National Institute of Electricity and Electronics

INELEC - BOUMERDES

DEPARTMENT OF RESEARCH

THESIS

Presented in partial fulfilment of the requirements of the

DEGREE OF MAGISTER

in Electronic System Engineering

by

Rachid BOUCHETA

Statistical Smoothing Methods For stationary random signals

Defended on June 01, 1994 before the jury:

President: Dr A. HOUACINE, Maitre de Conferences, U.S.T.H.B.

Members: Dr M. DJEDDI, INELEC.

Dr K. HARICHE, INELEC.

Dr A. KEZOUH, Maitre de conferences, U.S.T.H.B.

Registration Number 01/94.

DEDICATIONS

To my

parents

brothers

and

sisters

ACKNOWLEDGEMENTS

I want to express my most sincere gratitude to my supervisor Dr. M. Djeddi for his continued and untiring interest and research guidance throughout this work. I would like also to acknowledge Dr. H. Bourdoucen for his material assistance and comprehension.

I am specially grateful to Dr. K. Harich, head of the research department, for providing all necessary research facilities and advise to make our project a success.

I wish also to thank all the staff of INELEC for providing us all the necessary help.

Finally, I would like to thank all my friends who provided help and material support for the best achievement of this work.

TABLE OF CONTENTS

LIST OF SYMBOLS

$A(n+1,n)$	transition matrix
$C(n)$	observation matrix
E	expectation operator
$K(n)$	Kalman gain
$g(n)$	input matrix
I_n	n by n identity matrix
$J_{1,2}$	object functions to minimize
M	smoothing matrix
m	integer
n	integer " discrete time index"
N	number of data samples
p	signal model order
$P_{1,2}$	smoothing parameters
$P(n)$	one step prediction error covariance matrix
$R(n)$	output noise covariance matrix
$r(n)$	residuals
$Q(n)$	input noise covariance matrix
$s(n)$	the actual signal
$\hat{s}(n)$	filtered signal
S	smoothing parameter
$Z(n)$	space of the measurements
$x(n)$	state vector

$\hat{x}(n/n-k)$	k-step-ahead state vector predictor
$v(n)$	system noise
$V_e(n)$	innovation covariance
$w(n)$	output noise
$z(n)$	measurement signal
$\alpha(n)$	innovations
$\beta_{1,2}$	dummy variables
$\Gamma(n)$	filtering error covariance matrix
$\delta(n)$	Dirac function
π	product
$\sum$	summation
λ	smoothing parameter
ρ	smoothing parameter
σ	output noise variance
$\xi(./.)$	conditional mean

ABBREVIATIONS:

AMSE	average mean squared estimation error
Cov	covariance
CV	cross validation
Eq	equation
GXV	generalized cross validation
IMSE	integrated mean square error
lim	limite
LSS	local smoothing spline

MKS	modified Kalman smoother
MSE	mean squared estimation error
SS	smoothing spline
OGSP	optimum global smoothing parameter

CHAPTER I:

INTRODUCTION

In many applications, engineers and scientists are faced with the problem of measuring a quantity to predict a physical phenomenon, control a specified mechanism, or infer a quantity which is not directly measurable. In these cases, a filter or an estimator is built to remove unwanted information which may be deterministic or random. If the measured signal varies extraneously and is not repeatable, then it is called a random signal. In this case, an estimator must be built to extract useful information. In contrast to filters which are built to remove deterministic disturbances, scientists are, in general, concerned with the development of signal-processing techniques to extract pertinent signal information from random signals using prior useful information. This is known as the construction of an estimator. The estimation problem can be thought of as a procedure which consists of two main parts:

(1)_ Choice of the family model,

(2)_ Selection of the best element according to a criterion function.

The most popular criterion function in signal extraction is the minimisation of the mean squared error of the estimator. Models represent a broad class of information formalizing the apriori knowledge about the process generating the signal and noise characterization. A popular family of models is that of an observed signal with an additive noise:

Measurement = signal + noise

where the noise statistics and the signal model are to be specified.

The estimation procedure is usually chosen taking into account relevant practical specifications and constraints. One important step in the estimation procedure is the choice of the technique used. The choice of the technique can take many different forms depending on the model, the representation of the solution which is seeked, and the chosen criterion function. For example, one may decide, based on computational consideration, to calculate an estimate recursively rather than as a result of a bath process because we require an on-line estimate. Furthermore, one may choose a nonparametric estimator rather than a model-based processor because the measurement at hand arise from a physical phenomenon which is not well understood so that the choice of a parametric family of models will be difficult to justify.

Each estimation procedure must provide a measure of estimation

quality, usually in terms of the expected estimation error which provides a means for comparing the efficiency of different estimators.

In the present work, we are interested in the performance of a recent parametric method of estimation known as the modified Kalman smoother, and that of a nonparametric statistical method of estimation known as the smoothing spline.

In chapter 2, we present a detailed derivation of the Kalman smoother which is used to obtain an algorithm for the modified Kalman smoother (MKS). In chapter 3, we present the smoothing spline and the generalized cross validation technique which is used to determine the correct degree of smoothing applied to the measurements. In chapter 4, The comparison of the modified Kalman smoother and the smoothing spline methods is carried out using simulated stationary smooth random signal observed in a noisy environment over a fixed period of time. The criterion used for the comparison is the average mean squared estimation error (AMSE). Furthermore, the robustness of the modified Kalman smoother against errors which may occur in the identification step of the parameters will be investigated. A new statistical test based on the bootstrap techniques is also suggested to investigate the validity of the model used in the construction of the modified Kalman smoother. The computational complexity will also be presented. In chapter 5, a modification of the standard smoothing spline will be proposed. It is a new estimator based on a local smoothing

parameter in stead of a global smoothing parameter used by the standard smoothing spline (SS). Its properties will be illustrated using simulated random processes. In chapter 6, concluding remarks are drawn and further work is proposed.

CHAPTER II:

THE MODIFIED KALMAN SMOOTHER

2.1 INTRODUCTION

The problem of statistical signal estimation from recorded data has been studied for several decades. The mostly used signal processing model is that of the signal corrupted with an additive noise:

$$z(t) = s(t) + w(t)$$

where $s(t)$ is the signal, $w(t)$ is the noise that corrupts the signal, and $z(t)$ is the observed measurement at time t.

The two random variables $s(t)$ and $w(t)$ are not observable. The only observed random variable is the measurements. The measurements $z(t)$, the models of $s(t)$ and $w(t)$ are used to obtain an estimate of the signal $s(t)$. In the litterature, there are two approaches for the estimation of $s(t)$:

1)_ The transfer function approach: In this approach, estimators are built by estimating the parameters of the transfer function of the assumed model of $s(t)$.

2)_ The state space representation approach: In this approach

a state space representation is used for modelling. In other words, the estimation of $s(t)$ reduces to that of the estimation of the state vector.

The chronological improvement of the model-based processor, known as the parametric methods, can be summarized as follows:

Kolmogorov [2], Wiener [2] and kalman [2,20,26] have reintroduced and reformulated the linear least square approach for the estimation of the signal $s(t)$.

For the first approach, Kolmogorov [2] studied discrete time problems ie: the time t is assumed to be discrete. He obtains the estimator using a recursive orthogonalisation procedure known as the Wold decomposition. Weiner [2] studied mainly the continuous time problem and arrived at the famous Wiener-Hopf integral equation. The solution of this equation is difficult to solve except in the case where the signal $s(t)$ and the noise $w(t)$ are stationary processes. The equivalent of this integral equation in the discrete case are the so-called "normal equations", which can be solved by a fast algorithm given by Levinson [6] in the case of stationary time series.

For the second approach, Kalman [20] has used a state space representation as a model for the signal $s(t)$ and treated the discrete time problem. Kalman and Bucy [21] have extended the theory to continuous time problem. A recursive estimate was proposed by Kalman and Bucy [21] to treat the cases where the process is generated by passing a white noise through a linear

time varying system and assuming a complete knowledge of the system described by state space representation, and the input and output noise covariance matrices. They approached the estimation problem from a statistical point of view (conditional distributions and expectations), and they used the concept of orthogonal projections to solve the problem. D. C. Fraser et al [13] have considered the estimation of $s(t)$ as a smoothing problem. In this case all the data (measurements) must be collected before the use of the estimation procedure. One of the optimal smoothers is the Kalman smoother [13] which is basically a combination of two optimum filters. The first one is run in a forward direction and the other one is run in a backward direction. The two filters are then combined in an optimal manner.

Kailath and Gevers [19] solved the recursive least squares estimation problem using the innovation process. Recently J. Vaccaro and Fu Li [32] have made a modification to the Kalman smoother and suggested an elegant and efficient algorithm for the case where the signal $s(t)$ is stationary, and it is observed in a noisy environment. The obtained estimator is a time-invariant Kalman smoothing estimator modified by a data-dependent correction term involving an estimated initial state vector.

In this chapter, we derive the Kalman filter, the Kalman smoother and we end up with the modified Kalman smoother.

2.2 STATE SPACE REPRESENTATION OF RANDOM SIGNALS.

Generally, we model a random signal as the output of a dynamical system driven by a white noise. The system is called dynamical when the output at any instant t depends not only on the input at time t but it also depends on the prior history of the system. One representation of dynamical systems which enables us to take into acount the intial condition of the system is the state space representation. Any set of variables that summarizes the past history of the system affecting its future behavior is said to be state variables of the system. More precisely, we define the state of a system at time n_o as the information at n_o, other than the input at time n_o, which is used to determine uniquely the output for $n > n_o$.

Given at time n_o a set of p state variables $x_1(n_o), x_2(n_o), \ldots, x_p(n_o)$, we can define a state vector $x(n_o)$ as

$$x(n_o) = [x_1(n_o) \quad x_2(n_o) \quad \ldots \quad x_p(n_o)]^T$$

The number p of state variables used in a description of a system is called the order of the system. The vector x(n) is called the state vector of a p-dimensional dynamical system with a state equation of the form which is a common representation of a p-order Markov process:

$$x(n+1) = A(n+1,n)x(n) + g(n)v(n) \qquad (2.2.1)$$

where A(n+1,n) is a pxp state transition matrix and g(n) is a pxl$_g$ input matrix. $v(n)$ is a sequence of zero mean l$_g$-vector random processes satisfying

$$cov\{v(k), v(n)\} = E\{v(k)v^T(n)\} = Q(n)\delta(n-k) \qquad (2.2.2)$$

where $Q(n)$ is the covariance matrix of the input noise, δ is the Dirac function, and l_g is the dimension of the input noise vector.

The observed data vector z(n) will be assumed to have the form

$$z(n) = C(n)x(n) + w(n) \tag{2.2.3}$$

where C(n) is an $l_z \times p$ matrix called the observation matrix and $w(n)$ denotes the noise introduced in the observation process. $w(n)$ is assumed to be a zero mean vector random process with

$$cov\{w(k), w(n)\} = E\{w(k)w^T(n)\} = R(n)\delta(n-k) \tag{2.2.4}$$

where $R(n)$ is the covariance matrix of the output noise.

The input and output noises, $v(n)$ and $w(n)$, are assumed to be uncorrelated. That is:

$$cov\{w(k), v(n)\} = 0 \tag{2.2.5}$$

We will assume that the initial state $x(1)$ is a zero mean vector random variable with

$$cov\{x(1), x(1)\} = P(1) \tag{2.2.6}$$

and it is uncorrelated with v(k) and w(n) for $k, n \geq 1$.

This model is illustrated in Fig:2.1

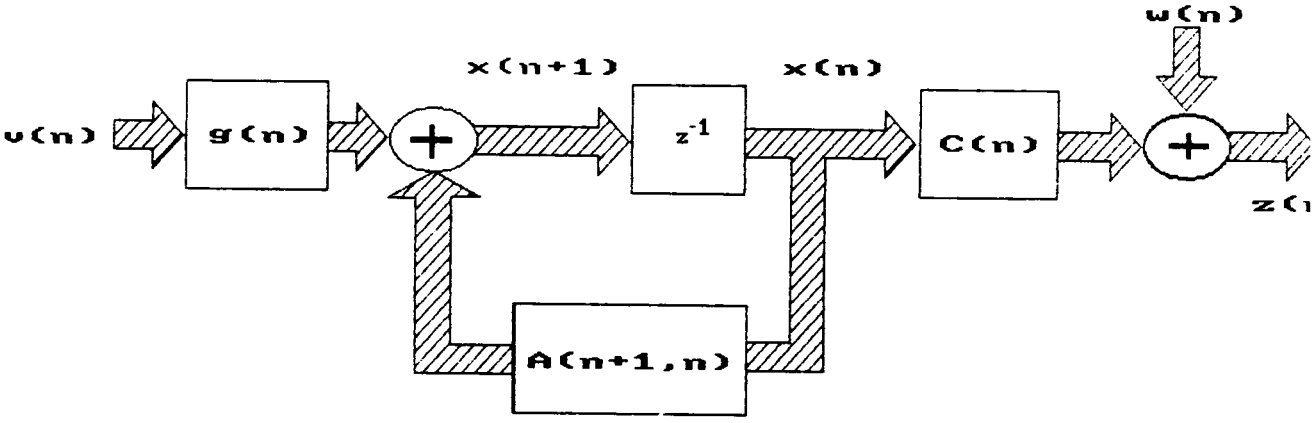


Fig:2.1 Model for the estimation problem.

In figure 2.1, z^{-1} stands for a backward unit time shift operator.

The problem is to find, for each $n \geq 1$, the linear minimum mean square error estimates of the components of x(i) from the data, $\{z(1),....,z(n)\}$, observed from time 1 up to the present time n $\{z(1),....,z(n)\}$ with $1 \leq i \leq n$. Therefore, this problem is a smoothing problem.

2.3 ESTIMATION USING THE INNOVATION PROCESS [18].

Consider the space Z(n-1) of scalar random variables generated by the components of the observed data $\{z(1),....,z(n-1)\}$. The optimum least square estimates of the components of x(i) from the observed data are the projections of the components of x(i) onto the space Z(n-1). We will denote these estimates by

$$\hat{x}(i/n-1) = \xi\{x(i)/Z(n-1)\} \qquad for \quad i = 1,...,n-1 \qquad (2.3.1)$$

We will use the innovation process to obtain the estimates $\hat{x}(i/n-1)$.

Definition 2.3.1: Assume that z(n) is a measurement vector process. Its innovation process $\alpha(n)$ is given by:

$$\alpha(n) = z(n) - \hat{z}(n/n-1) \qquad (2.3.2)$$

where

$$\hat{z}(n/n-1) = \xi\{z(n)/Z(n-1)\} \qquad (2.3.3)$$

which is the optimum least square prediction of z(n) given the measurement data observed up through time n-1. See Fig.2.2.

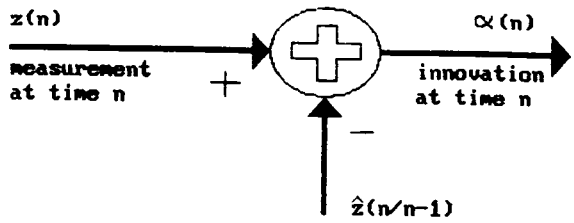


Fig.2.2: The innovation process

Since Z(n-1) is the space generated by the measurements $\{z(1),\ldots,z(n-1)\}$, using the projection theorem we see that $\alpha(n)$ is orthogonal to the space $Z(n-1)$. Therefore, we can easily derive that:

$$E\{\alpha(n)z^T(k)\} = 0 \qquad for \quad 1 \le k < n \qquad (2.3.4)$$

Hence $\alpha(n)$ is orthogonal to $Z(k)$ for any k such that $1 \le k < n$. In other words, $\alpha(n)$ is the new information contained in the observation z(n). See Fig:2.3.

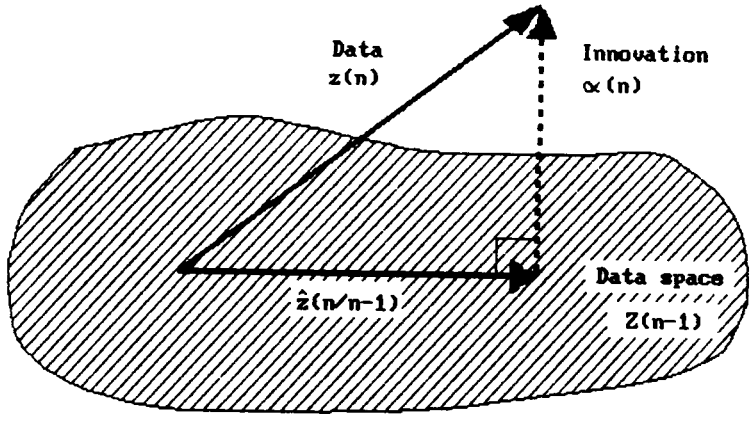


Fig.2.3:Orthogonal decomposition of the measurement space.

From Eq(2.3.2) it follows that $\alpha(k)$ lies in the space Z(k) which is spanned by $z(j)$ for $1 \le j \le k$. As a result

$$E\{\alpha(k)\alpha^T(n)\} = 0 \qquad for \quad k \neq n \qquad (2.3.5)$$

Therefore, the innovation process $\{\alpha(n)\}$ is a sequence of uncorrelated vector random variables.

Since the process $\{\alpha(1),\ldots,\alpha(n)\}$ spans the space generated by the measurement process $\{z(1),\ldots,z(n)\}$ and it forms an orthogonal set in this space, each zero mean finite variance vector process $x(i)$ can be written as a sum of a linear combination of $\alpha(1),\ldots,\alpha(n)$ and a term $x_{out}(i)$ describing the part of $x(i)$ that falls outside the space generated by $z(1),\ldots,z(n)$. That is:

$$x(i) = \sum_{k=1}^{n} L_i(k)\alpha(k) + x_{out}(i) \qquad (2.3.6)$$

where $\{L_i(k)\}$ is a set of pxl$_c$ constant matrices.

The linear least square estimate of this process based on the measurements $\{z(1),\ldots,z(n)\}$ is given by

$$\hat{x}(i/n) = \sum_{k=1}^{n} L_i(k)\alpha(k) \qquad for \quad i = 1,\ldots,n \qquad (2.3.7)$$

According to the projection theorem:

$$E\left\{\left[x(i) - \sum_{k=1}^{n} L_i(k)\alpha(k)\right]\alpha^T(m)\right\} = 0 \quad for \quad m = 1,\ldots,n \qquad (2.3.8)$$

Using Eq(2.3.5), Eq(2.3.8) reduces to

$$E\{x(i)\alpha^T(m)\} = L_i(m)E\{\alpha(m)\alpha^T(m)\}$$

If we denote by $V_\alpha(n) = cov\{\alpha(n),\alpha(n)\}$, we obtain:

$$L_i(m) = E\{x(i)\alpha^T(m)\}V_\alpha^{-1}(m)$$

Therefore,

$$\hat{x}(i/n) = \sum_{k=1}^{n} E\{x(i)\alpha^T(k)\}V_e^{-1}(k)\alpha(k) \qquad (2.3.9)$$

The k^{th} term in the sum is the projection of x(i) onto the subspace generated by $\alpha(k)$.

2.4 KALMAN FILTER

We derive the Kalman filter algorithm from the innovations view-point following the approach given by Kailath [18,19]. The Kalman filter will be, then, used for constructing an algorithm for the Kalman smoother [13,21,22,30].

The state equation (2.2.1) can be recursively solved to yield

$$x(k) = A(k,1)x(1) + \sum_{i=1}^{k-1} A(k,i+1)g(i)v(i) \quad for \ k>1 \qquad (2.4.1)$$

where $A(k,1) = \prod_{i=1}^{k-1} A(i+1,i)$ with $A(k,k) = I_p$, and I_p is the identity matrix of order p.

We notice that x(k) is a linear combination of $v(1),\ldots,v(k-1)$ and $x(1)$. Since the output noise w(n) is uncorrelated with $x(1)$ and v(k) for $n,k \geq 1$, we deduce that:

$$E\{x(k)w^T(n)\} = 0 \qquad for \quad n,k \geq 1 \qquad (2.4.2)$$

Therefore, according to Eq(2.2.3), the measurement $z(k)$ can be expressed as a linear combination of $x(1)$, $v(1),\ldots,v(k-1)$ and $w(k)$.

Using Eq(2.4.2), we can derive that

$$E\{z(k)w^T(n)\} = 0 \quad for \quad 1 \leq k \leq n-1 \qquad (2.4.3)$$

so that

$$\xi\{w(n)/Z(k)\} = 0 \qquad for \ 1 \leq k \leq n-1 \qquad (2.4.4)$$

and

$$E\{z(k)v^T(n)\} = 0 \qquad for \quad 1 \le k \le n \tag{2.4.5}$$

so that

$$\xi\{v(n)/Z(k)\} = 0 \qquad for \quad 1 \le k \le n \tag{2.4.6}$$

Consequently, $\hat{z}(n/n-1)$ can be expressed as:

$$\begin{aligned} \hat{z}(n/n-1) &= \xi\{C(n)x(n)+w(n)/Z(n-1)\} \\ &= C(n)\xi\{x(n)/Z(n-1)\}+\xi\{w(n)/Z(n-1)\} \\ &= C(n)\hat{x}(n/n-1) \end{aligned} \tag{2.4.7}$$

where $\hat{x}(n/n-1) = \xi\{x(n)/Z(n-1)\}$ is the optimum estimate of the vector x(n) obtained from the observed data vectors $\{z(1),\ldots,z(n-1)\}$ and the innovation can be written as:

$$\alpha(n) = z(n) - C(n)\hat{x}(n/n-1) \tag{2.4.8}$$

substituting Eq(2.2.3) for z(n), Eq(2.4.8) becomes

$$\alpha(n) = C(n)\tilde{x}(n/n-1)+w(n) \tag{2.4.9}$$

where

$$\tilde{x}(n/n-1) = x(n) - \hat{x}(n/n-1) \tag{2.4.10}$$

we can also deduce that

$$\begin{aligned} V_e(n) &= C(n)E\{\tilde{x}(n/n-1)\tilde{x}^T(n/n-1)\}C^T(n) \\ &\quad +C(n)E\{\tilde{x}(n/n-1)w^T(n)\} \\ &\quad +E\{w(n)\tilde{x}^T(n/n-1)\}C^T(n)+E\{w(n)w^T(n)\} \end{aligned} \tag{2.4.11}$$

The middle two terms on the right-hand side of Eq(2.4.11) are equal to zero as a result of Eq(2.4.2) and Eq(2.4.4). The quantity

$$P(n) = E\{\tilde{x}(n/n-1)\tilde{x}^T(n/n-1)\} \tag{2.4.12}$$

is the error covariance matrix for the optimum prediction of x(n) from the data observed up to time n-1. Thus

$$V_a(n) = C(n)P(n)C^T(n) + R(n) \tag{2.4.13}$$

According to Eq(2.3.9), the optimal estimate of x(n+1) obtained from the data observed up to time n is given by:

$$\hat{x}(n+1/n) = \sum_{k=1}^{n} E\{x(n+1)\alpha^T(k)\}V_a^{-1}(k)\alpha(k) \tag{2.4.14}$$

which can be written as

$$\hat{x}(n+1/n) = \sum_{k=1}^{n-1} E\{x(n+1)\alpha^T(k)\}V_a^{-1}(k)\alpha(k) + E\{x(n+1)\alpha^T(n)\}V_a^{-1}(n)\alpha(n) \tag{2.4.15}$$

Using Eq(2.4.6) and the fact that $\alpha(k)$ is in the space Z(k), it follows that for $1 \le k \le n$

$$\begin{aligned} E\{x(n+1)\alpha^T(k)\} &= E\{[A(n+1,n)x(n) + g(n)v(n)]\alpha^T(k)\} \\ &= A(n+1,n)E\{x(n)\alpha^T(k)\} \end{aligned} \tag{2.4.16}$$

Therefore,

$$\begin{aligned} \sum_{k=1}^{n-1} E\{x(n+1)\alpha^T(k)\}V_a^{-1}(k)\alpha(k) &= A(n+1,n)\sum_{k=1}^{n-1} E\{x(n)\alpha^T(k)\}V_a^{-1}(k)\alpha(k) \\ &= A(n+1,n)\hat{x}(n/n-1) \end{aligned} \tag{2.4.17}$$

We define K(n) by:

$$K(n) = E\{x(n+1)\alpha^T(n)\}V_a^{-1}(n) \tag{2.4.18}$$

The matrix K(n) is referred to as the Kalman gain for one step prediction.

Using Eqs(2.4.8), (2.4.15), and (2.4.17), we have:

$$\hat{x}(n+1/n) = A(n+1,n)\hat{x}(n/n-1) + K(n)[z(n) - C(n)\hat{x}(n/n-1)] \tag{2.4.19}$$

The form of the solution in Eq(2.4.19) is computationally very useful. The result of Eq(2.4.19) is a sequential algorithm for determining $\hat{x}(n+1/n)$ based on $\hat{x}(n/n-1)$ and the new observation $z(n)$. The new estimate is formed by predicting forward from the old estimate and then correcting it by the innovation term which contains the new information. On the other hand, the Kalman gain matrix K(n) can also be expressed in terms of the one-step predictor error covariance matrix P(n) defined by Eq(2.4.12).

