WIDE-SENSE MARTINGALE APPROACH TO LINEAR DISCRETE-TIME OPTIMAL ESTIMATION

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY
HALİT KARA

1971

This is to certify that the

thesis entitled

Wide-Sense Martingale Approach to Linear
Discrete-Time Optimal Estimation

presented by

Halit Kara

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Systems Science

Gerald J Park

Major professor

Date May 19, 1971

O-7639

WIDE-SENSE MARTINGALE APPROACH TO LINEAR DISCRETE-TIME OPTIMAL ESTIMATION

By

Halit Kara

AN ABSTRACT OF A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering and System Science

1971

ABSTRACT

WIDE-SENSE MARTINGALE APPROACH TO LINEAR DISCRETE-TIME OPTIMAL ESTIMATION

By

Halit Kara

The dissertation considers the minimum mean-square error estimation of the signal $x_k = \Phi(k)u_k$ where $\Phi(k)$ is an $n \times n$ matrix and u_k is a wide-sense martingale process. The optimal estimation equations are derived for prediction, filtering and smoothing based on noisy observations.

Along with the statement of the problem, the historical and mathematical background upon which the derivations of the optimal estimation equations are based is presented. The general formulas for the optimal estimation equations for a second-order discrete-time stochastic process are derived assuming that the observation process has full rank. Then, the recursive and algebraic estimation equations are derived for the signal when the observations are corrupted by additive white, cross-correlated white and cross-correlated colored noises. The recursive nature of these equations follows easily from wide-sense martingale property of u_k.

The thesis gives a purely orthogonal projection approach in solving the prediction, filtering and smoothing problems of Kalman and their extensions, for a more general class of signals.

The main object is to remove unnecessary analytic complications introduced by stochastic difference equations and to use a conceptually simpler geometric approach which provides a unified attack in all three cases (uncorrelated white, cross-correlated white and cross-correlated colored observation noises). However, at the same time, analytic solutions which can be studied numerically are obtained.

WIDE-SENSE MARTINGALE APPROACH TO LINEAR DISCRETE-TIME OPTIMAL ESTIMATION

By

Halit Kara

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering and System Science

1971

To my professors M.N.Özdaş and T. Özker whose encouragement and insistence made it all possible.

ACKNOWLEDGEMENTS

The constant help, encouragement, and guidance given by Drs. R.O. Barr and G.L. Park have made this work possible. Besides the help extended on the actual writing of the thesis, Dr. Park is responsible for procuring the financial aid without which nothing could have been achieved. The excellent academic background, whether it be through teaching or advising, provided by Dr. Barr; and his friendly and encouraging disposition have also had a great influence on the initiation and fruition of this work.

The problem was initially brought to the author's attention by Dr. V. Mandrekar whose suggestions and assistance throughout the writing of the dissertation needs special mention.

Thanks are also due to Drs. I. Korn and H. Salehi for their critical reviews and helpful suggestions.

The author would also like to take this opportunity to acknowledge the unselfish interest and encouragement extended by Dr. M.Z. Akcasu and his wife Melahat of the University of Michigan, and to Dr. P.K. Wong for providing him with mathematical background.

Support was granted by Consumer's Power Company Cooperative Research at Michigan State University and the Turkish Sugar Corporation. This support is gratefully adknowledged.

Last but, definitely not least, the author wishes to give special mention and thanks for the love, patience, and understanding put forth by his wife, Ayten, and his children, Tarik and Aysen.

TABLE OF CONTENTS

LIST OF FIGURES

GENERAL NOTATION

1. The discrete-time is denoted by $i,j,k,\ell,\ldots,s$
2. Vectors are denoted by small letters, such as u, v, w, x, y and z. The transpose of a vector is denoted by superscript T, for example x^T denotes the transpose of the vector x.
3. Matrices are denoted by capital letters, such as D, F, G, H,...,Φ. The transpose and trace of a matrix are denoted by superscript T and by tr respectively.
4. The symbols o denotes the scalar zero, or the null vector, or the null matrix, depending on the context.
5. The proof of a theorem will be introduced by the word PROOF and terminated by the abbreviation QED. If the proof is omitted the statement of the theorem will be terminated by the symbol .

LIST OF SYMBOLS

Symbol	Usage	Meaning	Reference
$\stackrel{d}{=}$	$\ldots \stackrel{d}{=} -$	... by definition equal to	Ch. 1
$\in$	$k \in Z$	k is an element of Z	Ch. 1
$\notin$	$k \notin Z$	k is not an element of Z	Ch. 2
$\subset$	$A \subset B$	A is a subset of B	Ch. 2
$\ni$	$\ldots \ni -$	... such that -	Ch. 2
$\forall$	$\ldots \forall -$	... for each (for all) -	Ch. 2
$\hat{}$	$\hat{x}_{k\|\ell}$	Optimal estimate x_k based on the observation $Y(\ell)$	Ch. 1
$\tilde{}$	$\tilde{x}_{k\|\ell}$	$\tilde{x}_{k\|\ell} \stackrel{d}{=} x_k - \hat{x}_{k\|\ell}$, estimation error	Ch. 1
$\{ \mid \}$	$\{x \mid -\}$	All x such that -	Ch. 2
$\|\cdot\|$	$\|\alpha\|$	Absolute value of α	Ch. 2
	$\|x\|$	Euclidean norm of vector x	Ch. 2
$< , >_L$	$<x,y>_L$	Inner product of x,y in L_2	Ch. 2
$< , >$	$<x,y>$	Inner product of x,y in L_2^n	Ch. 2
$[,]$	$[x,y]$	Gramian matrix of x,y	Ch. 2
$\| \cdot \|$	$\|x\|$	Norm of vector x induced by <,>	Ch. 2
$\Rightarrow$	$\ldots \Rightarrow -$	... implies -	Ch. 2
$\rightarrow$	$X \rightarrow Y$	Function on X to Y	Ch. 2
$\mapsto$	$w \mapsto x(w)$	Function assigning x(w) to w	Ch. 2
$\oplus$	$A \oplus B$	Direct sum of A, B	Ch. 2

$\otimes$	$A \otimes B$	Direct product of A, B	Ch. 2
$\perp$	$x \perp y$	x, y are orthogonal	Ch. 2
	$x \perp M$	x is orthogonal to subspace M	Ch. 2
$\perp$	$A^{\perp}$	Orthogonal complement of A	Ch. 2
$+$	P^{+}	Generalized inverse of P	Ch. 2
$(\mid)$	$(x\mid M)$	Orthogonal projection of x onto M	Ch. 2
$\wedge$	$i \wedge j$	$i \wedge j \stackrel{d}{=} \min\{i,j\}$	Ch. 2
$>$	$x > y$	y is less than of x	Ch. 1
	$P > 0$	P is positive definite	Ch. 1
$\geq$	$x \geq y$	y is less than or equal to x	Ch..1
	$P \geq 0$	P is positive semidefinite	Ch. 1

Abbreviations	Meaning	Reference
BP	Basic problem	Ch. 3
CCP	Cross-correlated noise problem	Ch. 4
CNP	Colored-noise problem	Ch. 5
iff	if and only if	Ch. 2
OPE	Orthogonal projection estimator	Ch. 2
H-space	Hilbert space	Ch. 2
$L_2(L_2(\Omega,\mathcal{A},p))$	Hilbert space of square integrable random variables (on the probability space $(\Omega,\mathcal{A},p)$)	Ch. 2
$L_2^n(L_2^n(\Omega,\mathcal{A},p))$	Hilbert space of square integrable random n-vector (on the probability space $(\Omega,\mathcal{A},p)$)	Ch. 2
(mod P)	With probability one	Ch. 2

P_M	$P_M \stackrel{d}{=} (\cdot \mid M)$ orthogonal projection operator onto subspace M	Ch. 2
r.v.	Random variable	
Z	$Z \stackrel{d}{=} \{\ldots -1,0,1,\ldots\}$, the set of all integers	Ch. 1
δ_{kj}	Kronecker delta	Ch. 1
I_n	n × n identity matrix	Ch. 3
R^n	n-dimensional Euclidean space over reals	Ch. 2
$Y(\ell)$	Observation record $Y(\ell) \stackrel{d}{=} \{y_1, y_2, \ldots, y_\ell\}$	Ch. 1

CHAPTER 1

INTRODUCTION

The optimal estimation problem is encountered under different forms in many branches of science as well as in a variety of engineering disciplines. The discrete-time linear estimation problem which is an important special case of the general problem is the topic of this study. It can be described quite generally in simple terms with reference to the block diagram in Figure 1.1. In this block diagram x_k and z_k denote, respectively, the input and output signals of a memoryless, non-random linear transformation, H(k), so that

$$z_k = H(k)x_k \ .$$

The output is observed in a noisy environment which is assumed to be an additive random signal v_k, called the output noise (or measurement noise or observation noise). Thus, the (actually) observed signal y_k can be represented as

$$y_k = z_k + v_k \ , \tag{1.1}$$

where the subscript refers to discrete-time, i.e., $k \in Z = \{\ldots,-1,0,1,2,\ldots\}$, the set of all integers.

It is assumed that the observations are available over a set of integers $\{k_1,k_1+1,\ldots,\ell\}$ where k_1 is an arbitrary

starting time (for the sake of simplicity k_1 is chosen to be unity) and ℓ moves along in discrete-time as additional data are recorded. The problem can now be stated as follows:

(1.2) OPTIMAL DISCRETE-TIME LINEAR ESTIMATION PROBLEM. Given:

(a) The relationship between x_k and y_k, $k \in Z$, i.e. $H(k)$, $k \in Z$ and (1.1).

(b) The means and covariance matrices of the stochastic processes $\{x_k, k \in Z\}$ and $\{v_k, k \in Z\}$.

Problem: Given an observation record (data) $Y(\ell) \stackrel{d}{=} \{y_1, y_2, \ldots, y_\ell\}$; find an optimal realizable estimate $\hat{x}_{k|\ell}$ of the signal x_k which is a linear function of the data $y_1, y_2, \ldots, y_\ell$, i.e.

$$\hat{x}_{k|\ell} \stackrel{d}{=} \sum_{i=1}^{\ell} A(i) y_i \quad 1)$$

for $k \geq 0$, where $A(i)$, $i = 1,2,\ldots,\ell$ are matrices in appropriate dimension. For $k > \ell$ the problem is called <u>prediction</u>, for $k = \ell$ <u>filtering</u>, and for $k < \ell$ <u>smoothing</u>.

The terms <u>optimal</u> and <u>realizable</u> that occur in the description of the problem are defined as follows:

OPTIMAL. The estimate $\hat{x}_{k|\ell}$ of x_k is optimal if it satisfies some specified criterion of optimality. The criterion used in this dissertation is the minimum mean-square error, i.e., the minimization of the mean-square error risk function (or performance measure):

$$(1.3) \qquad J(\tilde{x}_{k|\ell}) \stackrel{d}{=} \mathcal{E}\{(x_k - \hat{x}_{k|\ell})^T (x_k - \hat{x}_{k|\ell})\} = \operatorname{tr} \mathcal{E}\{(x_k - \hat{x}_{k|\ell})(x_k - \hat{x}_{k|\ell})^T\} \quad 2)$$

---

1) $- \stackrel{d}{=} \ldots$ means, by definition - equals .....

2) tr A denotes the trace of the matrix A.

where T denotes the transpose of a vector (or matrix), $\tilde{x}_{k|\ell} \overset{d}{=} x_k - \hat{x}_{k|\ell}$ is the estimation error and $\mathcal{E}\{\cdot\}$ denotes the expectation operator. This criterion is not unduly restrictive

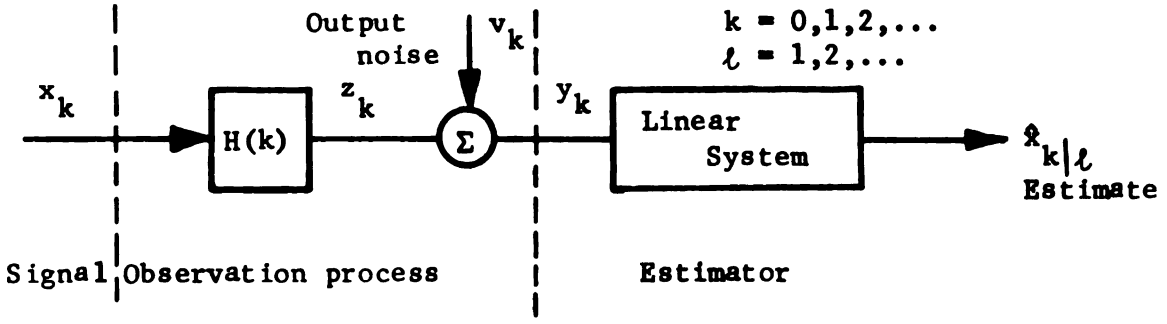


Figure 1.1. Block diagram for linear estimation problem (1.2).

because the estimate that is optimal for the minimum mean-square error criteria (hereafter will be called "minimum mean-square estimate") is often optimal for other criteria as well (cf. [B-4], [D-1], [K-4], [S-3], [Z-4]).

REALIZABILITY. The realizability of the estimate $\hat{x}_{k|\ell}$ of x_k means that the estimate depends only on present and past data $y_1, y_2, \ldots, y_\ell$; but not on future data $y_{\ell+1}, y_{\ell+2}, \ldots$. The estimate $\hat{x}_{k|\ell}$ can therefore be generated in discrete-time as the output of a physical system called the estimator (cf. Figure 1.1). For $k > \ell$ the estimator is called a predictor, for $k = \ell$ a filter, and for $k < \ell$ a smoother.

The purpose of this research is to study Problem (1.2) where the signal is an n-variate, discrete-time, second-order stochastic process $\{x_k = \Phi(k)u_k,\ k = 0,1,2,\ldots\}$ where $\forall k = 0,1,2,\ldots, \Phi(k)$ is a non-random $n \times n$ matrix and u_k, $k = 0,1,2,\ldots$ is a wide-sense martingale process[1]. The major

---

1) The idea of approaching to optimal linear estimation problem from a wide-sense martingale process approach was first suggested to the author by V. Mandrekar.

contribution of the present work lies in the characterization of the signal x_k as a linear transformation of a wide-sense martingale rather then as a solution of a given difference equation. This characterization leads directly to recursive estimates and also allows direct alternative derivations and extensions of earlier well-known results in discrete-time linear estimation (cf. [B-5], [B-6], [C-1], [K-3], [K-4]).

1.1 HISTORICAL BACKGROUND AND LITERATURE SURVEY

The problem of estimating a stochastic signal from a noisy observation record has been intensively studied since the appearance, during the early 1940's, of the classical work of A.N. Kolmogorov [K-9] and N. Wiener [W-1]. Kolmogorov studied only discrete-time stationary processes and solved the problem of linear estimation of such processes using a technique which was based on the time-domain recursive orthogonalization of the observed data. This technique was suggested by Wold [W-6] in his doctoral dissertation in 1938, and hence is known as the Wold decomposition method. Kolmogorov's theory extended to vector valued random elements by Wiener and Masani [W-3], [W-4] in 1957-58.

On the other hand, Wiener [W-1] studied the linear estimation problem of continuous-time processes and reduced it to the problem of solving a certain integral equation, the so-called "Wiener-Hopf equation". This equation was already studied by Wiener and Hopf [W-3] in 1931. It can only be solved explicitly for certain special cases of the general estimation problem. The solution involves formidable mathematics, which was beyond the

reach of most engineers at that time, even though Wiener undertook this work in response to an engineering problem. Because of its complexity, Wiener's theory did not receive proper attention and recognition for many years.

In 1950, H. Bode and C. Shannon [B-3] gave a different derivation of Wiener's results based on ideas in a report by Blackman, H. Bode, and C. Shannon [B-2]. The work of Bode and Shannon was instrumental in popularizing Wiener's theory. The same approach was independently discovered by L. Zadeh and R. Ragazzini [Z-3].

In 1950's the idea of generating linear estimates recursively was introduced. Such algorithms were used by Gauss in 1809 [G-1] in his numerical calculations of the orbit of the astereoid Ceres. But the modern interest in recursive estimation was stimulated by the increased usage of digital computers. The first modern work on this subject was done by R. Kalman [K-3], in 1960. The practicality of the Kalman approach to the estimation problem has made it immensely popular among engineers.

The paper by Kalman in 1960 introduced a different approach to the linear estimation problem of Kolmogorov and Wiener in the case of a special class of discrete-time stochastic processes. In 1961, R. Kalman and R.S. Bucy [K-6] generalized Kalman's results to continuous-time processes. The novelty of their formulation was the representation of all stochastic processes by state equations that are driven by additive white input noises rather than correlation functions. By restricting their attention to Gauss-Markov processes given by difference or differential eqautions, in

particular, they derived difference (for discrete-time) and differential (for continuous-time) equations for the filters, which are called Kalman filter and Kalman-Bucy filter, respectively. These equations can be used to construct a linear filter that is, of course, identical to the one specified by the Wiener-Hopf equation. However, there is a definite practical advantage in having a differential (or difference) equation for the estimate instead of an integral equation for the estimator. Specifically, it is much easier to solve a differential equation by analog or digital techniques than to solve an integral equation and then perform a convolution. This computational advantage of the Kalman approach to the linear estimation problem has stimulated a great number of papers, providing alternative derivations, extensions and relationships to classical parameter estimation techniques. It may be an overstatement to suggest that there are as many derivations of the Kalman filter equations as there are workers in the field. Most of these works are now in standard texts [A-1], [A-2], [B-7], [B-9], [D-1], [J-1], [L-1], [L-2], [N-1], [S-1], [S-2].

1.2 STATEMENT OF THE PROBLEM

A precise statement of the problem considered in this study will now be presented. We first note that, in the last decade, all the work on the Kalman filtering theory assumes that the signal is generated by a given linear stochastic difference equation driven by a white-noise process. It is known that such a stochastic difference equation generates a wide-sense Markov process (and it always can be written as $x_k = \Phi(k)u_k$ where $\Phi(k)$ is an invertible

matrix and u_k is a wide-sense martingale process (cf. [M-1], [M-2] and see also Chapter 2, Section 2). This remark motivates us to the following problem which is characterized by the assumptions on the signal x_k, $k = 0,1,2,\ldots$ and output noise v_k, $k = 0,1,2,\ldots$ in the general problem (1.2) described in Figure 1.1.

(1.4) SIGNAL. The signal x_k, $k = 0,1,2,\ldots$ is assumed to be an n-variate, second order stochastic process in discrete-time, and given by

$$x_k = \Phi(k)u_k \qquad k = 0,1,2,\ldots$$

where $k = 0,1,2,\ldots,\Phi(k)$ is an $n \times n$ matrix of known functions of discrete-time, and u_k, $k = 0,1,2,\ldots$ is a wide-sense martingale (see Section 2 of Chapter 2) with zero mean and known $n \times n$ positive semi-definite covariance matrix sequence $\{P_u(k), k = 0,1,2,\ldots\}$ where (cf. Section 2 of Chapter 2)

$$P_u(k) \stackrel{d}{=} \mathcal{E}\{u_k\, u_i^T\}$$

for $k \leq i$, $k,i = 0,1,2,\ldots$. As a notational convenience, for any integer k, the definition:

$$Q(k) \stackrel{d}{=} \mathcal{E}\{(u_{k+1} - u_k)(u_{k+1} - u_k)^T\}$$

$$= P_u(k+1) - P_u(k)$$

is made.

In addition, without loss of generality, it is assumed that the initial value u_o is orthogonal (cf. Section 1 of Chapter 2) to the output noise. Notice that, if not, define $\bar{u}_k = u_k - u_o$;

then $\bar{u}_o = 0$ and therefore it is orthogonal to any output noise. So that this assumption is not a restriction; it is just a notational convenience.

(1.5) OBSERVATIONS. The observations are corrupted by additive noise such that

$$y_k = H(k)x_k + v_k$$
$$= M(k)u_k + v_k$$

for $k = 1,2,...$ (thus the starting time is $k = 1$) where

$y_k \in R^m$ $(m \leq n)$, observation vector,

$v_k \in R^m$, output noise vector .

$H(k)$ is an $m \times n$ matrix of known functions of discrete-time and $M(k) \stackrel{d}{=} H(k)\Phi(k)$, $k = 0,1,2,...$. Note that, for $k = 0$, $M(0) = H(0) = 0$ since at the time $k = 0$, there is no observation.

The following assumptions are made on the output noise v_k, $k = 0,1,2,...$ each of which leads to a problem in the estimation theory:

(A.1.1) BASIC PROBLEM. The output noise v_k, $k = 0,1,2,...$ is a zero mean white noise (cf. Section 2 of Chapter 2) and that

$$\mathcal{E}\{v_k v_j^T\} = P_v(k)\delta_{kj}$$

for $k,j = 0,1,2,...$, where $P_v(k)$ is an $m \times m$ matrix of known functions of discrete-time and it is assumed that $P_v(k) > 0$ [1)]

---

1) If P is a symmetric matrix, $P > 0$ $(P \geq 0)$ means P is positive (semi) definite.

$\forall k = 0,1,2,\ldots$, and δ_{kj} is the Kronecker delta:

$$\delta_{kj} = \begin{cases} 1 & \text{if} \quad k = j \\ 0 & \text{if} \quad k \neq j \end{cases}$$

In addition, it is assumed that the process $\{v_k,\ k = 0,1,2,\ldots\}$ and $\{u_{k+1} - u_k,\ k = 0,1,\ldots\}$ are orthogonal, i.e.,

$$\mathcal{E}\{v_k(u_{j+1} - u_j)^T\} = 0 \quad \text{if} \quad j \neq k \ .$$

(A.1.2) CROSS-CORRELATED NOISE PROBLEM. The process v_k, $k = 0,1,2,\ldots$ is a zero-mean white noise process as in (A.1.1), except that it is correlated with the process $\{u_{k+1} - u_k,\ k = 0,1,\ldots\}$ such that

$$\mathcal{E}\{(u_{k+1} - u_k)v_j^T\} = C(k)\delta_{kj}$$

for $k,j = 0,1,2,\ldots$ where $C(k)$ is an $n \times m$ matrix of known functions of discrete-time.

(A.1.3) COLORED NOISE PROBLEM. The output noise process is the output of a known linear system with a white noise input (see Figure 1.2):

$$v_{k+1} = \Psi(k+1,k)v_k + n_k$$

for $k = 0,1,2,\ldots$ with the initial condition v_o, where

$v_k \in R^m$, output noise,

$n_k \in R^m$, white noise,

and $\Psi(k+1,k)$ is the $m \times m$ transition matrix of known function

of discrete-time. It is assumed:

1. The initial condition v_o is a square-integrable random vector with zero mean and known $m \times m$ covariance matrix

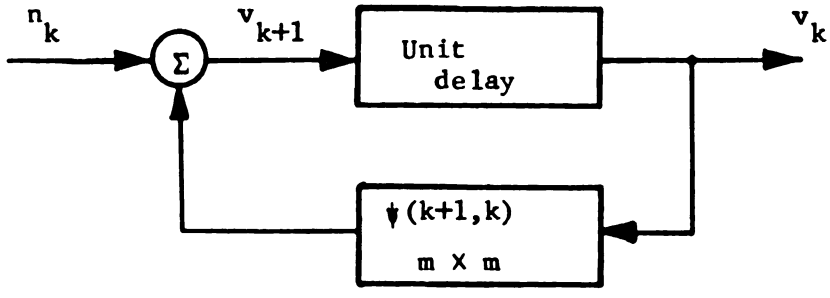


Figure 1.2. Block diagram for the output noise.

$P_v(0) \stackrel{d}{=} \mathcal{E}\{v_o\, v_o^T\}$ which is orthogonal to the processes u_k, $k = 0,1,2,\ldots$ and n_k, $k = 0,1,2,\ldots$

2. The process n_k, $k = 0,1,2,\ldots$ is a white noise process with zero mean and that

$$\mathcal{E}\{n_k\, n_j^T\} = P_n(k)\delta_{kj}$$

where $P_n(k)$ is an $m \times m$ matrix of known functions of discrete-time. In addition, the processes $\{u_{k+1} - u_k,\ k = 0,1,2,\ldots\}$ and $\{n_k,\ k = 0,1,2,\ldots\}$ are correlated such that

$$\mathcal{E}\{(u_{k+1} - u_k)n_j^T\} = C(k)\delta_{kj}$$

for $k,j = 0,1,2,\ldots$, where $C(k)$ is an $n \times m$ matrix of known functions of discrete-time.

3. The matrix

$$M(k+1)C(k) + C^T(k)M^T(k+1) + P_n(k)$$

is positive definite for all $k = 0,1,2,\ldots$.

The problem is now, under one of the assumptions (A.1.1), or (A.1.2) or (A.1.3), to find the minimum mean-square estimate of the signal $x_k = \Phi(k)u_k$ defined by (1.4) and observed via (1.5). Our main interest is to compute the optimal estimate $\hat{x}_{k|\ell}$ based on the observation record $Y(\ell)$ and the corresponding estimation error, $\tilde{x}_{k|\ell} \stackrel{d}{=} x_k - \hat{x}_{k|\ell}$, covariance matrix:

$$P_{\tilde{x}}(k|\ell) \stackrel{d}{=} \mathcal{E}\{\tilde{x}_{k|\ell}\, \tilde{x}^T_{k|\ell}\} .$$

So, by the solution of the optimal estimation problem we mean a set of equations which allow us to compute the pair $(\tilde{x}_{k|\ell}, P_{\tilde{x}}(k|\ell))$. We shall refer to this pair as prediction if $k > \ell$, filtering if $k = \ell$, and smoothing if $k < \ell$.

1.3 OUTLINE OF THE THESIS

The outline of the dissertation is as follows. Chapter 2 contains the mathematical background upon which the derivations of the estimation equations are based and the basic results. In Chapter 3, the optimal estimation equations for the signal $x_k = \Phi(k)u_k$ are derived under Assumption (A.1.1). Optimal estimation equations for the signal $x_k = \Phi(k)u_k$ under Assumptions (A.1.2) and (A.1.3) are derived in Chapters 4 and 5 respectively.

The results of the thesis are reviewed in Chapter 6 and conclusions are drawn concerning the application of this approach. A number of extensions of the present research are proposed.

CHAPTER 2

MATHEMATICAL BACKGROUND AND BASIC RESULTS

This chapter is devoted to the basic mathematical notions which are used throughout the dissertation and a new solution of the general discrete-time optimal estimation problem (cf. Section 3). The material presented in Section 1 and Section 2 is based on the works of Wiener and Masani [W-3], and Mandrekar [M-2], respectively. The proofs of the new results and the known results whose proof belong to the author are presented.

2.1 HILBERT SPACE OF RANDOM VECTORS

Let $(\Omega,\mathcal{A},P)$ be a probability space; that is, Ω is a set of points w, $\mathcal{A}$ is a σ-algebra of subsets of Ω, and P is a probability measure on Ω. A certain property is said to hold P-almost-everywhere on Ω (or with probability one) if the probability of the set of points w at which this property does not hold equals zero. We indicate this property with the expression (mod P).

Let $(X,\mathcal{B})$ be a measurable space (cf. [R-3], p. 217). A function $x : \Omega \to X$ [1] is called a <u>random element</u> with range in X, if it is $\mathcal{A}$-measurable; i.e. $\forall B \in \mathcal{B} : \{w \mid x(w) \in B\} \in \mathcal{A}$. A random element with range in a finite n-dimensional linear space

---

1) That is $w \mapsto x(w) \in X$.

is called a <u>random n-vector</u>. In the case in which X is the real line R and $\mathcal{B}$ is the σ-algebra of Borel subsets of R, the function x is called a <u>random variable</u> (abbreviation: r.v.). In this study, the range space X is chosen to be the Euclidean n-space R^n, and σ-algebra $\mathcal{B}$ is chosen to be the σ-algebra of Borel subsets of R^n, so that a random element x with range in R^n is a random n-vector (column) with r.v. components $x^{(i)}$, $i = 1,2,\ldots,n$.

We denote by $L_2(\Omega,\mathcal{A},P)$ (or L_2) the set of all r.v. x defined on $(\Omega,\mathcal{A},P)$ which are square integrable:

$$\mathcal{E}\{|x|^2\} \stackrel{d}{=} \int_\Omega |x(w)|^2 P(dw) < \infty.$$

The set $L_2(\Omega,\mathcal{A},P)$ is a Hilbert space (abbreviation: H-space) with usual operations and inner product (cf. [G-2], Theorem 7, Section 5 of Chapter II).

Now, let $L_2^n(\Omega,\mathcal{A},P)$ (or L_2^n) be the set of all random n-vectors x on Ω, with components $x^{(i)} \in L_2(\Omega,\mathcal{A},P)$, $i = 1,2,\ldots,n$. Thus $x \in L_2^n(\Omega,\mathcal{A},P)$ iff[1] $x^{(i)}$, $i = 1,2,\ldots,n$ are random variables and x is square integrable; i.e.

$$\mathcal{E}\{|x|^2\} \stackrel{d}{=} \int_\Omega |x(w)|^2 P(dw) < \infty ,$$

where $|\cdot|$ denotes the Euclidean norm:

$$|x|^2 \stackrel{d}{=} x^T x = \sum_{i=1}^{n} |x^{(i)}|^2.$$

The space $L_2^n(\Omega,\mathcal{A},P)$ is a direct-product H-space (cf. [P-1], p. 75) with usual operations and the inner product

---

1) "iff" is shorthand for "if, and only if".

$$
\begin{aligned}
\langle x,y\rangle &\stackrel{d}{=} \sum_{i=1}^{n} \langle x^{(i)},y^{(i)}\rangle_L \\
&= \int_{\Omega} \sum_{i=1}^{n} x^{(i)}(w)y^{(i)}(w)\,P(dw) \qquad (2.1)\\
&= \mathcal{E}\{x^T y\} \ .
\end{aligned}
$$

This inner product generates the norm

$$
\begin{aligned}
\|x\| = (\langle x,x\rangle)^{\frac{1}{2}} &= \Big(\sum_{i=1}^{n} \|x^{(i)}\|_L^2\Big)^{\frac{1}{2}} \qquad (2.2)\\
&= (\mathcal{E}\{|x|^2\})^{\frac{1}{2}} \ .
\end{aligned}
$$

which in turn induces a topology in L_2^n : a sequence $\{x_k\}$ in L_2^n is said (i) to converge to a vector $x \in L_2^n$ iff $\|x_k - x\| \to 0$ as $k \to \infty$, and (ii) to be a Cauchy sequence iff $\|h_k - h_m\| \to 0$ as $k,m \to \infty$.

