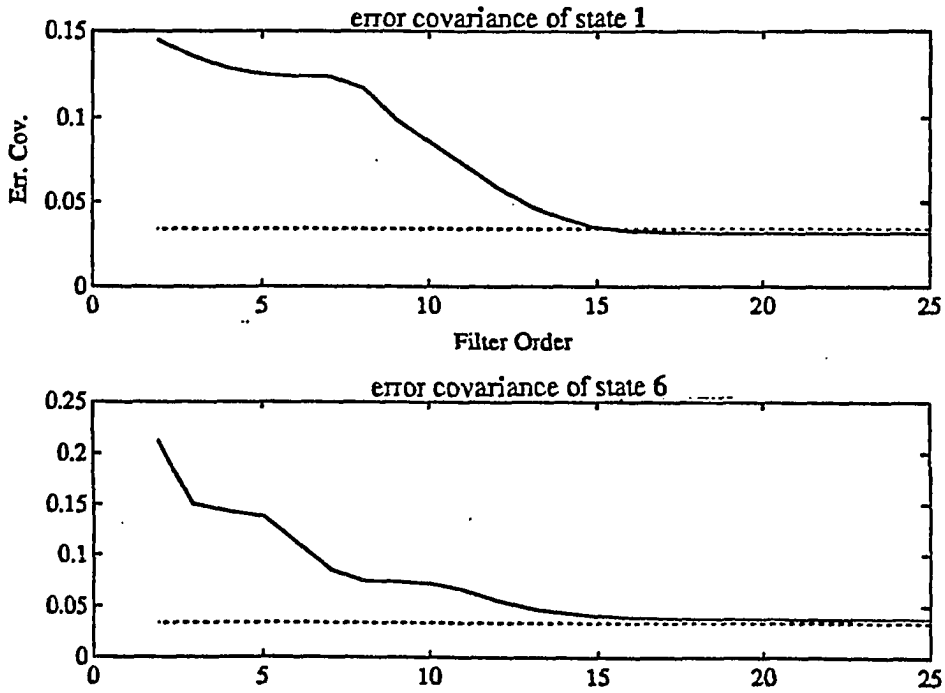


Fig. 3.7 State error variance by constant projection filter as function of order.

Chapter 4

LINEAR STATE ESTIMATION UNDER UNKNOWN NOISE COVARIANCES

—— Direct Least-squares Approach

4.1 Introduction

In Chapter 3 system model and noise covariances are assumed known in advance in estimating state information of a linear time-invariant dynamical system. In this chapter the noise covariances are assumed unknown. This situation poses the second stage of the state estimation problem: how to conduct state estimation under unknown noise covariances. The conventional Kalman filter under this situation cannot be used directly because without noise covariances the filter gain needed for operation can not be computed. In order to conduct state estimation, therefore, one should either get something done before really using the Kalman filter, or consider some approaches other than the conventional Kalman filter.

The work needed to be finished before running the Kalman filter is tuning the filter, or estimating either the noise covariances or the Kalman filter gain.

Tuning the filter basically consists of choosing a set of guessed noise covariances by "engineering judgment". In practice this is a cut-and-try process, which is subjective, experience-required and time-consuming. The criterion for adjusting parameters is to make the resulting residual sequence as white as possible, because the residual of the optimal Kalman filter is white.[28] Once the filter can produce near white residual, the parameters used can be regarded as proper. The quality of the estimation in the above approach relies fully on the quality of the adjustment, and because many parameters must be adjusted, it is very difficult to attain optimality. Therefore, this approach usually results in a suboptimal filter.

If the ad-hoc filter tuning process described above is not considered for a Kalman filter application, estimation of noise covariance or Kalman filter gain should be conducted and will be discussed in the next chapter. In this chapter we discard the Kalman filter structure and use a direct least-squares approach, which is basically different from the other approaches presented in the next chapter.

In this chapter the time-variant state estimation problem is re-phrased into a time-invariant linear parameter estimation problem, and least-squares techniques are then used to solve it. Since there is no statistical information about input/output data and noise, the least-squares technique is used because it does not require the noise statistics and the initial values of the state and its error covariance as Kalman filter does. It is also simple to understand and easy to use. However, the trade-off is that it can provide only suboptimal results when process noise exists. Kalman filter has been used to replace least-squares in parameter estimation problem.[16] However, to the author's knowledge, using least-squares filter to replace Kalman filter in state estimation has never been addressed.

Section 4.2 is devoted to deriving the least-squares filters for various conditions. A close relationship is found between the least-squares filters and the Kalman filter, which is discussed in Sections 4.3 and 4.4; therefore, this chapter

also provides material for a better understanding of the Kalman filter. Numerical examples are given in Section 4.5 to verify the derivations.

4.2 A Least-Squares Approach

In section 4.2.1 filters for linear systems without process noise are derived. These include the fixed-ordered filter using a fixed number of previous data to make the estimate, the infinite-ordered filter using all the data available, and the recursive weighted least-squares filter with different weighting for the measurement from each sensor when the measurement noise covariance is known. Section 4.2.2 derives the fading memory least-squares filter for systems with both process and measurement noises.

4.2.1 Least-squares Filters for Linear Systems without Process Noise

First, consider the case of a system without process noise. This system can be represented by the following state space model, in which the input force term is not included for simplicity:

$$x_{k+1} = Ax_k, \tag{4.1}$$

$$y_k = Cx_k + v_k. \tag{4.2}$$

Based on this model the following equations in matrix form can be derived:

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+1} \end{bmatrix} = \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+1} \end{bmatrix} x_k + \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-q+1} \end{bmatrix} \tag{4.3}$$

or, in short,

$$Y_{q,k} = H_q x_k + V_{q,k}, \tag{4.4}$$

where q denotes the number of the successive data used in the formulation. This equation relates $q-1$ previous data and their corresponding noises to the current state vector in a linear form. This linear equation, as can be seen in the next sections, is very useful in seeking a solution to the current state.

4.2.1.1 Fixed-ordered Least-squares Filter

Suppose the number q is fixed; then this situation is the same as in Section 3.3.2 except now the measurement noise covariance is unknown. State vector x_k has a least-squares solution

$$\hat{x}_k = (H_q^T H_q)^{-1} H_q^T Y_{q,k} = H_q^\dagger Y_{q,k}, \tag{4.5}$$

where $\hat{x}_k$ is the optimal estimate of the state vector x_k and $H_q^\dagger$ is the pseudo-inverse of the observability-type matrix H_q. This result is identical to (3.25).

To study the statistical properties of the estimate, define the estimation error e_k by

$$e_k \triangleq x_k - \hat{x}_k, \tag{4.6}$$

and substitute (4.4) into (4.5) for $Y_{q,k}$ to yield

$$\hat{x}_k = H_q^\dagger H_q x_k + H_q^\dagger V_{q,k} = x_k + H_q^\dagger V_{q,k}. \tag{4.7}$$

Because $E[H_q^\dagger V_{q,k}] = 0$ by the zero-mean assumption on v_k, one has $E[\hat{x}_k] = E[x_k]$, which indicates that the estimate is unbiased. Furthermore, comparing (4.7) with (4.6), apparently $e_k = -H_q^\dagger V_{q,k}$. Hence, the error covariance of the estimation can be calculated as

$$P_k = E[e_k e_k^T] = H_q^\dagger E[V_{q,k} V_{q,k}^T](H_q^\dagger)^T, \tag{4.8}$$

where $E[V_{q,k} V_{q,k}^T]$ is the covariance of the collective noise vector $V_{q,k}$. If all the measurement noises are uncorrelated and equally strong, which implies that the

noise covariance R is of the form $\sigma^2 \times I_p$, then the error covariance of the estimation becomes

$$P_k = \sigma^2 H_q^{\dagger}(H_q^{\dagger})^T = \sigma^2 (H_q^T H_q)^{-1}, \tag{4.9}$$

where $(H_q^T H_q)^{-1}$ is constant. Therefore, the quality of the estimate is directly proportional to the variance of the noise of a single measurement. Furthermore, because

$$\begin{aligned} H_q^T H_q &= [C^T,\ (CA^{-1})^T, \cdots,\ (CA^{-q+1})] \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+1} \end{bmatrix} \\ &= \sum_{i=1}^{q} (A^{-i+1})^T C^T C A^{-i+1}. \end{aligned} \tag{4.10}$$

When q approaches infinity, $H_q^T H_q$ becomes infinity and thus $(H_q^T H_q)^{-1}$ becomes zero, which in turn means P_k becomes zero. Therefore, the estimate is perfect when q approaches infinity. In this case, the method can be viewed as choosing a state vector $\hat{x}_k$ to fit q sets of measurement data optimally in the least-squares output error sense using the relations provided by the system dynamic equation. The pseudo-inverse matrix $H_q^{\dagger}$ is fixed and only needs to be calculated once. Hence, the filter is actually a time invariant finite impulse response (FIR) filter, which receives y_k and its $q-1$ delay versions as input and yields estimated state $\hat{x}_k$ as output. The filter order q should be sufficiently large to make matrix H_q full-column-ranked, yet not so large as to cause very heavy computational load.

4.2.1.2 Infinite-ordered Least-squares Filter

If the filter order q is not fixed but increases with time index k, a recursive least-squares filter can be derived from the ordinary least-squares as follows. Suppose at time k the least-squares solution of the state is

$$\hat{x}_k = (H_k^T H_k)^{-1} H_k^T Y_k,$$

(Y_k is used to replace $Y_{k,k}$ for convenience), then at next step $k+1$, the estimated state should be

$$\hat{x}_{k+1} = (H_{k+1}^T H_{k+1})^{-1} H_{k+1}^T Y_{k+1},$$

where

$$H_{k+1} = \begin{bmatrix} H_k \\ CA^{-k} \end{bmatrix} = \begin{bmatrix} C \\ H_k A^{-1} \end{bmatrix} \tag{4.11}$$

and

$$Y_{k+1} = \begin{bmatrix} y_{k+1} \\ Y_k \end{bmatrix}. \tag{4.12}$$

Hence,

$$\begin{aligned} \hat{x}_{k+1} &= \left(\left[C^T \vdots A^{-T} H_k^T \right] \begin{bmatrix} C \\ H_k A^{-1} \end{bmatrix} \right)^{-1} \left[C^T \vdots A^{-T} H_k^T \right] \begin{bmatrix} y_{k+1} \\ Y_k \end{bmatrix} \\ &= \left[C^T C + A^{-T} H_k^T H_k A^{-1} \right]^{-1} \left[C^T \vdots A^{-T} H_k^T \right] \begin{bmatrix} y_{k+1} \\ Y_k \end{bmatrix}. \end{aligned} \tag{4.13}$$

At this point the matrix inversion lemma is used to expand the matrix inversion part of (4.13). This is a crucial step of the derivation. Therefore, for better understanding the lemma is briefly stated here.

Matrix Inversion Lemma:

Let D and H be two positive-definite, $m \times m$ matrices related by

$$D = H + EFG \tag{4.14}$$

where F is another positive-definite, $n \times n$ matrix, and E an $m \times n$, G an $n \times m$ matrix, we may express the inverse of the matrix D as follows:

$$D^{-1} = H^{-1} - H^{-1}E(F^{-1} + GH^{-1}E)^{-1}GH^{-1}. \tag{4.15}$$

In the matrix inversion part of (4.13), with $H = A^{-T} H_k^T H_k A^{-1}$, $E = C^T$, $F = I_p$, $G = C$, the equation becomes

$$\hat{x}_{k+1} = [A(H_k^T H_k)^{-1} A^T - A(H_k^T H_k)^{-1} A^T C^T (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1}$$

$$
\begin{aligned}
&\times CA(H_k^T H_k)^{-1} A^T][C^T y_{k+1} + A^{-T} H_k^T Y_k] \\
=&[A(H_k^T H_k)^{-1} A^T - A(H_k^T H_k)^{-1} A^T C^T (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1} \\
&\times CA(H_k^T H_k)^{-1} A^T] C^T y_{k+1} + \\
&[A(H_k^T H_k)^{-1} A^T - A(H_k^T H_k)^{-1} A^T C^T (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1} \\
&\times CA(H_k^T H_k)^{-1} A^T] A^{-T} H_k^T Y_k \\
=&A(H_k^T H_k)^{-1} A^T C^T [(I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1} (I_p + CA(H_k^T H_k)^{-1} A^T C^T) \\
&- (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1} CA(H_k^T H_k)^{-1} A^T C^T] y_{k+1} + \\
&[A(H_k^T H_k)^{-1} H_k^T Y_k - A(H_k^T H_k)^{-1} A^T C^T (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1} \\
&\times CA(H_k^T H_k)^{-1} H_k^T Y_k] \\
=&A(H_k^T H_k)^{-1} A^T C^T (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1} y_{k+1} + \\
&[A(H_k^T H_k)^{-1} H_k^T Y_k - A(H_k^T H_k)^{-1} A^T C^T (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1} \\
&\times CA(H_k^T H_k)^{-1} H_k^T Y_k]. \qquad (4.16)
\end{aligned}
$$

Note that $(H_k^T H_k)^{-1} H_k^T Y_k = \hat{x}_k$, and if we denote

$$
\Pi_{k+1} = A(H_k^T H_k)^{-1} A^T C^T (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1}, \qquad (4.17)
$$

then (4.16) reduces to

$$
\begin{aligned}
\hat{x}_{k+1} &= \Pi_{k+1} y_{k+1} + A\hat{x}_k - \Pi_{k+1} CA\hat{x}_k \\
&= A\hat{x}_k + \Pi_{k+1}(y_{k+1} - CA\hat{x}_k). \qquad (4.18)
\end{aligned}
$$

If we further denote $\Phi_{k+1} = A(H_k^T H_k)^{-1} A^T$, then (4.17) becomes

$$
\Pi_{k+1} = \Phi_{k+1} C^T (I_p + C\Phi_{k+1} C^T)^{-1}. \qquad (4.19)
$$

By defining $\Psi_k \triangleq (H_k^T H_k)^{-1}$, similar to (4.16), we obtain the following equation:

$$
\Psi_{k+1} = (H_{k+1}^T H_{k+1})^{-1}
$$

$$
\begin{aligned}
&= A(H_k^T H_k)^{-1} A^T - A(H_k^T H_k)^{-1} A^T C^T (I_p + CA(H_k^T H_k)^{-1} A^T C^T)^{-1} \\
&\quad \times CA(H_k^T H_k)^{-1} A^T \\
&= A\Psi_k A^T - \Pi_{k+1} CA\Psi_k A^T \\
&= (I_n - \Pi_{k+1} C) A\Psi_k A^T = (I_n - \Pi_{k+1} C)\Phi_{k+1}. \qquad (4.20)
\end{aligned}
$$

It is interesting to see that (4.18) to (4.20) have exactly the same form as the Kalman filter, where Π_{k+1} is equivalent to the Kalman gain and the value in the parenthesis is the residual. The value of Ψ_k can be interpreted as the a posteriori error covariance and Φ_{k+1} the a priori error covariance.

4.2.1.3 Recursive Weighted Least-Squares Filter

The above formulations are derived under the criterion of the least-squares of output error, which implies that the estimate $\hat{x}_k$ can fit k equations in (4.3) ($q = k$) with a minimum sum of the squares of the output deviations. Every element of the output deviations is equally weighted, which is equivalent to assuming that the measurement noises from different sensor channels are equally strong and are uncorrelated with each other. If no other information about the properties of the output data is available, this is the best result one can obtained. However, if the covariance of the measurement noise is known, the result can be improved. In general, the covariance of measurement noise is much easier to obtain than process noise, if the latter one exists.

In this section we assume that the covariance of the measurement noise is given, which is the same situation as in Section 3.3.2, and, hence, the weighted least-squares should be employed. In this section a solution in recursive form is derived, and the results are very similar to the last section. It is found that, even though the measurement noises are white, if the noise from different sensors are not equally strong and/or are correlated with each other, results can be degraded if weighting is not used.

According to the theory of the weighted least-squares, the optimal weighting matrix is the inverse of the covariance of the measurement noise (see section 3.3.2). Therefore, (4.5) should be modified as

$$\hat{x}_k = (H_k^T \bar{R}_k^{-1} H_k)^{-1} H_k^{-1} \bar{R}_k^{-1} Y_k \tag{4.21}$$

where $\bar{R}_k = E[V_k V_k^T]$. Denoting the measurement noise covariance by R, the optimal estimate of the state at the next time step is

$$\begin{aligned}
\hat{x}_{k+1} =& (H_{k+1}^T \bar{R}_{k+1}^{-1} H_{k+1})^{-1} H_{k+1}^T \bar{R}_{k+1}^{-1} Y_{k+1} \\
=& \left([C^T \vdots A^{-T} H_k^T] \begin{bmatrix} R^{-1} & 0 \\ 0 & \bar{R}_k^{-1} \end{bmatrix} \begin{bmatrix} C \\ H_k A^{-1} \end{bmatrix} \right)^{-1} \\
& \times [C^T \vdots A^{-T} H_k^T] \begin{bmatrix} R^{-1} & 0 \\ 0 & \bar{R}_k^{-1} \end{bmatrix} \begin{bmatrix} y_{k+1} \\ Y_k \end{bmatrix} \\
=& [C^T R^{-1} C + A^{-T} H_k^T \bar{R}_k^{-1} H_k A^{-1}]^{-1} [C^T \vdots A^{-T} H_k^T] \begin{bmatrix} R^{-1} y_{k+1} \\ \bar{R}_k^{-1} Y_k \end{bmatrix} \\
=& \cdots \\
=& A(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T (R + CA(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T)^{-1} y_{k+1} \\
& + [A(H_k^T R^{-1} H_k)^{-1} H_k^T \bar{R}_k^{-1} Y_k - A(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T \\
& \times (R + CA(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T)^{-1} \times CA(H_k^T \bar{R}_k^{-1} H_k)^{-1} H_k^T \bar{R}_k^{-1} Y_k] \\
=& A\hat{x}_k + \Pi_{k+1}(y_{k+1} - CA\hat{x}_k)
\end{aligned} \tag{4.22}$$

where

$$\begin{aligned}
\Pi_{k+1} &= A(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T (R + CA(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T)^{-1} \\
&= \Phi_{k+1} C^T (R + C\Phi_{k+1} C^T)^{-1}
\end{aligned} \tag{4.23}$$

and

$$\Phi_{k+1} = A(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T = A\Psi_k A^T \tag{4.24}$$

$$\Psi_k = (H_k^T \bar{R}_k^{-1} H_k)^{-1} = (I - \Pi_k C)\Phi_k. \tag{4.25}$$

Equations (4.22) to (4.25) constitute the formulations for the recursive weighted least-squares state estimation. Note the criterion of optimality here is in the

least squares sense for the state estimation error as well as for the measurement error.[29] In the case of no process noise, this method yields the same results as the Kalman filter does. However, the Kalman filter requires a priori knowledge of the initial state and its corresponding error covariance. A poor estimate of the initial state and its corresponding error covariance may degrade the filter performance during the transient period. This method does not require such initial conditions. After H_k becomes full-column-ranked, (4.21) and (4.25) can provide optimal initial estimate of the state and its covariance, the recursive algorithm can then be put into operation. However, before H_k becomes full rank, no estimate can be made.

4.2.2 Least-squares Filter for Linear Systems with Both Process and Measurement Noises

When both process and measurement noises exist, the system dynamics can be modeled by the first Markov random process as

$$x_{k+1} = Ax_k + w_k, \tag{4.26}$$

where the sequence $\{w_k\}$ is assumed to be a zero-mean, stationary, Gaussian white noise with constant covariance Q. The relation between the current state and q successive measurements, including the current one, can be written as

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+2} \\ y_{k-q+1} \end{bmatrix} = \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+2} \\ CA^{-q+1} \end{bmatrix} x_k - \begin{bmatrix} 0 & \cdots & 0 & 0 \\ CA^{-1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ CA^{-q+2} & \cdots & CA^{-1} & 0 \\ CA^{-q+1} & \cdots & CA^{-2} & CA^{-1} \end{bmatrix} \begin{bmatrix} w_{k-1} \\ \vdots \\ w_{k-q+2} \\ w_{k-q+1} \end{bmatrix} + \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-q+2} \\ v_{k-q+1} \end{bmatrix}, \tag{4.27}$$

where the second term in the right hand side arises because of the existence of process noise. Since this noise term is correlated with the current state, which violates the "independent noise" assumption of the ordinary least-squares, the method does not apply in this case. If we attempt to cancel the correlation using correlation cancelation technique introduced in Section 3.5, the statistics of state x_k and noise sequence W_q ($= [\, w_{k-1}^T, \cdots, w_{k-q+1}^T]^T$) should be known in advance, which is not only impractical but also undesirable for a deterministic approach. Therefore, we should approach the problem in a different way.

4.2.2.1 Fading Memory Least-squares Filter

The degree of uncertainty of the current measurement due to measurement noise is indicated by noise covariance R. Because of the influence of process noise, the previous data are more uncertain as compared to the current data in terms of bearing the information about the current state and is therefore less reliable. A weighting technique should be used to account for this factor. By assigning each one-step-past data a larger noise covariance as compared to that of the current data, for instance by multiplying a factor λ^{-1} to the covariance matrix of the last noise, where λ is a number close to but less than 1, we can make the previous data less important during least-squares fitting. Through the recursive least-squares method the weighting of the previous data will be exponentially reduced. Therefore, the method has an ability to gradually "forget" the old data and emphasize the new data. The recursive forgetting algorithm is derived as follows.

Suppose at time step k we have the estimation equation as (4.21), then at time step $k+1$, by introducing the forgetting factor λ, we have

$$\begin{aligned}\hat{x}_{k+1} =& (H_{k+1}^T \bar{R}_{k+1}^{-1} H_{k+1})^{-1} H_{k+1} \bar{R}_{k+1}^{-1} Y_{k+1} \\ =& \left([C^T \vdots A^{-T} H_k^T] \begin{bmatrix} R^{-1} & 0 \\ 0 & \lambda \bar{R}_k^{-1} \end{bmatrix} \begin{bmatrix} C \\ H_k A^{-1} \end{bmatrix} \right)^{-1}\end{aligned}$$

$$
\begin{aligned}
&\times [C^T \vdots A^{-T}H_k^T] \begin{bmatrix} R^{-1} & 0 \\ 0 & \lambda\bar{R}_k^{-1} \end{bmatrix} \begin{bmatrix} y_{k+1} \\ Y_k \end{bmatrix} \\
=&[C^T R^{-1} C + \lambda A^{-T} H_k^T \bar{R}_k^{-1} H_k A^{-1}]^{-1} [C^T \vdots A^{-T} H_k^T] \begin{bmatrix} R^{-1} y_{k+1} \\ \lambda \bar{R}_k^{-1} Y_k \end{bmatrix} \\
=&\cdots \\
=&\lambda^{-1} A (H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T (R + \lambda^{-1} CA(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T)^{-1} y_{k+1} \\
&+ [A(H_k^T \bar{R}_k^{-1} H_k)^{-1} H^T \bar{R}_k^{-1} Y_k - \lambda^{-1} A (H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T \\
&\times (R + \lambda^{-1} CA(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T)^{-1} \times CA(H_k^T \bar{R}_k^{-1} H_k)^{-1} H_k^T \bar{R}_k^{-1} Y_k] \\
=&A\hat{x}_k + \Pi_{k+1}(y_{k+1} - CA\hat{x}_k) \qquad (4.28)
\end{aligned}
$$

where

$$
\begin{aligned}
\Pi_{k+1} &= \lambda^{-1} A (H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T (R + \lambda^{-1} CA(H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T C^T)^{-1} \\
&= \Phi_{k+1} C^T (R + C\Phi_{k+1} C^T)^{-1} \qquad (4.29)
\end{aligned}
$$

and

$$
\Phi_{k+1} = \lambda^{-1} A (H_k^T \bar{R}_k^{-1} H_k)^{-1} A^T = \lambda^{-1} A \Psi_k A^T \qquad (4.30)
$$

$$
\Psi_{k+1} = (H_{k+1}^T \bar{R}_{k+1}^{-1} H_{k+1})^{-1} = (I - \Pi_{k+1} C)\Phi_{k+1}. \qquad (4.31)
$$

Equations (4.28) to (4.31) constitute the fading memory least-squares filter for state estimation, which are also written in the Kalman filter form.

Parameter λ can be used to adjust the memory length. If $\lambda = 1$, which means infinite memory, the filter will rely on all the data equally as that in the no-process-noise case. Reducing λ will reduce the memory length and the filter will "forget" old data faster, corresponding to the case when the process noise is significantly large.

The question of how to choose the value of λ remains. This problem is similar to the problem of determining the process noise covariance Q in the Kalman filter. The usual way is to "tune" the filter by observing the whiteness of its residual

sequence. After all, residual sequence is the only information available for judging the performance of the filter. From the study of the Kalman filter one knows that if the filter is optimal the residual sequence should be white and zero-mean, which can be interpreted intuitively that no signal is left in the residual and hence the residual is totally unpredictable (white). Therefore, in practice we can adjust λ by monitoring the residual sequences. Although we may not be able to obtain a white residual by adjusting only one variable λ, we can expect to have satisfactory results if all the measurement residuals are quite random.

4.3 The Relation Between Fading Memory Least-squares Filter and Kalman Filter

The optimal Kalman filter is derived based on the optimality criterion of least-mean-squares of state error, while the fading memory least-squares filter (FMLS) is based on the least-squares of measurement error. However, it is interesting to see that they produce the same filter forms.

From the formulations, we note that Φ_k in the FMLS filter is equivalent to a priori state error covariance P_k^-, and Ψ_k to a posteriori error covariance P_k^+ in the Kalman filter. The formula for P_{k+1}^- in the Kalman filter is

$$P_{k+1}^- = AP_k^+A^T + Q. \tag{4.32}$$

Comparing this to (4.30) we note that, instead of adding process noise covariance Q to the propagation of state error covariance as in the Kalman filter, the FMLS simply multiplies a factor λ^{-1} (which is larger than one) to account for the effect of the process noise. By writing (4.30) as

$$\Phi_{k+1} = A\Psi_kA^T + (\lambda^{-1} - 1)A\Psi_kA^T \tag{4.33}$$

it can be clearly seen that the FMLS implicitly assigns the value of process noise covariance as $Q' = (\lambda^{-1} - 1)A\Psi_k A$. Therefore, if the process noise is small or not significantly different from Q', the FMLS can give reasonably good results.

The equivalent process noise covariance Q' is obtained by multiplying the propagated state error covariance by a constant $(\lambda^{-1}-1)$, and, therefore, the states which have larger error variances are assigned stronger process noise automatically and have less weighting. This appears to be intuitively correct. Hence, through the relation between the Kalman filter and the FMLS filter we can justify the use of the forgetting factor λ.

4.4 Another View of the Relation between Least-squares Filter and the Kalman Filter

In this section the Kalman filter is re-derived from another least-squares approach, which provides a better understanding of the relation between the Kalman filter and the least-squares filter.

Assume all the information carried by the data from the beginning till time $k-1$ can somehow be "compressed" into an estimated state and its corresponding error covariance by an optimal linear filter which is unknown at this stage. That is,

$$\hat{x}_{k-1}^{+} = HY_{k-1}, \tag{4.34}$$

where H is an appropriate matrix and Y_{k-1} a data vector stacking up all the data from the beginning till $k-1$. Denote the a posteriori prediction error and its covariance by e_{k-1}^{+} and P_{k-1}^{+}, respectively (i.e., $e_{k-1}^{+} \triangleq x_{k-1} - \hat{x}_{k-1}^{+}$ and $P_{k-1}^{+} = Cov[e_{k-1}^{+}]$). Using $\hat{x}_{k-1}^{+}$ in (4.34) we can make an optimal prediction of x_k based

on the system model:

$$\hat{x}_k^- = A\hat{x}_{k-1}^+ = AHY_{k-1}. \tag{4.35}$$

Denote the a priori prediction error by e_k^- (i.e., $e_k^- \triangleq x_k - \hat{x}_k^-$); then its covariance

$$\begin{aligned} P_k^- &= Cov[e_k^-] \\ &= Cov[Ae_{k-1}^+ + w_{k-1}] \\ &= AP_{k-1}^+ A^T + Q, \end{aligned} \tag{4.36}$$

where Q is the covariance of the process noise. Sequence e_k^- should be zero-mean if the estimation is unbiased. By definition,

$$x_k = \hat{x}_k^- + e_k^- = AHY_{k-1} + e_k^-;$$

hence,

$$HY_{k-1} = A^{-1}x_k - A^{-1}e_k^- \tag{4.37}$$

Combining (4.37) and (4.2) we have

$$\begin{bmatrix} y_k \\ HY_{k-1} \end{bmatrix} = \begin{bmatrix} C \\ A^{-1} \end{bmatrix} x_k + \begin{bmatrix} v_k \\ -A^{-1}e_k^- \end{bmatrix}$$

or

$$\begin{bmatrix} y_k \\ \hat{x}_{k-1}^+ \end{bmatrix} = \begin{bmatrix} C \\ A^{-1} \end{bmatrix} x_k + \begin{bmatrix} v_k \\ -A^{-1}e_k^- \end{bmatrix}. \tag{4.38}$$

Now seeking a weighted least-square solution of x_k from this equation, we have

$$\hat{x}_k^+ = \left([C^T \vdots A^{-T}] \begin{bmatrix} R^{-1} & 0 \\ 0 & \bar{P}_k^{-1} \end{bmatrix} \begin{bmatrix} C \\ A^{-1} \end{bmatrix} \right)^{-1} [C^T \vdots A^{-T}] \begin{bmatrix} R^{-1} & 0 \\ 0 & \bar{P}_k^{-1} \end{bmatrix} \begin{bmatrix} y_k \\ \hat{x}_{k-1}^+ \end{bmatrix}, \tag{4.39}$$

where

$$\bar{P}_k = Cov[-A^{-1}e_k^-] = A^{-1}P_k^- A^{-T}.$$

Simplifying (4.39) as in the previous sections, we obtain

$$\hat{x}_k^+ = (I_n - \Pi_k C)A\hat{x}_{k-1}^+ + \Pi_k y_k = A\hat{x}_{k-1}^+ + \Pi_k(y_k - \hat{y}_k^-) \tag{4.40}$$

where

$$\Pi_k = P_k^- C^T (R + C P_k^- C^T)^{-1},$$

and $\hat{y}_k^- = CA\hat{x}_{k-1}^+$. The state error covariance P_k^+ is

$$\begin{aligned} P_k^+ &= E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] = H^\dagger E[e_k e_k^T](H^\dagger)^{-T} \\ &= (H^T \bar{R}_k^{-1} H)^{-1} H^T \bar{R}_k^{-1} \bar{R}_k \bar{R}_k^{-1} H (H^T \bar{R}_k^{-1} H)^{-1} = (H^T \bar{R}_k^{-1} H)^{-1} \\ &= (I_n - \Pi_k C) A P_{k-1} A^T = (I_n - \Pi_k C) P_k^- \end{aligned} \tag{4.41}$$

These equations are exactly the same as the Kalman filter. This fact tells us that *though the Kalman filter is derived under the criterion of least-mean-squares of state error, in fact, it is also a least-squares filter, which provides least squares of output error.*

4.5 Numerical Examples

In the numerical examples the simulated three-mode dynamical system is the same as that used in Chapter 3. The results are also compared with the Kalman filter. Three cases are investigated.

Case 1: Recursive least-squares filter for systems without process noise

In this case the measurement is the impulse response or free decay data. The measurement noise covariance is

$$R = \begin{bmatrix} 0.5016 & 0.3742 \\ 0.3742 & 0.3308 \end{bmatrix},$$

which is intentionally set to a rather large value to show the effectiveness of the filter. To give a feeling of the intensity of the noise, Fig. 4.1 shows one of the measurement data, where (a) is the "clean" data while (b) is its noise-corrupted version. Fig. 4.2 shows the estimations of states 1, 2, 4 and 6, where the solid lines

represent the true states and the dashed lines represent the estimated values. It can be seen that the estimation is good after a transient period and becomes perfect as time goes on (the estimated states virtually coincide with the true states). The diagonal terms of the calculated estimation error covariance (calculated from 500 data points, skipping the transient period in the beginning) are

$$P_1 = [5.6764,\quad 5.8270,\quad 0.2889,\quad 0.3278,\quad 0.0147,\quad 0.0149] \times 10^{-3}.$$

Case 2: Recursive weighted least-squares filter for systems without process noise but with known measurement covariance

In this case the measurement noise covariance is assumed known. Therefore, the recursive weighted least-squares filter is used. The system and settings are the same as in case 1. The estimations of states 1, 2, 4 and 6 are shown in Fig. 4.3, and the diagonal terms of the corresponding error covariance are

$$P_2 = [4.9945,\ 5.1256,\ 0.0556,\ 0.0535,\ 0.0034,\ 0.0031] \times 10^{-5}$$

Compared with Case 1 it can be seen that the results are improved.

Case 3: Fading memory least-squares filter for systems with both process and measurement noises

In this case the process noise is set to be about 5% of the initial state and the measurement noise is also about 5% of the measurement in variance ratio.

For the optimal Kalman filter, the initial state is set to zero, and the initial error covariance is set to $10 \times I_n$. The results of the estimation of the first state and the auto-correlation function of the first output residuals are shown in Fig. 4.4. Again, the solid lines in the state estimation plots represent the true state histories, and the dashed lines represent the estimated ones. For the optimal Kalman filter these two lines almost coincide with each other. Theoretically, the residual of the optimal Kalman filter is a white sequence and hence its auto-correlation function

should be zero everywhere except at $\tau = 0$. However, the auto-correlation function shown in Fig. 4.4(b) is obtained from a finite number of samples and not normalized (the same for auto- correlation in other figures); therefore, it has ripples at $\tau \neq 0$.

For the fading memory least-squares filter (FMLS), the results of simulations with three different forgetting factors ($\lambda = 1.0,\ 0.90,\ 0.50$) are shown in Figs. 4.5 to 4.7. For $\lambda = 1.0$, which corresponds to an infinite memory case, the estimates deteriorate gradually due to the effect of neglecting process noise and finally fail to track the states. The auto-correlation function of the residual, Fig. 4.5(b), is clearly non-white. For the case of $\lambda = 0.90$, the state estimation is rather good when compared to the optimal Kalman filter. Auto-correlation of the residual is also very close to that of the Kalman filter. For the case of $\lambda = 0.50$, the memory was apparently too short and the estimates relied too heavily on a short period of recent measurements; therefore, the result is sensitive to measurement noise. The residual in this case is less white, as shown in Fig. 4.7(b). Therefore, by monitoring the whiteness of the residual, a proper forgetting factor can be chosen.

4.3 Concluding Remarks

A few conclusions can be drawn and are listed below:

(1) For systems without process noise, the fixed-ordered least-squares filter is the same as the fixed-ordered projection filter when the measurement noise of each output sensor is equally strong. The estimation is unbiased and consistent, so, for a sufficiently large order the estimate can approach perfectness.

(2) For systems without process noise, the recursive least-squares can provide unbiased and consistent estimations. The filter has the same structure as the Kalman filter. If the covariance of the measurement noise is known, the recursive weighted least-squares filter should be used. If the intensity of the

noises of each output sensor is significantly different or highly correlated to each other, the recursive weighted least-squares yield better results than the non-weighted approach.

(3) For systems having both process and measurement noises, the fading memory least-squares filter can be used. However, in general only suboptimal results can be obtained. The forgetting factor allows one to adjust memory length so as to cope with system process noise with different intensity.

(4) The least-squares filter is closely related to the Kalman filter. All the recursive least-squares filters in this chapter have structures similar to the Kalman filter. In fact, even though the Kalman filter is originally derived under the criterion of least-mean-squares of state error, it can also be derived under the criterion of weighted least squares of output error.

(5) The advantage of using the least-squares filters is its simplicity. Though it may produce suboptimal estimates, in general the results are reasonably accurate.

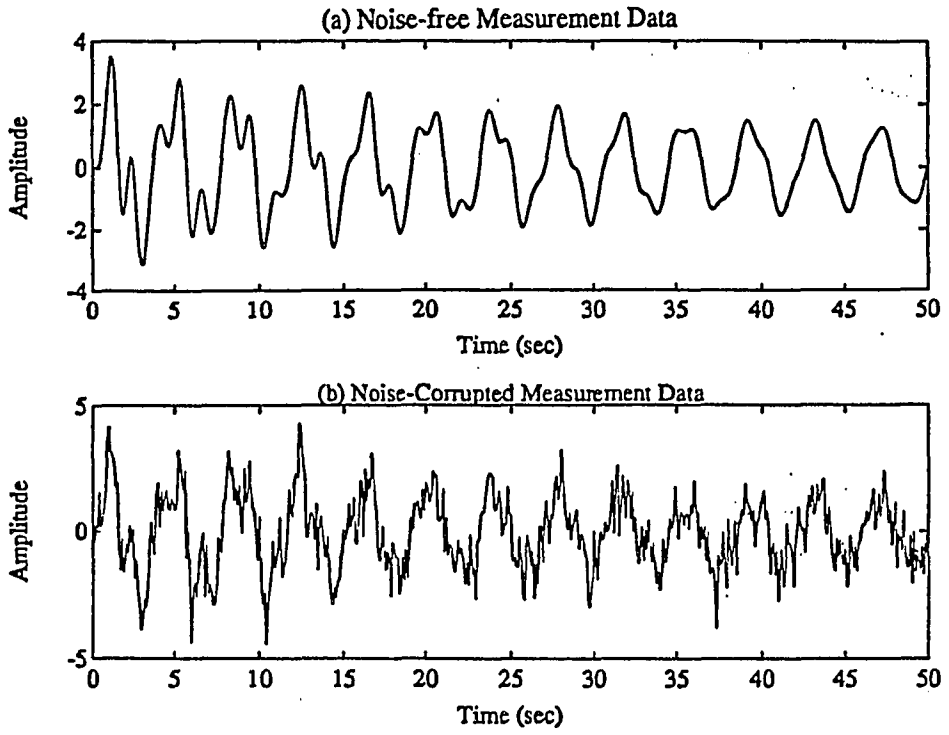


Fig. 4.1 Measurement of first output.

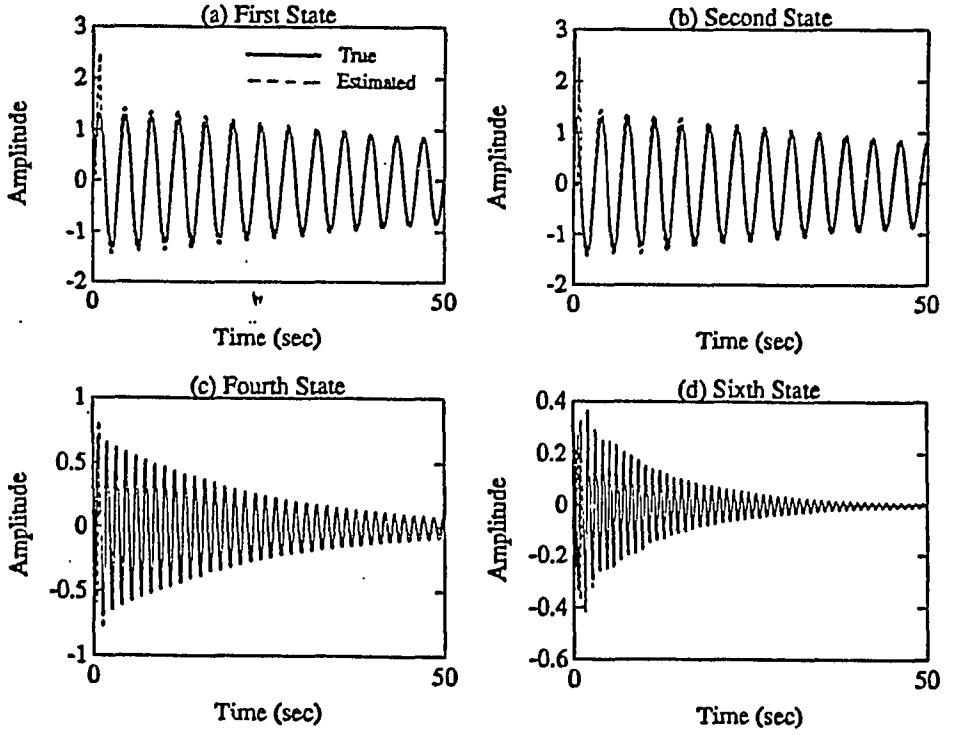


Fig. 4.2 State estimation of a system without process noise by a recursive least-squares filter.