Using Eqs(2.4.9) and (2.4.16), we can deduce that:

$$E\{x(n+1)\alpha^T(n)\} = A(n+1,n)E\{x(n)[C(n)\hat{x}(n/n-1)+w(n)]^T\}$$
$$= A(n+1,n)E\{x(n)\hat{x}^T(n/n-1)\}C^T(n) \qquad (2.4.20)$$

Since $x(n)$ and $w(n)$ are uncorrelated, we can write

$$x(n) = \hat{x}(n/n-1)+\tilde{x}(n/n-1)$$

Since $\tilde{x}(n/n-1)$ is orthogonal to $\alpha(k)$ *for* $k \le n-1$, it is also orthogonal to $\hat{x}(n/n-1)$ (since $\hat{x}(n/n-1)$ is a linear combination of $\alpha(k)$ *for* $k \le n-1$). Equation (2.4.20) can then be written as:

$$E\{x(n+1)\alpha^T(n)\} = A(n+1,n)E\{\hat{x}(n/n-1)\hat{x}^T(n/n-1)\}C^T(n)$$
$$= A(n+1,n)P(n)C^T(n) \qquad (2.4.21)$$

Therefore, using Eq(2.4.13), Eq(2.4.18) and Eq(2.4.21), we can see that:

$$K(n) = A(n+1,n)P(n)C^T(n)[C(n)P(n)C(n)^T+R(n)]^{-1} \qquad (2.4.22)$$

This formula for K(n) is not very useful because P(n) must be known. To overcome this problem we will derive a recursive

formula for computing P(n). The error in the one-step prediction of x(n+1) can be written as

$$\tilde{x}(n+1/n) = x(n+1) - \hat{x}(n+1/n)$$
$$= A(n+1,n)x(n) + g(n)v(n) - A(n+1,n)\hat{x}(n,n-1)$$
$$-K(n)[z(n) - C(n)\hat{x}(n/n-1)] \qquad (2.4.23)$$

Substituting Eq(2.2.3) for z(n) and rearranging terms yields to:

$$\tilde{x}(n+1/n) = [A(n+1,n) - K(n)C(n)]\tilde{x}(n/n-1) + g(n)v(n) - K(n)w(n) \qquad (2.4.24)$$

since $\tilde{x}(n/n-1)$, v(n) and w(n) are mutually uncorrelated, we find that

$$P(n+1) = cov\{\tilde{x}(n+1/n)\}$$
$$= [A(n+1,n) - K(n)C(n)]P(n)[A(n+1,n) - K(n)C(n)]^T$$
$$+ g(n)Q(n)g^T(n) + K(n)R(n)K^T(n) \qquad (2.4.25)$$

We will now derive a set of equations for recursively calculating the filtered estimate $\hat{x}(n/n)$. According to Eq(2.3.1) we have

$$\hat{x}(n+1/n) = \xi\{x(n+1)/Z(n)\}$$
$$= \xi\{A(n+1,n)x(n) + g(n)v(n)/Z(n)\}$$
$$= A(n+1,n)\hat{x}(n/n) \qquad (2.4.26)$$

and the filtered estimate can be computed as

$$\hat{x}(n/n) = A^{-1}(n+1,n)\hat{x}(n+1/n)$$
$$= A(n,n+1)\hat{x}(n+1/n) \qquad (2.4.27)$$

This expression is helpful in the derivation of the Kalman smoother where we assume $A(n+1,n)$ is nonsingular. Though the final form of the Kalman smoother does not require this

assumption.

According to Eq(2.4.19)

$$\hat{x}(n/n) = \hat{x}(n/n-1) + H(n)[z(n) - C(n)\hat{x}(n/n-1)] \qquad (2.4.28)$$

where $H(n) = A(n, n+1)K(n)$

$$= P(n)C^T(n)[C(n)P(n)C^T(n) + R(n)]^{-1} \qquad (2.4.29)$$

It follows from Eq(2.4.26) that

$$\hat{x}(n/n-1) = A(n, n-1)\hat{x}(n-1/n-1) \qquad (2.4.30)$$

Therefore, Eq(2.4.28) can be written as

$$\hat{x}(n/n) = A(n, n-1)\hat{x}(n-1, n-1) + H(n)[z(n) - C(n)A(n, n-1)\hat{x}(n-1, n-1)] \qquad (2.4.31)$$

which is the desired recursive formula for $\hat{x}(n/n)$.

A block diagram of the estimator, specified by Eq(2.4.31), including the signal model is shown in Fig:2.4.

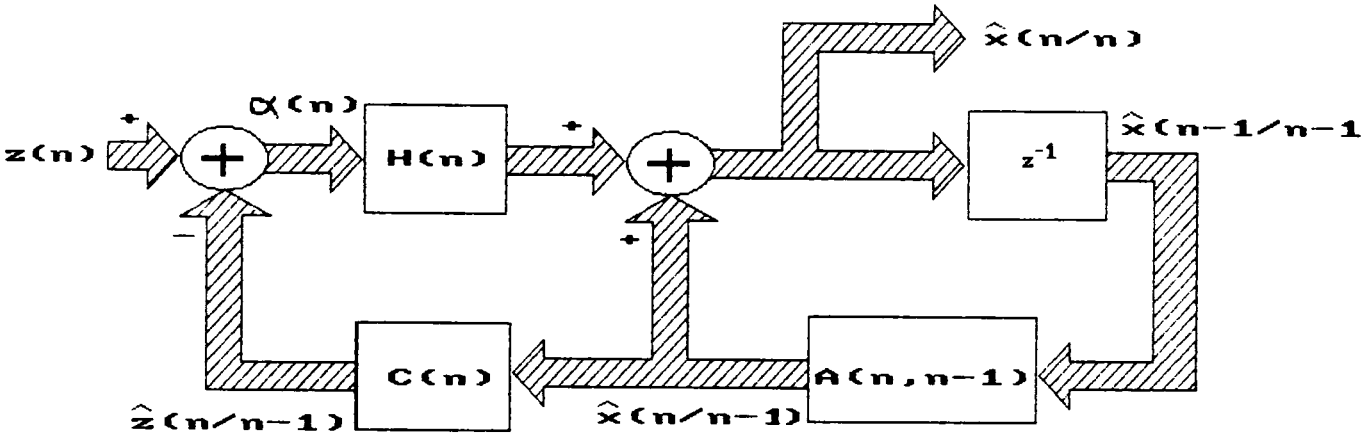


Fig:2.4 Optimum filter for direct computation of $\hat{x}(n/n)$.

Using Eq(2.4.28) for $\hat{x}(n/n)$, the estimation error can be written as

$$\tilde{x}(n/n) = x(n) - \hat{x}(n/n)$$
$$= \tilde{x}(n/n-1) - H(n)\alpha(n) \qquad (2.4.32)$$

Therefore, the filtering error covariance matrix is

$$\Gamma(n) = E\{\tilde{x}(n/n)\tilde{x}^T(n/n)\}$$
$$= P(n) - E\{\tilde{x}(n/n-1)\alpha^T(n)\}H^T(n)$$
$$-H(n)E\{\alpha(n)\tilde{x}^T(n/n-1)\} + H(n)V_e(n)H^T(n) \qquad (2.4.33)$$

Using Eq(2.4.16) and Eq(2.4.21) and the fact that $\tilde{x}(n/n-1)$ is orthogonal to $\alpha(n)$, it follows that

$$E\{\tilde{x}(n/n-1)\alpha^T(n)\} = E\{x(n)\alpha^T(n)\}$$
$$= A(n,n+1)E\{x(n+1)\alpha^T(n)\}$$
$$= P(n)C^T(n) \qquad (2.4.34)$$

Using Eq(2.4.29), Eq(2.4.13) and Eq(2.4.34), we find that the filtering error covariance matrix can be written as:

$$\Gamma(n) = P(n) - P(n)C^T(n)[C(n)P(n)C^T(n) + R(n)]^{-1}C(n)P(n) \qquad (2.4.35)$$

Using the expression of K(n) given by Eq(2.4.22), and expanding the first term on the right hand side of Eq(2.4.25), $P(n)$ can be written in the following consice form:

$$P(n+1) = A(n+1,n)\Gamma(n)A^T(n+1,n) + g(n)Q(n)g^T(n) \qquad (2.4.36)$$

As a consequence, the up to date estimate $\hat{x}(n/n)$ given by Eq(2.4.31) can be computed using the following algorithm which consist of 5 steps:

STEP_1: compute $H(n)$ for $n \geq 1$ using

$$H(n) = P(n)C^T(n)[C(n)P(n)C^T(n) + R(n)]^{-1}$$

with the initial condition $P(1) = cov\{x(1), x(1)\}$

STEP_2: compute the up to date estimate $\hat{x}(n/n)$ using

$$\hat{x}(n/n) = A(n, n-1)\hat{x}(n-1/n-1) + H(n)[z(n) - C(n)A(n, n-1)\hat{x}(n-1/n-1)]$$

with the initial condition $\hat{x}(0/0) = E\{x(0)\} = 0$

STEP_3: compute $\Gamma(n)$ using

$$\Gamma(n) = P(n) - P(n)C^T(n)[C(n)P(n)C^T(n) + R(n)]^{-1}C(n)P(n)$$

STEP_4: compute $P(n+1)$ using

$$P(n+1) = A(n+1, n)\Gamma(n)A^T(n+1, n) + g(n)Q(n)g^T(n)$$

STEP_5: increment n to n+1 and repeat the sequence.

2.5 KALMAN SMOOTHER

We will now investigate the smoothing problem, ie: the problem of estimating x(i) from the available data vectors $\{z(1), \ldots, z(n)\}$ for $1 \le i < n$. The general estimation equation (2.3.9) can be written in the form:

$$\begin{aligned}\hat{x}(i/n) &= \sum_{k=1}^{i} E\{x(i)a^T(k)\}V_e^{-1}(k)a(k) \\ &\quad + \sum_{k=i+1}^{n} E\{x(i)a^T(k)\}V_e^{-1}(k)a(k) \\ &= \hat{x}(i/i) + \sum_{k=i+1}^{n} E\{x(i)a^T(k)\}V_e^{-1}(k)a(k) \qquad (2.5.1)\end{aligned}$$

Thus the smoothed estimate can be computed by modifying the filtered estimate $\hat{x}(i/i)$.

We will derive an explicit computational formula by converting Eq(2.5.1) into an alternative form from which we will derive a

recursive formula for computing the smoothed estimates backward from time n. As a first step, the innovation sample $\alpha(k)$ can be expressed as

$$\alpha(k) = C(k)A(k,k-1)\hat{x}(k-1/k-1) + C(k)g(k-1)v(k-1) + w(k)$$

where $$\tilde{x}(k-1/k-1) = x(k-1) - \hat{x}(k-1/k-1)$$

Then for $i<k$

$$E\{x(i)\alpha^T(k)\} = E\{\hat{x}(i/i)\alpha^T(k)\}$$

$$= E\{\hat{x}(i/i)\tilde{x}^T(k-1/k-1)\}A^T(k,k-1)C^T(k) \qquad (2.5.2)$$

If we substitute this expression in Eq(2.5.1), we get:

$$\hat{x}(i/n) = \hat{x}(i/i) + \sum_{k=i+1}^{n} E\{\hat{x}(i/i)\tilde{x}^T(k-1/k-1)\}A^T(k,k-1)C^T(k)V_e^{-1}(k)\alpha(k) \qquad (2.5.3)$$

It follows from Eq(2.4.28) and Eq(2.4.29) that

$$C^T(k)V_e^{-1}(k)\alpha(k) = P^{-1}(k)[\hat{x}(k/k) - \hat{x}(k/k-1)] \qquad (2.5.4)$$

Therefore, Eq(2.5.3) can be written as

$$\hat{x}(i/n) = \hat{x}(i/i) + \sum_{k=i+1}^{n} E\{\hat{x}(i/i)\tilde{x}^T(k-1/k-1)\}A^T(k,k-1)P^{-1}(k)*$$

$$[\hat{x}(k/k) - \hat{x}(k/k-1)] \qquad (2.5.5)$$

When i=n-1 Eq(2.5.5) reduces to

$$\hat{x}(n-1/n) = \hat{x}(n-1/n-1) + \Gamma(n-1)A^T(n,n-1)P^{-1}(n)[\hat{x}(n/n) - \hat{x}(n/n-1)] \qquad (2.5.6)$$

where

$$\Gamma(n-1) = E\{\tilde{x}(n-1/n-1)\tilde{x}^T(n-1/n-1)\}$$

Similarly when i=n-2, Eq(2.5.5) becomes

$$\hat{x}(n-2/n)=\hat{x}(n-2/n-2)+\Gamma(n-2)A^{T}(n-1,n-2)P^{-1}(n-1)$$
$$*[\hat{x}(n-1/n-1)-\hat{x}(n-1/n-2)]+E\{\hat{x}(n-2/n-2)\hat{x}^{T}(n-1/n-1)\}A^{T}(n,n-1)P^{-1}(n)$$
$$*[\hat{x}(n/n)-\hat{x}(n/n-1)] \qquad (2.5.7)$$

Using the facts that:

$$\hat{x}(n-1/n-1)=A(n-1,n-2)\hat{x}(n-2/n-2)+H(n-1)\alpha(n-1)$$

and

$$x(n-1)=A(n-1,n-2)x(n-2)+g(n-2)v(n-2)$$

we deduce that

$$\tilde{x}(n-1/n-1)=x(n-1)-\hat{x}(n-1/n-1)$$
$$=A(n-1,n-2)\tilde{x}(n-2/n-2)+g(n-2)v(n-2)-H(n-1)\alpha(n-1) \qquad (2.5.8)$$

Therefore,

$$E\{\tilde{x}(n-2/n-2)\hat{x}^{T}(n-1/n-1)\}=\Gamma(n-2)A^{T}(n-1/n-2)$$
$$-E\{\tilde{x}(n-2/n-2)\alpha^{T}(n-1)\}H^{T}(n-1) \qquad (2.5.9)$$

The expression on the right-hand side of Eq(2.5.9) can be evaluated by letting i=n-2 and k=n-1 in Eq(2.5.2).

Using the obtained expression, we derive that:

$$E\{\tilde{x}(n-2/n-2)\hat{x}^{T}(n-1/n-1)\}=\Gamma(n-2)A^{T}(n-1/n-2)[I-C^{T}(n-1)H^{T}(n-1)]$$
$$=\Gamma(n-2)A^{T}(n-1/n-2)P^{-1}(n-1)$$
$$*[P(n-1)-P(n-1)C^{T}(n-1)H^{T}(n-1)] \qquad (2.5.10)$$

from Eq(2.4.29) and Eq(2.4.35), it follows that the expression which is within the brackets on the right-hand side of Eq(2.5.10) is $\Gamma(n-1)$ which has been defined in Eq(2.4.35) for $n-1$.

Thus

$$\hat{x}(n-2/n)=\hat{x}(n-2/n-2)+\Gamma(n-2)A^T(n-1,n-2)P^{-1}(n-1)$$
$$*[\hat{x}(n-1/n-1)-\hat{x}(n-1/n-2)]+E\{\tilde{x}(n-2/n-2)\tilde{x}^T(n-1/n-1)\}A^T(n,n-1)P^{-1}(n)$$
$$*[\hat{x}(n/n)-\hat{x}(n/n-1)] \qquad (2.5.7)$$

Using the facts that:

$$\hat{x}(n-1/n-1)=A(n-1,n-2)\hat{x}(n-2/n-2)+H(n-1)\alpha(n-1)$$

and

$$x(n-1)=A(n-1,n-2)x(n-2)+g(n-2)v(n-2)$$

we deduce that

$$\tilde{x}(n-1/n-1)=x(n-1)-\hat{x}(n-1/n-1)$$
$$=A(n-1,n-2)\tilde{x}(n-2/n-2)+g(n-2)v(n-2)-H(n-1)\alpha(n-1) \qquad (2.5.8)$$

Therefore,

$$E\{\tilde{x}(n-2/n-2)\tilde{x}^T(n-1/n-1)\}=\Gamma(n-2)A^T(n-1/n-2)$$
$$-E\{\tilde{x}(n-2/n-2)\alpha^T(n-1)\}H^T(n-1) \qquad (2.5.9)$$

The expression on the right-hand side of Eq(2.5.9) can be evaluated by letting i=n-2 and k=n-1 in Eq(2.5.2).

Using the obtained expression, we derive that:

$$E\{\tilde{x}(n-2/n-2)\tilde{x}^T(n-1/n-1)\}=\Gamma(n-2)A^T(n-1/n-2)[I-C^T(n-1)H^T(n-1)]$$
$$=\Gamma(n-2)A^T(n-1/n-2)P^{-1}(n-1)$$
$$*[P(n-1)-P(n-1)C^T(n-1)H^T(n-1)] \qquad (2.5.10)$$

from Eq(2.4.29) and Eq(2.4.35), it follows that the expression which is within the brackets on the right-hand side of Eq(2.5.10) is $\Gamma(n-1)$ which has been defined in Eq(2.4.35) for $n-1$.

Thus

$$E\{\hat{x}(n-2/n-2)\hat{x}^T(n-1/n-1)\} = \Gamma(n-2)A^T(n-1,n-2)P^{-1}(n-1)\Gamma(n-1) \tag{2.5.11}$$

The substitution of Eq(2.5.11) into Eq(2.5.7) gives

$$\begin{aligned}\hat{x}(n-2/n) &= \hat{x}(n-2/n-2)+\Gamma(n-2)A^T(n-1,n-2)P^{-1}(n-1)\\ &\quad *\{\hat{x}(n-1/n-1)+\Gamma(n-1)A^T(n,n-1)P^{-1}(n)\\ &\quad *[\hat{x}(n/n)-\hat{x}(n/n-1)]-\hat{x}(n-1/n-2)\}\end{aligned} \tag{2.5.12}$$

According to Eq(2.5.6), the sum of the first two terms within the braces on the right-hand side of Eq(2.5.12) is equal to $\hat{x}(n-1/n)$.

Therefore, we deduce that:

$$\begin{aligned}\hat{x}(n-2/n) &= \hat{x}(n-2/n-2)+\Gamma(n-2)A^T(n-1,n-2)P^{-1}(n-1)\\ &\quad *[\hat{x}(n-1/n)-\hat{x}(n-1/n-2)]\end{aligned} \tag{2.5.13}$$

Continuing in the same manner, we can see that:

$$\begin{aligned}&\hat{x}(i/n) = \hat{x}(i/i)+\Phi(i)[\hat{x}(i+1/n)-\hat{x}(i+1/i)] \qquad \text{for } i=1,\ldots,n-1\\ &\quad = \Phi(i)\hat{x}(i+1/n)+G(i)*\hat{x}(i/i) \qquad \text{for } i=1,\ldots,n-1\end{aligned} \tag{2.5.14}$$

where

$$\Phi(i) = \Gamma(i)A^T(i+1,i)P^{-1}(i+1) \tag{2.5.15}$$

and

$$G(i) = I-\Phi(i)A(i+1,i) \tag{2.5.16}$$

Equations (2.5.14), (2.5.15), and (2.5.16) provide a mean for computing recursively the smoothed estimates backward from time n as desired once the solution $\hat{x}(i/i)$ *for* $i=1,\ldots,n$ has been computed and stored.

The error of the smoothed estimate is given by:

$$\tilde{x}(i/n) = x(i) - \hat{x}(i/n)$$

$$= \tilde{x}(i/i) - \Phi(i)[\hat{x}(i+1/n) - \hat{x}(i+1/i)]$$

which can be rearranged into the form

$$\tilde{x}(i/n) + \Phi(i)\hat{x}(i+1/n) = \tilde{x}(i/i) + \Phi(i)\hat{x}(i+1/i) \qquad (2.5.17)$$

We will denote the smoothing error covariance matrix by:

$$\Delta(i/n) = cov\{\tilde{x}(i/n), \tilde{x}(i/n)\} \qquad (2.5.18)$$

From the orthogonality condition of the projection theorem, it follows that

$$cov\{\tilde{x}(i/n), \hat{x}(i+1/n\} = cov\{\tilde{x}(i/i), \hat{x}(i+1/i)\} = 0, \qquad (2.5.19)$$

$$cov\{\hat{x}(i+1/n), \hat{x}(i+1/n)\} = cov\{x(i+1), x(i+1)\} - \Delta(i+1/n), \qquad (2.5.20)$$

and $$cov\{\hat{x}(i+1/i), \hat{x}(i+1/i)\} = cov\{x(i+1), x(i+1)\} - P(i+1). \qquad (2.5.21)$$

Taking the covariance of both sides of Eq(2.5.17) and using Eqs (2.5.19 to 2.5.21), it follows that

$$\Delta(i/n) = \Gamma(i) + \Phi(i)[\Delta(i+1/n) - P(i+1)]\Phi^T(i) \qquad (2.5.22)$$

This equation can be solved backward in time starting with i=n-1 and the terminal condition $\Delta(n/n) = \Gamma(n)$ since $\Gamma(n)$ is a best estimate from the recorded data.

PARTICULAR CASE:

When a stationary zero mean signal process is observed over a fixed interval $[n=1, n=N]$ in a white noise environment, the signal in state-space form is

$$x(n+1) = Ax(n) + gv(n)$$

$$s(n) = Cx(n)$$

$$z(n) = s(n) + w(n) \qquad (2.5.23)$$

$$E\left\{\begin{bmatrix} v(k) \\ w(k) \end{bmatrix}[v(n)^T \quad w(n)^T]\right\} = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}\delta(n-k) \qquad (2.5.24)$$

where $x(n)$ is the state vector, $s(n)$ is the signal, $z(n)$ is the measured data and $v(n)$, $w(n)$ are respectively the input and output noises. The system parameters $(A^{p\times p}, g^{p\times 1}, C^{1\times p})$ are constants due to the assumption of stationarity.

Using the Eqs(2.4.22),(2.4.31), (2.4.35), (2.5.14), (2.5.15) and (2.5.16), the Kalman smoother takes the form

$$\hat{x}(n/n) = F(n)\hat{x}(n-1/n-1) + K(n)z(n) \tag{2.5.25}$$

$$\hat{x}(n/N) = \Phi(n)\hat{x}(n+1/N) + G(n)\hat{x}(n/n) \tag{2.5.26}$$

$$\hat{s}(n/N) = C\hat{x}(n/N) \tag{2.5.27}$$

where

$$K(n) = P(n)C^T[CP(n)C^T + R]^{-1} \tag{2.5.28}$$

$$\Phi(n) = \Gamma(n)A^T P^{-1}(n+1) \tag{2.5.29}$$

$$G(n) = I - \Phi(n)A \tag{2.5.30}$$

$$F(n) = [I - K(n)C]A \tag{2.5.31}$$

$$\Gamma(n) = P(n) - P(n)C^T[CP(n)C^T + R]^{-1}CP(n) \tag{2.5.32}$$

All parameters $F(n)$, $K(n)$, $\Phi(n)$, $G(n)$, $\Gamma(n)$ are time-varying associated with the state error covariance matrix $P(n)$ which is solved recursively from the difference Riccati equation:

$$P(n+1) = A\Gamma(n)A^T + gQg^T \tag{2.5.33}$$

with the initial conditions

$$\hat{x}(0/0) = E\{x(0)\} = 0$$

and

$$P(1) = cov\{x(1), x(1)\}$$

we can easily see that although Kalman smoothing estimates are optimum, they are computationaly expensive because at each step of the estimation algorithm the parameters

$F(n)$, $K(n)$, $\Phi(n)$, $G(n)$, $\Gamma(n)$, *and* $P(n)$ must be computed. A modified Kalman smoother which consists of a steady-state Kalman smoothing algorithm and a correction for transient effects using an initial state vector will be discused, in the next section, to reduce the computational burden.

2.6 THE MODIFID KALMAN SMOOTHER

The modified Kalman smoother provides an efficient and computationally cheap algorithm. The approach of this section follows the work of J. Vaccaro. and Fu Lu [32].

Equations (2.5.25) and (2.5.26) can be written as

$$\hat{x}(n/n) = \left[\prod_{j=1}^{n} F(i)\right]\hat{x}(0/0) + \sum_{j=1}^{n}\left[\prod_{i=1}^{n-j} F(i)\right]K(j)z(j) \qquad (2.6.1)$$

$$\hat{x}(n/N) = \left[\prod_{j=n}^{N} \Phi(j)\right]\hat{x}(N/N) + \sum_{j=n}^{N-1}\left[\prod_{i=j}^{N-1} \Phi(i)\right]G(j)\hat{x}(j/j) \qquad (2.6.2)$$

where the initial state vector $\hat{x}(0/0)$ is assumed to be zero due to missing data $z(n)$, *for* $n \le 0$.

It is well known that if we use the infinite-past data, the Kalman filter is then time-invariant, and Eq(2.6.1) becomes

$$\hat{x}(n/n) = F^{n}\hat{x}(0/0) + \sum_{j=1}^{n} F^{n-j}Kz(j) \quad for \quad 1 \le n \le N \qquad (2.6.3)$$

where $\hat{x}(0/0)$ is no longer 0.