The inner product (2.1) does not play any significant role in the stochastic theory, although the corresponding norm (2.2) and topology it induces do. Rather than inner product we often use rectangular Gramian matrices.

(2.3) DEFINITION. The $n \times m$ matrix

$$
\begin{aligned}
[x,y] &\stackrel{d}{=} [\langle x^{(i)}, y^{(j)}\rangle_L] \\
&= [\int_{\Omega} x^{(i)}(w)y^{(j)}(w)\,P(dw)] \qquad \begin{matrix} i = 1,2,\ldots,n; \\ j = 1,2,\ldots,m \end{matrix}
\end{aligned}
$$

that is defined for $x \in L_2^n$, $y \in L_2^m$ is called the Gramian of x,y.[1)]

---

1) In what follows we assume implicitly that the random vectors x and y are defined on the same probability space. This assumption is made throughout of this work.

In the next two definitions we introduce the concepts of orthogonality and subspace [W-3]. These definitions differ from the usual ones in that Gramians replace inner products, and matrix coefficients replace scalar coefficients in linear combinations.

(2.4) DEFINITION. We say that:

(a) two vectors x,y in L_2^n are orthogonal, written as $x \perp y$, iff $[x,y] = 0$

(b) two sets M,N contained in L_2^n are orthogonal, written as $M \perp N$, iff each vector in M is orthogonal to each vector in N.

(2.5) REMARK. (a) Note that this concept of orthogonality is stronger than the usual one. For $x \perp y$, it is not sufficient that $\langle x,y \rangle = 0$.

(b) From (2.3) we see that $[Ax, By] = A[x,y]B^T$, $\forall n \times p$ and $m \times q$ real matrices A,B; and vecotrs $x \in L_2^p$, $y \in L_2^q$. Hence, if $x \perp y$, then $Ax \perp By$.

(2.6) DEFINITION. A non-empty subset M of L_2^n is said to be:

(a) a linear manifold if $x,y \in M \Rightarrow Ax + By \in M$, $\forall n \times n$ real matrices A,B.

(b) a subspace of L_2^n if it is a linear manifold, which is closed in the topology of the norm (2.2). The subspace spanned by the family of random-vectors $\{x_k, k \in \Lambda\}$ where Λ is an index set will be denoted by $\mathcal{L}\{x_k, k \in \Lambda\}$.

The basic facts governing the notions just introduced which are referred in this study are given in the next two lemmas. These lemmas are quoted here from the work of N. Wiener and P. Masani ([W-3], Lemmas 5.8 and 5.9).

(2.7) ORTHOGONAL PROJECTION LEMMA (Wiener-Masani, 1957). (a) M is a subspace of L_2^n iff there is a subspace M_o of $L_2 \ni M = M_o^n$, where M_o^n denotes the direct-product $M_o \otimes \ldots \otimes M_o$ with n factors. M_o is the set of all components of all vectors in M.

(b) If M is a subspace of L_2^n and $x \in L_2^n$, then $\exists$ a unique (mod P) $x_M \in M \ni$

$$\|x - x_M\| = \min_{y \in M} \|x - y\| .$$

A vector $x \in M$ satisfies this equality iff it satisfies the following equivalent conditions:

$$x - x_M \perp M \quad \text{or} \quad [x , y] = [x_M, y] \quad \forall\, y \in M.$$

(c) If M,N are subspaces of L_2^n and $M \subset N$, then $\exists$ a unique subspace $M' \subset N \ni N = M \oplus M'$, $M \perp M'$ where $\oplus$ denotes a direct sum of vector spaces (cf. [P-2], p. 38). In particular, $M \subset L_2^n$ is a subspace $\Rightarrow$ $\exists$ a unique $M^\perp$, called the orthogonal complement of M such that $L_2^n = M \oplus M^\perp$, $M \perp M^\perp$.

(2.8) DEFINITION. The unique vector x_M of (2.7b) is called the orthogonal projection of x onto M, and denoted by $(x|M)$. The operator P_M on L_2^n defined by $P_M(x) = (x|M)$ $\forall\, x \in L_2^n$ is called the orthogonal projection operator.

(2.9) LEMMA (Wiener-Masani, 1957). (a) If M,N are orthogonal subspaces of L_2^n, then $M \oplus N$ is a subspace of L_2^n and for any $x \in L_2^n$

$$(x|M \oplus N) = (x|M) + (x|N) \quad \text{(mod P)}$$

(b) If M,N are subspaces $\ni M \subset N$, then for any $x \in L_2^n$

$$\|(x|M)\| \leq \|(x|N)\| \ .$$

In the study of the properties of orthogonal projection, a basic role is played by (2.7b). For example, from (2.7b) we may deduce the following:

(2.10) COROLLARY. Let M,N be subspaces of $L_2^n \ni M \subset N$. Then the following holds (mod P):

(a) $(Ax + By|M) = A(x|M) + B(y|M)$, $\forall n \times n$ real matrices A,B and $x,y \in L_2^n$

(b) $((x|M)|N) = ((x|N)|M) = (x|M) \quad \forall x \in L_2^n$

(c) $\forall\ x \in M$, $(x|M) = x$ and $\forall\ y \in M^{\perp}$, $(y|M) = 0$.

We conclude this section with the remark that the orthogonal projection of any random vector is defined only (mod P). Such a projection ought therefore to be viewed as any one of an equivalence class of random vectors differing from one another only on sets of zero probability. Since these sets of probability zero do not, in general, play an essential role in this study, the phrase "(mod P)" will usually be omitted.

2.2 DISCRETE-TIME WIDE-SENSE MARTINGALE AND MARKOV PROCESSES

Let $(\Omega,\mathcal{A},P)$ be a probability space. By an n-variate discrete-time stochastic process on $(\Omega,\mathcal{A},P)$ we mean a family of random vectors $\{x_k,\ k \in Z\}$. If for each $k \in Z$, $x_k \in L_2^n$ then the process is said to be a second-order process.

Associated with a second-order n-variate process $\{x_k,\ k \in Z\}$ we denote the following:

(i) the mean-value function $m_x(\cdot)$ by

$$m_x(k) \stackrel{d}{=} \mathcal{E}\{x_k\},\ k \in Z,$$

(ii) the correlation matrix function (or simply the correlation matrix) $C_x(\cdot,\cdot)$ by

$$C_x(i,j) \stackrel{d}{=} \mathcal{E}\{x_i\ x_j^T\} = [x_i,\ x_j^T]\quad i,j \in Z,$$

(iii) the covariance matrix function (or simply the covariance matrix) $P_x(\cdot,\cdot)$ by

$$P_x(i,j) \stackrel{d}{=} \mathcal{E}\{(x_i - m_x(i))(x_j - m_x(j))^T\}$$

$$= [x_i - m_x(i),\ x_j - m_x(j)]\ .$$

If $m_x(i) = 0\ \forall\ i \in Z$, then the process is said to have zero mean. For a zero mean process, it is clear that $P_x(i,j) = C_x(i,j)$ $\forall i,j \in Z$. If $C_x(i,j) = m_x(i)m_x^T(j)$ that is, if $P_x(i,j) = 0$ $\forall i,j \in Z$ then the process is said to be uncorrelated. From now on we assume, without loss of generality, that all the processes have zero mean unless otherwise stated.

(2.11) DEFINITION. A second-order multi-variate process $\{x_k,\ k \in Z\}$ is said to be:

(a) a white noise (or orthogonal) process iff $[x_i,x_j] = P_x(j)\delta_{ij}$ $\forall i,j \in Z$, where $P_x(j) \stackrel{d}{=} P_x(j,j)$,

(b) a process with orthogonal increments iff the process $\{x_{k+1} - x_k,\ k \in Z\}$ is orthogonal.

Let $\{x_k,\ k \in Z\}$ be a stochastic process in L_2^n. We shall denote by $L(x;\ell)$ and $L(x_\ell)$ the spaces $\mathcal{L}\{x_i,\ i \le \ell\}$ and $\mathcal{L}\{x_\ell\}$ respectively (cf. Definition 2.6b). These are called the past-present and the present of the process $\{x_k,\ k \in Z\}$ respectively. Obviously $L(x_\ell) \subset L(x;\ell) \subset L(x;\ell+1)$.

(2.12) DEFINITION. A discrete-time process $\{x_k, k \in Z\}$ in L_2^n is said to be:

(a) a wide-sense martingale process, iff $\forall k,\ell \in Z$ with $\ell \leq k$,

$$(x_k | L(x;\ell)) = x_\ell ;$$

(b) a wide-sense Markov process, iff $\forall\ k,\ell \in Z$, with $\ell \leq k$

$$(x_k | L(x;\ell)) = (x_k | L(x_\ell)) .$$

We see from (2.7b), (2.12a) that a process $\{x_k, k \in Z\}$ in L_2^n is a wide-sense martingale process iff $\forall k,\ell \in Z$ with $\ell \leq k$, $x_k - x_\ell \perp x_i \quad \forall i \leq \ell$. Hence for a wide-sense martingale process $P_x(i,j) = P_x(i \wedge j)$, where $i \wedge j \overset{d}{=} \min\{i,j\}$.

A necessary and sufficient condition for a process $\{x_k, k \in Z\}$ in L_2^n to be a wide-sense martingale is given in the following lemma. This lemma extends to n-variate case the result of Doob ([D-2], p. 166). The proof of the lemma is similar to the one that is given by Doob, so it is omitted.

(2.13) LEMMA. A discrete-time process $\{x_k, k \in Z\}$ in L_2^n is wide-sense martingale iff it satisfies the first-order linear vector difference equation

$$x_{k+1} = x_k + w_{k+1} \qquad x_{-\infty} = w_{-\infty} \qquad k > -\infty$$

where $\{w_k, k \in Z\}$ is a white-noise process.

(2.14) REMARK. (a) In continuous-time case the "only if" part of Lemma (2.13) does not hold, i.e. in continuous-time a wide-sense martingale process cannot be generated by a linear differential equation unless it is differentiable in the sense [M-3]. When that is so, the approach of the thesis can be applied to continuous time.

(b) We see from (2.5b) that if $\{w_k, k \in Z\}$ is a white noise process in L_2^n then $\{\Gamma(k)w_k, k \in Z\}$, where $\Gamma(k)$ is an $n \times m$ matrix function, is a white-noise process in L_2^n. Hence the solution of

$$x_{k+1} = x_k + \Gamma(k+1)w_{k+1} \quad x_{-\infty} = w_{-\infty} \quad k > -\infty$$

is a wide-sense martingale, by virtue of (2.13).

One important wide-sense martingale process for our purpose is the orthogonal projection of a process.

(2.15) LEMMA. Let $y \in L_2^n$ be fixed and $\{x_k, k \in Z\}$ be a process in L_2^n. Define

$$u_k = (y|L(x;k)), \ k \in Z \ .$$

The process $\{u_k, k \in Z\}$ is a wide-sense martingale.

PROOF. Since $y \in L_2^n$, so is u_k, $k \in Z$ by virtue of (2.9b). Also from (2.10b)

$$\begin{aligned}(u_k|L(x;\ell)) &= ((y|L(x;k))|L(x;\ell)) \\ &= (y|L(x;\ell)) \\ &= u_\ell\end{aligned}$$

for $\forall\, \ell \le k$, where we used the fact that $L(x;\ell) \subset L(x;k)$, $\ell \le k$. Since $L(u;\ell) \subset L(x;k)$, $\ell \le k$, by (2.10b)

$$\begin{aligned}(u_k|L(u;\ell) &= (u_k|L(x;\ell)|L(u;\ell)) \\ &= (u_\ell|L(u;\ell)) \\ &= u_\ell \quad \forall\, \ell \le k \ .\end{aligned}$$

QED

Next we study discrete-time wide-sense Markov processes. The concept of wide-sense Markov processes first was introduced by Doob [D-2]. Let $\{x_k, k \in Z\}$ be a wide-sense Markov process in L_2^n. The definition of wide-sense Markov process implies that $(x_k | L(x;\ell)) = A(k,\ell)x_\ell$ for $\ell \le k$, where $A(k,\ell)$ is an $n \times n$ matrix. Beutler [B-1] proved that

$$A(k,\ell) = P_x(k,\ell)P_x^+(\ell,\ell)$$

where $P_x(k,\ell) = [x_k, x_\ell]$ and P^+ denotes the pseudo-inverse (cf. A.3) of the matrix P. He also showed that a multivariate second-order process is a wide-sense Markov process iff

$$A(k,\ell) = A(k,j)A(j,\ell)$$

for $\ell \le j \le k$. The function $A(k,\ell)$ is called a transition matrix.

Mandrekar and Salehi ([M-3], Theorems 2.11 and 2.12) recently gave a representation to a wide-sense Markov process, extending to a singular case the work of Mandrekar [M-1], [M-2] which shows the connection between the signal process (1.4) of the problem that is considered in this study, and wide-sense Markov processes.

(2.16) THEOREM (Mandrekar-Salehi, 1971). Let $\{x_k, k \in Z\}$ be a discrete-time stochastic process in L_2^n. Then the process $\{x_k, k \in Z\}$ is a wide-sense Markov process with the transition function $A(k,\ell) = P(k,\ell)P^+(\ell,\ell)$ such that $P_x(k,\ell)P^+(\ell,\ell)$ is one-one on $R(P_x(k,\ell))$ onto $R(P_x(\ell,\ell)$ [1] for $k \le \ell$ iff

---

1) Here $R(\Phi)$ denotes the range space of the linear operator Φ (see Appendix A).

$x_k = \Phi(k)u_k$ where $\{u_k, k \in Z\}$ is a wide-sense martingale and $\Phi(k)$ is an $n \times n$ matrix function such that $u_k \in R(\Phi^T(k))$, $R(\Phi(k))$ is independent of $k \in Z$. In either case $A(k,\ell) = \Phi(k)\Phi^+(\ell)$. This representation is unique in the sense that if $x_k = \Psi(k)v_k$ with $v \in R(\Psi^T(k))$ then there exists an $n \times n$ matrix K such that $\forall k \in Z$ we have $u_k = Kv_k$ and $\Phi(k) = \Psi(k)K^+$.

(2.17) EXAMPLE: The Kalman filtering theory (cf. [K-3], [K-4]) assumes that the signal process $\{x_k, k = 0,1,2,\ldots\}$ is generated by a given linear difference equation of the form

$$x_{k+1} = F(k)x_k + \Gamma(k)w_k \quad k = 0,1,2,\ldots \tag{2.17a}$$

where $F(k)$ and $\Gamma(k)$ are matrices of functions of discrete-time and w_k is a white noise process. In addition $F(k)$ is invertible for all $k = 0,1,\ldots$ and w_k is orthogonal to a given any initial condition. Given an initial condition x_o, then the unique solution of (2.17a) is given by the expression (cf. [K-3])

$$x_k = \Phi(k,o)x_o + \sum_{i=1}^{k} \Phi(k,i)\Gamma(i-1)w_{i-1} \quad k = 0,1,2,\ldots \tag{2.17b}$$

where $\Phi(\cdot,\cdot)$ is the fundamental solution of

$$\Phi(k+1,\ell) = F(k)\Phi(k,\ell) \ , \ \Phi(\ell,\ell) = I$$

Define

$$u_k \overset{d}{=} x_o + \sum_{i=1}^{k} \Phi(o,i)\Gamma(i-1)w_{i-1}$$

$$= u_{k-1} + \Phi(o,k)\Gamma(k-1)w_{k-1} \quad \text{for} \quad k \geq 1$$

$$u_o \overset{d}{=} x_o \quad \text{for} \quad k = 0 \ .$$

Then (2.17b) can be written as $x_k = \Phi(k,o)u_k$. It follows from the definition u_k and (2.14b) that the stochastic process $\{u_k, k = 0,1,2,\ldots\}$ is a zero-mean wide-sense martingale whose covariance matrix

$$(2.17c) \quad P_u(k) = P_x(o) + \sum_{i=1}^{k} \Phi(o,i)\Gamma(i-1)P_w(i-1)\Gamma^T(i-1)\Phi^T(o,i)$$

$$= P_u(k-1) + \Phi(o,k)\Gamma(k-1)P_w(k-1)\Gamma^T(k-1)\Phi^T(o,k) \ .$$

Thus, by (2.16), the process $\{x_k, k = 0,1,2,\ldots\}$, which is defined by (2.17a) with the initial condition x_o is a wide-sense Markov process. We shall refer to this signal as a Kalman signal.

Observe that the class of Kalman signals is a special class of wide-sense Markov processes. It is obviously contained in the class of signal defined by (1.4). Thus the optimal estimation equations of a Kalman signal may be obtained from the optimal estimation equations for a signal defined by (1.4), but not conversely.

2.3 A SOLUTION OF THE GENERAL MINIMUM MEAN-SQUARE ESTIMATION PROBLEM

The problem we will discuss in this section can be formulated as follows: Consider two related stochastic processes $\{x_k, k = 0,1,2,\ldots\}$ and $\{y_k, k = 1,2,\ldots\}$ in L_2^n and L_2^m $(m \leq n)$ respectively. We will refer to $\{x_k, k = 0,1,2,\ldots\}$ as signal and to $\{y_k, k = 1,2,\ldots\}$ as the observation process. The problem is to find the minimum mean-square estimate of the signal x_k based on the observation record $Y(\ell) \overset{d}{=} \{y_1,\ldots,y_\ell\}$; that is, to find that vector $\hat{x}_{k|\ell}$ of the form $\sum_{i=1}^{\ell} A(i)y_i$, where $A(i)$, $i = 1,2,\ldots,\ell$ are $n \times m$ real matrices which minimizes the mean-

square error risk function (1.3):

$$J[\tilde{x}_{k|\ell}] = \mathcal{E}\{(x_k - \hat{x}_{k|\ell})^T (x_k - \hat{x}_{k|\ell})\}$$

$$= \|x_k - \hat{x}_{k|\ell}\|^2 \quad . \tag{2.18}$$

The solution of this problem is as follows: Let

$$L_2^n(y;\ell) \stackrel{d}{=} \mathcal{L}\{K(i)y_i,\ i = 1,2,\ldots,\ell\}$$

where $K(i)$, $i = 1,2,\ldots,\ell$ are $n \times m$ real matrices. Note that $L_2^n(y;\ell)$ is a subspace of L_2^n and $L_2^n(y;\ell) \neq L(y;\ell)$ (cf. Definition 2.6b). So that the problem is reduced to finding that vector $\hat{x}_{k|\ell}$ in $L_2^n(y;\ell)$ such that (2.18) is minimum. From the orthogonal projection lemma (cf. 2.7b), we know that $\hat{x}_{k|\ell}$ is given by

$$\hat{x}_{k|\ell} = (x_k | L_2^n(y;\ell))^{1)}$$

$$= P_{L_2^n(y;\ell)} x_k \quad . \tag{2.19}$$

That is the linear minimum mean-square estimate $\hat{x}_{k|\ell}$ of the signal x_k, based on the observation record $Y(\ell)$, is the orthogonal projection x_k onto the subspace $L_2^n(y;\ell)$ generated by the vectors $K(i)y_i$, $i = 1,2,\ldots,\ell$. We will refer to the operator $(\cdot | L_2^n(y;\ell)) = P_{L_2^n(y;\ell)}$ defined by (2.19) as the orthogonal projection estimator (abbreviation: OPE).

---

1) We recall that given processes have zero-means, otherwise

$$\hat{x}_{k|\ell} = \tilde{x}_{k|o} + (x_k | L_2^n(y;\ell))$$

where $\hat{x}_{k|\ell}$ = optimal estimate x_k given no observation $\stackrel{d}{=} \mathcal{E}\{x_k\}$.

From now on the notation $\hat{x}_{k|\ell}$ will denote the minimum mean-square estimate of the signal x_k based on the observation record $Y(\ell)$ which is given by (2.19), and the expression "optimal estimate" will mean this estimate. The pair $(\hat{x}_{k|\ell}, P_{\tilde{x}}(k|\ell))$ will denote the optimal estimate and the corresponding error covariance matrix of the estimation problem. We refer to the pair $(\hat{x}_{k|\ell}, P_{\tilde{x}}(k|\ell))$ as optimal estimation of the signal x_k, $k = 0,1,2,\ldots$ based on the observation record $Y(\ell)$, $\ell = 1,2,\ldots$ if k,ℓ are not specified, as optimal prediction if $k > \ell$, as optimal filtering if $k = \ell$, and as optimal smoothing if $k < \ell$.

Now let $\{x_k, k = 0,1,2,\ldots\}$ and $\{u_k, k = 0,1,2,\ldots\}$ be two signal processes such that

$$x_k = \Phi(k)u_k \qquad k = 0,1,2,\ldots, \tag{2.20}$$

where $\Phi(k)$ is a linear transformation which may not be invertible and $x_k, u_k \in L_2^n$ $\forall k = 0,1,2,\ldots$. Let $\{y_k, k = 1,2,\ldots\}$ in L_2^m $(m \leq n)$ be the observation process for the signal $\{x_k, k = 0,1,2,\ldots\}$. Then the optimal estimates of the signals x_k and u_k based on the observation record $Y(\ell)$ are

$$\hat{x}_{k|\ell} = (x_k | L_2^n(y;\ell)) \quad \text{and} \quad \hat{u}_{k|\ell} = (u_k | L_2^n(y;\ell))$$

respectively. Since $x_k = \Phi(k)u_k$ we have

$$\begin{aligned} \hat{x}_{k|\ell} &= (\Phi(k)u_k | L_2^n(y;\ell)) \\ &= \Phi(k)(u_k | L_2^n(y;\ell)) \\ &= \Phi(k)\hat{u}_{k|\ell} \,. \end{aligned} \tag{2.21a}$$

Thus the OPE commute with any linear deformation of the signal

process. Furthermore if $u_k \in R(\Phi^T(k))$ $\forall$ $k = 0,1,2,\ldots$, then

$$\begin{aligned}\Phi^+(k)x_k &= \Phi^+(k)\Phi(k)u_k \\ &= P_{R(\Phi^T(k))}u_k \quad , \text{ by (A.3)} \\ &= u_k\end{aligned}$$

since $u_k \in R(\Phi^T(k))$ $\forall$ $k = 0,1,2,\ldots$. Hence, it follows from (2.20) that

$$\hat{u}_{k|\ell} = \Phi^+(k)\hat{x}_{k|\ell} \cdot \tag{2.21b}$$

From the definitions of estimation error and covariance matrix of error, and (2.21a,b) similar results can easily be derived for the estimation error and its covariance matrix. These results are summarized in the following lemma.

(2.22) LEMMA. Let $\{x_k, k = 0,1,2,\ldots\}$ and $\{u_k, k = 0,1,2,\ldots\}$ be two signal processes as above. The optimal estimations $(\hat{x}_{k|\ell}, P_{\tilde{x}}(k|\ell))$ and $(\hat{u}_{k|\ell}, P_{\tilde{u}}(k|\ell))$, $k,\ell = 0,1,2,\ldots$ based on the observation record $Y(\ell)$ of the processes x_k, $k = 0,1,\ldots$ and u_k, $k = 0,1,\ldots$ are related to each other via the following relations:

(a) The optimal estimates:

$$\hat{x}_{k|\ell} = \Phi(k)\hat{u}_{k|\ell} \qquad k,\ell = 0,1,2,\ldots .$$

The estimation errors:

$$\tilde{x}_{k|\ell} = \Phi(k)\tilde{u}_{k|\ell} \qquad k,\ell = 0,1,2,\ldots .$$

The error covariance matrices:

$$P_{\tilde{x}}(k|\ell) = \Phi(k)P_{\tilde{u}}(k|\ell)\Phi^T(k) \qquad k,\ell = 0,1,2,\ldots .$$

(b) If in addition $u_k \in R(\Phi^T(k))$ $\forall$ $k = 0,1,2,\ldots$ then for $k,\ell = 0,1,2,\ldots$ we have

$$\hat{u}_{k|\ell} = \Phi^+(k)\hat{x}_{k|\ell} ,$$

$$\tilde{u}_{k|\ell} = \Phi^+(k)\tilde{x}_{k|\ell} ,$$

$$P_{\tilde{u}}(k|\ell) = \Phi^+(k) P_{\tilde{u}}(k|\ell) \Phi^{+T}(k) .$$

(2.23) NOTE. From this lemma we conclude that if we are given a signal process which can be written as a linear deformation of another signal (i.e. can be written in the form (2.20)) then it is sufficient to derive estimation equations for the latter signal. Thus in our problem (cf. Section 2 of Chapter 1) we may consider signal process as $\{u_k, k = 0,1,2,\ldots\}$ and derive the estimation equations for this signal then use (2.22a) to obtain the required equations for the original process. Since the process $\{u_k, k = 0,1,\ldots\}$ is a wide-sense martingale this approach will simplify the derivation of estimation equations as will be demonstrated.

Following Wiener and Masani [W-3] (also see [K-8]) we shall say that the stochastic process $\{y_k, k = 0,1,2,\ldots\}$ in L_2^m is purely non-deterministic iff for all k, $y_k \notin L(y,k-1)$ where $L(y,k-1)$ is as defined in (2.6b). Hence for any purely non-deterministic process $\{y_k, k = 0,1,\ldots\}$,

$$\tilde{y}_{k+1|k} = y_{k+1} - \hat{y}_{k+1|k} \neq 0 \quad k = 0,1,2,\ldots \tag{2.24}$$

where $\hat{y}_{k+1|k} \stackrel{d}{=} (y_k | L(y,k))$. For any second order process, we shall call the process $\{\tilde{y}_{k+1|k}, k = 0,1,\ldots\}$ the <u>innovation</u>

process[1] associated with $\{y_k,\ k = 0,1,2,\ldots\}$. We observe that for purely non-deterministic second order process $\tilde{y}_{k+1|k} \neq 0$ for all k. This process plays very important roles in this study because of its simple structure, as shown in the following lemma. (For the previous results in this line see [C-2], [C-3], [W-3]).

(2.25) LEMMA. If $\{\tilde{y}_{k+1|k},\ k = 0,1,2,\ldots\}$ is the innovation-process of a stochastic process $\{y_k,\ k = 0,1,..\}$ in L_2^m then it is an orthogonal process, i.e.

$$[\tilde{y}_{k+1|k},\ \tilde{y}_{j+1|j}] = P_{\tilde{y}}(k+1)\delta_{kj}.$$

PROOF. In view of (2.11a) we must show that

$$[\tilde{y}_{k+1|k},\ \tilde{y}_{j+1|j}] = 0$$

for $k \neq j$.

From (2.24) we have if $k > j$

$$\begin{aligned}[\tilde{y}_{k+1|k},\ \tilde{y}_{j+1|j}] &= [\tilde{y}_{k+1|k},\ y_{j+1}] - [\tilde{y}_{k+1|k},\ \hat{y}_{j+1|j}] \\ &= [\tilde{y}_{k+1|k},\ y_{j+1}] = [\tilde{y}_{k+1|k},\ \hat{y}_{j+1|j}] \\ &= 0\end{aligned}$$

Since $\tilde{y}_{k+1|k} \perp L(y,j+1)$ for $j+1 \leq k$. Similarly we find that $[\tilde{y}_{k+1|k},\ \tilde{y}_{j+1|j}] = 0$ for $k+1 \leq j$. So that

$$[\tilde{y}_{k+1|k},\ \tilde{y}_{j+1|j}] = P_{\tilde{y}}(k+1)\delta_{kj},$$

---

1) Note that this definition of innovation process is different than one that was recently given by Kailath (cf. [K-1]). Our definition is the one that was given by Cramér [C-3]. (Also see [W-3]).

where

$$P_{\tilde{y}}(k+1) \stackrel{d}{=} [\tilde{y}_{k+1|k}, \tilde{y}_{k+1|k}] \geq 0 . \qquad \text{QED}$$

It is obvious that the stochastic process $\{y_k, k = 0,1,...\}$ is purely non-deterministic if rank $P_{\tilde{y}}(k+1) \geq 1$ for all $k = 0,1,...$. Let rank $P_{\tilde{y}}(k) = r(k)$ $k \geq 1$. We shall refer to $r(k)$ as the rank of the stochastic process $\{y_k, k = 0,1,2,...\}$. It is clear that $r(k) \leq m, \forall k = 0,1,2,...,$ if $r(k) = m, \forall k$ then we say that the stochastic process has full rank. We note that $P_{\tilde{y}}(k+1)$ is invertible iff $r = m$, that is, the stochastic process $\{y_k, k = 0,1,2,...\}$ has full rank (cf. [C-3], [W-3]).

Related to the innovation process associated with a signal process $\{y_k, k = 0,1,2,...\}$ we have the following result (see also [C-3], [W-3]).

(2.26) LEMMA. Let $\{\tilde{y}_{k+1|k}, k = 0,1,...\}$ be the innovation process associated with a stochastic process $\{y_k, k = 0,1,2,...\}$. Then for $\ell < k$, $\ell,k = 0,1,2,...,$

$$L(y;\ell) \perp L(\tilde{y}_{\ell+1|\ell}) \oplus ... \oplus L(\tilde{y}_{k|k+1})$$

and

$$L(y;k) = L(y;\ell) \oplus L(\tilde{y}_{\ell+1|\ell}) \oplus L(\tilde{y}_{\ell+2|\ell+1}) \oplus ... \oplus L(\tilde{y}_{k|k-1}).$$

PROOF. In view of (2.25) the subspace $L(\tilde{y}_{\ell+1|\ell})$, $L(\tilde{y}_{\ell+2|\ell+1})$,..., $L(\tilde{y}_{k|k-1})$ are orthogonal to each other. So that

$$L(\tilde{y}_{\ell+1|\ell}) \oplus ... \oplus L(\tilde{y}_{k|k-1})$$

is a subspace by virtue of (2.9a). Since by (2.24)

$\tilde{y}_{\ell+1|\ell}, \tilde{y}_{\ell+2|\ell+1}, \ldots, \tilde{y}_{k|k-1} \perp L(y;\ell), L(y;\ell+1), \ldots, L(y;k-1)$ respectively, and $L(y,\ell)$ is contained in all these subspaces, it follows that

$$L(y;\ell) \perp L(\tilde{y}_{\ell+1|\ell}) \oplus \ldots \oplus L(\tilde{y}_{k|k-1}) ,$$

and therefore by (2.9a)

$$L(y;\ell) \oplus L(\tilde{y}_{\ell+1|\ell}) \oplus \ldots \oplus L(\tilde{y}_{k|k-1})$$

is a subspace. Now it remains to show that this subspace is equal to $L(y;k)$. Since $L(y,k-1) \subset L(y,k)$ and $\tilde{y}_{k|k-1} \in L(y;k)$, we have

$$L(y,k) \supset L(y,k-1) \oplus L(\tilde{y}_{k|k-1}) .$$

On the other hand, by (2.24)

$$(*) \qquad y_k = \tilde{y}_{k|k-1} + (y_k | L(y;k-1)) \in L(\tilde{y}_{k|k-1}) \oplus L(y,k-1).$$

and for $\ell < k$, $y_\ell \in L(y;k-1) \subset L(\tilde{y}_{k|k-1}) \oplus L(y;k-1)$. It follows that

$$(**) \qquad L(y,k) \subset L(y,k-1) \oplus L(\tilde{y}_{k|k-1}) .$$

Combining (*) and (**) we obtain

$$L(y,k) = L(y,k-1) \oplus L(\tilde{y}_{k|k-1}).$$

By iteration of this equality we get

$$\begin{aligned} L(y,k) &= L(y,k-1) \oplus L(\tilde{y}_{k|k-1}) \\ &= L(y,k-2) \oplus L(\tilde{y}_{k-1|k-2}) \oplus L(\tilde{y}_{k|k-1}) \\ &\vdots \\ &= L(y,\ell) \oplus L(\tilde{y}_{k+1|\ell}) \oplus \ldots \oplus L(\tilde{y}_{k|k-1}). \end{aligned}$$

QED

Now, consider the signal process $\{x_k,\ k = 0,1,2,\ldots\}$ in L_2^n and the associated observation process $\{y_k,\ k = 1,2,\ldots\}$ in L_2^m $(m \le n)$. Suppose that the process $\{y_k,\ k = 1,2,\ldots\}$ has full rank; then we have the following fundamental result:

(2.27) THEOREM. The optimal estimation $(\hat{x}_{k|\ell},\ P_{\tilde{x}}(k|\ell))$ $k,\ell = 0,1,2,\ldots$ of the signal $x_k,\ k = 0,1,2,\ldots$ is accomplished as follows.