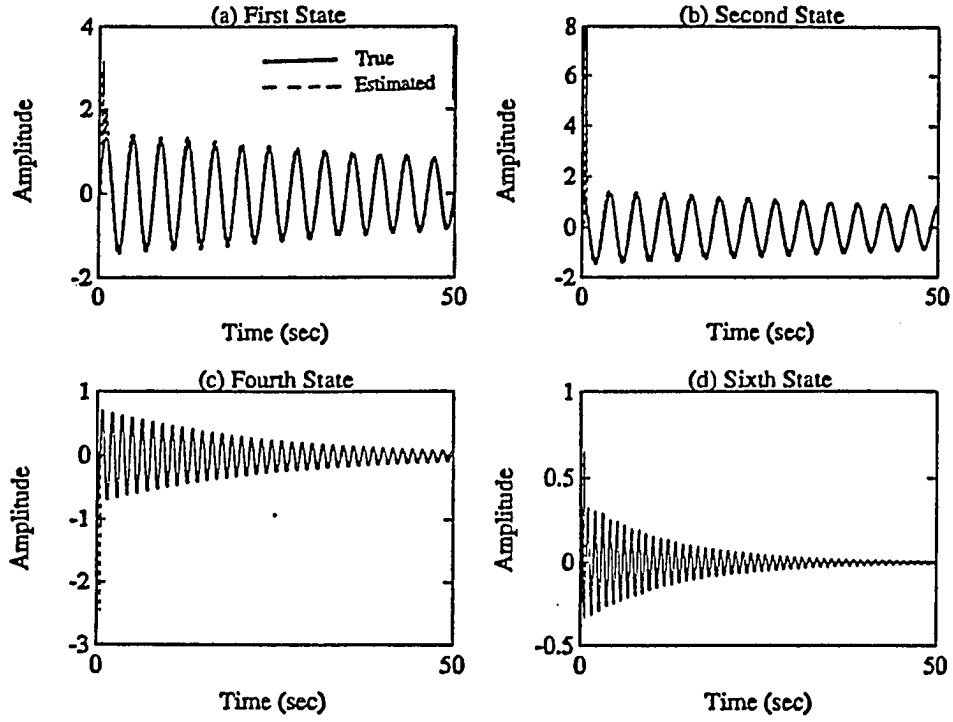


Fig. 4.3 State estimation of a system without process noise by a recursive weighted least-squares filter.

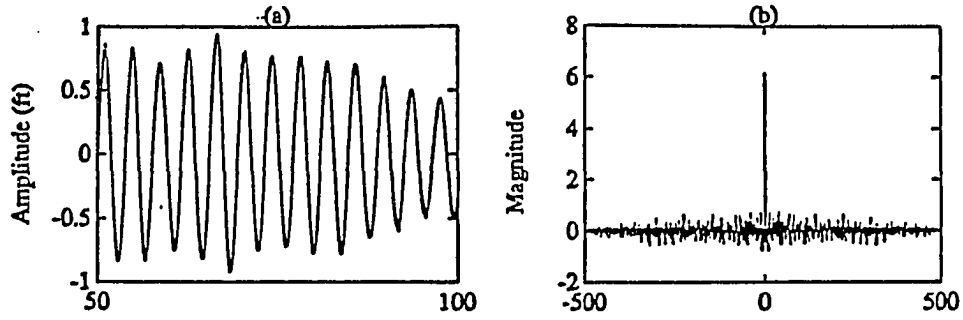


Fig. 4.4 (a) Estimation of the first state by an optimal Kalman filter and (b) Auto-correlation function of the corresponding residual.

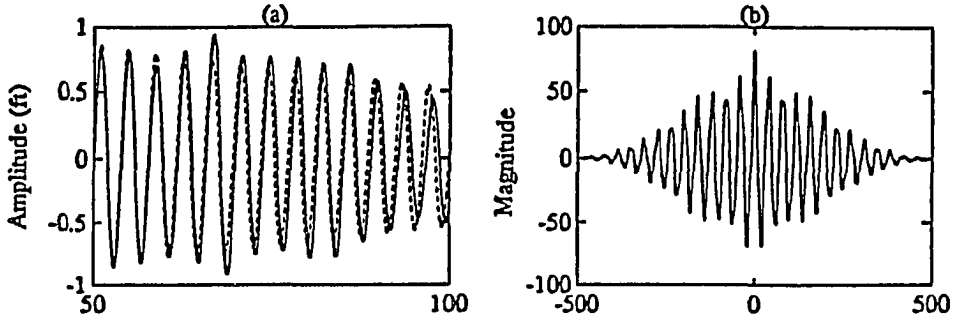


Fig. 4.5 (a) Estimation of the first state by FMLS filter ($\lambda = 1.0$) and (b) Auto-correlation function of the corresponding residual.

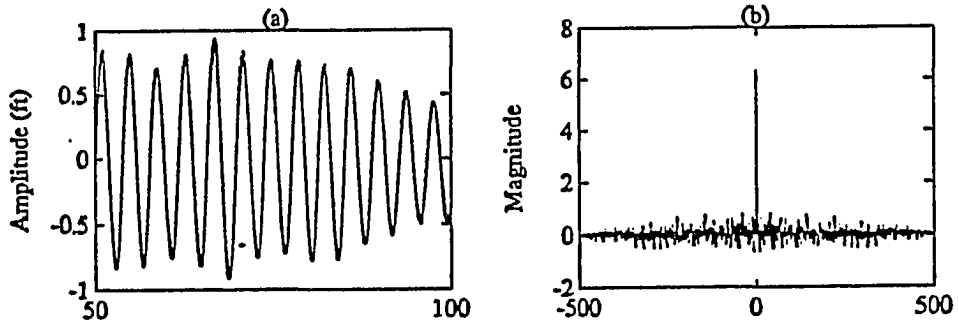


Fig. 4.6 (a) Estimation of the first state by FMLS filter ($\lambda = 0.9$) and (b) Auto-correlation function of the corresponding residual.

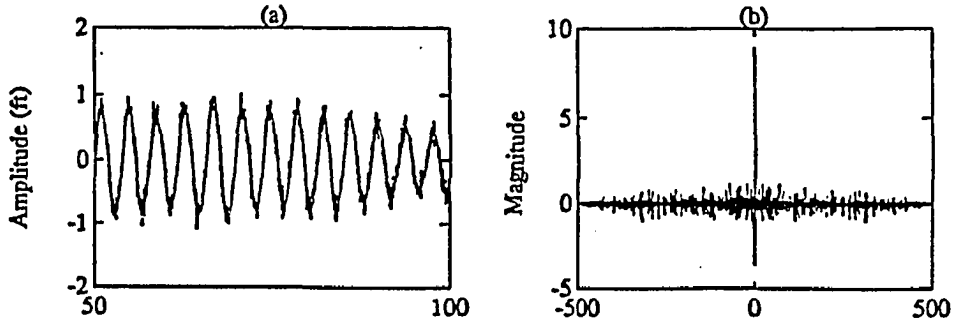


Fig. 4.7 (a) Estimation of the first state by FMLS filter ($\lambda = 0.5$) and (b) Auto-correlation function of the corresponding residual.

Chapter 5

LINEAR STATE ESTIMATION UNDER UNKNOWN NOISE COVARIANCES

—— Optimal Filter Gain Approach

5.1 Introduction

This chapter continues the topic of solving the problem of state estimation under unknown noise covariances in the last chapter. In Chapter 4, the problem is solved by treating previous measurements as of decaying importance, backwards in time, in determining the current state. The method results in a suboptimal filter in general. In this chapter the problem is solved by directly estimating the optimal Kalman filter gain utilizing the relation between state space models and matrix polynomial models of linear systems. Here matrix polynomial model means a system equation whose z-transform is a matrix polynomial equation. The approach in this chapter, therefore, is based on fundamentally different philosophy.

Methods for conducting state estimation under unknown noise covariances can be classified into two categories; one performs estimation of noise covariances or

filter gain once and for all and is suitable for off-line application; the other performs the same estimation continuously or continually during the filter operation and is suitable for on-line application. The latter one is called adaptive filtering, or adaptive Kalman filtering, if the Kalman filter structure is used.

The key point in adaptive Kalman filtering is to find a proper Kalman filter gain corresponding to the current stochastic environment. Usually there are two different approaches. One starts from estimating the covariances of process and measurement noises, and then uses the estimates to compute the filter gain according to the Kalman filter formulations.[10-12] The other estimates the optimal Kalman filter gain directly.[13,14] The first approach is usually theoretically complicated and computationally tedious. Moreover, the estimation of the covariance of process noise results in a non-unique solution, unless some restrictions are imposed on the covariance matrix to reduce the number of unknowns in the covariance matrix.[18] Furthermore, the number of parameters needed to be estimated is usually significantly larger than that in the second approach. For the first approach, two square matrices (the covariance matrices of the process and measurement noises) need to be estimated; while for the second approach, there is only one unknown matrix (the optimal Kalman filter gain). The second approach, by contrast, is simpler and more direct. After all, for the purpose of state estimation, the information of noises is needed for calculating a proper Kalman filter gain only; therefore, it is desirable to achieve the ultimate goal of obtaining the proper filter gain directly without going through the intermediate steps of estimating noise covariances. All the methods developed in this chapter belong to the second approach.

The relation between a state space model and a matrix polynomial model can be derived through the Kalman filter formulations. This relation is very useful in combining advantages inherent in these two different model structures. Generally speaking, a state space model is essential in state estimation. However, estimating

state space parameters or Kalman filter gain using model in state space format directly is a nonlinear optimization problem and is difficult to solve. On the other hand, although a matrix polynomial model can not provide state information, it has a great property of having linear relation between the model parameters and input/output data. Consequently, the estimation of the parameters is a linear optimization problem, which can be solved analytically. For instance, least-squares techniques can be easily employed in estimating the system parameters, or performing linear output predictions. The advantage of using least-squares is that it does not require a priori knowledge about the system and noise, and so the parameter estimation and linear prediction can be performed adaptively. Therefore, identifying a matrix polynomial model is much easier than identifying a state space model. Moreover, the relation between these two models provides ways to extract state space parameters and steady state Kalman filter gain of the system from matrix polynomial parameters.

In section 5.2 the relation between a state space model and an autoregressive with exogeneous input (ARX) model, a special matrix polynomial model, is derived. The least-squares method for identifying scalar linear equations (equations with scalar coefficients) can be easily found in many textbooks.[16,21,22] However, the least-squares for matrix equations are not available. Section 5.3 extends the scalar case to derive a least-squares method for estimating the matrix coefficients of the ARX model. The properties of the estimation are also discussed.

Three methods are developed in this chapter. Section 5.4 describes the first method of this chapter, which utilizes the property of equivalent prediction. This property says that the optimal linear output predictions made by the predictor (filter) in a state space structure and the predictor in a matrix polynomial structure are equivalent.[15] Since the noise covariances are unknown, predictors in a state space structure cannot be constructed. However, an adaptive transversal predictor (ATP), a predictor based on a matrix polynomial model, can be used to yield

optimal one- to r-step-ahead output predictions adaptively using input/output data. Taking these predictions as a reference, the gain of a Kalman filter can be adjusted such that the Kalman filter can produce approximately the same output prediction. Thus, an estimate of the optimal steady state Kalman filter gain can be obtained. This method is suitable for both off-line and on-line use.

Section 5.5 provides the second method of this chapter, which utilizes the relation between the state space parameters and the ARX coefficients. The optimal Kalman filter gain is calculated directly from the estimated coefficients of the ARX model.

Section 5.6 derives the third method of estimating the optimal Kalman filter gain, which utilizes the property of whiteness of the optimal residual. The inversion of a polynomial matrix is used to identify a moving average (MA) model from an autoregressive (AR) model.

Section 5.7 discusses the problem of obtaining the covariances of process and measurement noises after having an estimate of the optimal steady state Kalman filter gain. Section 5.8 gives a summary of this chapter.

5.2 A Relation between State Space and Matrix Polynomial Models

The state space model of a finite-dimensional, linear, discrete, time-invariant stochastic system is re-written for convenience:

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{5.1}$$

$$y_k = Cx_k + v_k, \tag{5.2}$$

where the input term, Bu_k, is included for generality. The assumptions and the meanings of all the notations are the same as those in Sections 2.3 and 3.1.

The same system can also be represented by a stochastic AutoRegressive Moving Average with eXogeneous input model,[15,22,23] or ARMAX for short,

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})e_k, \tag{5.3}$$

where

$$A(q^{-1}) = I_p + A_1 q^{-1} + \cdots + A_{na} q^{-na}, \tag{5.4}$$

$$B(q^{-1}) = B_1 q^{-1} + \cdots + B_{nb} q^{-nb}, \tag{5.5}$$

$$C(q^{-1}) = I_p + C_1 q^{-1} + \cdots + C_{nc} q^{-nc}, \tag{5.6}$$

q^{-1} is a backward shift operator (i.e. $q^{-1}y_k = y_{k-1}$), and na, nb, nc are the orders of the polynomials $A(q^{-1})$, $B(q^{-1})$ and $C(q^{-1})$, respectively. The sequence $\{e_k\}$ is a Gaussian, white noise with zero mean. The term $A(q^{-1})y_k$ is the autoregressive (AR) part, $C(q^{-1})e_k$ the moving average (MA), and $B(q^{-1})u_k$ the eXogeneous (X) because the control signal in economics literature is known as the exogeneous variable.

The state space model provides "inner" messages about the states of the system in addition to the input-output information, while the ARMAX model gives the relation between the input and the output only. Since these two different models describe the same system, they must be related. Indeed, the relation can be obtained through the Kalman filter.

One can write the following filter innovation model, which describes the filter system as driven by the innovation sequence, and the output is the measurement data,

$$\hat{x}_{k+1}^- = A\hat{x}_k^- + Bu_k + AK_k\varepsilon_k \tag{5.7}$$

$$y_k = C\hat{x}_k^- + \varepsilon_k \tag{5.8}$$

where $\hat{x}_k^-$ is the a priori estimate of the state x_k, and the term ε_k is the residual. The quantity ε_k contains the "new" information in the sense that it can not be

obtained from the previous data. Therefore, it is also called "innovation".[30] The $n \times m$ matrix K_k is the Kalman filter gain. Introducing (5.8) into (5.7) yields:

$$\begin{aligned} \hat{x}_{k+1}^- &= A(I_n - K_k C)\hat{x}_k^- + Bu_k + AK_k y_k \\ &= \bar{A}\hat{x}_k^- + Bu_k + AK_k y_k \end{aligned} \tag{5.9}$$

where

$$\bar{A} = A(I_n - K_k C). \tag{5.10}$$

This equation provides another system dynamic equation of the filter other than (5.7), where $\bar{A}$ is the system matrix and y_k the input.

The existence of a steady state Kalman filter gain, K, is guaranteed if the system is detectable and $(A, Q^{1/2})$ is stabilizable.[26] In the implementation of the Kalman filter, one can start from an arbitrary guess of the initial state value and its corresponding error covariance. For a stable filter, the Kalman filter gain will converge exponentially to its steady state value independently of the initial condition.

Introducing (5.9) into (5.8) iteratively, with K_k replaced by the steady state gain K, one can obtain the following input-output description:

$$\begin{aligned} y_k =& C\hat{x}_k^- + \varepsilon_k \\ =& C\bar{A}\hat{x}_{k-1}^- + CBu_{k-1} + CAKy_{k-1} + \varepsilon_k \\ =& \cdots \\ =& CAKy_{k-1} + C\bar{A}AKy_{k-2} + \cdots + C\bar{A}^{q-1}AKy_{k-q} + CBu_{k-1} \\ &+ C\bar{A}Bu_{k-2} + \cdots + C\bar{A}^{q-1}Bu_{k-M} + C\bar{A}^q\hat{x}_{k-q}^- + \varepsilon_k \\ =& \sum_{i=1}^{q} C\bar{A}^{i-1}AKy_{k-i} + \sum_{i=1}^{q} C\bar{A}^{i-1}Bu_{k-i} + C\bar{A}^q\hat{x}_{k-q}^- + \varepsilon_k \end{aligned} \tag{5.11}$$

for some integer q. Note that, although the steady state Kalman filter gain might not be known at the very beginning, it has already existed. This implies that (5.11) is a valid relation even for the very beginning data. In other words, once

the value of every input-output term in (5.11) is known, this equation holds. Matrix $\bar{A}$ in (5.11) is the system matrix of the filter dynamical equation (5.9), where the steady state gain is used instead, and also is the system matrix of the filter error dynamical system.[26] For a stable filter the matrix $\bar{A}$ is asymptotically stable. Therefore, for a sufficiently large number q the term next to the last one of (5.11) is negligibly small and can be dropped out from the equation. Moving all the terms containing output y to the left hand side, (5.11) becomes

$$y_k - \sum_{i=1}^{q} C\bar{A}^{i-1}AKy_{k-i} = \sum_{i=1}^{q} C\bar{A}^{i-1}Bu_{k-i} + \varepsilon_k \tag{5.12}$$

which is a special form of an ARMAX model with $C(q^{-1}) = I_p$, hence called ARX because it has no moving average part. All the coefficients of this model are expressed in terms of the state space parameters A, B, C and the Kalman gain K. Note that the noise term here is the residual of the optimal Kalman filter, which is zero-meaned, white and not correlated with previous output data according to the orthogonality principal in estimation.

5.3 Estimation of the Coefficients of an ARX Model

Estimation of the coefficient matrices of an ARX model given in (5.12) can be accomplished by using an adaptive transversal predictor[16] (ATP) which is shown in Fig. 5.1. The name arose due to the structure of the filter. In fact, it is a recursive least-squares filter. This filter sequentially feeds the measurements, inputs and their delay versions to its q tap inputs for a filter of order q. Each tap input signal is multiplied by the tap coefficient matrix, and the results are summed up to yield the filter output, which is a one-step-ahead output prediction. The prediction error is then fed back to modify the tap coefficient matrices in the next recursion. The adaptive process can be explained by the following least-squares method.

Equation (5.12) can be written as

$$\begin{aligned} y_k &= \sum_{i=1}^{q} C\bar{A}^{i-1}AKy_{k-i} + \sum_{i=1}^{q} C\bar{A}^{i-1}Bu_{k-i} + \varepsilon_k \\ &= \hat{y}_k + \varepsilon_k, \end{aligned} \tag{5.13}$$

which can be interpreted in two different ways. On one hand, it can be regarded as a signal generator, where y_k is synthesized by using finite previous input/output data $\{u_{k-1}, \cdots, u_{k-q}, y_{k-1}, \cdots, y_{k-q}\}$, and white noise ε_k. On the other hand, it can also be viewed as a linear predictor, where $\hat{y}_k$ is a prediction of y_k and ε_k is the prediction error. The output y_k can be thought as truly coming out from a linear transversal process generator driven by the known deterministic input and unknown white noise.

Equation (5.13) can also be written in a compact form:

$$y_k = \Theta_0^T \Phi_k^T + \varepsilon_k$$

or

$$y_k^T = \Phi_k \Theta_0 + \varepsilon_k^T, \tag{5.14}$$

where

$$\Theta_0^T = [CAK, \cdots, C\bar{A}^{q-1}AK, CB, \cdots, C\bar{A}^{q-1}B], \tag{5.15}$$

$$\Phi_k = [y_{k-1}^T, \cdots, y_{k-q}^T, u_{k-1}^T, \cdots, u_{k-q}^T]. \tag{5.16}$$

Vector Φ, called a regressor, is composed of q previous input/output data. The parameter matrix Θ_0 is to be estimated from output data $\{y_k\}$ and regressors $\{\Phi_k\}$ ($k = q, \cdots, N$, N is the number of total data used in the estimation). To this end, first we define a scalar cost function

$$C_N = \frac{1}{N-q+1} \sum_{k=q}^{N} \lambda_k (y_k^T - \Phi_k \hat{\Theta})(y_k^T - \Phi_k \hat{\Theta})^T \tag{5.17}$$

then we minimize this with respect to $\hat{\Theta}$, where $\hat{\Theta}$ denotes the estimated parameter matrix and $\{\lambda_k\}$ is a sequence of weighting factors. These weighting factors allow

us to give different weights to measurements of different time, thus providing the capability of identifying slowly time-variant systems.

Denoting the measurement vector, y_k, by $y_k = [y_{k1}, \cdots, y_{kp}]^T$ and the parameter matrix, $\hat{\Theta}$, by $\hat{\Theta} = [\hat{\theta}_1, \cdots, \hat{\theta}_p]$, where y_{ki} is the i-th entry of y_k and $\hat{\theta}_j$ the j-th column of $\hat{\Theta}$, (5.17) can be rewritten as

$$C_N = \frac{1}{N-q+1} \sum_{i=1}^{p} \sum_{k=q}^{N} \lambda_k (y_{ki} - \Phi_k \hat{\theta}_i)^2. \tag{5.18}$$

Note that the cost function C_N is composed of p summation and each summation is a quadratic function of a different column vector of $\hat{\Theta}$. Therefore, it can be minimized analytically.

Minimizing C_N with respect to $\hat{\theta}_j$ gives

$$\hat{\theta}_j = \left(\sum_{k=1}^{N} \lambda_k \Phi_k^T \Phi_k \right)^{-1} \sum_{k=q}^{N} \lambda_k \Phi_k^T y_{kj}, \tag{5.19}$$

provided the inverse exists. Therefore,

$$\hat{\Theta}_N = [\hat{\theta}_1, \cdots, \hat{\theta}_p] = \left(\sum_{k=q}^{N} \lambda_k \Phi_k^T \Phi_k \right)^{-1} \sum_{k=q}^{N} \lambda_k \Phi_k^T y_k^T, \tag{5.20}$$

and this is the basic formulation for the parameter estimation.

To discuss the properties of the parameter estimation, substitute y_k in (5.14) into (5.20) to give

$$\hat{\Theta}_N = \Theta_0 + \left(\sum_{k=q}^{N} \lambda_k \Phi_k^T \Phi_k \right)^{-1} \sum_{k=q}^{N} \lambda_k \Phi_k^T \varepsilon_k^T, \tag{5.21}$$

where Θ_0 is the true parameter matrix. According to the orthogonal principle of optimal estimation, the residual sequence $\{\varepsilon_k\}$ is uncorrelated with the previous measurements. If the input $\{u_k\}$ is uncorrelated with $\{\varepsilon_k\}$ also, then

$$E[\hat{\Theta}_N] = \Theta_0 + E\left[\left(\sum_{k=q}^{N} \lambda_k \Phi_k^T \Phi_k \right)^{-1} \sum_{k=q}^{N} \lambda_k \Phi_k^T \right] E\left[\varepsilon_k^T\right] = \Theta_0, \tag{5.22}$$

where $E[\varepsilon_k] = 0$ is used. This equation indicates that the estimation is unbiased. Furthermore, the term $\sum_{k=q}^{N} \Phi_k \varepsilon_k^T$ in (5.21) can be viewed as calculating the sample correlation, which will approach its expectation value (zero), as the number of data N increases to infinity. This in turn means that the estimated parameter matrix will asymptotically converge to the true value. This property is referred to as *"p-consistent"* [31] in the literature, which means that as the number of data tends to infinity the estimate converges almost surely to a matrix which, in turn, converges to the true parameter matrix as the order of the ARX model tends to infinity. The order of the ARX model in the original paper [31] is denoted by p and hence the name.

In implementing the transversal predictor, its order should be determined in advance. It should be sufficiently large in order to yield satisfactory result. A method based on the information-theoretic criterion (AIC) is commonly used.[32,33]

5.4 Equivalent Prediction Method

Linear optimal prediction can be made by using either a state space model or a matrix polynomial model.[15] In this section, based on the equivalence of the predictions made by different models, the Kalman filter gain is chosen so that the predictions made by the Kalman filter can match that made by the adaptive transversal predictor optimally in a least-squares sense. The Kalman filter is based on a state space model, and the transversal predictor is based on ARX. The reason for using the adaptive transversal predictor is to take advantage of its adaptive feature which requires no a priori knowledge about the system.

5.4.1 Linear Prediction by a State Space Model

If a state space model is chosen for linear prediction, a simple mechanism based on the Kalman filter can be used. For one-step-ahead prediction, the Kalman filter innovation model shown in (5.7) and (5.8) can provide the answer:

$$\hat{x}_{k+1}^{-} = A\hat{x}_k^{-} + Bu_k + AK_k\varepsilon_k,$$

$$y_k = C\hat{x}_k^{-} + \varepsilon_k = \hat{y}_k + \varepsilon_k,$$

where $\hat{y}_k$ is the optimal prediction of output y_k based on all the previous data.

For r-step-ahead prediction ($r \geq 2$) the system dynamical equation (5.1), omitting the input and noise terms, is used to propagate the state to the future, and the output equation (5.2), omitting the noise term, is used to yield the prediction. For instance, the predicted r-step-ahead state and output are

$$\hat{x}_{k+r-1}^{(r)} = A^{r-1}\hat{x}_k^{-}, \tag{5.23}$$

$$\hat{y}_{x+r-1}^{\prime(r)} = C\hat{x}_{k+r-1}^{(r)}, \tag{5.24}$$

where $\hat{x}_i^{(j)}$ and $\hat{y}_i^{\prime(j)}$ denote the j-step-ahead predictions of x_i and y_i, respectively. To avoid confusion, the superscript $'$ in $\hat{y}_i^{\prime(r)}$ is used to denote the prediction made by the Kalman filter, thus distinguishing it from that made by a ARX model.

Unfortunately, to perform prediction, the optimal Kalman filter gain should be obtained first, which requires the knowledge of the statistics of the process and measurement noises. This is the restriction in using state space model for linear output prediction.

5.4.2 Linear Prediction by an ARX Model

If an ARX model is chosen in linear prediction, (5.13) is used to have

$$\hat{y}_k = \sum_{i=1}^{q} C\bar{A}^{i-1}AKy_{k-i} + \sum_{i=1}^{q} C\bar{A}^{i-1}Bu_{k-i} \tag{5.25}$$

where $\hat{y}_k$ represents the one-step-ahead output prediction of y_k.

Two or more steps-ahead predictions using the ARX model can be made by iteratively using (5.25), replacing true measurements with their predicted versions and omitting the future input terms, because the future input is not known for the current moment. By doing so the model plays a role in propagating the output to the future based on the currently available data. For instance, the r-step-ahead prediction ($r \geq 2$) is

$$\begin{aligned}\hat{y}_{k+r-1}^{(r)} =& CAK\hat{y}_{k+r-2}^{(r-1)} + \cdots + C\bar{A}^{r-2}AK\hat{y}_k + C\bar{A}^{r-1}AKy_{k-1} \\ &+ \cdots + C\bar{A}^{q-1}AKy_{k-m+r-1} + C\bar{A}^{r-1}Bu_{k-1} \\ &+ \cdots + C\bar{A}^{q-1}Bu_{k-m+r-1}. \end{aligned} \tag{5.26}$$

To perform prediction using (5.26), the coefficients of the ARX model should be known also, which is impossible if the noise covariance is unknown. However, by using the adaptive transversal predictor, the optimal one- to r-step-ahead predictions can be obtained adaptively. The "optimal" prediction here is in the sense that the current estimated coefficients of the ARX model, which are used in making the prediction, can fit the input/output data with the least sum of the squares of the error. The adaptive prediction uses the currently estimated model instead of the true one. As a result, the prediction might not be accurate in the beginning; however, better predictions can be expected as more input/output data are processed.

5.4.3 Equivalence of Linear Predictions

This section provides a proof of the equivalence between the predictions made by the Kalman filter and that made by the ARX model. For one-step-ahead prediction shown in (5.11), the Kalman filter prediction $C\hat{x}_k^-$ can be approximately expanded to a matrix polynomial of a finite order q. For a sufficiently large integer

q the difference is negligibly small. We can, therefore, say they are equivalent.

For r-step-ahead prediction ($r \geq 2$), from the Kalman filter prediction the following equation can be derived:

$$\begin{aligned}\hat{y}^{\prime(r)}_{k+r-1} &= CA^{r-1}\hat{x}^{-}_{k} \\ &= \sum_{i=1}^{q} CA^{r-1}\bar{A}^{i-1}AKy_{k-i} + \sum_{i=1}^{q} CA^{r-1}\bar{A}^{i-1}Bu_{k-i}\end{aligned} \tag{5.27}$$

where $\hat{x}^{-}_{k}$ is expanded as shown in (5.11).

On the other hand, if the prediction is made based on ARX model according to (5.26), for $r = 2$ we have

$$\begin{aligned}\hat{y}^{(2)}_{k+1} =& CAK\hat{y}_k + \sum_{i=2}^{q} C\bar{A}^{i-1}AKy_{k-i+1} + \sum_{i=2}^{q} C\bar{A}^{i-1}Bu_{k-i+1} \\ =& CAK\left(\sum_{i=1}^{q} C\bar{A}^{i-1}AKy_{k-i} + \sum_{i=1}^{q} C\bar{A}^{i-1}Bu_{k-i}\right) \\ &+ \sum_{i=2}^{q} C\bar{A}^{i-1}AKy_{k-i+1} + \sum_{i=2}^{q} C\bar{A}^{i-1}Bu_{k-i+1} \\ =& \sum_{i=1}^{q-1}\left(CAKC\bar{A}^{i-1}AK + C\bar{A}^{i}AK\right)y_{k-i} \\ &+ \sum_{i=1}^{q-1}\left(CAKC\bar{A}^{i-1}B + C\bar{A}^{i}B\right)u_{k-i} \\ &+ CAKC\bar{A}^{q-1}AKy_{k-q} + CAKC\bar{A}^{q-1}Bu_{k-q},\end{aligned} \tag{5.28}$$

where $\hat{y}_k$ is expanded using (5.25), and the coefficients of the same variables are collected together. The coefficients of the first summation can be simplified to be

$$\begin{aligned}CAKC\bar{A}^{i-1}AK + C\bar{A}^{i}AK &= C(AKC + \bar{A})\bar{A}^{i-1}AK \\ &= C[AKC + A(I_n - KC)]\bar{A}^{i-1}AK \\ &= CA\bar{A}^{i-1}AK,\end{aligned} \tag{5.29}$$

and for the second summation, similarly,

$$CAKC\bar{A}^{i-1}B + C\bar{A}^{i}B = CA\bar{A}^{i-1}B. \tag{5.30}$$

Moreover, for a sufficiently large q, the last two terms in (5.28) can be neglected. Therefore, (5.28) becomes

$$\hat{y}_{k+1}^{(2)} = \sum_{i=1}^{q-1} CA\bar{A}^{i-1}AKy_{k-i} + \sum_{i=1}^{q-1} CA\bar{A}^{i-1}Bu_{k-i}, \tag{5.31}$$

which is the same as (5.27) for $r = 2$ except the upper limit of the summation is one term less. The difference can be negligibly small by having large q.

Following the same pattern, the p-step-ahead prediction using the ARX model can be written as

$$\begin{aligned}\hat{y}_{k+r-1}^{(r)} &= \sum_{i=1}^{r-1} C\bar{A}^{i-1}AK\hat{y}_{k+r-1-i}^{(r-i)} + \sum_{i=r}^{q} C\bar{A}^{i-1}AKy_{k+r-1-i} \\ &\quad + \sum_{i=r}^{q} C\bar{A}^{i-1}Bu_{k+r-1-i} \\ &= \sum_{i=1}^{q-r+1} CA^{r-1}\bar{A}^{i-1}AKy_{k-i} + \sum_{i=1}^{q-r+1} CA^{r-1}\bar{A}^{i-1}Bu_{k-i},\end{aligned} \tag{5.32}$$

which is essentially the same as (5.27), provided q is large and $q >> r$. Therefore, it has been proved that the prediction made by the Kalman filter according to (5.24) and that made by the ARX model according to (5.26) are equivalent.

5.4.4 Obtaining Preliminary State Estimation

The optimal Kalman filter is derived under the stochastic framework. In other words if the initial values and the noise statistics are unchanged, the Kalman filter is the same for all realizations of input/output set of a process. On the other hand, the adaptive transversal predictor, a recursive least-squares filter, is derived under the deterministic framework and yields different filter for each realization of input/output data. Nevertheless, the recursive least-squares algorithm utilizes all the information contained in the input/output data to update its filter parameters. Therefore, although in the beginning the adaptive transversal predictor yields less

accurate predictions compared to those from the optimal Kalman filter, as more data processed, the prediction will gradually converge to that made by the optimal Kalman filter. This fact can be stated as follows:

Denote the one-step-ahead prediction made by the ARX model of order q with accurate coefficients by $\hat{y}_{q,k}$ (i.e., $\hat{y}_{q,k} = \Theta_{q,0}^T \Phi_{q,k}$), and that made by adaptive transversal predictor of order q with estimated coefficients by $\hat{\hat{y}}_{q,k}$ (i.e., $\hat{\hat{y}}_{q,k} = \hat{\Theta}_{q,k}^T \Phi_{q,k}$). The matrices $\Theta_{q,0}$ and $\Phi_{q,k}$ are the same as Θ_0 and Φ_k, respectively, in (5.15) and (5.16) except the subscript q is used here to explicitly specify the order of the ARX model. From (5.21) we have

$$\hat{\Theta}_{q,N} \longrightarrow \Theta_{q,0} \qquad \text{as} \quad N \longrightarrow \infty, \tag{5.33}$$

therefore,

$$\hat{\hat{y}}_{q,k} = \hat{\Theta}_{q,k}^T \Phi_{q,k}^T \longrightarrow \Theta_{q,0}^T \Phi_{q,k}^T = \hat{y}_{q,k} \qquad \text{as} \quad k \longrightarrow \infty, \tag{5.34}$$

where "$\longrightarrow$" means "converges to". However, from (5.27) and (5.32) it is obvious that

$$\hat{y}_{q,k} \longrightarrow \hat{y}'_k \qquad \text{as} \quad q \longrightarrow \infty. \tag{5.35}$$

Therefore, for a sufficiently large q and long input/output data, the one-step-ahead prediction of the adaptive transversal predictor will converge to that of the optimal Kalman filter. This is also true for r-step-ahead prediction because r-step-ahead prediction is based on one-step-ahead prediction. As the coefficient matrices converge to their true values all the one- to r-step-ahead predictions converge to the optimal Kalman filter predictions.

Without a priori information about the statistics of the noises the optimal Kalman filter prediction cannot be made. However, based on the theory just proved, the adaptive transversal predictor can provide approximate answers. Therefore, one can write

$$\hat{\hat{y}}_{k+r-1}^{(r)} \approx CA^{r-1}\hat{x}_k^-, \tag{5.36}$$

where the subscript q is omitted for simplicity. From (5.36) one can derive the following equation:

$$\begin{bmatrix} \hat{\hat{y}}_k \\ \hat{\hat{y}}^{(2)}_{k+1} \\ \vdots \\ \hat{\hat{y}}^{(r)}_{k+r-1} \end{bmatrix} \approx \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r-1} \end{bmatrix} \hat{x}_k^-, \tag{5.37}$$

or in short,

$$\hat{\hat{Y}}_r \approx V\hat{x}_k^- \tag{5.38}$$

where $\hat{\hat{Y}}_r$ denotes the prediction vector obtained from the adaptive transversal predictor and V the observability-type matrix. If the system is observable, and integer r is sufficiently large to make matrix V full-column-ranked, a least-square solution of $\hat{x}_k^-$, can be obtained by

$$\hat{\hat{x}}_k = V^{\dagger}\hat{\hat{Y}}_r, \tag{5.39}$$

where $V^{\dagger}$ is the pseudoinverse of V and $\hat{\hat{x}}_k$ an estimate of the one-step-ahead optimal Kalman filter prediction, $\hat{x}_k^-$.

At this stage an estimation of $\hat{x}_{k-1}^+$ can be obtained by propagating $\hat{\hat{x}}_k$ backward as

$$\hat{\hat{x}}_{k-1}^+ = A^{-1}(\hat{\hat{x}}_k - Bu_{k-1}), \tag{5.40}$$

where $\hat{\hat{x}}_{k-1}^+$ denotes an estimate of the optimal Kalman filter a posteriori estimation $\hat{x}_{k-1}^+$. However, since $\hat{\hat{x}}_k$ is only an estimate of $\hat{x}_k^-$, the estimated state $\hat{\hat{x}}_{k-1}^+$ could be rather fluctuating. Besides, the computational load for each estimate is much heavier than that of the Kalman filter if the state estimation works entirely by this way. A better alternative is to use the estimated prediction to "train" the Kalman filter gain, and the trained filter gain, after it converges to some extent, is used to run constant gain Kalman filter state estimation. By this way a smoother state estimation with less computation can be obtained.

5.4.5 Estimating Optimal Kalman Filter Gain

From (5.7), for steady state, one can have

$$K\varepsilon_{k-1} = A^{-1}(\hat{x}_k^- - Bu_{k-1}) - \hat{x}_{k-1}^- \triangleq d_{k-1}. \tag{5.41}$$

Replacing $\hat{x}_i^-$ by $\hat{\hat{x}}_i$, ε_i by $\hat{\varepsilon}_i$ ($\hat{\varepsilon}_i = y_i - \hat{\hat{y}}_i$), and collecting records through time, one can have the following equation:

$$K[\hat{\varepsilon}_{k-1},\ \hat{\varepsilon}_{k-2}, \cdots,\ \hat{\varepsilon}_{k-s}] = [\hat{d}_{k-1},\ \hat{d}_{k-2},\ \cdots,\ \hat{d}_{k-s}], \tag{5.42}$$

or in short,

$$KE_s = D_s \tag{5.43}$$

where E_s denotes $[\hat{\varepsilon}_{k-1},\ \hat{\varepsilon}_{k-2}, \cdots,\ \hat{\varepsilon}_{k-s}]$, D denotes $[\hat{d}_{k-1},\ \hat{d}_{k-2},\ \cdots,\ \hat{d}_{k-s}]$, $\hat{d}_k$ is an approximation of d_k defined in (5.41) when the optimal Kalman predictions are replaced by their estimates, and s the number of data point. Matrix E_s has a dimension $p \times s$, where p is the number of output. The sequence $\{\hat{\varepsilon}_i\}$ is approximately white; therefore, for $s > p$, E_s is full-row-ranked in general. Then from (5.43) a least-square solution of K is

$$\hat{K} = D_s E_s^{\dagger} \tag{5.44}$$

where $E_s^{\dagger} = E_s^T(E_s E_s^T)^{-1}$ denotes the pseudoinverse of E_s.