$\hat{x}(0/0)$ is a function of past data $z(n)$ *for* $n \le 0$. F, K are now constants and, therefore, they are defined only once from P which is the solution of the steady-state algebraic Riccati equation [1,33].

$$P = A[P - PC^T(CPC^T + R)^{-1}CP]A^T + gQg^T \qquad (2.6.4)$$

Thus the time-invariant estimation is much more efficient than time-varying estimation of the previous section.

However, the problem is that the infinite-past data from which $\hat{x}(0/0)$ is required to be calculated is missing. Common practice is to use $\hat{x}(0/0) = 0$ which is a poor estimate because a good time-invariant estimation procedure needs a nonzero initial value.

Since the initial state vector $x(0)$ summarizes the past history of the signal for $n \leq 0$, the infinite past data is not needed if the value of $x(0)$ is known. Furthermore, since $x(0)$ contributes to the signal in a known way, an estimate $\hat{x}(0)$ can be calculated from the given data which minimizes the sum of the squared error between the estimated signal and the data.

Using Eqs (2.5.23) and (2.6.3), and if we denote by $e(n)$ *for* $n = 1,\ldots,N$ the error of the estimated signal estimated from the forward filtering with the initial state $\hat{x}(0/0) = 0$, we have

$$\begin{aligned} z(n) &= Cx(n) + w(n) && 1 \leq n \leq N \\ &= C\hat{x}(n/n) + e(n) + w(n) \\ &= CF^n\hat{x}(0/0) + C\sum_{j=1}^{n} F^{n-j}Kz(i) + e(n) + w(n) && (2.6.5) \end{aligned}$$

We define the residual data $r(n)$ by:

$$r(n) = z(n) - C\sum_{j=1}^{n} F^{n-j}Kz(j) \qquad (2.6.6)$$

If we denote by $u(n)$ the total error which is the sum of the observation and the estimation errors:

$$u(n) = e(n) + w(n) \tag{2.6.7}$$

We deduce from Eqs (2.6.5 to 2.6.7) that:

$$r(n) = CF^n \hat{x}(0/0) + u(n) \tag{2.6.8}$$

Assume that Eq(2.6.8) has been computed for sufficient time to form N consecutive residual data vector. The obtained residuals allows Eq(2.6.8) to be expressed in the following vector form:

$$\begin{bmatrix} r(1) \\ r(2) \\ . \\ . \\ . \\ . \\ r(N) \end{bmatrix} = \begin{bmatrix} CF^1 \\ CF^2 \\ . \\ . \\ . \\ . \\ CF^N \end{bmatrix} \hat{x}(0/0) + \begin{bmatrix} u(1) \\ u(2) \\ . \\ . \\ . \\ . \\ u(N) \end{bmatrix} \tag{2.6.9}$$

In order to estimate the initial state vector $\hat{x}(0/0)$ uniquely the number of data records N must not be less than the signal model order p.

Rewriting Eq(2.6.9) in a stacked notation we obtain:

$$\mathbf{r} = \theta_f \hat{x}(0/0) + \mathbf{u} \tag{2.6.10}$$

in which:

$$\theta_f^T = (CF^1, \ldots, CF^N)$$

$$\mathbf{r}^T = (r(1), \ldots, r(N))$$

$$\mathbf{u}^T = (u(1), \ldots, u(N))$$

rearranging Eq(2.6.10) in terms of the total error $\mathbf{u}$:

$$\mathbf{u} = \mathbf{r} - \theta_f \hat{x}(0/0) \tag{2.6.11}$$

and select an estimate $\hat{x}(0/0)$ which minimizes J, the sum of squares of the total error.

$$J = \sum_{n=1}^{N} u^2(n) = \mathbf{u}^T\mathbf{u} \tag{2.6.12}$$

To find the least square estimate, we rewrite Eq(2.6.12) in terms of the residual data vector and the initial state vector:

$$\begin{aligned} J &= (\mathbf{r} - \theta_f \hat{x}(0/0))^T(\mathbf{r} - \theta_f \hat{x}(0/0)) \\ &= \mathbf{r}^T\mathbf{r} - \hat{x}^T(0/0)\theta_f^T\mathbf{r} - \mathbf{r}^T\theta_f\hat{x}(0/0) + \hat{x}^T(0/0)\theta_f^T\theta_f\hat{x}(0/0) \end{aligned} \tag{2.6.13}$$

setting to zero the derivative of J with respect to $\hat{x}(0/0)$:

$$\frac{\partial J}{\partial \hat{x}(0/0)} = -2\theta_f^T\mathbf{r} + 2\theta_f^T\theta_f\hat{x}(0/0) = 0 \tag{2.6.14}$$

yields:

$$\theta_f^T\theta_f\hat{x}(0/0) = \theta_f^T\mathbf{r} \tag{2.6.15}$$

The solution is unique if the second derivative matrix:

$$\frac{\partial^2 J}{\partial \hat{x}^2(0/0)} = 2\theta_f^T\theta_f \tag{2.6.16}$$

is positive definite.

Hence the least squares estimator under this assumption for the initial state vector $\hat{x}(0/0)$ is:

$$\hat{x}(0|0) = (\theta_f^T\theta_f)^{-1}\theta_f^T\mathbf{r} \tag{2.6.17}$$

Thus the forward filtering estimate given in Eq(2.6.1) is completely obtained.

The backward filtering parameters $\Phi(n)$, $G(n)$ are only a function of forward filtering parameters. Once the forward

filter is in steady-state, by approximating $\hat{x}(0/0)$ the backward filter is also time-invariant. This completes the derivation of the modified Kalman smoothing algorithm.

The algorithm is summarized as follows:

STEP_1: solve the following algebraic Reccati equation to get the state error covariance matrix P.

$$P = A(P - PC^T(CPC^T + R)^{-1}CP)A^T + gQg^T$$

Then compute the parameters Φ, K, F and G by:

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

$$F = (I - KC)A$$

$$\Phi = \Gamma A^T P^{-1}$$

$$G = I - FA$$

$$\Gamma = P - PC^T(CPC^T + R)^{-1}CP.$$

STEP_2: Estimate the state vector by forward time-invariant filtering with zero initial conditions and compute the residual data $r(n)$, i.e.

$$\bar{x}(n|n) = F\bar{x}(n-1|n-1) + Kz(n) \qquad for \quad n = 1, \ldots, N.$$

$$r(n) = z(n) - C\bar{x}(n|n)$$

STEP_3: Approximate the initial state estimate $\hat{x}(0|0)$ from $r(n)$ by least-squares estimation (LSE). The LSE solution is

$$\hat{x}(0|0) = (\theta_f^T\theta_f)^{-1}\theta_f^T\mathbf{r}$$

Where

$$\theta_f^T = (CF, \ldots, CF^N)$$

$$\mathbf{r}^T = (r(1), \ldots, r(N))$$

Correct the state estimates obtained by forward filtering by

$$\hat{x}(n|n) = \bar{x}(n|n) + F^{n}\hat{x}(0|0) \qquad for \quad n = 1,\ldots,N.$$

STEP_4: Calculate the smoothed state vector estimates using backward time-invariant filtering with final condition $\hat{x}(N|N)$ obtained from step_3:

$$\hat{x}(n|N) = \Phi\hat{x}(n+1|N) + G\hat{x}(n|n) \quad for \quad n = N-1,\ldots,1.$$

STEP_5: Compute the signal estimate by

$$\hat{s}(n) = C\hat{x}(n|N) \qquad n = 1,\ldots,N.$$

2.7 ILLUSTRATIVE EXAMPLE.

In this section we compare the performance of the modified Kalman smoother (MKS) to that of the time-varying Kalman filter (KF) in the case of estimating a stationary signal observed in a noisy envirement.

The signal is modelled as the output of first order linear time-invariant filter driven by a zero mean white noise, $v(n)$, of variance $\sigma^2 = 1$. The signal represents a first order Markov process. Its description in a state space representation is given by:

$$x(n+1) = 0.91x(n) + v(n)$$
$$s(n) = x(n) \qquad (2.7.1)$$

with

$x(1) = 0.3$

The measurement model is described by the signal with an additive zero mean white, $w(n)$, noise of variance $\sigma^2 = 1$:

$$z(n) = s(n) + w(n) \qquad (2.7.2)$$

Both the modified Kalman smoother and the Kalman filter have been applied to the measurement in order to estimate the signal $s(n)$.

The estimated signals obtained by the two above methods are ploted together with the actual and the measurement signals. See Fig:2.5

The mean squared estimation error is computed for both estimators. The result for the (MKS) and the (KF), in the case of long data records, were 0.446, and 0.547 respectively. In the case of short data records, where only 16 data samples are used, the mean squared estimation error is computed for both the (MKS) and (KF) and the result were 0.375 and 0.404, respectively. It is worthnoting that the clear superiority of the (MKS) over the (KF) decreases in the case of short data records. This may be due to the inaccuracy of the least square estimation in the case of short data records.

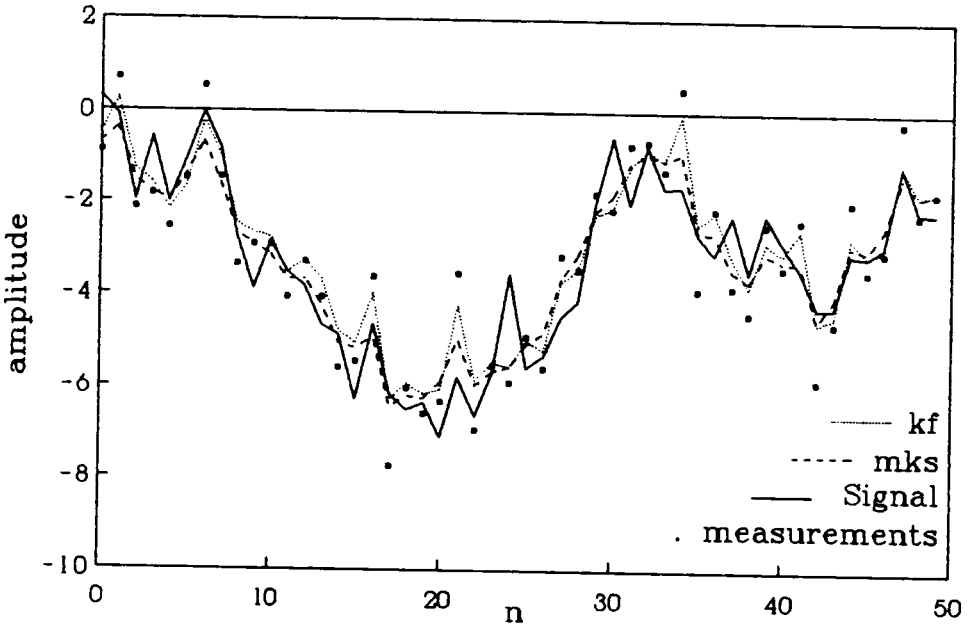


Fig.2.5 : signal estimation σ =1

CHAPTER III:

SMOOTHING SPLINE

3.1 INTRODUCTION

Consider the classical input-response model where for given inputs at $t_1,\ldots,t_N$ in some interval $[t_1, t_N]$, the corresponding output $z(t_n)$ satisfy:

$$z(t_n) = s(t_n) + w(t_n) \tag{3.1.1}$$

where $s(t_n)$ is a smooth signal observed in noisy environment described by a zero mean white noise $w(t_n)$ with variance σ^2. The random variables $s(t_n)$, and $w(t_n)$ are not observed. The only observed random variable sequence is $\{z(t_n)\}$ *for* $n = 1,\ldots,N$ which is recorded to obtain an estimate $\hat{s}(t_n)$ of the signal $s(t_n)$.

Many nonparametric approaches had been proposed to obtain an estimator $\hat{s}(t)$ of the signal $s(t)$, among which:

1) The polynomial fit: In this approach we use an approximation to the model described by Eq(3.1.1) of the form:

$$z(t_n) = p(t_n) + w(t_n) \tag{3.1.2}$$

where $p(t)$ is a polynomial of a certain predefined order m, that is:

$$p(t) = \sum_{i=0}^{m} a_i t^i \qquad (3.1.3)$$

This has the advantage that the unknown parameters, a_i's, in the approximate model enter in a linear fashion. Hence, they may be estimated by ordinary least square method. The polynomial regression approach models provide an entirely satisfactory description of a set of data. However, it fails to fit satisfactorily data which have a more rapid variation in one region than other regions since polynomials along with most other mathematical functions, have the property that their behaviour in a small region determines their behaviour everywhere. Another drawback is that it is not suitable for fitting a set of data that possesses several high peaks.

2) Cubic smoothing spline fit: This nonparametric approach quantifies the competition between two conflicting aims in smooth signal: one wants to produce a good fit to the data, and to avoid too much rapid local fluctuation. This phenomenon is encountered in the case of noisy data interpolation.

A measure of the rapid local variation of the signal can be given by the roughness penalty defined as the integrated square second derivative.

$$\int |s''(t)|^2 dt \qquad (3.1.4)$$

when the roughness penality is small, the signal fluctuates considerably.

Using this measure, the estimation of the signal $s(t)$ reduces to the minimisation of:

$$\int |\ddot{s}(t)|^2 dt + \lambda \left\{ \sum_{n=1}^{N} (z(t_n) - \hat{s}(t_n))^2 \right\} \qquad (3.1.5)$$

where λ is a smoothing parameter which represents the rate of exchange between residual error and local variation. The minimization of the expression Eq(3.1.5) over the class of all twice-differentiable functions $s(t)$ will yield an estimate $\hat{s}(t)$ which, for a given value of λ, gives the best compromise between smoothness and goodness of fit.

In the spline approach, the estimated signal is, assumed to be, composed of piecewise polynomials of degree three as given by the following expression:

$$\hat{s}(t) = a_n + b_n(t-t_n) + c_n(t-t_n)^2 + d_n(t-t_n)^3 \quad for \quad t_n \le t < t_{n+1} \qquad (3.1.6)$$

The consecutive piecewise polynomials join at the knot t_n and fulfils the continuity condition for the signal and its first two derivatives. The piecewise nature of the estimated signal $\hat{s}(t_n)$ makes it a reasonable approximation for models with changing structure since it provides more locally adaptive fits. As a result, the smoothing spline regression does not suffer from the two drawbacks of the polynomial regression approach. It can be regarded as an extention of polynomial regression.

Smoothing splines which is a nonparametric estimator have

enjoyed increased popularity as a tool in both theoretical and applied statistical research. They have been found useful for handling problems such as nonparametric regression [8,14], data smoothing [7,27], numerical differentiation [27], model validation [9,11], and nonparametric density function estimation [28]. They have also been used in many applications to image processing [23] including enhancement, interpolation, enlargement, and reduction [17]. In other words, they are an essential tool for data analysis [36].

In this chapter, a detailed development of smoothing spline is presented along with the generalized cross-validation for choosing the right value of the smoothing parameter λ which is crucial for smoothing spline.

3.2 SMOOTHING SPLINES

Suppose that the measurement model is given by

$$z(t_n) = s(t_n) + w(t_n) \tag{3.2.1}$$

where $s(t)$ is a signal observed over a fixed interval $[t_1, t_N]$ in a noisy environment described by an uncorrelated zero mean noise $w(t_n)$ with variance σ^2. In order to filter the signal, we consider the fidelity criterion

$$minimize \int_{t_1}^{t_N} |\hat{s}''(t)|^2 dt \tag{3.2.2}$$

which is defined among the class of all function for which $\hat{s}(t)$ and $\hat{s}'(t)$ are absolutely continuous and $\hat{s}''(t)$ is square-integrable, such that

$$\sum_{n=1}^{N}\left(\frac{z(t_n)-\hat{s}(t_n)}{\delta_n}\right)^2 \le S \qquad (3.2.3)$$

where the positive quantities δ_n controls locally the smoothing window at time t_n and S controls the overall extent of smoothing. If available one should use for δ_n the estimate of the standard deviation of the noise at time t_n. According to Reinsch [24], natural values of S lie within the confidence interval corresponding to the left hand side of (3.2.3) that is

$$N-2N^{1/2} \le S \le N+2N^{1/2} \qquad (3.2.4)$$

where N is the number of data records during the time interval $[t_1,t_N]$. The solution to the constrained problem Eqs:(3.2.2) and (3.2.3) was shown by Reinsch [24] and Woodford [37], to be a cubic spline and more generally it is a spline of degree **2q-1** for least square minimisation of the **qth** derivative [25]. The case where **q=2** is computationaly attractive since it leads to a very simple algorithm for the estimation of $s(t)$.

Applying the well known Lagrange multiplier method after normalisation of the inequality (3.2.3) constrained by introducing a dummy variable β. The object function

$$J = \int_{t_1}^{t_N} |\hat{s}''(t)|^2 dt + \lambda\left\{\sum_{i=1}^{N}\left(\frac{z(t_n)-\hat{s}(t_n)}{\delta_n}\right)^2 + \beta^2 - S\right\} \qquad (3.2.5)$$

must be minimized.

The optimal solution $\hat{s}(t)$ of the constrained problem given by

Eqs(3.2.2-3) was shown by Reinsch [24] to satisfy the following conditions.

$$\hat{s}(t_n)_- - \hat{s}(t_n)_+ = 0 \quad for \quad n = 2,\ldots,N-1 \qquad (3.2.6)$$

$$\hat{s}^{(1)}(t_n)_- - \hat{s}^{(1)}(t_n)_+ = 0 \quad for \quad n = 2,\ldots,N-1 \qquad (3.2.7)$$

$$\hat{s}^{(2)}(t_n)_- - \hat{s}^{(2)}(t_n)_+ = 0 \quad for \quad n = 1,\ldots,N \qquad (3.2.8)$$

$$\hat{s}^{(3)}(t_n)_- - \hat{s}^{(3)}(t_n)_+ = 2\lambda\left(\frac{\hat{s}(t_n) - z(t_n)}{\delta_n^2}\right) \quad for \quad n = 1,\ldots,N \qquad (3.2.9)$$

$$\hat{s}^{(k)}(t_n)_\pm = \lim_{\epsilon \to 0} \hat{s}^{(k)}\left(t_n \pm \epsilon\right)$$

Moreover,

$$\hat{s}^{(4)}(t) = 0 \quad for \quad t_n \le t < t_{n+1} \quad n = 1,\ldots,N-1 \qquad (3.2.10)$$

$s^{(k)}(t)$ stands for the k^{th} derivative of $s(t)$.

Eqs(3.2.6 to 3.2.10) indicate that the function $\hat{s}(t)$ is a piecewise polynomial of degree 3 in each interval $[t_n, t_{n+1}]$. See Fig.3.1.

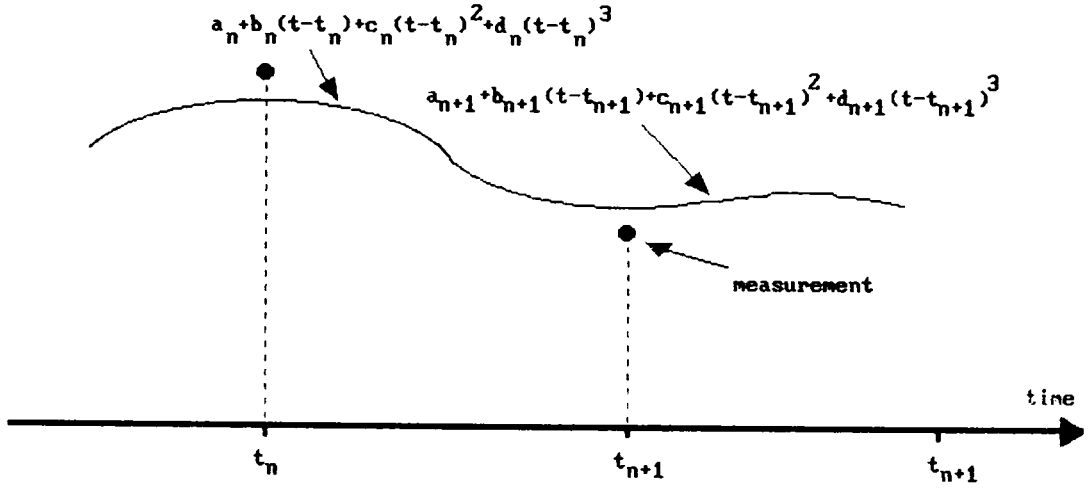


Fig:3.1 piecewise polynomials of the smoothing spline estimates.

The pieces are joined in the knots t_n and fulfil continuity condition for the function itself and the first two derivatives. Hence $s(t)$ is of the form

$$s(t) = a_n + b_n(t-t_n) + c_n(t-t_n)^2 + d_n(t-t_n)^3 \quad for \quad t_n \le t < t_{n+1} \quad (3.2.11)$$

The following two extra conditions are needed to determine completely the spline.

$$s^{(2)}(t_1) = s^{(3)}(t_1) = s^{(2)}(t_N) = s^{(3)}(t_N) = 0 \quad (3.2.12)$$

In this case, the function is called natural spline of degree 3.

The unknown coefficients of the spline are found by inserting Eq(3.2.11) into Eqs(3.2.6 to 3.2.9) and Eq(3.2.12) as follows: from Eq(3.2.8) and Eq(3.2.12), we have:

$$c_1 = c_N = d_1 = d_N = 0, \quad and \quad d_n = \frac{c_{n+1} - c_n}{3h_n} \qquad n = 2,\ldots,N-1 \quad (3.2.13)$$

from Eq(3.2.6)

$$b_n = \frac{a_{n+1} - a_n}{h_n} - c_n h_n - d_n h_n^2 \quad for \qquad n = 1,\ldots,N-1 \quad (3.2.14)$$

where $h_n = t_{n+1} - t_n$.

using Eq(3.2.7), the coefficient $c = (c_2,\ldots,c_{N-1})^T$ and $a = (a_1,\ldots,a_N)^T$ satisfy:

$$Tc = Q^T a \quad (3.2.15)$$

where **T** is positive definite, tridiagonal matrix of order **N-2**. **T** is given by:

$$T(i,i) = 2(h_{i-1} + h_i)/3, \quad and \quad T(i+1,i) = T(i,i+1) = h_i/3$$

and **Q** is a tridiagonal matrix with **N** rows and **N-2** columns. **Q** is given by:

$$Q(i-1,i) = 1/h_{i-1}, \quad Q(i,i) = -1/h_{i-1} - 1/h_i, \text{ and } Q(i+1,i) = 1/h_i$$

For a uniformly sampled signal with unit sampling interval, $t_n = n$, **Q** and **T** have the form

$$\mathbf{Q} = \begin{bmatrix} 1 & & & \\ -2 & \ddots & & \\ 1 & & \ddots & 1 \\ & \ddots & & -2 \\ & & & 1 \end{bmatrix}; \text{ and } \mathbf{T} = \begin{bmatrix} 4/3 & 1/3 & & & \\ 1/3 & 4/3 & 1/3 & & \\ & \ddots & \ddots & \ddots & \\ & & 1/3 & 4/3 & 1/3 \\ & & & 1/3 & 4/3 \end{bmatrix} \tag{3.2.16}$$

from Eq(3.2.9)

$$Qc = \lambda D^{-2}(z-a) \tag{3.2.17}$$

where

$$h_n = t_{n+1} - t_n$$

$$c = (c_2, \ldots, c_{N-1})^T$$

$$z = (z_1, \ldots, z_N)^T$$

$$a = (a_1, \ldots, a_N)^T$$

$$D = diag(\delta_1, \ldots, \delta_N)$$

Using Eq(3.2.15), and a left hand multiplication of Eq(3.2.17) by $Q^T D^2$ separates the variable **c** :

$$(Q^T D^2 Q + \lambda T)c = \lambda Q^T z \tag{3.2.18}$$

we can obtain an expression for α by using Eqs(3.2.17) and (3.2.18). The result is given by Eq(3.2.19).

$$a = z - D^2Q(Q^TD^2Q+\lambda T)^{-1}Q^Tz \qquad (3.2.19)$$

If we denote by s_λ the estimated signal evaluated at the design points t_n, using the smoothing parameter λ, then the optimal smoother with respect to the constraints (3.2.2) and (3.2.3) is obtained as:

$$s_\lambda = (I - D^2Q(Q^TD^2Q+\lambda T)^{-1}Q^T)z \qquad (3.2.20)$$

the matrix $(Q^TD^2Q+\lambda T)$ has an inverse provided that λ is positive which is a common assumption. The object function J given by expression (3.2.5) has to be minimized also with respect to β and λ, leading to the conditions

$$\frac{\partial J}{\partial \beta} = 0 \qquad ie: \qquad \lambda\beta = 0 \qquad (3.2.21)$$

and

$$\frac{\partial J}{\partial \lambda} = 0 \qquad ie: \qquad \sum_{n=1}^{N}\left(\frac{s(t_i)-z(t_i)}{\delta_i}\right)^2 = S-\beta^2 \qquad (3.2.22)$$

Using the expression of s_λ given in Eq(3.2.20), the left-hand side of Eq(3.2.22) can be written as the square of the function $F(\lambda)$ given by:

$$F(\lambda) = \|DQ(Q^TD^2Q+\lambda T)^{-1}Q^Tz\|_2 \qquad (3.2.23)$$

from Eq(3.2.21) we conclude that either $\lambda = 0$ or $\beta = 0$. A solution corresponding to $\lambda = 0$ would be a single polynomial of degree at most 1 and the integral smoothing criterion would be a trivial minimum.