(a) If $k > \ell = 1,2,\ldots$, optimal prediction:

$$\hat{x}_{k|\ell} = \hat{x}_{k|\ell-1} + G(k|\ell)\tilde{y}_{\ell|\ell-1}\ , \quad \text{optimal predicted estimate,} \tag{2.28}$$

$$G(k|\ell) = [\tilde{x}_{k|\ell-1},\ \tilde{y}_{\ell|\ell-1}][\tilde{y}_{\ell|\ell-1},\ \tilde{y}_{\ell|\ell-1}]^{-1}\ , \quad \text{predictor gain matrix} \tag{2.29}$$

$$P_{\tilde{x}}(k|\ell) = P_{\tilde{x}}(k|\ell-1) - G(k|\ell)[\tilde{y}_{\ell|\ell-1},\ \tilde{x}_{k|\ell-1}], \quad \text{prediction error covariance matrix.} \tag{2.30}$$

For $\ell = 0$, $\hat{x}_{k|o} = \mathcal{E}\{x_k\} = 0$ and $P_{\tilde{u}}(k|o) = P_u(k)$.

(b) If $k = \ell = 1,2,\ldots$, optimal filtering:

$$\hat{x}_{k|k} \overset{d}{=} \hat{x}_k = \hat{x}_{k|k-1} + G(k)\tilde{y}_{k|k-1}\ , \quad \text{optimal filtered estimate,} \tag{2.31}$$

$$G(k) = [\tilde{x}_{k|k-1},\ \tilde{y}_{k|k-1}][\tilde{y}_{k|k-1},\ \tilde{y}_{k|k-1}]^{-1}, \quad \text{filter gain matrix,} \tag{2.32}$$

$$P_{\tilde{x}}(k|k) \overset{d}{=} P_{\tilde{x}}(k) = P_{\tilde{x}}(k|k-1) - G(k)[\tilde{y}_{k|k-1},\ \tilde{x}_{k|k-1}], \quad \text{filtering error covariance matrix.} \tag{2.33}$$

For $k = 0$, $\hat{x}_o = \mathcal{E}\{x_o\} = 0$ and $P_{\tilde{u}}(0) = P_u(0)$.

(c) If $k < \ell$, optimal smoothing:

$$\hat{x}_{k|\ell} = \hat{x}_k + \sum_{i=k+1}^{\ell} G(k,i|i-1)\tilde{y}_{i|i-1}, \quad \text{optimal smoothed estimate,} \tag{2.34}$$

(2.35) $G(k,i|i-1) = [\tilde{x}_{k|i-1}, \tilde{y}_{i|i-1}][\tilde{y}_{i|i-1}, \tilde{y}_{i|i-1}]^{-1}$, smoother gain matrix,

(2.36) $P_{\tilde{x}}(k|\ell) = P_{\tilde{x}}(k) - \sum_{i=k+1}^{\ell} G(k,i|i-1)[\tilde{y}_{i|i-1}, \tilde{x}_{k|i-1}]$, smoothing error covariance matrix.

For $k = \ell$, $\hat{x}_{k|k} = \hat{x}_k$ and $P_{\tilde{x}}(k|k) = P_{\tilde{x}}(k)$.

PROOF. (a) Since by (2.26)

$$L_2^n(y;\ell) = L_2^n(y;\ell-1) \oplus L_2^n(\tilde{y}_{\ell|\ell-1}),\ L_2^n(y;\ell-1) \perp L_2^n(\tilde{y}_{\ell|\ell-1}) \quad 1)$$

we have

$$\begin{aligned}\hat{x}_{k|\ell} &= (x_k | L_2^n(y;\ell-1) \oplus L_2^n(\tilde{y}_{\ell|\ell-1})) \\ &= (x_k | L_2^n(y;\ell-1)) + (x_k | L_2^n(\tilde{y}_{\ell|\ell-1})) \text{ , by (2.28)} \\ &= \hat{x}_{k|\ell-1} + G(k|\ell)\tilde{y}_{\ell|\ell-1} \qquad \ell = 1,2,\ldots \quad k \geq \ell\end{aligned}$$

where $G(k|\ell)$ is the $n \times m$ gain matrix to be determined.

To determine the gain matrix $G(k|\ell)$, notice that from (2.7b)

$$x_k - G(k|\ell)\tilde{y}_{\ell|\ell-1} \perp \tilde{y}_{\ell|\ell-1} .$$

So that

$$[x_k, \tilde{y}_{\ell|\ell-1}] = G(k|\ell)[\tilde{y}_{\ell|\ell-1}, \tilde{y}_{\ell|\ell-1}] .$$

---

1) Actually by (2.26) we have $L(y;\ell) = L(y;\ell-1) \oplus L(\tilde{y}_{\ell|\ell-1})$, it easily follows from the definition $L_2^n(y;\ell)$ that Lemma 2.26 holds for this subspace too.

Since the observation process has full rank, we get

$$G(k|\ell) = [x_k, \tilde{y}_{\ell|\ell-1}][\tilde{y}_{\ell|\ell-1}, \tilde{y}_{\ell|\ell-1}]^{-1}$$

for $\ell = 1,2,\ldots,\ k \geq \ell$. Noting that $x_k = \tilde{x}_{k|\ell-1} + \hat{x}_{k|\ell-1}$ and $\hat{x}_{k|\ell-1} \perp \tilde{y}_{\ell|\ell-1}$ [1] we may write the expression for $G(k|\ell)$ as

$$G(k|\ell) = [\tilde{x}_{k|\ell-1}, \tilde{y}_{\ell|\ell-1}][\tilde{y}_{\ell|\ell-1}, \tilde{y}_{\ell|\ell-1}]^{-1}. \tag{2.29}$$

The prediction error is by definition

$$\begin{aligned}\tilde{x}_{k|\ell} &= x_k - \hat{x}_{k|\ell} \\ &= \tilde{x}_{k|\ell-1} - G(k|\ell)\tilde{y}_{\ell|\ell-1}\ ,\ \ell = 1,2,\ldots\end{aligned}$$

Therefore the error covariance matrix is given by

$$\begin{aligned}P_{\tilde{x}}(k|\ell) &= [\tilde{x}_{k|\ell-1} - G(k|\ell)\tilde{y}_{\ell|\ell-1}, \tilde{x}_{k|\ell-1} - G(k|\ell)\tilde{y}_{\ell|\ell-1}] \\ &= [\tilde{x}_{k|\ell-1}, \tilde{x}_{k|\ell-1}] + G(k|\ell)[\tilde{y}_{\ell|\ell-1}, \tilde{y}_{\ell|\ell-1}]G^T(k|\ell) \\ &\quad - G(k|\ell)[\tilde{y}_{\ell|\ell-1}, \tilde{x}_{k|\ell-1}] - [\tilde{x}_{k|\ell-1}, \tilde{y}_{\ell|\ell-1}]G^T(k|\ell)\ .\end{aligned}$$

Using (2.29) and noting that $[\tilde{x}_{k|\ell-1}, \tilde{y}_{\ell|\ell-1}]^T = [\tilde{y}_{\ell|\ell-1}, \tilde{x}_{k|\ell-1}]$ we get

$$P_{\tilde{x}}(k|\ell) = P_{\tilde{x}}(k|\ell-1) - G(k|\ell)[\tilde{y}_{\ell|\ell-1}, \tilde{x}_{k|\ell-1}] \tag{2.30}$$

for $k > \ell = 1,2,\ldots$.

For $\ell = 0$, obviously

---

1) $\tilde{y}_{\ell|\ell-1} \perp L(y,\ell-1) \Rightarrow \tilde{y}_{\ell|\ell-1} \perp L_2^n(y,\ell-1)$, by virtue of (2.5b)

$\Rightarrow \tilde{y}_{\ell|\ell-1} \perp \hat{x}_{k|\ell-1}$, since $\hat{x}_{k|\ell-1} \in L_2^n(y,\ell-1)$.

$$\hat{x}_{k|o} = \text{optimal estimate } x_k \text{ given no observation}$$
$$= \mathcal{E}\{x_k\} = 0$$

and

$$P_{\tilde{x}}(k|o) \stackrel{d}{=} [x_k - \hat{x}_{k|o},\ x_k - \hat{x}_{k|o}]$$
$$= [x_k, x_k]$$
$$\stackrel{d}{=} P_x(k) .$$

(b) Since the above equations hold for $k \geq \ell$, by letting $k = \ell = 1,2,\ldots$ we obtain expression for filtering (2.31-2.33). For $k = \ell = 0$, $\hat{x}_o = 0$, since

$$\hat{x}_o = \text{optimal estimate } x_o \text{ given no observation} =$$
$$= \mathcal{E}\{x_o\}$$
$$= 0 .$$

Thus $P_{\tilde{x}}(0) \stackrel{d}{=} [\tilde{x}_o, \tilde{x}_o] = [x_o, x_o] \stackrel{d}{=} P_x(0)$.

(c) Since by (2.26)

$$L_2^n(y;\ell) = L_2^n(y;k) \oplus L_2^n(\tilde{y}_{k+1|k}) \oplus \ldots \oplus L_2^n(\tilde{y}_{\ell|\ell-1}),\ \ell > k$$

and the subspaces on the right of this equality are orthogonal to each other, the optimal smoothed estimate is

$$\hat{x}_{k|\ell} = (x_k | L_2^n | y;\ell)$$
$$= (x_k | L_2^n(y;k) \oplus L_2^n(\tilde{y}_{k+1|k}) \oplus \ldots \oplus L_2^n(\tilde{y}_{\ell|\ell-1}))$$
$$= \hat{x}_k + \sum_{i=k+1}^{\ell} G(k,i|i-1)\tilde{y}_{i|i-1} \qquad (2.34)$$

where $G(k,i|i-1)$, $i = k+1,\ldots,\ell$ are the $n \times m$ gain matrices to be determined. If the steps lead to (2.29) repeated here, their

results

$$(2.35)\quad G(k,i|i-1) = [\tilde{x}_{k|i-1},\ \tilde{y}_{i|i-1}][\tilde{y}_{i|i-1},\ \tilde{y}_{i|i-1}]^{-1}$$

for $i = k+1,\ldots,\ell$.

To obtain an expression for the smoothing error covariance matrix we note that

$$\begin{aligned}\tilde{x}_{k|\ell} &= x_k - \hat{x}_{k|\ell} \\ &= \tilde{x}_k - \sum_{i=k+1}^{\ell} G(k,i|i-1)\tilde{y}_{i|i-1}\end{aligned}$$

for $\ell > k = 0,1,2,\ldots$. Hence

$$\begin{aligned}P_{\tilde{x}}(k|\ell) &= [\tilde{x}_k - \sum_{i=k+1}^{\ell} G(k,i|i-1)\tilde{y}_{i|i-1},\ \tilde{x}_k - \sum_{i=k+1}^{\ell} G(k,i|i-1)\tilde{y}_{i|i-1}] \\ &= [\tilde{x}_k,\ \tilde{x}_k] + \sum_{i=k+1}^{\ell} G(k,i|i-1)[\tilde{y}_{i|i-1},\tilde{y}_{i|i-1}]G^T(k,i|i-1) - \\ &\quad - \sum_{i=k+1}^{\ell} G(k,i|i-1)[\tilde{y}_{i|i-1},\tilde{x}_k] \\ &\quad - \sum_{k+1}^{\ell} [\tilde{x}_k,\tilde{y}_{i|i-1}]G^T(k,i|i-1)\end{aligned}$$

where we made use of (2.25). Substituting (2.35) into this expression and noting that $[\tilde{x}_k,\tilde{y}_{i|i-1}] = [\tilde{x}_{k|i-1},\tilde{y}_{i|i-1}]$ we obtain

$$(2.36)\quad P_{\tilde{x}}(k|\ell) = P_{\tilde{x}}(k) - \sum_{i=k+1}^{\ell} G(k,i|i-1)[\tilde{y}_{i|i-1},\tilde{x}_{k|i-1}]$$

for $\ell > k = 0,1,2,\ldots$. For $k = \ell$, it is obvious that $\hat{x}_{k|k} = \hat{x}_k$ and $P_{\tilde{x}}(k|k) = P_{\tilde{x}}(k)$. QED

This theorem is basic in the study of discrete-time linear estimation. To the best of the author's knowledge, the results of this theorem do not exist in the literature. At this point, we

make the following remarks related to this theorem.

(2.37) REMARK. The only assumption related to the signal process is that the signal process is second order. The observation process is assumed to be second order and have full rank. The full-rank assumption can be dropped by using the generalized inverse (cf. Appendix A) instead of the inverse. If the process has constant rank r, then by using a suitable invertible transformation on the observation process one may obtain a full rank equivalent observation process.

(2.38) REMARK. To determine explicitly the estimation equations in Theorem 2.27 we need only determinethe following matrices:

$$[\tilde{x}_{k|i-1}, \tilde{y}_{i|i-1}] \quad \text{and} \quad [\tilde{y}_{i|i-1}, \tilde{y}_{i|i-1}] \, .$$

Since the process $\{\tilde{y}_{k|k-1}, k = 1,2,\ldots\}$ is a white-noise process, the computation of the second matrix above is easy, actually it is usually given as part of the problem. The determination of the first matrix is rather involved and usually is possible by making further assumptions on the signal process, such as being generated by a given linear difference equation driven by a white-noise process (Kalman filtering theory) or as in (1.4) (cf. Section 2 of Chapter 1).

CHAPTER 3

BASIC PROBLEM (BP)

This chapter is devoted to the derivation of optimal estimation equations for the basic problem (BP). In this problem the signal and observation equation are described by (1.4) and (1.5), which are repeated here for convenience

$$\text{signal:} \qquad x_k = \Phi(k)u_k, \ \{u_k, k = 0,1,2,\ldots\} \text{ is a wide-sense martingale}$$

$$\text{observation:} \qquad y_k = M(k)u_k + v_k, \ k = 1,2,\ldots$$

The assumptions on the initial signal x_o (or u_o), output noise are the same as those stated in (1.4) and (A.1.1). The matrices $\Phi(k)$ and $M(k)$ have been described in (1.4) and (1.5). The problem is simply to find the minimum mean-square estimation equations for the signal $x_k = \Phi(k)u_k$ based on the observation record $Y(\ell)$, for $k,\ell = 0,1,2,\ldots$. In view of Section 3 of Chapter 2, we need to derive the equations for the stochastic process $\{u_k, k = 0,1,2,\ldots\}$, then use (2.22a) to get the equations for the process $\{x_k = \Phi(k)u_k, k = 0,1,2,\ldots\}$.

The estimation equations are derived in the following three sections. Section 1 is devoted to the prediction problem. Where three distinct classes of prediction are defined and a recursive algorithm developed for the single-stage prediction. Sections 2 and 3, the filtering and smoothing problems are

considered, and recursive algorithms are derived for the filter and smoother. A simple example is given in Section 4 to illustrate the application of the results of the earlier sections.

We need to note here that the results of the present Chapter can be obtained from the results of the following chapter. The primary reason, however, for its separate treatment is to provide a full exposition of the new approach, with an explicit statement of the terminology, followed by the derivations of major estimation equations.

3.1 OPTIMAL PREDICTION FOR BP

We recall from Chapter 1 that in the prediction problem, we wish to obtain the optimal estimate $\hat{u}_{k|\ell}$ of the signal u_k, based on the observation record $Y(\ell) = \{y_1, y_2, \ldots, y_\ell\}$, where $k > \ell$. In other words, we wish to obtain the estimate of the signal at a time in future in terms of the existing data at the present time.

Our primary interest here is to obtain data prediction algorithms for the signal u_k. In particular, we wish to develop algorithms which are recursive in time, thereby permitting us to perform prediction efficiently with a digital computer. Before attempting this, however, it will prove expedient first to classify predicted estimates according to the possible relationship between the two time indices, k and ℓ. The need for this classification arises because both indices are variables or one may be fixed and the other may be allowed to vary. Depending on how the time indices vary, three distinct classes of prediction can be defined:

(3.1) DEFINITION. (a) Fixed-interval prediction: $\hat{u}_{k|\ell}$, $\ell = L$ = fixed positive integer, $k = L+1, L+2, \ldots$

(b) Fixed-point prediction: $\hat{u}_{k|\ell}$, $k = N$ = fixed positive integer, $\ell = 0,1,2,\ldots,N-1$.

(c) Fixed-lead prediction: $\hat{u}_{k|\ell}$, $\ell = 0,1,2,\ldots$, $k = \ell + L$ where L is a fixed positive integer.

Having introduced three distinct classes of prediction we now derive a formula for the general prediction problem, and then proceed to obtain algorithms for computing the optimal fixed-interval, fixed-point, and fixed-lead predictions.

From the orthogonal projection lemma, we know that the predicted estimate of the signal u_k based on the observation record $Y(\ell)$ which is optimal for the mean-square error risk function is

$$\hat{u}_{k|\ell} = (u_k | L_2^n(y;\ell))$$

for $k > \ell$, $k,\ell = 0,1,2,\ldots$. In addition, since the orthogonal projection is unique, all the predicted estimates are unique.

Using the fact that $u_k - u_\ell \perp L_2^n(y;\ell)$ $\forall$ $k \geq \ell$ (cf. (1.4) and (A.1.1)) we have

$$\begin{aligned} \hat{u}_{k|\ell} &= (u_k - u_\ell + u_\ell | L_2^n(y;\ell)) \\ &= (u_k - u_\ell | L_2^n(y;\ell)) + (u_\ell | L_2^n(y;\ell)) \\ &= \hat{u}_\ell \qquad k \geq \ell \end{aligned} \tag{3.2}$$

where $\hat{u}_\ell \overset{d}{=} \hat{u}_{\ell|\ell}$ is the filtered estimate. Thus the optimal predicted estimate is equal to the optimal filtered estimate at the present time. The filtered estimate $\hat{u}_\ell$ is obtained by the use

of a filtering algorithm.

The algorithm (3.2) is valid for all the classes of prediction, although the computational procedure must be altered slightly. We now consider the above three cases separately.

FIXED-INTERVAL PREDICTION. Let $\ell = L$ = fixed positive integer, then the fixed interval predicted estimate is

$$\hat{u}_{k|L} = \hat{u}_L \tag{3.3}$$

for $k = L, L+1, \ldots$. The corresponding error is by definition

$$\begin{aligned} \tilde{u}_{k|L} &\stackrel{d}{=} u_k - \hat{u}_L = u_k - u_L + u_L - \hat{u}_L \\ &= \tilde{u}_L + u_k - u_L \,, \quad k = L, L+1, \ldots \end{aligned} \tag{3.4}$$

or

$$\begin{aligned} \tilde{u}_{k|L} &= u_k - u_{k-1} + u_{k-1} - \hat{u}_{k-1|L} \,, \text{ since } \hat{u}_{k|L} = \hat{u}_{k-1|L} = \hat{u}_L \\ &= \tilde{u}_{k-1|L} + u_k - u_{k-1} \end{aligned} \tag{3.5}$$

for $k = L+1, L+2, \ldots$ with the initial condition $\tilde{u}_{L|L} = \tilde{u}_L$.

From (3.4) and (3.5) we obtain the following expressions for the covariance matrix $P_{\tilde{x}}(k|\ell)$ of the fixed-interval prediction error:

$$\begin{aligned} P_{\tilde{u}}(k|\ell) &\stackrel{d}{=} [\tilde{u}_{k|\ell}, \tilde{u}_{k|\ell}] \\ &= P_{\tilde{u}}(L) + P_u(k) - P_u(L), \text{ by using (3.4)} \\ &= P_{\tilde{u}}(L) + \sum_{i=L}^{k-1} Q(i) \,, \text{ since } Q(i) = P_u(i+1) - P_u(i) \end{aligned} \tag{3.6}$$

for $k = L, L+1, \ldots$, where $P_{\tilde{u}}(L)$ is the filtering error covariance matrix at time L; or

$$P_{\tilde{u}}(k|\ell) = P_{\tilde{u}}(k-1|L) + P_u(k) - P_u(k-1) \quad \text{, by using (3.5)}$$

$$= P_{\tilde{u}}(k-1|L) + Q(k-1) \tag{3.7}$$

for $k = L+1, L+2, \ldots$ with the initial condition $P_{\tilde{u}}(L|L) = P_{\tilde{u}}(L)$.

(3.8) REMARK. It is easily shown that (3.6) is the solution of the linear matrix difference equation (3.7) with the initial condition $P_{\tilde{u}}(L|L) = P_{\tilde{u}}(L)$.

(3.9) REMARK. In order to compute the fixed-interval estimate $\hat{u}_{k|L}$, the only value of the filtered estimate we must know is $\hat{u}_L$, where L is the present time. Hence we do not need to continue to process the filter algorithm once all the data have been received.

(3.10) REMARK. It is easily verified that the process $\{\hat{u}_{k|L},\ k = L, L+1, \ldots\}$ and $\{\tilde{u}_{k|L},\ k = L, L+1, \ldots\}$ be defined by (3.3) and (3.5), respectively, are zero mean wide-sense martingale processes. The process $\{\tilde{u}_{k|L},\ k = L, L+1, \ldots\}$ being a zero mean wide-sense martingale process implies that

$$\tilde{u}_{k|L} - \tilde{u}_{j|L} \perp \tilde{u}_{j|L} \qquad j \le k \ ,$$

so that

$$\begin{aligned} \|\tilde{u}_{k|L}\|^2 &= \|\tilde{u}_{k|L} - \tilde{u}_{j|L} + \tilde{u}_{j|L}\|^2 \\ &= \|\tilde{u}_{k|L} - \tilde{u}_{j|L}\|^2 + \|\tilde{u}_{j|L}\|^2 \\ &\ge \|\tilde{u}_{j|L}\|^2 \ . \end{aligned}$$

This result implies, as one would expect on physical grounds that the magnitude of the norm of the covariance matrix

increases with increasing k for a fixed L. On simpler terms, the error of the future estimates with a fixed observation data is larger for more distant future estimates.

(3.11) REMARK. As shown above the fixed-interval prediction $(\hat{u}_{k|L}, P_{\tilde{u}}(k|L))$, $k > L$, based on $Y(L)$ is accomplished via (3.3) and (3.6) with the initial condition $(\hat{u}_L, P_{\tilde{u}}(L))$. If we are given $(\hat{u}_L, P_{\tilde{u}}(L))$ then the fixed-interval predicted estimate $\hat{u}_{k|L}$ is obtained without calculation, in fact $\hat{u}_{k|L} = \hat{u}_L$, and the fixed-interval prediction error covariance matrix is computed through simple algebraic operations. But, the only value of L for which $(\hat{u}_L, P_{\tilde{u}}(L))$ is known without processing the filter algorithm is $L = 0$. On the other hand for $L = 0$,

$$\hat{u}_{o|o} \overset{d}{=} \hat{u}_o = \text{optimal estimate } u_o \text{ given no observation}$$
$$= \mathcal{E}\{u_o\} = 0$$

and

$$P_{\tilde{u}}(0) \overset{d}{=} [\tilde{u}_o, \tilde{u}_o]$$
$$= P_u(0) \ , \quad \text{since } \tilde{u}_o \overset{d}{=} u_o - \hat{u}_o = u_o \ .$$

So that

$$\hat{u}_{k|o} = 0 \quad \text{and} \quad P_{\tilde{u}}(k|o) = P_u(k) \ .$$

This result, though trivial, shows that the future estimates without observation is zero, and the covariance matrix of the estimate is the same as that of the signal process. Thus expressions (3.3) and (3.6) have limited practical use as far as performing the fixed-interval prediction is concerned.

FIXED-POINT PREDICTION. In the fixed-point prediction problem, we wish to obtain the optimal estimate of the signal at a given time in the future as function of the current time. Here the form of (3.2) that is of interest is

$$\hat{u}_{N|\ell} = \hat{u}_{\ell} \qquad \ell < N = \text{fixed positive integer.} \tag{3.12}$$

In this case we continue to process the filter algorithm as the data arrives as opposed to the fixed-interval prediction, where we stop processing the filter algorithm once all the data are received.

The equations of the error and the error covariance matrix of the fixed point prediction are easily obtained from their definitions as:

$$\tilde{u}_{N|\ell} = \tilde{u}_{\ell} + u_N - u_{\ell} \tag{3.13}$$

$$\begin{aligned} P_{\tilde{u}}(N|\ell) &= P_{\tilde{u}}(\ell) + P_u(N) - P_u(\ell) \\ &= P_{\tilde{u}}(\ell) + \sum_{i=\ell}^{N-1} Q(i) \end{aligned} \tag{3.14}$$

for $\ell = 0,1,2,\ldots,N-1$ with the boundary conditions $\tilde{u}_{N|N} = \tilde{u}_N$ and $P_{\tilde{u}}(N|N) = P_{\tilde{u}}(N)$.

Note that the fixed-point prediction error covariance matrix eventually is equal to the filtering error covariance matrix, since N is fixed and ℓ eventually is equal to N.

FIXED-LEAD PREDICTION. This is probably the most used case as it has application to control systems for "lead" correction action, etc. [L-I]. Here, we wish to predict the value of the signal a fixed amount of discrete-time L in the future from the current

time ℓ, i.e., we wish to predict the signal with lead L. Hence the form of (3.2) that is of interest is

$$\hat{u}_{L+\ell|\ell} = \hat{u}_{\ell} \qquad \ell = 0,1,2,... \tag{3.15}$$

So that as in the fixed-point prediction, we must continue to process the filter algorithm as the data arrives.

The error and error covariance matrix of the fixed-lead prediction, respectively, are given by the following equations:

$$\tilde{u}_{L+\ell|\ell} = \tilde{u}_{\ell} + u_{L+\ell} - u_{\ell} , \quad \ell = 0,1,2,... \tag{3.16}$$

$$\begin{aligned} P_{\tilde{u}}(L+\ell|\ell) &= P_{\tilde{u}}(\ell) + P_u(L+\ell) - P_u(\ell) \\ &= P_{\tilde{u}}(\ell) + \sum_{i=\ell}^{L+\ell-1} Q(i) , \quad \ell = 0,1,2,... \end{aligned} \tag{3.17}$$

with the initial conditions $\hat{u}_{L|o} = 0$ and $P_{\tilde{u}}(L|o) = P_u(L)$.

From (3.16) we see that the magnitude of the fixed-lead prediction error depends upon the amount of lead L. When $L = 1$, the magnitude has its smallest value. For that reason, this special case has merit attention, and is called the <u>single-stage prediction</u>. We shall obtain a recursive algorithm for computing the single-stage prediction $\hat{u}_{k+1|k}$, $k = 0,1,2,...,$. In developing the algorithm for the single-stage prediction, we assume, only the initial estimate $\hat{u}_{1|o} = 0$ and the corresponding covariance matrix of the error $P_{\tilde{u}}(1|o) = P_u(1)$ are given. The algorithm will depend only on the previous estimate $\hat{u}_{k|k-1}$ and the new observation y_k. The result is given in the following theorem. To prove the theorem, we need the following technical lemma, which will be referred to

as the innovation lemma.

(3.18) INNOVATION LEMMA. The innovation process $\{\tilde{y}_{k+1|k},\ k = 0,1,2,\ldots\}$ associated with the observation process $\{y_k,\ k = 1,2,\ldots\}$ of the BP defined by

$$y_k = M(k)u_k + v_k \qquad k = 1,2,\ldots \tag{3.19}$$

(see (A.1.1) for the meaning of the symbols) is generated by

$$\tilde{y}_{k+1|k} = y_{k+1} - M(k)\hat{u}_{k+1|k} \qquad k = 0,1,2,\ldots \tag{3.20}$$

The process $\{y_k,\ k = 1,2,\ldots\}$ has full rank, i.e.,

$$[\tilde{y}_{k+1|k},\ \tilde{y}_{k+1|k}] > 0 \quad \forall k = 0,1,2,\ldots\ .$$

PROOF. By definition

$$\begin{aligned}\tilde{y}_{k+1|k} &= y_{k+1} - (y_{k+1}|L(y;k)) \\ &= y_{k+1} - (M(k+1)u_{k+1} + v_{k+1}|L(y;k)) \\ &= y_{k+1} - M(k+1)\hat{u}_{k+1|k} - (v_{k+1}|L(y;k))\end{aligned}$$

for $k = 0,1,2,\ldots$. Since v_k, $k = 0,1,2,\ldots$ is a white-noise process, and by (A.1.1) $v_k \perp u_j$ $\forall k,j = 0,1,2,\ldots$, we have $v_{k+1} \perp L(y,k)$. So that $(v_{k+1}|L(y;k)) = 0$ by virtue of (2.10c) and therefore

$$\tilde{y}_{k+1|k} = y_{k+1} - M(k+1)\hat{u}_{k+1|k} \qquad k = 0,1,2,\ldots\ .$$

Note that for $k = 0$, $\tilde{y}_{1|o} = y_1$, since $\hat{u}_{1|o} = \mathcal{E}\{u_1\} = 0$ (cf. (3.11)).