Equation (5.44) can also be solved recursively. By doing so, as more data processed, the estimated gain $\hat{K}$ can be improved, and the state estimation is thus in turn improved.

With this estimated gain, state estimation can be carried out by using a constant gain Kalman filter. The state estimator is as the followings:

$$\hat{x}_k^- = A\hat{x}_{k-1}^- + Bu_{k-1}, \tag{5.45}$$

$$\hat{x}_k^+ = \hat{x}_k^- + \hat{K}(y_k - C\hat{x}_k^-). \tag{5.46}$$

5.4.6 Forgetting Factor

A well-known feature of the adaptive transversal filter is that, by properly introducing a forgetting factor, the filter can tract a slow time varying system.[21-23] The forgetting factor has a function to let the filter put less weighting on the older data and thus gradually "forget" them. It is actually a recursive weighted least-squares method where the importance of the data at a specific point decreases exponentially as time increases. The value of the forgetting factor is a number very close to but less than 1. As the factor is closer to 1, the filter has less forgetting function. It is somewhat subjective to choose the value. The proper choice can be obtained by monitoring the output prediction error. For a suitable value, the prediction error should be close to a white sequence.

If the forgetting factor is used in the recursive process of updating a Kalman filter gain described in section 5.4.5, the proposed method can deal with the situation when the noise statistics are slowly changing. This is a great advantage of this approach.

5.4.7 Numerical Examples

In the numerical examples, the dynamical system is the same as that in the last two chapters, except input force is added this time. The system is excited by random force u at node 3, while the responses are measured at nodes 1 and 2 (see Fig. 5.2). The state space parameters including B are re-listed here:

$$A = diag\left\{\begin{bmatrix} 0.9856 & 0.1628 \\ -0.1628 & 0.9856 \end{bmatrix} \begin{bmatrix} 0.8976 & 0.4305 \\ -0.4305 & 0.8976 \end{bmatrix} \begin{bmatrix} 0.8127 & 0.5690 \\ -0.5690 & 0.8127 \end{bmatrix}\right\}$$

$$B = [0.0011 \quad 0.0134 \quad -0.0016 \quad -0.0072 \quad 0.0011 \quad 0.0034]^T$$

$$C = \begin{bmatrix} 1.5119 & 0.0000 & 2.0000 & 0.0000 & 1.5119 & 0.0000 \\ 1.3093 & 0.0000 & 0.0000 & 0.0000 & -1.3093 & 0.0000 \end{bmatrix}$$

where A is a block diagonal matrix. The sampling frequency is 10 Hz, which is sufficient high for estimating the state of the highest frequency (0.97 Hz) of the system.

The variance of the random excitation force, σ_u^2, is set to 40. The covariances of process noise and measurement noise are

$$Q = diag[\ 0.0024,\ \ 0.3593,\ \ 0.0053,\ \ 0.1035,\ \ 0.0023,\ \ 0.0228\ ] \times 10^{-3},$$

$$R = 0.0279 \times I_2.$$

The standard deviation of the process noise w_k is about 23% of that of the input influence Bu_k; the standard deviation of the measurement noise v_k is about 10% of that of the output measurement y_k. Under these settings, the theoretical optimal steady state Kalman filter gain is

$$K = \begin{bmatrix} 0.0604 & 0.0279 & 0.0471 & 0.0146 & 0.0132 & 0.0055 \\ 0.0648 & 0.0366 & -0.0059 & 0.0143 & -0.0162 & -0.0011 \end{bmatrix}^T.$$

The filter order of the adaptive transversal predictor is set to 100. After the adaptive transversal predictor has processed 1,000 input/output data, its yields an estimation of the coefficients of the ARX model as shown in Figs. 5.3 and 5.4. Figure 5.3 shows the four elements of matrix sequence $C\bar{A}^{i-1}AK$ while Fig. 5.4 shows the two elements of $C\bar{A}^{i-1}B$, where $i = 1, \cdots, 100$. The estimated sequences are still very "noisy". To show the estimation is convergent, Figs. 5.5 and 5.6 show the results after processing 5,000 data. Apparently the estimation has been greatly improved. The estimation error variances of the sequences against the number of data processed are shown in Figs. 5.7 and 5.8, where Fig. 5.7 shows the sequences of $C\bar{A}^{i-1}AK$ and Fig. 5.8 shows that of $C\bar{A}^{i-1}B$. The convergence of the estimation is clearly shown in this figure. Using the method derived in this section, two cases are studied.

Case 1: *Off-line Batch Estimation*

In the first case a batch type treatment is performed. First, a batch of input/output data is used to "train" the adaptive transversal predictor; in other words, the data is processed using the ATP to estimate the ARX model. Then the estimated ARX model is used to perform one- to r-step-ahead output predictions using the same set of input/output data. The optimal Kalman filter gain is then obtained from the output predictions based on the method.

In this simulation, one- to ten-step-ahead predictions are performed. Depending on the size of the batch, different results are obtained. In general, the bigger the size is, the better the result will be. One way to compare the qualities of the estimated gains is to compare their "distances" to their theoretical optimal value. The distance of two matrices can be represented by the 2-norm of their difference matrix. The batch sizes of 1,000 to 5,000 data, with an increment of 1,000 data, are examined. The norms of the difference matrices of the results obtained based on different batch sizes are plotted against the batch size in Fig. 5.9. As the batch size increases, the norm reduces; however, it seems that the results saturate after 3000 data.

Another way of verifying the quality of the estimated gain is to use it in the Kalman filter and check the whiteness of the residual, or, in a simulation case, compare the state estimation and the residual with those of the optimal Kalman filter. Using the gains obtained from the five different batch sizes in the Kalman filter, the trace of the corresponding state error covariances along with that of the optimal filter are listed below. For comparison's sake, the same input/output data are used in each case and the state error covarances are calculated by averaging 700 samples.

Batch size	1,000	2,000	3,000	4,000	5,000	optimal
State error	0.0084	0.0072	0.0068	0.0068	0.0070	0.0070

From the state errors we can see that though the estimated gains obtained from

2000 data and more are not exactly the same as the theoretical optimal one, the state errors are almost the same as that of the optimal filter. From this fact we can say that the optimality of state estimation is not sensitive to some small deviations in the optimal gain.

For comparison's sake, the state estimation of the first state, a part of the first residual sequence, and its corresponding auto- correlation function (calculated from 500 samples) of the optimal Kalman filter is shown in Fig. 5.10. The counterparts of the case which uses the gains estimated from batches of 1,000 and 5,000 data are shown in Figs. 5.11 and 5.12. In the residual plots, Figs. 5.11(b) and 5.12(b), the solid lines represent the optimal residual from the optimal Kalman filter (the same as in Fig. 5.10(b)), and the dashed line represents the estimated residuals (from the filters using the estimated gains). For these two cases, the estimated states and residuals almost coincide with the real state and the optimal residual respectively, and the auto-correlation functions show the residuals are rather white. Therefore, we can conclude that the estimated gains are satisfactory. The corresponding estimated gains in these cases are

$$\hat{K}_{1000} = \begin{bmatrix} 0.0887 & 0.0089 & 0.0242 & -0.0007 & 0.0135 & 0.0147 \\ 0.0668 & 0.0583 & -0.0151 & 0.0358 & -0.0218 & -0.0038 \end{bmatrix}^T .$$

$$\hat{K}_{5000} = \begin{bmatrix} 0.0712 & 0.0132 & 0.0402 & 0.0077 & 0.0100 & 0.0033 \\ 0.0650 & 0.0280 & -0.0073 & 0.0220 & -0.0241 & 0.0011 \end{bmatrix}^T .$$

Case 2: *On-line Recursive Estimation*

In the on-line case, when the input/output data are available, the ATP updates its tap parameters and makes one to r-step-ahead output predictions based on the current values of the parameters. The predictions are inaccurate in the beginning but will keep on improving. Skipping the transient period where the predictions could be extremely bad (indicated by the non-stationary part in the beginning of the ATP residual), a Kalman gain estimator starts to estimate the Kalman filter gain using the predictions. Because the estimation of the ARX

model is convergent, the quality of the current predictions is always better than those made earlier. Therefore, a forgetting factor is used in calculating the Kalman gain to put emphasis on recent predictions. After converging to a certain degree, the estimated gain can be adopted in a Kalman filter and starts state estimation. Before the adoption, the stability of the filter in using the gain should be checked to avoid divergence. Thereafter, the filter gain can be replaced from time to time by a newer gain provided by the Kalman gain estimator.

This on-line process is simulated using 5,000 data. The Kalman gain estimator starts to operate after 500 data has been processed in the ATP. The forgetting factor is set to 0.999. The norms of the difference matrices between the optimal and estimated gains are plotted against the number of data processed in the ATP and shown in Fig. 5.13. As expected, the result improves as the number of data increases. Compared with the norm in Case 1, the on-line approach needs more data to achieve the same value of norm. This is reasonable because the on-line approach estimates the gain adaptively, that is, it cannot re-process previous data. Therefore, the predictions made earlier cannot take advantage of the currently updated model. Consequently, the predictions from which the gain is estimated are less accurate than that in the batch-type approach.

Similarly, the state estimations using estimated gains and the corresponding residual comparisons are shown in Figs. 5.14 and 5.15. Figure 5.14 shows the result of using a gain obtained after 2,000 data are processed, while Fig. 5.15 shows that after 5,000 data are processed. We skip the gain obtained with 1,000 data because the filter is not stable with that gain. For comparison, we use the same set of data as in case one to conduct state estimation. In Fig. 5.14, though the state estimation of the first state (Fig. 5.14(a)) seems good, the estimated residual (dashed line) does not quite agree with the optimal version (solid line) in Fig. 5.14(b). This is caused by the estimation errors in some other state. As a result, the residual is not white, which can be seen from its auto-correlation

function (Fig. 5.14(c)). The estimated gain is not good enough in this case. In Fig. 5.15, the results have been greatly improved. The traces of the state error covariances of these two cases are 0.0409 and 0.0092, respectively. The gains used in these two cases are

$$\hat{K}_{2000} = \begin{bmatrix} 0.0797 & 0.0314 & 0.0725 & -0.0159 & 0.0122 & -0.0159 \\ 0.1590 & -0.0324 & -0.0767 & -0.0378 & -0.0332 & 0.0619 \end{bmatrix}^T,$$

$$\hat{K}_{5000} = \begin{bmatrix} 0.0622 & 0.0270 & 0.0613 & 0.0178 & 0.0124 & -0.0100 \\ 0.0983 & -0.0001 & -0.0329 & 0.0007 & -0.0265 & 0.0228 \end{bmatrix}^T.$$

5.5 The ARX Coefficient Method

The ARX coefficient method calculates the optimal steady state Kalman filter gain directly from the tap parameters of the adaptive transversal predictor, the estimated coefficients of the ARX model, using the relation between the ARX model and the state space model. It is simpler than the method described in the last section.

5.5.1 Obtaining Kalman Gain from the Estimated ARX Coefficients

From the adaptive transversal predictor two sets of coefficient matrices are obtained:

$$\hat{S}_1 = \{\widehat{CAK},\ \widehat{C\bar{A}AK}, \cdots,\ \widehat{C\bar{A}^{q-1}AK}\} \tag{5.47}$$

$$\hat{S}_2 = \{\widehat{CB},\ \widehat{C\bar{A}B}, \cdots,\ \widehat{C\bar{A}^{q-1}B}\}, \tag{5.48}$$

where "$\wedge$" denotes estimated value. The following relations provide a method of obtaining an estimate of the optimal Kalman filter gain from set $\hat{S}_1$.

Denote the true value set corresponding to $\hat{S}_1$ by S_1. From S_1 the elements of the matrix sequence CA^iK, $i = 1, \cdots, q$ can be calculated recursively. Note the first element of S_1 is CAK, and

$$CA^{j+1}K = C\bar{A}^jAK + \sum_{i=1}^{j} C\bar{A}^{j-i}AKCA^iK, \tag{5.49}$$

where $j = 1, \cdots, q-1$. Denoting the j-th element in S_1 by $S_{1,j}$, (5.49) can be written as

$$CA^{j+1}K = S_{1,j+1} + \sum_{i=1}^{j} S_{1,j+1-i}CA^iK, \tag{5.50}$$

which clearly shows that all the information needed to compute $CA^{j+1}K$ can be obtained from S_1 and previous calculations.

Proof of (5.49):

$$\begin{aligned}
&C\bar{A}^jAK + \sum_{i=1}^{j} C\bar{A}^{j-i}AKCA^iK \\
&\quad = C\bar{A}^jAK + C\bar{A}^{j-1}AKCAK + C\bar{A}^{j-2}AKCA^2K + \cdots + CAKCAK \\
&\quad = C(\bar{A}^j + \bar{A}^{j-1}AKC + \bar{A}^{j-2}AKCA + \cdots + AKCA^{j-1})AK \\
&\quad = C(\bar{A}^{j-1}A + \bar{A}^{j-2}AKCA + \cdots + AKCA^{j-1})AK \\
&\quad = C(\bar{A}^{j-2}A^2 + \cdots + AKCA^{j-1})AK \\
&\quad = \cdots \\
&\quad = C(\bar{A}A^{j-1} + AKCA^{j-1})AK \\
&\quad = CA^{j+1}K.
\end{aligned} \tag{5.51}$$

Q.E.D.

Using CA^iK, $i = 1, \cdots, q$ a matrix M_s can be formed:

$$M_s = \begin{bmatrix} CAK \\ \vdots \\ CA^sK \end{bmatrix} \triangleq H_sK, \tag{5.52}$$

where $s \leq q$ and $H_s = \left[(CA)^T, \cdots, (CA^s)^T\right]^T$ is an observability-type matrix of the system. For an observable system, if s is large enough, matrix H_s will have a rank of n (full-column-ranked).

Replacing all $C\bar{A}^{i-1}AK$'s from (5.49) to (5.51) by their estimated values in set $\hat{S}_1$, $\{\widehat{C\bar{A}^{i-1}AK},\ i = 1, \cdots,\ q\}$, an estimated $\hat{M}_s$ of M_s can be obtained. Then the least-squares solution of (5.52) is an estimate of the steady state Kalman filter gain:

$$\hat{K} = H_s^{\dagger}\hat{M}_s, \tag{5.53}$$

where $H_s^{\dagger}$ is the pseudo-inverse of matrix H_s.

Since the estimated coefficients of the ARX model will converge to their true values eventually when the order q is large enough and the number of data goes to infinity, we can also expect the estimate of the steady state Kalman filter gain will converge to its optimal value.

5.5.2 Numerical Example

In this numerical example, the dynamical system, the noises, the input/output data and the adaptive transversal predictor (ATP) are all the same as that in Section 5.4.7. The tap parameters of the ATP are taken out periodically during operation for estimating the Kalman filter gain. In calculating the gain according to (5.51), the number of s is set to 20. The norms of the difference matrices between the optimal gain and the estimated gains obtained based on different number of data are shown in Fig. 5.16. The first estimate is made after the ATP has processed 500 data.

Similar to the numerical example in the last section, the state estimation of the first state, the residual comparison, and the residual auto-correlation function of using the gain obtained after processing 1,000 and 5,000 data are shown in

Figs. 5.17 and 5.18 respectively. The results in both cases are satisfactory. The estimated gains in this two cases are

$$\hat{K}_{1000} = \begin{bmatrix} 0.0565 & 0.0359 & 0.0513 & -0.0164 & 0.0046 & -0.0023 \\ 0.0966 & 0.0461 & -0.0193 & 0.0111 & -0.0250 & -0.0043 \end{bmatrix}^T,$$

$$\hat{K}_{5000} = \begin{bmatrix} 0.0579 & 0.0222 & 0.0445 & 0.0142 & 0.0126 & 0.0030 \\ 0.0667 & -0.0252 & -0.0094 & 0.0228 & -0.0235 & 0.0025 \end{bmatrix}^T.$$

This method is simpler than the equivalent prediction method introduced in the last section. It is also suitable for on-line application.

5.6 Inverse Filter Method

The inverse filter method utilizes the fact that the residual of an optimal Kalman filter is white. Through the innovation model the output is formulated as the sum of a deterministic part and a stochastic part. The deterministic part is driven by known input force, while the stochastic part is driven by the residual. The deterministic part can be subtracted out from the output, and the remaining signal can be modeled by a moving average (MA) model whose coefficients are in terms of the state space parameters and the optimal steady state Kalman filter gain. Hence the optimal Kalman filter gain can be calculated from the coefficients of the MA model. To identify the MA model, we identify a corresponding autoregressive (AR) model first, which is a filter whitening the remaining signal. The inverse of the AR model gives the MA model.

5.6.1 Obtaining Kalman Filter Gain from the Inverse Filter

From the Kalman filter formulations, an innovation model can be derived (see (5.7) and (5.8)), which are re-written here for convenience:

$$\hat{x}_{k+1}^- = A\hat{x}_k^- + Bu_k + AK_k\varepsilon_k \tag{5.7}$$

$$y_k = C\hat{x}_k^- + \varepsilon_k \tag{5.8}$$

For an optimal Kalman filter, the sequence $\{\varepsilon_k\}$ is white. In the steady state the filter gain becomes constant and thus the subscript can be deleted.

From the innovation model, the Kalman filter can be viewed as driven by the deterministic input u_k through B and by the stochastic input ε_k through AK. Hence, the filter state and output can be decomposed into two parts, one caused by the deterministic input and the other caused by the stochastic input. Accordingly, the innovation model can be divided into two models:

$$\hat{x}_{k+1,1}^- = A\hat{x}_{k,1}^- + Bu_k \tag{5.54}$$

$$y_{k,1} = C\hat{x}_{k,1}^- \tag{5.55}$$

and

$$\hat{x}_{k+1,2}^- = A\hat{x}_{k,2}^- AK_k\varepsilon_k \tag{5.56}$$

$$y_{k,2} = C\hat{x}_{k,2}^- + \varepsilon_k \tag{5.57}$$

where $x_k^- = x_{k,1}^- + x_{k,2}^-$ and $y_k = y_{k,1} + y_{k,2}$. Expanding (5.55) and (5.57) based on (5.54) and (5.56), respectively, one can derive

$$y_{k,1} = \sum_{i=1}^{k} CA^{i-1}Bu_{k-i}, \tag{5.58}$$

$$y_{k,2} = \sum_{i=1}^{k-1} CA^{i}K\varepsilon_{k-i} + \varepsilon_k. \tag{5.59}$$

Combining the above two equations, one obtains

$$y_k = \sum_{i=1}^{k} CA^{i-1}Bu_{k-i} + \sum_{i=1}^{k-1} CA^{i}K\varepsilon_{k-i} + \varepsilon_k. \tag{5.60}$$

Equation (5.60) clearly shows the two parts of which the output is composed. Since the state space parameters $[A,\ B,\ C]$ are known, one can subtract the deterministic component out from the output. That is, by defining

$$s_k = y_k - \sum_{i=1}^{k} CA^{i-1}Bu_{k-i},$$

(5.60) becomes

$$s_k = \sum_{i=1}^{k-1} CA^i K\varepsilon_{k-i} + \varepsilon_k = \sum_{i=0}^{k-1} C_i \varepsilon_{k-i}. \tag{5.61}$$

where $C_0 = I_p$, and $C_i = CA^iK$ for $i > 1$. The signal s_k is solely driven by sequence $\{\varepsilon_k\}$. For a stable system all the terms CA^iK, $i > q$, are negligibly small when q is sufficiently large; therefore, when k is large, the upper limit of the summation in the right hand side of (5.61) can be replaced by q. Equation (5.61) describes the signal s_k as linear transformation of a white sequence $\{\varepsilon_k\}$; therefore, it is called a Moving Average (MA) model. The matrices $C_1, \cdots, C_q$ are constants called the MA parameters. The term "moving average" arose because s_k can be regarded as a weighted average of $\varepsilon_k, \cdots, \varepsilon_{k-q}$. Note that the MA parameters are expressed in terms of the state space parameters A, C and steady state Kalman filter gain K. Knowing the MA parameters, one can compute the filter gain.

The problem of estimating the MA model in (5.61) is that the white sequence $\{\varepsilon_k\}$ is not readily available; therefore, the ordinary least-squares method frequently used to estimate the coefficients of linear equations cannot be used directly. However, we can estimate the MA model by estimating a corresponding autoregressive (AR) model first, and then seek the inverse of the AR model to find the MA model. To highlight this point, we take the z-transform of both sides of (5.61) to become

$$S = \sum_{i=0}^{q} C_i z^{-i} E = M(z^{-1})E, \tag{5.62}$$

where $M(z^{-1})$ is a polynomial matrix in z^{-1} (a matrix whose entries are polynomials in z^{-1}). Matrix $M(z^{-1})$ can be regarded as a filter which receives ε_k and its delayed versions as inputs and yields s_k as output. If we can find the inverse filter $N(z^{-1})$ of $M(z^{-1})$ such that $N(z^{-1})M(z^{-1}) = I_p$, by pre-multiplying Eq. (5.62) with $N(z^{-1})$ we have

$$N(z^{-1})S = E. \tag{5.63}$$

Matrix $N(z^{-1})$ usually is an infinite-ordered polynomial matrix in z^{-1}. In (5.63)

$N(z^{-1})$ can be viewed as a whitening filter which receives s_k and its delayed versions as inputs and yields white sequence $\{\varepsilon_k\}$ as output.

To obtain a whitening filter for the signal s_k, we can write an AutoRegressive model of s_k with order r in time domain as

$$\sum_{i=0}^{r} N_i s_{k-i} = \varepsilon_k, \tag{5.64}$$

where $N_0 = I_p$, and estimate the AR parameters $N_i, \cdots, N_r$.[16] Comparing (5.63) with (5.64) it can be seen that the infinite-ordered polynomial matrix $N(z^{-1})$ is approximated by a finite-ordered polynomial matrix $\sum_{i=0}^{r} N_i z^{-i}$. The parameter estimation of the AR model can be accomplished by using the ordinary least-squares method, which is well developed in the literature for scalar cases. For a matrix AR model, the extension of the method is straightforward.

After obtaining $N(z^{-1})$, we can inverse it to find $M(z^{-1})$. The operation of inversing a square polynomial matrix is similar to the inverse of an ordinary square matrix (i.e., a matrix with scalar entries) and the result is the adjoint matrix of the matrix divided by its own determinant of the matrix. In the operation multiplication of two polynomials can be calculated by convoluting the coefficient sequences of the two polynomial; division of two polynomials can be calculated by deconvoluting the coefficient sequence of the numerator polynomial over that of the denominator polynomial, expanding to as many terms as desired.

After obtaining the estimated MA model, collecting $q1$ coefficients one can form a matrix

$$M = \begin{bmatrix} CAK \\ CA^2K \\ \vdots \\ CA^{q1}K \end{bmatrix} = HK \tag{5.65}$$

where

$$H = \left[(CA)^T,\ (CA^2)^T,\ \cdots,\ (CA^{q1})^T\right]^T. \tag{5.66}$$

Note that H is an observability-like matrix, which is full column-ranked for an observable system and a sufficiently large $q1$. The least-squares solution of K is

$$\hat{K} = (H^T H)^{-1} H^T M = H^{\dagger} M \tag{5.67}$$

where $H^{\dagger}$ is the pseudo-inverse of H and $\hat{K}$ is the estimated optimal steady state Kalman filter gain.

5.6.2 Numerical Example

The same dynamical system and input/output data in the last two sections are used again in this example. The order of the AR model is set to 100, and the inverse of the polynomial matrix is also expanded to have 100 terms. For calculating the filter gain, the number $q1$ in (5.65) is also set to 100. After processing 5,000 data, the four elements of the estimated matrix sequence $\widehat{CA^iK}$, $(i = 1, \cdots, 100)$ are plotted along with their theoretical true values in Fig. 5.19. The variances of estimation error of the sequences against the number of data processed are shown in Fig. 5.20, where we can see the estimation is converging. Figure 5.21 shows the norms of the difference matrix between the estimated and the optimal gain against the number of data processed. This figure shows the convergence of the estimation of the filter gain. Similar to the last section, the gains estimated after processing 1,000 and 5,000 data are used in state estimation, and the results are shown in Figs. 5.22 and 5.23 respectively. The results show that the gain estimated using 1,000 data is fairly good already. The estimated gains in these two cases are

$$\hat{K}_{1000} = \begin{bmatrix} 0.0691 & 0.0303 & 0.0471 & 0.0184 & 0.0102 & -0.0051 \\ 0.0880 & 0.0447 & -0.0107 & 0.0143 & -0.0251 & 0.0054 \end{bmatrix}^T,$$

$$\hat{K}_{5000} = \begin{bmatrix} 0.0604 & 0.0279 & 0.0471 & 0.0146 & 0.0132 & 0.0055 \\ 0.0648 & 0.0366 & -0.0059 & 0.0143 & -0.0162 & -0.0011 \end{bmatrix}^T.$$

Figure 5.24 shows the reconstructed sequences $CA^i\hat{K}_{5000}(\ i = 1, \cdots, 100)$ using the gain $\hat{K}_{5000}$ and their theoretical true values. It indicates how well the MA model in (5.61) is estimated.

5.7 Estimation of Measurement and Process Noise Covariances from the Kalman Filter Gain

Though a well-estimated optimal steady state Kalman filter gain is sufficient for state estimation purpose, covariances of process and measurement noises might be of interest to those who want to know the stochastic properties of the system and the sensor. If we can derive this information from the estimated filter gain, it will be useful in practice. Indeed, we can easily derive the measurement noise covariance, but for process noise, the solution is not unique. However, one of the possible covariances has been derived.

5.7.1 Estimation of Measurement Noise Covariance

From Kalman filter formulations, the steady state filter gain is

$$K = P^{-}C^{T}(CP^{-}C^{T} + R)^{-1}, \tag{5.68}$$

where P^{-} is the a priori state error covariance. Without confusion, the superscript is omitted for simplicity. The covariance of the optimal residual sequence is

$$\begin{aligned} \Phi_k &\triangleq Cov[y_k - \hat{y}_k] \\ &= Cov[Ce_k^{-} + v_k] \\ &= CP_kC^{T} + R, \end{aligned} \tag{5.69}$$

where e_k^{-} is the a priori state error. In steady state the subscript k can be dropped. From (5.68) and (5.69), it is obvious that

$$PC^{T} = K\Phi; \tag{5.70}$$

therefore,

$$R = \Phi - CK\Phi = (I_p - CK)\Phi. \tag{5.71}$$

The optimal residual can be estimated by running a Kalman filter using the estimated filter gain. For a well-estimated gain, the covariance of the estimated residual $\hat{\Phi}$ calculated by sample averaging should be very close to the true one. Replacing all the true values on the right hand side of (6.71) by the estimated versions, an estimate of measurement noise covariance, $\hat{R}$, is obtained.

5.7.2 Estimation of Process Noise Covariance

From Kalman filter formulations in steady state, we have

$$P = AP^{+}A^{T} + Q \tag{5.72}$$

$$P^{+} = (I_n - KC)P, \tag{5.73}$$

where P^{+} is the a posteriori state error covariance. Combining the above two equations yields

$$P = A(I_n - KC)PA^{T} + Q = APA^{T} - AKCPA^{T} + Q. \tag{5.74}$$

According to (5.74), suppose K is known, by assigning any value to P, Q is uniquely determined, and vice versa. However, P and Q cannot be chosen arbitrarily. First, both covariance matrices should be symmetric and positive definite. In addition, since the covariance of the optimal residual is fixed for a fixed gain, the value PC^{T} is fixed from (5.70). Let

$$PC^{T} = X, \tag{5.75}$$

where X is a fixed matrix. Any symmetric positive definite matrix P' which satisfies (5.75) and produces a symmetric positive definite matrix Q' by (5.74) is a qualified candidate solution of P, and so is Q' a qualified candidate of Q. Using Q' thus found and R back to the Kalman filter will result in an identical steady state filter gain K.

Solving P and Q from (5.74) and (5.75) under the constraints of being both symmetric and positive definite is not a trivial task. Moreover, the solution in general is not unique, which can be proved by a simple numerical example.

5.7.3 Numerical Examples

For estimating measurement noise covariance, the example in Section 5.6.2 is used. Using the estimated optimal Kalman filter gain obtained after processing 5,000 data, the Kalman filter yields a residual having a covariance matrix (calculated by averaging 700 samples)

$$\hat{\Phi} = \begin{bmatrix} 0.0349 & 0.0031 \\ 0.0031 & 0.0306 \end{bmatrix}.$$

According to (5.71), the estimated measurement noise covariance is

$$\hat{R} = \begin{bmatrix} 0.0278 & 0.0010 \\ 0.0006 & 0.0279 \end{bmatrix}$$

while the theoretical true value is

$$R = \begin{bmatrix} 0.0279 & 0 \\ 0 & 0.0279 \end{bmatrix}.$$

We can see that the estimation is very accurate. If calculated from the optimal Kalman filter residual, the residual covariance and estimated measurement noise covariance are

$$\hat{\Phi}_{opt} = \begin{bmatrix} 0.0333 & 0.0024 \\ 0.0024 & 0.0312 \end{bmatrix}, \quad \hat{R}_{opt} = \begin{bmatrix} 0.0263 & 0.0000 \\ 0.0001 & 0.0278 \end{bmatrix};$$

therefore, even from the optimal Kalman filter, we cannot estimate the covariance exactly. This is caused by finite length of data processed.

To prove the solution of process noise covariance is not unique, we use a simple numerical example. The state space parameters of a single-input single-output dynamical system with just one mode are

$$A = \begin{bmatrix} 0.9801 & 0.1772 \\ -0.1772 & 0.9801 \end{bmatrix}, \quad B = \begin{bmatrix} 0.0049 \\ 0.0434 \end{bmatrix}, \quad C = [0.8163, \quad -0.0183].$$

The noise covariances are

$$Q = \begin{bmatrix} 0.0050 & 0.0000 \\ 0.0000 & 0.0050 \end{bmatrix}, \quad R = 0.0162.$$

The optimal steady state Kalman filter gain and a priori state error covariance under this situation are

$$K_{opt} = \begin{bmatrix} 0.5248 \\ 0.2594 \end{bmatrix}.$$

$$P = \begin{bmatrix} 0.0183 & 0.0098 \\ 0.0098 & 0.0369 \end{bmatrix}.$$

We can find two sets of P and Q (denoted by P_1, Q_1 and P_2, Q_2) which can satisfy

$$PC^T = \begin{bmatrix} 0.0148 \\ 0.0073 \end{bmatrix}$$

and (5.74), and be symmetric and positive definite:

$$P_1 = \begin{bmatrix} 0.0183 & 0.0104 \\ 0.0104 & 0.0648 \end{bmatrix}, \quad Q_1 = \begin{bmatrix} 0.0183 & 0.0104 \\ 0.0104 & 0.0648 \end{bmatrix};$$
$$P_2 = \begin{bmatrix} 0.0183 & 0.0091 \\ 0.0091 & 0.0094 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} 0.0061 & 0.0047 \\ 0.0047 & 0.0037 \end{bmatrix}.$$

Using Q_1 or Q_2 along with R in the Kalman filter, we can obtain the same gain K_{opt}. In other words, with the same measurement covariance, there is more than one process noise covariance which can result in the same Kalman filter gain. This example proves that, given an optimal Kalman filter gain, the solution of process noise covariance is not unique.

5.8 Concluding Remarks

This chapter presents three methods of estimating Kalman filter gain to solve the problem of state estimation under unknown noise covariances. The main points can be summarized as follows:

(1) A linear stochastic system can be represented by a state space model or a matrix polynomial model. An autoregressive with exogeneous input (ARX) model of a linear system whose parameters are expressed in terms of state space parameters and optimal steady state Kalman filter gain can be derived through the Kalman filter formulations. The parameters of the ARX model can be estimated using a recursive least-squares filter called adaptive transversal predictor (ATP), which requires no initial information about the system and noise. The estimation of the ARX parameters is p-consistent.

(2) The first method of estimating optimal Kalman filter gain uses the equivalence in output predictions based on a state space model and on the ARX model. The ARX model obtained by the ATP is used to generate multiple steps-ahead output predictions and from which the optimal steady state Kalman filter gain is calculated. The method is suitable for on-line adaptive applications as well as off-line batch-type analyses.

(3) The second method utilizes the relation between the state space model and the ARX model and calculates the optimal Kalman filter gain directly from the ARX parameters. The method also uses the ATP to estimate the ARX parameters. This method is simpler than the first method, yet more effective. It is also suitable for both on-line and off-line applications.

(4) The third method uses the concepts of decomposing output measurement into deterministic and stochastic parts, and estimates a moving average (MA) model through inversing a whitening filter. It subtracts the deterministic part out from the output first, and uses the ATP to whiten the remaining signal. The optimal filter gain is calculated from the inverse filter of the ATP. The estimation of filter gain is p-consistent. The method is also suitable for both on-line and off-line applications.

(5) The covariance of measurement noise can be obtained after the optimal

Kalman filter gain is estimated. However, the covariance of process noise cannot be uniquely determined.

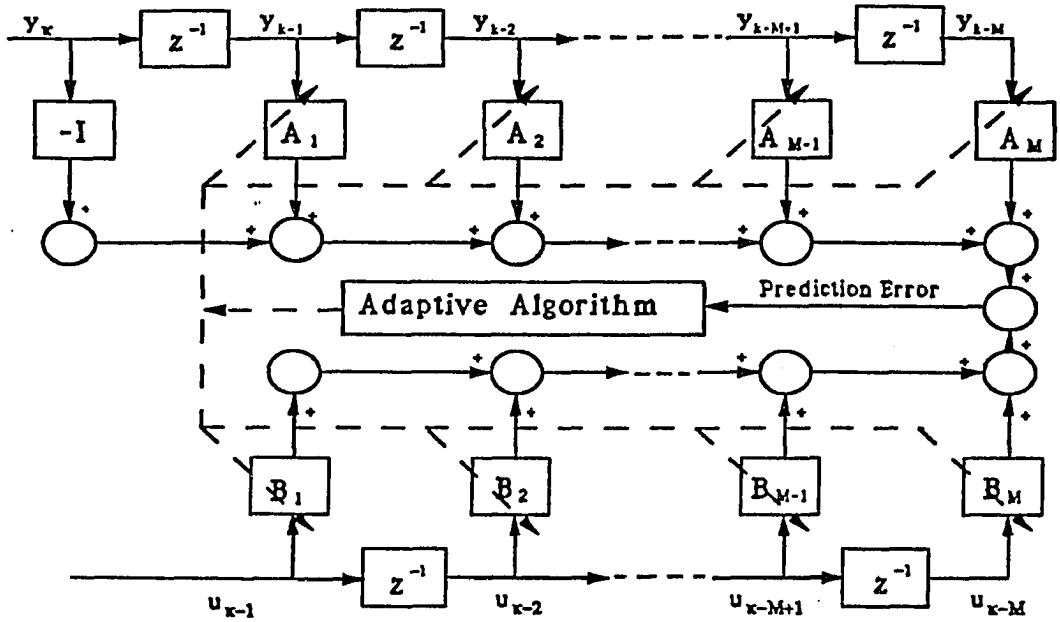
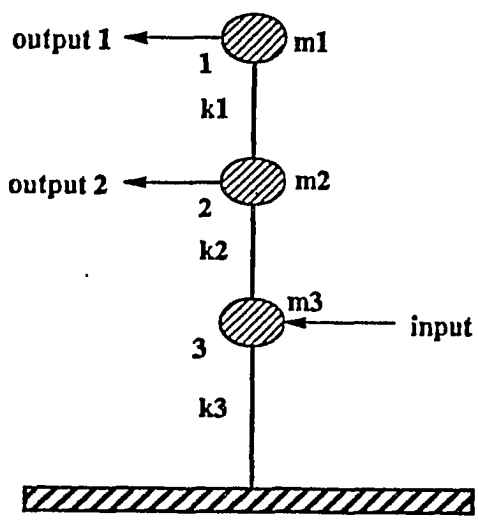


Fig. 5.1 Adaptive transversal predictor



m1	0.5
m2	1.0
m3	1.0
k1	10.0
k2	10.0
k3	10.0

Fig. 5.2 A simulated lumped-mass beam-like system with input force.

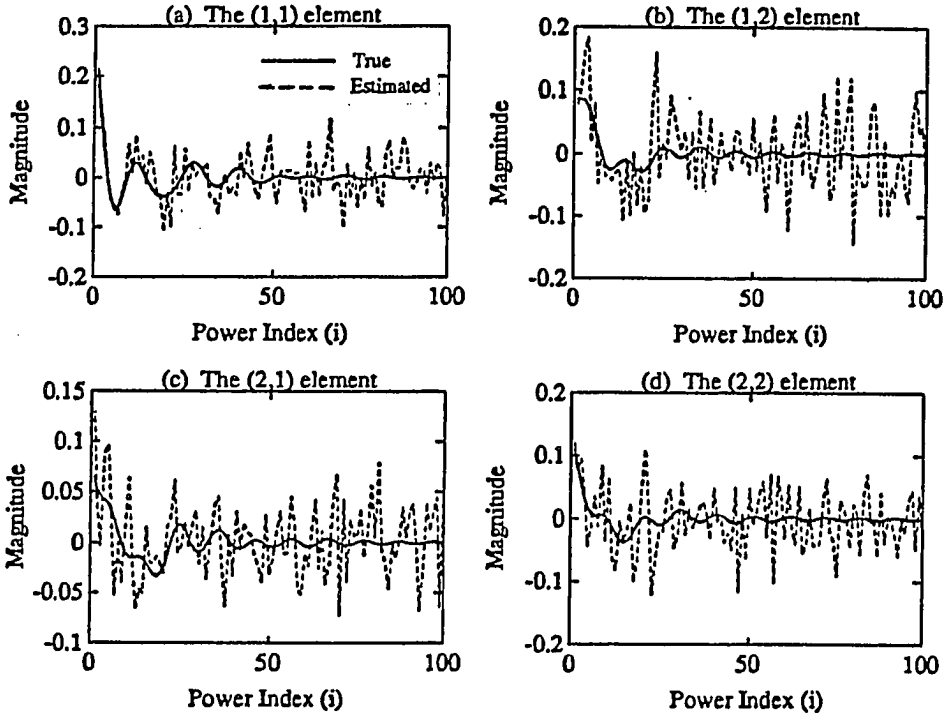


Fig. 5.3 Estimation of matrix sequence $C\bar{A}^{i-1}AK$ using 1,000 data.

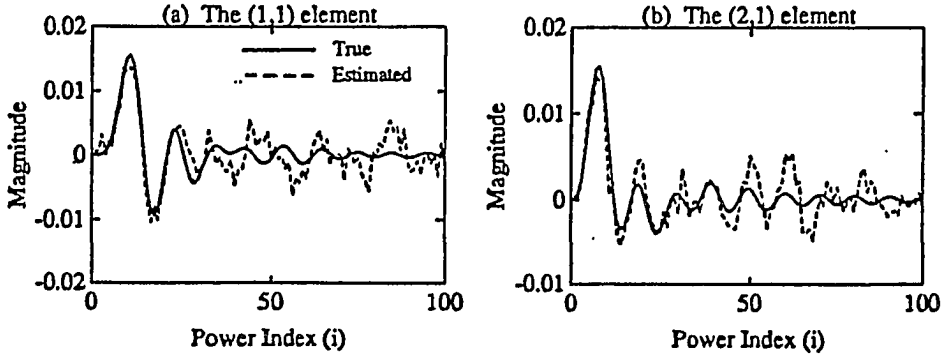


Fig. 5.4 Estimation of matrix sequence $C\bar{A}^{i-1}B$ using 1,000 data.