If $F(0) > S$, then $\lambda \neq 0$ and $\beta = 0$. The inequality constraint

(3.2.3) becomes an equality constraint. In this case we have to find a value of λ such that

$$F(\lambda) = S^{1/2} \tag{3.2.24}$$

Equation (3.2.24) has at least one root since $F(\lambda)$ is a decreasing function. The uniqueness of the root was shown by Woodford [37]. This positive unique solution can be determined by using Newton's method.

The parameter λ and S are interchanged by Eq(3.2.24) and they both have the same effect on smoothing. According to expression (3.2.5) the parameter λ controls the tradeoff between the roughness of the smoothed signal as measured by $\int_{t_1}^{T_N} |\hat{s}''(t)|^2 dt$ and the fidelity to the data as measured by $\sum_{n=1}^{N} \left(\frac{\hat{s}(t_n) - x(t_n)}{\delta_n} \right)^2$.

As $\lambda \to \epsilon$, where ϵ is very small number $\hat{s}_\lambda$ becomes increasingly smooth, and the limiting function $\hat{s}_\epsilon$ is the least squares straight line. As a result, the shape of the signal $s(t)$ will be lost. On the other hand, as $\lambda \to \infty$, $\hat{s}_\lambda$ passes through the data . $\hat{s}_\infty$ is the natural cubic spline of interpolation to the measured data in this case $\hat{s}_\infty$ is too wiggly and picks up too much noise.

For many reasons, the variance of the measurement noise is not available in practice. This forces us to set $D = I_N$, where I is the identity matrix of order N. Hence, the linear smoother can be written in more concise form [23] as:

$$\hat{s}_\rho = A(\rho) z \tag{3.2.25}$$

where

$$A(\rho) = \left(I + \frac{1}{\rho}\Omega\right)^{-1} \tag{3.2.26}$$

and

$$\Omega = QT^{-1}Q^T \tag{3.2.27}$$

The lack of information on the common observation noise variance is incorporated in a new smoothing parameter ρ. To use the smoothing spline method in practice, it is convenient to have an automatic method for choosing the optimal parameter ρ. A strong candidate for choosing ρ is the cross-validation method [31,34,35] which is described in the next section.

3.3 THE GENERALIZED CROSS-VALIDATION METHOD

Ideally, we might wish the selected ρ to minimize the true mean square error (mse) given by

$$mse(\rho) = \frac{1}{N}\sum_{n=1}^{N}(s(t_n) - \hat{s}_\rho(t_n))^2 \tag{3.3.1}$$

In practice, the random variable $s(t_n)$ is not observed which makes Eq(3.3.1) impossible to apply.

An obvious alternative is to use its sample estimate

$$\hat{mse}(\rho) = \frac{1}{N}\sum_{n=1}^{N}(z(t_n) - \hat{s}_\rho(t_n))^2 \tag{3.3.2}$$

Unfortunately, Eq(3.3.2) is always minimized at $\rho = \infty$, since $\hat{s}_\infty$ interpolates the data. This fact led Wahba and Wold [35] to suggest to use the smoothing parameter value which minimize the cross-validation (CV) criterion

$$CV(\rho) = \frac{1}{N}\sum_{n=1}^{N}(\hat{s}_\rho^n(t_n) - z(t_n))^2 \tag{3.3.3}$$

where s_ρ^n is the smoothing spline fit to $\{z(t_1),\ldots,z(t_{n-1}),z(t_{n+1}),\ldots,z(t_N)\}$ for $n=1,\ldots,N$.

The use of Eq(3.3.3) is justified by the belief that a good value of ρ should be the one for which $s_\rho^n(t_n)$ is a good prediction of the omitted data value $z(t_n)$. In 1979, Wahba and Craven [34] has proposed an automatic method called the generalized cross-validation (GXV) for choosing the optimal smoothing parameter which has computational advantages over the cross-validation [35] and good statistical properties. They have suggested an estimate $\hat{\rho}$, called the generalized cross-validation estimate, for the minimiser of $mse(\rho)$. The estimate $\hat{\rho}$ is just the minimiser of $V(\rho)$ defined by

$$V(\rho)=\frac{1}{N}\frac{\sum_{n=1}^{N}(s_\rho(t_n)-z(t_n))^2}{\left(1-\frac{1}{N}trace(A(\rho))\right)^2} \tag{3.3.2}$$

where $A(\rho)$ is as defined by Eq(3.2.26) .

They have shown that the minimiser of $V(\rho)$ converges asymptotically to the minimiser of $mse(\rho)$. The main aim of Wahba and Craven's paper was to obtain a good estimate of the minimiser of mse_ρ from the data which does not require knowledge of the observation noise variance. This feature makes the smoothing spline attractive in many fields of application. We can see that (GXV) is meanly a weighted version of the $mse(\rho)$ which utilizes the weight $\left(1-\frac{1}{N}trace(A(\rho))\right)^{-2}$ to counterpart the tendency to choose ρ as infinity. On the other hand, the (GXV) criterion also prevents against the choice of small values for ρ since the increase in $\left(1-\frac{1}{N}trace(A(\rho))\right)^{-2}$ obtained by decreasing

ρ tends to be counteracted by an increase in the average size of the residuals. Hopefully, the selected value will then reflect the correct balance, between total fidelity to the data obtained when $\rho = \infty$, and the smoothest possible fit realised at $\rho = \epsilon$

3.4 UTRERAS [31] IMPROVEMENT

The computation of $V(\rho)$ for any particular value of ρ using Eq(3.3.2) requires the determination of s_ρ and the trace of $A(\rho)$ which is computationaly expensive. Utreras [31] proposed an algorithm that devides by the factor N the number of operations used to obtain $\hat{\rho}$ in the case of equally spaced data. The main idea of the method proposed by Utreras is the expression

$$trace(A(\rho)) = \sum(eigenvalues\ of\ A) \tag{3.4.1}$$

From Eq(3.2.27) we see that the matrix Ω does not depend on ρ. It is clear that we must approximate its eigenvalues to find the trace of $A(\rho)$ with negligible computational cost. If we denote by ω_i's the eigenvalues of Ω, Utreras has found an approximation to the eigenvalues of Ω given by

$$\omega_i \sim \frac{1}{(hN)^{2q}} \frac{\alpha_i}{N} \tag{3.4.2}$$

where α_i are the eigenvalues of the differential operator that occurs in classical mechanics when the vibrations of a rod with free ends are considered. The first ten eigenvalues α_i are tabulated by Utreras [31] page 23 and he noticed that for

$i > 10$ *the* α_i's satisfy the following relation

$$\alpha_i = \pi^4 (i - 1.5)^4 \tag{3.4.3}$$

.

As a result the trace of A will be computed as follows:

$$trace(A) = \sum_{i=1}^{N} \frac{1}{1 + \frac{1}{\rho}\omega_i} \tag{3.4.4}$$

Finally, the smoothing spline algorithm is summarized as follows:

STEP_1: form the matrices **Q** and **T**.

STEP_2: obtain the minimiser of the following expression.

$$V(\rho) = \frac{1}{N} \frac{\sum_{n=1}^{N} (\hat{s}_\rho(t_n) - z(t_n))^2}{\left(1 - \frac{1}{N} trace(A(\rho))\right)^2}$$

STEP_3: insert the value of ρ, found in STEP_2, in the following expression to obtain the smoothing spline estimates evaluated at the design points $\{t_n\}$.

$$\hat{s}_\rho(t_n) = (I - Q(Q^T Q + \rho T)^{-1} Q^T) z(t_n)$$

STEP_4: in order to completely determine the smoothing spline the parameters a_n , b_n , c_n , d_n are computed as follows

$$a = \hat{s}_\rho$$

$$c = T^{-1} Q^T a$$

$$c_1 = c_N = d_1 = d_N = 0 \qquad d_n = \frac{c_{n+1} - c_n}{3h_n} \qquad for \qquad n = 2, \ldots, N-1.$$

$$b_n = \frac{a_{n+1} - a_n}{h_n} - c_n h_n - d_n h_n^2 \qquad for \qquad n = 1, \ldots, N-1.$$

where

$$h_n = t_{n+1} - t_n$$

and b_n should be initialised.

3.5 ILLUSTRATIVE EXAMPLE

In this section, the smoothing spline estimator will be applied to 50 data samples simulated by adding normally distributed errors, with zero mean and standard deviation of .1, to a signal obtained by passing a zero mean white noise of variance $(.001)^2$ through a linear time-invariant filter. The signal and the measurement model are given, in state space representation, by Eq(3.5.1).

$$\begin{aligned} x(n+1) &= \begin{bmatrix} .91 & .3 \\ 0 & .74 \end{bmatrix} x(n) + \begin{bmatrix} 2.2 \\ 1.5 \end{bmatrix} w(n) \\ s(n) &= [1.8 \quad 0]\, x(n) \\ z(n) &= s(n) + v(n) \end{aligned} \tag{3.5.1}$$

with

$x(0) = [.5 \quad .5]^T$

Figure 3.2 displays the signal estimated by smoothing spline, the actual signal and the measurements. The optimal smoothing parameter ρ is found by minimizing expression (3.3.2) as shown in Fig.3.3.

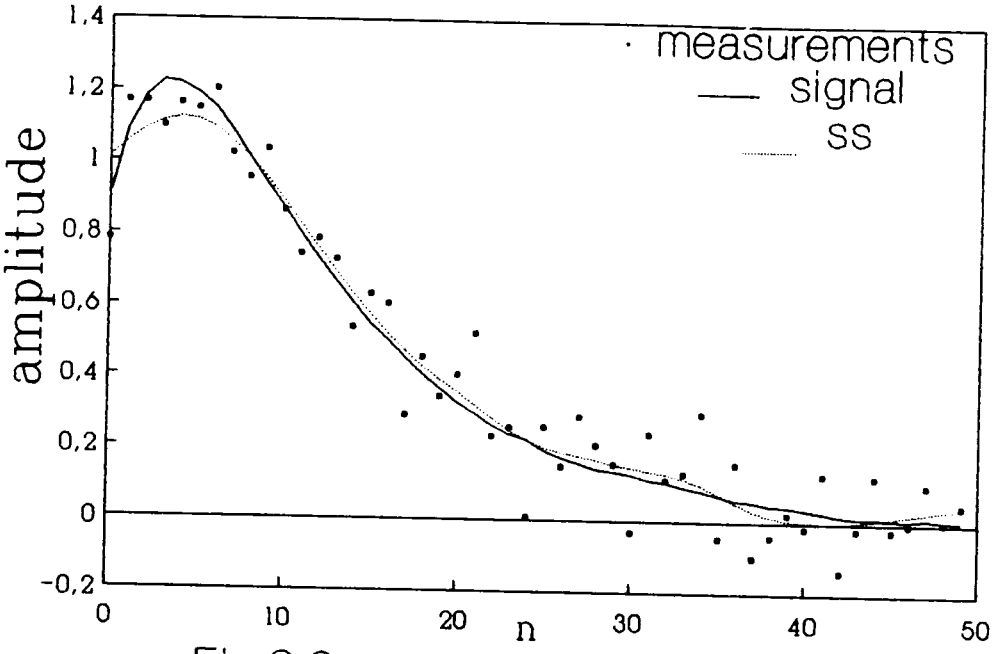


Fig:3.2 : signal estimation when σ =.1

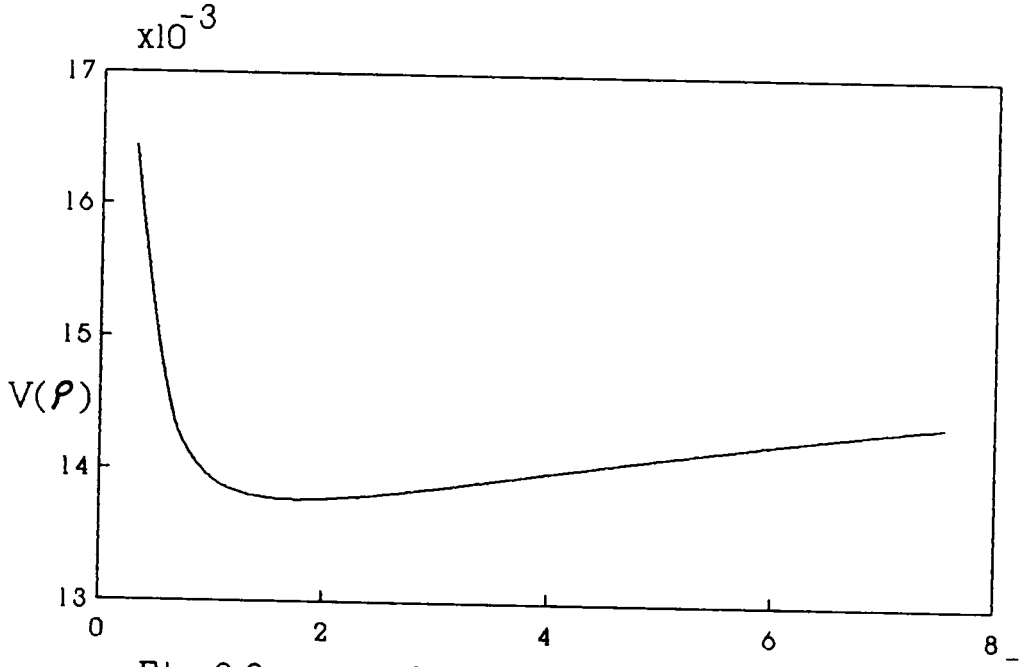


Fig.3.3: smoothing parameter by (GXV)
ρ_{opt} =1.8 E-2

Note that the signal estimated by the smoothing spline exhibits some oscillation in the last 30 data samples. On the otherhand, the estimated signal is not able to follow the actual peak. These two important remarks are the result of the global smoothing parameter found by (GXV) which gives the best compromise between the overall smoothness and goodness of fit. This drawback is remedied by the use of local smoothing parameter which is the basic idea on which a new estimator, called local smoothing spline, is based. A detailed description of the local smoothing spline will be given in chapter 5. For the sake of comparison, the signal estimated by both the smoothing spline and the modified Kalman smoother are plotted together with the measurement and the actual signal as depicted in Fig.3.4. The mean squared estimation error (MSE) has been computed and the results were 14.7992E-4 and 4.9032E-4, respectively. According to this single example, the modified Kalman smoother excels on the smoothing spline as far as the (MSE) is concerned. A detailed comparison of the two estimators will be given in the next chapter.

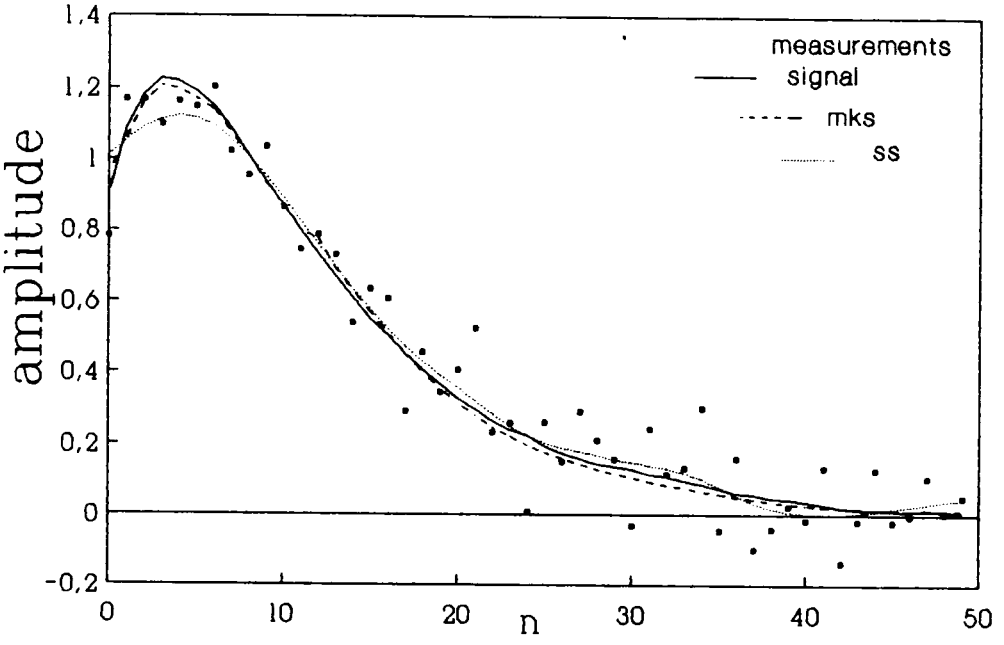


Fig.3.4: signal estimation when σ =.1

CHAPTER IV:

COMPARISON OF THE PERFORMANCES OF THE MODIFIED KALMAN SMOOTHER AND SMOOTHING SPLINE

4.1 INTRODUCTION

In this chapter, a number of simulation experiments have been conducted to compare the performances of the modified Kalman smoother estimator to that of the smoothing spline estimator. Long and short data records have been considered in the simulation study to see how good the performances of the estimators are in function of the data size. Furthermore, the robustness of the two estimators is investigated when certain required assumptions under which the two estimators are optimum, have been violated such as: the smoothness of the signal to be estimated in the case of the smoothing spline, and the knowledge of the signal parameters A, g, and C in the case of the modified Kalman smoother. Some diagnostics, such as innovations plots and the whiteness test are, also, provided to investigate the validity of the models used. Computational complexity is also given to compare the efficiency of the two

estimators. Finally, a new method based on the bootstrap technique is suggested to investigate the validity of the model used. The motivation was based on the fact that the whiteness test may lead to erronous conclusions.

4.2 SIMULATION STUDY:

In this section, we present simulation results based on ten different linear time-invariant signal models driven by a white noise of variance $(0.001)^2$. An initial state vector $x(0)=[.5\ \ .5]^T$ is used for the signals whose transfer-function is given by:

$$H(z)=C[zI-A]^{-1}g \qquad (4.2.1)$$

where H(z) is the input output model for the signal known as the transfer function, A, g, and C are the signal parameters when the state space representation for the signal is used. z is the forward time shift operator, and I is the pxp identity matrix.

The poles and the zeros of the transfer function for each of the ten simulated signals are given in table_4.2.1:

The ten simulated signals are chosen to cover a wide range of signals which satisfy both the smoothness property required by the (SS) estimator and the stationarity property required by the (MKS).

The measurement model consists of the signal corrupted with

white noise of variance σ^2. In other words, the measurement model is given by the following equation:

$$z(n) = s(n)+v(n) \qquad (4.2.2)$$

where s(n) represent the signal evaluated at time n, $v(n)$ is measurement noise added to the signal at time n and the only observed random variable is z(n) which is the measurement recorded at time n.

SIGNAL NUMBER	ZERO	POLE 1	POLE 2
SIGNAL 1	0.77	0.97	0.87
SIGNAL 2	0.67	0.99	0.87
SIGNAL 3	0.80	0.999	0.90
SIGNAL 4	0.98	0.999	0.90
SIGNAL 5	0.65	0.91	0.74
SIGNAL 6	0.83	0.99	0.90
SIGNAL 7	0.54	0.91	0.74
SIGNAL 8	0.81	0.85	0.92
SIGNAL 9	0.81	0.88	0.98
SIGNAL 10	0.30	0.42	0.87

Table_4.2.1: The poles and zeros locations for the ten different signals.

We simulated ten different realisations from each of the ten considered signals. For each of the ten realisations of the

measurement noise, we take three different values of the output noise variance ($\sigma^2=(0.05)^2$, $\sigma^2=(0.1)^2$, and $\sigma^2=(0.5)^2$). For each realisation, corresponding to a row in the tables:(4.2.2 to 4.2.7), the mean-squared estimation error (MSE), given by Eq(4.2.3), is computed for both the modified Kalman smoother (MKS) and the smoothing spline (SS) for the ten different signals numbered from **1** to **10**.

$$MSE=\frac{1}{N}\sum_{n=1}^{N}(s(n)-\hat{s}(n))^2 \qquad (4.2.3)$$

where N is the data size. For short data records, N is taken to be equal to 16 while for long data records N is taken to be equal to 50. $\hat{s}(n)$ is the estimated signal which is obtained either by the (MKS) or the (SS), and $s(n)$ is one of the ten simulated signals.

The last row of the tables:(4.2.2 to 4.2.7) gives the average, AMSE, of the mean squared estimation errors for the ten realisations. The plots of the (AMSE) as a function of the signals numbered from **1** to **10**, are shown in Figs:4.1, 4.2 and 4.3.

We notice that the (AMSE) decreases, in general, when the measurement noise variance decreases and that is for both the modified Kalman smoother and smoothing spline. More importantly, these figures indicate clearly the superiority of the performance of the (MKS) on that of the (SS). However, for signal number **8** the performance of the (SS) is close to that of the (MKS) because this signal does not have large local variations.

TABLE_4.2.2: This table gives 1.E+5x(MSE) of the (SS) and the (MKS) when $\sigma^2 = (0.05)^2$.

	signal1		*signal2*		*signal3*		*signal4*		*signal5*	
ESTI-MATO-RS	SS	MKS	SS	MKS	SS	MKS	SS	MKS	SS	MKS
noise 1	19.33	10.13	24.42	10.23	13.45	6.58	18.08	5.98	29.28	20.75
noise 2	17.12	3.96	24.17	4.31	10.27	3.03	12.08	2.24	10.96	13.21
noise 3	18.15	10.22	26.75	12.35	12.60	7.85	10.97	5.92	32.29	12.17
noise 4	25.63	6.67	32.29	8.79	17.82	3.99	12.05	2.67	40.38	22.95
noise 5	19.05	14.31	23.28	17.96	16.80	7.23	17.75	4.16	26.63	13.03
noise 6	23.02	7.72	35.77	9.79	13.19	5.39	16.98	3.61	41.07	12.73
noise 7	21.69	5.42	26.14	5.60	18.09	2.52	16.60	2.05	27.59	20.30
noise 8	9.95	6.03	13.13	6.41	7.12	3.55	21.10	2.29	20.66	12.42
noise 9	28.78	9.01	39.17	12.14	15.97	6.36	10.96	4.57	41.77	22.01
noise 10	18.23	7.05	24.30	7.27	11.86	4.81	15.08	4.11	20.12	16.98
AMSE	20.10	8.10	26.94	9.49	13.72	5.13	15.17	3.76	29.08	16.63

TABLE_4.2.3: This table gives 1.E+5x(MSE) of the (SS) and the (MKS) when $\sigma^2 = (0.05)^2$.

	signal6		*signal7*		*signal8*		*signal9*		*signal10*	
ESTI-MATO-RS	SS	MKS	SS	MKS	SS	MKS	SS	MKS	SS	MKS
noise 1	16.77	7.84	37.33	19.37	13.40	9.42	16.81	13.81	19.59	19.21
noise 2	13.72	3.81	42.04	11.42	10.92	4.45	15.24	7.56	15.05	21.03
noise 3	15.38	9.93	46.64	10.81	6.20	7.29	15.19	14.55	13.90	13.30
noise 4	22.30	5.67	44.39	20.46	10.83	3.65	23.18	14.13	15.12	27.55
noise 5	18.54	11.55	38.56	11.00	3.10	11.03	18.73	21.31	22.31	8.29
noise 6	18.30	7.01	50.80	11.84	8.49	5.63	16.96	11.50	22.14	12.42
noise 7	20.20	3.77	38.89	18.15	14.27	5.43	20.62	11.79	18.78	17.23
noise 8	82.40	4.73	28.82	11.85	6.93	5.34	8.67	7.14	25.11	13.25
noise 9	22.72	8.28	52.36	20.47	9.00	4.49	22.51	15.57	15.36	25.47
noise 10	15.25	5.83	39.69	15.40	12.16	6.94	16.03	10.69	17.32	20.12
AMSE	24.56	6.84	41.95	15.08	9.53	6.37	17.39	12.81	18.47	17.79

TABLE_4.2.4: This table gives 1.E+4x(MSE) of the (SS) and the (MKS) when $\sigma^2 = (0.1)^2$.

	signal1		*signal2*		*signal3*		*signal4*		*signal5*	
ESTI-MATO-RS	SS	MKS	SS	MKS	SS	MKS	SS	MKS	SS	MKS
noise 1	5.75	3.27	7.54	3.39	3.50	2.27	5.68	2.14	8.44	5.25
noise 2	4.45	1.05	6.53	1.31	2.84	0.80	4.33	0.64	3.21	3.06
noise 3	5.06	3.13	7.11	3.76	4.01	2.39	3.41	2.16	7.41	3.47
noise 4	7.36	1.28	9.95	2.06	4.41	1.00	3.63	0.80	10.43	4.25
noise 5	6.62	3.79	7.54	4.31	5.98	1.78	4.58	1.48	7.32	3.82
noise 6	5.33	1.90	9.06	2.62	3.42	1.47	4.73	1.25	9.05	3.65
noise 7	7.36	1.41	8.54	1.42	7.72	0.65	5.17	0.35	8.33	4.02
noise 8	2.98	1.55	3.86	1.94	2.45	0.94	5.87	0.75	4.97	4.32
noise 9	6.79	1.75	11.08	2.91	4.14	1.85	3.21	1.63	10.37	4.76
noise 10	5.10	2.16	7.04	2.35	3.17	1.54	5.01	1.39	5.83	4.16
AMSE	5.68	2.13	7.83	2.61	4.16	1.47	4.56	1.26	7.54	4.08

TABLE_4.2.5: This table gives 1.E+4x(MSE) of the (SS) and the (MKS) when $\sigma^2=(0.1)^2$.

	signal6		*signal7*		*signal8*		*signal9*		*signal10*	
ESTI-MATO-RS	SS	MKS	SS	MKS	SS	MKS	SS	MKS	SS	MKS
noise 1	4.83	2.61	14.80	4.90	4.68	3.98	4.89	3.86	6.41	4.19
noise 2	3.64	0.99	11.14	2.80	3.63	1.22	3.68	1.64	4.96	4.26
noise 3	4.60	2.86	11.95	3.27	1.81	2.33	4.43	4.32	4.06	3.91
noise 4	6.38	1.26	13.81	3.76	3.08	4.55	6.44	2.81	4.54	6.52
noise 5	6.43	2.59	9.79	3.63	0.74	3.61	6.34	5.80	6.26	2.05
noise 6	4.53	1.78	15.64	3.43	2.75	1.50	4.11	3.20	6.35	3.13
noise 7	6.79	0.94	10.61	3.61	4.77	1.40	6.81	2.06	5.76	2.54
noise 8	2.72	1.23	7.14	4.17	2.07	1.39	2.66	2.30	7.80	4.23
noise 8	5.52	2.10	16.05	4.30	2.91	0.63	5.36	3.84	3.94	6.57
noise 10	4.24	1.80	12.97	3.85	4.16	2.60	4.29	2.75	5.69	4.23
AMSE	4.97	1.82	12.39	3.77	3.06	2.32	4.90	3.23	5.58	4.16

TABLE_4.2.6: This table gives 1.E+3x(MSE) of the (SS) and the (MKS) when $\sigma^2 = (0.5)^2$.