Substituting (3.19) into (3.20) and rearranging the terms we get

$$\tilde{y}_{k+1|k} = y_{k+1} - M(k+1)\hat{u}_{k+1|k}$$
$$= M(k+1)\tilde{u}_{k+1|k} + v_{k+1} \qquad k = 0,1,2,...$$

Thus, for $k = 0,1,2,...$

$$[\tilde{y}_{k+1|k}, \tilde{y}_{k+1|k}] = [M(k+1)\tilde{u}_{k+1|k} + v_{k+1}, M(k+1)\tilde{u}_{k+1|k} + v_{k+1}]$$
$$= M(k+1)[\tilde{u}_{k+1|k}, \tilde{u}_{k+1|k}]M^T(k+1) + [v_{k+1}, v_{k+1}],$$

since $v_{k+1} \perp \tilde{u}_{k+1|k}$ $\forall$ $k = 0,1,2,...$. Noting that

$$[\tilde{u}_{k+1|k}, \tilde{u}_{k+1|k}] \overset{d}{=} P_{\tilde{u}}(k+1|k) \geq 0$$

and

$$[v_{k+1}, v_{k+1}] = P_v(k+1) > 0 \text{ , by (A.1.1) ,}$$

we obtain

(3.21) $$[\tilde{y}_{k+1|k}, \tilde{y}_{k+1|k}] = M(k+1)P_{\tilde{u}}(k+1|k)M^T(k+1) + P_v(k+1)$$

for $k = 0,1,2,...$. Since $\forall$ $k = 0,1,2,...$, $P_v(k+1) > 0$ and $P_{\tilde{u}}(k+1|k) \geq 0$, by (B.2) the inverse of $[\tilde{y}_{k+1|k}, \tilde{y}_{k+1|k}]$ exists for all $k = 0,1,2,...$. Thus the observation process has full rank. QED

(3.22) THEOREM. The single-stage prediction $(\hat{u}_{k+1|k}, P_{\tilde{u}}(k+1|k))$ $k = 0,1,2,...$ for the process $\{u_k, k = 0,1,2,...\}$ is accomplished via the following equations:

(a) The stochastic process $\{\hat{u}_{k+1|k}, k = 0,1,2,...\}$, which is defined by the single-stage prediction estimate, is a zero-mean

wide-sense martingale, and is generated by the recursive equation

$$\hat{u}_{k+1|k} = \hat{u}_{k|k-1} + G(k+1|k)[y_k - M(k)\hat{u}_{k|k-1}] \tag{3.23}$$

for $k = 1,2,\ldots$ with the initial condition $\hat{u}_{1|o} = 0$, where $G(k+1|k)$ is the $n \times m$ gain matrix and is given by the following expression:

$$G(k+1|k) = P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1} \tag{3.24}$$

$$k = 1,2,\ldots$$

(b) The stochastic process $\{\tilde{u}_{k+1|k}, k = 0,1,2,\ldots\}$, which is defined by the single-stage prediction error $\tilde{u}_{k+1|k}$, is the solution of the following stochastic linear difference equation

$$\tilde{u}_{k+1|k} = (I_n - G(k+1|k)M(k))\tilde{u}_{k|k-1} - G(k+1|k)v_k + u_{k+1} - u_k \tag{3.25}$$

for $k = 1,2,\ldots$ with the initial condition $\tilde{u}_{1|o} = u_1$. This process is a zero mean wide-sense Markov process whose covariance matrix is given by the recursive equation

$$P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k|k-1) - P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}M(k)P_{\tilde{u}}(k|k-1) + Q(k) \tag{3.26}$$

for $k = 1,2,\ldots$ with the initial condition $P_{\tilde{u}}(1|o) = P_u(1)$.

PROOF. (a) We first prove that the process $\{\hat{u}_{k+1|k}, k = 0,1,2,\ldots\}$ is a zero mean wide-sense martingale. Obviously, the process has zero-mean. To prove that it is a wide-sense martingale, we must show that (i) $\hat{u}_{k+1|k} \in L_2^n \quad \forall k = 0,1,2,\ldots$, and (ii)

$$(\hat{u}_{k+1|k} | L(\hat{u}, \ell+1|\ell)) = \hat{u}_{\ell+1|\ell} , \quad \ell \le k .$$

Since $u_k \in L_2^n$, so that $\hat{u}_{k+1|k}$ by virtue of (3.9b). It remains to verify the requirement (ii). The verification of (ii) is as follows: For $k \ge \ell$

$$\begin{aligned}
(\hat{u}_{k+1|k} | L(\hat{u},\ell+1|\ell)) &= ((u_{k+1} | L_2^n(y;k)) | L(\hat{u},\ell+1|\ell)) \\
&= (u_{k+1} | L(\hat{u},\ell+1|\ell)), \text{ by (2.10b), since} \\
&\qquad L(\hat{u},\ell+1|\ell) \subset L_2^n(y;k) \quad \text{for} \quad \ell \le k \\
&= ((u_{k+1} | L_2^n(y;\ell)) | L(\hat{u},\ell+1|\ell)), \text{ same reasoning as above} \\
&= (\hat{u}_{k+1|\ell} | L(\hat{u},\ell+1|\ell) \\
&= (\hat{u}_{\ell+1|\ell} | L(\hat{u},\ell+1|\ell) , \text{ by (3.2)} \\
&= \hat{u}_{\ell+1|\ell} , \text{ since } \hat{u}_{\ell+1|\ell} \in L(\hat{u},\ell+1|\ell) .
\end{aligned}$$

Since the observation process has full rank, by (3.18), the single-stage predicted estimate is given by

$$\hat{u}_{k+1|k} = \hat{u}_{k+1|k-1} + G(k+1|k)\tilde{y}_{k|k-1}$$

for $k = 1,2,\ldots$, which is obtained by letting $\ell = k$ in (2.28). For $k = 0$, by (3.11), $\hat{u}_{1|0} = 0$. From (3.2) and (3.20) we see that

$$\hat{u}_{k+1|k-1} = \hat{u}_{k|k-1} \quad \text{and} \quad \tilde{y}_{k|k-1} = y_k - M(k)\hat{u}_{k|k-1} ,$$

and therefore,

$$\hat{u}_{k+1|k} = \hat{u}_{k|k-1} + G(k+1|k)[y_k - M(k)\hat{u}_{k|k-1}], \quad k = 1,2,\ldots$$

where the gain matrix $G(k+1|k)$ is given by (2.29):

$$G(k+1|k) = [\tilde{u}_{k+1|k-1}, \tilde{y}_{k|k-1}][\tilde{y}_{k|k-1}, \tilde{y}_{k|k-1}]^{-1}$$

From (3.21) we know that

$$[\tilde{y}_{k|k-1}, \tilde{y}_{k|k-1}] = M(k) P_{\tilde{u}}(k|k-1) M^T(k) + P_v(k) \quad . \tag{3.21}$$

On the other hand,

$$[\tilde{u}_{k+1|k-1}, \tilde{y}_{k|k-1}] = [\tilde{u}_{k|k-1} + u_{k+1} - u_k,\ M(k)\tilde{u}_{k|k-1} + v_k],$$

by (3.19)

$$= P_{\tilde{u}}(k|k-1) M^T(k) \tag{3.27}$$

since $u_{k+1} - u_k$, $\tilde{u}_{k|k-1}$ and v_k are orthogonal.
Substituting (3.21) and (3.27) into the expression for $G(k+1|k)$ above we obtain

$$G(k+1|k) = P_{\tilde{u}}(k|k-1) M^T(k) [M(k) P_{\tilde{u}}(k|k-1) M^T(k) + P_v(k)]^{-1} \quad .$$

(b) The expression for the single-stage prediction error is obtained as follows:

$$\begin{aligned}
\tilde{u}_{k+1|k} &\stackrel{d}{=} u_{k+1} - \hat{u}_{k+1|k} \\
&= u_{k+1} - \hat{u}_{k|k-1} - G(k+1|k)[y_k - M(k)\hat{u}_{k|k-1}] \quad , \text{ by (3.23)} \\
&= \tilde{u}_{k|k-1} - G(k+1|k)[M(k)\tilde{u}_{k|k-1} + v_k] + u_{k+1} - u_k \quad , \\
&\qquad \text{by letting } u_{k+1} = u_{k+1} - u_k + u_k \text{ and substituting (3.20)} \\
&= (I_n - G(k+1|k)M(k))\tilde{u}_{k|k-1} - G(k+1|k)v_k + u_{k+1} - u_k \qquad (3.25)
\end{aligned}$$

for $k = 1,2,\ldots$. For $k = 0$, it is obvious that $\tilde{u}_{1|o} = u_1$, since $\hat{u}_{1|o} = 0$.

To show that the process $\{\tilde{u}_{k+1|k},\ k = 0,1,2,\ldots\}$, which is generated by (3.25), is a zero-mean wide-sense Markov process, define

$$F(k) \triangleq I_n - G(k+1|k)M(k),$$

$$\Gamma(k) \triangleq [-G(k+1|k),\ I_n],$$

and

$$r_k = \begin{bmatrix} v_k \\ u_{k+1} - u_k \end{bmatrix}$$

then (3.25) can be written as

$$\tilde{u}_{k+1|k} = F(k)\tilde{u}_{k|k-1} + \Gamma(k)r_k\ ,\quad k = 1,2,\ldots \tag{3.28}$$

From the definitions $F(k)$ and r_k it is clear that $F(k)\ \forall\ k = 1,2,\ldots$ is invertible and the stochastic process $\{r_k,\ k = 1,2,\ldots\}$ is a zero-mean white-noise process. Furthermore, since $\tilde{u}_{1|o} = u_1$ and therefore $\tilde{u}_{1|o} \perp r_k$, $k = 1,2,\ldots$, the process $\{\tilde{u}_{k+1|k},\ k = 0,1,\ldots\}$ is defined by (3.28) is of the same form and is subject to the same conditions as the process defined by (2.17). Hence, it is a zero-mean wide-sense Markov process.

It now remains to determine the single-stage prediction error covariance matrix $P_{\tilde{u}}(k+1|k)$. To obtain an equation for $P_{\tilde{u}}(k+1|k)$, we may multiply (3.25) with its transpose and take mathematical expectations, or let $k = \ell$ in (2.30) and then substitute (3.21) and (3.27) into it to get

$$P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k+1|k-1) - P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P(k|k-1)M^T(k) + P_v(k)]^{-1}M(k)P_{\tilde{u}}(k|k-1)$$

for $k = 1,2,\ldots$. Since, by (3.7) $P_{\tilde{u}}(k+1|k-1) = P_{\tilde{u}}(k|k-1) + Q(k)$, we have

$$(3.26)\quad P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k|k-1) - P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P(k|k-1)M^T(k) + P_v(k)]^{-1}M(k)P_{\tilde{u}}(k|k-1) + Q(k)$$

for $k = 1,2,\ldots$. For $k = 0$, obviously $P_{\tilde{u}}(1|o) = P_u(1)$, since $\tilde{u}_{1|o} = u_1$. QED

This theorem gives a recursive algorithm for computing the single-stage prediction. The recursive algorithms are given in Theorem 3.22 are extremely useful in processing observations to obtain the predicted estimate utilizing a digital computer. The observations can be processed as they occur, and there is no need to store any observation data. In fact, so far as storage of the observations and the signal is concerned, only $\hat{u}_{k|k-1}$ need to be stored in proceding from time k to time $k+1$. An additional feature is that the error covariance $P_{\tilde{u}}(k+1|k)$ is computed as a direct part of the estimator, and may be used to judge the accuracy of the estimation procedure as in the Kalman filtering theory. This is based on the assumption that the observation models and the means and covariances of related processes are correctly known.

A block diagram of the single-stage predictor is shown in Figure 3.1. The information flow in the predictor can be explained very simply by considering this block diagram, which

is a representation of (3.23). From this figure we see that single-stage predictor consists of a model discrete-time linear dynamical system $\hat{\Sigma}_p = (I_n, G(\cdot|\cdot), I_n)$ (cf. [K-5]):

$$\hat{\Sigma}_p: \quad \begin{aligned} \hat{u}_{k+1|k} &= \hat{u}_{k|k-1} + G(k+1|k)\tilde{y}_{k|k-1} \quad , \text{ state equation} \\ \hat{z}_k &= \hat{u}_{k+1|k} \quad , \text{ output equation} \end{aligned}$$

in which the gain-times-innovation term is applied to the model as a forcing function. Observe that the predictor operates in a predict-correct fashion. That is, the correction term $G(k+1|k)\tilde{y}_{k+1|k}$ is added to the predicted estimate $\hat{u}_{k|k-1}$ to determine the current predicted estimate. The correction term involves a weighting of the innovation associated with the observation process by the gain matrix $G(k+1|k)$.

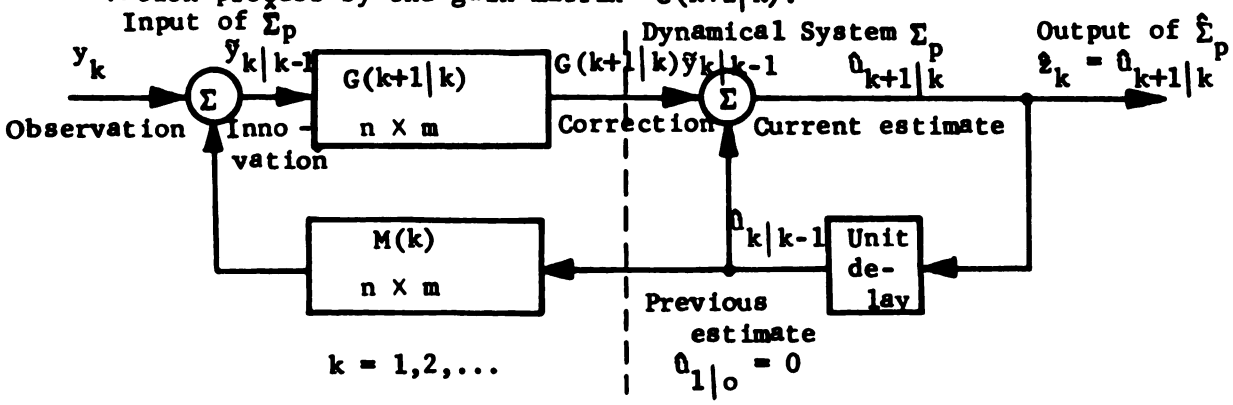


Figure 3.1. Block diagram of single-stage predictor.

The estimation equations derived above, can be used to estimate the signal $\{x_k = \Phi(k)u_k, k = 0,1,2,...\}$ as pointed out in Section 3 of Chapter 2. If we wish to have these equations involving only the estimates of the signal $\{x_k = \Phi(k)u_k, k = 0,1,2,...\}$, then as shown there we need the assumption that

$u_k \in R(\Phi^T(k))$ $\forall k = 0,1,2,\ldots$. Assuming that is so, we state the result in the following corollary for the single-stage predictor. The proof of this corollary easily follows from (2.22) and (3.22).

(3.29) COROLLARY. The single-stage prediction $(\hat{x}_{k+1|k}, P_{\tilde{x}}(k+1|k))$ $k = 0,1,2,\ldots$ for the process $\{x_k = \Phi(k)u_k, k = 0,1,2,\ldots\}$ (cf. (1.4)) such that $u_k \in R(\Phi^T(k))$ $\forall$ $k = 0,1,2,\ldots$, is accomplished via the following equations:

(a) The stochastic process $\{\hat{x}_{k+1|k}, k = 0,1,2,\ldots\}$, which is defined by the single-stage prediction estimate is a zero-mean wide-sense Markov process, and is generated by the recursive equation

$$\hat{x}_{k+1|k} = \Phi(k+1)\Phi^+(k)\hat{x}_{k|k-1} + K(k+1|k)[y_k - H(k)\hat{x}_{k|k-1}] \tag{3.30}$$

for $k = 1,2,\ldots$ with the initial condition $\hat{x}_{1|o} = 0$, where $K(k+1|k)$ is the $n \times m$ gain matrix and is given by

$$K(k+1|k) = \Phi(k+1)\Phi^+(k)P_{\tilde{x}}(k|k-1)H^T(k)[H(k)P_{\tilde{x}}(k|k-1)H^T(k) + P_v(k)]^{-1}. \tag{3.31}$$

(b) The stochastic process $\{\tilde{x}_{k+1|k}, k = 0,1,2,\ldots\}$, which is defined by the single-stage prediction error $\tilde{x}_{k+1|k}$, is the solution of the linear stochastic difference equation

$$\tilde{x}_{k+1|k} = (\Phi(k+1) - K(k+1|k)H(k))\tilde{x}_{k|k-1} - K(k+1|k)v_k + \Phi(k+1)(u_{k+1} - u_k) \tag{3.32}$$

for $k = 1,2,\ldots$, with the initial condition, $\tilde{x}_{1|o} = \Phi(1)u_1$. This process is a zero-mean wide-sense Markov process whose

covariance matrix is given by the recursive equation

$$(3.33)\quad P_{\tilde{x}}(k+1|k) = \Phi(k+1)\Phi^{+}(k)P_{\tilde{x}}(k|k-1)\Phi^{+T}(k)\Phi^{T}(k+1) -$$

$$- \Phi(k+1)\Phi^{+}(k)P_{\tilde{x}}(k|k-1)H^{T}(k)[H(k)P_{\tilde{x}}(k|k-1)H^{T}(k) +$$

$$+ P_{v}(k)]^{-1}H(k)P_{\tilde{x}}(k|k-1)\Phi^{+T}(k)\Phi^{T}(k+1) + \Phi(k+1)Q(k)\Phi^{T}(k+1)$$

for $k = 1,2,\ldots$, with the initial condition $P_{\tilde{x}}(1|o) = \Phi(1)P_{u}(1)\Phi^{T}(1)$.

(3.34) REMARK. It is shown in (2.17) that the signals of the Kalman filtering theory can be written in the form $x_k = \Phi(k)u_k$ for $k = 1,2,\ldots$ and $x_o = u_o$ where $\Phi(k) = \Phi(k,o)$ is a transition matrix. By letting $\Phi(k) = \Phi(k,o)$ and $\Phi^{+}(k) = \Phi^{-1}(k,o) = \Phi(o,k)$ in (3.29) and noting that $\Phi(k,j)\Phi(j,k) = I_n$, one obtains the results of the Kalman filtering theory (cf. [K-3], [K-4]) for the BP.

(3.35) REMARK. The algorithms given in Corollary 3.29 are not available in the current literature to the best of my knowledge, and cannot be obtained directly from the Kalman filtering theory.

(3.36) REMARK. A comparison of (3.22) and (3.29) shows that in the estimation of the signal $x_k = \Phi(k)u_k$, $k = 0,1,2,\ldots$ computation time is saved if the algorithms given in (3.22) is first used followed by (2.22a) to obtain the desired results.

3.2 OPTIMAL FILTERING FOR BP

We now examine the problem of obtaining an algorithm for computing the basic problem of interest, namely, the filtering problem. In developing the algorithm for optimal filtering for the signal $\{u_k, k = 0,1,2,\ldots\}$ and therefore for the signal

$\{x_k = \Phi(k)u_k,\ k = 0,1,2,\ldots\}$, we assume that only the initial estimate $\hat{u}_o = 0$, and the filtering error covariance matrix at the initial time, $P_{\tilde{u}}(0) = P_u(0)$, are given.

From (3.2), we observe that prediction and filtering are interdependent in terms of the determination of the predicted estimate given the filtered estimate and vice versa. In fact,

$$\hat{u}_{k+1|k} = \hat{u}_k \qquad \forall\ k = 0,1,2,\ldots$$

Hence, the single-stage predicted estimate algorithm (3.23) can be used to compute the filtered estimate. Of course, the filtering error covariance matrix will not be the same one that is given by (3.26) for the single-stage predictor error. It must be computed to judge the accuracy of the estimation procedure.

With these preliminaries completed, we now state and prove the basic theorem of optimal filtering for the signal $\{u_k,\ k = 0,1,2,\ldots\}$.

(3.37) THEOREM. The filtering $(\hat{u}_k, P_{\tilde{u}}(k))$, $k = 0,1,2,\ldots$ for the stochastic process $\{u_k,\ k = 0,1,2,\ldots\}$ is accomplished via the following equations:

(a) The stochastic process $\{\hat{u}_k,\ k = 0,1,2,\ldots\}$, which is defined by the filtered estimate, is a zero-mean wide-sense martingale. It is generated by the recursive equation

$$\hat{u}_k = \hat{u}_{k-1} + G(k)[y_k - M(k)\hat{u}_{k-1}] \tag{3.38}$$

for $k = 1,2,\ldots$ with the initial condition $\hat{u}_o = 0$, where $G(k)$ is the $n \times m$ gain matrix and is given by

$$G(k) = P_{\tilde{u}}(k)M^T(k)P_v^{-1}(k) \quad . \tag{3.39}$$

(b) The stochastic process $\{\tilde{u}_k,\ k = 0,1,2,\ldots\}$, which is defined by the filtering error $\tilde{u}_k$ that satisfies the stochastic linear difference equation

$$\tilde{u}_k = (I_n - G(k)M(k))\tilde{u}_{k-1} + [I_n - G(k)M(k)](u_k - u_{k-1}) - G(k)v_k \tag{3.40}$$

for $k = 1,2,\ldots$ with the initial condition $\tilde{u}_o = u_o$, is a zero-mean wide-sense Markov process. Its covariance matrix is given by the following recursive equation:

$$P_{\tilde{u}}(k) = P_{\tilde{u}}(k|k-1) - P_{\tilde{u}}(k)M^T(k)P_v^{-1}(k)M(k)P_{\tilde{u}}(k|k-1) \tag{3.41}$$

for $k = 1,2,\ldots$ with the initial condition $P_{\tilde{u}}(0) = P_u(0)$.

PROOF. (a) Since, by (3.2) $\hat{u}_{k+1|k} = \hat{u}_k$, it follows from (3.22a) that the process $\{\hat{u}_k,\ k = 0,1,2,\ldots\}$ is a zero-mean wide-sense martingale and

$$\hat{u}_k = \hat{u}_{k-1} + G(k+1|k)[y_k - M(k)\hat{u}_{k-1}]$$

for $k = 1,2,\ldots$. For $k = 0$, obviously $\hat{u}_o = 0$ (cf. 3.11).

The filter gain matrix $G(k)$ is obviously equal to the single-stage predictor gain matrix $G(k+1|k)$. Thus, from (3.24)

$$G(k) = G(k+1|k) = P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1} \quad .$$

To obtain an expression in terms of the filtering error covariance matrix for the gain matrix $G(k)$, note that since $P_{\tilde{u}}(k|k-1) \geq 0$ and $P_v(k) > 0$, from (B.2) and (B.1)

$$P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}$$

$$= P_{\tilde{u}}(k|k-1)M^T(k)R^{-1}(k) - P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) +$$

$$+ P_v(k)]^{-1}M(k)P_{\tilde{u}}(k|k-1)M^T(k)P_v^{-1}(k).$$

It is shown in part (b) of this theorem that

$$P_{\tilde{u}}(k) = P_{\tilde{u}}(k|k-1) - P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}$$

$$\times M(k)P_{\tilde{u}}(k|k-1) .$$

Hence

$$G(k) = P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}$$

$$= P_{\tilde{u}}(k)M^T(k)P_v^{-1}(k) ,$$

(b) The filtering error is by definition

$$\tilde{u}_k = u_k - \hat{u}_k, \quad k = 0,1,2,\ldots .$$

Substituting (3.38) for $\hat{u}_k$, and rearranging the terms we get

$$\tilde{u}_k = u_k - \hat{u}_{k-1} - G(k)[y_k - M(k)\hat{u}_{k+1}]$$

$$= \tilde{u}_{k-1} - G(k)[M(k)\tilde{u}_{k-1} + M(k)(u_k - u_{k-1}) + v_k] + u_k - u_{k-1}$$

$$= [I_n - G(k)M(k)]\tilde{u}_{k-1} + [I_n - G(k)M(k)](u_k - u_{k-1}) - G(k)v_k \tag{3.40}$$

for $k = 1,2,\ldots$. For $k = 0$, it is clear that $\tilde{u}_o = u_o$, since $\hat{u}_o = 0$.

A procedure similar to that used in (3.22b), shows that the stochastic process $\{\tilde{u}_k, k = 0,1,2,\ldots\}$, which is generated by (3.40), is a zero-mean wide-sense Markov process.

To obtain an expression for the filtering error covariance matrix $P_{\tilde{u}}(k)$, we note from (3.7) that

$$P_{\tilde{u}}(k) = P_{\tilde{u}}(k+1|k) - Q(k) .$$

Substituting (3.26) for $P_{\tilde{u}}(k+1|k)$ and noting that

$$G(k) = G(k+1|k) = P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}$$

$$= P_{\tilde{u}}(k)M^T(k)P_v^{-1}(k)$$

we obtain

$$P_{\tilde{u}}(k) = P_{\tilde{u}}(k|k-1) - G(k)M(k)P_{\tilde{u}}(k|k-1)$$

$$= P_{\tilde{u}}(k|k-1) - P_{\tilde{u}}(k)M^T(k)P_v^{-1}(k)M(k)P_{\tilde{u}}(k|k-1)$$

for $k = 1,2,\ldots$. For $k = 0$, $P_{\tilde{u}}(0) = P_u(0)$, since $\tilde{u}_o = u_o$. QED

A block diagram of the filter is shown in Figure 3.2 which is a representation of (3.38). By comparing Figures 3.1 and 3.2, we see that the single-stage predictor and filter for the stochastic process $\{u_k, k = 0,1,2,\ldots\}$ based on the same observation record have exactly the same structure:

$\hat{\Sigma}_P = (I_n, G(\cdot|\cdot), I_n)$, the single-stage predictor

$\hat{\Sigma}_F = (I_n, G(\cdot), I_n)$, the filter .

The computation, of filtering differs from the computation of single-stage prediction, in the determination of the filtering error covariance matrix. To compute the filtering error covariance matrix $P_{\tilde{u}}(k)$, one may use (3.41) or the following

equality (cf. (3.7)):

$$P_{\tilde{u}}(k) = P_{\tilde{u}}(k+1|k) - Q(k) \qquad k = 0,1,2,\ldots .$$

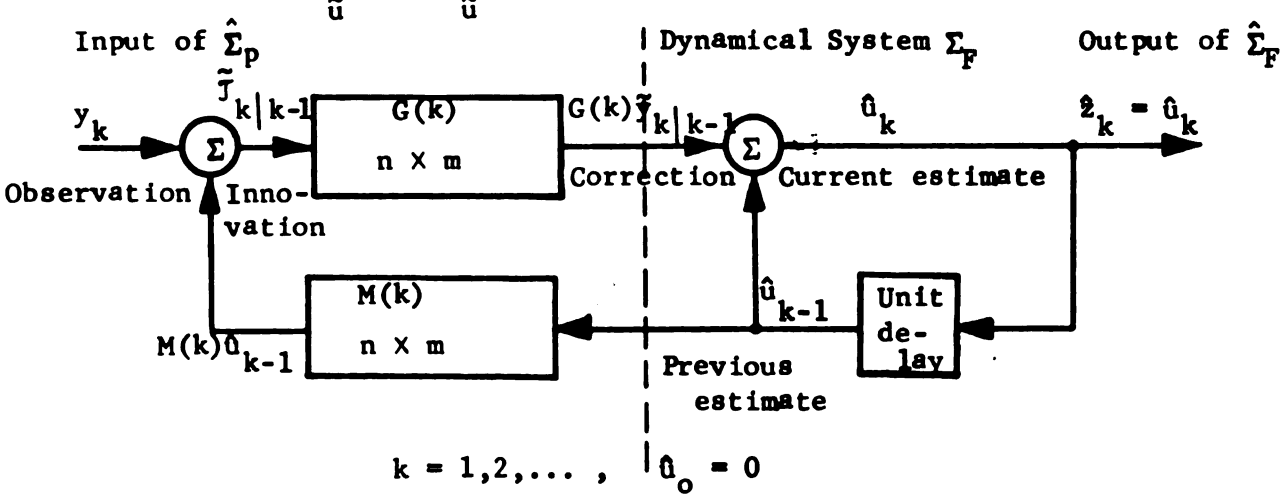


Figure 3.2. Block diagram of filter.

We now state without proof the result of Theorem (3.37) for the signal $\{x_k = \Phi(k)u_k,\ k = 0,1,2,\ldots\}$ (cf. (1.4)), assuming that $u_k \in R(\Phi^T(k))$, $k = 0,1,2,\ldots$.

(3.42) COROLLARY. The filtering $(\hat{x}_k,\ P_{\tilde{x}}(k))$, $k = 0,1,2,\ldots$ for the signal $\{x_k = \Phi(k)u_k,\ k = 0,1,2,\ldots\}$ is accomplished via the following equations:

(a) The stochastic process $\{\hat{x}_k,\ k = 0,1,2,\ldots\}$, which is defined by the single-stage prediction estimate, is a zero-mean wide-sense Markov process, and is generated by the recursive equation

$$\hat{x}_k = \Phi(k)\Phi^+(k-1)\hat{x}_{k-1} + K(k)[y_k - H(k)\Phi(k)\Phi^+(k-1)\hat{x}_{k-1}] \tag{3.43}$$

for $k = 1,2,\ldots$. The initial condition is $\hat{x}_o = 0$, where $K(k)$ is the $n \times m$ gain matrix and is given by

$$K(k) = P_{\tilde{x}}(k)H^T(k)P_v^{-1}(k) \tag{3.44}$$

(b) The stochastic process $\{\tilde{x}_k,\ k = 0,1,2,\ldots\}$, which is defined by the filtering error $\tilde{x}_k$ that satisfies the linear stochastic difference equation

$$(3.45)\quad \tilde{x}_k = [\Phi(k)\Phi^{+}(k-1) - K(k)H(k)\Phi(k)\Phi^{+}(k-1)]\tilde{x}_{k-1} + [\Phi(k) - K(k)H(k)\Phi(k)](u_k - u_{k-1}) - K(k)v_k$$

for $k = 1,2,\ldots$ with the initial condition $\tilde{x}_o = x_o = \Phi(0)u_o$, is a zero-mean wide-sense Markov process whose covariance matrix is given by the following equation:

$$(3.46)\quad P_{\tilde{x}}(k) = \Phi(k)\Phi^{+}(k-1)P_{\tilde{x}}(k-1)\Phi^{+^T}(k-1)\Phi^{T}(k) - P_{\tilde{x}}(k)H^{T}(k)P_{v}^{-1}(k)H(k)P_{\tilde{x}}(k|k-1)\Phi^{+^T}(k-1)\Phi^{T}(k)$$

for $k = 1,2,\ldots$ with the initial condition $P_{\tilde{x}}(0) = P_x(0) = \Phi(0)P_u(0)\Phi^{T}(0)$.