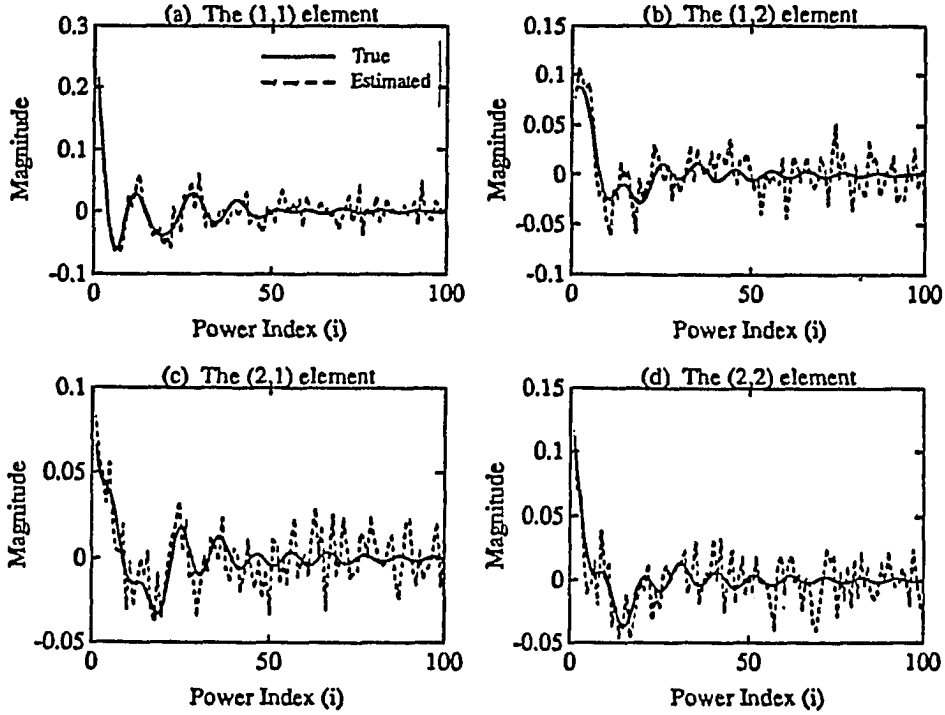


Fig. 5.5 Estimation of matrix sequence $C\bar{A}^{i-1}AK$ using 5,000 data.

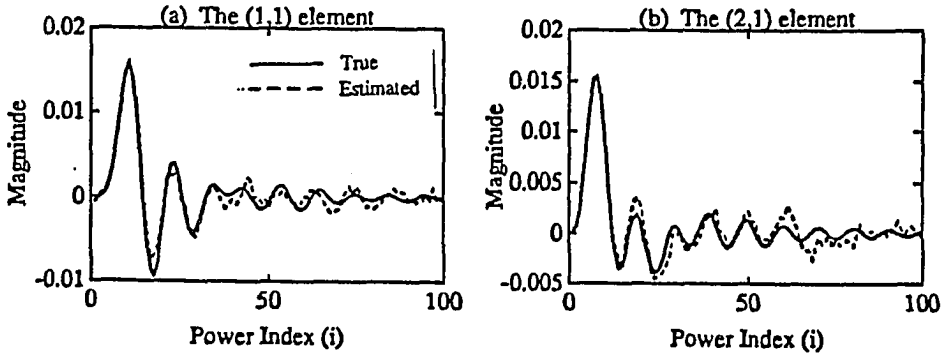


Fig. 5.6 Estimation of matrix sequence $C\bar{A}^{i-1}B$ using 5,000 data.

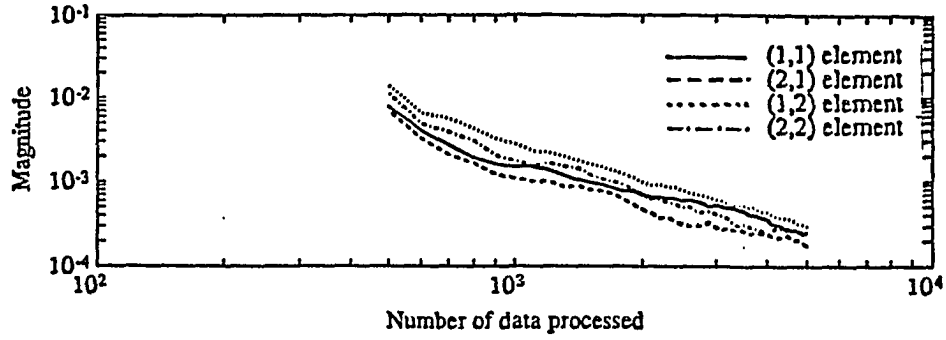


Fig. 5.7 Estimation error variance of $C\bar{A}^{i-1}AK$.

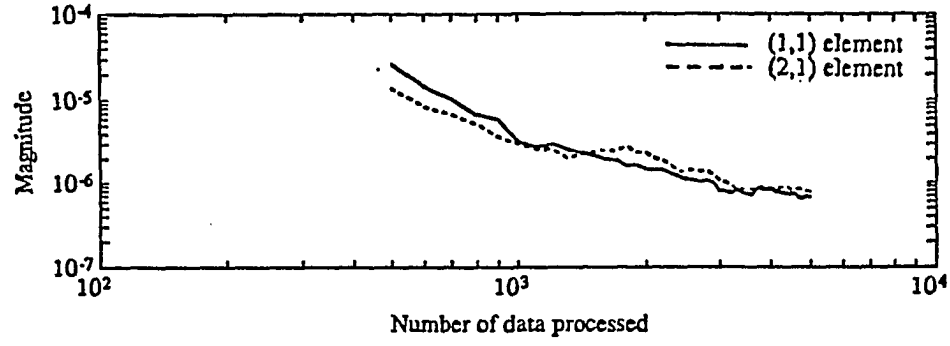


Fig. 5.8 Estimation error variance of $C\bar{A}^{i-1}B$.

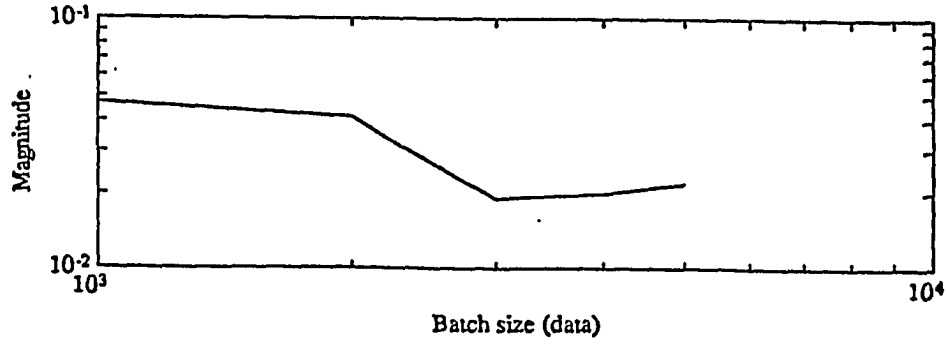


Fig. 5.9 Norm of estimated filter gains error as function of different batch sizes.

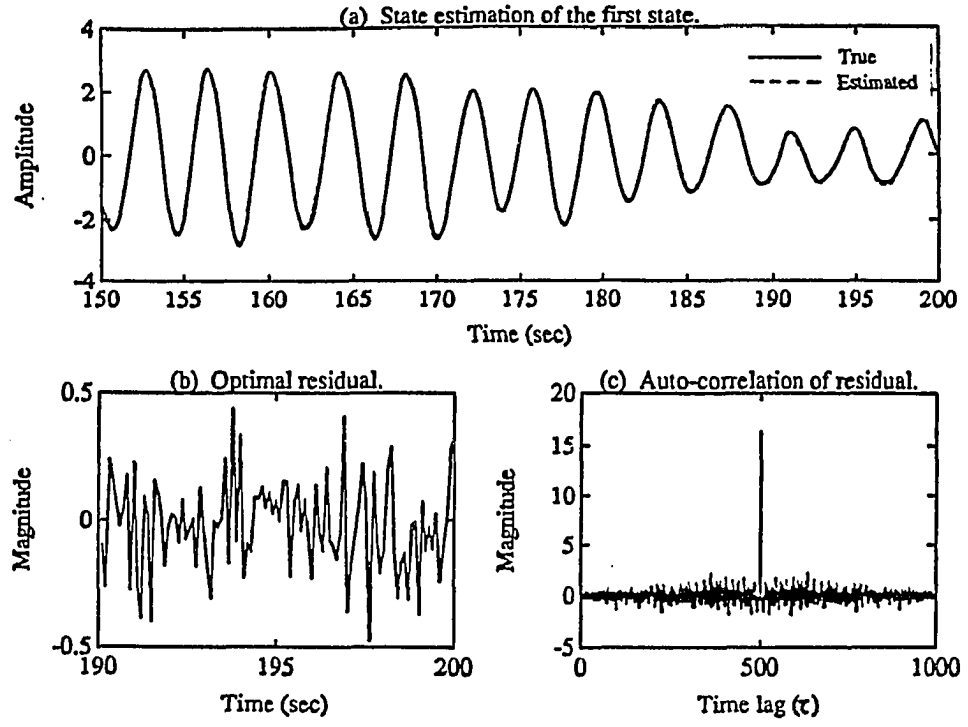


Fig. 5.10 Estimation results using the optimal Kalman filter.

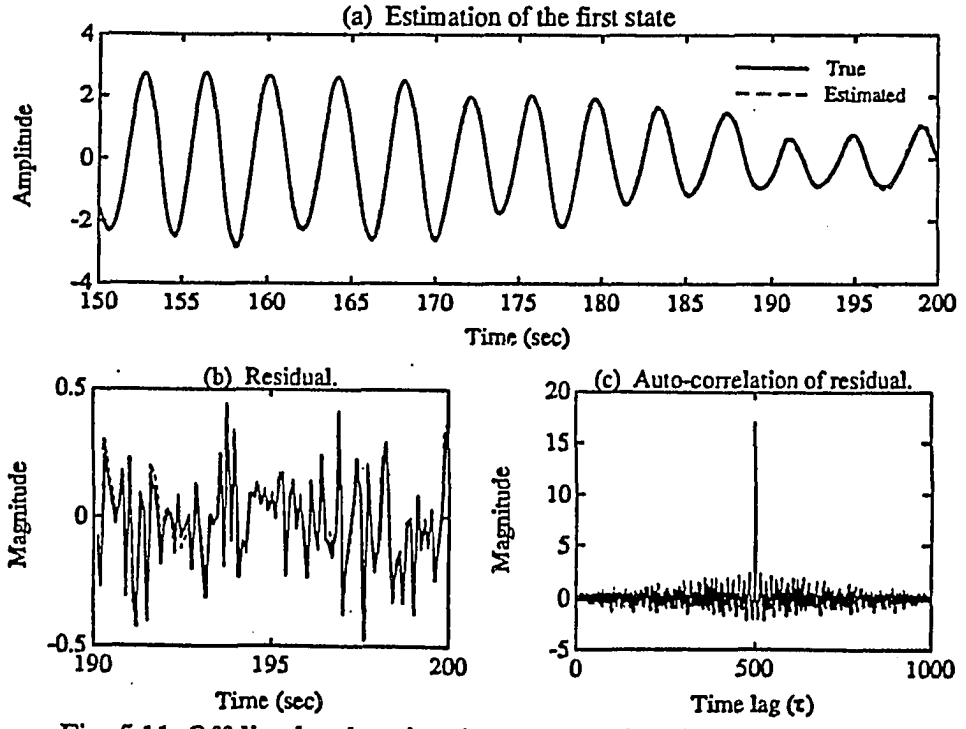


Fig. 5.11 Off-line batch estimation results using the equivalent prediction method with 1,000 data samples.

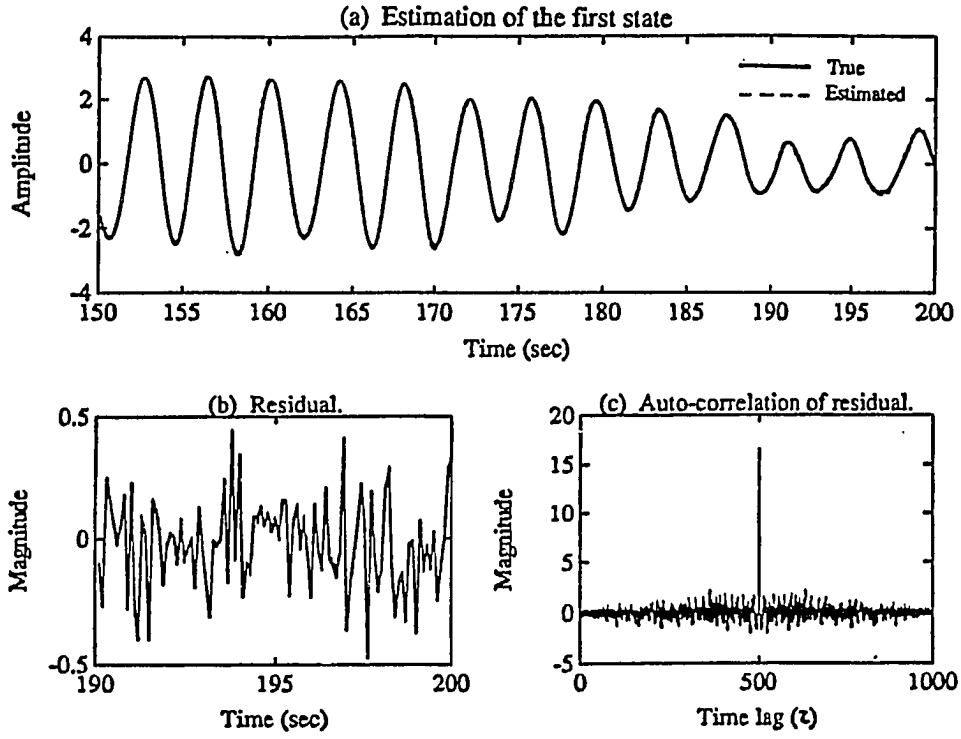


Fig. 5.12 Off-line batch estimation results using the equivalent prediction method with 5,000 data samples.

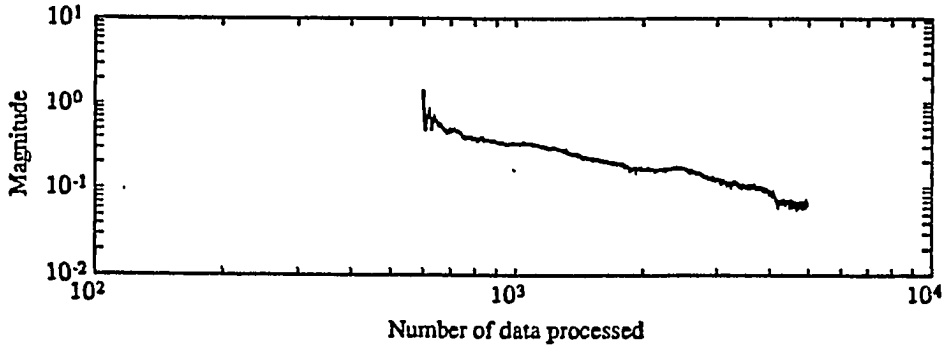


Fig. 5.13 Norm of the estimated filter gain error using an on-line recursive gain estimator as function of number of data samples processed.

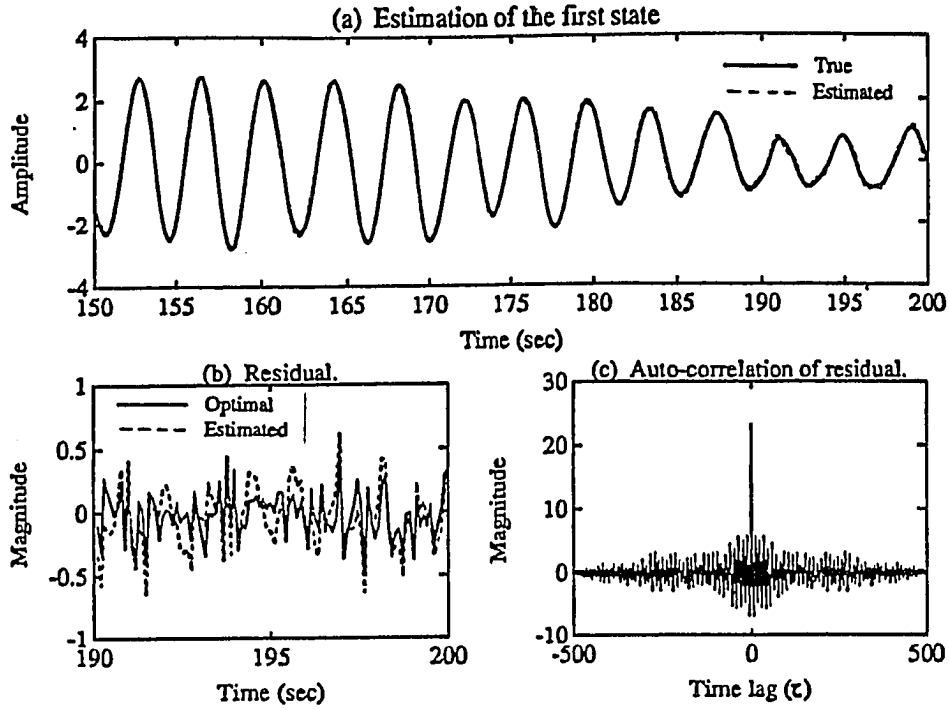


Fig. 5.14 On-line recursive estimation results using the equivalent prediction method with 2,000 data samples.

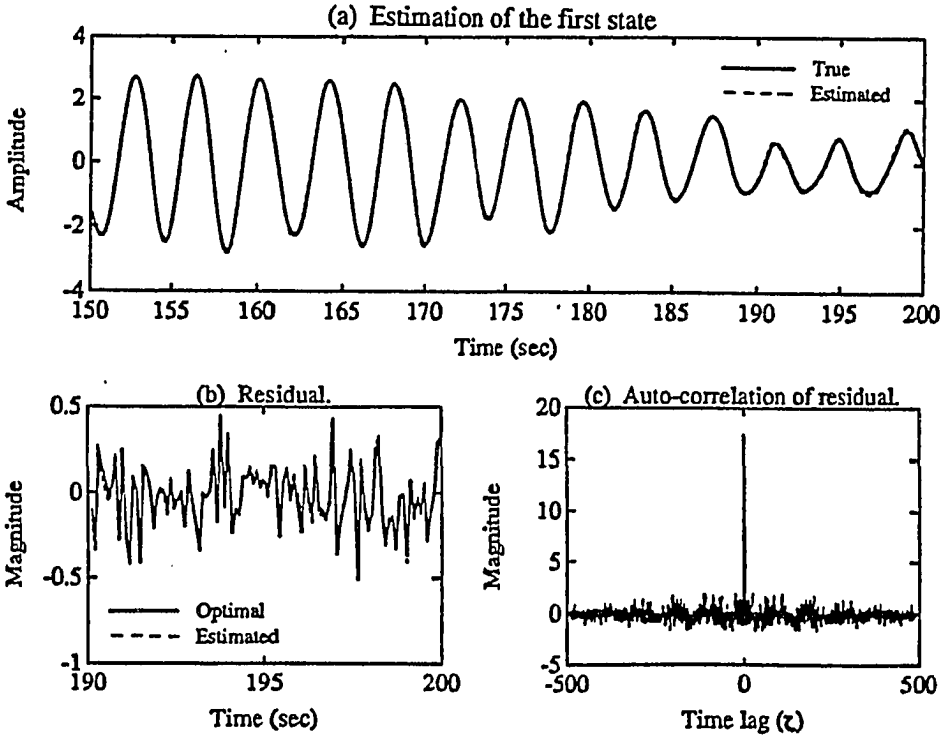


Fig. 5.15 On-line recursive estimation results using the equivalent prediction method with 5,000 data samples.

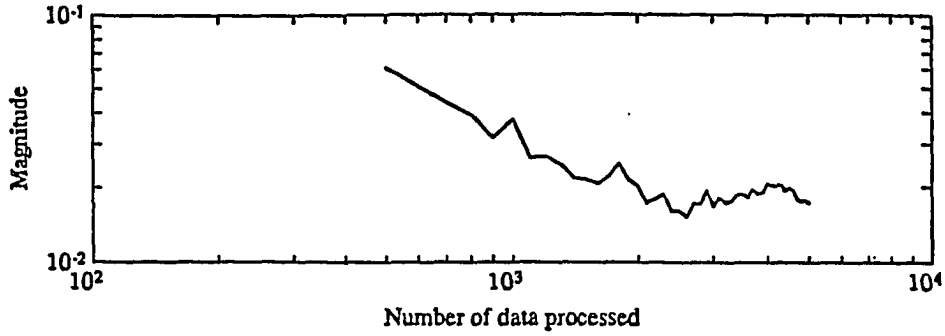


Fig. 5.16 Norm of the estimated filter gain error using the ARX coefficient method as function of number of data samples processed.

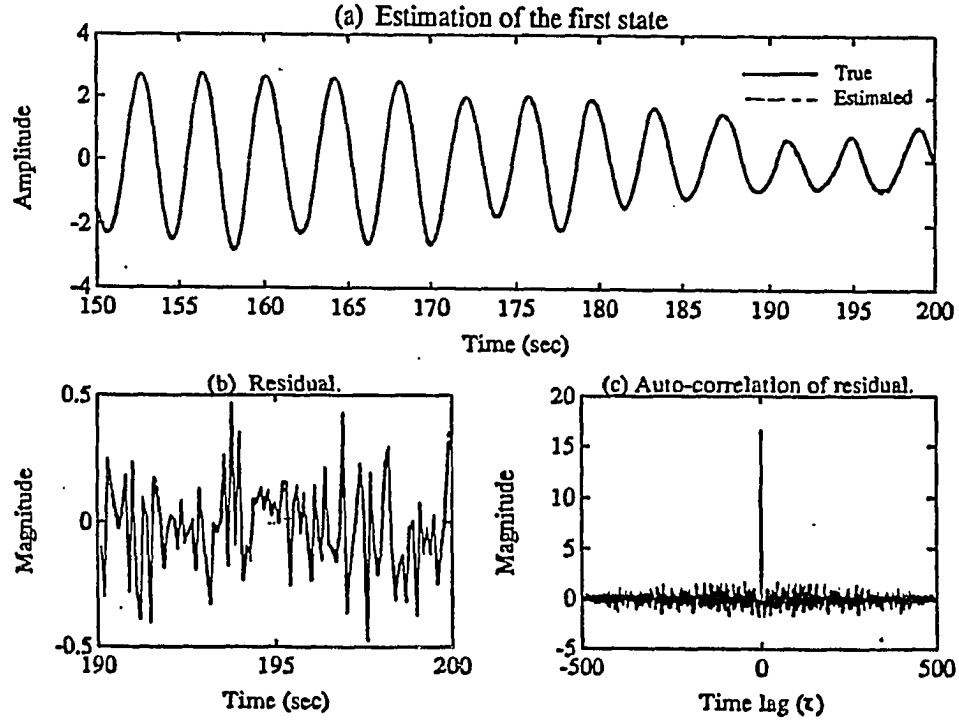


Fig. 5.17 Estimation results using the ARX coefficient method with 1,000 data samples.

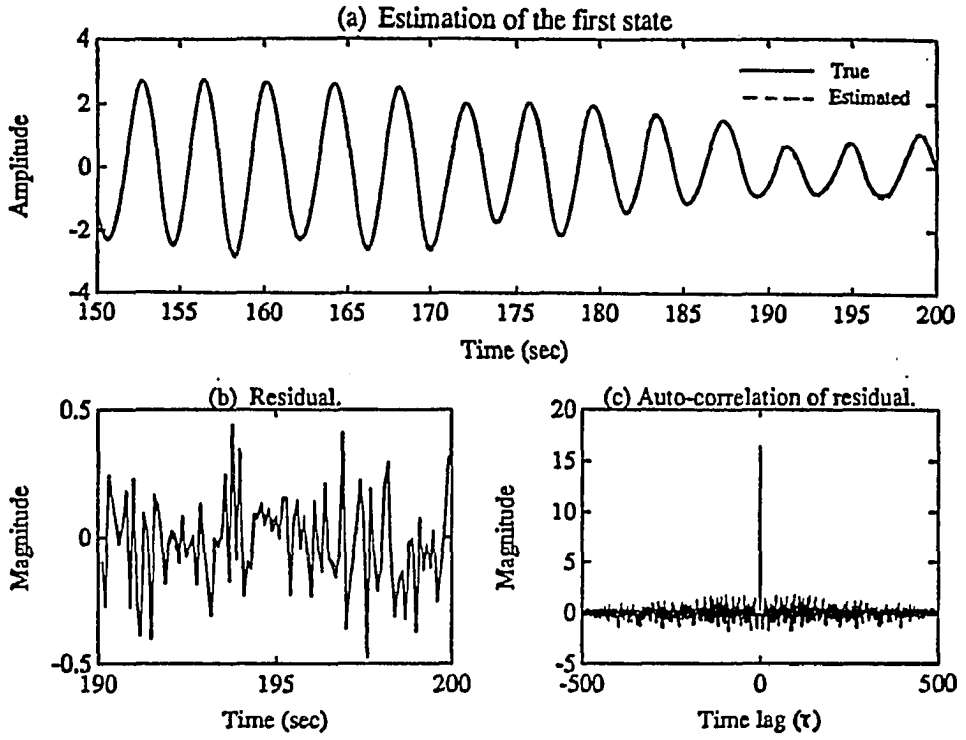


Fig. 5.18 Estimation results using the ARX coefficient method with 5,000 data samples.

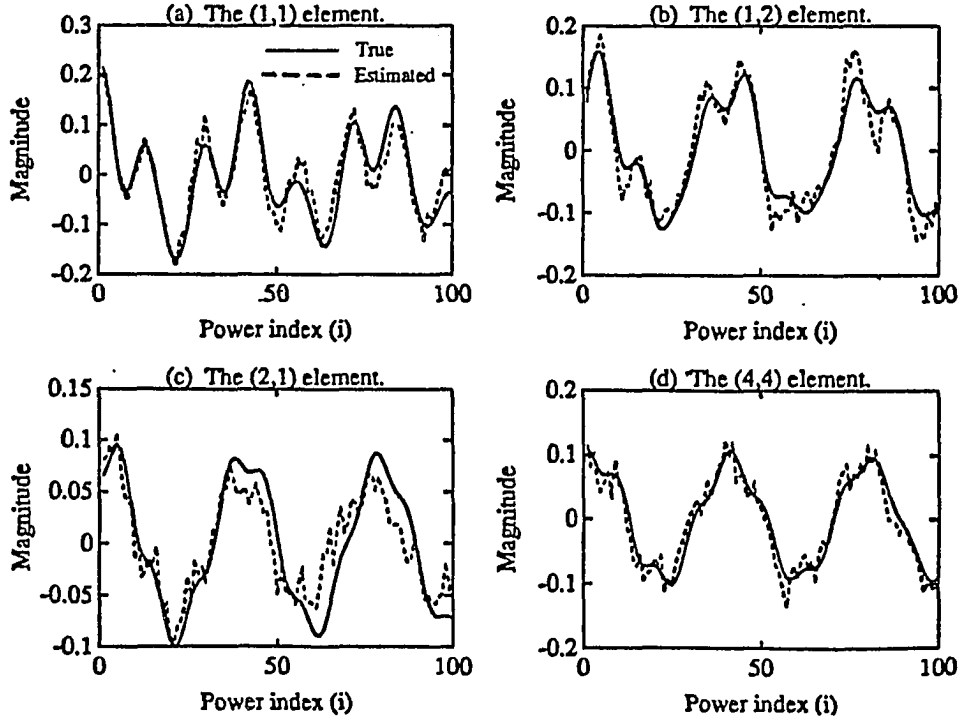


Fig. 5.19 Esimation of matrix sequence CA^iK using the inverse filter method with 5,000 data samples.

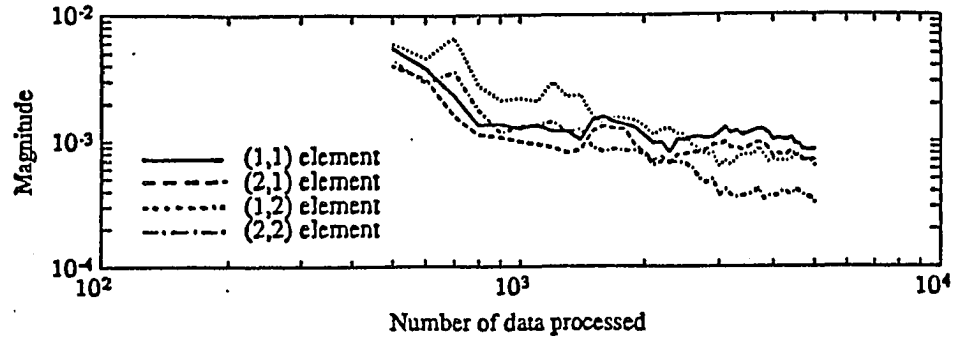


Fig. 5.20 Error variances of the elements of CA^iK using the inverse filter method.

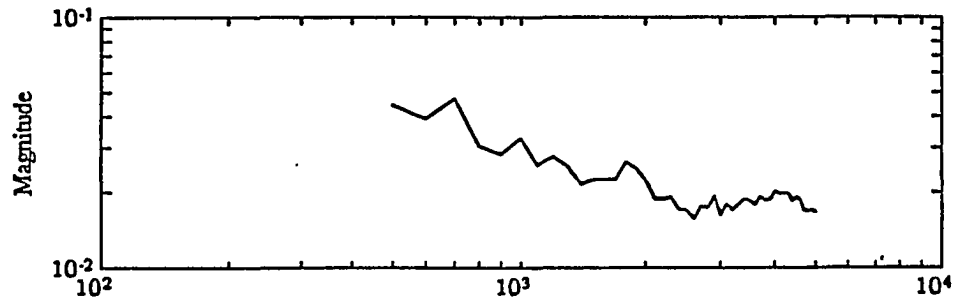


Fig. 5.21 Norm of the estimated filter gain error using the inverse filter method as function of number of data samples processed.

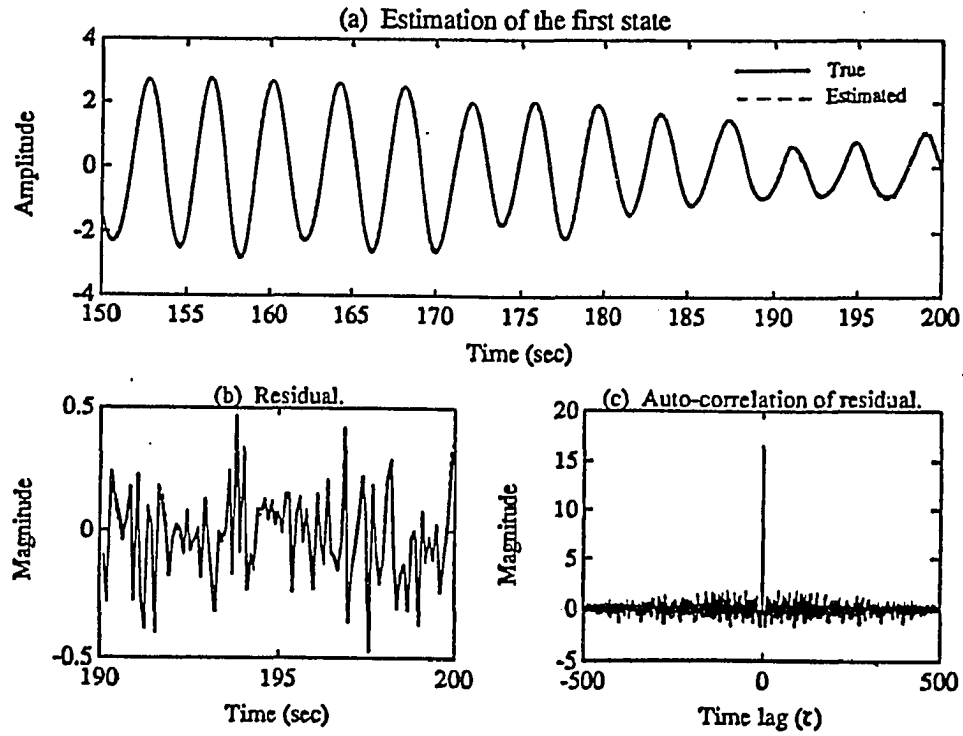


Fig. 5.22 Estimation results using the inverse filter method with 1,000 data samples.

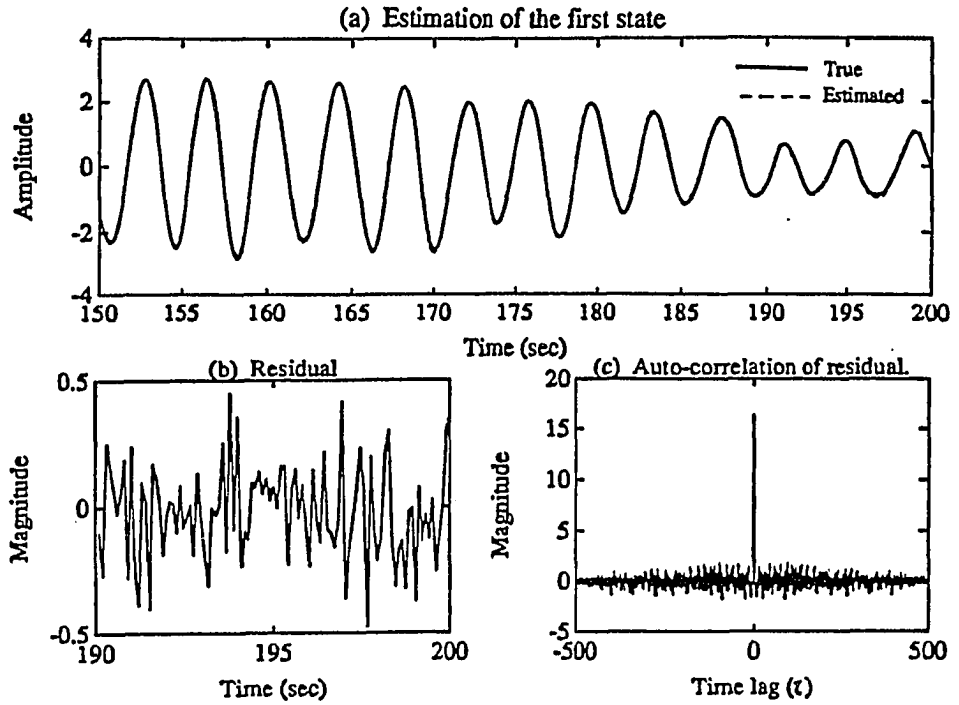


Fig. 5.23 Estimation results using the inverse filter method with 5,000 data samples.

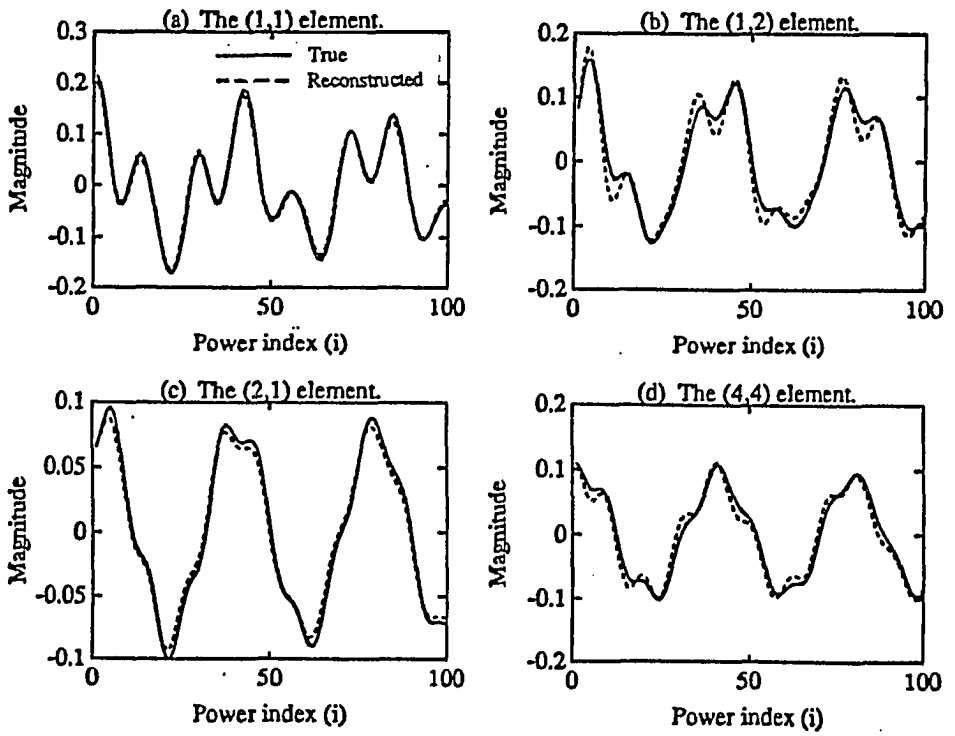


Fig. 5.24 Reconstruction of matrix sequence CA^iK using the inverse filter method with 5,000 data.

Chapter 6

LINEAR STATE ESTIMATION UNDER UNKNOWN SYSTEM MODEL AND NOISE COVARIANCES

6.1 Introduction

In Chapters 4 and 5 state estimation under unknown noise covariances was studied, where state space models of the linear systems are assumed known. Going one step further, in this chapter state estimation under unknown system models and unknown noise covariances is investigated. In other words, the problem is posed as follows: assuming the system is linear and the process and measurement noises are stationary, zero-mean and white with unknown covariances, given a set of input/output data, how can state estimation be conducted?

The usual way of approaching the problem is to divide it into two separate steps. First, system identification is conducted and a set of state space parameters $[A,\ B,\ C]$ is identified. Second, the identified system parameters are used in estimating either noise covariances or filter gain as described in the last chapter, and then the results are used in estimating states; or alternatively, after the first

step the least-squares approach introduced in Chapter 4 is used to estimate state directly, by-passing the estimation of noise covariances or filter gain.

Regardless of what approache one might take, the step of identifying a state space model is inevitable because without a state space model it is impossible to extract state information from input/output data. However, the above two steps can be combined together. In other words, the state space parameters and the corresponding optimal steady state Kalman filter gain can be identified simultaneously.[45] In fact, the steady state Kalman filter gain can be regarded as a parameter which characterizes the stochastic properties of the system. This point can be seen clearly from (5.7), (5.8) and (5.60), where the system can be viewed as driven by a deterministic force $\{u_k\}$ through input matrix B and by a stochastic force $\{\varepsilon_k\}$ through an equivalent input matrix AK. For deterministic systems, the state space parameters $[A,\ B,\ C]$ are sufficient for characterizing the system; however, for stochastic systems one needs parameters $[A,\ B,\ C,\ K]$ to characterize the system. Therefore, conducting stochastic system identification one should have a quadruplet $[A,\ B,\ C,\ K]$ as a result. Having this quadruplet, a state estimator can be readily constructed.