	*signal*1		*signal*2		*signal*3		*signal*4		*signal*5	
ESTI-MATO-RS	SS	MKS	SS	MKS	SS	MKS	SS	MKS	SS	MKS
noise 1	4.25	6.74	7.23	0.63	2.10	5.16	9.32	5.14	10.30	5.51
noise 2	3.69	1.40	5.93	1.30	3.51	1.42	5.53	1.40	5.47	2.54
noise 3	7.82	6.17	9.41	6.04	7.53	5.23	4.08	5.21	3.92	4.55
noise 4	3.87	1.30	8.05	1.77	2.78	1.83	4.65	1.81	6.04	1.20
noise 5	6.70	7.14	14.32	5.48	7.7	3.56	2.43	3.54	2.41	4.73
noise 6	5.28	2.89	7.46	2.96	4.70	2.97	5.94	2.94	6.21	3.44
noise 7	4.99	2.19	8.88	1.56	3.12	1.13	8.03	1.11	9.88	2.35
noise 8	4.49	2.22	5.66	2.20	2.70	1.73	3.68	1.71	3.77	5.60
noise 9	7.27	2.09	9.24	3.09	5.35	3.92	5.89	3.89	7.35	1.80
noise 10	3.97	4.07	6.58	0.97	2.81	3.29	7.43	3.27	7.89	4.03
AMSE	5.23	3.62	8.28	2.60	4.23	3.02	5.70	3.00	6.32	3.58

TABLE_4.2.7: This table gives 1.E+3x(MSE) of the (SS) and the (MKS) when $\sigma^2 = (0.5)^2$.

	signal6		*signal7*		*signal8*		*signal9*		*signal10*	
ESTI-MATO-RS	SS	MKS	SS	MKS	SS	MKS	SS	MKS	SS	MKS
noise 1	2.94	5.57	12.34	5.44	7.85	6.70	3.01	6.63	11.24	4.91
noise 2	3.96	1.46	7.67	2.48	3.17	2.37	3.43	1.59	9.74	2.60
noise 3	8.08	5.63	8.17	4.50	2.32	5.13	5.10	6.39	7.12	4.13
noise 4	3.62	1.78	9.34	1.13	3.50	0.26	3.19	1.86	6.41	5.88
noise 5	12.06	4.44	7.79	4.67	1.29	8.24	4.25	6.42	5.14	1.10
noise 6	5.18	2.98	9.11	3.38	4.65	2.98	4.84	3.22	7.71	2.99
noise 7	4.07	1.41	12.65	2.28	6.86	2.75	3.98	2.05	10.04	1.81
noise 8	4.01	1.86	6.11	5.54	2.96	2.68	3.35	2.38	6.09	11.4
noise 9	6.67	3.71	10.97	1.70	5.89	0.60	6.14	3.14	6.35	6.53
noise 10	3.45	3.51	10.00	3.96	5.51	4.53	3.22	4.11	10.49	3.75
AMSE	5.40	3.23	9.41	3.50	4.40	3.62	4.05	3.77	8.03	4.51

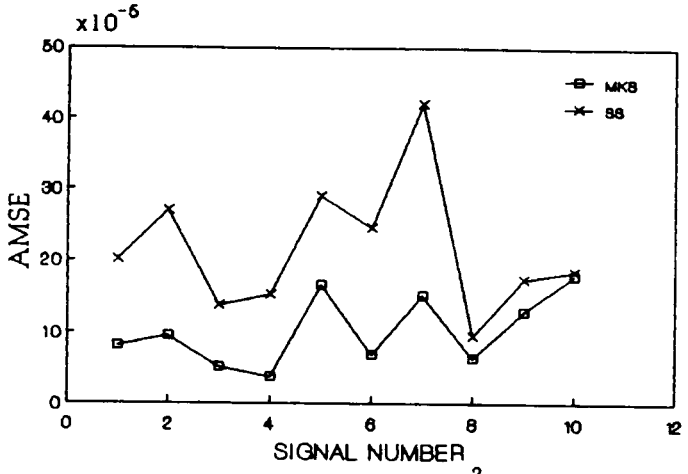


Fig.4.1: (AMSE) plot when $\sigma^2=(0.05)^2$

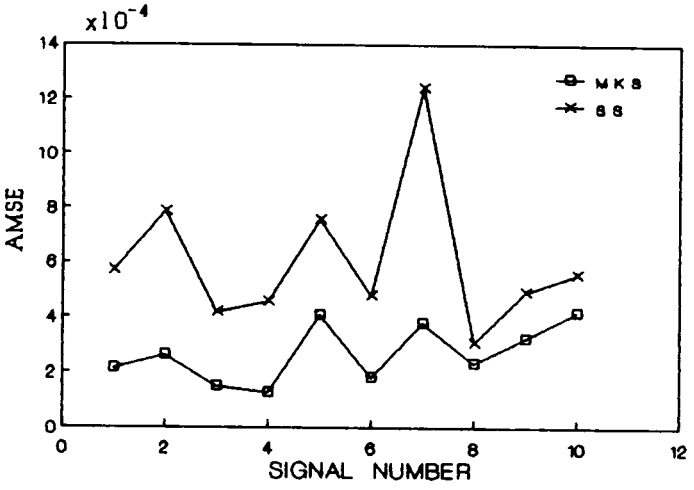


Fig.4.2: (AMSE) plot when $\sigma^2=(0.1)^2$

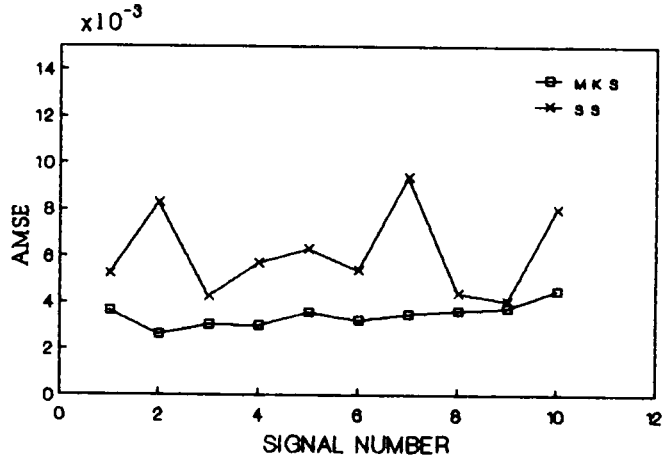


Fig.4.3:(AMSE) plot when $\sigma^2=(0.5)^2$

As an example, the signal and the measurement models in state space representation of the signal number 1 are given by Eq(4.2.4):

$$x(n+1) = \begin{bmatrix} .97 & .1 \\ 0 & .87 \end{bmatrix} x(n) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} w(n)$$

$$s(n) = [2 \quad 0]x(n)$$

$$z(n) = s(n) + v(n) \tag{4.2.4}$$

The transfer function of the signal can be obtained from Eq(4.2.1).

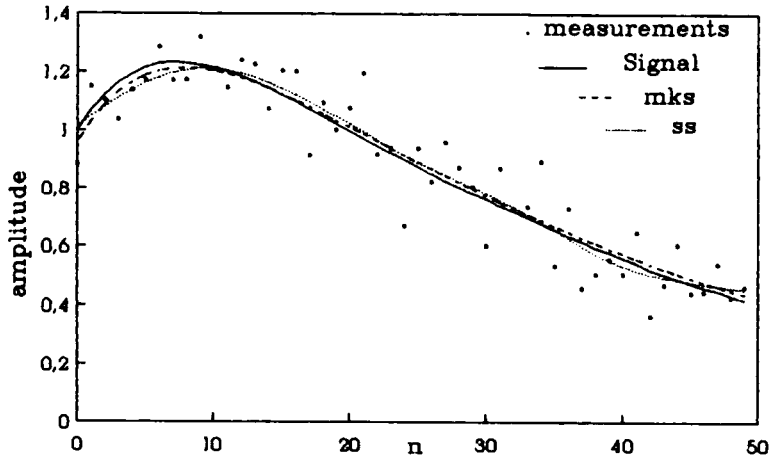
$$H(z) = \frac{z - 0.77}{(z - 0.97)(z - 0.87)} \tag{4.2.5}$$

In this illustrative example, the first realisation for the measurement noise is used which is denoted by *noise1* in the tables:(4.2.2 to 4.2.7).

The estimated signal by (SS) and (MKS) are plotted together with the measurements and the actual signal for $\sigma^2 = (0.1)^2$ and $\sigma^2 = (0.5)^2$, as shown in Fig:4.4 and Fig:4.5, respectively.

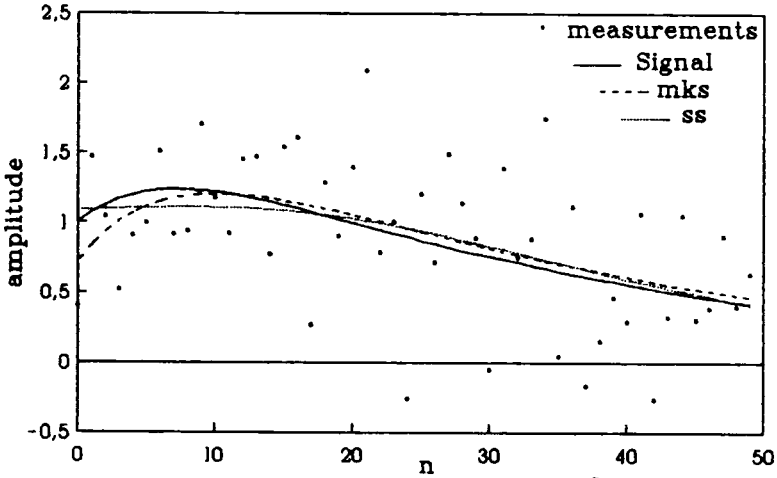
Note that in Fig.4.4, obtained for $\sigma^2 = (0.1)^2$, both the (MKS) and (SS) follows satisfactorily the noise-free signal and the transient effect are not visible. This is due to a good estimation of the initial state vector. However, in Fig.4.5 we increase our incertitude on the measurement by increasing the measurement noise variance to $\sigma^2 = (0.5)^2$. This has bad effect on the (MKS) estimator since the transient effects have not been decreased at the beginning of the estimation. The transient effects are visible because the (MKS) uses a suboptimal gain with a poor estimation of the initial state vector due to the large incertitude on the measurement. As more data records are used the estimated signal becomes closer and closer to the

noise free-signal. On the other hand, the (SS) appears like a straight line trough the measurement which is confirmed by a small value of the smoothing parameter, $\rho = 6.6E-6$, found by (GXV).



smoothing parameter by (GXV) ρ =0.006

Fig.4.4:plot of the example when $\sigma^2=(0.1)^2$



smoothing parameter by (GXV) ρ =6.6 1.E-6

Fig.4.5: plot of the example when $\sigma^2=(0.5)^2$

In the case of short data records only the first sixteen measurement data and only the first realisation of the measurement noise is used to compare the performance of the two estimators for the ten different signals. The criterion used for comparison is the (MSE). The (MSE) computed for both the (MKS) and (SS), for the three different measurement noise variance, is given in tables:(4.2.8 to 4.2.10) and plotted as function of the ten signals numbered from **1** to **10** in Figs:4.6, 4.7 and 4.8. We also notice that, the (MSE) decreases when the measurement noise variance decreases. Furthermore, the modified Kalman smoother appears to perform better than the smoothing spline although the superiority is not as obvious as in the case of long data records.

TABLE_4.2.8: This table gives 1.E+4x(MSE) of the (MKS) and the (SS), when $\sigma^2 = (0.05)^2$.

	signal 1	*signal* 2	*signal* 3	*signal* 4	*signal* 5	*signal* 6	*signal* 7	*signal* 8	*signal* 9	*signal* 10
MKS	1.92	2.06	2.07	0.71	2.84	2.03	2.55	1.80	2.53	2.80
SS	3.09	3.63	2.79	2.21	3.78	3.02	3.73	2.41	3.09	3.09

TABLE_4.2.9: This table gives 1.E+4x(MSE) of the (MKS) and the (SS), when $\sigma^2=(0.1)^2$.

	signal 1	*signal* 2	*signal* 3	*signal* 4	*signal* 5	*signal* 6	*signal* 7	*signal* 8	*signal* 9	*signal* 10
MKS	7.34	7.56	7.59	8.72	7.78	7.72	7.31	6.81	8.16	7.91
SS	10.3	11.4	10.4	9.96	13.0	9.3	13.2	9.01	12.2	9.76

TABLE_4.2.10: This table gives 1.E+3x(MSE) of the (MKS) and the (SS), when $\sigma^2=(0.5)^2$.

	signal 1	*signal* 2	*signal* 3	*signal* 4	*signal* 5	*signal* 6	*signal* 7	*signal* 8	*signal* 9	*signal* 10
MKS	21.3	18.3	21.0	19.1	15.2	18.9	15.1	16.8	18.4	20.6
SS	21.5	34.2	21.1	22.5	22.9	23.1	35.2	20.6	22.0	27.8

For comparative purposes we define the following ratios:

$$R_{SS}=\frac{MSE_{SS_L}}{MSE_{SS_S}},\quad R_{MKS}=\frac{MSE_{MKS_L}}{MSE_{MKS_S}}$$

$$R_L=\frac{MSE_{MKS_L}}{MSE_{SS_L}},\quad R_S=\frac{MSE_{MKS_S}}{MSE_{SS_S}}$$

where:

MSE_{ss_s} stands for the (MSE) computed for the (SS) in the case of short data records.

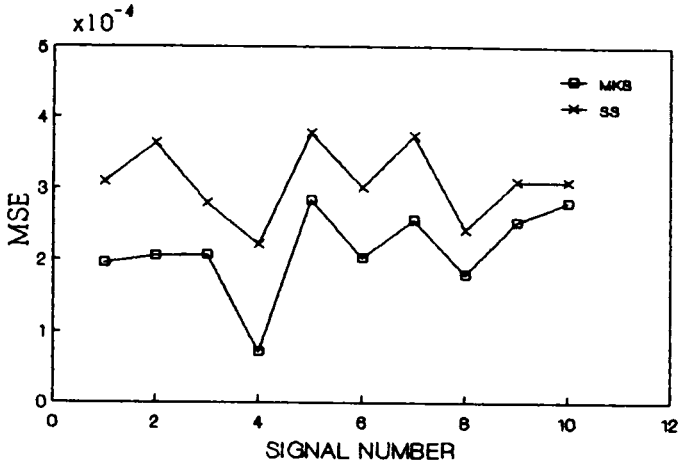


Fig.4.6:(MSE) plot when $\sigma^2=(0.05)^2$

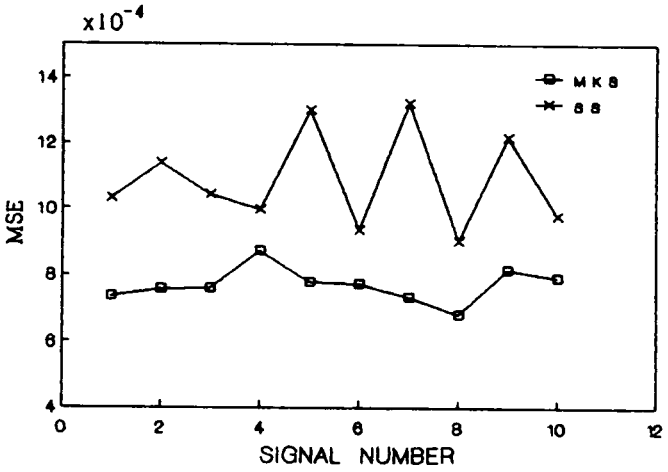


Fig.4.7:(MSE) plot when $\sigma^2=(0.1)^2$

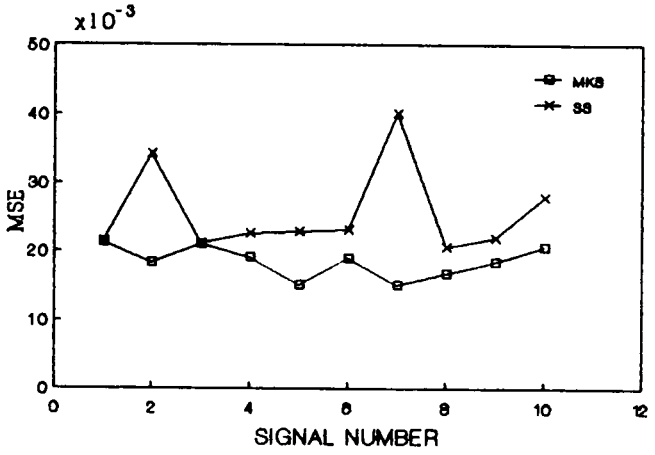


Fig.4.8:(MSE) plot when $\sigma^2=(0.5)^2$

MSE_{SS_L} stands for the (MSE) computed for the (SS) in the case of long data records.

MSE_{MKS_S} stands for the (MSE) computed for the (MKS) in the case of short data records.

MSE_{MKS_L} stands for the (MSE) computed for the (MKS) in the case of long data records.

The ratios are computed for each signal, and for the three different measurement noise variances $\sigma^2=(0.05)^2$, $\sigma^2=(0.1)^2$, and $\sigma^2=(0.5)^2$. The results are given in tables 4.2.11, 4.2.12, and 4.2.13, respectively.

Table_4.2.11: this table gives the values of the ratios when $\sigma^2=(0.05)^2$.

	signal 1	*signal* 2	*signal* 3	*signal* 4	*signal* 5	*signal* 6	*signal* 7	*signal* 8	*signal* 9	*signal* 10
R_{SS}	0.63	0.67	0.48	0.82	0.77	0.56	1.00	0.56	0.54	0.63
R_{MKS}	0.53	0.50	0.32	0.84	0.73	0.39	0.76	0.52	0.55	0.69
R_L	0.52	0.42	0.49	0.33	0.71	0.47	0.52	0.70	0.82	0.98
R_S	0.62	0.57	0.74	0.32	0.75	0.67	0.68	0.75	0.82	0.91

Table_4.2.12: this table gives the values of the ratios when $\sigma^2 = (0.1)^2$.

	signal 1	signal 2	signal 3	signal 4	signal 5	signal 6	signal 7	signal 8	signal 9	signal 10
R_{SS}	0.56	0.66	0.34	0.57	0.65	0.52	1.12	0.52	0.40	0.66
R_{MKS}	0.45	0.45	0.30	0.25	0.67	0.34	0.67	0.58	0.48	0.53
R_L	0.57	0.45	0.65	0.38	0.62	0.54	0.33	0.85	0.79	0.65
R_S	0.71	0.66	0.73	0.88	0.60	0.83	0.55	0.76	0.67	0.81

Table_4.2.13: this table gives the values of the ratios when $\sigma^2 = (0.5)^2$.

	signal 1	signal 2	signal 3	signal 4	signal 5	signal 6	signal 7	signal 8	signal 9	signal 10
R_{SS}	0.20	0.21	0.10	0.41	0.45	0.13	0.35	0.38	0.14	0.40
R_{MKS}	0.32	0.03	0.25	0.27	0.36	0.29	0.36	0.40	0.36	0.24
R_L	1.59	0.09	2.46	0.55	0.53	1.89	0.44	0.85	2.20	0.44
R_S	0.99	0.54	1	0.85	0.66	0.81	0.43	0.82	0.83	0.74

Using the three tables, we deduce that the ratio R_S is larger than R_L. This suggests that the superiority of the performance

of the modified Kalman smoother on that of the smoothing spline decreases as N decreases. On the other hand, the comparison of the ratios R_{ss} with R_{MKS} shows that the smoothing spline is slightly less sensitive than the modified Kalman smoother to short data records.

4.3 INTERPRETATION.

According to the simulation study results reported in the tables:(4.2.2 to 4.2.7), there exist some realisations of the measurement noise, in the case where $\sigma^2=(0.5)^2$, for which the (MSE) of the (SS) is smaller than the one of the (MKS). In general, According to the values of the (AMSE) reported in the last row of the tables:(4.2.2 to 4.2.7) we can conclude that the performance of the modified Kalman smoother using the true signal model is better than that of the smoothing spline.
Signal number 7 has a large variation in the degree of smoothing, see Fig.5.1 of the next chapter. This degrades the performance of the (SS). It is expected because of the pronounced peak in the graphs shown in Figs:(4.1 and 4.2).
Comparing, for the ten different signals, the (mse) of the (SS) and that of the (MKS) given for the noise1 in the case of short and long data records, we note that the performance of the (MKS) are more sensitive to short data records than that of the (SS) when the measurement noise variance is $\sigma^2=(0.05)^2$ or $\sigma^2=(0.1)^2$. While for the measurement noise variance of $(0.5)^2$, the (SS) estimator, using the value of the smoothing

parameter ρ found by the (GXV), is unable to recover the noise-free signal. This is illustrated by plotting the first 16 data samples of signal number 1 with the measurement, the estimated signal by the (SS) using two different smoothing parameters, and the signal estimated by the (MKS), as shown in Fig.4.9. We notice that the estimated signal by (SS), using a smoothing parameter $\rho_{(GXV)}$ found by the (GXV), is undersmoothed. When $\rho = \rho_{(GXV)}/5$, the estimated signal presents less fluctuations and captures the shape of the noise-free signal. On the other hand, (MKS) estimator behaves as if it is a biased estimator because it captures the shape of the signal but the estimated signal is either boosted up or down. The conclusion is that both the (MKS) and the (SS) are unable to follow closely the noise-free signal in the case of short data records observed with large incertitude.

The pronounced peak that occurs in Figs:(4.1 to 4.2) for long data records has been minimized for short data records Figs(4.6 to 4.7). This suggests the use of local smoothing which will be covered in the next chapter.

The (SS) is more sensitive to outlying data samples than the (MKS) since, in the case of the (MKS), the gain matrix used to weight the innovations is suboptimal. This means that the (MKS) is more confident on the signal model than it is on the measurements, so that outlying data samples will not affect too much the estimator performances. However, the outlying data could have an effect in the identification procedure when

the model is identified (not known) which may lead to the use of an incorrect model. The effect of the case of an incorrect model will be investigated in the next section.

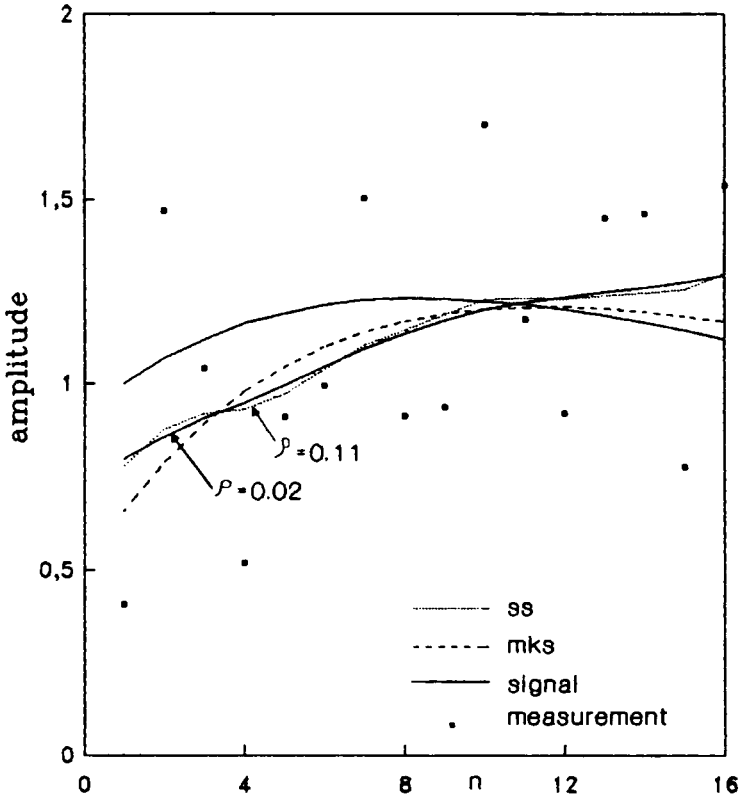


Fig.4.9: estimation of the example when σ^2=(0.5) and N=16.