3.3 OPTIMAL SMOOTHING FOR BP.

Recall from Chapter 1 that the optimal smoothing problem deals with estimate of the signal at a time k based on the observation record $Y(\ell) = \{y_1, y_2, \ldots, y_\ell\}$, where $k < \ell$: that is, the time at which it is desired to estimate the signal precedes the time of the last observation y_ℓ.

Just as in the case of prediction, the smoothed estimate of a signal is classified according to possible relationships between the two time indices, ℓ and k. Depending upon how the time indices ℓ and k vary, analogous to the prediction, three classes of smoothing can be defined (cf. [M-6], [M-7]):

(3.47) DEFINITION. (a) Fixed-interval smoothing: $\hat{u}_{k|\ell}$, $\ell = L$ = fixed-positive integer, $k = 0,1,2,\ldots,L-1$.

(b) Fixed-point smoothing: $\hat{u}_{k|\ell}$, $k = N$ = fixed-positive integer, $\ell = N+1, N+2,\ldots$.

(c) Fixed-lag smoothing: $\hat{u}_{k|\ell}$, $\ell = k+L$, L = fixed-positive integer, $k = 0,1,2,\ldots$.

Before examining each of these classes separately in terms of developing algorithms, we seek a general formula to the optimal smoothing. First we recall from (3.18) that the observation process has full rank, so, by (2.27c) the optimal smoothing $(\hat{u}_{k|\ell}, P_{\tilde{u}}(k|\ell))$, $k < \ell$ is accomplished via the following equations with the initial condition $(\hat{u}_k, P_{\tilde{u}}(k))$:

$$\hat{u}_{k|\ell} = \hat{u}_k + \sum_{i=k+1}^{\ell} G(k, i|i-1)\tilde{y}_{i|i-1} ,$$

$$G(k, i|i-1) = [\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}][\tilde{y}_{i|i-1}, \tilde{y}_{i|i-1}]^{-1} ,$$

$$P_{\tilde{u}}(k|\ell) = P_{\tilde{u}}(k) - \sum_{i=k+1}^{\ell} G(k, i|i-1)[\tilde{y}_{i|i-1}, \tilde{u}_{k|i-1}]$$

where, by (3.18),

$$\tilde{y}_{i|i-1} = y_i - M(i)\hat{u}_{i|i-1}$$

and

$$[\tilde{y}_{i|i-1}, \tilde{y}_{i|i-1}] = M(i)P_{\tilde{u}}(i|i-1|)M^T(i) + P_v(i) .$$

Notice that in these equations, which are valid for all the classes of smoothing, the only unknown is the $n \times m$ matrix $[\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}]$ $i = k+1,k+2,\ldots,\ell$. This matrix is determined as follows:

v_k, $k = 0,1,2,\ldots$ is a white-noise process and $v_i \perp u_j$,

$$i,j = 0,1,2,\ldots .$$

$$\Rightarrow v_i \perp \tilde{u}_{k|i-1} \qquad i = 1,2,\ldots, \quad k = 0,1,2,\ldots$$

$$\Rightarrow [\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}] = [\tilde{u}_{k|i-1}, M(i)\tilde{u}_{i|i-1} + v_i]$$

$$= [\tilde{u}_{k|i-1}, \tilde{u}_{i|i-1}]M^T(i)$$

$$= P_{\tilde{u}}(k, i|i-1)M^T(i) \tag{3.48}$$

where $P_{\tilde{u}}(k, i|i-1) \overset{d}{=} [\tilde{u}_{k|i-1}, \tilde{u}_{i|i-1}]$. The problem is now to find an expression for the cross-covariance matrix $P_{\tilde{u}}(k,i|i-1)$.

Noting that

$$\tilde{u}_{k|i} \overset{d}{=} u_k - \hat{u}_{k|i}$$

$$= \tilde{u}_k - \sum_{j=k+1}^{i} G(k, j|j-1)\tilde{y}_{j|j-1}, \text{ for } k < i \text{ by (2.27c)}$$

$$= \tilde{u}_{k|i-1} - G(k, i|i-1)\tilde{y}_{i|i-1}, \quad k < i$$

and

$$\tilde{u}_{i+1|i} \overset{d}{=} u_i - \hat{u}_{i+1|i}$$

$$= \tilde{u}_{i|i-1} - G(i+1|i)\tilde{y}_{i|i-1} + u_{i+1} - u_i \quad \text{by (3.25)}$$

we have

$$P_{\tilde{u}}(k, i+1|i) \overset{d}{=} [\tilde{u}_{k|i}, \tilde{u}_{i+1|i}]$$

$$= [\tilde{u}_{k|i-1} - G(k, i|i-1)\tilde{y}_{i|i-1}, \tilde{u}_{i|i-1} - G(i+1|i)\tilde{y}_{i|i-1} + u_{i+1} - u_i]$$

$$= [\tilde{u}_{k|i-1}, \tilde{u}_{i|i-1}] - G(k,i|i-1)[\tilde{y}_{i|i-1}, \tilde{u}_{i|i-1}]$$

$$- [\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}]G^T(i+1|i) +$$

$$+ G(k,i|i-1)[\tilde{y}_{i|i-1}, \tilde{y}_{i|i-1}]G^T(i+1|i)$$

for $i = k+1, k+2, \ldots$. In obtaining the last equality we used the fact that

$$u_{i+1} - u_i \perp \tilde{u}_{k|i-1},\ \tilde{y}_{i|i-1} \quad \text{for} \quad k < i$$

which is an easy consequence of (1.4) and (A.1.1). Substituting (2.34) for $G(k, i|i-1)$ and (2.29) for $G(i+1|i)$ into this last equation, and performing the necessary operations we obtain

$$P_{\tilde{u}}(k, i+1|i) = [\tilde{u}_{k|i-1},\ \tilde{u}_{i|i-1}] - [\tilde{u}_{k|i-1},\ \tilde{y}_{i|i-1}]$$
$$[\tilde{y}_{i|i-1},\ \tilde{y}_{i|i-1}]^{-1}[\tilde{y}_{i|i-1},\ \tilde{u}_{i|i-1}]$$

for $i = k+1, k+2, \ldots$. Since

$$[\tilde{u}_{k|i-1},\ \tilde{u}_{i|i-1}] \overset{d}{=} P_{\tilde{u}}(k, i|i-1),$$

$$[\tilde{y}_{i|i-1}, \tilde{y}_{i|i-1}] = M(i)P_{\tilde{u}}(i|i-1)M^T(i) + P_v(i) \quad \text{by (3.18)},$$

$$[\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}] = P_{\tilde{u}}(k, i|i-1)M^T(i) \quad \text{by (3.48)},$$

$$[\tilde{y}_{i|i-1}, \tilde{u}_{i|i-1}] = M(i)P_{\tilde{u}}(i|i-1) \quad \text{by (3.27)}$$

we can express $P_{\tilde{u}}(k, i|i-1)$ as

$$P_{\tilde{u}}(k, i+1|i) = P_{\tilde{u}}(k, i|i-1) - P_{\tilde{u}}(k, i|i-1)M^T(i)[M(i)P_{\tilde{u}}(i|i-1)M^T(i) +$$
$$+ P_v(i)]^{-1}M(i)P_{\tilde{u}}(i|i-1)$$

for $i = k, k+1, \ldots$ with $P_{\tilde{u}}(k, k|k-1) = P_{\tilde{u}}(k|k-1)$ as the initial condition.

This completes our derivation of a general formula for optimal smoothing, and we summarize our results below.

(3.49) THEOREM. Optimal smoothing $(\hat{u}_{k|\ell}, P_{\tilde{u}}(k|\ell))$, $k < \ell$ for the signal u_k, $k = 0,1,2,\ldots$ is accomplished via the following equations with $(\hat{u}_k, P_{\tilde{u}}(k))$ as the initial condition:

(a) Optimal smoothed estimate is given by the algebraic equation

$$\hat{u}_{k|\ell} = \hat{u}_\ell + \sum_{i=k+1}^{\ell} G(k, i|i-1)[y_i - M(i)\hat{u}_{i|\ell-1}] \tag{3.50}$$

where the $n \times m$ matrix $G(k, i|i-1)$ is

$$G(k, i|i-1) = P_{\tilde{u}}(k, i|i-1)M^T(i)[M(i)P_{\tilde{u}}(i|i-1)M^T(i) + P_v(i)]^{-1}. \tag{3.51}$$

(b) The $n \times n$ cross-covariance matrix $P_{\tilde{u}}(k, i|i-1)$ satisfies the recursion

$$P_{\tilde{u}}(k,i+1|i) = P_{\tilde{u}}(k,i|i-1) - P_{\tilde{u}}(k,i|i-1)M^T(i)[M(i)P_{\tilde{u}}(i|i-1)M^T(i) + P_v(i)]^{-1}M(i)P_{\tilde{u}}(i|i-1) \tag{3.52}$$

for $i = k,k+1,1,\ldots$ with the initial condition $P_{\tilde{u}}(k, k|k-1) = P_{\tilde{u}}(k|k-1)$.

(c) Optimal smoothing error covariance matrix is given by the algebraic equation

$$P_{\tilde{u}}(k|\ell) = P_{\tilde{u}}(k) - \sum_{i=1}^{\ell} G(k, i|i-1)M(i)P_{\tilde{u}}(i, k|i-1) . \tag{3.53}$$

This theorem provides a solution to general optimal smoothing problem for the signal u_k. It is an easy task to show that the optimal smoothing $(\hat{x}_{k|\ell}, P_{\tilde{x}}(k|\ell))$, for the signal $x_k = \Phi(k)u_k$ such that $u_k \in R(\Phi^T(k))$, is accomplished via Theorem 3.49 where u is replaced by x and M is replaced by H. Thus the optimal smoothing equations for the Kalman signal (cf.

(2.17)) are exactly those given in (3.49) with x replaced by u and H is replaced by M. This solution, for discrete-time, Kalman signals, is not known at the present time to the best of the author's knowledge. (For previous works in this area see e.g. [H-6], [K-2], [M-5], [K-1], [W-5].)

The optimal estimation equations given in Theorem 3.49 are obviously valid for all the classes of optimal smoothing. They may be written in different forms for each class as in the case of optimal prediction. In the following we shall give the forms of these equations for the single-stage smoothing (fixed-lag smoothing with lag 1) and fixed-point smoothing.

SINGLE-STAGE SMOOTHING. Here we wish to obtain the optimal estimate $\hat{u}_{k|k+1}$ for $k = 0,1,2,\ldots$. By letting $\ell = k+1$ in (3.49) and noting (3.2), (3.37) we get the results:

$$\hat{u}_{k|k+1} = \hat{u}_k + G(k, k+1|k)[y_{k+1} - M(k+1)\hat{u}_k],$$

$$G(k, k+1|k) = P_{\tilde{u}}(k)M^T(k+1)[M(k+1)P_{\tilde{u}}(k+1|k)M^T(k+1) + P_v(k+1)]^{-1}$$

$$P_{\tilde{u}}(k, k+1|k) = P_{\tilde{u}}(k) ,$$

$$P_{\tilde{u}}(k|k+1) = P_{\tilde{u}}(k) - G(k, k+1|k)M(k+1)P_{\tilde{u}}(k)$$

for $k = 0,1,2,\ldots$ with the initial conditions $\hat{u}_{o|o} = \hat{u}_o = 0$, $P_{\tilde{u}}(o|o) = P_u(0)$.

FIXED-POINT SMOOTHING. Recall from (3.47b) that the optimal fixed-point smoothing problem deals with the estimate $\hat{u}_{N|\ell}$ where N = positive-integer and $\ell = N+1,N+2,\ldots$. Let $k = N$ be fixed in (3.49), then from (3.50) it is seen that

$$\hat{u}_{N|\ell} = \hat{u}_N - \sum_{i=N+1}^{\ell-1} G(N,\, i|i-1)[y_i - M(i)\hat{u}_{i|i-1}]$$

$$+ G(N,\, \ell|\ell-1)[y_\ell - M(\ell)\hat{u}_{\ell|\ell-1}]$$

$$= \hat{u}_{N|\ell-1} + G(N,\, \ell|\ell-1)[y_\ell - M(\ell)\hat{u}_{\ell|\ell-1}] \tag{3.54}$$

where, by (3.51),

$$G(N,\ell|\ell-1) = P_{\tilde{u}}(N,\ell|\ell-1)M^T(\ell)[M(\ell)P_{\tilde{u}}(\ell|\ell-1)M^T(\ell) + P_v(\ell)]^{-1} \tag{3.55}$$

where $P_{\tilde{u}}(N,\, \ell|\ell-1)$ is given by (3.52).

To complete the derivation of optimal fixed-point smoothing equations we need to find an expression for the smoothing error-covariance matrix $P_{\tilde{u}}(N|\ell)$. From (3.53) it follows that the covariance matrix satisfies the recursion

$$P_{\tilde{u}}(N|\ell) = P_{\tilde{u}}(N|\ell-1) - G(N,\, \ell|\ell-1)M(\ell)P_{\tilde{u}}(\ell,\, k|\ell-1) \tag{3.56}$$

for $\ell = N+1, N+2, \ldots$ with $P_{\tilde{u}}(N|N) = P_{\tilde{u}}(N)$ as the initial condition.

Thus we have found that the optimal fixed-point smoothing $(\hat{u}_{N|\ell},\, P_{\tilde{u}}(N|\ell))$, $\ell = N+1, N+2, \ldots$ for the signal u_k is accomplished via the recursive equations (3.54), (3.55), (3.52) and (3.56), with the initial condition $(\hat{u}_N,\, P_{\tilde{u}}(N))$. When equations are written for a Kalman signal, i.e. u is replaced by x (cf. (2.17)) and M is replaced by H, one obtains a new procedure in smoothing of Kalman signals.

3.4 AN EXAMPLE

Let us consider a scalar stochastic process $\{x_k, k = 0,1,2,\ldots\}$, which has zero mean and whose covariance function is $P_x(k,j) = \sigma e^{-(k+i)}$, where σ = constant > 0. Suppose that we observe this process in the presence of a zero mean white noise process $\{v_k, k = 1,2,\ldots\}$ for which $\mathcal{E}\{v_k v_j\} = r(k)\delta_{kj}$ and $\mathcal{E}\{v_k x_j\} = 0 \quad \forall\, k,j$, where $0 \le r(k) < \infty \quad \forall\, k = 1,2,\ldots$. Then the observation equation is

$$y_k = x_k + v_k \qquad k = 1,2,\ldots \tag{3.57}$$

Since only the first and second moments of the signal process $\{x_k, k = 0,1,2,\ldots\}$ are given, we attempt to determine a wide-sense Markov process with the same properties. To do so, in view of (2.16) it is sufficient to show that $x_k = \Phi(k)u_k$, $k = 0,1,2,\ldots$, where $\Phi(k)$ is a scalar function of discrete-time, and $\{u_k, k = 0,1,2,\ldots\}$ is a wide-sense martingale process. If this is so, then (cf. [M-1])

$$\begin{aligned}\Phi(k) &= P_x(k,o)P_x^{-1}(o,o) \qquad k \ge 0\\ &= e^{-k}.\end{aligned}$$

Thus, assuming $x_k = e^{-k}u_k \quad \forall\, k = 0,1,2,\ldots$ we obtain

$$e^{-k-i}\sigma \overset{d}{=} [x_k,x_j] = e^{-k-i}P_u(k,i)$$

$$\Rightarrow P_u(k,i) = \sigma = \text{constant} > 0 \ .$$

Since $[u_k - u_i, u_i] = \sigma - \sigma = 0 \quad \forall\, k,i = 0,1,2,\ldots$, the stochastic process $\{u_k, k = 0,1,2,\ldots\}$ is a wide-sense martingale with zero mean and constant covariance function σ. Hence,

$x_k = e^{-k}u_k$, and it is a wide-sense Markov process.

Now, we are ready to compute the optimal estimation equations for the process $\{x_k, k = 0,1,2,\ldots\}$ which is observed by (3.57). We first note that since $P_u(k) = \sigma$ = constant, $Q(k) \stackrel{d}{=} P_u(k+1) - P_u(k) = 0$ (cf. (1.4)). Therefore it follows from (3.17) that

$$P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k) \quad \forall k = 0,1,2,\ldots . \tag{3.58}$$

Hence, the optimal single-stage prediction and filtering for the process $\{u_k, k = 0,1,2,\ldots\}$ are accomplished via the same set of equations (cf. (3.22), (3.37)). Note that for the process $\{x_k, k = 0,1,2,\ldots\}$ we have (cf. (2.22), (3.58))

$$\begin{aligned} P_{\tilde{x}}(k+1|k) &= e^{-2(k+1)}e^{+2k}P_{\tilde{x}}(k) \\ &= e^{-2}P_{\tilde{x}}(k) \qquad \forall k = 0,1,2,\ldots \end{aligned} \tag{3.59}$$

We summarize the results below. The computations here are exceedingly simple hence the derivations of these results will not be demonstrated in detail.

OPTIMAL PREDICTION. In the following, N and L denote fixed-positive integers

(a) Fixed-interval prediction (cf. (3.3), (3.6)): For $k = L+1, L+2, \ldots$

$$\left.\begin{aligned} \hat{u}_{k|L} &= \hat{u}_L , \\ P_{\tilde{u}}(k|L) &= P_{\tilde{u}}(L). \end{aligned}\right\} \quad \Rightarrow \quad \begin{aligned} \hat{x}_{k|L} &= e^{-(k-L)}\hat{x}_L \\ P_{\tilde{x}}(k|L) &= e^{-2(k-L)}P_{\tilde{x}}(L) \end{aligned}$$

(b) Fixed-point prediction (cf. (3.13), (3.14)):

$$\left.\begin{aligned} \hat{u}_{N|\ell} &= \hat{u}_{\ell}\ , \\ P_{\tilde{u}}(N|\ell) &= P_{\tilde{u}}(\ell)\ . \end{aligned}\right\} \Rightarrow \begin{aligned} \hat{x}_{N|\ell} &= e^{-(N-\ell)}\hat{x}_{\ell} \\ P_{\tilde{x}}(N|\ell) &= e^{-2(N-\ell)}P_{\tilde{x}}(\ell) \end{aligned}$$

(c) Fixed-lead prediction (cf. (3.15), (3.17)): For $\ell = 0,1,2,\ldots$

$$\left.\begin{aligned} \hat{u}_{L+\ell|\ell} &= \hat{u}_{\ell}\ , \\ P_{\tilde{u}}(L+\ell|\ell) &= P_{\tilde{u}}(\ell)\ . \end{aligned}\right\} \Rightarrow \begin{aligned} \hat{x}_{L+\ell|\ell} &= e^{-L}\hat{x}_{\ell}\ , \\ P_{\tilde{x}}(L+\ell|\ell) &= e^{-2L}P_{\tilde{x}}(\ell)\ . \end{aligned}$$

OPTIMAL SINGLE-STAGE PREDICTION AND FILTERING.

(a) Optimal estimate: From (3.38) (or (3.23)) we have

$$\hat{u}_k = \hat{u}_{k-1} + G(k)[y_k - e^{-k}\hat{u}_{k-1}] \Rightarrow \hat{x}_k = e^{-1}\hat{x}_{k-1} + K(k)[y_k - e^{-1}\hat{x}_{k-1}]$$

for $k = 1,2,\ldots$ with $\hat{u}_o = 0$.

(b) Gain matrix $G(k)$: From (3.39) (or (3.38))

$$G(k) = \frac{P_{\tilde{u}}(k)e^{-k}}{r(k)} \quad \Rightarrow \quad K(k) = \frac{P_{\tilde{x}}(k)}{r(k)}$$

(c) Error covariance matrix: From (3.41) and (3.58)

$$\begin{aligned} P_{\tilde{u}}(k+1|k) &= P_{\tilde{u}}(k) \\ &= P_{\tilde{u}}(k-1) - \frac{P_{\tilde{u}}(k-1)e^{-2k}}{e^{-2k}P_{\tilde{u}}(k-1)+r(k)}\ P_{\tilde{u}}(k-1) \\ &= \frac{r(k)P_{\tilde{u}}(k-1)}{e^{-2k}P_{\tilde{u}}(k-1)+r(k)} \end{aligned} \tag{3.60}$$

for $k = 1,2,\ldots$ with $P_{\tilde{u}}(0) = \sigma$. It is easily shown that (3.60) with the initial condition $P_{\tilde{u}}(0) = \sigma$ has the unique solution

$$(3.61)\qquad P_{\tilde{u}}(k) = \frac{1}{\sum_{i=1}^{k} \frac{e^{-2i}}{r(i)} + \frac{1}{\sigma}} \qquad k = 0,1,2,\ldots .$$

Hence, from (2.22) we have

$$(3.62)\quad P_{\tilde{x}}(k) = \frac{e^{-2k}}{\sum_{i=1}^{k} \frac{e^{-2(i-k)}}{r(i)} + \frac{e^{2k}}{\sigma}} = \frac{1}{\sum_{i=1}^{k} \frac{e^{-2(i-k)}}{r(i)} + \frac{e^{2k}}{\sigma}} \qquad k = 0,1,2,\ldots .$$

Notice that

$$\text{(b), (3.61) and (3.62)} \Rightarrow \begin{cases} G(k) = \dfrac{1}{r(k)\left(\sum_{i=1}^{k} \frac{e^{-(2i-k)}}{r(i)} + \frac{e^{k}}{\sigma}\right)} , \\ K(k) = \dfrac{1}{r(k)\left(\sum_{i=1}^{k} \frac{e^{-2(i-k)}}{r(i)} + \frac{e^{2k}}{\sigma}\right)} . \end{cases}$$

OPTIMAL SMOOTHING. We shall only derive the optimal estimation equations for general smoothing (cf. (3.49)). We see from the equations given (3.49) that the only quantity that needs to be determined is $P_{\tilde{u}}(k, j|j-1)$, the other quantities in these equations are computed above for the filtering.

From (3.52) and (3.58) we have

$$P_{\tilde{u}}(k,j+1|j) = P_{\tilde{u}}(k,j|j-1)\left[1 - \frac{e^{-2j}P_{\tilde{u}}(j-1)}{e^{-2j}P_{\tilde{u}}(j-1)+r(j)}\right]$$

$$= P_{\tilde{u}}(k,j|j-1)\,\frac{r(j)}{e^{-2j}P_{\tilde{u}}(j-1)+r(j)}$$

for $j = k,k+1,\ldots$, with the initial condition $P_{\tilde{u}}(k,k|k-1) = P_{\tilde{u}}(k|k-1)$. It is shown that the unique solution of this equation is

$$P_{\tilde{u}}(k,j|j-1) = \frac{P_{\tilde{u}}(k-1)\prod_{i=1}^{j-1} r(i)}{\prod_{i=k}^{j-1}\left[e^{-2i}P_{\tilde{u}}(i-1)+r(i)\right]} \qquad j \geq k+1 .$$

For the stochastic process $\{x_k,\ k = 0,1,2,\ldots\}$ this equation becomes

$$P_{\tilde{x}}(k,j|j-1) = \frac{e^{-(j+1)} P_{\tilde{x}}(k-1) \prod_{i=1}^{j-1} r(i)}{\prod_{i=k}^{j-1} [P_{\tilde{x}}(i-1)+r(i)]} \qquad j \geq k+1 \quad .$$

This completes our discussion on the example.

We note here that the above solutions for this simple example can be obtained by other techniques as well. One will obtain the same result.

CHAPTER 4

CROSS-CORRELATED NOISE PROBLEM (CCP)

In this chapter, we derive the optimal estimation equations for the cross-correlated noise problem (A.1.2), that is, the estimation equations for the process $\{x_k = \Phi(k)u_k,\ k = 0,1,\ldots\}$ where the process $\{u_{k+1} - u_k,\ k = 0,1,2,\ldots\}$ and output noise $\{v_k,\ k = 0,1,2,\ldots\}$ are correlated (cf. Assumption A.1.2). As we did in the preceeding chapter, we shall derive the equations for the process $\{u_k,\ k = 0,1,2,\ldots\}$ then use (2.22a) to obtain equations for the process $\{x_k = \Phi(k)u_k,\ k = 0,1,\ldots\}$.

4.1 OPTIMAL PREDICTION FOR CCP

We first derive a general formula of the prediction for the problem of interest then seek the form of this formula for the three distinct classes of the prediction problems introduced in Chapter 3. The general formula is derived easily by using the orthogonal projection lemma as follows:

The linear minimum mean-square predicted estimate of the signal u_k based on the observation record $Y(\ell)$ is

$$\hat{u}_{k|\ell} = (u_k | L_2^n(y;\ell)) \qquad k > \ell$$

where $k,\ell = 0,1,2,\ldots$. By adding and subtracting the term $u_{\ell+1}$ to u_k and using the linearity of the orthogonal projection we get

$$\hat{u}_{k|\ell} = (u_k - u_{\ell+1}|L_2^n(y;\ell)) + (u_{\ell+1}|L_2^n(y;\ell))$$

$k > \ell$. From the assumptions made in (1.4), (A.1.2) we have

$$u_k - u_{\ell+1} \perp u_\ell \quad \text{and} \quad u_k - u_{\ell+1} \perp v_\ell \ ,$$

for all $k > \ell$. Therefore $u_k - u_{\ell+1} \perp L_2^n(y;\ell)$ and

$$\hat{u}_{k|\ell} = (u_{\ell+1}|L_2^n(y;\ell))$$

$$= \hat{u}_{\ell+1|\ell} \tag{4.1}$$

for $k > \ell$. From this result we observe that the general predicted estimate is equal to the single-stage predicted estimate. So we need to develop an algorithm for the single-stage predicted estimate. From this algorithm we may obtain the three classes of prediction by using (4.1) as shown below.

We now state the predictor (4.1) and the corresponding covariance matrices of the estimation errors for the classes of prediction, that is, the equations for prediction $(\hat{u}_{k|\ell}, P_{\tilde{u}}(k|\ell))$ $k > \ell$ are developed for the three distinct classes. The derivation of these equations is straightforward and will not be demonstrated here.

FIXED-INTERVAL PREDICTION. Fixed-interval prediction $(\hat{u}_{k|L}, P_{\tilde{u}}(k|L))$ $k > L$, L = fixed positive integer, is accomplished via the following equations with the initial condition $(\hat{u}_{L+1|L}, P_{\tilde{u}}(L+1|L))$:

$$\hat{u}_{k|L} = \hat{u}_{L+1|L} \ , \tag{4.2}$$

$$(4.3)\qquad P_{\tilde{u}}(k|L) = P_{\tilde{u}}(L+1|L) + P_u(k) - P_u(L+1), \quad k \geq L+1$$
$$= P_{\tilde{u}}(L+1|L) + \sum_{i=L+1}^{k-1} Q(i).$$

FIXED-POINT PREDICTION. Fixed-point prediction $(\hat{u}_{k|\ell}, P_{\tilde{u}}(N|\ell))$ $\ell = 0,1,\ldots,N-1$, N = fixed positive integer, is accomplished via the following equations with the boundary condition $(\hat{u}_N, P_{\tilde{u}}(N))$:

$$(4.4)\qquad \hat{u}_{N|\ell} = \hat{u}_{\ell+1|\ell},$$

$$(4.5)\qquad P_{\tilde{u}}(N|\ell) = P_{\tilde{u}}(\ell+1|\ell) + P_u(N) - P_u(\ell+1).$$

FIXED-LEAD PREDICTION. Fixed-lead prediction $(\hat{u}_{\ell+L|\ell}, P_{\tilde{u}}(\ell+L|\ell))$ $\ell = 0,1,2,\ldots$ with lead L = fixed positive integer, is accomplished via the following equations with the initial condition $(\hat{u}_{L|o}, P_{\tilde{u}}(L|o)) = (0, P_u(L))$ (cf. Remark (3.10) of Chapter 2):

$$(4.6)\qquad \hat{u}_{\ell+L|\ell} = \hat{u}_{\ell+1|\ell},$$

$$(4.7)\qquad P_{\tilde{u}}(\ell+L|\ell) = P_{\tilde{u}}(\ell+1|\ell) + P_u(\ell+L) - P_u(\ell+1).$$

Note that in order to compute the above three classes of prediction we need to know the fixed-lead prediction with lead one, i.e. the single-stage prediction. In the fixed-interval prediction, the only value of the single-stage estimate we must know is $\hat{u}_{L+1|L}$, where L is the end-point of the observation interval and it is fixed. For the fixed-point and fixed-lead prediction we need to process the single-stage prediction algorithm as the data arrives.

In the following,wwe develop a recursive algorithm for single-stage prediction based on the previous estimate and the new observation. To do this we need the innovation lemma for the observation process defined by (1.5) and (A.1.2).

(4.8) INNOVATION LEMMA. The observation process $\{y_k,\ k = 1,2,\ldots\}$ defined by (1.5) and (A.1.2) has full rank, i.e.,

$$[\tilde{y}_{k+1|k},\tilde{y}_{k+1|k}] = M(k+1)P_{\tilde{u}}(k+1|k)M^T(k+1) + P_v(k+1) > 0$$

for all $k = 0,1,2,\ldots$.