This chapter develops two methods for state estimation under unknown system model and noise covariances. The first one, called simultaneous method, identifies a state space model of the system and the corresponding optimal steady state Kalman filter gain simultaneously, using the adaptive transversal predictor (ATP). The second, called sequential method, identifies a state space model using the ATP with shorter order first, and then uses the methods developed in Chapter 5 to estimate the filter gain. A method for state estimation under uncertain system models was introduced in Ref. 15, where the state vector is augmented to include the uncertain system parameters. In this way the system parameters and state can be estimated at the same time. However, the formulations are complicated and rather difficult to implement, especially for high-order systems. Nonlinear

state estimation techniques, such as extended Kalman filter and its varieties, have to be used in this approach because the system model becomes nonlinear due to state augmentation. For nonlinear estimation, a system is usually linearized at each estimated state, which is very time-consuming especially for large order systems. Moreover, the convergence of the estimate is not guaranteed. The approach has inherent problems of bias and divergence. The methods introduced in this chapter can overcome these difficulties. On the other hand, least-squares lattice filter has been used in identifying structural dynamics;[34] however, deriving a state space model and a Kalman filter gain from a least-squares filter has never been addressed.

Section 6.2 briefly introduces the meaning, classification and applications of system identification. A system identification method, the eigensystem realization algorithm (ERA), is especially introduced because it is frequently used in the methods derived later in this chapter. Section 6.3 describes the method which can identify a state space model and the corresponding optimal steady state Kalman filter gain of a linear system simultaneously from input/output data. Modal transformation which transforms the identified parameters to modal coordinates is also discussed. In Section 6.4, another method of system identification is derived through the projection filter theory.

6.2 System Identification

System identification, sometimes called *modelling* or *time series analysis*, is important in many fields, e.g. economics, biology, physiology, ecology and process control. System identification deals with the problem of building mathematical models of dynamical systems based on observed data. In a sense, inferring models from observations and studying their properties is really what science is about;

therefore, system identification is actually part of basic scientific methodology.

Model building is an important method to understand the dynamic behavior of the system under study, and a mathematical model enables mathematical treatments of the system. Especially in technical application, system identification is the most important step for making use of control theory. Without an adequate model of the system to be controlled, the synthesis of a control algorithm is not possible.

The construction of a model from data involves three basic entities: the data, a set of candidate models, and a rule by which candidate models can be assessed using the data. The requirement of the data is that they should be "rich", in other words, they should contain sufficient information about the characteristics of the system under investigation.[21,22] In choosing a model type one should consider subsequent application, and the difficulty inherent to different model types. The rule for assessing a candidate model is usually to check how well the model can predict the output, given the input/output data of the system. Mathematically, for a least-squares method, the model parameters are chosen to minimize the sum of the squares of all the prediction errors.

There are several ways to classify system identification methods.[19,20,35,36] For example, from the type of the resulting model, there are Parametric models and Nonparametric models. Parametric models are such as algebraic equations, differential equations, a system of differential equations, and transfer functions. A nonparametric model is the response obtained directly or indirectly from an experimental analysis of a system, such as the recorded step response of a system is a nonparametric model. Other examples are the results from frequency domain analysis. From the signals used, there are continuous-time and discrete-time models. Continuous-time models are identified using continuous measurement, while discrete time models use sampled data. From the property of linearity of the systems under study, there are linear and non-linear system identifications. From

whether the disturbance from environment is considered, there are deterministic and stochastic system identifications. Besides, there are frequency-domain and time-domain system identifications. Frequency-domain methods use Fast Fourier Transform (FFT) technique to analyze the frequency property of the data and derive a model from it, while time-domain methods derive a model from input/output data directly, and are easier to formulate an adaptive method. For the purpose of subsequent applications in control design, a system identification method which can give a linear, parametric, stochastic and discrete-time model is preferable.

The Eigensystem Realization Algorithm (ERA) is a simple and powerful algorithm for identification of linear deterministic systems from impulse responses. It has been proved valuable for modal parameter identification from test data.[37-41] The algorithm uses the impulse responses (i.e., the Markov parameters for discrete systems) to form a large block data matrix which is referred to as the general Hankel matrix. Then the technique of singular value decomposition is used to decompose the Hankel matrix. The system order is determined by counting the number of singular values retained. The small singular values are attributed to noises and are truncated. The state space model can be computed from the decomposed matrices. Identifying a state space model from input/output data is also called "realization", because after having the state space model the system can be simulated, or "realized", using electric circuits. The realized model is not unique, or is unique only under equivalent transformation; but the Markov parameters are unique. For further details, readers are referred to Ref. 25.

In the methods developed later in this chapter, ERA is used as a standard tool for decomposing a Markov-parameter-type matrix sequence CA^iB, $i = 0, 1, \cdots$ into a triplet $[A', B', C']$, where A is a square matrix and the superscript $'$ denotes a version under equivalent transformation. Matrices A, B, C are not necessarily the state space parameters of a system. They can have or have not physical meanings.

6.3 The Simultaneous Method

This section introduces a method which can simultaneously identify a state space model and the corresponding optimal steady state Kalman filter gain of a linear system from input/output data. The simultaneous identification provides great advantages for the purpose of state estimation, for a state estimator can be readily constructed. The method is capable of updating the identified model and filter gain continually, which is desirable for on-line applications.

6.3.1 Coefficients of an ARX Model Estimated by ATP

In section 5.2 an autoregressive with exogeneous input (ARX) model whose coefficients are expressed in terms of state space parameters and the steady state Kalman filter gain is derived through the Kalman filter formulations. Section 5.3 introduced the adaptive transversal predictor (ATP), a recursive least-squares filter, to estimate the coefficients of the ARX model. The ATP requires no a priori information about the system and noises, except the assumptions that the system is linear and noises are stationary, zero-mean and white. The estimation is proved to be "p-consistent". After processing a number of N input/output data, the ATP gives a matrix $\hat{\Theta}_N$ containing the estimated coefficient matrices,

$$\hat{\Theta}_N^T = [\widehat{CAK}, \cdots, \widehat{C\bar{A}^{q-1}AK}, \widehat{CB}, \cdots, \widehat{C\bar{A}^{q-1}B}], \tag{6.1}$$

where q is the order of the ATP, and "$\wedge$" denotes estimated value.

Matrix $\hat{\Theta}_N$ can be divided into two sets of coefficient matrices, i.e.,

$$\hat{S}_1 = \{\widehat{CAK}, \widehat{C\bar{A}AK}, \cdots, \widehat{C\bar{A}^{q-1}AK}\} \tag{6.2}$$

$$\hat{S}_2 = \{\widehat{CB}, \widehat{C\bar{A}B}, \cdots, \widehat{C\bar{A}^{q-1}B}\} \tag{6.3}$$

where

$$\bar{A} = A(I_n - KC). \tag{6.4}$$

Note that sets $\hat{S}_1$ and $\hat{S}_2$ contain estimated coefficient matrices. Now denote the two sets which contain the corresponding true values by S_1 and S_2, respectively. The matrix sequences in S_1 and S_2 have the same form as Markov parameters. In fact, S_1 is the Markov parameters of the filter system described by (5.9) when driven by y_k only; while S_2 is the Markov parameters when driven by u_k only. Since the elements of the two sets are all represented in terms of the state space parameters and the Kalman filter gain, we can obtain these parameters from the two sets.

6.3.2 Identification of a State Space Model and Steady State Kalman Filter Gain Via ERA

Though there are several methods of realizing the state space parameters A, B, C and the steady state Kalman filter gain K from S_1 and S_2, they can be classified into two different major approaches. One approach processes on S_1 and S_2 individually and sequentially, while the other combines them together to form another sequence. They are introduced as follows.

6.3.2.1 Method 1

In the first method, the Eigensystem Realization Algorithm (ERA) is applied to decompose S_1 or S_2 directly. Since S_1 and S_2 are Markov-parameter-type matrix sequences, the ERA can be used to decompose them into triplets containing equivalent system matrix, output matrix and input matrix. However, since the Markov parameters belong to the filter system rather than the original system, further treatment on the the identified triplets is needed to obtain the parameters of the original system.

The ERA is used to decompose S_1 into a triplet $[\bar{A}',\ (AK)',\ C']$, which is related to $[\bar{A},\ AK,\ C]$ via some unknown equivalent transformation P. Specifically,

$$\bar{A}' = P^{-1}\bar{A}P \tag{6.5}$$

$$(AK)' = P^{-1}(AK) \tag{6.6}$$

$$C' = CP. \tag{6.7}$$

Note that with respect to the equivalent transformation P, a realized output matrix C' is readily available as given in (6.6). The corresponding realized system matrix $A' = P^{-1}AP$ can be obtained from (6.4) to (6.7) as

$$A' = \bar{A}' + (AK)'C'. \tag{6.8}$$

To obtain the realized input matrix under the same transformation, namely $B' = P^{-1}B$, first note that

$$\begin{aligned} CB &= CPP^{-1}B = C'B' \\ C\bar{A}^iB &= CPP^{-1}\bar{A}^iPP^{-1}B = C'(\bar{A}^i)'B' \\ &= C'(\bar{A}')^iB', \end{aligned}$$

hence, from the set S_2 one can write

$$\begin{bmatrix} CB \\ C\bar{A}B \\ \vdots \\ C\bar{A}^{q-1}B \end{bmatrix} = \begin{bmatrix} C'B' \\ C'\bar{A}'B' \\ \vdots \\ C'(\bar{A}')^{q-1}B' \end{bmatrix} = \begin{bmatrix} C' \\ C'\bar{A}' \\ \vdots \\ C'(\bar{A}')^{q-1} \end{bmatrix} B'$$

$$\triangleq HB', \tag{6.9}$$

where V denotes the observability-like matrix. Thus B' can then be calculated as

$$B' = H^{\dagger}\begin{bmatrix} CB \\ C\bar{A}B \\ \vdots \\ C\bar{A}^{q-1}B \end{bmatrix} \tag{6.10}$$

where $H^{\dagger}$ denotes the pseudo-inverse of H.

An alternative way to obtain matrix B is to apply ERA to the second set S_2. However, the realized matrix B is not necessary in the same coordinates as the realized system matrix A' and output matrix C' obtained using the first set S_1. A coordinate transformation is needed to bring the two sets of identified matrices to the same coordinates. The coordinate transformation will be discussed in the next section.

Under different coordinates for the state variable, the steady state Kalman gain should be transformed accordingly. If the state is transformed by a non-singular matrix P, i.e. $\hat{x}_k^- = P(\hat{x}_k^-)'$, then (5.7) becomes

$$\begin{aligned}(\hat{x}_{k+1}^-)' &= P^{-1}AP(\hat{x}_k^-)' + P^{-1}Bu_k + P^{-1}APP^{-1}K\varepsilon_k \\ &= A'(\hat{x}_k^-)' + B'u_k + A'K'\varepsilon_k\end{aligned}$$

where $K' = P^{-1}K$. Note that K is transformed exactly the same way as B.

Since $(AK)' = A'K'$, K' can be calculated as

$$K' = (A')^{-1}(AK)'. \tag{6.11}$$

Having these relations between parameters, we can substitute the true values, S_1 and S_2, by their estimated versions, $\hat{S}_1$ and $\hat{S}_2$, to obtain a quadruplet $[\hat{A}', \hat{B}', \hat{C}', \hat{K}']$.

6.3.2.2 Method 2

In the second method, the two sets S_1 and S_2 are combined to yield the Markov parameters of the original system. After decomposing the combined sequence, the state space parameters of the original system are derived.

The first element in S_2, CB, is also the first element of the Markov parameters. Starting from this point, we can calculate the Markov parameters recursively according to the following equation

$$CA^kB = C\bar{A}^kB + \sum_{i=1}^{k} C\bar{A}^{k-i}AKCA^{i-1}B, \tag{6.12}$$

where $k = 1, \cdots, q-1$. Denoting the j-th element in S_1 and S_2 by $S_{1,j}$ and $S_{2,j}$, respectively, (6.12) can be written as

$$CA^kB = S_{2,k+1} + \sum_{i=1}^{k} S_{1,k+1-i} CA^{i-1}B, \tag{6.13}$$

which clearly shows that all the information needed to compute CA^kB can be obtained from S_1, S_2 and the previous calculation of CA^lB, $l = k-1,\ k-2, \cdots, 1$.

Proof of (6.12) :

$$\begin{aligned}
&C\bar{A}^kB + \sum_{i=1}^{k} C\bar{A}^{k-i}AKCA^{i-1}B \\
&\quad = C\bar{A}^kB + C\bar{A}^{k-1}AKCB + C\bar{A}^{k-2}AKCAB + \cdots + CAKCA^{k-1}B \\
&\quad = C(\bar{A}^k + \bar{A}^{k-1}AKC + \bar{A}^{k-2}AKCA + \cdots + AKCA^{k-1})B \\
&\quad = C\left[\bar{A}^{k-1}(\bar{A} + AKC) + \bar{A}^{k-2}AKCA + \cdots + AKCA^{k-1}\right]B \\
&\quad = C\left[\bar{A}^{k-2}(\bar{A} + AKC)A + \cdots + AKCA^{k-1}\right]B \\
&\quad = \cdots \\
&\quad = C\left[\bar{A}(\bar{A} + AKC)A^{k-2} + AKCA^{k-1}\right]B \\
&\quad = C\left(\bar{A} + AKC)A^{k-1} + AKCA^{k-1}\right]B \\
&\quad = CA^kB.
\end{aligned} \tag{6.14}$$

Q.E.D.

After identifying a number of the Markov parameters, ERA can be used to realize a state space model. To obtain the corresponding filter gain, the method presented in Section 5.5 can be used.

6.3.3 Modal State Estimation

For a given realized quadruplet $[A',\ B',\ C',\ K']$, a state estimator for the original system can be constructed using this quadruplet. Because the sets $\hat{S}_1$

and $\hat{S}_2$ can be adaptively improved when more data are processed, ERA must be used from time to time to update the quadruplet. However, the quadruplet may not necessarily belong to the same coordinates each time . Therefore, in order to compare the realized quadruplet and estimated states successively , they should be transformed to the same coordinates. The modal coordinate, in which the system matrix A is block diagonal and the matrices C and B are normalized in some sense, is an appropriate choice.

A model in modal form and modal state estimation are of interest among researchers in structural dynamics,[25,42,43] because modal state bears good physical meanings. For a state space model in modal form, system matrix A gives the information of modal frequencies and dampings; matrix C gives mode shapes at the locations of output sensors.[25] If the mode shapes have been normalized, the magnitude of each state indicates the amount of mechanical energy allocated in that mode. Modal frequencies, dampings and mode shapes are called modal parameters, which are the main interests of conducting modal analysis. However, one should remember that, for control purpose, a state space model in modal form is not necessary. All that a controller design requires is a state space model of the system, no matter what coordinate it might refer to. Therefore, the modal transformation discussed in the followings is of interest in modal analysis.

Consider the case when all eigenvalues of the system matrix A are distinct. Then the normalized eigenvector matrix $V = [v_1, v_2, \cdots, v_n]$ can be used to diagonalize A, i.e. $V^{-1}AV = \Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$ where λ_i $(i = 1, 2, \cdots, n)$ denotes the i-th eigenvalues of A. Since a scalar multiple of an eigenvector is still an eigenvector, any $T = VK_c$ can also diagonalize A, where K_c is any (non-singular) diagonal matrix. Furthermore, any matrix T that diagonalizes A can be written as $T = VK_c$ for some K_c. A three-step procedure is presented in the following to transform a realized triplet [A, B, C] to its modal coordinates.

Step 1: Diagonalization

Let the system matrix A be diagonalized by a matrix T such that

$$A^* = T^{-1}AT = \Lambda. \tag{6.15}$$

Correspondingly, B and C are transformed according to

$$B^* = T^{-1}B, \qquad C^* = CT. \tag{6.16}$$

Given two sets of realized $[A,\ B,\ C]$ that are equivalent, i.e., they are related by some equivalent transformation, the above transformation will uniquely recover $A^* = \Lambda$ but not necessarily B^* and C^* because of the freedom in T. In order to uniquely recover B^* and C^*, they must be normalized in a certain way. The following describes such a normalization procedure.

Step 2: Normalization

The normalization is defined so that each column of the normalized matrix has unit length, and the first element of the column is a positive real number. Noting that the elements could be complex numbers, this procedure can be accomplished by the following steps. First, find a constant diagonal matrix M, $M = diag[m_1, m_2, \cdots, m_n]$, such that $C^*M = [c_1m_1, \cdots, c_nm_n]$ and $m_i^2 c_i^T c_i = 1$, where c_i denotes the i-th column of C^*. Next, find a constant diagonal matrix R, $R = diag[r_1, \cdots, r_n]$, where r_i is a pure complex number or ± 1 which rotates a complex number without changing its length, such that $C_n = C^*MR = diag[c_1m_1r_1, \cdots, c_nm_nr_n]$ and the first element of vector $c_im_ir_i$ is a positive real number. Then C^* is normalized such that

$$C_n = C^*MR. \tag{6.17}$$

Accordingly, B^* is transformed to B_n so that

$$B_n = R^{-1}M^{-1}B^*. \tag{6.18}$$

Step 3: Modal Transformation

For vibratory systems such as flexible space structures, the eigenvalues often appear as complex conjugate pairs. After the above transformation the states are complex numbers, and so are the elements of B_n and C_n. Another transformation can be used to further transform $[\Lambda,\ B_n,\ C_n]$ to their modal forms, $[A_m,\ B_m,\ C_m]$, where A_m is block diagonal and all the matrices are real. The realized steady state Kalman filter gain K is transformed in the same way as B.

It can be shown that the above transformation procedure will recover a unique set of $[A_m,\ B_m,\ C_m,\ K_m]$ from any equivalent sets of $[A,\ B,\ C,\ K]$ (see Appendix for proof). With the quadruplet in modal form, the modal space state estimation can be carried out by using a constant gain Kalman filter. The modal state estimator is described by the following two equations,

$$\hat{x}_k^- = A_m \hat{x}_{k-1}^- + B_m u_{k-1}, \tag{6.19}$$

$$\hat{x}_k^+ = \hat{x}_k^- + K_m(y_k - C_m \hat{x}_k^-), \tag{6.20}$$

where $\hat{x}_k^+$ is the estimated state. Note that A_m is a block diagonal matrix which makes this state estimator easier for implementation. To this end, the integrated system identification and state estimation scheme has been accomplished.

6.3.4 Numerical Examples

To study the numerical properties of the method, two sample problems are presented. In the first example the system is a beam-like structure (the same one used in the previous chapters) characterized by non-repeated low frequencies and low dampings. In the second example the system is a simulated Mini-Mast structure, characterized by repeated low frequencies and low dampings.

Example 1: A Beam-Like Structure

In the first example the lumped-mass beam-like simulated system with three masses (see Fig. 5.2) is considered again. The settings of the system and the environment are re-stated here for convenience. The modal frequencies and damping factors are shown in Table 6.1. The variance of the random excitation force, σ_u^2, is set to 40. The standard deviation of the process noise w_k is about 23% of that of the input influence Bu_k, and the standard deviation of the measurement noise v_k is about 10% of that of the output measurement y_k. The sampling frequency is 10 Hz. The filter order of the adaptive transversal filter is set to 100. The identified modal frequencies and damping factors after processing 1,000 to 5,000 data (with an increment of 1,000 data), using both method 1 and method 2, respectively, are listed in Table 6.1 and compared with the true values. From the results, we can see that method 2 is more effective in identifying modal parameters. The true quadruplet $[A_m,\ B_m,\ C_m,\ K_m]$ and its identified versions using the second method with 1,000 and 5,000 data are shown here for comparison. Note that the Kalman filter gains are obtained using the method described in Section 5.5 after obtaining the state space modal models, and 20 terms are used for pseudoinverse ($s = 20$, see (5.52)).

$$A_m = diag\left\{\begin{bmatrix} 0.9856 & 0.1628 \\ -0.1628 & 0.9856 \end{bmatrix}\begin{bmatrix} 0.8976 & 0.4305 \\ -0.4305 & 0.8976 \end{bmatrix}\begin{bmatrix} 0.8127 & 0.5690 \\ -0.5690 & 0.8127 \end{bmatrix}\right\}$$

$$B_m = [0.0011 \quad 0.0134 \quad -0.0016 \quad -0.0072 \quad 0.0011 \quad 0.0034]^T$$

$$C_m = \begin{bmatrix} 1.5119 & 0.0000 & 2.0000 & 0.0000 & 1.5119 & 0.0000 \\ 1.3093 & 0.0000 & 0.0000 & 0.0000 & -1.3093 & 0.0000 \end{bmatrix}$$

$$K_m = \begin{bmatrix} 0.0604 & 0.0279 & 0.0471 & 0.0146 & 0.0132 & 0.0055 \\ 0.0648 & 0.0366 & -0.0059 & 0.0143 & -0.0162 & -0.0011 \end{bmatrix}^T$$

where m denotes modal form and A_m is a block diagonal matrix.

$$\hat{A}_{m,1000} = diag\left\{\begin{bmatrix} 0.9851 & 0.1630 \\ -0.1630 & 0.9851 \end{bmatrix}\begin{bmatrix} 0.8976 & 0.4313 \\ -0.4313 & 0.8976 \end{bmatrix}\begin{bmatrix} 0.8115 & 0.5686 \\ -0.5686 & 0.8115 \end{bmatrix}\right\}$$

$$\hat{B}_{m,1000} = [0.0013 \quad 0.0137 \quad -0.0016 \quad -0.0071 \quad 0.0010 \quad 0.0033]^T$$

$$\hat{C}_{m,1000} = \begin{bmatrix} 1.5131 & 0.0000 & 1.9986 & 0.0000 & 1.5219 & 0.0000 \\ 1.3078 & -0.0173 & -0.0746 & -0.0004 & -1.2976 & 0.0050 \end{bmatrix}$$

$$\hat{K}_{m,1000} = \begin{bmatrix} 0.0567 & 0.0358 & 0.0520 & 0.0165 & 0.0050 & -0.0039 \\ 0.0972 & 0.0460 & -0.0196 & 0.0109 & -0.0254 & -0.0049 \end{bmatrix}^T$$

and

$$\hat{A}_{m,5000} = diag\left\{ \begin{bmatrix} 0.9852 & 0.1631 \\ -0.1631 & 0.9852 \end{bmatrix} \begin{bmatrix} 0.8966 & 0.4306 \\ -0.4306 & 0.8966 \end{bmatrix} \begin{bmatrix} 0.8121 & 0.5689 \\ -0.5689 & 0.8121 \end{bmatrix} \right\}$$

$$\hat{B}_{m,5000} = [0.0011 \quad 0.0134 \quad -0.0016 \quad -0.0074 \quad 0.0010 \quad 0.0034]^T$$

$$\hat{C}_{m,5000} = \begin{bmatrix} 1.5113 & 0.0000 & 2.0000 & 0.0000 & 1.4999 & 0.0000 \\ 1.3099 & -0.0030 & -0.0107 & -0.0056 & -1.3227 & -0.0244 \end{bmatrix}$$

$$\hat{K}_{m,5000} = \begin{bmatrix} 0.0578 & 0.0224 & 0.0449 & 0.0144 & 0.0124 & 0.0029 \\ 0.0669 & 0.0254 & -0.0093 & 0.0230 & -0.0235 & 0.0021 \end{bmatrix}^T .$$

Note the true and the identified quadruplet are in fairly good agreement. Theoretically, as more data are processed these parameters converge to their true values if the filter order is sufficiently large.

Using the identified quadruplet, a modal state estimator is constructed as shown in (6.19) and (6.20). The estimated modal states are then compared with their true values. Figures 6.1 to 6.3 show the state estimations and the corresponding residuals of three cases: after 500, 1,000, and 5,000 data processed (using the second method). In state estimation figures, the solid lines represent the true values and the dashed lines represent the estimated modal state using the identified quadruplet. Three modal states are shown in the figure and demonstrate that as more data are used the state estimations are improved. In the residual plots, the solid lines represent the optimal Kalman filter residuals and the dashed lines represent the estimated ones, and in Figs. 6.2(d) and 6.3(d), they are in good agreement. The auto-correlation functions of the residual show that the state estimation using the model and filter gain obtained with 1,000 data is satisfactory already.

Example 2: Mini-Mast

In the second example a simulated Mini-Mast model is considered. Mini-Mast is a 20-meter-long generic space truss built in NASA Langley Research Center for control experiments of flexible structures[44] (see Fig. 6.4). It is deployed vertically and cantilevered from its base on a rigid foundation. A five-mode (ten states) model of Mini-Mast which includes two repeated frequencies is used to generate simulated data. The modal frequencies and damping factors of the model are listed in Table 6.2. The first two modes are closely-spaced representing the first bending mode in x and y axes (see Fig. 6.4) with the same mode shapes in different phases, and similar are the last two closely-spaced modes, which are the second bending mode. The third mode represents the first Mini-Mast torsion mode. The simulated system has two inputs (torque wheels) and two outputs (Kaman sensors). The inputs are random forces with unit strength. The process noise is set at approximately 23 % of the input influence and the measurement noise about 10 % of the output, both in the standard deviation ratio. The sampling time is 33.3 Hz (0.03 sec). The order of the adaptive transversal filter is set at 100. Method 2 is used in identifying a system model, and the corresponding Kalman filter gain is obtained according to Section 5.5. Again the number of terms used in pseudoinverse is 20. The identified modal frequencies and damping factors with the corresponding number of data processed are listed in Table 6.2. The results are fairly accurate even in the presence of repeated modes.

When a system has repeated eigenvalues, the mode shapes of the repeated frequency are not unique even though they are normalized as described in the above section. For a specific repeated frequency any linear combination of the identified mode shapes may be used as a mode shape. This does not impose any problem if the system identification and state estimation are conducted for control purpose because a state feedback controller design only requires a set of state space model and filter gain, regardless of what coordinate the model might refer to. However,

the Markov parameters of the system are unique because they are independent of the state coordinate. Hence, the quality of the identified model can be evaluated by comparing the reconstructed Markov parameters ($\hat{C}\hat{A}^{i-1}\hat{B}$) with that of the original system ($CA^{i-1}B$). Figures 6.5 and 6.6 show the comparison of the four elements of the Markov parameter matrices after 1,000 and 5,000 data processed, respectively. They are in good agreement even in the 1,000 data case. Note that the scales in Figs. 6.5(d) and 6.6(d) are different from others. Another alternative for evaluating the quality of the identified model is to compare the singular values of the corresponding transfer function matrix over an interested frequency range with that of the original system. Figures 6.7 and 6.8 show such comparison, in which the identified model is obtained with 1,000 and 5,000 data, respectively. Because there are two inputs and two outputs, the transfer function matrix has two singular values for each frequency; therefore, there are two figures for each case. A good agreement is obtained with some deviation caused by noise-induced errors in the identified frequencies, dampings and mode shapes. The improvement in 5,000 data case is apparent. To examine the quality of the identified filter gain, the elements of the reconstructed filter Markov parameter matrices ($\hat{C}\bar{A}^{i-1}\hat{A}\hat{K}$, $i = 1, \cdots, 100$) is compared to the true one, which is shown in Figs. 6.9 and 6.10 for the above two cases, respectively. They are in good agreement but not totally converged yet. Figures 6.11 and 6.12 show the estimated first, fifth and ninth states of the above two cases. They represent the modal states of the first bending, the first torsion and the second bending modes, respectively. Because the coordinates of the identified models are different from that used in simulation, the estimated states do not have true version for comparison. Note that the corresponding modal states of these two cases are also different, because they are in different modal coordinates. However, because the modal models have been normalized, the amplitudes of the modal states are the same. From this we can see the energy allocated in each mode. Though we can not compare the estimated state with its true value, we can compare the estimated output (calculated from the estimated state) with the true output, and the corresponding residual. Figures 6.13 and 6.14

show such comparison for the above two cases. The true and estimated output coincide together in both cases, however, the auto-correlation functions of residual show the estimation of the second case is better. From Figs. 6.5 to 6.14, it is fair to say that the identified model including an optimal filter in this example is very good after 5,000 data processed.

6.4 The Sequential Method

The simultaneous system identification and state estimation method developed in the last section is effective, yet there is a drawback. The method relies on identifying an ARX model of the system using the adaptive transversal predictor (ATP). The ARX model is derived from Kalman filter formulations. If the system is flexible (low-damped) and the process noise is not strong (thus the Kalman filter gain K is small), according to the derivation in (5.11) the ARX model should have a large order q to make the truncation insignificant. However, as the order of ARX model increases, the computational load in identifying the model using the ATP also increases rapidly. The rate of increase is about in the order of q^2.[16] This is not desirable for practical application.

One question arises: if the order of ATP is assigned much shorter than needed to make the truncation insignificant, what values will the parameters converge to and what is the relation between these parameters and the system state space model. Fortunately, through the projection filter derived in Chapter 3, we found the answer to the question and derived a system identification method similar to that in the last section. This method can derive a state space model of the system using an ARX model of smaller order, but it cannot derive the corresponding Kalman filter gain at the same time. To estimate the Kalman filter gain, methods introduced in Chapter 5 can be used after obtaining a system model.

6.4.1 A Relation Between the Projection Filter and Matrix Polynomial Models of a Linear System

The projection filter was developed in Chapter 3; however, because the input term Bu_k is added to the model in this section, the formulations are a little different. Therefore, they are briefly listed here for clarity. To derive the relation between the projection filter and matrix polynomial models of a linear system, we start from a simple case and gradually move to more general ones.

Consider a noise-free dynamic system, which can be represented by a model:

$$x_{k+1} = Ax_k + Bu_k \tag{6.21}$$

$$y_k = Cx_k. \tag{6.22}$$

From Eqs. (6.21) and (6.22) it is easy to follow that

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+2} \\ y_{k-q+1} \end{bmatrix} = \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+2} \\ CA^{-q+1} \end{bmatrix} x_k - \begin{bmatrix} 0 & \cdots & 0 & 0 \\ CA^{-1}B & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ CA^{-q+2}B & \cdots & CA^{-1}B & 0 \\ CA^{-q+1}B & \cdots & CA^{-2}B & CA^{-1}B \end{bmatrix} \begin{bmatrix} u_{k-1} \\ \vdots \\ u_{k-q+2} \\ u_{k-q+1} \end{bmatrix}, \tag{6.23}$$

or in short,

$$Y_{q,k} = H_q x_k - G_q U_k, \tag{6.24}$$

or in a normal form

$$H_q x_k = Y_{q,k} + G_q U_k, \tag{6.25}$$

where q denotes the number of data stacked up to form the equation, and the meanings of the matrices are self-evident. For a sufficiently large q which can

makes H_q full-column-ranked, the unique least-squares solution of x_k is

$$\hat{x}_k = F_q(Y_{q,k} + G_q U_k), \tag{6.26}$$

where

$$F_q = (H_q^T H_q)^{-1} H_q^T \tag{6.27}$$

is the pseudo-inverse of H_q and also the projection filter in this case. The solution $\hat{x}_k$ is identical with the true value x_k for this noise-free case. The number q can be any integer bigger than a integer q_{min}, which is the minimum number required to make H_q full-column-ranked.

To write a matrix polynomial model of the system which expresses the current output as a linear transformation of finite previous input/output data, one can use (6.21), (6.22) and (6.26):

$$\begin{aligned} y_k &= Cx_k \\ &= CAx_{k-1} + CBu_{k-1} \\ &= CA[F_q(Y_{q,k-1} + G_q U_{k-1})] + CBu_{k-1} \\ &= \sum_{i=1}^{q} CAF_{qi} y_{k-i} + CBu_{k-1} + \sum_{i=2}^{q} CAF_q G_{q(i-1)} u_{k-i}, \end{aligned} \tag{6.28}$$

where F_{qi} and G_{qi} are the i-th partitions of F_q and G_q, respectively, defined as

$$F_q = [F_{q1},\ F_{q2},\ \cdots, F_{qq}] \tag{6.29}$$

$$G_q = [G_{q1},\ G_{q2},\ \cdots, G_{q(q-1)}], \tag{6.30}$$

matrix F_{qi} has a dimension of $n \times p$, and G_{qi} of $(p \times q) \times m$.

Next, consider a system without process noise but with additive, white, Gaussian, and zero-mean measurement noise which is not correlated with the state variable. Then we can derive a matrix equation

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+2} \\ y_{k-q+1} \end{bmatrix} = \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+2} \\ CA^{-q+1} \end{bmatrix} x_k$$

$$- \begin{bmatrix} 0 & \cdots & 0 & 0 \\ CA^{-1}B & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ CA^{-q+2}B & \cdots & CA^{-1}B & 0 \\ CA^{-q+1}B & \cdots & CA^{-2}B & CA^{-1}B \end{bmatrix} \begin{bmatrix} u_{k-1} \\ \vdots \\ u_{k-q+2} \\ u_{k-q+1} \end{bmatrix} + \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-q+2} \\ v_{k-q+1} \end{bmatrix}, \quad (6.31)$$

or in short,

$$Y_{q,k} = H_q x_k - G_q U_k + V_{q,k}, \quad (6.32)$$

or in a normal form

$$H_q x_k = Y_{q,k} + G_q U_k - V_{q,k}. \quad (6.33)$$

Here the unknown variable x_k is a deterministic variable. By the theory of parameter estimation for deterministic parameters from a linear equation with independent white noise, one can write the optimal estimate of x_k as

$$\hat{x}_k = F_q(Y_{q,k} + G_q U_k), \quad (6.34)$$

where

$$F_q = (H_q^T \bar{R}^{-1} H_q)^{-1} H_q^T \bar{R}^{-1} \quad (6.35)$$

is a weighted pseudo-inverse of H_q and the projection filter for this case; $\bar{R} = R \otimes I_q$, $\otimes$ is the Kronecker product, R the covariance of the measurement noise. Note the optimality is defined by the minimum variance of state estimation error.

To derive a model in matrix polynomial format using the projection filter, we can form a one-step-ahead output prediction using the last estimated state,

$$\hat{y}_k = CA\hat{x}_{k-1} + CBu_{k-1} \quad (6.36)$$

and define

$$y_k = \hat{y}_k + \eta_k \quad (6.37)$$

where η_k is prediction error. Therefore,

$$\begin{aligned} y_k &= CA\hat{x}_{k-1} + CBu_{k-1} + \eta_k \\ &= CA\left[F_q(Y_{q,k-1} + G_q U_{k-1})\right] + CBu_{k-1} + \eta_k \\ &= \sum_{i=1}^{q} CAF_{qi} y_{k-i} + CBu_{k-1} + \sum_{i=2}^{q} CAF_q G_{q(i-1)} u_{k-i} + \eta_k, \end{aligned} \quad (6.38)$$

where F_{qi} and G_{qi} are defined in the same way as in Eqs. (6.29) and (6.30), but F_q is defined by Eq. (6.35) in this case.

Next, consider a general case of a system with both process and measurement noises. By writing the previous output in terms of the current state, one can derive

$$\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-q+2} \\ y_{k-q+1} \end{bmatrix} = \begin{bmatrix} C \\ CA^{-1} \\ \vdots \\ CA^{-q+2} \\ CA^{-q+1} \end{bmatrix} x_k$$
$$- \begin{bmatrix} 0 & \cdots & 0 & 0 \\ CA^{-1}B & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ CA^{-q+2}B & \cdots & CA^{-1}B & 0 \\ CA^{-q+1}B & \cdots & CA^{-2}B & CA^{-1}B \end{bmatrix} \begin{bmatrix} u_{k-1} \\ \vdots \\ u_{k-q+2} \\ u_{k-q+1} \end{bmatrix}$$
$$- \begin{bmatrix} 0 & \cdots & 0 & 0 \\ CA^{-1} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ CA^{-q+2} & \cdots & CA^{-1} & 0 \\ CA^{-q+1} & \cdots & CA^{-2} & CA^{-1} \end{bmatrix} \begin{bmatrix} w_{k-1} \\ \vdots \\ w_{k-q+2} \\ w_{k-q+1} \end{bmatrix} + \begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-q+2} \\ v_{k-q+1} \end{bmatrix}, \quad (6.39)$$

or in short,

$$Y_{q,k} = H_q x_k - G_q U_k - M_q W_{q,k} + V_{q,k}, \quad (6.40)$$

where the process noise vector is denoted by $W_{q,k}$ and its coefficient matrix by M_q. Equation (6.40) can be further simplified to

$$H_q x_k = Y'_{q,k} + \xi_{q,k} \quad (6.41)$$

where

$$Y'_{q,k} = Y_{q,k} + G_q U_k, \quad (6.42)$$

$$\xi_{q,k} = -M_q W_{q,k} + V_{q,k}. \quad (6.43)$$

The unknown variable x_k is a random variable in this case. The overall noise vector $\xi_{q,k}$ is Gaussian and zero-mean because $W_{q,k}$ and $V_{q,k}$ are Gaussian and zero-mean. It is also correlated with the unknown variable x_k because $W_{q,k}$ is

correlated with x_k. By denoting the covariance between x_k and $\xi_{q,k}$ by $\Omega_{x\xi}$, the optimal estimate of x_k can be obtained by

$$\hat{x}_k = \bar{x}_k + F_q(Y'_{q,k} - \bar{Y}'_{q,k}), \tag{6.44}$$

where the overbar – denotes the expectation value, and

$$F_q = (\Omega_x H_q^T + \Omega_{x\xi})(H_q \Omega_x H_q^T + H_q \Omega_{x\xi} + \Omega_{x\xi}^T H_q^T + R_\xi)^{-1} \tag{6.45}$$

is the projection filter in this case, where R_ξ denotes the covariance of $\xi_{q,k}$. The optimality is defined by the minimum variance of state estimation error.

Similarly, to write a matrix polynomial model using the relation provided by the projection filter, we can use one-step-ahead output prediction of the current output as (6.36) and have

$$\begin{aligned}
\hat{y}_k &= CA\hat{x}_{k-1} + CBu_{k-1} \\
&= CA\left[\bar{x}_{k-1} + F_q(Y'_{q,k-1} - \bar{Y}'_{q,k-1})\right] + CBu_{k-1} \\
&= CAF_qY_{q,k-1} + CBu_{k-1} + CAF_qG_qU_{q,k-1} + CA(I_n - F_qH_q)\bar{x}_{k-1} \\
&= \sum_{i=1}^{q} CAF_{qi}y_{k-i} + CBu_{k-1} + \sum_{i=2}^{q} CAF_qG_{q(i-1)}u_{k-i} + CAL\bar{x}_{k-1},
\end{aligned} \tag{6.46}$$

where

$$L = I_n - F_qH_q,$$

F_{qi} and G_{qi} are again defined in the same way as in (6.29) and (6.30) but F_q is defined by (6.45) instead.