4.4 MODEL MISMATCH.

Model mismatch comes from three main points. It can come either from lack of knowledge on the physical system under study or the use of a reduced-order model in order to decrease computational complexity or linearizing nonlinear model. It is clear that an imprecise estimator model degrades estimator performance. These modeling errors may even cause the estimator to diverge.

In this section we investigate the effects of using signal models that do not fit the exact assumption needed by the modified Kalman smoother or the smoothing spline estimators. The analysis of the performance of the two estimators is presented under the effect of model mismatch.

4.4.1 MODIFIED KALMAN SMOOTHER

In applying the modified Kalman smoother to a given signal, the state-space model parameter [A,g,C], noise statistics (Q,R), and initial condition (P(1), x(0)) must be specified. In real situations we either not have these parameters or they are estimated which can introduce some errors in the final model. In the following simulation example we consider the process modeling error of the signal given by Eq(4.3.1). For this purpose, the (MKS) estimator is computed using the model set $A+\Delta A$, g, C, R, Q, $P(1)$, ie, only the process dynamics modeling error is considered. ΔA_1 is given by:

$$\Delta A_1 = \begin{bmatrix} -.02 & -.01 \\ .08 & .01 \end{bmatrix}$$

with the measurement noise variance $\sigma^2 = (0.1)^2$.

In figure 4.10, the plots of the signal estimated by the (MKS) under process dynamics modelling error, the measurement that have been obtained from the exact signal model, the true signal, and the used signal. The mean squared estimation error (MSE) is computed under model mismatch and the obtained result was 68.344E-3 while the (MSE) obtained for the (SS) was 5.75E-4. Using this ΔA_1 the signal model has became unstable since one of the poles has been shifted out of the unit circle. It is clearly shown, in Fig.4.10, by the increase in the amplitude of the signal used for the (MKS) computation. An other example for process modelling error, ΔA_2, is taken where the poles and zeros of the modified model are in the unit circle. In this case, the process modelling error ΔA_2 is given by:

$$\Delta A_2 = \begin{bmatrix} -.02 & .05 \\ 0 & -.01 \end{bmatrix}$$

As the previous example; the plots of the measurements, the true signal, the used signal, and the estimated signal are given in Fig.4.11. The mean squared estimation error is computed for this second example and the obtained result was 10.316E-3.

For the diagnosis of the quality of the (MKS) estimator we use the bias test [5] as well as the whiteness test which will be introduced in the next section.

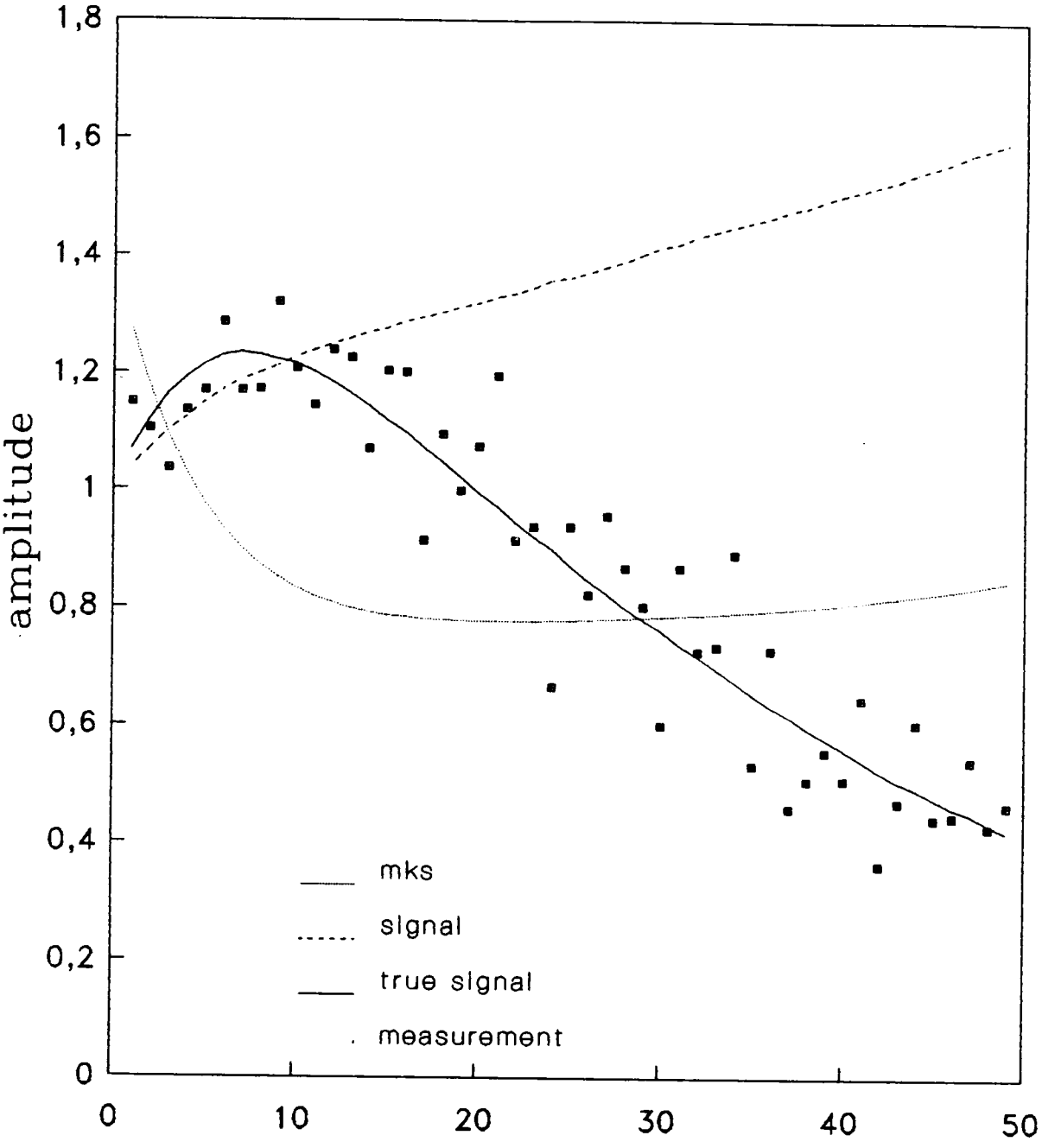


Fig.4.10: estimation under model mismatch ΔA_1

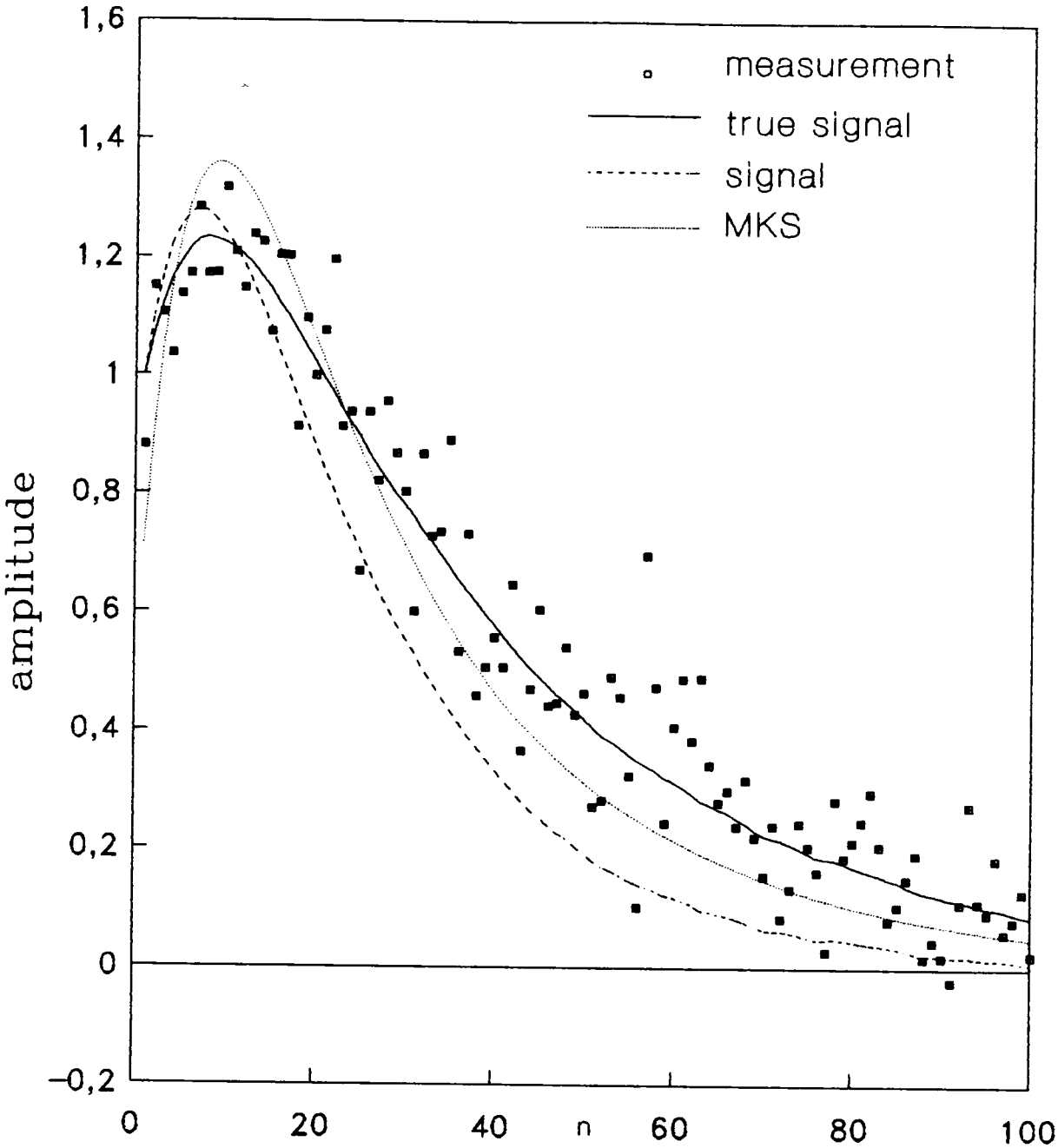


Fig.4.11: estimation under model mismatch ΔA_4

BIAS TEST:

The bias test is to check statistically whether the mean of the innovation is zero or not. When the innovations sequence is ergodic and gaussian, The sample mean for the innovation vector $\alpha(n)$ is given by:

$$m_\alpha = \frac{1}{N}\sum_{n=1}^{N}\alpha(n) \qquad (4.4.1)$$

A test is performed to decide if the innovation mean is zero, ie:

we test the null hypothesis H_o: $m_\alpha = 0$ against the alternative hypothesis H_1: $m_\alpha \neq 0$.

The zero-mean test [5] on the innovation $\alpha(n)$ rejects H_o if $m_\alpha > \mu$ and accepts it if $m_\alpha < \mu$.

Under the null hypothesis H_o, m_α is assumed zero. At 5 percent significant level, the confidence interval for the innovation is given by:

$$\mu = [m_\alpha - 1.96*\sqrt{r_\alpha(0)},\ m_\alpha + 1.96*\sqrt{r_\alpha(0)}] \qquad (4.4.2)$$

where $r_\alpha(0)$ is the sample variance given by:

$$r_\alpha(0) = \frac{1}{N}\sum_{n=1}^{N}\alpha^2(n) \qquad (4.4.3)$$

This test is applied for the first and the second process dynamics modelling error illustrative examples and the results are given in Figs.(4.12 and 4.13), respectively.

Note that, in Fig.(4.12) 36 data samples are out of the band which gives 18 out 100 are out of the band. Sine 18% > 5% we conclud that the estimator is biased. The same conclusion is

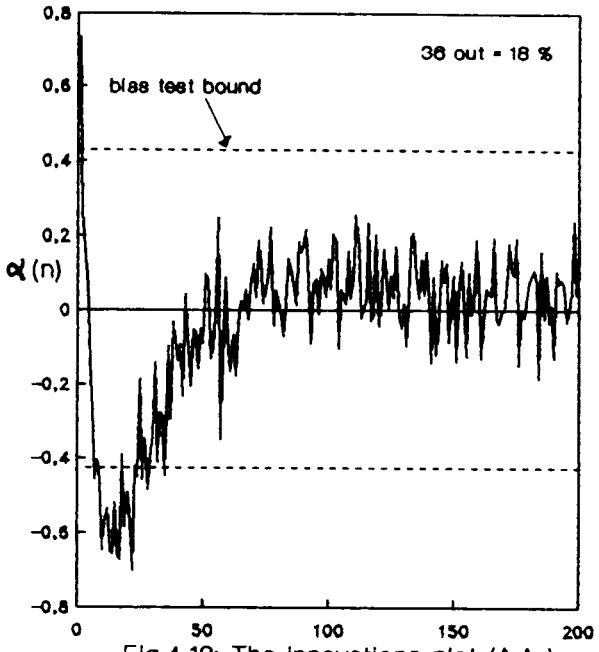


Fig.4.12: The innovations plot (ΔA_1)

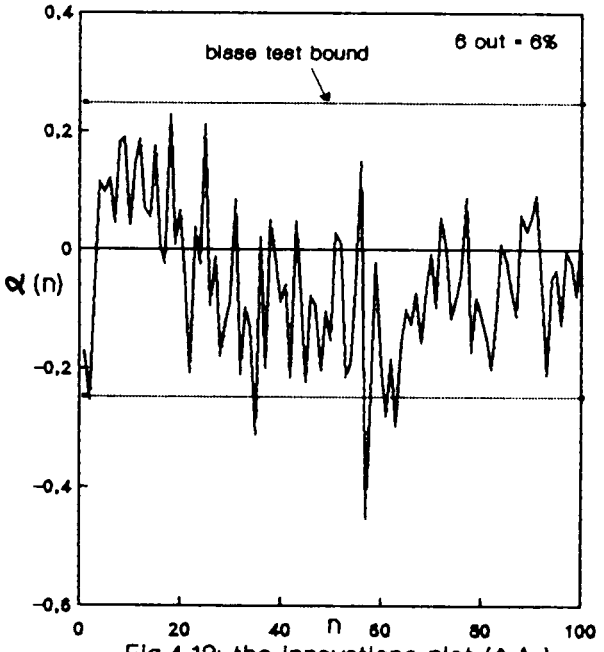


Fig.4.13: the innovations plot (ΔA_2)

found for the second example since 6%>5% as shown in Fig.4.13.

In general, the bias test is not sufficient for the analysis of the estimator performances. Another test is applied to innovation which is the whiteness test to decide on whether the estimator is good or not.

WHITENESS TEST

The whiteness test is to check statistically whether the covariance function of the innovation is that of a white noise or not. When the innovations sequence is ergodic and gaussian, the τ^{th} component covariance is given by

$$r(\tau) = \frac{1}{N}\sum_{j=1}^{N-\tau} \alpha(j)*\alpha(j+\tau) \qquad (4.4.4)$$

where $\alpha(n)$ is the innovation. Furthermore, the 95 percent confidence interval estimate of $r(\tau)$ is given by [5]:

$$I_r(\tau) = \left[r(\tau) - \frac{1.96*r(0)}{\sqrt{N}}, r(\tau) + \frac{1.96*r(0)}{\sqrt{N}} \right] \qquad (4.4.5)$$

When the innovation is white, $r(\tau)=0$ *for* $\tau \neq 0$.

The whiteness test is applied for the first and the second process dynamics modelling error illustrative examples and the results are given in Figs.(4.14 and 4.15), respectively.

Observing the two Figs.(4.14 and 4.15), we notice that the autocovariance functions are out of the bands this means that the innovations are nonwhite.

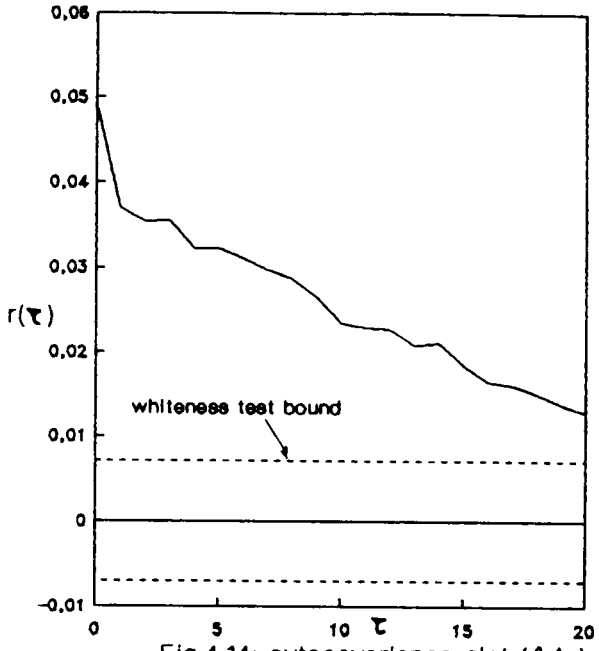


Fig.4.14: autocovariance plot (ΔA_1)

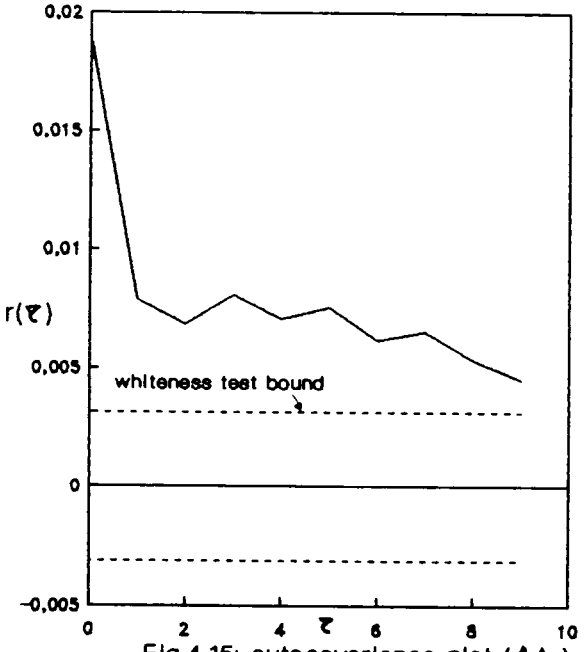


Fig.4.15: autocovariance plot (ΔA_2)

From these simulation runs, we conclude that under small process dynamics modeling inaccuracies, the innovations becomes biased and nonwhite. Although they were judged to be white and unbiased when the true signal has been used. See Fig.4.16 and Fig.4.17 for the bias test and the whiteness test, respectively.

The effect of input modeling error Δg and measurement modeling error ΔC are also investigated. The results of the run show that, for small measurement modelling error, ie, ΔC=[-0.1 0.08], the innovations are unbiased and white. The estimated signal by the (MKS) under small measurement modeling error is shown in Fig.4.18. The mean squared estimation error is computed under model mismatch and the obtained result was 3.28E-4. Under small output modelling error the (MKS) follows closely the noise-free signal. The same result was obtained when taking small input modeling error. However, when either the input or the output modelling errors are large the (MKS) diverges. As an example, taking ΔC=[0.6 0.9], The obtained mean squared estimation error under large output modeling error was 2.155. Similarly, When Δg=[5 -3], The mean squared estimation error is computed under input modeling error and the obtained result was 61.260. Figure 4.19 provides the plots of the signal estimated by the (MKS) under large input modeling error, ie Δg=[5 -3], the measurement that comes from the exact signal model, the true signal, and the signal

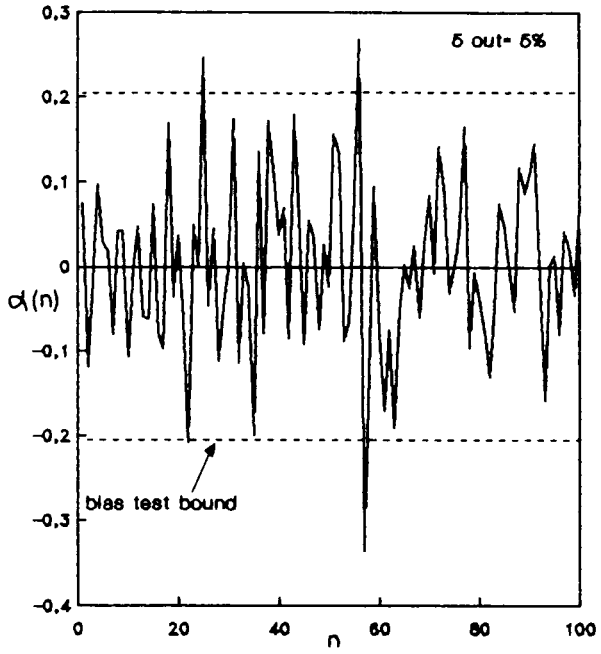


Fig.4.16: the innovations plot

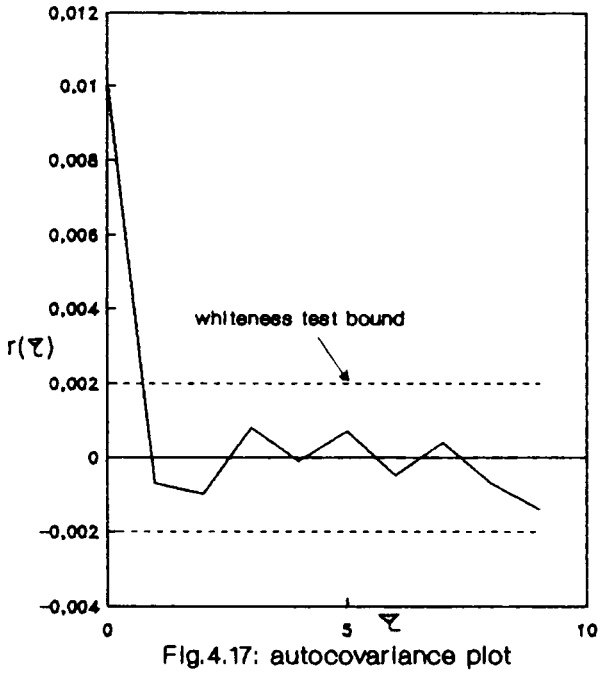


Fig.4.17: autocovariance plot

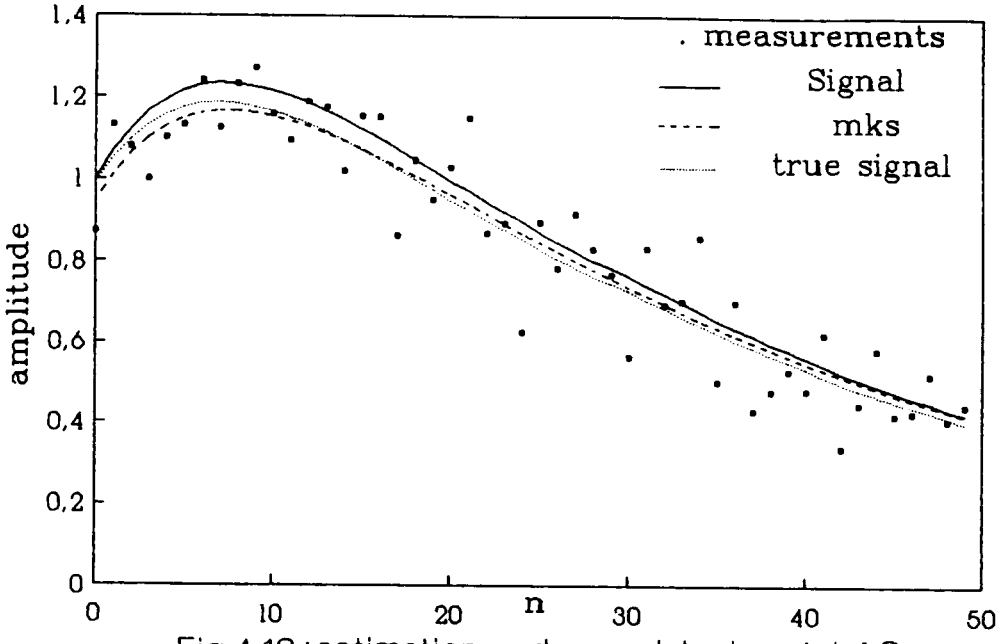


Fig.4.18 : estimation under model mismatch ΔC

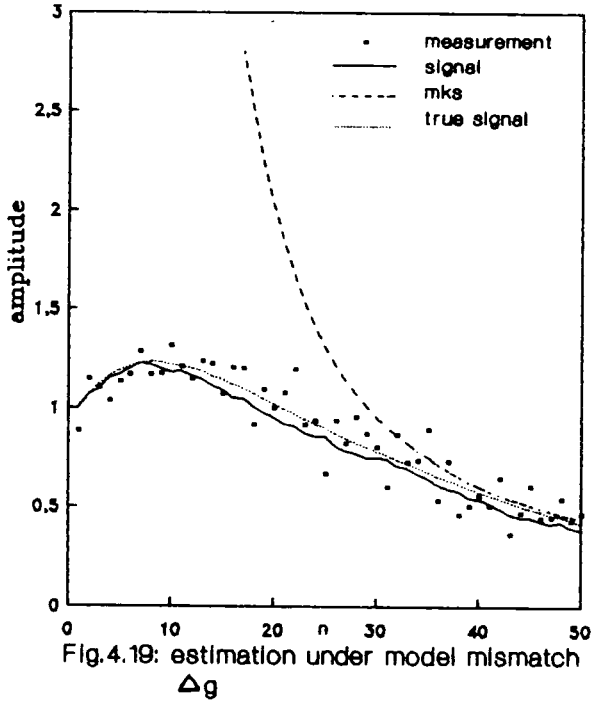


Fig.4.19: estimation under model mismatch Δg

used that comes from the model under input modelling error. The estimated signal is truncated, in the figure 4.19, because of its large amplitude compared to the other signals shown on the same figure.