PROOF. By definition

$$\tilde{y}_{k+1|k} \stackrel{d}{=} y_{k+1} - (y_{k+1}|L(y;k)),$$

Since $y_k = M(k)u_k + v_k$, we have

$$\tilde{y}_{k+1|k} = y_{k+1} - M(k+1)\hat{u}_{k+1|k} + \hat{v}_{k+1|k},\ k = 0,1,2,\ldots$$

Recall from (1.4) and (A.1.2) that v_j, $j = 0,1,2,\ldots$ is a white-noise process such that

$$[u_{k+1} - u_k, v_j] = C(k)\delta_{kj} \quad \text{and} \quad u_o \perp v_j$$

for all $k,j = 0,1,2,\ldots$. It follows that

$$[u_k, v_j] = 0 \quad \forall\ k \le j,\ k,j = 0,1,2,\ldots . \tag{4.9}$$

Thus $v_{k+1} \perp L(y;k)$ and therefore $(v_{k+1}|L(y;k)) = 0$. Then the innovation vector is

$$\tilde{y}_{k+1|k} = y_{k+1} - M(k+1)\hat{u}_{k+1|k} = M(k+1)\tilde{u}_{k+1|k} + v_{k+1},\ k = 0,1,2,\ldots \tag{4.10}$$

Using the expression (4.10) for the innovation vector we obtain

$$[\tilde{y}_{k+1|k},\tilde{y}_{k+1|k}] = [M(k+1)\tilde{u}_{k+1|k} + v_{k+1}, M(k+1)\tilde{u}_{k+1|k} + v_{k+1}]$$

$$= M(k+1)P_{\tilde{u}}(k+1|k)M^T(k+1) + P_v(k+1)$$

for $k = 0,1,2,\ldots$. Since $P_{\tilde{u}}(k+1|k) \geq 0$ and by (A.1.2), $P_v(k+1) > 0$ the matrix

$$[\tilde{y}_{k+1|k}, \tilde{y}_{k+1|k}] > 0 \quad \forall\, k = 0,1,2,\ldots$$

as desired. QED

(4.11) THEOREM. The single-stage prediction $(\hat{u}_{k+1|k}, P_{\tilde{u}}(k+1|k))$ $k = 0,1,2,\ldots$ of the signal $\{u_k, k = 0,1,2,\ldots\}$ is accomplished as follows:

(a) The stochastic process $\{\hat{u}_{k+1|k}, k = 0,1,2,\ldots\}$, which is defined by the single-stage predicted estimate $\hat{u}_{k+1|k}$,

$$(4.12) \quad \hat{u}_{k+1|k} = \hat{u}_{k|k+1} + G(k+1|k)[y_k - M(k)\hat{u}_{k|k-1}]$$

for $k = 1,2,\ldots$ with $\hat{u}_{1|o} = 0$ as the initial condition, is a zero-mean wide-sense martingale. The predictor gain matrix $G(k+1|k)$ is

$$(4.13) \quad G(k+1|k) = [P_{\tilde{u}}(k|k-1)M^T(k) + C(k)][M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}.$$

(b) The stochastic process $\{\tilde{u}_{k+1|k}, k = 0,1,2,\ldots\}$, which is defined by the single-stage prediction error $\tilde{u}_{k+1|k}$ satisfying the linear stochastic difference equation

$$(4.14) \quad \tilde{u}_{k+1|k} = [I_n - G(k+1|k)M(k)]\tilde{u}_{k|k-1} - G(k+1|k)v_k + u_{k-1} - u_k$$

for $k = 1,2,\ldots$ with the initial condition $\tilde{u}_{1|o} = u_1$, is a wide-sense Markov process. The covariance matrix for this process is determined by the recursive equation

$$(4.15)\quad P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k|k-1) - [P_{\tilde{u}}(k|k-1)M^T(k) + C(k)][M(k)P_{\tilde{u}}(k|k-1)M^T(k) + R(k)]^{-1}[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + C^T(k)] + Q(k)$$

for $k = 1,2,\ldots$ with $P_{\tilde{u}}(1|0) = P_u(1)$ as the initial condition.

PROOF. (a) The proof of the fact that the process $\{\hat{u}_{k+1|k}, k = 0,1,2,\ldots\}$ is a zero-mean wide-sense martingale is similar to the one that is given in (3.22a) and hence omitted.

From (2.27a), (4.1) and (4.10) it is seen that,

$$\hat{u}_{k+1|k} = \hat{u}_{k|k-1} + G(k+1|k)[y_k - M(k)\hat{u}_{k|k-1}]$$

for $k = 1,2,\ldots$, and $\hat{u}_{1|o} = 0$ for $k = 0$, since by (4.8) the observation process has full rank. Now, it remains to find an expression for the predictor gain matrix $G(k+1|k)$. From (2.29) and (4.8), we have

$$G(k+1|k) = [\tilde{u}_{k+1|k-1}, \tilde{y}_{k|k-1}][\tilde{y}_{k|k-1}, \tilde{y}_{k|k-1}]^{-1}$$

$$= [\tilde{u}_{k+1|k-1}, \tilde{y}_{k|k-1}][M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}.$$

To complete the determination of the gain matrix we must compute the matrix $[\tilde{u}_{k+1|k-1}, \tilde{y}_{k|k-1}]$. This is done as follows:

$$
\begin{aligned}
[\tilde{u}_{k+1|k-1},\ \tilde{y}_{k|k-1}] &= [u_{k+1} - \hat{u}_{k+1|k-1},\ \tilde{y}_{k|k-1}] \\
&= [u_{k+1},\ \tilde{y}_{k|k-1}], \text{ since } \tilde{y}_{k|k-1} \perp \hat{u}_{k+1|k-1} \in L_2^n(y;k-1) \\
&= [u_{k+1},\ M(k)\tilde{u}_{k|k-1}] + [u_{k+1},\ v_k], \text{ by (4.10)} \\
&= [u_k,\ \tilde{u}_{k|k-1}]M^T(k) + [u_{k+1} - u_k,\ v_k], \text{ by (4.9)} \\
&= P_{\tilde{u}}(k|k-1)M^T(k) + C(k), \text{ since } u_k = \tilde{u}_{k|k-1} + \hat{u}_{k|k-1} \\
&\qquad \text{and } \tilde{u}_{k|k-1} \perp \hat{u}_{k|k-1}\,. \qquad (4.16)
\end{aligned}
$$

Substitution of this result into the expression for the predictor gain matrix, yields

$$G(k+1,k) = [P_{\tilde{u}}(k|k-1)M^T(k) + C(k)][M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}.$$

(b) By definition

$$\tilde{u}_{k+1|k} \triangleq u_{k+1} - \hat{u}_{k+1|k}\,, \quad k = 0,1,2,\ldots$$

Substitute (4.12) into this equation, and rearrange the terms to obtain

$$
\begin{aligned}
\tilde{u}_{k+1|k} &= u_{k+1} - \hat{u}_{k|k-1} - G(k+1|k)[y_k - M(k)\hat{u}_{k|k-1}] \\
&= \tilde{u}_{k|k-1} - G(k+1|k)[M(k)\tilde{u}_{k|k-1} + v_k] + u_{k+1} - u_k \\
&= [I_n - G(k+1|k)M(k)]\tilde{u}_{k|k-1} - G(k+1|k)v_k + u_{k+1} - u_k \qquad (4.14)
\end{aligned}
$$

for $k = 1,2,\ldots$. For $k = 0$, obviously $\tilde{u}_{1|o} = u_1$, since $\hat{u}_{1|o} = 0$.

A procedure analogous to the one that is used in (3.22b) shows that the stochastic process $\{\tilde{u}_{k+1|k},\ k = 0,1,2,\ldots\}$ is a zero-mean wide-sense Markov process. The covariance matrix of this process is given by (2.30):

$$P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k+1|k-1) - G(k+1|k)[\tilde{y}_{k|k-1}, \tilde{x}_{k+1|k-1}],\ k = 1,2,\ldots$$

for $k = 1,2,\ldots$, and $P_{\tilde{u}}(1|o) = P_u(1)$ for $k = 1$. Noting that

$$P_{\tilde{u}}(k+1|k-1) = P_{\tilde{u}}(k|k-1) + Q(k), \text{ by (4.7)}$$

and

$$[\tilde{y}_{k|k-1}, \tilde{x}_{k+1|k-1}] = [\tilde{x}_{k+1|k-1}, \tilde{y}_{k|k-1}]^T$$

$$= M(k)P_{\tilde{u}}(k|k-1) + C^T(k), \text{ by (4.16)}$$

we obtain

$$P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k|k-1) - G(k+1|k)[M(k)P_{\tilde{u}}(k|k-1) + C^T(k)] + Q(k)$$

$$= P_{\tilde{u}}(k|k-1) - [P_{\tilde{u}}(k|k-1)M^T(k) + C(k)][M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}$$

$$\times [M(k)P_{\tilde{u}}(k|k-1) + C^T(k)] + Q(k)$$

for $k = 1,2,\ldots$. QED

An examination of the results of this theorem reveals that a predict-correct concept is present as in the BP, and the predictor that is given by (4.12) has the same structure as the one given by (3.22).

The only difference between the algorithms given in (3.22) for BP and in (4.11) for the CCP is the difference between the expressions for the gain matrices $G(k+1,k)$. We expect this on the grounds that the gain matrix $G(k+1|k)$ is indicative of the amount of information contained in the innovation $\tilde{y}_{k|k-1}$ about the signal u_{k+1}. Since $u_{k+1} - u_k$ is correlated with v_k we expect that the expression for $G(k+1|k)$ of CCP may include the cross-correlation matrix $C(k) \triangleq [u_{k+1} - u_k, v_k]$.

We complete the discussion about (4.11) with the following remarks:

(4.17) REMARK. For the Kalman signal (2.17), the results of Theorem 4.11 can easily be written by using (4.22) and one obtains the estimation equations first derived by Kalman [K-4].

(4.18) REMARK. A different expression for the optimal single-stage prediction $\hat{u}_{k+1|k}$ can be obtained as follows:

$$\begin{aligned}\hat{u}_{k+1|k} &= (u_{k+1}|L_2^n(y;k)) \\ &= (u_{k+1} - u_k + u_k|L_2^n(y;k)) \\ &= \hat{u}_k + (u_{k+1} - u_k|L_2^n(y;k))\end{aligned}$$

Since by (2.26) $L_2^n(y,k) = L_2^n(y;k-1) \oplus L_2^n(\tilde{y}_{k|k-1})$ and since it can easily be shown that $u_{k+1} - u_k \perp L_2^n(y;k-1)$,

$$\begin{aligned}\hat{u}_{k+1|k} &= \hat{u}_k + (u_{k+1} - u_k|L_2^n(\tilde{y}_{k|k-1})) \\ &= \hat{u}_k + S(k+1|k)[y_k - M(k)\hat{u}_{k|k-1}] \qquad (4.19)\end{aligned}$$

for $k = 1,2,\ldots$, where $S(k+1|k)$ is the $n \times m$ predictor gain matrix to be determined. An easy computation shows that

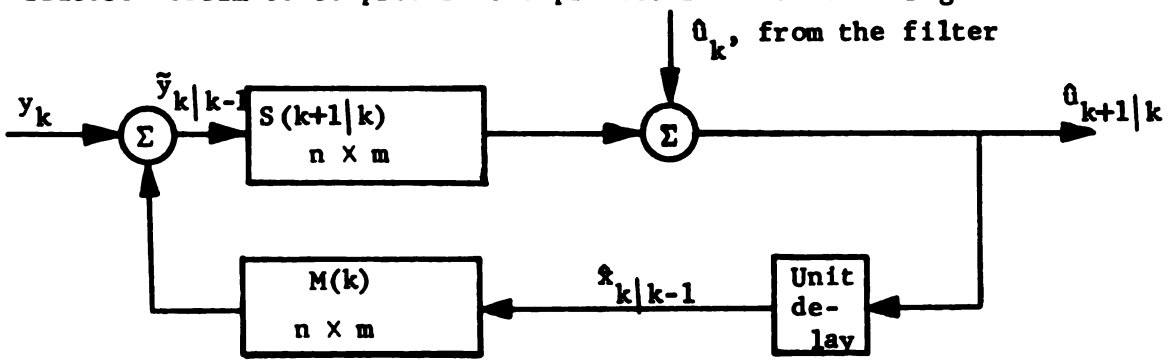
$$\begin{aligned}S(k+1|k) &= [u_{k+1} - u_k, \tilde{y}_{k|k-1}][\tilde{y}_{k|k-1}, \tilde{y}_{k|k-1}]^{-1} \\ &= C(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1} . \qquad (4.20)\end{aligned}$$

After a moderate amount of algebraic labor, we find that the optimal single-stage predictor error covariance matrix is given by

$$\begin{aligned}P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k) &- C(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}C^T(k) + Q(k) \\ &- P_{\tilde{u}}(k)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}C^T(k) - \\ &- C(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}M(k)P_{\tilde{u}}(k). \qquad (4.21)\end{aligned}$$

for $k = 1,2,\ldots$ with $P_{\tilde{u}}(1|o) = P_u(1)$ as the initial condition.

At this point we note that the optimal estimation equations (4.19)-(4.20) describe a procedure for the optimal single-stage prediction $(\hat{u}_{k+1|k}, P_{\tilde{u}}(k+1|k))$ in which one needs the optimal filtered estimate to process the predictor as shown in Figure 4.1.



$k = 1,2,\ldots$, $\hat{u}_{1|o} = 0$

Figure 4.1. Optimal single-stage predictor for CCP defined by (4.19).

4.2 OPTIMAL FILTERING FOR CCP

We now consider the case where we are only interested in the filtering problem. We shall derive a recursive equation for the filtered estimate that is based on the predicted estimate based on the previous observation record and the present observation. We summarize the results in the following theorem:

(4.22) THEOREM. The optimal filtering $(\hat{u}_k, P_{\tilde{u}}(k))$ $k = 0,1,2,\ldots$ of the signal process $\{u_k, k = 0,1,2,\ldots\}$ is accomplished via the following equations:

(a) The optimal filtered estimate is given by the expression

$$\hat{u}_k = \hat{u}_{k|k-1} + G(k)[y_k - M(k)\hat{u}_{k|k-1}] \tag{4.23}$$

for $k = 1,2,\ldots$ with $\hat{u}_o = 0$ as the initial condition, where

$G(k)$ is the filter gain matrix, and is given by

$$(4.24)\quad G(k) = P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}$$

$$= P_{\tilde{u}}(k)M^T(k)P_v(k).$$

(b) The filtering error covariance matrix is given by the expression

$$(4.25)\quad P_{\tilde{u}}(k) = P_{\tilde{u}}(k|k-1) - P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}M(k)P_{\tilde{u}}(k|k-1)$$

for $k = 1,2,\ldots$ with $P_{\tilde{u}}(0) = P_u(0)$ as the initial condition.

PROOF. (a) Note that

$$(4.8) \Rightarrow (2.27) \text{ holds,}$$

$$\Rightarrow \begin{cases} \hat{u}_k = \begin{cases} \hat{u}_{k|k-1} + G(k)[y_k - M(k)\hat{u}_{k|k-1}] & \text{for } k = 1,2,\ldots \\ \hat{u}_o = 0 \quad \text{for } k = 0 \end{cases} \\ G(k) = [\tilde{u}_{k|k-1}, \tilde{y}_{k|k-1}][\tilde{y}_{k|k-1}, \tilde{y}_{k|k-1}]^{-1}. \end{cases}$$

From (4.8), we know that

$$[\tilde{y}_{k|k-1}, \tilde{y}_{k|k-1}] = M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k).$$

On the other hand,

$$[\tilde{u}_{k|k-1}, \tilde{y}_{k|k-1}] = [\tilde{u}_{k|k-1}, M(k)\tilde{u}_{k|k-1} + v_k]$$

$$(4.26)\qquad = P_{\tilde{u}}(k|k-1)M^T(k),$$

since $v_k \perp u_k, \hat{u}_{k|k-1}$. Thus

$$G(k) = P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}.$$

Note that this expression has exactly the same form as that derived for BP (cf. (3.22)). Computations analogous to those in (3.22a) lead to

$$G(k) = P_{\tilde{u}}(k)M^T(k)P_v^{-1}(k).$$

(b) From (2.27b), it is seen that the filtering error covariance matrix is given by

$$P_{\tilde{u}}(k) = P_{\tilde{u}}(k|k-1) - G(k)[\tilde{y}_{k|k-1}, \tilde{u}_{k|k-1}]$$

for $k = 1,2,\ldots$ and $P_{\tilde{u}}(0) = P_u(0)$ for $k = 0$. Substituting (4.24) for $G(k)$ and (4.26) for $[\tilde{y}_{k|k-1}, \tilde{u}_{k|k-1}]^T$, we obtain

$$P_{\tilde{u}}(k) = P_{\tilde{u}}(k|k-1) - P_{\tilde{u}}(k|k-1)M^T(k)[M(k)P_{\tilde{u}}(k|k-1)M^T(k) + P_v(k)]^{-1}M(k)P_{\tilde{u}}(k|k-1)$$

for $k = 1,2,\ldots$, with $P_{\tilde{u}}(0) = P_u(0)$ as the initial condition. QED

We observe that the cross-correlation matrix $C(k)$ does not appear explicitly in the estimation equations (4.23)-(4.25). This matrix affects the estimation equations through the single-stage prediction-error covariance matrix. We must expect this fact because of the correlation between $u_{k+1} - u_k$ and v_k. Since v_k is uncorrelated with u_j for $j \leq k$, the innovation vector $\tilde{y}_{k|k-1}$ does not contain any information through $C(k)$ about the signal u_k at the time k. That is why in the expression for the gain matrix $G(k)$, $C(k)$ does not appear explicitly.

The block diagram of the filter is shown in Figure 4.2. It is a dynamical system $(I_n, G(k), I_n)$ and operates in a predict-correct fashion.

The estimation equations given in (4.22) can easily be written for the signal $x_k = \phi(k)u_k)$ by using (2.22). For the Kalman signal (2.17), the usual Kalman filter will be obtained [K-4].

$$k = 1,2,\ldots\ , \hat{u}_o = 0$$

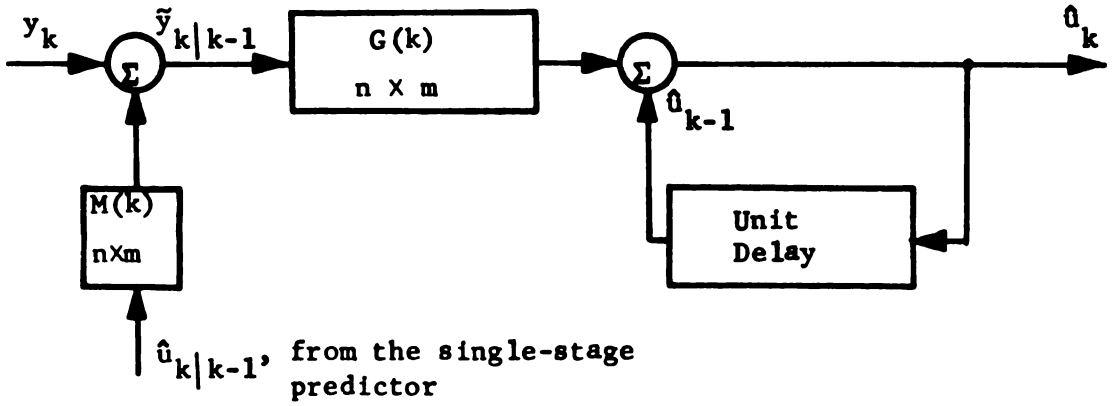


Figure 4.2. Block diagram of optimal filter for CCP.

4.3 OPTIMAL SMOOTHING FOR CCP

We continue our study of optimal estimation for the CCP with an examination of the smoothing problem. We recall from Section 3 of Chapter 3 that this problem can be classified into three distinct classes. We shall examine here only optimal single-stage and fixed-point smoothing. To do this we first derive a general formula analogous to (3.49) for the present case.

By the innovation lemma, (4.8), the observation process has full rank. Therefore from (2.27c) it is seen that the optimal smoothing $(\hat{u}_{k|\ell}, P_{\tilde{u}}(k|\ell))$ $k < \ell$ is accomplished via the following equations with $(\hat{u}_k, P_{\tilde{u}}(k))$ as the initial condition:

$$\hat{u}_{k|\ell} = \hat{u}_k + \sum_{i=k+1}^{\ell} G(k,\, i|i-1)\tilde{y}_{i|i-1} ,$$

$$G(k,\, i|i-1) = [\tilde{u}_{k|i-1},\, \tilde{y}_{i|i-1}][\tilde{y}_{i|i-1},\, \tilde{y}_{i|i-1}]^{-1} ,$$

$$P_{\tilde{u}}(k|\ell) = P_{\tilde{u}}(k) - \sum_{i=k+1}^{\ell} G(k,i|i-1)[\tilde{y}_{i|i-1},\, \tilde{u}_{k|i-1}]$$

where, by (4.8),

$$\tilde{y}_{i|i-1} = y_i - M(i)\hat{u}_{i|i-1}$$

and

$$[\tilde{y}_{i|i-1},\, \tilde{y}_{i|i-1}] = M(i)P_{\tilde{u}}(i|i-1)M^T(i) + P_v(i) .$$

If the steps which lead to (3.48) and (3.52) are repeated for the present case, their results

$$[\tilde{u}_{k|i-1},\tilde{y}_{i|i-1}] = P_{\tilde{u}}(k,\, i|i-1)M^T(i), \text{ by noting } v_i \perp \tilde{u}_{k|i-1}$$

and

$$P_{\tilde{u}}(k,i+1|i)=P_{\tilde{u}}(k,i|i-1)-P_{\tilde{u}}(k,i|i-1)M^T(i)[M(i)P_{\tilde{u}}(i|i-1)M^T(i)+P_v(i)]^{-1}M(i)P_{\tilde{u}}(i|i-1)$$

for $i = k, k+1, \ldots$ with the initial condition $P_{\tilde{u}}(k,k|k-1) = P_{\tilde{u}}(k|k-1)$.

Thus we have found the following optimal estimation equations for the smoothing $(\hat{u}_{k|\ell},\, P_{\tilde{u}}(k|\ell))$ $k < \ell$ with the initial condition $(\hat{u}_k,\, P_{\tilde{u}}(k))$:

$$(4.27)\quad \hat{u}_{k|\ell} = \hat{u}_k + \sum_{i=k+1}^{\ell} G(k,\, i|i-1)[y_i - M(i)\hat{u}_{i|i-1}] ;$$

$$(4.28)\quad G(k,i|i-1) = P_{\tilde{u}}(k,i|i-1)M^T(i)[M(i)P_{\tilde{u}}(i|i-1)M^T(i) + P_v(i)]^{-1} ;$$

$$(4.29)\quad P_{\tilde{u}}(k,i+1|i)=P_{\tilde{u}}(k,i|i-1)-P_{\tilde{u}}(k,i|i-1)M^T(i)[M(i)P_{\tilde{u}}(i|i-1)M^T(i) + P_v(i)]^{-1}M(i)P_{\tilde{u}}(i|i-1),\ P_{\tilde{u}}(k,k|k-1) = P_{\tilde{u}}(k|k-1);$$

$$(4.30)\qquad P_{\tilde{u}}(k|\ell) = P_{\tilde{u}}(k) - \sum_{i=k+1}^{\ell} G(k,i|i-1)M(i)P_{\tilde{u}}(k,i|i-1).$$

Note that these equations have exactly the same form of those obtained for the BP (cf. (3.50)-(3.53)). The cross-correlation effects these estimation equations only through the single-stage prediction-error covariance matrix $P_{\tilde{u}}(i|i-1)$. We expect that because v_i is correlated with u_j if $j \geq i+1$.

The optimal smoothing equations (4.27)-(4.29) for the signal u_k, which are valid for all the classes of smoothing, hold for the signal $x_k = \Phi(k)u_k$ with $u_k \in R(\Phi^T(k))$ and hence for Kalman signals (cf. (2.17)).

The form of these equations for single-stage smoothing and fixed-point smoothing is easily derived by repeating the steps leading to the analogous results given in Chapter 3. The equations derived here for Kalman signals extend well-known results (cf. [M-4]) to the cross-correlated noise case which was not previously solved to the author's knowledge.

SINGLE-STAGE SMOOTHING. Optimal single-stage smoothing $(\hat{u}_{k|k+1}, P_{\tilde{u}}(k|k+1))$, $k = 0,1,2,\ldots$, for the signal u_k is accomplished via the following equations with $(\hat{u}_o, P_{\tilde{u}}(0)) = (0, P_u(0))$ as the initial condition:

$$\hat{u}_{k|k+1} = \hat{u}_k + G(k, k+1|k)[y_{k+1} - M(k)\hat{u}_{k+1|k}],$$

$$G(k,k+1|k) = P_{\tilde{u}}(k)M^T(k+1)[M(k+1)P_{\tilde{u}}(k+1|k)M^T(k+1) + P_v(k+1)]^{-1},$$

$$P_{\tilde{u}}(k, k+1|k) = P_{\tilde{u}}(k),$$

$$P_{\tilde{u}}(k|k+1) = P_{\tilde{u}}(k) - G(k,k+1|k)M(k+1)P_{\tilde{u}}(k).$$

FIXED-POINT SMOOTHING. Optimal fixed-point smoothing, $(\hat{u}_{N|\ell}, P_{\tilde{u}}(N|\ell))$, $\ell = N+1, N+2, \ldots$ for the signal u_k, is accomplished via the following equations with $(\hat{u}_N, P_{\tilde{u}}(N))$ as the initial condition:

$$\hat{u}_{N|\ell} = \hat{u}_{N|\ell-1} + G(N, \ell|\ell-1)[y_\ell - M(\ell)\hat{u}_{\ell|\ell-1}] ,$$

$$G(N,\ell|\ell-1) = P_{\tilde{u}}(N,\ell|\ell-1)M^T(\ell)[M(\ell)P_{\tilde{u}}(\ell|\ell-1)M^T(\ell) + P_v(\ell)]^{-1} ,$$

$$P_{\tilde{u}}(N,\ell+1|\ell) = P_{\tilde{u}}(N,\ell|\ell-1) - P_{\tilde{u}}(N,\ell|\ell-1)M^T(\ell)[M(\ell)P_{\tilde{u}}(\ell|\ell-1)M^T(\ell)+P_v(\ell)]^{-1}$$

$$\times M(\ell)P_{\tilde{u}}(\ell|\ell-1), \quad P_{\tilde{u}}(N,N|N-1) = P_{\tilde{u}}(N|N-1),$$

$$P_{\tilde{u}}(N|\ell) = P_{\tilde{u}}(N|\ell-1) - G(N,\ell|\ell-1)M(\ell)P_{\tilde{u}}(\ell, N|\ell-1) .$$

CHAPTER 5

COLORED NOISE PROBLEM (CNP)

Having treated the problem of optimal estimation for uncorrelated and cross-correlated observation noises (cf. (A.1.1), (A.1.2)) in the preceding two chapters, we turn now to the estimation problem for cross-correlated colored observation noise. This is the most general problem studied in this dissertation.

The colored noise problem in Kalman filtering theory was first discussed by Cox [C-1], Bryson and Johanson [B-5], and Bucy [B-9], and by Bryson and Henrikson [B-6], Stear and Stubberud [S-4] and others [M-7], [Z-5], [F-2]. Bryson and Johansen's work was based on (i) the "augmented state" procedure suggested by Kalman [K-4], and (ii) the assumption that colored noise is generated by a given linear difference (or differential) equation forced by white noise. The augmented state procedure has not been widely used because it leads to ill-conditioned computations in constructing the filter. Assumption (ii) has been used in all investigations published to date.

Here we solve the colored noise problem by reducing it to a wide-sense martingale noise problem and then applying the technique developed in the preceding chapters. This is a more direct method than the previous work in this area and gives a different perspective and yields new results. An advantage of this approach is that the

colored noise, which is a wide-sense Markov process by assumption, need not necessarily be given by a linear difference equation with white noise input. In addition, the "martingale" approach provides optimal estimation algorithms for observations corrupted by additive wide-sense martingale noise. This case has not been considered in the existing literature as far as the author knows.

We shall begin our study by reformulating the problem in the following section.

5.1 REFORMULATION OF THE PROBLEM.

Recall from Section 2 of Chapter 1 that the signal and observation are described by, respectively,

$$x_k = \Phi(k)u_k \qquad k = 0,1,2,\ldots \tag{5.1}$$

$$y_k = M(k)u_k + v_k \tag{5.2}$$

where

$$v_{k+1} = \Psi(k+1, k)v_k + n_k \qquad k = 0,1,2,\ldots \tag{5.3}$$

The assumptions on the initial conditions x_o (or u_o), v_o and output noise are the same as those stated in (1.4) and (A.1.3). The matrices $\Phi(k)$ and $M(k)$ have been defined in (1.4) and (1.5) respectively.

Note that the unique solution of (5.3), with the initial condition v_o, is

$$\begin{aligned} v_k &= \Psi(k,o)v_o + \sum_{i=1}^{k} \Psi(k,i)n_{i-1} \\ &= \Psi(k,o)\left[v_o + \sum_{i=1}^{k} \Psi(o,i)n_{i-1}\right], \text{ since } \Psi^{-1}(k,o)\Psi(k,i) = \Psi(o,i) \end{aligned} \tag{5.4}$$

for $k = 1,2,\ldots$. Now, define

$$m_k = \begin{cases} v_o + \sum_{i=1}^{k} \psi(o,i) n_{i-1} , & \text{if } k = 1,2,\ldots \\ v_o & \text{if } k = 0 \end{cases} \tag{5.5}$$

then it follows that m_k satisfies the linear stochastic difference equation

$$m_{k+1} = m_k + \psi(o, k+1) n_k \tag{5.6}$$

with $m_o = v_o$ as the initial condition. By Assumption A.1.2, $v_o \perp n_k$, $\forall\, k = 0,1,2,\ldots$, hence, in view of (2.14), the stochastic process $\{m_k, k = 0,1,2,\ldots\}$ is a wide-sense martingale process with zero mean, and also

$$[m_{k+1} - m_k, m_{k+1} - m_k] = \psi(o,k+1) P_n(k) \psi^T(o,k+1) \quad k = 0,1,2,\ldots \tag{5.7}$$

where $P_n(k) \overset{d}{=} [n_k, n_k]$. In addition,

$$[u_{k+1} - u_k, m_{j+1} - m_j] = C(k) \psi^T(o,k+1) \delta_{kj} \quad k,j = 0,1,2,\ldots \tag{5.8}$$

since, by (A.1.3), $[u_{k+1} - u_k, n_j] = C(k)\delta_{kj}$. Finally we note from (1.4) that $[u_o, n_k] = 0 \quad \forall\, k = 0,1,2,\ldots$ which implies

$$[u_o, m_{k+1} - m_k] = 0 \quad k = 0,1,2,\ldots \tag{5.9}$$

From (A.1.3) and (5.5) we conclude the following useful results:

$$u_{k+1} - u_k \perp m_j \quad \text{and} \quad m_{k+1} - m_k \perp u_j \tag{5.10}$$

$\forall\, j \le k$, $k,j = 0,1,2,\ldots$.

Observe from (5.4) and (5.5) that $v_k = \Psi(k,o)m_k$, $k = 0,1,2,\ldots$. Hence, (5.2) and (5.3) can be combined in one equation as

$$y_k = M(k)u_k + \Psi(k,o)m_k \qquad k = 1,2,\ldots \tag{5.11}$$

The problem is thus to find the minimum mean-square estimate of the signal $x_k = \Phi(k)u_k$, or, equivalently, in view of Section 3 of Chapter 2, of the signal u_k, from the data $Y(\ell) = \{y_1, y_2, \ldots, y_\ell\}$ when u_k is related to the data $Y(\ell)$ by (5.11). The solution to this problem is given in the following three sections.