Equation (6.46) represents the best prediction of y_k one can make using q previous input/output data. If the prediction is made once and for all, namely, no prediction of previous state is made, the best value assigned to $\bar{x}_k$ is zero. However, if previous state estimation has been made out, the best choice for $\bar{x}_k$ is the a priori Kalman filter estimate. Note that for the Kalman filter

$$\hat{x}_{k-1}^- = A\hat{x}_{k-2}^- + AK(y_{k-2} - C\hat{x}_{k-2}^-) + Bu_{k-2}$$

$$= \cdots$$
$$= \sum_{i=1}^{q-1} \bar{A}^{i-1} AK y_{k-1-i} + \sum_{i=1}^{q-1} \bar{A}^{i-1} B u_{k-1-i} + \bar{A}^q \hat{x}_{k-q}, \tag{6.47}$$

where

$$\bar{A} = A(I_n - KC).$$

Based on the argument above, we can replace $\bar{x}_{k-1}$ in (6.46) by (6.47) and obtain

$$\begin{aligned} y_k &= \hat{y}_k + \eta_k \\ &= CAF_{q1} y_{k-1} + \sum_{i=2}^{q} CA\left(F_{qi} + L\bar{A}^{i-2} AK\right) y_{k-i} + CBu_{k-1} \\ &+ \sum_{i=2}^{q} CA\left(F_q G_{q(i-1)} + L\bar{A}^{i-2} B\right) + \eta_k. \end{aligned} \tag{6.48}$$

Equations (6.28), (6.38) and (6.48) represent the AutoRegressive with eXogeneous input (ARX) models of linear systems in various different noise situations. The equation in each case provides a best prediction of the output measurement at time k in the sense of minimum state error at time $k-1$ using q previous input/output data.

6.4.2 Least-squares Identification for an ARX Model

A general ARX model of a linear system can be written as

$$y_k = \sum_{i=1}^{q1} A_i y_{k-i} + \sum_{i=1}^{q2} B_i u_{k-i} + \epsilon_k, \tag{6.49}$$

where $(q1, q2)$ is the order of the model. Given a set of input/output data $\{y_k, \cdots, y_0, u_k, \cdots, u_0\}$ of the system, we can use the least-squares method to find a set of matrix coefficients $\{\hat{A}_1, \cdots, \hat{A}_{q1}, \hat{B}_1, \cdots, \hat{B}_{q2}\}$ which fits the equation optimally in the least-squares error of output prediction sense. The least-squares method for single-input single-output ARX model (a scalar equation) can be found in many text books;[16] for a multi-input multi-output ARX model, one can refer

to Section 5.3. The ARX models derived in the last section have order (q, q), or just q in short, which are special cases of the general ARX model.

We claim that if we do least-squares input/output data fitting using an ARX model as (6.49) with $q1 = q2 = q$, the ARX model will converge to that derived from a projection filter of the same order (q) if the number of the data sample is sufficiently large. In other words, though the projection filter is derived based on the criterion of resulting least-mean-square state error, it also provides least-squares output error, which can be proved as follows.

The a priori output estimation by a projection filter of order q is

$$\hat{y}_k \triangleq C\hat{x}_k^- = CA\hat{x}_{k-1} + CBu_{k-1} \tag{6.50}$$

where $\hat{x}_{k-1}$ is the estimate of x_{k-1} made by the projection filter based on q previous input/output data including the output at time $k-1$. The output prediction error ϵ is

$$\begin{aligned} \epsilon_k \triangleq y_k - \hat{y}_k &= [CAx_{k-1} + CBu_{k-1} + Cw_{k-1} + v_k] - [CA\hat{x}_{k-1} + CBu_{k-1}] \\ &= CAe_{k-1} + Cw_{k-1} + v_k \end{aligned} \tag{6.51}$$

where e_{k-1} is the state estimation error. Since w_k and v_k are zero-mean, uncorrelated sequences by assumption and are also independent of e_{k-1}, we have

$$E[\epsilon_k \epsilon_k^T] = CAE[e_{k-1}e_{k-1}^T]A^T C^T. \tag{6.52}$$

Therefore, minimizing the mean-square of the state estimation error is equivalent to minimizing the mean-square of the output estimation error. Since the steady state output estimation error is a stationary sequence, by the assumption of ergodicity, minimizing the mean-square of it is equivalent to minimizing the sum of the squares of all the error, provided the number of the samples is large enough. On the other hand, since (6.49) is a linear equation, the solution of the least-squares output prediction is unique. Therefore, we conclude that the ARX model derived

from the projection filter also gives least-squares output predictions, and the ARX model identified using the least-squares method converges to that derived from the projection filter.

6.4.3 Obtaining System Markov Parameters from an ARX Model

There are some special relations between the Markov parameters of the system and the coefficient matrices of the ARX models derived in the previous section. Based on these relations we can obtain the Markov parameters from the ARX models.

For noise-free systems, from (6.28) if we denote the coefficient matrices of y_{k-j} and u_{k-j} by A_j and B_j, respectively, we can have

$$CA^jB = B_{j+1} + \sum_{i=1}^{j} A_i CA^{j-i}B. \tag{6.53}$$

This equation can be used iteratively to calculate the system Markov parameters CA^jB $(j = 1, \cdots, q-1)$ from the coefficient matrices of an ARX model of order q (note $B_1 = CB$).

Proof:

By definition

$$\begin{aligned} G_{qj} &= \begin{bmatrix} 0 \\ \vdots \\ CA^{-1}B \\ \vdots \\ CA^{-q+j}B \end{bmatrix} = \begin{bmatrix} C \\ \vdots \\ CA^{-j+1} \\ CA^{-j+2} \\ \vdots \\ CA^{-q+1} \end{bmatrix} A^{j-1}B - \begin{bmatrix} CA^{j-1} \\ \vdots \\ C \\ 0 \\ \vdots \\ 0 \end{bmatrix} B \\ &= H_q A^{j-1}B + D_{j-1}B \end{aligned} \tag{6.54}$$

where

$$D_{j-1} = \left[(CA^{j-1})^T, \cdots, C^T, 0, \cdots 0\right]^T.$$

Therefore,

$$\begin{aligned} B_{j+1} &= CAF_qGqj = CAF_qH_qA^{j-1}B + CAF_qD_{j-1} \\ &= CA^jB - \sum_{i=1}^{j} CAF_{qi}CA^{j-i}B \\ &= CA^jB - \sum_{i=1}^{j} A_iCA^{j-1}B, \end{aligned} \tag{6.55}$$

because

$$F_qH_q = I_n; \tag{6.56}$$

hence, (6.53) follows.

Q.E.D.

We can also iteratively calculate CA^jB ($j = q,\ q+1,\ ...$) by

$$\begin{aligned} CA^jB &= CA(F_qH_q)A^{j-1}B \\ &= \sum_{i=1}^{q} A_iCA^{j-i}B. \end{aligned} \tag{6.57}$$

Though derived from noise-free systems, the above equations ((6.53) to (6.57)) also hold for systems with additive white measurement noise. Because for systems with white measurement noise the projection filter F_q is a weighted pseudo-inverse of H_q (see (6.35)), hence (6.56) also holds.

It is interesting to see that (6.53) also holds for systems with both process and measurement noise even though (6.56) does not hold in this case. This can be proved as follows.

proof:

Using the expressions of the terms defined in (6.48), we have

$$B_{j+1} + \sum_{i=1}^{j} A_iCA^{j-i}B$$

$$
\begin{aligned}
&= CA(F_qG_{qj} + L\bar{A}^{j-1}B) + CAF_{q1}CA^{j-1}B + \\
&\quad \cdots + CA(F_j + L\bar{A}^{j-2}AK)CB \\
&= CAF_qG_{qj} + CAL\bar{A}^{j-1}B + \sum_{i=1}^{j} CAF_{qi}CA^{j-i}B \\
&\quad + \sum_{i=1}^{j-1} CAL\bar{A}^{i-1}AKCA^{j-i-1}B \\
&= CAF_qH_qA^{j-1}B + CAL\bar{A}^{j-1}B + \sum_{i=1}^{j-1} CAL\bar{A}^{i-1}AKCA^{j-i-1}B \\
&= CAF_qH_qA^{j-1}B + CAL\bar{A}^{j-1}B + CAL\bar{A}^{j-2}AKCB \\
&\quad + \sum_{i=1}^{j-1} CAL\bar{A}^{i-2}AKCA^{j-i-1}B \\
&= CAF_qH_qA^{j-1}B + CAL\bar{A}^{j-2}(\bar{A} + AKC)B \\
&\quad + \sum_{i=1}^{j-1} CAL\bar{A}^{i-1}AKCA^{j-i-1}B \\
&= CAF_qH_qA^{j-1}B + CAL\bar{A}^{j-2}AB + \sum_{i=1}^{j-2} CAL\bar{A}^{i-1}AKCA^{j-i-1}B \\
&= \cdots \\
&= CAF_qH_qA^{j-1}B + CALA^{j-1}B \\
&= CA(F_qH_q + L)A^{j-1}B \\
&= CA^jB
\end{aligned}
\tag{6.58}
$$

where the relations $\bar{A} + AKC = A$ and $F_qH_q + L = I_n$ are used.

Q.E.D.

However, (6.57) does not hold for systems with process noise. Hence, for an ARX model of order q, only q terms of the Markov parameters can be obtained.

Therefore, from the identified ARX model one can identified a set of Markov parameters of the system. To decompose the Markov parameters into state space parameters [A, B, C], the Eigensystem Realization Algorithm (ERA) can be used.

6.4.4 Numerical Example

In the numerical example, the simulated Mini-Mast model of the last section is used again. The order of the ATP is reduced from 100 to 50. The identified modal frequencies and damping factors with the corresponding number of data processed are listed in Table 6.3. The results are fairly accurate. The Kalman filter gain is calculated according to the inverse filter method introduced in Section 5.6 using the identified state space parameters, where $q1$ in (5.65) is set to 20.

The results of two cases are shown, that is, the cases of using 1,000 and 5,000 data. The results are arranged in the same way as in the Mini-Mast case of the last section, except the figures of estimated states are skipped. Figures 6.15 and 6.16 show the comparison of the true and the reconstructed system Markov parameters. Figures 6.17 and 6.18 show the comparison of singular value responses of the transfer functions. A good agreement is obtained. Figures 6.19 and 6.20 show the comparison of the true and the reconstructed filter Markov parameters ($C\hat{\bar{A}}^{i-1}AK$ and $\hat{C}\bar{A}^{i-1}\hat{A}\hat{K}$, $i = 1, \cdots, 100$), which can indicate the quality of the estimated Kalman filter gain. The comparison of the true and estimated outputs, the optimal and the estimated residual are shown in (a) and (b) parts of Figs. 6.21 and 6.22, and the auto-correlation functions of the residuals are in (c) parts. From all these results we can see the method works very well.

6.5 Concluding Remarks

This chapter presented two methods for state estimation under unknown system model and noise covariances. The main results are summarized as follows:

(1) The simultaneous method obtains simultaneously the state space parameters

and the corresponding optimal steady state Kalman filter gain of the system from the coefficients of an ARX model of large order identified through the adaptive transversal predictor. The ARX model is derived based on Kalman filter formulations. There are two different ways to calculate a state space model and the corresponding Kalman filter gain from the coefficients of the ARX model. One uses the eigensystem realization algorithm (ERA) to decompose the coefficient sequences directly and from which identifies all the parameters; the other combines the coefficients to form the Markov parameters of the system first, then uses ERA to realize a state space model, and finally the Kalman filter gain is obtained using the ARX coefficient method described in Section 5.5. The numerical example shows the latter is more effective in identifying system modal parameters.

(2) The sequential method also obtains state space parameters from the coefficients of an ARX model which is derived based on the projection filter; however, the order of the ARX model is smaller than that of the simultaneous method, and the residual might not be white. Similar to the second way of the previous method, the coefficients of the ARX model are combined to form the Markov parameters of the system first, and then ERA is used to obtain a state space model from it; the corresponding Kalman filter gain is obtained using the inverse filter method described in Section 5.6.

(3) In practical applications, because the information about process and measurement noises is usually not available, it is difficult to judge how large the order of the ARX model should be in order to apply the first method. In addition, if process noise is small, the order of the ARX model of the first method will be very large, and the computaional load can be excessive. Since the second method does not have such difficulty, it is more suitable for practical applications from this point of view. However, the second method cannot obtain a state space model and the corresponding Kalman filter gain at the same time.

Therefore, additional efforts are required to obtain the filter gain.

(4) If a system has no repeated eigenvalues, the state space model can be normalized to have a unique solution; however, having repeated eigenvalues, it does not have a unique solution even normalized as described in Section 6.3.3.

(5) The numerical examples show that both methods are very effective in system identification and modal state estimation for flexible structures such as Mini-Mast, even in the presence of closely-spaced or repeated modes.

Table 6.1 Identified modal parameters of the beam-like structure.

	# of data processed	Mode 1 Freq. (rad/sec)	Mode 1 Damp (%)	Mode 2 Freq. (rad/sec)	Mode 2 Damp (%)	Mode 3 Freq. (rad/sec)	Mode 3 Damp (%)
Method 1	1,000	*	*	3.0230	19.11	10.7293	2.43
	2,000	1.6711	16.06	4.5988	5.14	10.6371	4.15
	3,000	1.6920	6.59	4.5118	3.79	7.0251	11.29
	4,000	1.6405	5.22	4.4252	1.69	6.5016	5.73
	5,000	1.6245	4.73	4.4139	0.89	6.3176	4.75
Method 2	1,000	1.6397	0.93	4.4794	0.93	6.1116	1.49
	2,000	1.6347	0.87	4.4798	1.15	6.1063	1.47
	3,000	1.6371	0.95	4.4829	1.15	6.1080	1.42
	4,000	1.6414	0.89	4.4806	1.16	6.1080	1.41
	5,000	1.6407	0.86	4.4776	1.21	6.1112	1.39
	0 [a]	1.6369	0.63	4.4719	1.01	6.1085	1.30

* Cannot be identified

[a] True values

Table 6.2 Identified modal parameters of Mini-Mast using the simultaneous method.

# of data processed	Mode 1		Mode 2		Mode 3		Mode 4		Mode 5	
	Freq. (rad/sec)	Damp (%)	Freq. (rad/sec)	Damp (%)	Freq. (rad/sec)	Damp (%)	Freq. (rad/sec)	Damp (%)	Freq. (rad/sec)	Damp (%)
1,000	4.9863	2.81	5.0891	1.11	27.0820	1.11	37.5743	3.04	38.6144	1.13
2,000	5.0213	1.91	5.0609	1.94	27.4867	0.87	38.1814	1.45	38.7552	1.22
3,000	5.0246	1.97	5.0610	1.83	27.4548	1.02	38.2935	1.26	38.7079	1.15
4,000	5.0366	1.92	5.0459	1.60	27.5389	0.95	38.2625	1.27	38.7368	1.12
5,000	5.0263	1.97	5.0405	1.50	27.4746	1.13	38.3041	1.20	38.7314	1.09
0[a]	5.0318	1.80	5.0356	1.80	27.4201	1.20	38.3511	1.00	38.6823	1.00

[a]True values

Table 6.3 Identified modal parameters of Mini-Mast using the sequential method.

# of data processed	Mode 1		Mode 2		Mode 3		Mode 4		Mode 5	
	Freq. (rad/sec)	Damp (%)	Freq. (rad/sec)	Damp (%)	Freq. (rad/sec)	Damp (%)	Freq. (rad/sec)	Damp (%)	Freq. (rad/sec)	Damp (%)
1,000	4.9840	0.53	5.1264	1.01	26.6878	1.12	38.3373	2.49	38.6320	1.84
2,000	5.0364	1.17	5.0511	1.69	27.2379	0.49	38.2784	0.91	38.7403	2.00
3,000	5.0195	1.45	5.0659	1.14	27.2445	0.98	38.3330	1.35	38.6660	1.37
4,000	5.0342	1.09	5.0519	1.42	27.2904	0.89	38.2720	1.53	38.6566	1.13
5,000	5.0328	1.66	5.0328	0.90	27.3454	0.79	38.1475	1.61	38.6544	1.03
0[a]	5.0318	1.80	5.0356	1.80	27.4201	1.20	38.3511	1.00	38.6823	1.00

[a]True values

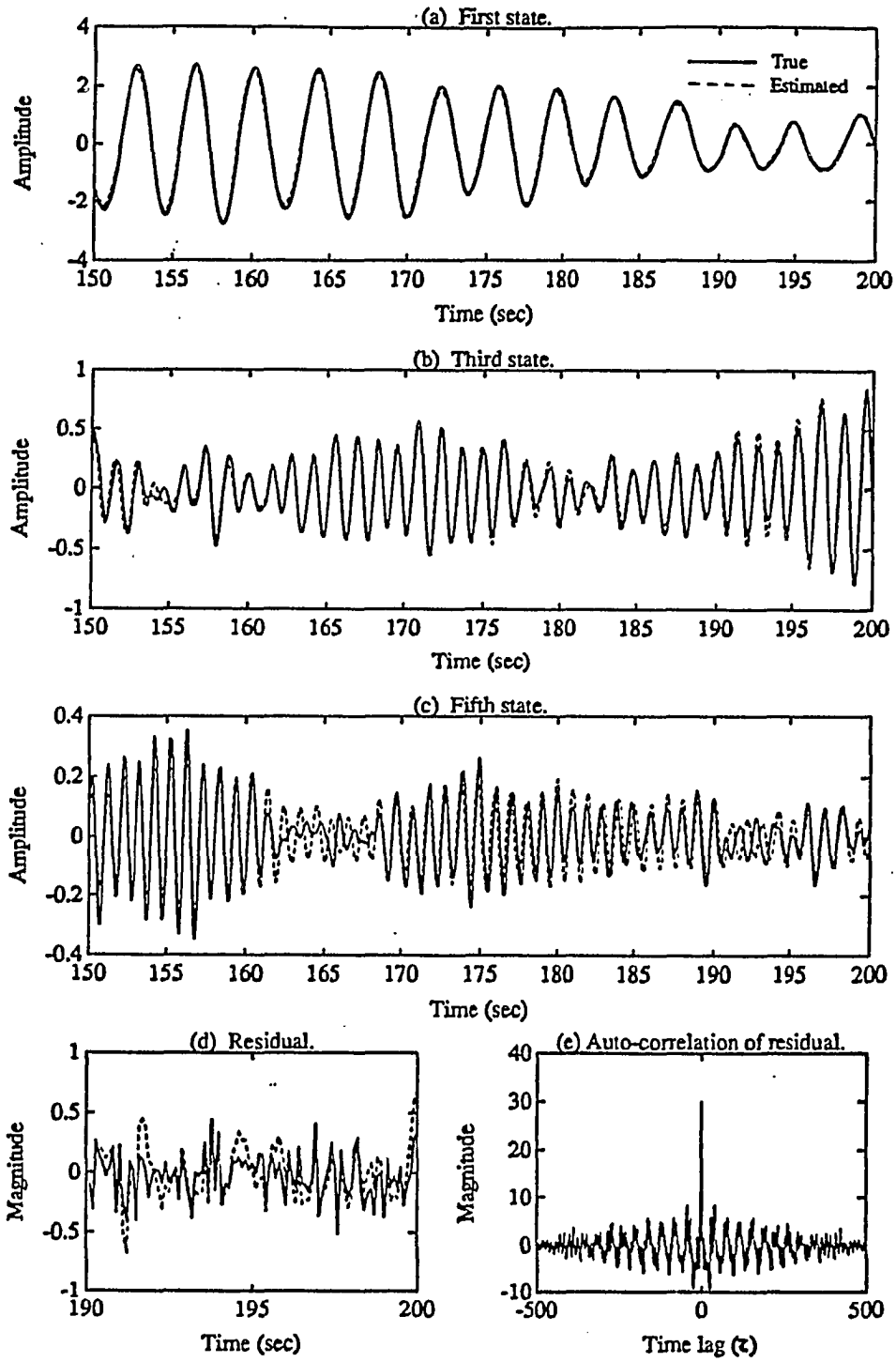


Fig. 6.1 State estimation results of the beam-like structure using 500 data samples.

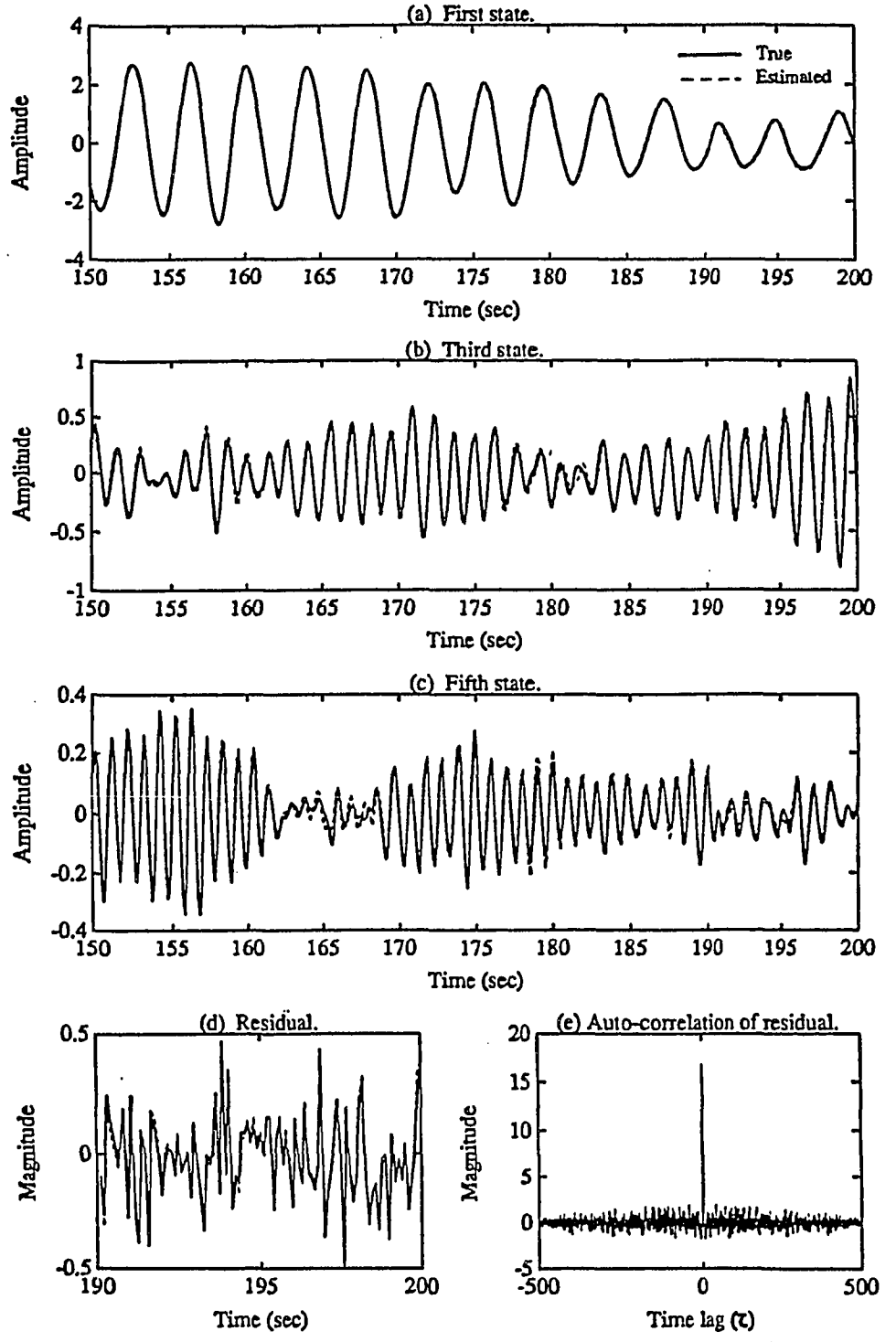


Fig. 6.2 State estimation results of the beam-like structure using 1,000 data samples.

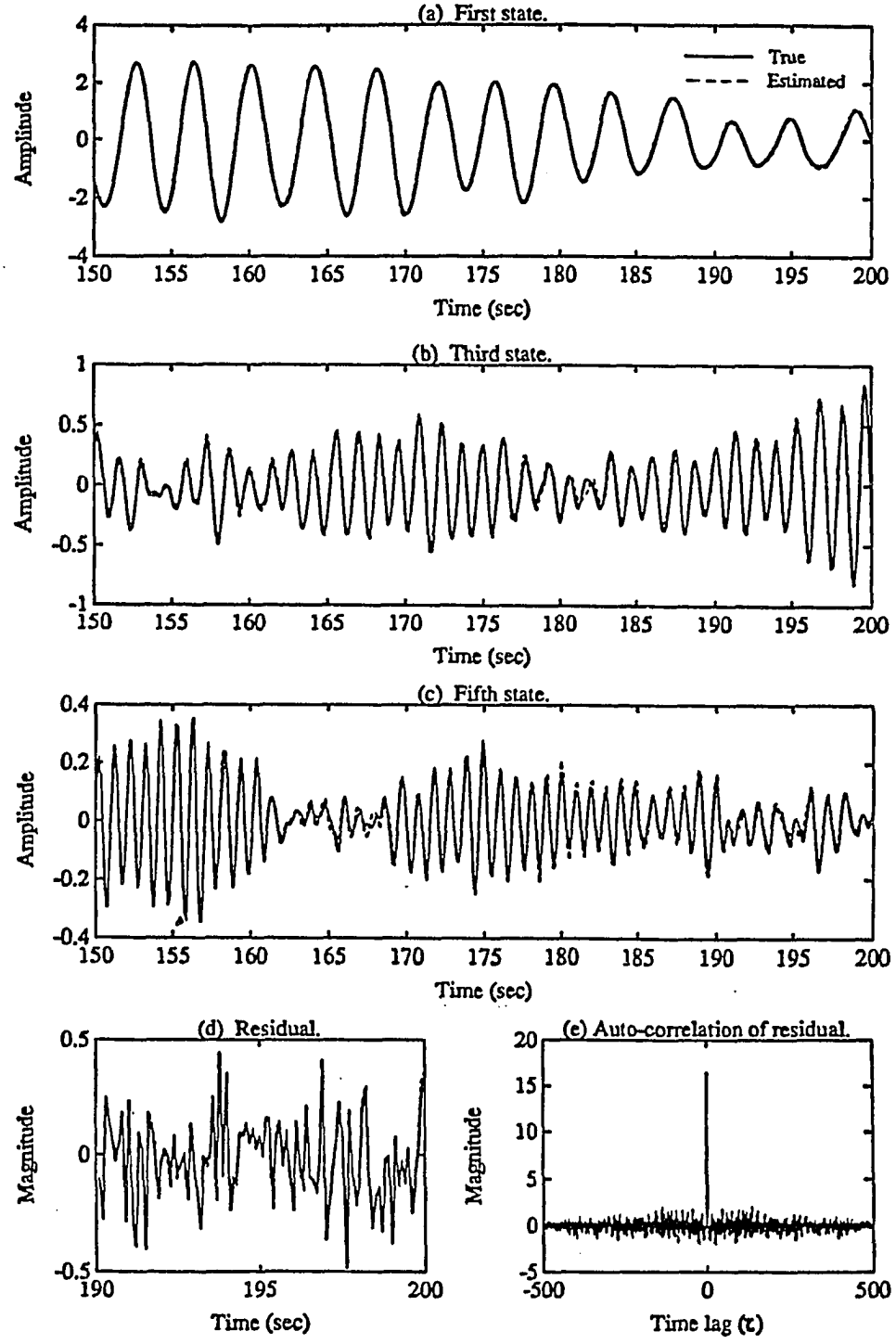


Fig. 6.3 State estimation results of the beam-like structure using 5,000 data samples.

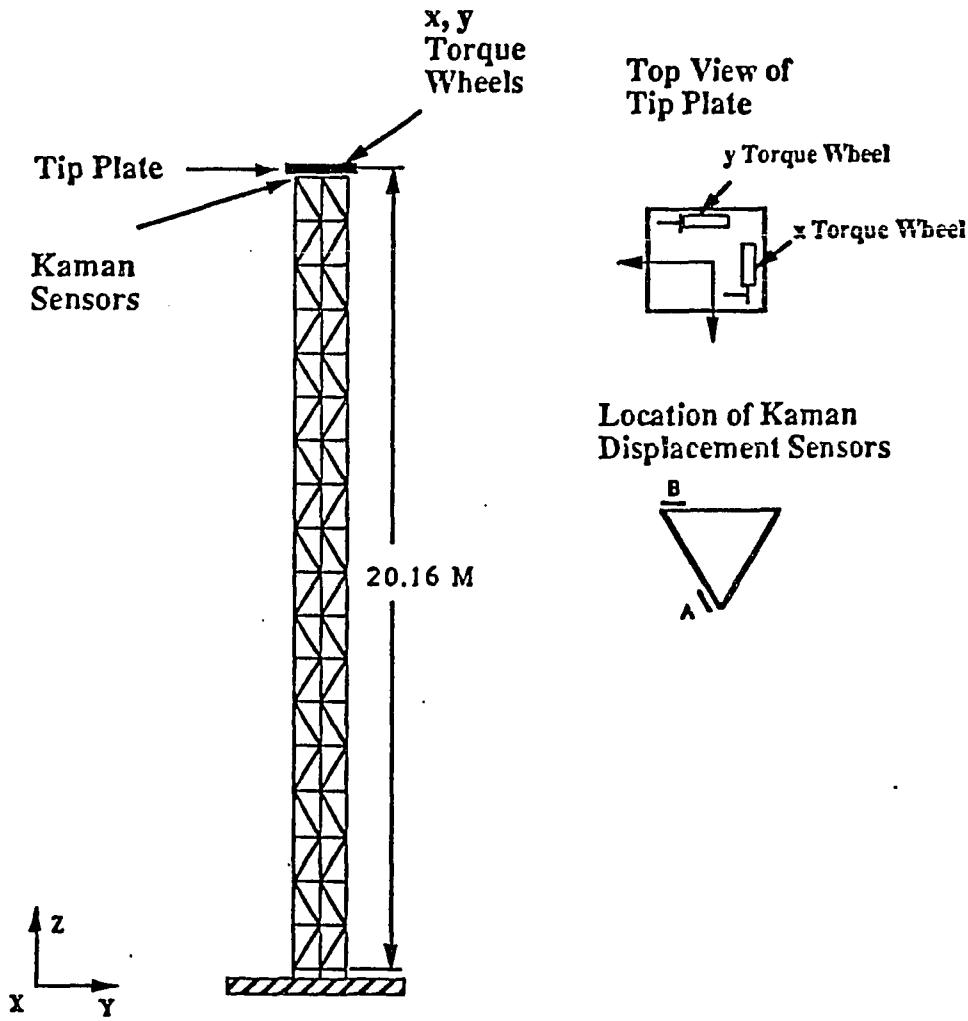


Fig. 6.4 Mini-Mast structure.

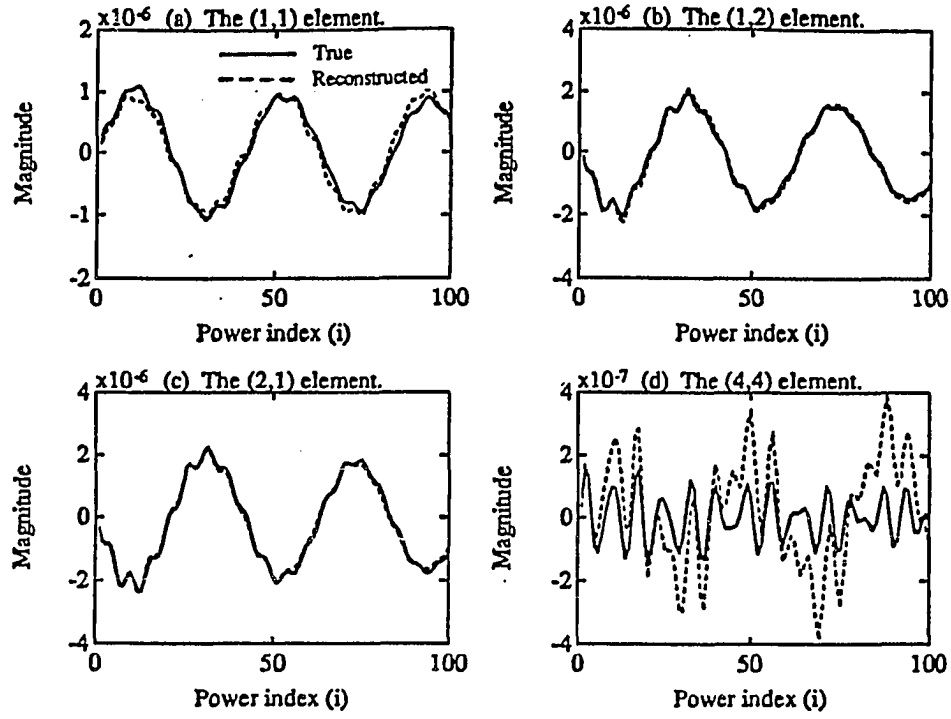


Fig. 6.5 Comparison of true and reconstructed system Markov parameters $CA^{i-1}B$ using the simultaneous method with 1,000 data samples.

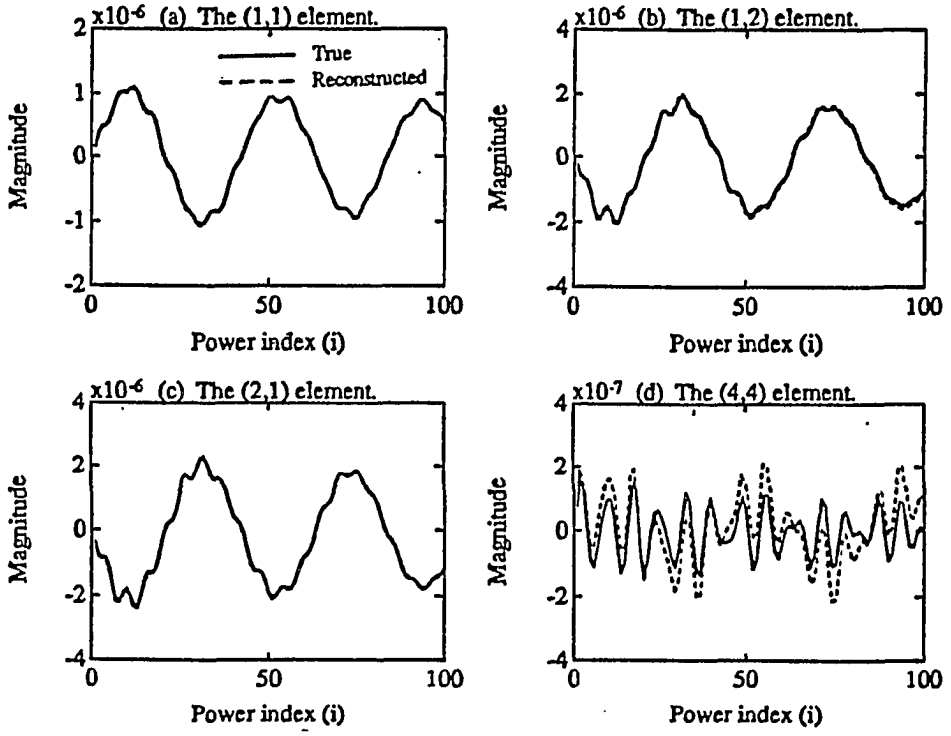


Fig. 6.6 Comparison of true and reconstructed system Markov parameters $CA^{i-1}B$ using the simultaneous method with 5,000 data samples.

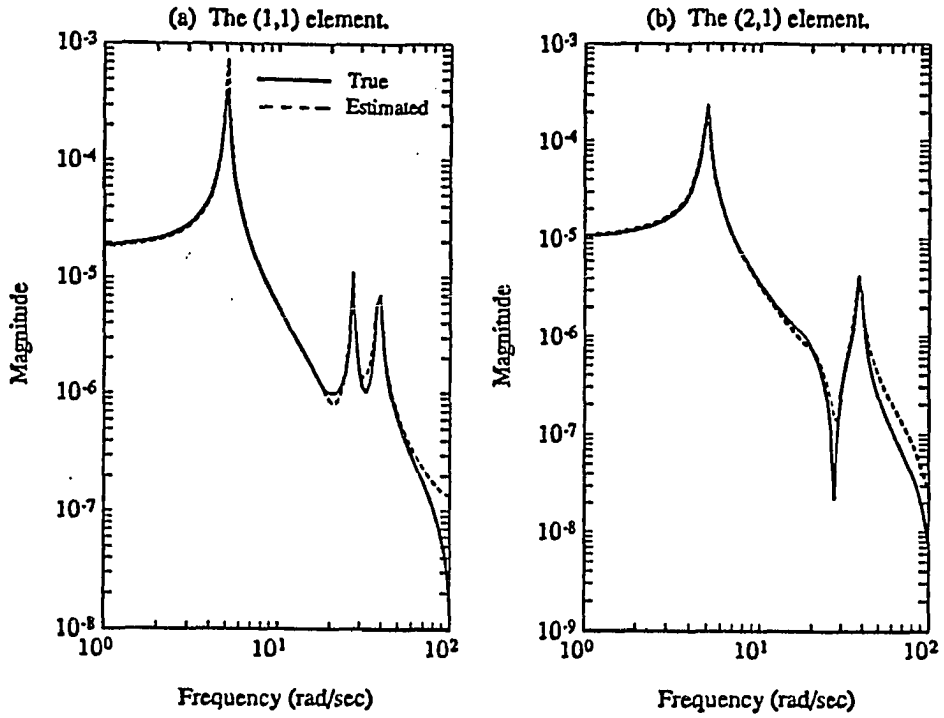


Fig. 6.7 Comparison of singular value responses of true and estimated transfer functions using the simultaneous method with 1,000 data samples.

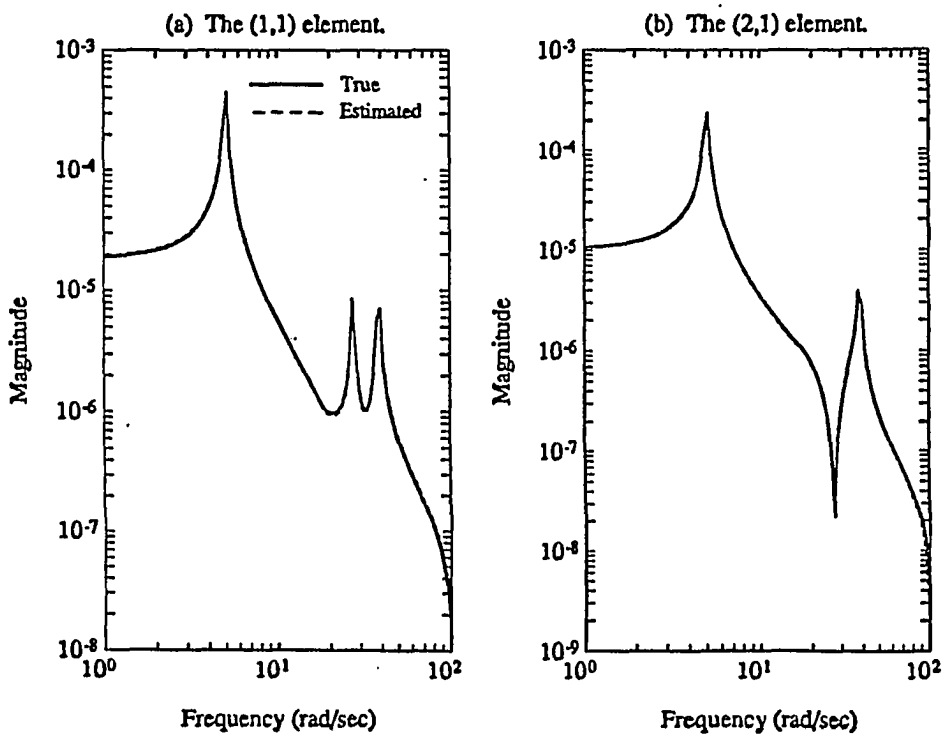


Fig. 6.8 Comparison of singular value responses of true and estimated transfer functions using the simultaneous method with 5,000 data samples.