From these simulation runs, we can conclude that the (MKS) diverges under small process modelling error and large input or output modelling error. However, under small input or output modelling error the (MKS) does not diverge.

4.4.2 SMOOTHING SPLINE

In the case of estimation of signals via smoothing spline the assumption of smoothness may not be true that is the measured data comes from a process that does not satisfy the assumptions imposed by the smoothing spline. The effect of model mismatch in the case of smoothing spline will be investigated by the simulation of two different first order Markov processes. The results are compared to the modified Kalman smoother.

The first simulation example is obtained by passing a white noise of variance 1 through a linear time invariant filter. The measurements consist of the signal corrupted by a white noise of variance $(0.1)^2$. the signal and the measurement models are described in state space representation by:

$$x(n+1) = 0.9802x(n) + 0.04w(n)$$
$$s(n) = x(n) \quad .$$
$$z(n) = s(n) + v(n) \quad . \qquad (4.4.6)$$

with

$x(0) = 0.7$

Figure 4.20 plots both the signal estimated by the (SS) and (MKS) with the actual signal and the measurements. The mean squared estimation error has been computed for both the (MKS) and (SS). The obtained (MSE) of the (MKS) and that of (SS) were 15.9418E-4 and 13.0576E-4, respectively.

The state space representation of the second first order Markov process used in the second simulation example is given by:

$$x(n+1) = 0.3679x(n) + 0.894w(n)$$
$$s(n) = x(n) \quad .$$
$$z(n) = s(n) + v(n) \quad . \qquad (4.4.7)$$

where

$x(0) = 0.7$

Figure 4.21 plots both the signal estimated by the (SS) and (MKS) with the actual signal and the measurement. The mean squared estimation error has been computed for both the (MKS) and (SS) and the results were 89.0650E-4 and 0.6779, respectively. In this example, the signal estimated by the (MKS) is undistinguishble while the (SS) estimator brakes down. From these two previous simulated examples we can conclude that the (SS) is unable to follow a signal which has a large local variations.

A third simulation example where a noisy portion of some triangular wave which is given to the (SS) estimator to see how the smoothing spline behaves in the region of the signal

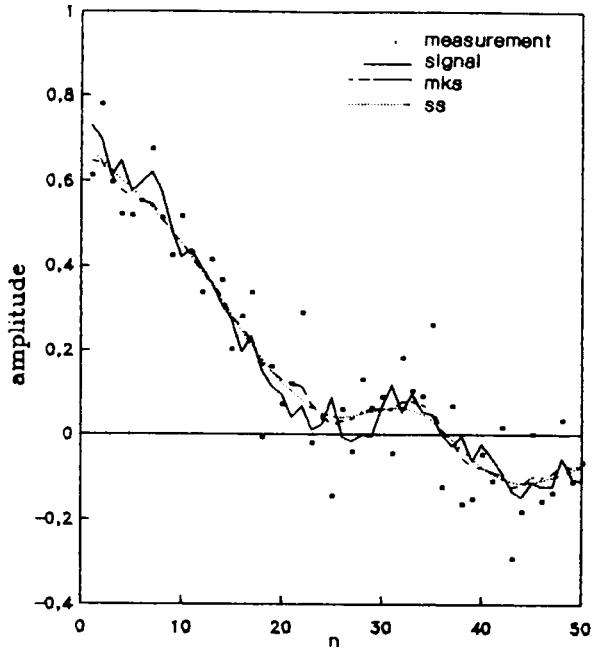


Fig.4.20: Markov process estimation

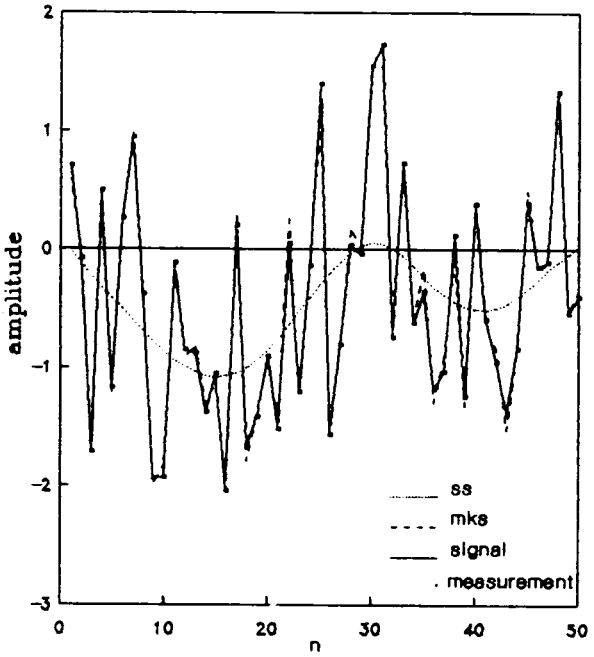


Fig.4.21: Markov process estimation

where the smoothness assumption is violated. For the simulation results see Fig.4.22 where the plots of the signal and the measurements which are obtained by corrupting the signal by white noise of variance $\sigma^2=(0.1)^2$, and the estimated signal are depicted. In figure 4.23, the error which is defined as follows:

$$e(n)=\hat{s}(n)-s(n) \tag{4.4.8}$$

is plotted. $\hat{s}(n)$ is the estimated signal and $s(n)$ is the signal. According to Figs: 4.22 and 4.23, we note that near the peak the (SS) estimator has larger error. The smoothing spline estimator will not follow closely the noise-free signal in the regions where the smoothness assumption is violated. The problem of minimizing the error near the peaks will be investigated in the next chapter.

The conclusion that we can draw form this section is that the smoothing spline estimator does not perform well when we estimate nonsmooth signals with large local variations. The performance of the modified Kalman smoother are better than that of the smoothing spline only when the signal model is precisely known. Therefore, in the situations where we may be unable to provide a precise model for a certain set of measurements; the smoothing spline will be a good alternative.

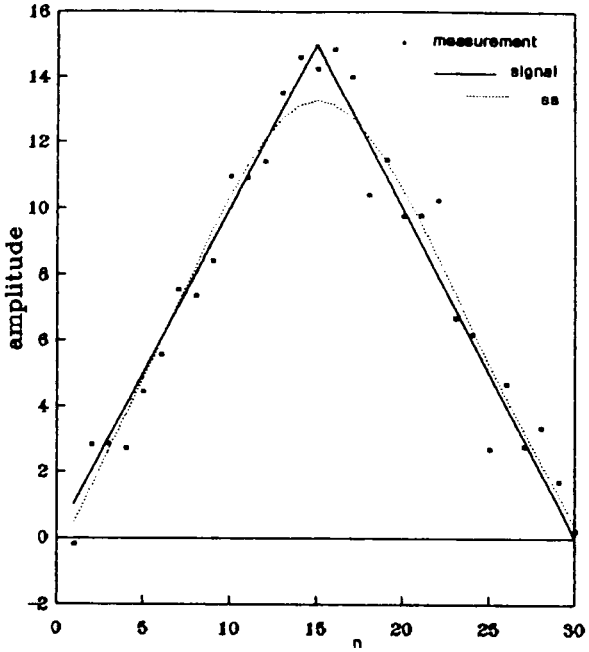


Fig.4.22: SS estimate with σ = 1

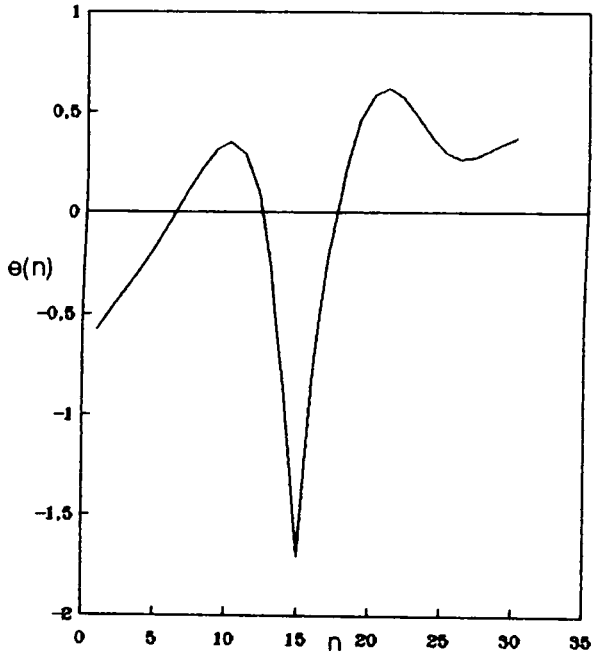


Fig.4.23: the error plot

4.5 COMPUTATIONAL COMPLEXITY

The analysis of computational complexity is presented here in order to compare the efficiency of the two methods. The computational complexity of the two methods are calculated in terms of the numbers of "flops". One flop counts one for a multiplication or a division in the case of real numbers. When the size of the measurement data is N and the underlying linear time invariant signal model has order p, the modified Kalman smoothing methods needs $(70p^3+12p^2)$ flops for the estimation of parameters P, Γ, G, Φ, K and F, and $(5p^2+6p)N$ for signal estimation [32]. On the other hand, the smoothing spline needs $25N$ flops for the signal estimation [27] and $(28N+3)$ flops for the computation of the (GXV) [31]. Therefore, there are about $((28N+3)m+25N)$ flops in total since the (GXV) needs to be evaluated m-times to find the optimal smoothing parameter. Summary of the computational complexity and their description is given in table_4.5.1:

To have a clear idea on the computational complexity, we take two values for m which are 15 and 30. The data size is taken to be N=16 for short data records and N=50 for long data records. The underlying linear time invariant signal model order is taken to be p=1, 2, 4, 6, and 8. The corresponding number of flops are given in table_4.5.2.

ESTIMATORS	DESCRIPTION	NUMBER OF OPERATIONS IN FLOPS
SMOOTHING SPLINE	FOR SIGNAL ESTIMATION	$25N$
	FOR THE (GXV) COMPUTATION	$28N+3$
	IN TOTAL	$(28N+3)m+25N$
MODIFIED KALMAN SMOOTHER	FOR PARAMETER ESTIMATION	$70p^3+12p^2$
	FOR SIGNAL ESTIMATION	$(5p^2+6p)N$
	IN TOTAL	$(70p^3+12p^2)+(5p^2+6p)N$

Table_4.5.1: Summary of the computational complexity of the two estimators.

Table_4.5.2 indicates that the (MKS) involves less computation compared to (SS) when the signals are modelled by small order linear time-invariant systems ($p \leq 4$). It is worth noting that the computation required for the identification of the signal are not included. The computational cost for smoothing spline is more sensitive to data size than that of the modified Kalman smoother.

		NUMBER OF OPERATIONS IN FLOPS		
DATA SIZE	MODEL ORDER	MODIFIED KALMAN SMOOTHER	SMOOTHING SPLINE	
			m=15	m=30
N=16	P=1	258	7165	13930
	P=2	1120	7165	13930
	P=4	6336	7165	13930
	P=6	19008	7165	13930
	P=8	42496	7165	13930
N=50	P=1	632	22295	43340
	P=2	2208	22295	43340
	P=4	9872	22295	43340
	P=6	26352	22295	43340
	P=8	55008	22295	43340

Table_4.5.2: Number of flops for various values of N, p, and m.

4.6 BOOTSTRAP BASED METHOD OF VALIDATION OF THE MODEL USED.

In the previous section, we illustrated the effect of the use of incorrect values of the parameters A, g, and C on the performance of the (MKS) which was shown to be affected seriously when some errors are introduced in the parameters. On the other hand, the application of the whiteness test may lead to erroneous conclusions under model mismatch. For instance, the output errors were judged as nonwhite despite the fact that they were simulated from a white gaussian process. These remarks motivated us to develop a new test which detects the nonvalidity of the model used rather than the nonwhiteness of the output noise.

4.6.1 DESCRIPTION OF THE MODEL

Consider the state-space representation, given in page 24, of the signal:

$$\begin{aligned} x(n+1) &= Ax(n) + gw(n) \\ s(n) &= Cx(n) \\ z(n) &= s(n) + v(n) \end{aligned} \qquad (4.6.1)$$

It is clear that the signal $s(n)$ could be estimated by the (MKS) and the (SS) methods simultaneously. On the other hand, if $\tilde{s}(n)$ denotes the (SS) estimator, the corresponding residuals are defined by:

$$\tilde{v}(n) = z(n) - \tilde{s}(n) \quad for \quad n = 1, \ldots, N \qquad (4.6.2)$$

The sequence $\{\tilde{v}(1),\ldots,\tilde{v}(N)\}$ is a consistant estimate of the unknown output noise $\{v(n)\}$. Therefore, it could be used to determine the required statistics of the unknown output noise $\{v(n)\}$. It is worthnothing that $\{\tilde{v}(n)\}$ is free independant on the true unknown model of the signal $s(n)$. On the other hand, if the model used

$$x(n+1) = Ax(n) + gw(n)$$
$$s(n) = Cx(n) \tag{4.6.3}$$

is a correct model, then the obtained estimator (MKS) fits well the measurements $z(n)$. Furthermore, the obtained estimator should be comparable with the (MKS) estimator computed from artificial measurements simulated from the model corrupted by a simulated noise from the model of $\{\tilde{v}(n)\}$. It is obvious that we can simulate, as many as we like, samples of artificial measurements from which we can obtain a sequence of (MKS) estimators.

The desicion on whether the obtained (MKS) is compared to the simulated ones, may be taken using Bootstrap techniques which consist of constructing a confidence interval of the (MKS) estimator under the hypothesis that the model used is a correct one. If the obtained (MKS) estimator computed from the original data does not belong to the confidence band, we deduce that the original data and the artificial data sets are not generated from the same model. Since the model of the residuals noise $\{\tilde{v}(n)\}$ is a consistant estimator of the model of the unknown output noise $v(n)$, the descripancy between the

original data and the simulated ones is due to the descripancy between the two models (4.6.3) ie the values A, g, and C, used in the computation of the (MKS) estimator, are not good estimates of the unknown parameters. On the other hand, it is natural to expect that the (MKS) computed from the original data will lie within the confidence band if the measurements are governed by model given by Eq(4.6.3) because the original data and the simulated ones are generated from the same model. The confidence band of the estimated (MKS) under the model used can be obtained using the following algorithm:

I_* Simulate a set of measurements $z^*(n)$, n=1,...,N, from the model given by Eq(4.6.3) to be corrupted by simulated noise from the model of the sequence $\{0(1),....,0(N)\}$

II_* Compute the (MKS) estimator MKS_i^* using the measurements $z^*(n)$

III_* Repeat step (I) and step (II) 95 times. 19 runs are sufficient in practice

IV_* Construct the 95% confidence interval using

$$UP\_MKS(n) = \mathrm{Max}_{i=1,\ldots,95}(MKS_i^*(n)) \quad for \quad n = 1,\ldots,N$$

$$LW\_MKS(n) = \mathrm{Min}_{i=1,\ldots,95}(MKS_i^*(n)) \quad for \quad n = 1,\ldots,N$$

4.6.2 SIMULATION OF THE ARTIFICIAL OUTPUT NOISE

A random sample of size N which is required in step 1 of the algorithm can be generated from the law of the sequence $\{0(1),....,0(N)\}$ easily using one of the Bootsrap techniques which

are summarized as follows:

A) PARAMETRIC BOOTSTRAP METHOD:

* Fit a parametric model (Guassian Law in average case) to the sample of residuals $\{\vartheta(1),\ldots,\vartheta(N)\}$, and then simulated samples from the estimated distribution $N(\hat{\mu},\hat{\sigma}^2)$ where

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\vartheta(i) \qquad and \qquad \hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\vartheta(i)-\hat{\mu})^2 \tag{4.6.4}$$

B) DISCRETE NONPARAMETRIC BOOTSTRAP METHOD:

* Simulate a random integer i* from {1,...,N} and take $v^*(i) = \vartheta(i^*)$

* Repeat the above step N times with replacement. This approach is justified by the fact that the empirical distribution function of the residuals is asymptotically a consistant estimate of the unknown distribution of the output noise.

C) SMOOTHED NONPARAMETRIC BOOTSTRAP METHOD:

One could estimate nonparametrically the density of $\{\vartheta(1),\ldots,\vartheta(N)\}$. One popular asymptotically consistant estimator is the Kernel estimator [10,12,29] defined by

$$f_n(v,h) = \frac{1}{\sqrt{2\pi}nh}\sum_{i=1}^{N}\exp\left(\frac{v-\vartheta(i)}{h}\right) \tag{4.6.5}$$

Simulated observations from $f_n(v,h)$ can be obtained using:

$$v^*(i) = \left(1+\frac{\hat{\sigma}^2}{\hat{h}^2}\right)^{-1/2}\{\vartheta(i)+\hat{h}\epsilon_i\} \tag{4.6.6}$$

where $\hat{h}$ is an estimator of the smoothing parameter and ϵ_i is a simulated number from the standard gaussian law.

The optimal value of $\hat{h}$ can be defined as the value which minimises the integrated mean square error of f_n defined by:

$$MISE = \int_R (f_n(v,h) - N_v(\hat{\mu},\hat{\sigma}^2))^2 dv \qquad (4.6.7)$$

The minimisation of the MISE leads to take:

$$\hat{h} \approx 1.06\hat{\sigma}N^{-1/5} \qquad (4.6.8)$$

4.6.3 ILLUSTRATIVE EXAMPLE:

In order to illustrate this test, we take the second example, described in the last section, of the process dynamics modelling error where

$$\Delta A = \Delta A_2 = \begin{bmatrix} -.02 & .05 \\ 0 & -.01 \end{bmatrix}$$

It is clear from Figs(4.13 and 4.15) that the innovations are biased and nonwhite. The answer to the question, "does the divergence of the (MKS) come from the model mismatch", is provided by the following bootstrap test for model validation. First we have applied the (SS) estimator to the original measurements and obtain an estimated signal $\hat{s}(n)$. From the residuals which are defined by

$$\hat{v}(n) = z(n) - \hat{s}(n) \quad for \quad n = 1,\ldots,N \qquad (4.6.7)$$

we obtained an estimate of the output noise statistics:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\hat{v}(i) \qquad and \qquad \hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\hat{v}(i)-\hat{\mu})^2 \qquad (4.6.8)$$

We found $\hat{\mu} = 3.3E-18$ and $\hat{\sigma} = 0.094$.

Then we simulate 19 realisations, using the Gaussian law, with

parameters $\tilde{\mu}$ and $\tilde{\sigma}$. Moreover, the (MKS) is run for each of the 19 realisations using the model under process dynamics modelling error, $\Delta\Lambda_2$. The 19 obtained (MKS) estimators enable us to construct a 95% confidance band to the (MKS) estimator of the original data under the model. For the graphical illustration see Fig.4.24 where the bands, the estimate of the (MKS) using the original data and the signal used are plotted. We notice that the (MKS) estimate using the original data does not lie within the constructed bands. This means that the model which had been used for the estimation is not the appropriate one. In figure 4.25 the same procedure has been conducted to obtain a 95% confidence band for the (MKS) estimate when the exact model parameter for the signal was used. This figure shows that the (MKS) estimate using the original data lies within the band which is an expected result under no model mismatch.

CHAPTER V:

LOCAL SMOOTHING SPLINE

5.1 INTRODUCTION

We were concerned with nonparametric regression and data smoothing for which smoothing spline (SS), an optimal solution of a constrained least square based on the minimisation of the second derivative [24], has been found useful. However, it has been shown through simulation studies of the previous chapter that the performance of the (SS) degrades when fitting processes whose degree of smoothness may not be constant [16,27]. This is due to the fact that the parameter which controls the degree of smoothing applied to the measurements is constant. In order to overcome this problem the smoothing parameter should not be kept constant. It should take different values depending on the curvature of the measurement pattern which is nothing else than the curvature of the actual signal when the noise is of zero mean. Our aim in this chapter is to develop a

new nonparametric estimator, local smoothing spline, for overcoming the problem of global smoothing. The proposed estimator which, is going to be denoted by (LSS), is based on local smoothing parameter that controls the smoothness of the measurement signal in different intervals. For example, near a peak a relatively large value of smoothing parameter is appropriate, whereas a small value should be used on approximately linear section. Other techniques have been developed for the adaptation of the local smoothing parameter in the case of the nonparametric estimator known as the kernel estimator [12,16,29].

In section two, we shall develop the (LSS). The superiority of the (LSS) over the (SS) is demonstrated in section three. Concluding remarks and further scopes are presented in section four.

5.2 LOCAL SMOOTHING SPLINE.

It can be easily seen that the smoothing parameter, in standard smoothing spline, is the Lagrange multiplier. Our idea is very simple one: in order to end up with, for example, two different smoothing parameters, the integral of Eq(3.2.2) is split into two integrals. As a result, we have two constrained problems which are solved "independently", following Reinsch approach, with an additional constraint that the two estimated signals and their first two derivatives must be continuous at

the point where the integral is split.

Suppose that the measurement model is given by

$$z(t_n) = s(t_n) + v(t_n) \tag{5.2.1}$$

where $s(t)$ is a signal observed over a fixed interval $[t_1, T_N]$ in a noisy environment described by an uncorrelated zero mean noise $v(t_i)$ with variance σ^2.

In order to filter the signal which is composed of two signals one observed in the interval $[t_1, T_j]$ and the other observed in the interval $[t_{j+1}, T_N]$.

the first fidelity criteria

$$minimize \int_{t_1}^{T_j} |\hat{s}_1''(t)|^2 dt \tag{5.2.2}$$

are defined among the class of all functions for which $\hat{s}_1(t)$ and $\hat{s}_1'(t)$ is absolutely continuous and $\hat{s}_1''(t)$ is square-integrable such that

$$\sum_{n=1}^{j} (z(t_n) - \hat{s}_1(t_n))^2 \le S_1 \tag{5.2.3}$$

and the second fidelity criteria

$$minimize \int_{t_{j+1}}^{T_N} |\hat{s}_2''(t)|^2 dt \tag{5.2.4}$$

are defined among the class of all functions for which $\hat{s}_2(t)$ and $\hat{s}_2'(t)$ is absolutely continuous and $\hat{s}_2''(t)$ is square-integrable such that

$$\sum_{n=j+1}^{N} (z(t_n) - \hat{s}_2(t_n))^2 \le S_2 \tag{5.2.5}$$

Applying the well known Lagrange method after normalisation of the inequality constrained Eq(5.2.3) by introducing a dummy variable β_1.

the object function

$$J_1 = \int_{t_1}^{t_j} |\hat{s}_1''(t)|^2 dt + \lambda_1 \left\{ \sum_{n=1}^{j} (z(t_n) - \hat{s}_1(t_n))^2 + \beta_1^2 - S_1 \right\} \qquad (5.2.6)$$

must be minimized.

The same procedure is done for $\hat{s}_2(t)$ then

the object function

$$J_2 = \int_{t_{j+1}}^{t_N} |\hat{s}_2''(t)|^2 dt + \lambda_2 \left\{ \sum_{n=j+1}^{N} (z(t_n) - \hat{s}_2(t_n))^2 + \beta_2^2 - S_2 \right\} \qquad (5.2.7)$$

where β_2 is a dummy variable; must be minimized.

According to Reinsch [24], $\hat{s}_1(t)$ and $\hat{s}_2(t)$ must satisfy conditions given by Eqs:(3.2.6 to 3.2.10) and Eq(3.2.12) except that

$$\hat{s}_1^{(2)}(t_j),\ \hat{s}_1^{(3)}(t_j),\ \hat{s}_2^{(2)}(t_{j+1}),\ and\ \hat{s}_2^{(3)}(t_{j+1}) \qquad (5.2.8)$$

are left to be determined from the continuity constrains imposed on the two estimated signals at t_{j+1} which are as follows:

$$\hat{s}_2^{(k)}(t_{j+1})_- - \hat{s}_2^{(k)}(t_{j+1})_+ = 0 \quad if \quad k = 0, 1, 2 \qquad (5.2.9)$$

See fig.5.1 for graphical interpretation.