5.2 OPTIMAL PREDICTION FOR CNP

As in the preceding two chapters, we first derive a formula for the general predicted estimate $\hat{u}_{k|\ell}$, $k > \ell$. To do so, we proceed as follows:

$$\begin{aligned}
\hat{u}_{k|\ell} &= (u_k | L_2^n(y;\ell)) \text{ , by the orthogonal projection lemma} \\
&= (u_k - u_\ell + u_\ell | L_2^n(y;\ell)) \\
&= \hat{u}_\ell + (u_k - u_\ell | L_2^n(y;\ell)), \text{ since } (u_\ell | L_2^n(y;\ell)) = \hat{u}_\ell
\end{aligned}$$

for $k > \ell$, $k,\ell = 0,1,2,\ldots$. From (1.4) and (5.10), we see that $u_k - u_\ell \perp L_2^n(y;\ell)$. Therefore

$$\hat{u}_{k|\ell} = \hat{u}_\ell, \quad k \geq \ell \qquad k,\ell = 0,1,2,\ldots \tag{5.12}$$

where $\hat{u}_\ell$ is the optimal filtered estimate at the present observation time ℓ. It is assumed that the filtered estimate $\hat{u}_\ell$ is obtained by using the filtering algorithms which will be derived in Section 3 of the present chapter.

Obviously (5.12) is valid for all the classes of prediction. At this point, we note that we obtained exactly the same result for the BP (cf. (3.2)), i.e., the general predicted estimate is equal to the filtered estimate at the last observation time. So, the form (5.12) for each class of prediction will be exactly the same one that was obtained for the corresponding class in Chapter 3. These forms are repeated here for the sake of completeness without further comment.

FIXED-INTERVAL PREDICTION. Optimal fixed-interval prediction $(\hat{u}_{k|L}, P_{\tilde{u}}(k|L))$, $k = L+1, L+2, \ldots$, L = fixed-positive integer is accomplished as follows with the initial condition $(\hat{u}_L, P_{\tilde{u}}(L))$:

$$\hat{u}_{k|L} = \hat{u}_L \,, \tag{5.13}$$

$$\begin{aligned} P_{\tilde{u}}(k|L) &= P_{\tilde{u}}(L) + P_u(k) - P_u(L) \\ &= P_{\tilde{u}}(L) + \sum_{i=L}^{k-1} Q(i) \,. \end{aligned} \tag{5.14}$$

FIXED-POINT PREDICTION. Optimal fixed-point prediction $(\hat{u}_{N|\ell}, P_{\tilde{u}}(N|\ell))$, $\ell = 0,1,2,\ldots,N-1$, N = fixed-positive integer is accomplished via the following equations with the boundary condition $(\hat{u}_N, P_{\tilde{u}}(N))$:

$$\hat{u}_{N|\ell} = \hat{u}_\ell \,, \tag{5.15}$$

$$P_{\tilde{u}}(N|\ell) = P_{\tilde{u}}(\ell) + P_u(N) - P_u(\ell) \,. \tag{5.16}$$

FIXED-LEAD PREDICTION. Optimal fixed-lead prediction $(\hat{u}_{L+\ell|\ell}, P_{\tilde{u}}(L+\ell|\ell))$, $\ell = 0,1,2,\ldots$, L = fixed-positive integer, is given by the following equation with $(\hat{u}_{L|o}, P_{\tilde{u}}(L|o)) = (0, P_u(L))$ as

the initial condition:

$$\hat{u}_{L+\ell|\ell} = \hat{u}_{\ell} , \tag{5.15}$$

$$P_{\hat{u}}(L+\ell|\ell) = P_{\tilde{u}}(\ell) + P_u(L+\ell) - P_u(\ell) . \tag{5.16}$$

Consider the special case $L = 1$, i.e., the single-stage prediction $(\hat{u}_{\ell+1|\ell}, P_{\tilde{u}}(\ell+1|\ell))$:

$$\hat{u}_{\ell+1|\ell} = \hat{u}_{\ell} ,$$

$$P_{\tilde{u}}(\ell+1|\ell) = P_{\tilde{u}}(\ell) + Q(\ell) .$$

As in the preceding two chapters, we now develop the recursive algorithms for this case based on the last observation and previous predicted estimate. To do so, we need the following lemma which was called the innovation lemma in the previous chapters.

(5.17) INNOVATION LEMMA. The observation process $\{y_k, k = 1,2,...\}$ which is defined by (5.11) has full rank, i.e.

$$[\tilde{y}_{k+1|k}, \tilde{y}_{k+1|k}] > 0 .$$

PROOF. We first derive an expression for the innovation vector $\tilde{y}_{k+1|k}$. Doing so, we note by definition

$$\tilde{y}_{k+1|k} = y_{k+1} - (y_{k+1}|L(y;k)) .$$

Since, by (5.11), $y_{k+1} = M(k+1)u_{k+1} + \Psi(k+1,o)m_{k+1}$ we have

$$\tilde{y}_{k+1|k} = y_{k+1} - M(k+1)\hat{u}_{k+1|k} - \Psi(k+1,o)\hat{m}_{k+1|k} \tag{5.18}$$

for $k = 0,1,2,...$. Note this expression involves the optimal

single-stage predicted estimate of the observation noise m_k. This estimate is computed as follows:

$$\begin{aligned}
\hat{m}_{k+1|k} &= (m_{k+1}|L(y;k)) \\
&= (m_{k+1} - m_k + m_k|L(y;k)) \\
&= (m_k|L(y;k)), \text{ since, by (5.10), } m_{k+1} - m_k \perp L(y;k) \\
&= \Psi(o,k)(\Psi(k,o)m_k|L(y;k)), \text{ by } \Psi(o,k)\Psi(k,o) = I_m \text{ and (2.21a)} \\
&= \Psi(o,k)(y_k - M(k)u_k|L(y;k)), \text{ by (5.11)} \\
&= \Psi(o,k)[y_k - M(k)\hat{u}_k] \text{ , since } y_k \in L(y;k) \text{ .} \qquad (5.19)
\end{aligned}$$

Substituting (5.19) into (5.18) and noting that $\hat{u}_k = \hat{u}_{k+1|k}$ and $\Psi(k+1,o)\Psi(o,k) = \Psi(k+1,k)$ we get the result

$$\begin{aligned}
\tilde{y}_{k+1|k} &= y_{k+1} - M(k+1)\hat{u}_{k+1|k} - \Psi(k+1,k)[y_k - M(k)\hat{u}_{k+1|k}] \\
&= y_{k+1} - \Psi(k+1,k)y_k - [M(k+1) - \Psi(k+1,k)M(k)]\hat{u}_{k+1|k} \qquad (5.20)
\end{aligned}$$

for $k = 0,1,2,\ldots$. Note that to process the innovation vector $\tilde{y}_{k+1|k}$ we need the observation vectors y_{k+1}, y_k (i.e. the last and preceding observation vectors) and the single-stage predicted estimate based on the observation record $Y(k)$.

In order to prove that the observation process has full rank, in view of Section 3 of Chapter 2, we must show that $[\tilde{y}_{k+1|k}, \tilde{y}_{k+1|k}] > 0$, $\forall\, k = 0,1,2,\ldots$. From (5.20) and (5.11), it follows that

$$\tilde{y}_{k+1|k} = (M(k+1) - \psi(k+1,k)M(k))\tilde{u}_{k+1|k} + \psi(k+1,k)M(k)(u_{k+1} - u_k)$$

$$+ \psi(k+1,o)(m_{k+1} - m_k)$$

$$(5.21) = \tilde{M}(k+1)\tilde{u}_{k+1|k} + \psi(k+1,k)M(k)(u_{k+1} - u_k) + \psi(k+1,o)(m_{k+1} - m_k)$$

for $k = 0,1,2,\ldots$, where the definition

$$(5.22) \qquad \tilde{M}(k+1) = M(k+1) - \psi(k+1,k)M(k)$$

is made as a notational convenience. Using (5.21) we may write

$$[\tilde{y}_{k+1|k},\tilde{y}_{k+1|k}]=[\tilde{M}(k+1)\tilde{u}_{k+1|k} + \psi(k+1,k)M(k)(u_{k+1} - u_k) + \psi(k+1,o)(m_{k+1}-m_k),$$

$$\tilde{M}(k+1)\tilde{u}_{k+1|k} + \psi(k+1,k)M(k)(u_{k+1}-u_k)+ \psi(k+1,o)(m_{k+1}-m_k)]$$

for $k = 0,1,2,\ldots$. Noting that

$$[\tilde{u}_{k+1|k},\tilde{u}_{k+1|k}] \stackrel{d}{=} P_{\tilde{u}}(k+1,k),$$

$$[\tilde{u}_{k+1|k},u_{k+1}-u_k] = [u_{k+1}-u_k,\tilde{u}_{k+1|k}]^T$$

$$= Q(k) \text{ , since } u_{k+1} - u_k \perp \hat{u}_{k+1|k} \text{ ,}$$

$$[\tilde{u}_{k+1|k},m_{k+1} - m_k] = [m_{k+1} - m_k, \tilde{u}_{k+1|k}]^T$$

$$= C(k)\psi^T(o,k+1) \text{ , since } m_{k+1}-m_k \perp \hat{u}_{k+1|k},u_k$$

$$[u_{k+1} - u_k, u_{k+1} - u_k] \stackrel{d}{=} Q(k) \text{ ,}$$

$$[u_{k+1}-u_k,m_{k+1}-m_k] = [m_{k+1}-m_k,u_{k+1}-u_k]^T$$

$$= C(k)\psi^T(o,k+1) \text{ by (5.8)}$$

$$[m_{k+1}-m_k,m_{k+1}-m_k] = \psi(o,k+1)P_n(k)\psi^T(o,k+1) \text{ , by (5.7)}$$

we obtain the result

$$
\begin{aligned}
[\tilde{y}_{k+1|k},\tilde{y}_{k+1|k}] &= \tilde{M}(k+1)\,P_{\tilde{u}}(k+1|k)\tilde{M}^T(k+1) + \tilde{M}(k+1)Q(k)M^T(k)\psi^T(k+1,k) \\
&\quad + \tilde{M}(k+1)C(k) + \psi(k+1,k)M(k)Q(k)\tilde{M}^T(k+1) \\
&\quad + \psi(k+1,k)M(k)Q(k)M^T(k)\psi^T(k+1,k) \\
&\quad + \psi(k+1,k)M(k)C(k) + C^T(k)\tilde{M}^T(k+1) + C^T(k)M^T(k)\psi^T(k+1,k) \\
&\quad + P_n(k) \\
&= \tilde{M}(k+1)\,P_{\tilde{u}}(k+1|k)\tilde{M}^T(k+1) + M(k+1)Q(k)M^T(k)\psi^T(k+1,k) \\
&\quad + \psi(k+1,k)M(k)Q(k)M^T(k+1) + M(k+1)C(k) + C^T(k)M^T(k+1) + P_n(k) \\
&\quad - \psi(k+1,k)M(k)Q(k)M^T(k)\psi^T(k+1,k)
\end{aligned} \tag{5.23}
$$

for $k = 0,1,2,\ldots$. From (5.16) we have $P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k) + Q(k)$. Substituting this into (5.23) and simplifying the result using (5.22), we get

$$
\begin{aligned}
[\tilde{y}_{k+1|k},\tilde{y}_{k+1|k}] &= \tilde{M}(k+1)\,P_{\tilde{u}}(k)\tilde{M}^T(k+1) + M(k+1)Q(k)M^T(k+1) \\
&\quad + M(k+1)C(k) + C^T(k)M^T(k+1) + P_n(k)
\end{aligned} \tag{5.24}
$$

for $k = 0,1,2,\ldots$. It is clear that this $m \times m$ matrix is invertible if

$$M(k+1)C(k) + C^T(k)M^T(k+1) + P_n(k) > 0 .$$

By assumption A.1.3 this is so, therefore

$$[\tilde{y}_{k+1|k},\ \tilde{y}_{k+1|k}] > 0$$

as desired. For notational convenience, we shall denote this matrix by $P_{\tilde{y}}(k+1|k)$. QED

(5.25) THEOREM. Optimal single-stage prediction $(\hat{u}_{k+1|k}, P_{\tilde{u}}(k+1|k))$ $k = 0,1,2,\ldots$ for the signal u_k is accomplished as follows:

(a) The stochastic process $\{\hat{u}_{k+1|k}, k = 0,1,2,\ldots\}$, which is defined by the single-stage prediction estimate, is a zero mean wide-sense martingale, and is generated by the recursion

$$(5.26)\quad \hat{u}_{k+1|k} = \hat{u}_{k|k-1} + G(k+1|k)[y_k - \Phi(k,k-1)y_{k-1} - \tilde{M}(k)\hat{u}_{k|k-1}]$$

for $k = 1,2,\ldots$ with $\hat{u}_{1|o} = 0$ as the initial condition. The $n \times m$ gain matrix $G(k+1|k)$ is given by

$$(5.27)\quad G(k+1|k) = [P_{\tilde{u}}(k|k-1)\tilde{M}^T(k) + C(k-1)]P_{\tilde{y}}^{-1}(k|k-1) .$$

(b) The stochastic process $\{\tilde{u}_{k+1|k}, k = 0,1,2,\ldots\}$, which is defined by the single-stage prediction error $\tilde{u}_{k+1|k}$ given by

$$(5.28)\quad \tilde{u}_{k+1|k} = [I_n - G(k+1)\tilde{M}(k)]\tilde{u}_{k|k-1} - G(k+1|k)M(k-1)(u_k - u_{k-1})$$
$$- G(k+1|k)\Phi(k,o)(m_k - m_{k-1}) + u_{k+1} - u_k$$

for $k = 1,2,\ldots$ with the initial condition $\tilde{u}_{1|o} = u_1$, is a zero-mean wide-sense Markov process. The covariance matrix of this process is given by the recursive equation

$$(5.29)\quad P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k|k-1) - G(k+1|k)[\tilde{M}(k)P_{\tilde{u}}(k|k-1) + C^T(k-1)] + Q(k)$$

for $k = 1,2,\ldots$ with $P_{\tilde{u}}(1|o) = P_u(1)$ as the initial condition.

PROOF. (a) That the stochastic process $\{\hat{u}_{k+1|k}, k = 0,1,2,\ldots\}$ is a zero-mean wide-sense martingale process follows from (5.12) and the properties of orthogonal projection as demonstrated in (3.22a).

From (5.17) we conclude that (2.27) holds, i.e.,

$$\hat{u}_{k+1|k} = \hat{u}_{k|k-1} + G(k+1|k)\tilde{y}_{k|k-1}$$

$$= \hat{u}_{k|k-1} + G(k+1|k)[y_k - \psi(k,k-1)y_{k-1} - \tilde{M}(k)\hat{u}_{k|k-1}]$$

for $k = 1,2,\ldots$ with $\hat{u}_{1|o} = 0$ as the initial condition. In obtaining the last equality, we used (5.12), (5.20) and (5.22). The gain matrix is given by (2.29):

$$G(k+1|k) = [\tilde{u}_{k+1|k-1},\tilde{y}_{k|k-1}][\tilde{y}_{k|k-1},\tilde{y}_{k|k-1}]^{-1}$$

where the only unknown is the $n \times m$ matrix $[\tilde{u}_{k+1|k-1},\tilde{y}_{k|k-1}]$. This matrix is determined as follows:

$$[\tilde{u}_{k+1|k-1},\tilde{y}_{k|k-1}] = [u_{k+1}-u_k + \tilde{u}_{k|k-1},\tilde{M}(k)\tilde{u}_{k|k-1} + \psi(k,k-1)(u_k-u_{k-1})$$

$$+ \psi(k,o)(m_k - m_{k-1})]$$

$$= P_{\tilde{u}}(k|k-1)\tilde{M}^T(k) + C(k-1) \quad . \tag{5.30}$$

Thus, we have found

$$G(k+1|k) = [P_{\tilde{u}}(k|k-1)\tilde{M}^T(k) + C(k-1)]P_{\tilde{y}}^{-1}(k|k-1) \tag{5.27}$$

where $P_{\tilde{y}}(k|k-1)$ is given by (5.23) (or (5.24)).

(b) By definition the single-stage prediction error is

$$\tilde{u}_{k+1|k} = u_{k+1} - \hat{u}_{k+1|k}, \quad k = 1,2,\ldots \text{ and } \tilde{u}_{1|o} = u_1 \text{ for } k = 0.$$

Noting that $\hat{u}_{k+1|k}$ is given by (5.26), we obtain

$$(5.28)\quad \tilde{u}_{k+1|k} = u_{k+1} - \hat{u}_{k|k-1} - G(k+1|k)[y_k - \psi(k,o)y_{k-1} - \tilde{M}(k)\hat{u}_{k|k-1}]$$

$$= [I_n - G(k+1|k)\tilde{M}(k)]\tilde{u}_{k|k-1} - G(k+1|k)M(k-1)(u_k - u_{k-1})$$

$$- G(k+1|k)\psi(k,o)(m_k - m_{k-1}) + u_{k+1} - u_k$$

for $k = 1,2,\ldots$ with $\tilde{u}_{1|o} = u_1$ as the initial condition. It is clear that $\tilde{u}_{k+1|k}$ has zero mean.

We now prove that the stochastic process $\{\tilde{u}_{k+1|k},\ k = 0,1,2,\ldots\}$ is a zero-mean wide-sense Markov process. To begin, we define

$$F(k) \triangleq I_n - G(k+1|k)\tilde{M}(k)$$

$$\Gamma(k) \triangleq [-G(k+1|k)M(k-1),\ -G(k+1|k)\psi(k,o),\ I_n]$$

and

$$w_k \triangleq \begin{bmatrix} u_k - u_{k-1} \\ m_k - m_{k-1} \\ u_{k+1} - u_k \end{bmatrix}$$

Then (5.26) can be written as

$$\tilde{u}_{k+1|k} = F(k)\tilde{u}_{k|k-1} + \Gamma(k)w_k \qquad k = 1,2,\ldots$$

From the definitions it is clear that $F(k)$ is invertible and the stochastic process $\{w_k,\ k = 1,2,\ldots\}$ is a white-noise process with zero mean and that

$$[w_k, w_j] = \begin{bmatrix} Q(k-1) & C(k-1)\psi^T(o,k) & 0 \\ \psi(o,k)C^T(k-1) & \psi(o,k)P_n(k-1)\psi^T(k,o) & 0 \\ 0 & 0 & Q(k) \end{bmatrix} \delta_{kj} .$$

Furthermore, since $M(0) = 0$ and $G(1|o) = 0$, $\tilde{u}_{1|o} = u_1 \perp w_1$ and therefore $\tilde{u}_{1|o} \perp w_k \quad \forall\, k = 1,2,\ldots$. Hence, the stochastic process $\{\tilde{u}_{k+1|k},\ k = 0,1,2,\ldots\}$ is a zero-mean wide-sense Markov process, by virtue of (2.17).

Now it remains to find an expression for the single-stage prediction error covariance matrix $P_{\tilde{u}}(k+1|k)$. From (2.30) we have

$$P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k+1|k-1) - G(k+1|k)[\tilde{y}_{k|k-1},\tilde{u}_{k|k-1}]$$

for $k = 1,2,\ldots$ with $P_{\tilde{u}}(1|o) = P_u(1)$ as the initial condition. Substituting (5.27) for $G(k+1|k)$, (5.30) for $[\tilde{y}_{k|k-1},\tilde{u}_{k|k-1}]^T$, and noting from (5.26) that

$$P_{\tilde{u}}(k+1|k-1) = P_{\tilde{u}}(k|k-1) + Q(k)$$

we obtain

$$P_{\tilde{u}}(k+1|k) = P_{\tilde{u}}(k|k-1) - [P_{\tilde{u}}(k|k-1)\tilde{M}^T(k) + C^T(k-1)]P_{\tilde{y}}^{-1}(k|k-1)[\tilde{M}(k)P_{\tilde{u}}(k|k-1) + C^T(k-1)] + Q(k)$$

for $k = 1,2,\ldots$. QED

This theorem gives the recursive algorithm for computing the single-stage prediction. The information flow in the predictor is shown in the Figure (5.1) which is a representation of (5.26). We observe that the predictor (5.26) requires the storage of one observation. This is the only difference between the computational procedure of the predictor given in Figure 5.1 and the preceding two given in Chapters 3 and 4.

We now write the optimal single-stage prediction equations for the signal $x_k = \Phi(k)u_k$ by assuming that $u_k \in R(\Phi^T(k))$

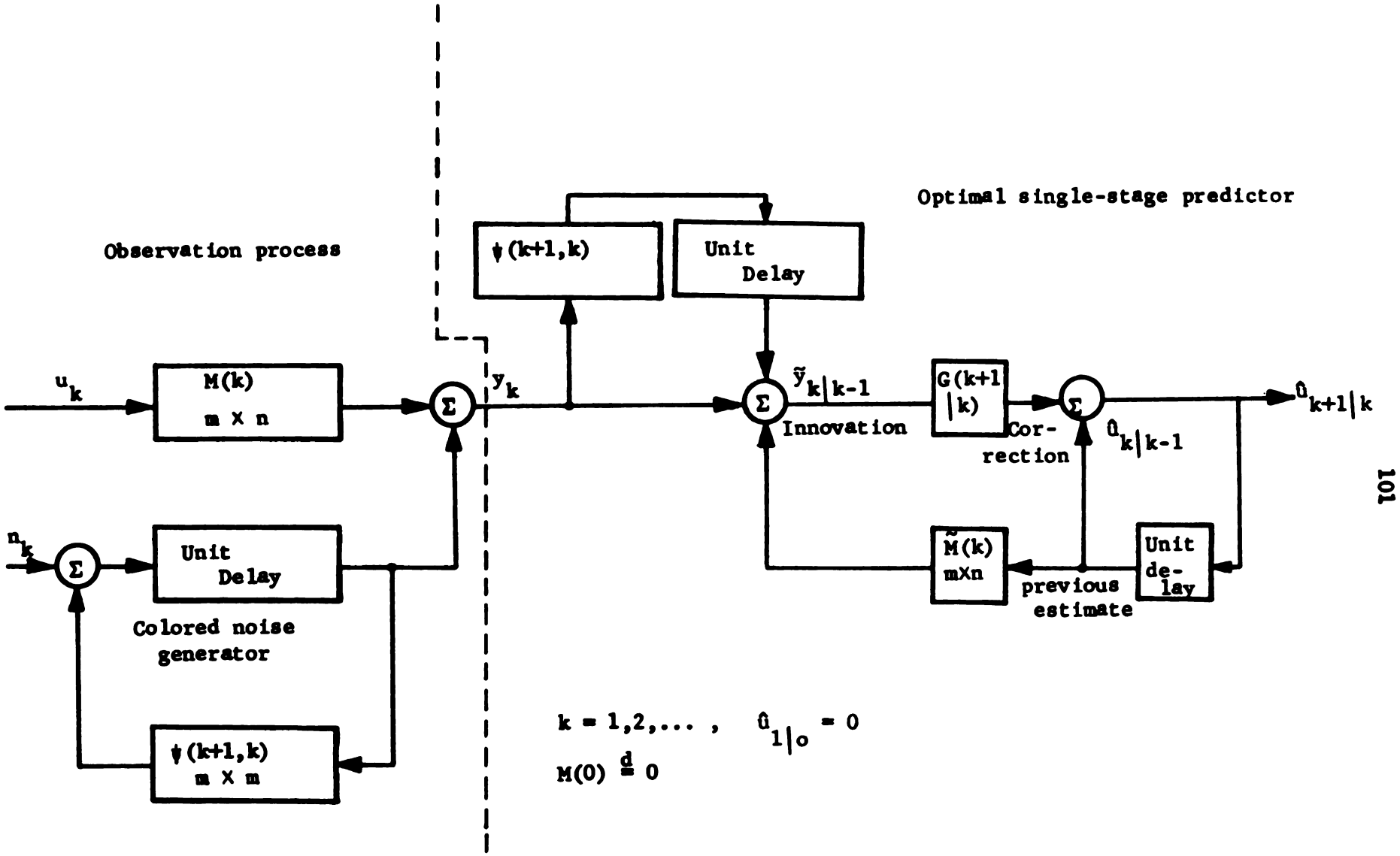


Figure 5.1. Block diagram for single-stage predictor for CNP.

$\forall\, k = 0,1,2,\ldots$. The derivations of these equations are straightforward and are omitted.

Optimal estimate:

$$(5.31)\quad \hat{x}_{k+1|k} = \Phi(k+1)\Phi^{+}(k)\hat{x}_{k|k-1} + K(k+1|k)[y_k - \Psi(k,k-1)y_{k-1} - \tilde{H}(k)\hat{x}_{k|k-1}]$$

for $k = 1,2,\ldots$ with $\hat{x}_{1|o} = 0$ as the initial condition, where

$$(5.32)\quad K(k+1|k) = \Phi(k+1)[\Phi^{+}(k)P_{\tilde{x}}(k|k-1)H^{T}(k) + C(k-1)]P_{\tilde{y}}^{-1}(k|k-1)$$

$$(5.33)\quad \tilde{H}(k) = H(k) - \Psi(k,k-1)H(k-1)\Phi(k-1)\Phi^{+}(k) \ .$$

Error covariance matrix:

$$\begin{aligned}(5.34)\quad P_{\tilde{x}}(k+1|k) &= \Phi(k+1)\Phi^{+}(k)P_{\tilde{x}}(k|k-1)\Phi^{+^{T}}(k)\Phi^{T}(k+1) \\ &\quad - K(k+1|k)[\tilde{H}(k)P_{\tilde{x}}(k|k-1)\Phi^{+^{T}}(k) + C^{T}(k-1)]\Phi^{T}(k+1) \\ &\quad + \Phi(k+1)Q(k)\Phi^{T}(k+1)\end{aligned}$$

for $k = 1,2,\ldots$ with $P_{\tilde{x}}(1|o) = P_x(1)$ as the initial condition.

We remark here that the proposed single-stage predictor (5.31)-(5.34) for the signal $x_k = \Phi(k)u_k$ is of dimension n instead of $(n+m)$ as in the augmented state predictor given by Bryson and Johanson [B-5] who considered a smaller class of signals (Kalman signals). The results of this chapter also extend Bryson and Henrikson's [B-6] work to cross-correlated colored noise in the larger signal class. The wide-sense martingale approach we develop is conceptually and computationally simpler.

5.3 OPTIMAL FILTERING FOR CNP

We continue our study by an examination of the optimal filtering problem. We wish to develop an algorithm for optimal filtering of the signal u_k. In doing so, we assume that only

the initial estimate $\hat{u}_{o|o} = \hat{u}_o$, and the filtering error covariance matrix at the initial time, $P_{\tilde{u}}(o|o) = P_{\tilde{u}}(0)$, are known.

As in Chapter 3, from (5.12) we observe that optimal prediction and filtering are interdependent in terms of the determination of the filtered estimate given the predicted estimate and vice-versa. In fact

$$\hat{u}_{k+1|k} = \hat{u}_k \qquad k = 0,1,2,\ldots .$$

Thus the single-stage predicted estimate algorithm (5.26) can be used to process the optimal filtered estimate. So, in order to solve the filtering problem, it remains to find a recursive expression for the filtering error covariance matrix. This is done in the following theorem.

(5.35) THEOREM. Optimal filtering $(\hat{u}_k, P_{\tilde{u}}(k))$ $k = 0,1,2,\ldots$ for the signal u_k is accomplished as follows:

(a) The stochastic process $\{\hat{u}_k, k = 0,1,2,\ldots\}$, which is defined by the filtered estimate, is a zero-mean wide-sense martingale, and is generated by the recursion

$$(5.36) \quad \hat{u}_k = \hat{u}_{k-1} + G(k)[y_k - \Phi(k,k-1)y_{k-1} - \tilde{M}(k)\hat{u}_{k-1}]$$

for $k = 1,2,\ldots$ with $\hat{u}_o = 0$ as the initial condition. The $n \times m$ matrix $G(k)$ is given by

$$(5.37) \quad G(k) = [P_{\tilde{u}}(k|k-1)M^T(k) + C(k-1)]P_{\tilde{y}}^{-1}(k|k-1)$$

where $P_{\tilde{y}}(k|k-1)$ is given by (5.24) .

(b) The stochastic process $\{\tilde{u}_k, k = 0,1,2,\ldots\}$, which is defined by the filtering error $\tilde{u}_k$ given by

$$(5.38)\quad \tilde{u}_{k+1} = [I_n - G(k)\tilde{M}(k)]\tilde{u}_{k-1} + [I_n - G(k)M(k)](u_k - u_{k-1}) + G(k)\psi(k,o)(m_k - m_{k-1})$$

for $k = 1,2,\ldots$ with the initial condition $\tilde{u}_o = u_o$, is a zero-mean wide-sense Markov process. The covariance matrix of this process is given by the recursion

$$(5.39)\quad P_{\tilde{u}}(k) = P_{\tilde{u}}(k|k-1) - G(k)[\tilde{M}(k)P_{\tilde{u}}(k|k-1) + C^T(k)]$$

for $k = 1,2,\ldots$ with $P_{\tilde{u}}(0) = P_u(0)$ as the initial condition.

PROOF. (a) It follows from (5.12) and (5.25a).

(b) By definition

$$\tilde{u}_k = u_k - \hat{u}_k,\ k = 0,1,2,\ldots .$$

Substitute (5.32) for $\hat{u}_k$ to obtain

$$\begin{aligned}
\tilde{u}_k &= u_k - \hat{u}_{k-1} - G(k)[y_k - \psi(k,k-1)y_{k-1} - \tilde{M}(k)\hat{u}_{k-1}] \\
&= \tilde{u}_{k-1} + u_k - u_{k-1} - G(k)[M(k)u_k - \psi(k,k-1)M(k-1)u_{k-1} \tilde{M}(k)\hat{u}_{k-1} \\
&\quad + \psi(k,o)m_k - \psi(k,o)m_{k-1}] \\
&= \tilde{u}_{k-1} - G(k)[(M(k) - \psi(k,k-1)M(k-1))u_{k-1} - \tilde{M}(k)\hat{u}_{k-1} - M(k)(u_k - u_{k-1}) \\
&\quad + \psi(k,o)(m_k - m_{k-1})] + u_k - u_{k-1} \\
&= [I_n - G(k)\tilde{M}(k)]\tilde{u}_{k-1} + [I_n - G(k)M(k)](u_k - u_{k-1}) + G(k)\psi(k,o)(m_k - m_{k-1})
\end{aligned}$$

for $k = 1,2,\ldots$. Obviously $\tilde{u}_o = u_o - \hat{u}_o = u_o$, since $\hat{u}_o = 0$.

Utilizing the same procedure as that in (5.32b), we prove that the stochastic process $\{\tilde{u}_k,\ k = 0,1,2,\ldots\}$ is defined by (5.34) is a zero-mean wide-sense Markov process.