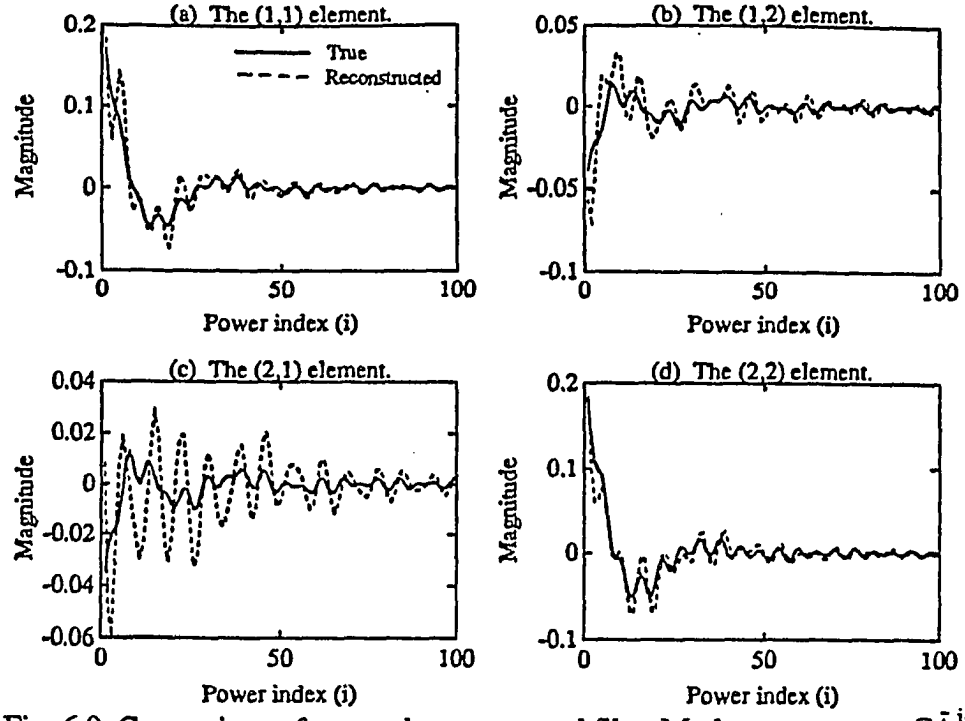


Fig. 6.9 Comparison of true and reconstructed filter Markov parameters $C\bar{A}^{i-1}AK$ using the simultaneous method with 1,000 data samples.

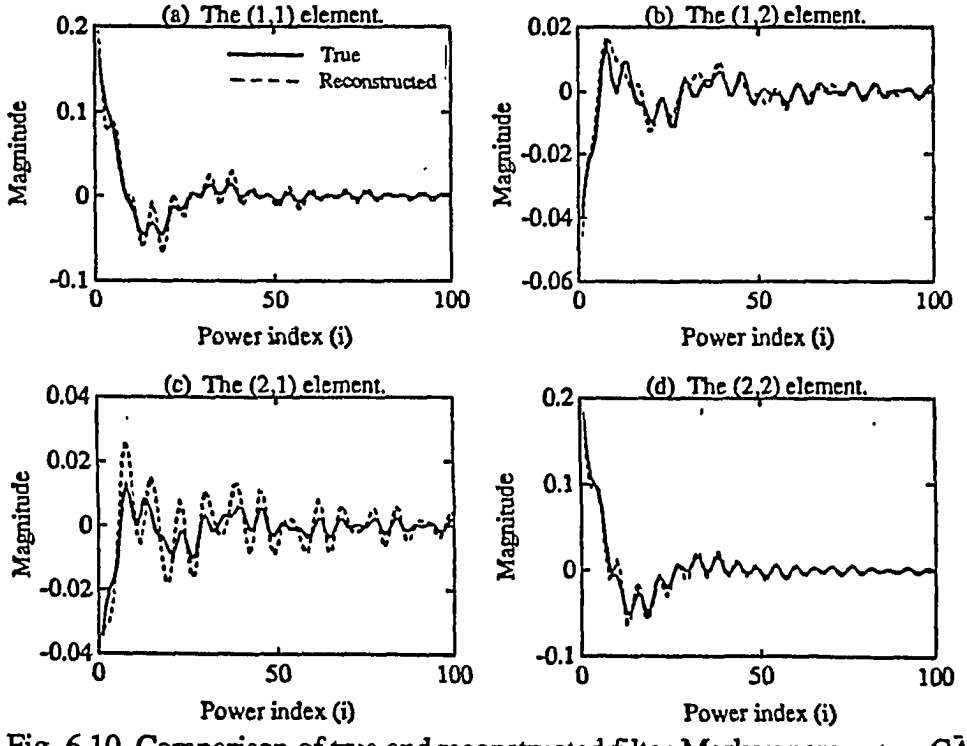


Fig. 6.10 Comparison of true and reconstructed filter Markov parameters $C\bar{A}^{i-1}AK$ using the simultaneous method with 5,000 data samples.

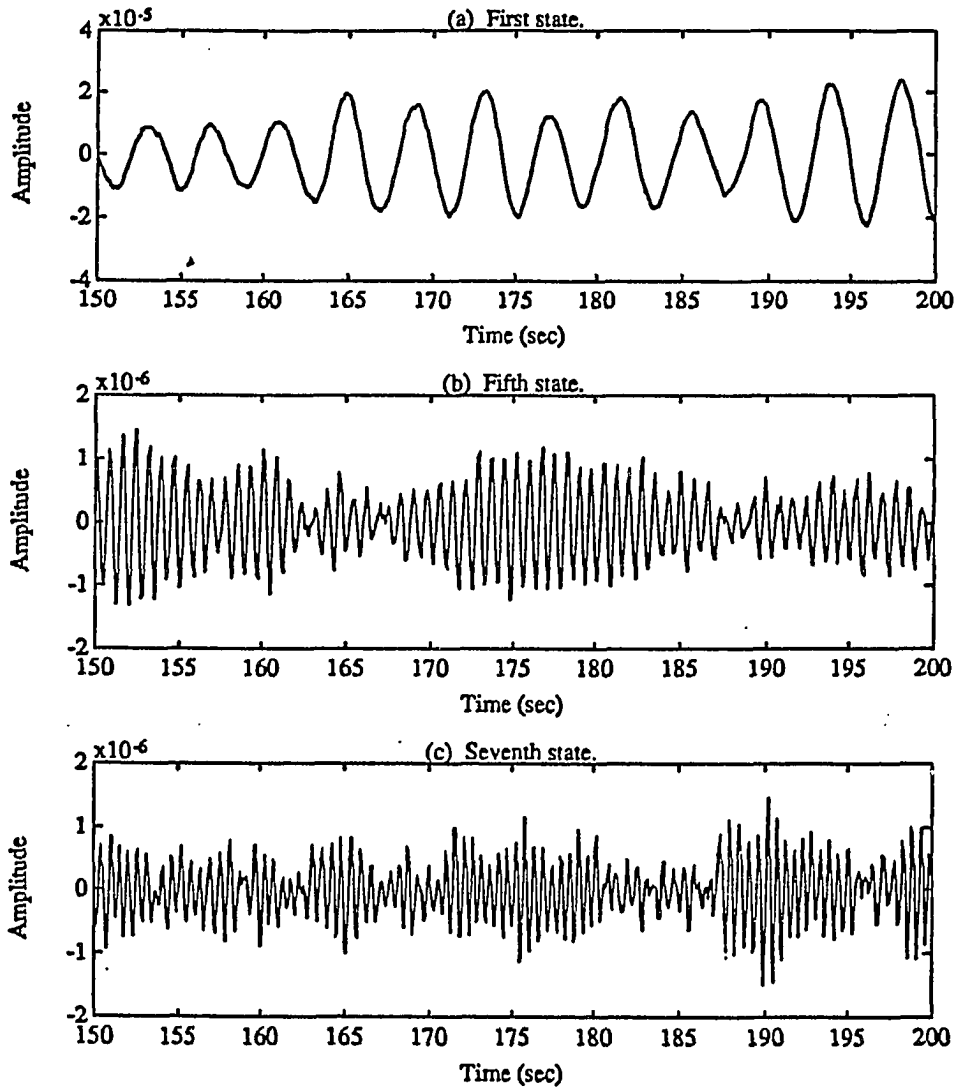


Fig. 6.11 Estimated states using the parameters obtained with 1,000 data samples using the simultaneous method.

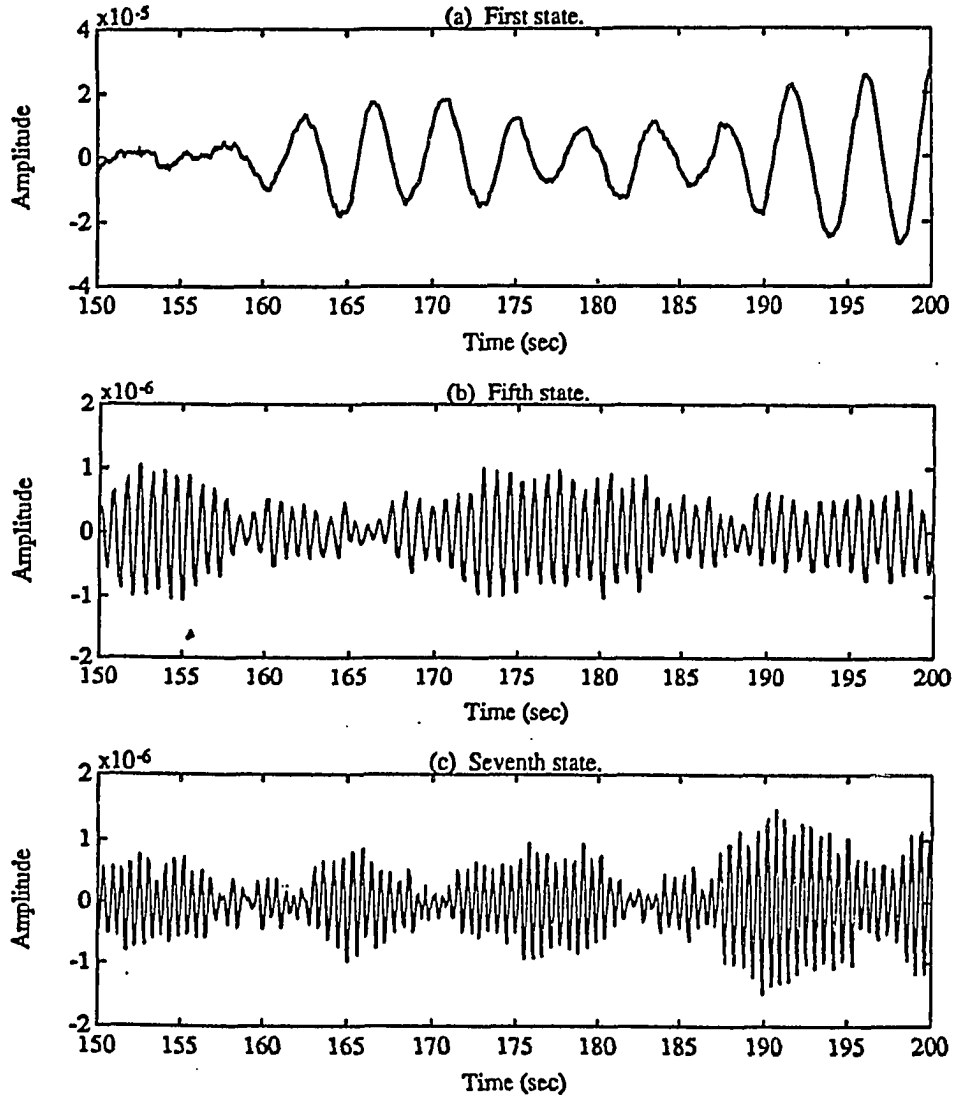


Fig. 6.12 Estimated states using the parameters obtained with 5,000 data samples using the simultaneous method.

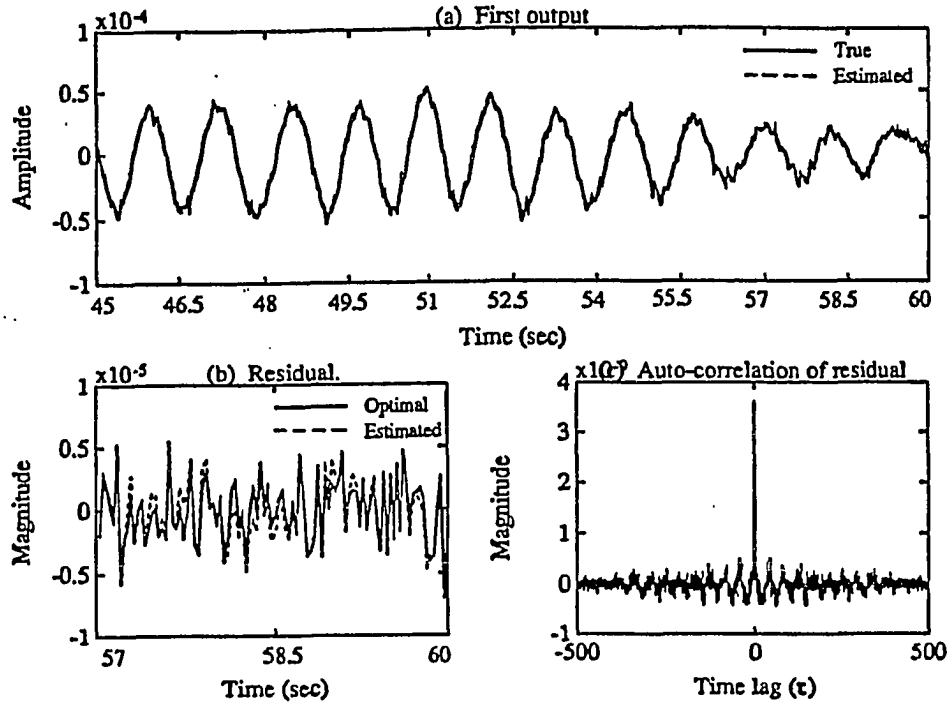


Fig. 6.13 Estimation of the first output using the simultaneous method with 1,000 data samples.

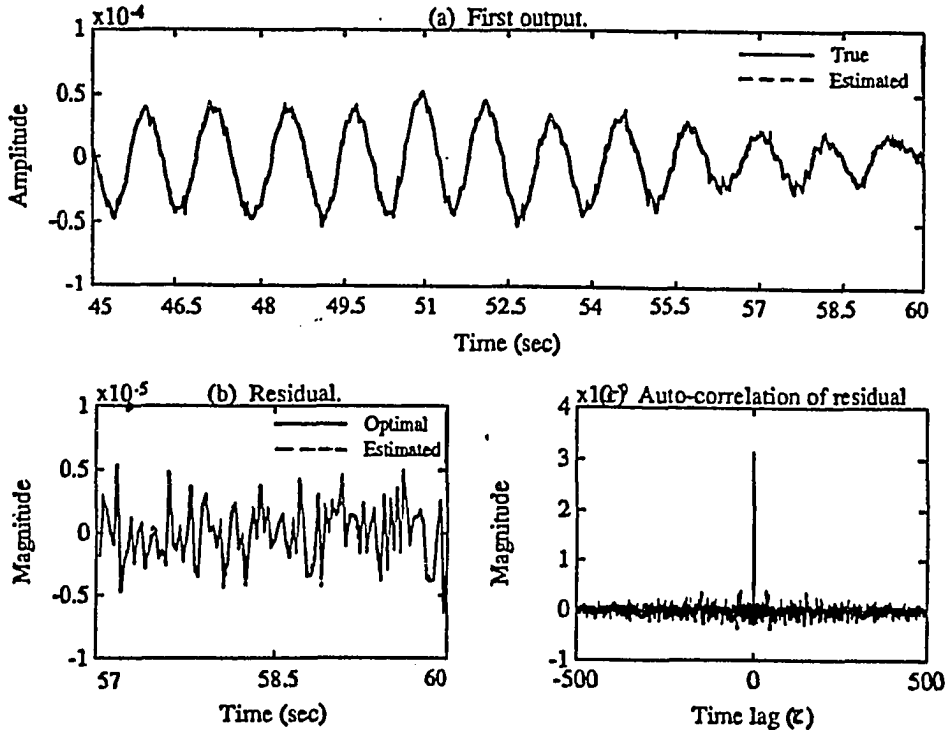


Fig. 6.14 Estimation of the first output using the simultaneous method with 5,000 data samples.

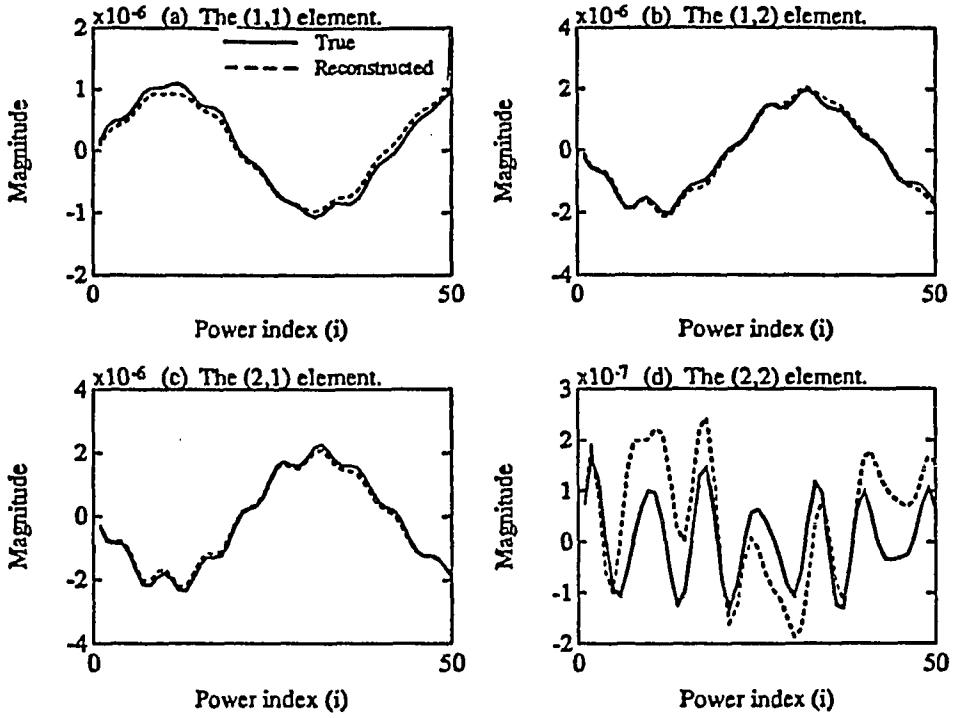


Fig. 6.15 Comparison of true and reconstructed system Markov parameters $CA^{i-1}B$ using the sequential method with 1,000 data samples.

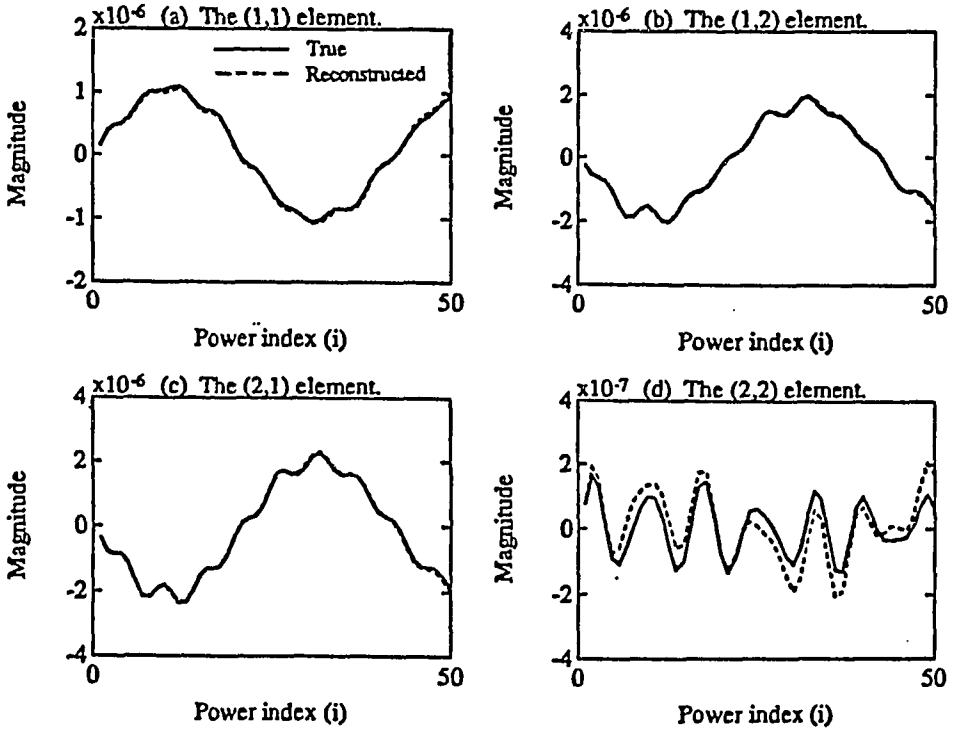


Fig. 6.16 Comparison of true and reconstructed system Markov parameters $CA^{i-1}B$ using the sequential method with 5,000 data samples.

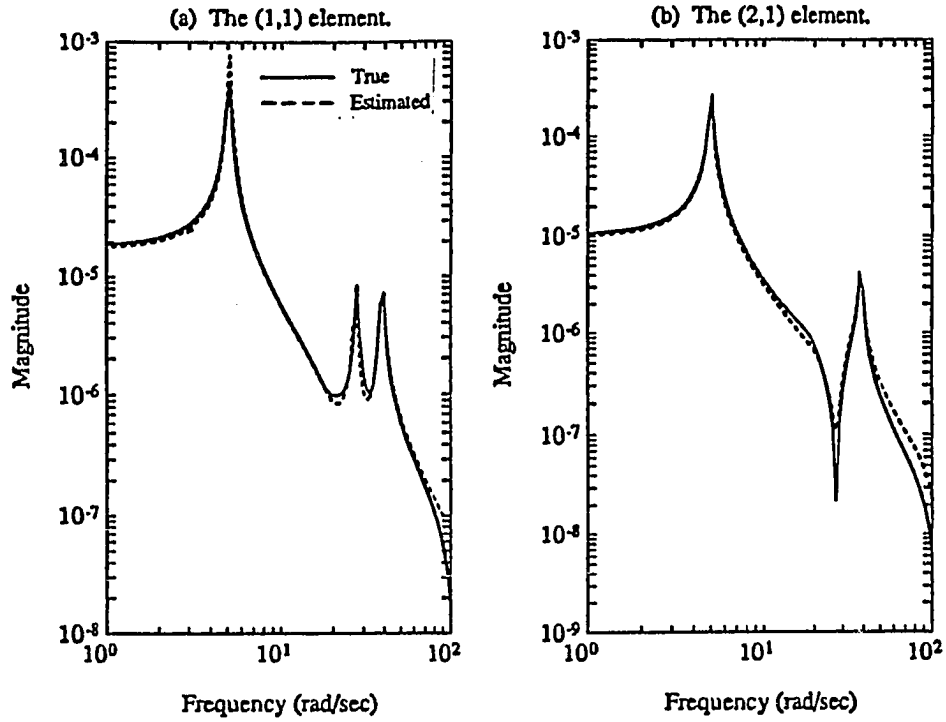


Fig. 6.17 Comparison of singular value responses of true and estimated transfer functions using the sequential method with 1,000 data samples.

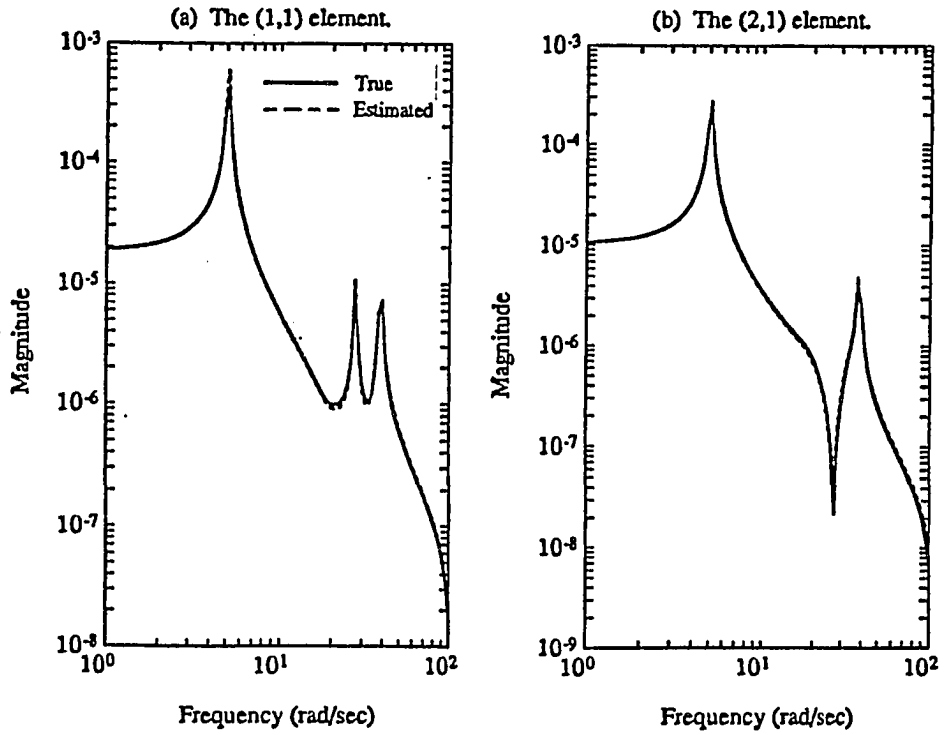


Fig. 6.18 Comparison of singular value responses of true and estimated transfer functions using the sequential method with 5,000 data samples.

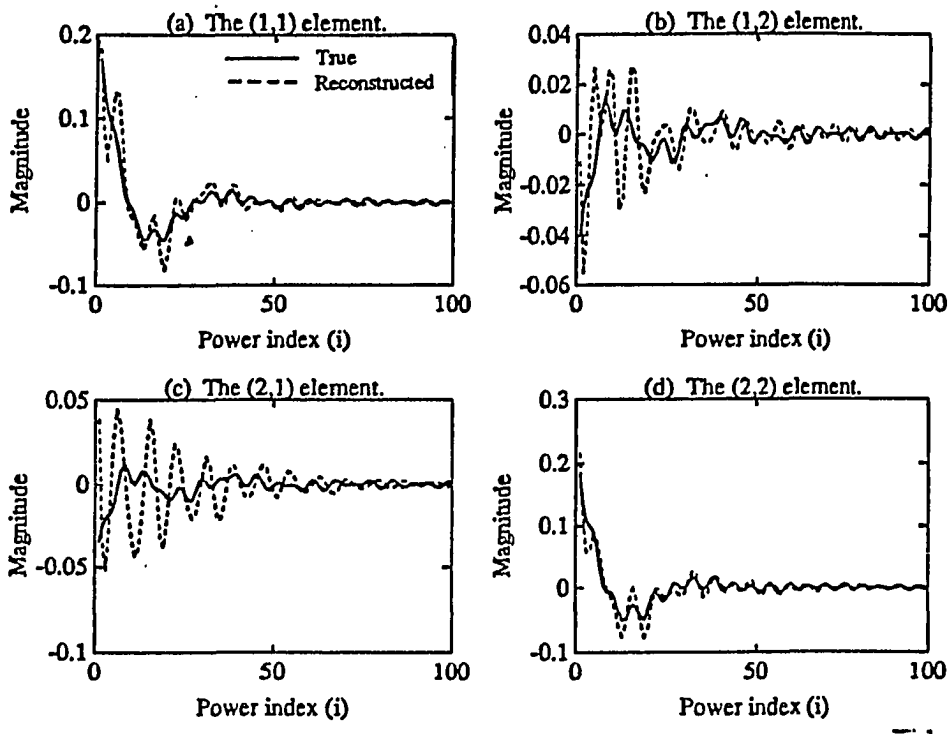


Fig. 6.19 Comparison of true and reconstructed filter Markov parameters $C\bar{A}^{i-1}AK$ using the sequential method with 1,000 data samples.

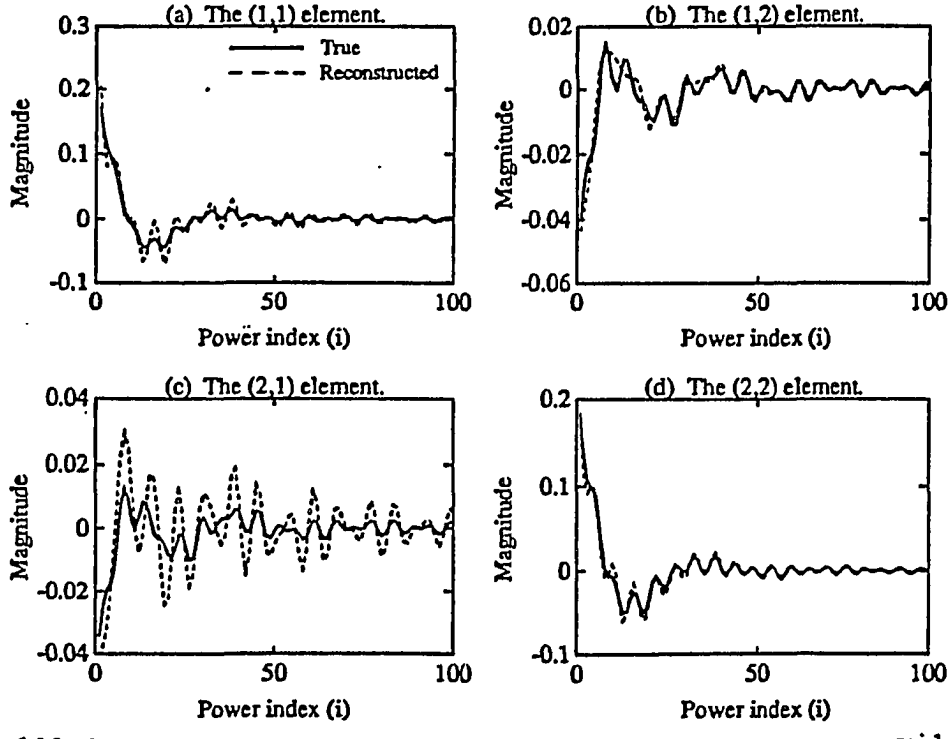


Fig. 6.20 Comparison of true and reconstructed filter Markov parameters $C\bar{A}^{i-1}AK$ using the sequential method with 5,000 data samples.

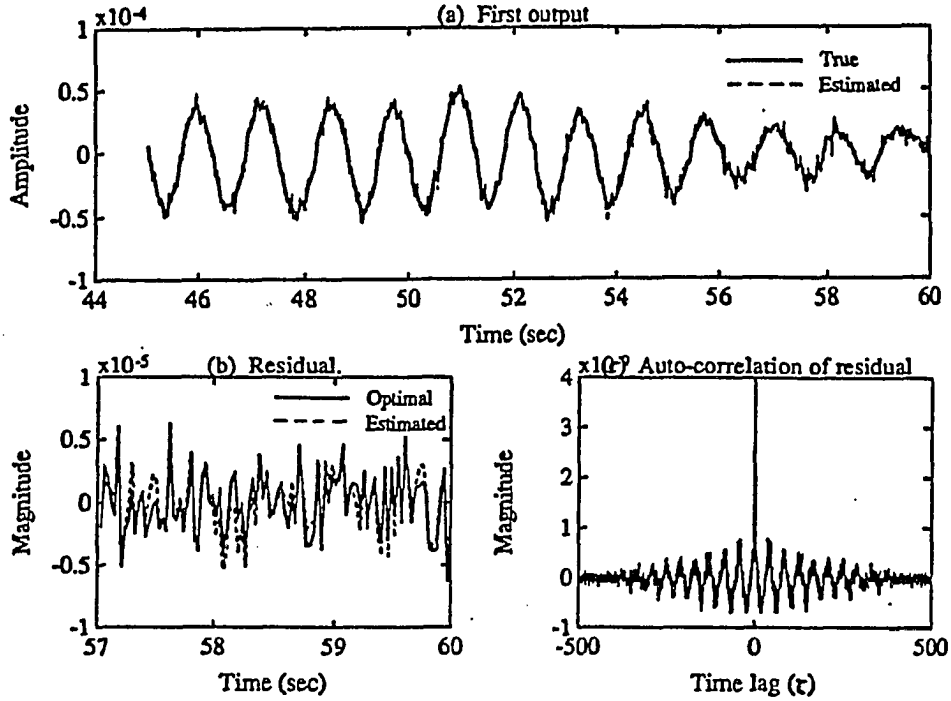


Fig. 6.21 Estimation of the first output using the sequential method with 1,000 data samples.

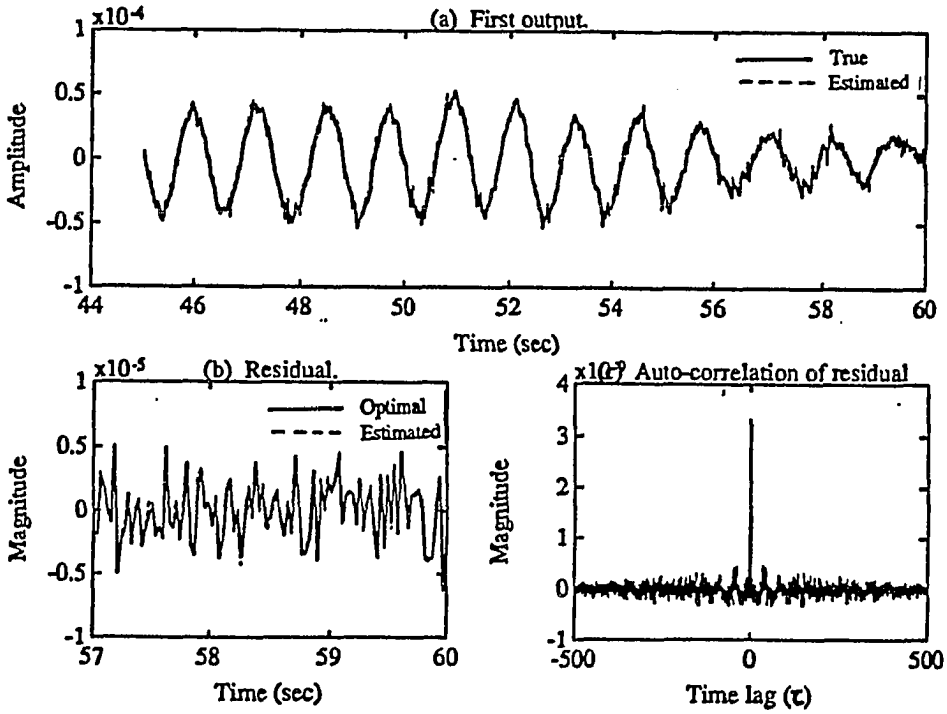


Fig. 6.22 Estimation of the first output using the sequential method with 5,000 data samples.

Chapter 7

EXPERIMENTAL VALIDATION

7.1 Introduction

This chapter uses an experimental example to demonstrate the feasibility of the integrated system identification and state estimation method developed in the previous chapters. A ten-bay structure as shown in Fig. 7.1 is considered. The truss is one of the structures built in NASA Langley Research Center for experiments in studies of control and structure interaction (CSI). It is 100 inches long, with a square cross section of 10 in $\times$ 10 in. All the tubing (longerons, battens, and diagonals) and ball joints are made of aluminum. The structure is in a vertical configuration attached from the top using an L-shaped fixture to a backstop. Two cold air thrusters acting in the same direction are placed at the tip. The thrusters which are used for excitation and control have a maximum thrust of 2.2 lb. each. A mass of approximately 20 lb. is attached at the beam tip to lower the fundamental frequency of the truss. Two servo accelerometers located at a corner of the square cross section provide the in-plan tip acceleration.

7.2 Modification for Using Acceleration Measurement

Since the outputs are acceleration signals in this experiment, the formulations derived in the previous chapters should be modified to accommodate this situation. The state space model used in the previous chapters assumes no direct influence from input force on output measurement. However, if the output measurement is acceleration instead of displacement, there is direct influence from input. Under this situation, the state space model should be modified to

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{6.1}$$

$$y_k = Cx_k + Du_k + v_k, \tag{6.2}$$

where Du_k represents the direct influence on output from input. The corresponding steady state Kalman filter innovation model becomes

$$\hat{x}_{k+1}^- = A\hat{x}_k^- + Bu_k + AK\varepsilon_k \tag{7.3}$$

$$y_k = C\hat{x}_k^- + Du_k + \varepsilon_k. \tag{7.4}$$

Introducing (7.4) into (7.3) yields:

$$\begin{aligned}\hat{x}_{k+1}^- &= A(I_n - KC)\hat{x}_k^- + (B - AKD)u_k + AKy_k \\ &= \bar{A}\hat{x}_k^- + (B - AKD)u_k + AKy_k\end{aligned} \tag{7.5}$$

where

$$\bar{A} = A(I_n - KC).$$

Similar to the derivation in Section 5.2 we can obtain the following input-output description:

$$\begin{aligned}y_k =& C\hat{x}_k^- + \varepsilon_k \\ =& C\bar{A}\hat{x}_{k-1}^- + C(B - AKD)u_{k-1} + CAKy_{k-1} + Du_k + \varepsilon_k \\ =& \cdots\end{aligned}$$

$$
\begin{aligned}
&= CAKy_{k-1} + C\bar{A}AKy_{k-2} + \cdots + C\bar{A}^{q-1}AKy_{k-q} + Du_k \\
&\quad + C(B - AKD)u_{k-1} + C\bar{A}(B - AKD)u_{k-2} + \cdots \\
&\quad + C\bar{A}^{q-1}(B - AK)u_{k-M} + C\bar{A}^q\hat{x}_{k-q}^- + \varepsilon_k \\
&\approx \sum_{i=1}^{q} C\bar{A}^{i-1}AKy_{k-i} + Du_k + \sum_{i=1}^{q} C\bar{A}^{i-1}(B - AKD)u_{k-i} + \varepsilon_k \\
&= \sum_{i=1}^{q} A_i y_{k-i} + \sum_{i=0}^{q} B_i u_{k-i} + \varepsilon_k \qquad (7.6)
\end{aligned}
$$

for a large integer q, where

$$A_i = \sum_{i=1}^{q} C\bar{A}^{i-1}AK \qquad (7.7)$$

$$B_i = C\bar{A}^{i-1}(B - AKD), \quad B_0 = D. \qquad (7.8)$$

Equation (7.6) is also an autoregressive with exogeneous input (ARX) model.

Compared to (5.12), (7.6) has an additional term Du_k, and B in (5.12) is replaced by $B - AKD$. However, these changes do not cause any difficulty in identifying the model parameter using the adaptive transversal predictor (ATP). Nevertheless, some of the identified parameters have different meanings; the parameters A_i's are the same as before, but B_i's are changed. However, since

$$
\begin{aligned}
C\bar{A}^{i-1}B &= C\bar{A}^{-1}B - C\bar{A}^{i-1}AKD + C\bar{A}^{i-1}AKD \\
&= C\bar{A}^{i-1}(B - AKD) + C\bar{A}^{i-1}AKD \\
&= B_i + A_iB_0, \qquad (i = 1, \cdots, q), \qquad (7.9)
\end{aligned}
$$

we can obtain an estimate of the matrix sequence $C\bar{A}^{i-1}B$ by combining the estimated parameters. After this procedure, all the methods developed in the previous chapters can be readily used.

7.3 Experimental Results

The structure was excited using random inputs to both thrusters for 30 seconds. The input signals were filtered to concentrate the energy in the low frequency range. A total of 7499 data points at sampling rate 250 Hz is taken. The two output acceleration signals were filtered using a three-pole Bessel filter with a break frequency of 20 Hz.

From the output we can tell the dominant mode is about 5 to 6 Hz. In order not to use too large an order in the adaptive transversal predictor (ATP), we reduced the sampling rate to 1/2 of the original one by choosing one out of every two samples. Hence, the sampling rate becomes 125 Hz and totals 3750 data. The order of the ATP is set to 100. The method introduced in Section 6.4 is used to identify a state space model. Figure 7.2 shows the system Markov parameters $C\widehat{A^{i-1}}B$ $(i = 1, \cdots, 100)$ identified from ATP. By ERA three modes are identified. The identified modal frequencies and dampings are listed as follows:

Mode	Frequency (rad/sec)	Damping (%)
1	37.0988	0.27
2	46.1175	2.87
3	304.4817	0.40

The corresponding state space parameters in normalized modal format are

$$\hat{A} = diag\left\{\begin{bmatrix} 0.9555 & 0.2922 \\ -0.2922 & 0.9555 \end{bmatrix} \begin{bmatrix} 0.9229 & 0.3568 \\ -0.3568 & 0.9229 \end{bmatrix} \begin{bmatrix} -0.7538 & 0.6423 \\ -0.6423 & -0.7538 \end{bmatrix}\right\}$$

$$\hat{B} = \begin{bmatrix} 0.1725 & -0.1117 & 0.1122 & -0.0321 & 0.3241 & -0.1871 \\ -0.1789 & 0.1267 & -0.1522 & 0.0556 & 0.3309 & -0.2725 \end{bmatrix}^T$$

$$\hat{C} = \begin{bmatrix} 1.7754 & 0.0000 & 1.0023 & 0.0000 & 1.3946 & 0.0000 \\ 0.9201 & 0.0362 & -1.6909 & -0.3692 & -1.4287 & 0.1185 \end{bmatrix}.$$

After obtaining state space parameters, the method introduced in Section 5.6 is used to estimate the corresponding Kalman filter gain. The identified stochastic

Markov parameters $\widehat{CA^iK}$ $(i = 1, \cdots, 100)$ are shown in Fig. 7.3. The estimated Kalman filter gains are

$$\hat{K} = \begin{bmatrix} 0.2787 & 0.1807 & 0.3437 & 0.1072 & 0.0065 & 0.0357 \\ 0.2277 & -0.0685 & 0.1236 & -0.0884 & -0.0874 & -0.0182 \end{bmatrix}^T$$

To show the results of state estimation, the first state of each mode is shown in Fig. 7.4. The shapes of higher frequency modes are not smooth, for the sampling rate is not high enough to give a good appearance. Since the modal model has been normalized, the amplitude of each modal state indicates the energy allocated in that mode. To evaluate the quality of the system identification and state estimation, the estimated outputs calculated based on the estimated state are compared to the true outputs. Because the true state is not available, the output comparison is the only way to validate the results. The output comparison, the residual and its corresponding auto-correlation functions are shown in Figs 7.5. The estimated and true outputs are in good agreement. The covariance of the difference between them (the residual) is less than 1.5 % of the covariance of the output. However, the auto-correlation function shows that the residual is not quite white. Given more data, the results can be improved.

Figure 7.6 is a frequency response diagram (Bode gain plot) of the structure obtained from frequency-domain modal analysis conducted in NASA. It is attached here for reference. The dashed lines represents those after curve fitting. The figure shows only a frequency range from 2 to 10 Hz, hence misses the third mode (48.4598 Hz) of the system. However, for the two modes shown, the frequencies are in good agreement with the identified frequencies (5.9045 Hz and 7.3398 Hz).

7.4 Concluding Remarks

Integrated system identification and state estimation has been successfully conducted for a ten-bay truss. The good agreement between the estimated and

the true output shows the identified model along with the filter gain are fairly good. The identified frequencies also agree with those obtained from frequency-domain analysis. This experimental example shows that the method is correct and has high potential in practical application.

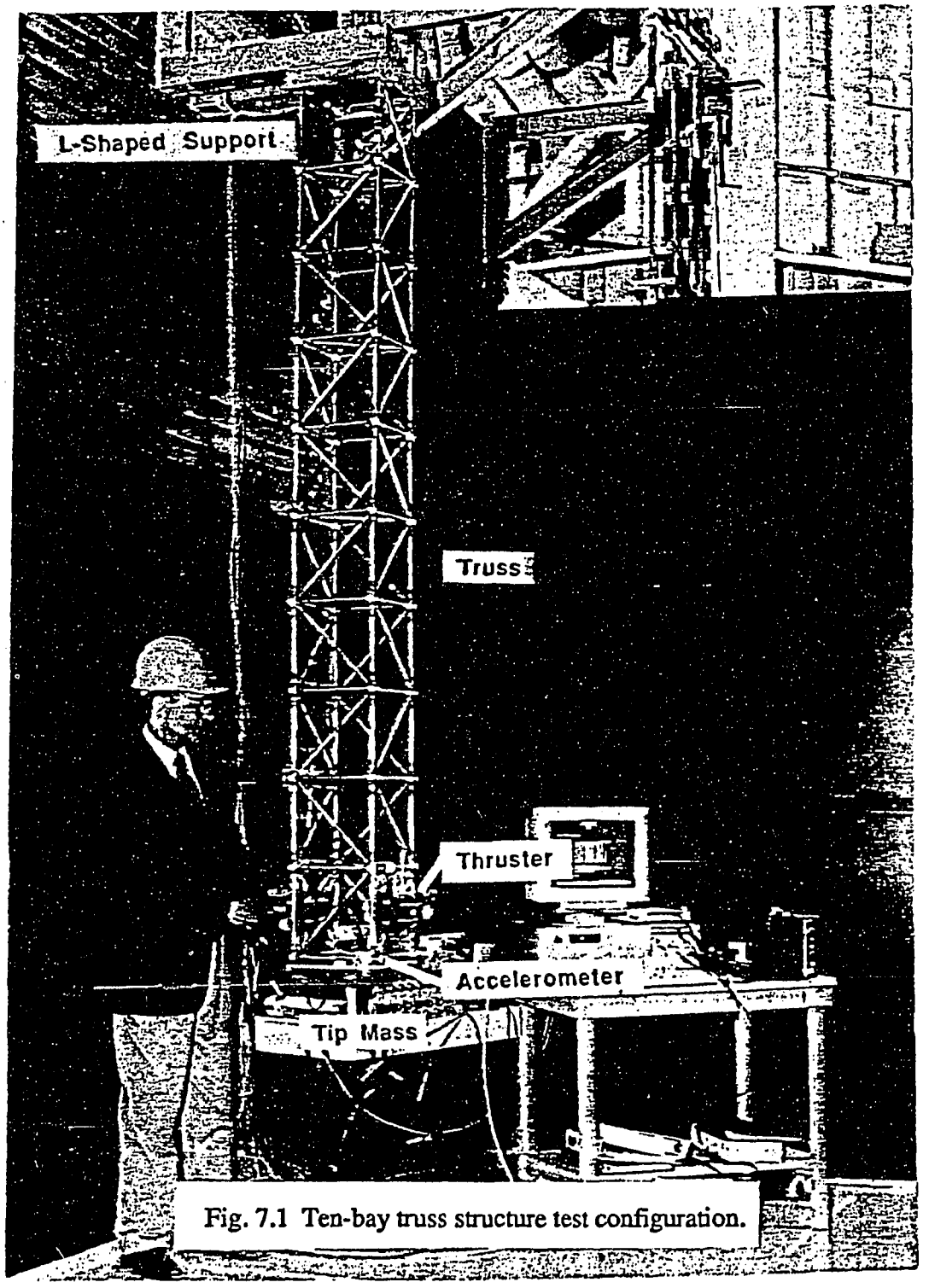


Fig. 7.1 Ten-bay truss structure test configuration.

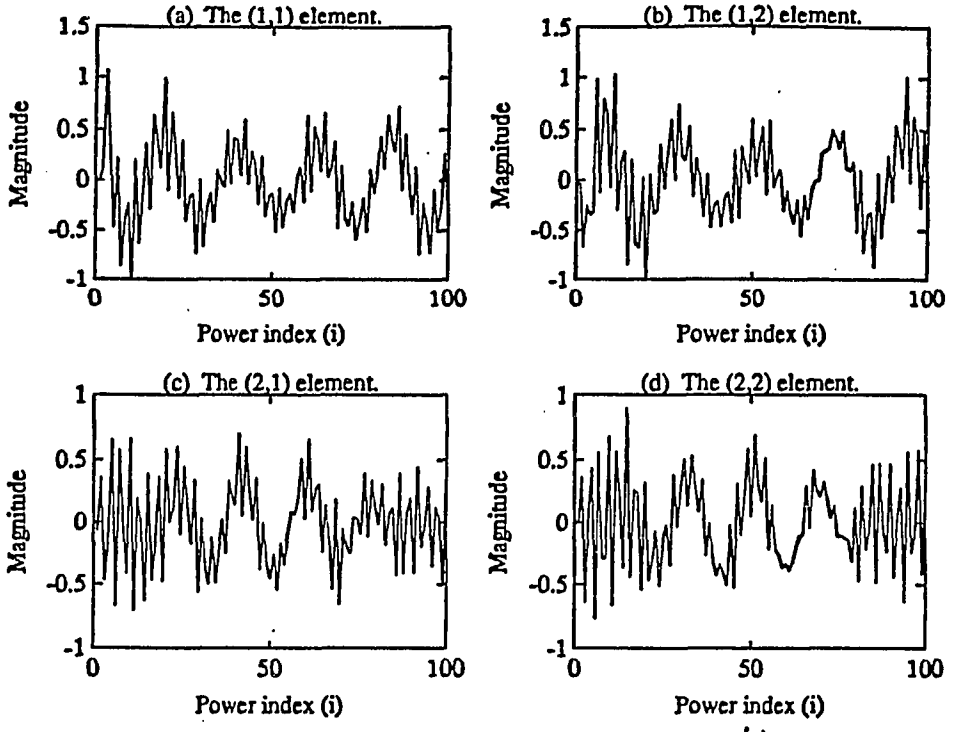


Fig. 7.2 Identified system Markov parameters $CA^{i-1}B$.

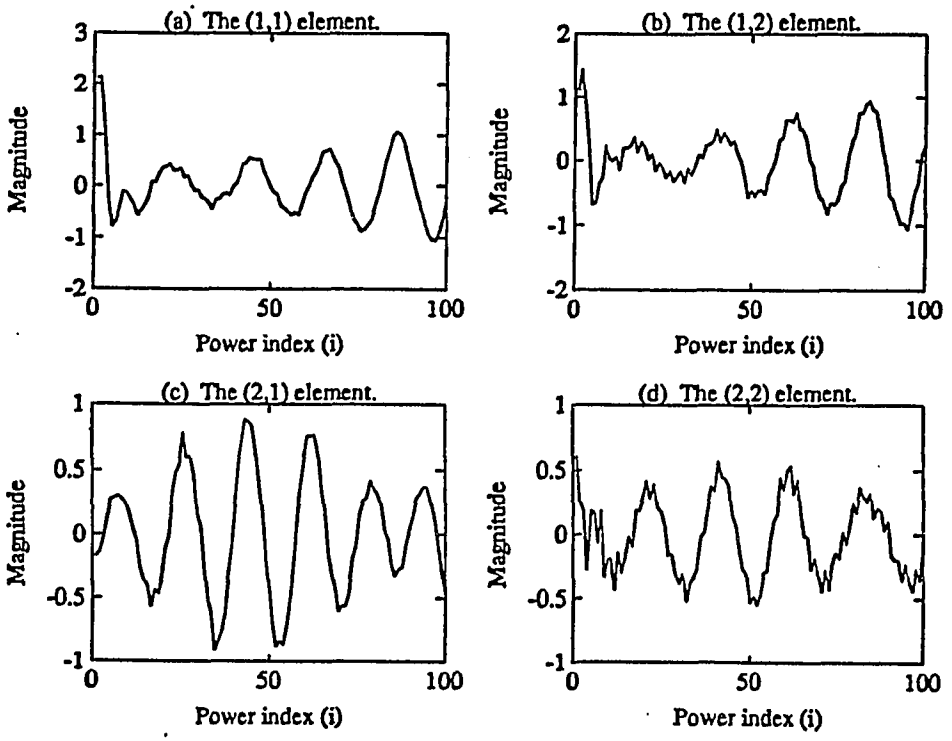


Fig. 7.3 Identified MA parameters $CA^{i}K$.

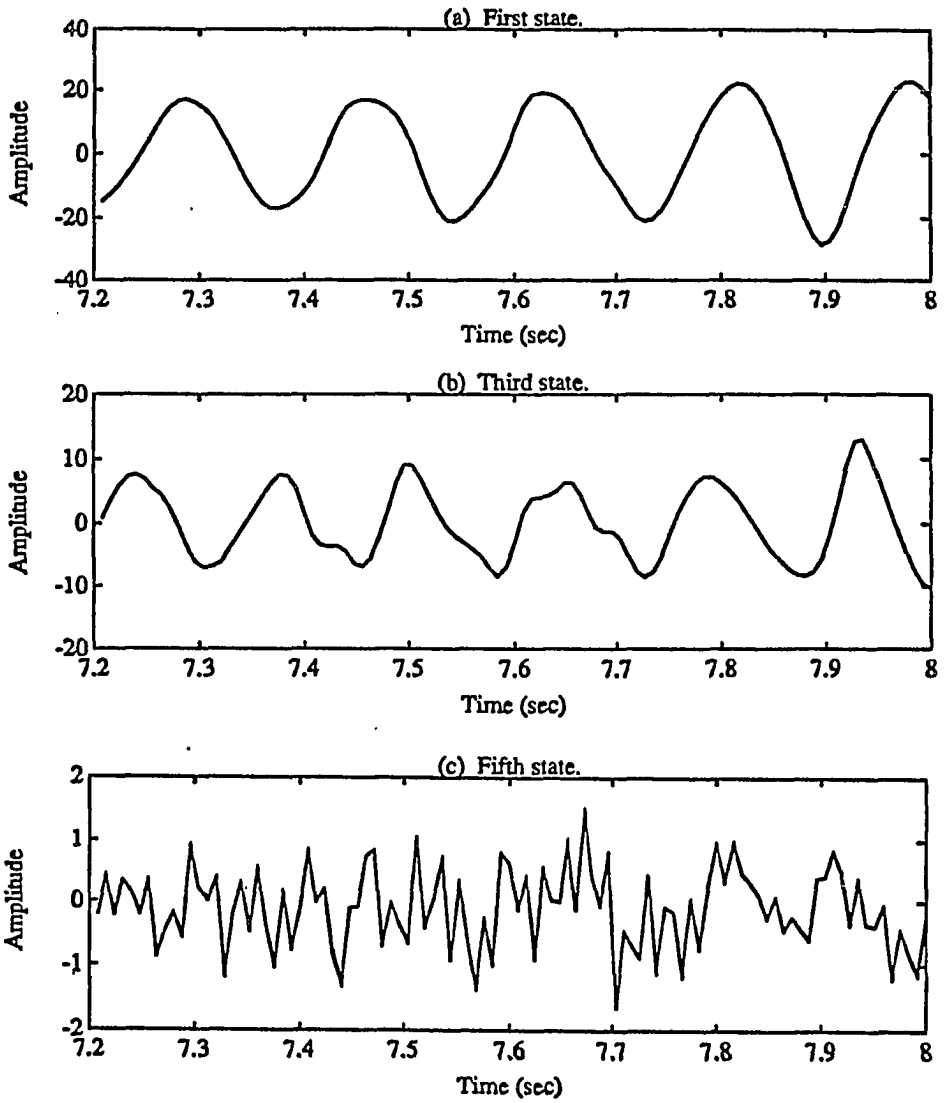


Fig. 7.4 Estimated modal states using 3,750 data.

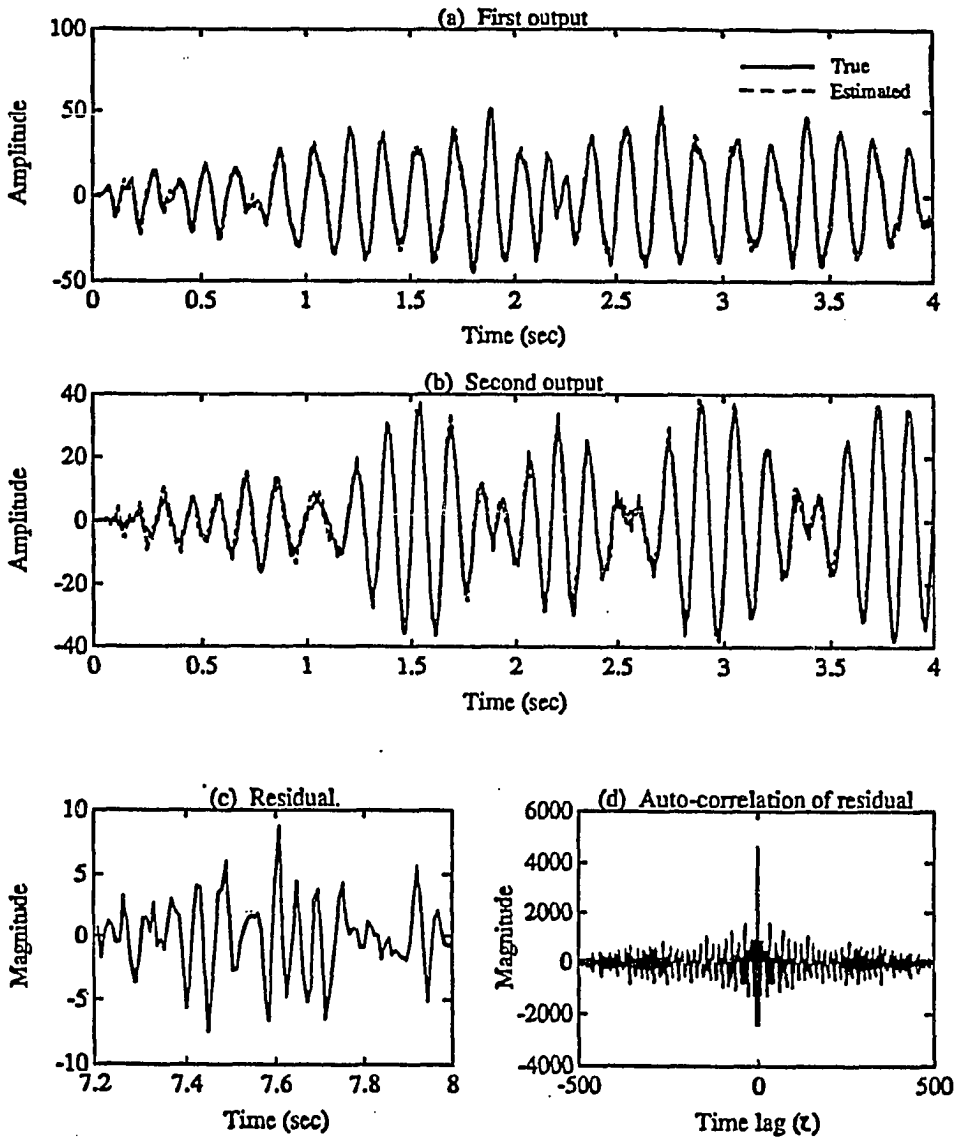
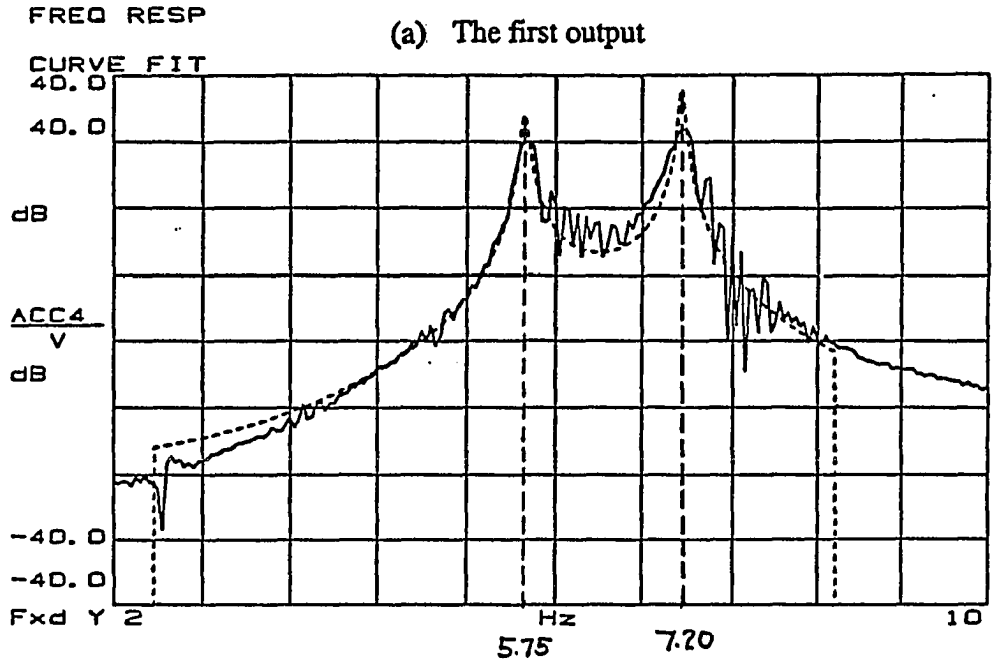
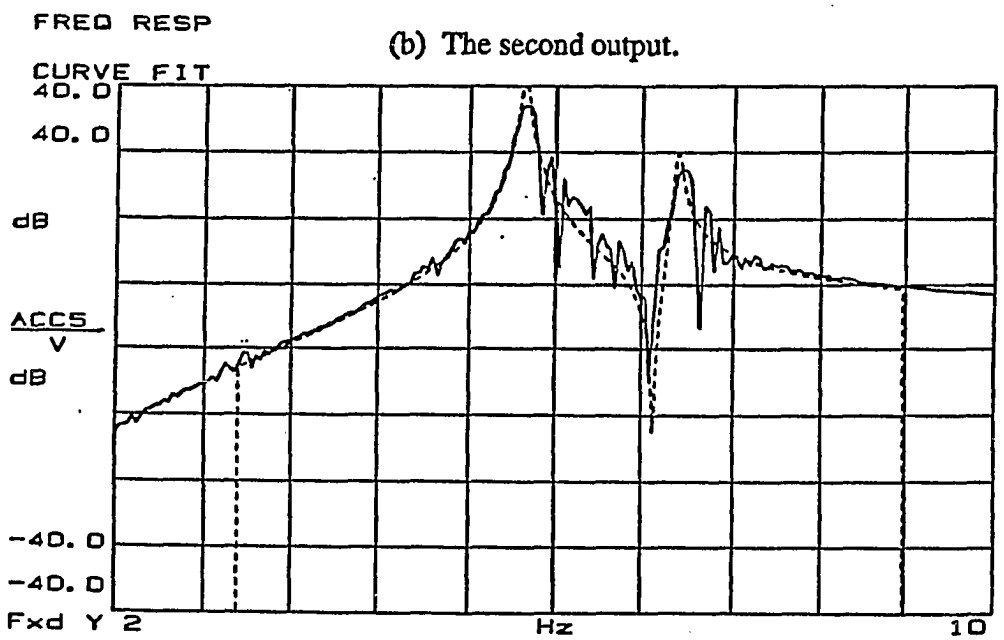


Fig. 7.5 Estimation of outputs using the sequential method with 3,750 data samples.



(a) The first output



(b) The second output.

Fig. 7.6 Frequency response functions using frequency-domain analysis.

Chapter 8

CONCLUSIONS

8.1 Review

We have systematically presented here some effective methods for estimating state information of linear, discrete-time, dynamical systems under three different situations. The situations are classified according to the amount of a priori knowledge one has about the systems, ranging from knowing both system state space model and noise covariances to having only input/output data. Areas covered include parameter estimation, signal processing, recursive estimation, adaptive Kalman filtering and system identification. Through all the methods, least-squares and its variations are the heart of the techniques.

For state estimation with sufficient a priori knowledge, we write the current state and the data used for estimation in linear equations and use linear parameter estimation theories to solve them. As a result, we derive the projection filter, which is useful in the later studies. The derivation of the projection filter also serves as

background material for understanding the link between Kalman filter and classic estimation theories.

For state estimation under unknown noise covariances, we have two different approaches; one is based on the same linear equations as the previous case but uses recursive least-squares to solve state information directly; the other estimated the optimal steady state Kalman filter gain first, and used the Kalman filter along with the estimated gain to perform state estimation. One method is derived for the former approach and three for the latter. The latter approach is also referred to as adaptive Kalman filtering. It turns out that estimating Kalman filter gain is an effective way of performing adaptive Kalman filtering.

For state estimation under unknown system model and unknown noise covariances, we are involved in a compound problem of system identification and state estimation. Two methods were derived to solve the problem: one identifies a state space model and the corresponding Kalman filter gain simultaneously; the other identifies a state space model first, and then use this identified model to estimate the Kalman filter gain. The former method utilizes the relation between the Kalman filter and a matrix polynomial model, while the latter uses the relation between the projection filter and a matrix polynomial model. For identification of a system model and the corresponding Kalman filter gain, the Markov parameters of the system and of the Kalman filter play important roles because they are unique with respect to the locations of input actuators and output sensors regardless of the dimension of the system. Obtaining Markov parameters is equivalent to obtaining system model, because they can be decomposed into state space parameters through the eigensystem realization algorithm.

With the objective of applying the derived methods for control of flexible space structures in mind, the derivations have been focusing on time-domain, recursive approaches, which are suitable for on-board adaptive applications. As a result, ordinary least-squares, recursive least-squares, and recursive weighting

least-squares are used as the fundamental techniques in most of the methods derived. The great advantage of using least-squares techniques is their capability of recursive operation in time-domain and their adaptability. In this dissertation we highly exploit this advantage and successfully derive a state space model and the corresponding Kalman filter gain from a matrix polynomial model. The latter is the natural model for using least-squares techniques to identify a mathematical model for a linear system, while the former is required for designing controller based on modern control theories. The success of deriving a state space model and the corresponding Kalman filter gain from a matrix polynomial model has provided effective integrated methods for adaptive state estimation and adaptive system identification. Moreover, it also provides a promising way towards adaptive control. This is the main contribution of this dissertation.

8.2 Further Extension of the Research

The theories derived in this dissertation can be extended further. One natural extension is to investigate its application in adaptive control, because it is well-known that control and observation are dual problems. The theory useful in state estimation (observation) should be useful in control also. Another possible extension is to look for computationally more effective ways in least-squares system identification by investigating the relation between a state space model and some fast least-squares filters, such as fast transversal filter and lattice filter,[16] because the computational load of the ordinary least-squares method increases rapidly as the order of the matrix polynomial model and/or the numbers of input and output increase. This drawback restricts on-line applications of the methods derived in this dissertation. Nevertheless, the ordinary least-squares method has been modified for faster operation in the field of adaptive filtering[16]. The relation between these fast filters and a state space model has not been clearly established

yet. Because both of them can provide optimal output predictions of a system, they must be somehow related to each other. A method which can obtain a state space model from those fast filters will be very useful in practice because of the fast operation feature. It is promising and worthwhile to explore this problem.

REFERENCES

1. Sorenson, H. W., "Least-squares Estimation: from Gauss to Kalman," *IEEE Spectrum*, Vol. 7, pp. 63-68, July 1970.
2. Kailath, T., "A View of Three Decades of Linear Filtering Theory," *IEEE Trans. Inform. Theory*, Vol. IT-20, 1974, pp. 146-181.
3. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problem," *Trans. ASME Journal of Basic Engineering*, Vol. 82D, pp. 35-45, March, 1960.
4. Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory." *Trans. ASME J. Basic Eng.*, Vol. 83, pp. 95-108, 1961.
5. Sorenson, H. W., *Kalman Filter: Theory and Application,* IEEE, Inc., 1985, New York.
6. Schlee, F. H., Standish, C. J., and Toda, N. F., " Divergence in the Kalman Filter," *AIAA J.*, Vol. 5, pp. 1114-1120, June 1976.
7. Fitzgerald, "Divergence of the Kalman Filter," *IEEE Trans. Automatic Control*, Dec. 1971.
8. Sangsuk-Iam and Bullock, T. E., "Anaysis of Continuous-time Kalman Filtering Under Incorrect Noise Convariances," *Automatica*, Vol. 24, No. 5, pp. 659-669, 1988.
9. Sangsuk-Iam and Bullock, T. E., "Behavior of the Discrete-time Kalman Filter Under Incorrect Noise Convariances," *Proc. 26th IEEE Conf. on Decision and Control*, Los Angels, California.
10. Lee. T. T., "A Direct Approach to Identify the Noise Covariances of Kalman Filtering," *IEEE Trans. Automatic Control*, AC-25, pp. 841-842. 1980.
11. Belanger, P. R., "Estimation of Noise Covariance Matrices for a Linear Time-varying Stochastic Process," *Automatica*, 10, pp. 267-275.
12. Sangsuk-Iam and Bullock, T. E., "Direct Estimation of Noise Covariances," *Proc of ACC conf.*, pp. 1289-1294, 1988, Atlanta, Geogia.
13. Carew, B., Belanger, P. R., "Identification of Optimal Filter Steady-state Gain for Systems with Unknown Noise Covariances," *IEEE Trans. Automatic Control*, AC-18, No. 6, pp. 582-587, Dec. 1973.
14. Tajima, K., "Estimation of Steady-state Kalman Filter Gain," *IEEE Trans. Automatic Control*, AC-23, pp. 944-945, Oct. 1978.
15. Goodwin, G. C., and Sin K. S., *Adaptive Filtering, Prediction and Control*, Prentice-Hall, Englewood Cliffs, New Jersey 07632, 1984.

16. Haykin, S., *Adaptive Filter Theory,* Prentice-Hall, Englewood Cliffs, New Jersey 07632, 1986.

17. Mehra, R. K. "Approach to Adaptive Filtering," *IEEE Trans. Automatic Control*, AC-17, pp. 693-698, Oct. 1972.

18. Mehra, R. K. "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Trans. Automatic Control*, AC-15, pp. 175-184, Apr. 1970.

19. Eykhoff, P., *System Identification - Parameter and State Estimation*, A Wiley-Interscience Publication, 1974.

20. Sage, A. P. and Melsa, J. L., *System Identification,* Academic Press, New York, 1971.

21. Hsia, T. C., *System Identification - Least-Squares Methods,* D. C. Heath and Company Lexington, Massachusetts Toronto, 1977.

22. Ljung, L. and Soderstrom, T., *Theory and Practice of Recursive Identification,* The M.I.T Press, 1983.

23. Ljung, L., *System Identification - Theory for the User,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

24. Sophocles J. Orfanidis, *Optimum Signal Processing - An Introduction*, Macmillan Pub. Co. London, 1985.

25. Juang, J.-N., and Pappa, R. S., "An Eigensystem Realization Algorithm for Modal Parameter Identification and Model Reduction," *Journal of Guidance, Control, and Dynamics,* Vol. 8, Sept.-Oct. 1985, pp. 620-627.

26. Lewis, F. L., *Optimal Estimation - with an Introduction to Stochastic Control Theory*, John Wiley & Sons, New York, 1986.

27. Elbert, T. F. *Estimation and Control of Systems*, Van Nostrand Reinhold Co., New York, 1984.

28. Gelb, A., *Applied Optimal Estimation*, The M.I.T. Press, Cambridge, Massachusetts, and London, England, 1974.

29. Deutsch, R., *Estimation Theory*, Pretice Hall, Englewood Cliffs, N.J., 1965.

30. Kailath, T., "An Innovations Approach to Least-squares Estimation – Part I: Linear Filtering in Additive White Noise," *IEEE Trans. Automatic Control*, AC-13, pp. 646-655, Dec. 1968.

31. Mayne, D. Q., Firoozan, F., "Linear Identification of ARMA Processes," *Automatica*, Vol. 18. No. 4, pp. 461-466, 1982.

32. Akaike, H., "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models," *Biometrika*, Vol. 60, pp. 255-265, 1973.

33. Akaike, H., "A New Look at the Statistical Model Identification," *IEEE Trans. Automatic Control*, AC-19, pp. 716-723, 1974.

34. Sundararajan, N, and Montgomery, R. C., "Identification of Structural Dynamics Systems Using Least-Square Lattice Filters," *Journal of Guidance, Control, and Dynamics,* Vol. 6, Sept.-Oct. 1983, pp. 374-381.

35. Strejc, V., "Trends in Identification" *International Federation of Automatic Control*, pp. 7-12, 1981.

36. Fasol, K. H. and Jorgl, H. P., "Principles of Model Building and Identification," *Automatica*, Vol. 16, pp. 505-518, 1980.

37. Juang, J.-N., and Pappa, R. S., "A Comparative Overview of Modal Testing and System Identification for Control of Structures," *The Shock and Vibration Digest*, Vol. 20, No. 6, June 1988, pp. 4-15.

38. Juang, J.-N., "Mathematical Correlation of Modal Parameter Identification Methods Via System Realization Theory," *International Journal of Analytical and Experimental Modal Analysis*, Vol. 2. Jan. 1987, pp. 1-18.

39. Juang, J.-N. and Pappa, R. S., "Effects of Noise on Modal Parameters Identified by the Eigensystem Realization Algorithm," *Journal of Guidance, Control, and Dynamics,* Vol. 9, May-June 1986, pp. 620-627.

40. Juang, J.-N. and Suzuki, H., "An Eigensystem Realization Algorithm in Frequency Domain for Modal Parameter Identification," *proceedings of the AIAA Guidance, Navigation and Control Conference*, New York, 1986.

41. Pappa, R. S. and Juang J.-N., "Some Experiences With the Eigensystem Realization Algorithm," *Sound and Vibration*, January 1988.

42. Huang, J.-K., Juang, J.-N. and Chen, C.-W., "Single Mode Projection FIlter for Modal Parameter Ientification," *AIAA J. of Guidance, Control and Dynamics,* Vol. 12, No. 4, July-August 1989, pp. 568-576.

43. Meirovitch, L., and Baruh, H., "The Implementation of Modal Filters for Control of Structures," *AIAA J. of Guidance, Control and Dynamics,* Vol. 8, No. 6, Nov-Dec 1985, pp. 707-716.

44. Pappa, R. S., and Schenk, A., "ERA Modal Identification Experiences with Mini-Mast," Presented at *the 2nd USAF/NASA Workshop on System Identification and Health Monitoring of Precision Space Structures*, Pasadena, CA, March 27-29, 1990.

45. Maybeck, P. S., "Stochastic Models, Estimations, and Control" Volumn 1-3, Academic Press Inc. New York, 1982.

46. Chen, C.-W., Huang, J.-K. "Adaptive state estimation for control of flexible structures," *Proceedings of SPIE's Conference on Controls for Optical Systems*, Orlando, Florida, April, 1990, pp. 252-263.

47. Chen, C.-W., Huang, J.-K., Phan, M. and Juang, J.-N., "Integrated System Identification and Modal State Estimation for Control of Flexible Space Structures," *Proceedings of AIAA Guidance, Navigation and Control Conference*, Portland, Oregon, August, 1990, pp. 1396-1404.

48. Chen, C.-W., Huang, J.-K., "State Estimation under Unknown Noises – A Least-Squares Approach" *Proceedings of the 5th International Conference on CAD/CAM Robotics & Factories of the Future*, Norfolk, Virginia, December, 1990.

49. Chen, C.-W., Huang, J.-K., Phan, M. and Juang, J.-N., "Integrated System Identification and Modal State Estimation for Control of Flexible Space Structures," to appear in *Journal of Guidance, Control and Dynamics.*

50. Chen, C.-W., Huang, J.-K. "Estimation of Optimal Kalman Filter Gain from Non-optimal Filter Residuals," *Proceedings of SPIE's Conference on Controls for Optical Systems*, Orlando, Florida, April, 1991.

APPENDIX

Any two equivalent systems, $[A_1,\ B_1,\ C_1]$ and $[A_2,\ B_2,\ C_2]$, where both A_1 and A_2 have distinct eigenvectors, can be transformed to the same triplet $[A_n,\ B_n,\ C_n]$ by a certain equivalent transformation. The proof is given in the following:

Proof:

After diagonalizing the system matrix and normalizing the output or input matrix respectively, $[A_1,\ B_1,\ C_1]$ and $[A_2,\ B_2,\ C_2]$ become $[\Lambda_1,\ B_{n1},\ C_{n1}]$ and $[\Lambda_2,\ B_{n2},\ C_{n2}]$, where

$$\Lambda_1 = D_1^{-1}A_1D_1,\quad B_{n1} = D_1^{-1}B_1,\quad C_{n1} = C_1D_1 \tag{A1}$$

$$\Lambda_2 = D_2^{-1}A_2D_2,\quad B_{n2} = D_2^{-1}B_2,\quad C_{n2} = C_2D_2 \tag{A2}$$

with D_1 and D_2 being the equivalent transformation matrices for the corresponding set.

Since $[A_1,\ B_1,\ C_1]$ and $[A_2,\ B_2,\ C_2]$ are equivalent, there also exists a nonsingular matrix P such that

$$A_1 = P^{-1}A_2P,\quad B_1 = P^{-1}B_2,\quad C_1 = C_2P. \tag{A3}$$

The similarity transformation will not change the eigenvalues; therefore, A_1 and A_2 have the same eigenvalues. This means that Λ_1 and Λ_2 are identical,

assuming that the eigenvalues have been sorted in the same order. Hence, from (A1)-(A3) one can have

$$\begin{aligned} \Lambda_1 &= D_1^{-1} A_1 D_1 = D_1^{-1} P^{-1} A_2 P D_1 \\ &= \Lambda_2 = D_2^{-1} A_2 D_2 = A_n. \end{aligned} \tag{A4}$$

From Eq. (A4), because both PD_1 and D_2 diagonalize A_2, the following relation must hold:

$$PD_1 = D_2 K_c \tag{A5}$$

where K_c is some non-singular constant diagonal matrix. Therefore, from Eqs. (A1)-(A5) one obtains

$$\begin{aligned} C_{n1} &= C_1 D_1 = C_1 P^{-1} D_2 K_c = C_2 P P^{-1} D_2 K_c \\ &= C_{n2} D_2^{-1} P P^{-1} D_2 K_c = C_{n2} K_c. \end{aligned} \tag{A6}$$

Since K_c is a constant diagonal matrix, C_{n1} and C_{n2} are scaled versions to each other. However, since both matrices have been normalized in the same way, the only possible solution for K is the identity matrix, which means $C_{n1} = C_{n2}$. Similarly, B_{n1} and B_{n2} can be proved to be identical. As a result, the unique set $[A_n, B_n, C_n]$ is obtained.