To complete the proof, notice from (5.16) that

$$\underset{\tilde{u}}{P}(k) = \underset{\tilde{u}}{P}(k+1|k) - Q(k) \qquad k = 0,1,2,\ldots .$$

Substituting (5.29) for $\underset{\tilde{u}}{P}(k+1|k)$ and noting that $G(k+1|k) = G(k)$, we obtain

$$\underset{\tilde{u}}{P}(k) = \underset{\tilde{u}}{P}(k|k-1) - G(k)[\tilde{M}(k)\underset{\tilde{u}}{P}(k|k-1) + C^T(k-1)]$$

or

$$\underset{\tilde{u}}{P}(k) = \underset{\tilde{u}}{P}(k-1) - G(k)[\tilde{M}(k)\underset{\tilde{u}}{P}(k-1) + \tilde{M}(k)Q(k-1) + C^T(k-1)] + Q(k-1)$$

for $k = 1,2,\ldots..$ For $k = 0$, it is clear that $\underset{\tilde{u}}{P}(0) = P_u(1)$ by (2.27b). QED

We see from the theorem that, as we expected from the results of Chapter 3, the proposed filter has exactly the same dynamical structure as the single-stage predictor discussed in the preceding section. Thus, the same dynamics can be used to generate the filtered and predicted estimates. The error covariance matrices of these estimates can be computed by the recursive equations (5.29) and (5.39) or after one of them is computed by one of these equations then the other follows from the relation (cf. (5.16)):

$$\underset{\tilde{u}}{P}(k+1|k) = \underset{\tilde{u}}{P}(k) + Q(k) \qquad k = 01,2,\ldots .$$

The optimal filtering equations for the signal $x_k = \Phi(k)u_k$, such that $u_k \in R(\Phi^T(k))$, is obtained from (5.35) by making use of (2.22) as follows:

Optimal estimate:

$$(3.40) \qquad \hat{x}_k = \Phi(k)\Phi^+(k-1)\hat{x}_{k-1} + K(k)[y_k - \Phi(k,k-1)y_{k-1} - \tilde{H}(k)\hat{x}_{k-1}]$$

for $k = 1,2,\ldots$ with the initial condition $\hat{x}_o = 0$, where

$$(3.41)\quad K(k) = [P_{\tilde{x}}(k|k-1)H^T(k) + \Phi(k)C(k-1)]P_{\tilde{y}}^{-1}(k|k-1)$$

$$(3.42)\quad \tilde{H}(k) = H(k) - \Psi(k,k-1)H(k-1) .$$

Error covariance matrix:

$$(3.43)\quad P_{\tilde{x}}(k) = P_{\tilde{x}}(k|k-1) - K(k)[\tilde{H}(k)P_{\tilde{x}}(k|k-1) + C^T(k)\Phi^T(k)]$$

for $k = 1,2,\ldots$ with $P_{\tilde{x}}(0) = P_x(0)$ as the initial condition.

We remark at this point that the filter and predictor for the signal $x_k = \Phi(k)u_k$ have different gain matrices as opposed to the filter and predictor for the signal u_k where they have the same gain matrices (cf. (5.25), (5.35)).

5.4 OPTIMAL SMOOTHING FOR CNP

To complete our study of optimal estimation of the signal $x_k = \Phi(k)u_k$ under colored noisy observation, we examine the optimal smoothing problem. We proceed as in the preceding chapter by deriving a general formula for optimal smoothing.

We observe from the innovation Lemma 5.17 that the observation process has full rank. Hence, by (2.27c), the optimal smoothing $(\hat{u}_{k|\ell},\ P_{\tilde{u}}(k|\ell))$ $k < \ell$ is accomplished as follows with $(\hat{u}_k,\ P_{\tilde{u}}(k))$ as the initial condition:

$$\hat{u}_{k|\ell} = \hat{u}_k + \sum_{i=k+1}^{\ell} G(k,\ i|i-1)\tilde{y}_{i|i-1} ,$$

$$G(k,i|i-1) = [\tilde{u}_{k|i-1},\ \tilde{y}_{i|i-1}][\tilde{y}_{i|i-1},\tilde{y}_{i|i-1}]^{-1},$$

$$P_{\tilde{u}}(k|\ell) = P_{\tilde{u}}(k) - \sum_{i=k+1}^{\ell} G(k,\ i|i-1)[\tilde{y}_{i|i-1},\tilde{u}_{k|i-1}]$$

where by (5.17)

$$\tilde{y}_{i|i-1} = y_i - \Psi(i,i-1)y_{i-1} - \tilde{M}(i)\hat{u}_{i|i-1}$$

and

$$\begin{aligned}[\tilde{y}_{i|i-1},\tilde{y}_{i|i-1}] &= P_{\tilde{y}}(i|i-1) \\ &= \tilde{M}(i)P_{\tilde{u}}(i)\tilde{M}^T(i) + M^T(i)Q(i-1)M^T(i) + M(i)C(i-1) \\ &\quad + C^T(i-1)M^T(i) + P_n(i-1)\end{aligned}$$

Note that the only unknown in the above equations is the $n \times m$ matrix $[\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}]$. As in the preceding two chapters, this matrix is determined as follows:

$$(5.21) \Rightarrow [\tilde{u}_{k|i-1},\tilde{y}_{i|i-1}] = [\tilde{u}_{k|i-1},\tilde{M}(i)\tilde{u}_{i|i-1} + \Psi(i,i-1)M(i)(u_i - u_{i-1}) + \Psi(i,o)(m_i - m_{i-1})]$$

$$(1.4),\ k \le i+1,\ (5.10) \Rightarrow \tilde{u}_{k|i-1} \perp u_i - u_{i-1},\ m_i - m_{i-1}$$

Therefore,

$$\begin{aligned}[\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}] &= [\tilde{u}_{k|i-1}, \tilde{M}(i)\tilde{u}_{i|i-1}] \\ &= P_{\tilde{u}}(k, i|i-1)\tilde{M}^T(i) \qquad (5.44)\end{aligned}$$

where $P_{\tilde{u}}(k, i|i-1) \triangleq [\tilde{u}_{k|i-1}, \tilde{u}_{i|i-1}]$. The problem is now to find the expression for the cross-covariance matrix $P_{\tilde{u}}(k,i|i-1)$. Noting that

$$\begin{aligned}\tilde{u}_{k|\ell} &\triangleq u_k - \hat{u}_{k|i} \\ &= \tilde{u}_{k|i-1} - G(k, i|i-1)\tilde{y}_{i|i-1} \qquad k < i\end{aligned}$$

and

$$\tilde{u}_{i+1|i} = \tilde{u}_{i|i-1} - G(i+1|i)\tilde{y}_{i|i-1} + u_{i+1} - u_i \quad \text{by use of (5.26)}$$

we write

$$[\tilde{u}_{k|i}, \tilde{u}_{i+1|i}] = [\tilde{u}_{k|i-1} - G(k,i|i-1)\tilde{y}_{i|i-1}, \tilde{u}_{i|i-1} - G(i+1|i)\tilde{y}_{i|i-1} + u_{i+1} - u_i].$$

From the above, we know that $\tilde{u}_{k|i-1} \perp u_{i+1} - u_i$ and it can easily be shown that $u_{i+1} - u_i \perp \tilde{y}_{i|i-1}$. Therefore

$$\begin{aligned} P_{\tilde{u}}(k,i+1|i) &\triangleq [\tilde{u}_{k|i}, \tilde{u}_{i+1|i}] \\ &= P_{\tilde{u}}(k,i|i-1) - G(k,i|i-1)[\tilde{y}_{i|i-1}, \tilde{u}_{i|i-1}] - [\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}]G^T(i+1|i) \\ &\quad + G(k,i|i-1)[\tilde{y}_{i|i-1}, \tilde{y}_{i|i-1}]G^T(i+1|i) \\ &= P_{\tilde{u}}(k,i|i-1) - [\tilde{u}_{k|i-1}, \tilde{y}_{i|i-1}][\tilde{y}_{i|i-1}, \tilde{y}_{i|i-1}]^{-1}[\tilde{y}_{i|i-1}, \tilde{u}_{i|i-1}] \end{aligned}$$

where to derive the last equality, we used (2.29) and (2.34). Substituting (5.24), (5.30) and (5.44) into this equation yields the recursion

$$P_{\tilde{u}}(k,i+1|i) = P_{\tilde{u}}(k,i|i-1) - P_{\tilde{u}}(k,i|i-1)\tilde{M}^T(i)P_{\tilde{y}}^{-1}(i|i-1)[\tilde{M}^T(i)P_{\tilde{u}}(i|i-1) + C^T(i-1)]$$

for $k = k, k+1, \ldots$ with $P_{\tilde{u}}(k,k+1|k) = P_{\tilde{u}}(k)$ as the initial condition.

In summary, we have found that the optimal smoothing $(\hat{u}_k, P_{\tilde{u}}(k|\ell))$ is accomplished via the following equations with $(\hat{u}_k, P_{\tilde{u}}(k))$ as the initial condition:

$$(5.44) \quad \hat{u}_{k|\ell} = \hat{u}_k + \sum_{i=k+1}^{\ell} G(k,i|i-1)[y_i - \Psi(i,i-1)y_{i-1} - \tilde{M}(i)\hat{u}_{i|i-1}]$$

$$(5.45) \quad G(k,i|i-1) = P_{\tilde{u}}(k,i|i-1)\tilde{M}^T(i)P_{\tilde{y}}^{-1}(i|i-1)$$

$$(5.46)\quad \underset{\tilde{u}}{P}(k,i+1|i-1) = \underset{\tilde{u}}{P}(k,i|i-1) - G(k,i|i-1)[\tilde{M}(i)\underset{\tilde{u}}{P}(i|i-1) + C^{T}(i-1)],\ \underset{\tilde{u}}{P}(k,k+1|k) = \underset{\tilde{u}}{P}(k),$$

$$(5.47)\quad \underset{\tilde{u}}{P}(k|\ell) = \underset{\tilde{u}}{P}(k) - \sum_{i=k+1}^{\ell} G(k,i|i-1)\tilde{M}(i)\underset{\tilde{u}}{P}(i,k|i-1)\ .$$

These optimal estimation equations for the signal u_k, which are valid for all the classes of smoothing hold for the signal $x_k = \Phi(k)u_k$ such that $u_k \in R(\Phi^{+}(k))$ as in the preceding two chapters when u is replaced by x and $\tilde{M}$ is replaced by $\tilde{H}$ which is defined by (5.33). For the class of Kalman signals, the proposed smoother, i.e., (5.44)-(5.47) is apparently new.

The forms of the optimal estimation equations (5.44)-(5.47) for the single-stage and fixed-point smoothing can easily be obtained by repeating the steps leading to analogous results in Chapter 3. We state only the results.

SINGLE-STAGE SMOOTHING. Optimal single-stage smoothing $(\hat{u}_{k|k+1},\ \underset{\tilde{u}}{P}(k|k+1))$ $k = 0,1,2,\ldots$ for the signal u_k is accomplished via the following equations with $(\hat{u}_o,\ \underset{\tilde{u}}{P}(0)) = (0,P_u(0))$ as the initial condition:

$$\hat{u}_{k|k+1} = \hat{u}_k + G(k,k+1|k)[y_{k+1} - \Phi(k+1,k)y_k - \tilde{M}(k+1)\hat{u}_{k+1|k}]$$

$$G(k,k+1|k) = \underset{\tilde{u}}{P}(k)\tilde{M}^{T}(k+1)P_{\tilde{y}}^{-1}(k+1|k)$$

$$\underset{\tilde{u}}{P}(k,k+1|k) = \underset{\tilde{u}}{P}(k)$$

$$\underset{\tilde{u}}{P}(k|k+1) = \underset{\tilde{u}}{P}(k) - G(k,k+1|k)\tilde{M}(k+1)\underset{\tilde{u}}{P}(k)$$

where all the terms are defined as before.

FIXED-POINT SMOOTHING. Optimal fixed-point smoothing $(\hat{u}_{k|\ell}, P_{\tilde{u}}(N|\ell))$ $\ell = N+1, N+2, \ldots$ N = fixed positive integer, is accomplished via the following equations with the initial condition $(\hat{u}_N, P_{\tilde{u}}(N))$:

$$\hat{u}_{k|\ell} = \hat{u}_{N|\ell-1} + G(N,\ell|\ell-1)[y_\ell - \Psi(\ell,\ell-1)y_{\ell-1} \tilde{M}(\ell)\hat{u}_{\ell|\ell-1}]$$

$$G(N,\ell|\ell-1) = P_{\tilde{u}}(N,\ell|\ell-1)\tilde{M}^T(\ell)P_{\tilde{y}}^{-1}(\ell|\ell-1)$$

$$P_{\tilde{u}}(N,\ell+1|\ell) = P_{\tilde{u}}(N,\ell|\ell-1) - G(N,\ell|\ell-1)[\tilde{M}(\ell)P_{\tilde{u}}(\ell|\ell-1) + C^T(\ell-1)]$$

$$P_{\tilde{u}}(N,N+1|N) = P_{\tilde{u}}(N)$$

$$P_{\tilde{u}}(N|\ell) = P_{\tilde{u}}(N|\ell-1) - G(N,\ell|\ell-1)\tilde{M}(\ell)P_{\tilde{u}}(\ell,N|\ell-1)$$

where all the terms are defined as before.

CHAPTER 6

CONCLUSIONS

This chapter includes a discussion of the main objectives of thesis and possible extensions.

6.1 CONCLUSIONS AND RESULTS

The dissertation presents the derivation of the optimal minimum mean-square estimation equations for the signal model $x_k = \Phi(k)u_k$, $k = 01,2,,...$ where $\Phi(k)$ is an $n \times n$ matrix and u_k is a wide-sense martingale process. The primary results are theorems which demonstrate the recursive and algebraic optimal estimation equations for prediction, filtering and smoothing when observations are corrupted by additive uncorrelated white, cross-correlated white, and cross-correlated colored noises.

After briefly describing the discrete-time linear estimation problem and a literature review, Chapter 1 discusses the statement of the problem. A literature review suggests that the signal used in the Kalman approach can always be written in the form $x_k = \Phi(k)u_k$ where $\Phi(k)$ is an $n \times n$ invertible matrix and u_k is a wide-sense martingale. This, in turn, suggests that the optimal estimation equations for a signal $x_k = \Phi(k)u_k$, where $\Phi(k)$ is not necessarily an invertible matrix and u_k is as above, could be applied in an investigation of the prediction, filtering and smoothing problems of Kalman.

In Chapter 2, the Hilbert space of random vectors, multi-variate wide-sense martingale and wide-sense Markov processes are briefly discussed. It is shown in this chapter also that the optimal estimation problem of a second order signal (stochastic process) is equivalent to determination of two matrices assuming that the observation process has full rank.

Chapters 3, 4 and 5 include the derivation of optimal prediction, filtering, and smoothing equations for the signal $x_k = \Phi(k)u_k$ when observations are corrupted by additive uncorrelated white noise, cross-correlated white noise (Ch. 4) and cross-correlated colored noise (Ch. 5). Several new results for Kalman signals have been obtained. For example, a new and simple approach to discrete-time linear smoothing problems is developed in Chapters 3, 4, 5. The optimal estimation equations for cross-correlated colored noise problems are apparently new for Kalman signals.

In summary, this thesis gives a new approach to solving the prediction, filtering and smoothing problems of Kalman and their extensions, for a more general class of signals. Existing derivations of Kalman filtering in the simplest case are complicated. The complications are due to unnecessary analytic assumptions on the signal model. The method is developed in this thesis is algebraic in nature and gives simple derivation of Kalman filtering in all different cases. This avoids analytic assumptions and a distinct approach to each case.

This work has been directed at the theoretical foundations of discrete-time linear estimation and has not considered detailed applications. It is hoped that this new approach will provide a

basis for such applications.

6.2 EXTENSIONS

There are a number of topics for further research which are suggested by this work, for example:

(1) The present approach can be applied to yield results for the infinite-dimensional discrete-time case.

(2) Because of the availability of innovation decomposition in continuous-time [K-8], this approach again can be extended to continuous-time case involving wide-sense Markov signals covering Falb's work [F-1], using [M-3].

(3) The impact on stochastic optimal control of these new approaches to linear estimation should be explored.

(4) The question of asymptotic behavior of the filter in view of this approach should be investigated.

The problems (1) and (2) are partially settled by the author and will be completed in a subsequent work.

REFERENCES

REFERENCES

[A-1] Aoki, M. (1967). "Optimization of Stochastic Systems, Topics in Discrete-Time Systems", Academic Press, N.Y.

[A-2] Åström, Karl J. (1970). "Introduction to Stochastic Control Theory", Academic Press, N.Y.

[B-1] Beutler, F.J. (1963). "Multivariate Wide-Sense Markov Processes and Prediction Theory"; Ann. Math. Statist. 34, (424-438).

[B-2] Blackman, R.B., H.W. Bode, and C.E. Shannon. (1948). "Data Smoothing and Prediction in Fire-Control Systems", Research and Development Board, Washington, D.C., August 1948.

[B-3] Bode, H.D. and C.E. Shannon (1950). "A Simplified Derivation of Linear Least Squares Smoothing and Prediction Theory"; Proc. IRE, 38, (417-425).

[B-4] Brown, J.L., Jr. (1962). "Asymmetric non-mean-square Error Criteria", IRE, Trans. PGAC, AC-7, (64-66).

[B-5] Bryson, A.E. and D.E. Johansen (1965). "Linear Filtering for Time Varying Systems Using Measurements Containing Colored Noise"; IEEE Trans. Automatic Control, AC-10, (4-10).

[B-6] Bryson, A.E. and L.J. Henrikson (1968). "Estimation Using Sampled-Data Containing Sequentially Correlated Noise"; J. Spacecraft 5, (662-665).

[B-7] Bryson, A.E., Jr. and Yu-Chi Ho (1969). "Applied Optimal Control, Optimization, Estimation, and Control"; Blaisdell Publ. Co.

[B-8] Bucy, R.S. (1967). "Optimal Filtering for Colored-Noise"; Journal of Mathematical Analysis and its Applications 20, (1-8).

[B-9] Bucy, R.S. and Peter D. Joseph (1968). "Filtering for Stochastic Processes with Applications to Guidance"; Interscience Publishers.

[C-1] Cox, H. (1964). "Estimation of State Variables via Dynamic Programming"; 1964 JACC Proceedings, June 1964, (376-381).

[C-2] Cramér, H. (1961). "On some classes of non-stationary stochastic processes"; Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press, 2, (57-78).

[C-3] Cramér, H. (1966). "A contribution to the multiplicity theory of stochastic process"; Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press, 2, (215-221).

[D-1] Deutsch, R. (1965). "Estimation Theory", Prentice-Hall.

[D-2] Doob, J.L. (1953). "Stochastic Processes"; John Wiley & Sons, Inc., N.Y., Third Printing.

[F-1] Falb, P.L. (1968). "Infinite Dimensional Filtering: The Kalman-Bucy Filter in Hilbert Space"; Information and Control, 11, (102-137).

[F-2] Fujita, S. and T. Fukas (1970). "Optimal Linear Fixed-Interval Smoothing for Colored Noise"; Information and Control, 17, (313-325).

[G-1] Gauss, K.F. (1963). "Theory of the Motion of the Heavenly Bodies about the Sun in Conic Sections"; New York: Dover Publications, Inc. (reprint).

[G-2] Gikhman, I.I. and A.V. Skorokhod (1969). "Introduction to the Theory of Random Processes"; W.B. Saunders Comp., Philadelphia.

[J-1] Jazwinski, A.H. (1970). "Stochastic Processes and Filtering Theory"; Academic Press.

[K-1] Kailath, T. (1968). "An Innovation Approach to Least-Square Estimation, Part I: Linear Filtering in Additive White Noise"; IEEE on Control, AC-13, (645-654).

[K-2] Kailath, T. and P. Frost (1968). "An Innovation Approach to Least-Square Estimation, Part II: Linear Smoothing in Additive White Noise"; IEEE on Control, AC-13, (655-660).

[K-3] Kalman, R.E. (1960). "A New Approach to Linear Filtering and Prediction Problems"; Trans. ASME, J. Basic Engrg., 82, (34-45).

[K-4] Kalman, R.E. (1963). "New Methods in Wiener Filtering Theory"; Proc. 1st Symp. on Engrg. Applications of Random Function Theory and Probability; J.L. Boydanoff and F. Kozin, Eds.: Wiley.

[K-5] Kalman, R.E. (1969). "Lectures on Controllability and Observability"; Lectures delivered at CENTRO INTERNAZIONALE MATEMATICO ESTIVO (C.I.M.E.), Corso tenuto a Sasso Marconi (Bologna).

[K-6] Kalman, R.E. and R.S. Bucy (1961). "New Results in Linear Prediction and Filtering Theory"; J. Basic Engr. (Trans. ASME, ser. D), 83, (95-100).

[K-7] Kalman, R.E. and T.S. Englar (1966). "A User's Manual for Automatic Synthesis Program"; Washington, D.C.

[K-8] Kallianpur, G. and V. Mandrekar (1965). "Multiplicity and Representation Theory of Purely Non-deterministic Stochastic Processes"; Teor. Veroyatnost. i Primenen, X, USSR.

[K-9] Kolmogorov, A.N. (1941). "Interpolation and Extrapolation of Stationary Random Sequences"; Bull. Acad. Sci. USSR, Math. Ser. 5, A translation has been published by the RAND Corp., Santa Monica, Calif., as Memo. RM-3090-PR.

[L-1] Leondes, C.T. (Editor) (1970). "Theory and Applications of Kalman Filtering"; AGARDograph, No. 139.

[L-2] Liebelt, P.B. (1967). "An Introduction to Optimal Estimation Theory"; Addison-Wesley.

[M-1] Mandrekar, V. (1968). "On Multivariate Wide-sense Markov Processes"; Nagoya Math. J., 33, (7-19).

[M-2] Mandrekar, V. (1970). "Probability and Stochastic Processes, Unpublished class notes", M.S.U., East Lansing, Michigan.

[M-3] Mandrekar, V. and H. Salehi (1970). "Operator-valued Wide-sense Markov Processes and Solutions of Infinite Dimensional Linear Differential Systems Driven by White Noise"; Mathematical Systems Theory, 4, (340-356).

[M-4] Meditch, J.S. (1967a). "Orthogonal Projection and Discrete Optimal Linear Smoothing"; SIAM J. Control, 5, (74-89).

[M-5] Meditch, J.S. (1967b). "On Optimal Linear Smoothing Theory"; Information and Control, 10, (598-615).

[M-6] Meditch, J.S. (1969). "Stochastic Optimal Linear Estimation and Control"; McGraw-Hill.

[M-7] Mehra, R.K. and A.E. Bryson (1968). "Linear Smoothing Using Measurements Containing Correlated Noise with an Application to Inertial Navigation"; IEEE Trans. Auto. Control, AC-13, (496-503).

[M-8] Miller, K.S. (1968). "Linear Difference Equations"; W.A. Benjamin, Inc., N.Y.

[N-1] Nahi, N.E. (1969). "Estimation Theory and Applications"; John Wiley and Sons, Inc., N.Y.

[P-1] Penrose, R. (1955). "A Generalized Inverse for Matrices"; Proc. Cambridge Philos. Soc., 51, (406-413).

[P-2] Porter, William A. (1967). "Modern Foundations of Systems Engineering"; Macmillan Company, N.Y.

[R-1] Rauch, H.E. (1963). "Solutions to the Linear Smoothing Problem"; IEEE Trans. on Automatic Control, AC-8, (371-372).

[R-2] Rauch, H.E., F. Tung, and C.T. Striebel (1965). "Maximum Likelihood Estimates of Linear Dynamic Systems"; AIAAJ., 3, (1445-1450).

[R-3] Royden, H.L. (1968). "Real Analysis"; Mcmillan Company, N.Y.

[S-1] Sage, A.P. (1968). "Optimum Systems Control"; Prentice-Hall, Inc.

[S-2] Sage, A.P. and J.L. Mepsa (1971). "Estimation Theory with Applications to Communication and Control"; McGraw-Hill.

[S-3] Sherman, S. (1958). "Non-Mean-Square Error Criteria"; Trans. IRE. Proc. Group on Information Theory, IT-4, (125-126).

[S-4] Stear, E.B. and A.R. Stubberud (1968). "Optimal Filtering for Gauss-Markov Noise"; Int. J. Contr. 8, (123-130).

[W-1] Wiener, N. (1949). "The Extrapolation, Interpolation and Smoothing of Stationary Time Series"; John Wiley and Sons, Inc., New York, N.Y.

[W-2] Wiener, N. and E. Hopf (1931). "On a Class of Singular Integral Equations"; Proc. Prussian Acad., Math.-Phys. Ser., (696).

[W-3] Wiener, N. and P. Masani (1957). "The Prediction Theory of Multivariate Stochastic Processes I"; Acta Math., 98, (111-149).

[W-4] Wiener, N. and P. Masani (1958). "The Prediction Theory of Multivariate Stochastic Processes II"; Acta Math. 99, (93-137).

[W-5] Willman, W.W. (1969). "On the Linear Smoothing Problem"; IEEE Trans. Automatic Control, AC-14, (116-117).

[W-6] Wold, H. (1938). "A Study in the Analysis of Stationary Time Series"; Sweden: Almgvist and Wiksell.

[Z-1] Zachrisson, L.E. (1969). "On Optimal Smoothing of Continuous Time Kalman Processes"; Inform. Sci., 1, (143-172).

[Z-2] Zadeh, L.A. and C.A. Dasoer (1963). "Linear System Theory: The State Space Approach"; McGraw-Hill Book Company, N.Y.

[Z-3] Zadeh, L.A. and J.R. Ragazzini (1950). "An Extension of Wiener's Theory of Prediction"; J. Appl. Phys., 21, (645-655).

[Z-4] Zakai, M. (1964). "General Error Criteria"; IEEE, Trans. PGIT, IT-10, (94-95).

[Z-5] Zimmerman, W. (1969). "On the Optimum Colored Noise Kalman Filter"; IEEE Trans. Automatic Control, AC-14, (194-196).

APPENDICES

APPENDIX A

GENERALIZED INVERSES

This appendix defines and reviews briefly some properties of generalized inverses of linear operators on a real Euclidean space R^n. Before starting with the definition of generalized inverse we recall some concepts of linear transformations.

Let A be a linear mapping with domain $D(A)$ in the n-dimensional space R^n into the m-dimensional space R^m. In the following we shall not distinguish between the linear transformation A and its $m \times n$ matrix representation.

(A.1) DEFINITION. (a) The null space of A is the set $N(A)$ defined by

$$N(A) = \{x \mid Ax = 0,\ x \in D(A)\}.$$

(b) The range of A is denoted by $R(A)$ and is given by

$$R(A) = \{y \mid y = Ax,\ x \in D(A)\}.$$

It is trivial that $N(A)$ and $R(A)$ are linear subspaces of R^n and R^m respectively (cf. [P-2], p. 94).

(A.2) THEOREM. Let A be a linear transformation from R^n into R^m. Then

(a) $R^n = R(A^T) \oplus N(A)$, and $R^m = R(A) \oplus N(A^T)$;

(b) $N^{\perp}(A) = R(A^T)$, and $N(A^T) = R^{\perp}(A)$;

(c) A is one-one mapping of $R(A^T)$ onto $R(A)$.

PROOF. (cf. [Z-2], Appendix C).

Now, let us turn to the generalized inverse. There are several ways of defining the generalized inverse of a linear transformation. Here we have chosen the following.

(A.3) DEFINITION. Let A be a linear transformation on R^n into R^m. A^+ is the generalized inverse of A if

$$AA^+ = P_{R(A)} ,$$

$$A^+A = P_{R(A^T)} .$$

where $P_{R(A)}$ is the orthogonal projection operator onto the subspace $R(A)$ (cf. Definition 2.).

Penrose [P-1] has an alternative definition which could be shown to be equivalent to (A.3).

Some properties of generalized inverse are given in the following theorem. The proof of this theorem can be found e.g. ([Z-2], [K-7]).

(A.4) THEOREM (Properties of A^+).

(a) A^+ is a linear transformation from R^m into R^m with the range $R(A^+) = R(A^T)$ and the null space $N(A^+) = N(A^T)$.

(b) $(A^+)^+ = A$ and $(A^T)^+ = (A^+)^T$.

(c) $AA^+A = A$ and $A^+AA^+ = A^+$.

(d) $A^+ = A^{-1}$ if A^{-1} exists.

APPENDIX B

MATRIX INVERSION LEMMA

In this appendix certain matrix equalities which are used in the dissertation will be derived. These results can be found in the standard textbooks on the estimation theory or control theory (e.g. see [J-1], Appendix 7B).

In the following P, R and M denote $n \times n$, $m \times m$ and $m \times n$ matrices respectively.

(B.1) LEMMA: $P \geq 0$ and $R > 0 \Rightarrow (I + PM^T R^{-1} M)^{-1} = I - PM^T (MPM^T + R)^{-1} M$.

PROOF. Since

$$\begin{aligned}(I + PM^T R^{-1} M)(I - PM^T (MPM^T + R)^{-1} M) &= I + PM^T R^{-1} M - PM^T (MPM^T + R)^{-1} M - PM^T R^{-1} MPM^T (MPM^T + R)^{-1} M \\ &= I + PM^T R^{-1} M - PM^T R^{-1} (R + MPM^T)(MPM^T + R)^{-1} M \\ &= I,\end{aligned}$$

and similarly $(I - PM^T (MPM^T + R)^{-1} M)(I + PM^T R^{-1} M) = I$, by the definition inverse of a matrix

$$(I + PM^T R^{-1} M)^{-1} = I - PM^T (MPM^T + R)^{-1} M . \qquad \text{QED}$$

(B.2) LEMMA. $P \geq 0$ and $R > 0 \Rightarrow (I + PM^T R^{-1} M)^{-1} PM^T R^{-1} = PM^T (MPM^T + R)^{-1}$.

PROOF. Multiply (B.1) on the right by $PM^T R^{-1}$; obtain

$$(I + PM^TR^{-1}M)^{-1}PM^TR^{-1} = PM^TR^{-1} - PM^T(MPM^T + R)^{-1}MPM^TR^{-1}$$

$$= PM^TR^{-1} - PM^T(MPM^T + R)^{-1}MPM^TR^{-1} - PM^T(MPM^T + R)^{-1}$$

$$- PM^T(MPM^T + R)^{-1}RR^{-1}$$

$$= PM^T(MPM^T + R)^{-1} . \qquad \text{QED}$$