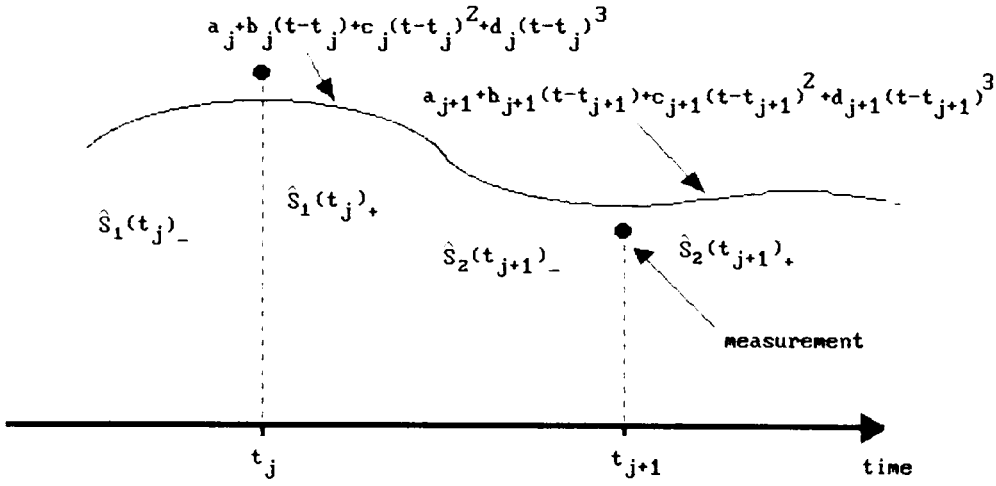


Fig:5.1 Graphical illustration of the continuity conditions.

Then the estimated signal over the interval $[t_1, T_N]$ is composed of $\hat{s}_1(t)$ *and* $\hat{s}_2(t)$. The resulting estimate is then formed by cubic polynomials.

$$\hat{s}(t) = a_n + b_n(t-t_n) + c_n(t-t_n)^2 + d_n(t-t_n)^3 \qquad for\ t_n \le t < t_{n+1} \qquad (5.2.10)$$

which join at their common endpoints such that $\hat{s}(t)$, $\hat{s}^{(1)}(t)$ and $s^{(2)}(t)$ are continuous. Hence, the solution is a cubic spline. The relation among the spline coefficients is obtained by inserting $\hat{s}_1(t)$ and $\hat{s}_2(t)$ into Eqs(3.2.6 to 3.2.10), and satisfying Eq(5.2.9). This yields to the following set of equations:

when $k = 2$

$$c_1 = c_N = d_1 = d_N = 0,\ d_n = (c_{n+1} - c_n)/(3h_n)\ for \quad n = 2, .., N-1. \qquad (5.2.11)$$

when $k = 0$

$$b_n = (a_{n+1} - a_n)/h_n - c_n h_n - d_n h_n^2 \qquad for \quad n = 1, \ldots, N-1. \tag{5.2.12}$$

when $k = 1$

$$Tc = Q^T a \tag{5.2.13}$$

when $k = 3$

$$Qc = M(z - a). \tag{5.2.14}$$

where

$$h_n = t_{n+1} - t_n$$

$$c = [c_2, c_3, \ldots, c_{N-1}]^T$$

$$z = [z_1, z_2, \ldots, z_N]^T$$

$$a = [a_1, a_2, \ldots, a_N]^T$$

and the smoothing matrix M has the following form:

$$\mathbf{M} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_1 & & \\ & & \lambda_2 & \\ 0 & & & \lambda_2 \end{bmatrix} \tag{5.2.15}$$

T and **Q** are as defined in Eq(3.2.16), for unit sampling period.

A left hand multiplication of Eq(5.2.14) by $Q^T M^{-1}$ separates the variable c:

$$(Q^T M^{-1} Q + T)c = Q^T z \tag{5.2.16}$$

$$a = z - M^{-1} Qc \tag{5.2.17}$$

If we denote by $\hat{s}_M$ the estimated signal evaluated at the

design points $\{t_n\}$, then the local smoothing spline estimator, can be written in the following matrix form:

$$\hat{s}_\mu = (I - M^{-1}Q(Q^T M^{-1}Q + T)^{-1}Q^T M^{-1})z \qquad (5.2.18)$$

and in more concise form the (LSS) can be written as

$$\hat{s}_\mu = A(M)z \qquad (5.2.19)$$

where

$$A(M) = (I + M^{-1}QT^{-1}Q^T)^{-1} \qquad (5.2.20)$$

If we denote by λ_{opt} the minimiser of the expression given by Eq(3.3.2), and if $\lambda_{1\ opt} = \lambda_{2\ opt} = \lambda_{opt}$ then the local smoothing spline (LSS) becomes the standard smoothing spline (SS).
$\lambda_{1\ opt}$ *and* $\lambda_{2\ opt}$ are found by the minimisation of the generalised cross validation (GXV) criterion for the first set of the measurements, $\{z(1),\ldots,z(j)\}$, and the second set of the measurements, $\{z(j+1),\ldots,z(N)\}$, respectively. A detailed illustration of the application of the (LSS) will be given in the next section.

5.3_ ILLUSTRATION OF THE METHOD

The signal is taken to be the output of a linear time-invariant filter driven by a zero mean white noise, $w(n)$, with variance $(0.001)^2$. The signal and the measurement models are given, in state space representation, by Eq(5.3.1). The actual signal has linear section as well as a curvature, as shown in Fig.5.2.

$$x(n+1) = \begin{bmatrix} .97 & .3 \\ 0 & .74 \end{bmatrix} x(n) + \begin{bmatrix} 2.2 \\ 1.5 \end{bmatrix} w(n)$$

$$s(n) = [1.8 \quad 0]x(n)$$

$$z(n) = s(n) + v(n) \qquad (5.3.1)$$

With

$x(0) = [.5 \quad .5]^T$

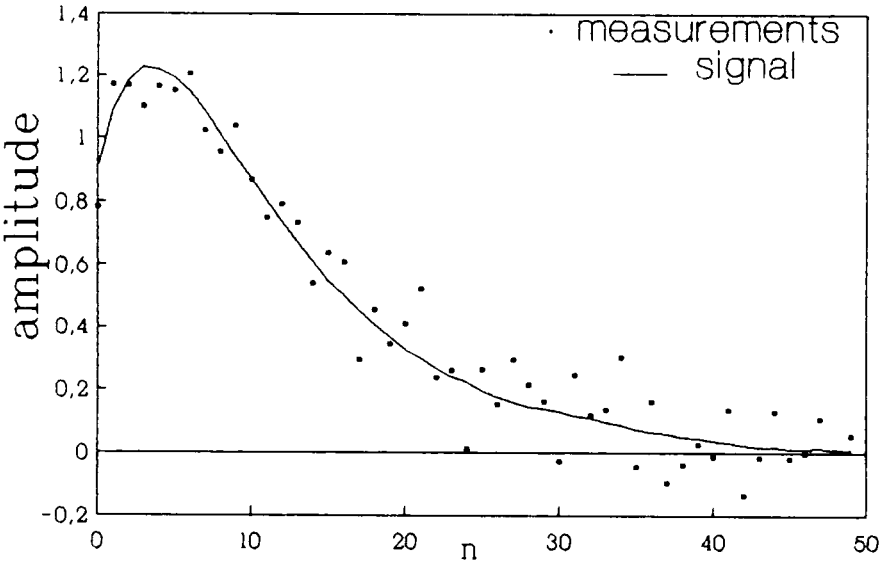


Fig:5.2 :signal and the measurement

Figure 5.2 displays also some data simulated by adding normally distributed errors, $v(n)$, with zero mean and a standard deviation of .1, to the signal $s(n)$. The size of the data samples is taken to be 50 that is (n=1,...,50). Note that from the measurement pattern seen over the 50 data samples two different section are well noticeable. The first section defined by the first 20 data points suggesting that the actual signal

has a certain curvature whereas, the second section which consists of the remaining data points suggest that the actual signal is approximately linear. This remark recommend the use of two different amounts of smoothing since the smoothing parameter represents the rate of exchange between residual error and local variations. Therefore, good estimation of the smoothing parameter, $\lambda_{1\ opt}$, for section one could be obtained by minimizing $V(\lambda)$ where only the first 20 points are used. The obtained value was $\lambda_1 opt = 4.8E-2$, see Fig.5.3. The same procedure is applied to the last 30 points to obtain a good estimation of $\lambda_{2\ opt}$, and the obtained value was $\lambda_2 opt = 5.8E-4$, see Fig.5.4.

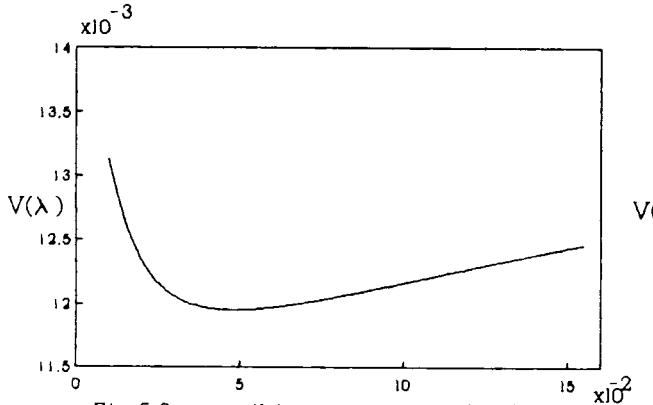


Fig:5.3: smoothing parameter by (GXV)
λ 1 opt=4.8 E-2

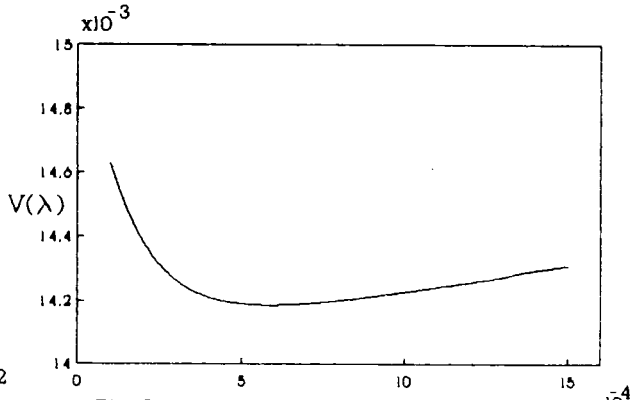


Fig:5.4: smoothing parameter by (GXV)
λ 2 opt=5.8 E-4

These two optimal local smoothing parameter, $\lambda_{1\ opt}$ and $\lambda_{2\ opt}$, obtained by the method described above are fed to the (LSS) through the smoothing matrix M. The estimated signal by (LSS) is obtained by applying Eq(5.2.19) using the obtained M. For the sake of comparison the signal estimated by (SS) is first found by obtaining the optimal global smoothing parameter (OGSP) λ_{opt} which is the minimiser of expression (3.3.2), as shown in Fig.3.3 page 49. Figure 5.5 plots the signal estimated by the (LSS) and (SS) together with the actual and the measurement signals. Note that section one of the signal estimated by (SS) is not able to follow the peak of the actual signal. On the other hand, section two of the (SS) estimate exhibits some oscillations. This comes from the fact the (OGSP) obtained by generalized cross-validation is not suitable in these sections since it is either over smoothed or under smoothed, respectively. However, the signal estimated by the (LSS) follows closely the actual signal. The mean squared error (MSE) is computed for both estimators. The result for the (SS) based on global smoothing parameter and the (LSS) based on local smoothing parameter were 14.7992E-4 and 8.4023E-4, respectively. This confirms the superiority of the (LSS) over the standard (SS).

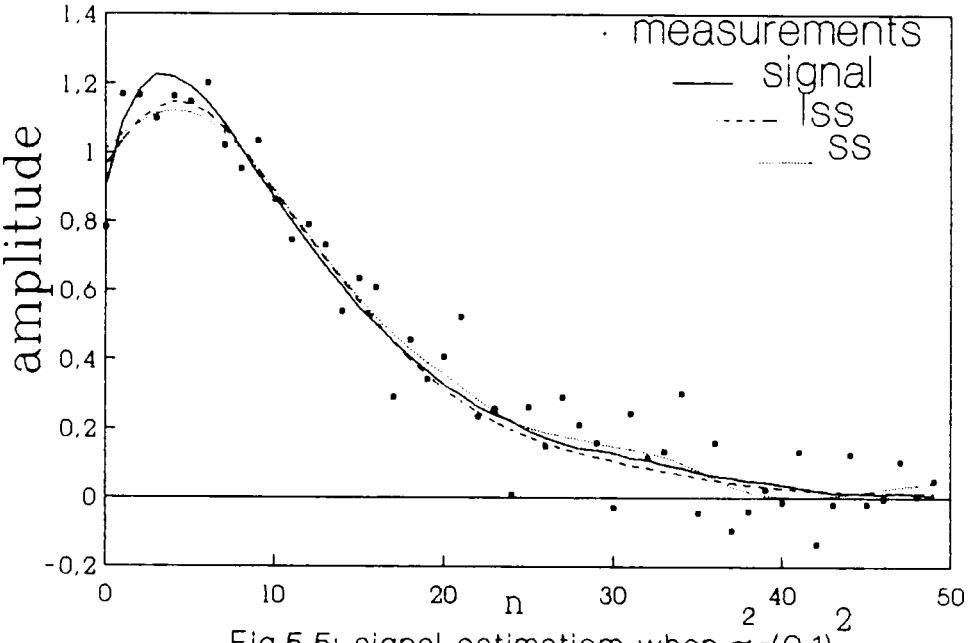


Fig.5.5: signal estimatiom when $\sigma^2 = (0.1)^2$

In order to justify the superiority of the method and the noncruciality of the choice of the cutting off point, two other sectioning points have been identified. The first was at the data point numbered 15 and the second was the data point numbered 25. The mean squared estimation error has been computed for the two estimators using the two different sectioning. The results were 6.0847E-4 and 8.6459E-4, respectively. These results indicate that the choice of the sectioning point is not crucial since for the three cases the (LSS) is better, in the (MSE) sense, than the standard (SS).

In the case where the data pattern of the measurement signal is not noticeable, we recommend to apply the standard smoothing spline directly to obtain an indication of the general pattern of the data so that good choose of the sections could be made and the application of the (LSS) would be feasible.

The algorithm of the local smoothing spline is summarized as follows:

STEP_1: Identify the different section of the measured signal (curvatures as well as linear regions). If necessary apply the standard smoothing spline to identify the different section of the signal.

STEP_2: Choose moderate data sample for each chosen section in step_1 and for each section obtain the optimum smoothing parameter by minimizing the following expression:

$$V(\lambda_i) = \frac{1}{N_i} \frac{\sum^{sn_i} (s_{i\lambda_i}(t_n) - z_i(t_n))^2}{\left(1 - \frac{1}{N_i} trace(A_i(\lambda_i))\right)^2}$$

where

$\sum^{sn}_i$ is the sum over the data samples of the section i.

STEP_3: Form the smoothing matrix M by inserting the optimum smoothing parameters, found for each section, obtained in step_2.

STEP_4: Compute the **Q** and **T** matrices. Then the local smoothing spline estimator evaluated at the design point $\{t_n\}$ is given by:

$$\hat{s}_M = (I + M^{-1} Q T^{-1} Q^T)^{-1} z$$

STEP_5: In order to completely determine the local smoothing spline, the parameters a_n , b_n , c_n , d_n are computed as follows

$$a = \hat{s}_M$$

$$c_1 = c_N = d_1 = d_N = 0 \qquad d_n = \frac{c_{n+1} - c_n}{3h_n} \qquad for \qquad n = 2, \ldots, N-1.$$

$$c = T^{-1} Q^T a$$

$$b_n = \frac{a_{n+1} - a_n}{h_n} - c_n h_n - d_n h_n^2 \qquad for \qquad n = 1, \ldots, N-1.$$

where

$$h_n = t_{n+1} - t_n$$

and

b_n should be initialised, as for the (SS) estimator.

5.4_ CONCLUSION

The development of the (LSS) shows that the (SS) is a special case of the (LSS). It has been shown that the (LSS) is successful, in the sense, that it provides a way through which visual information could be fed to the (LSS) estimator resulting in a good estimation compared to the standard smoothing spline which are based on global smoothing parameter. The adaptation of the smoothing parameter is achieved by the use of the visual information drawn from the measurement data pattern. Other technique could be applied to the local smoothing spline (LSS) by making the local smoothing parameter either dependent on the second derivative which is known as the plug-in method, or found through the bootstrap, which is a resampling technique. The bootstrap technique has been successfully applied to the kernel estimator in the case of nonparametric regression [10,15,16], and density estimation [12,29].

CHAPTER VI:

CONCLUSION

In this research work, two linear smoothers, the smoothing spline (SS) and the modified Kalman smoother (MKS), are compared through simulation study. The choice of which technique is suitable depends on the following concluding points which are meanly drawn from the simulation study.

1_ The (SS) estimator does not force the model into a rigid class and does not need information on the measurement noise variance, the only requirement is that the signal to be estimated is smooth. On the other hand, the (MKS) requires a complete information on the signal model, the output and input noise variance for closely optimal results.

2_ In the case of long data records, the performances of the (MKS) are found to be better than that of the (SS), when the signal parameter, the input and output noise statistics are known. However, in the case of short data records the superiority of the (MKS) over the (SS) decreases as compared

to long data records.

3_ Both the (SS) and the (MKS) do not give good results for a moderate number of data sample recorded with an additive noise of high variance.

4_ Non stationarity of the measurement noise is easily handled by (SS), where the δ_n are chosen to locally controls the smoothing window at time t_n as it has been used by Peyrovian and Sawchuk [23] in the restoration of a blurred image and Silverman [27] in the case of estimating smooth signal observed with an additive non stationary noise. Whereas (MKS) assumes a constant variance of the measurement noise.

5_ The (MKS) is suitable for adaptive processing application.

6_ The (MKS) involves much computations compared to the (SS), specially for signals modelled by a large signal model order. On the other hand, the computational complexity of the (SS) has been found more sensitive to the size of the data records than that of the (MKS).

7_ The (SS) are suitable for fitting empirical functions and for functional analysis [27,36] where estimation of higher order derivative is encountered.

8_ The (SS) is quite close to an ideal low-pass filter which make it resistant to brief spikes, but it is very sensitive to a set of out laying data samples and especially at the boundaries.

9_ The model mismatch has been investigated for both the (MKS) and the (SS), and it was shown through simulation studies that

the performance of both estimators has been degraded under model mismatch. In the case of the (MKS) divergence, a statistical test based on the bootstrap technique has been developed to investigate whether the signal parameter, found through identification, are not correct or the noise assumption is not appropriate.

10_ The computation of the global smoothing parameter by generalized cross-validation restricts the ability of the use of the (SS) to processes whose degree of smoothness may not be stationary. To overcome this problem, a new approach which is the local smoothing spline (LSS) which is based on a local smoothing parameter has been found to be feasible and effective.

11_ Other techniques such as the bootstrap [10,15,16] and the plug-in method [10] could know be used, for finding the local smoothing parameter, for the (LSS).

REFERENCES:

[1] B. D. O. Anderson and J. B. Moore: *Optimum Filtering.* pp.153-162, Englewood Cliffs, NJ. Prentice-Hall, Inc., (1979).

[2] A. J. Berkhout and P. R. Zaanen: A comparison between wiener filtering, kalman filtering, and deterministic least squares estimation. Geophysical Prospecting, **24**, pp.141-197. (1974).

[3] R. Boucheta and M. Djeddi: Parametric and nonparametric estimators: "simulation experiments based comparative study". Advances in Modelling and Analysis, B, vol.**30**, N°2, pp.19-29, (1994).

[4] R. Boucheta and M. Djeddi: Smoothing spline: "local adaptive smoothing parameter". Advances in Modelling and Analysis, A, vol.**20**, N°3, pp.55-64, (1994).

[5] James V. Candy: *Signal Processing: The Model-based Approach.* Mc Craw-Hill,Inc (1986).

[6] James V. Candy: *Signal Processing: The Modern Approach.* Mc Craw-Hill,Inc, pp.220-231, (1988).

[7] R. M. Clark: Non-parametric estimation of a smooth regression function. JRSSB, vol.**39**, pp.107-113, (1977).

[8] D. Cox: Asymptotics for M-type smoothing splines. Ann. statist, vol. **11**, No.2, pp.530-551, (1983).

[9] D. Cox and E. Koh: A smoothing spline based test of model adequacy in polynomial regression. Ann. Inst. Statist. Math.,

vol.**41**, No.2, pp.383-400, (1989).

[10] B. Efron: More efficient bootstrap computations. Journal of Amer Statis Ass, vol **85**, No.409, pp.79-89, (1990).

[11] R. L. Eubank: Approximate regression model and splines. Comm. statist. theory. math., **13**(4), pp.433-484, (1984).

[12] J. J. Faraway and M. Jhun: Bootstrap choise of bandwidth for density estimation. Amer Statist Ass, theory and methods vol. **85**, No 412, pp.1119-1122, (1990).

[13] D. C. Fraser et al: The optimum linear smoother as a combination of two optimum linear filters . IEEE Trans. on AC, Vol-**8**, pp.387-390, (1963).

[14] T. N. E. Greville: Introduction to spline functions: In theory and applications of spline functions. (T.N.E. Greville ed.),pp.1-36, New York: Academic Press, (1969).

[15] P. Hall: Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. Journal of Multivariate Analysis, vol **32**, pp.177-203, (1990).

[16] W. Hardle and A. W. Bowman: Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. Amer Statist Ass, theory and methods vol. **83**, No 401, pp.102-110, (1988).

[17] H. S. Hou and H. C. Andrews: Cubic splines for image interpolation and digital filtering. IEEE trans. on ASSP, vol **26**, No 6, pp.508-517, (1978).

[18] T. Kailath: An innovations approach to least-squares

estimation. Part I: Linear filtering in additive noise. IEEE trans. on Aut. Contr. AC-**18**, pp.601-607, (1973).

[19] T. Kailath and M. R. Gevers: An innovations approach to least-squares estimation. Part VI: Discrite-time innovations representation and recursive estimation, IEEE trans. on Aut. Contr. AC-**18**, pp.601-607, (1973).

[20] R. E. Kalman: A new approach to linear filtering and prediction problems. Trans. ASME, Journ. Basic Eng., vol.**82**, pp.34-45, (1960).

[21] R. E. Kalman and R. S. Bacy: New results in linear filtering and prediction theory. Trans. ASME, Journ. Basic Eng., vol.**83**, pp.95-107, (1961).

[22] F. L. Lewis: *Optimal Estimation*. John Wiley & Sons, Inc.(1986).

[23] M. J. Peyrovian and A. A. Sawchuk: Restoration of noisy blurred images by a smoothing spline filter. Applied Optics vol. **16**, No.12, pp.3147-3153, (1977).

[24] C. H. Reinsch: Smoothing by spline functions. Numer. Math. **10**, pp.177-183, (1967).

[25] C. H. Reinsch: Smoothing by spline functions II. Numer. Math. **16**, pp.451-454, (1971).

[26] B. G. Robert: *Introduction to Random Signal Analysis and Kalman Filtering*. John Wiley and Sons, Inc. (1983).

[27] B. W. Silverman: Some aspects of the spline smoothing approach to non-parametric regression curve fitting. J. R. statist, soc. B, **47**, No 1, pp.1-52, (1985).

[28] P. L. Smith: Splines as a useful and convenient statistical tool. The American Statistician, vol. **33**, No.2, pp.57-62, (1979).

[29] C. C. Taylor: Bootstrap choice of the smoothing parameter in kernel density estimation. Biometrika, **76**, No.4, pp.705-712, (1989).

[30] S. A. Tretter:*Introduction to Discrite-Time Signal Processing*. pp.371-387, John Wiley York, (1976).

[31] D. F. Uteras: Sur le choix du parametre d´ajustement dans le lissage par fonctions spline. Numer. Math. , **34**, pp.15-28, (1980).

[32] J. Richard Vaccaro and Li Fu: Signal estimation over short data records: A data-dependent time-invariant algorithm. Journal of the Franklin Institute Vol.**327**, No.3, pp.439-455, (1990).

[33] D. R. Vaughan: A nonrecursive algebraic solution for the discrete reccati equation. IEEE Trans. Automatic Control, Vol.AC-**15**, No.5, pp.597-599, (1970).

[34] G. Wahba and P. Craven: Smoothing noisy data with spline functions. Numer. Math. , **31**,pp.377-403, (1979).

[35] G. Wahba and S. Wold: A completely automatic frensh curve: fitting spline functions by cross-validation. comm. statist., **4**,pp.1-17, (1975).

[36] S. Wold: Spline functions in data analysis. Technomet-

rics. **16**, pp.1-11, (1974).

[37] C. H. Woodford: An algorithm for data smoothing using splines functions. BIT, **10**, pp.501-510, (1970).

A N N E X E

Liste et composition du jury en vue de la soutenance du Memoire de Magister en.....Ingineerie.des.Systemes.Electroniques.....

de/)/)r..BOUCHETTA..Rachid.....

PRESIDENT /: Dr.HOUACINE, Maitre de Conference U.S.T.H.B

RAPPORTEUR/: Dr.M.DJEDDI, Chargé de Recherche, I.N.E.L.E.C.

MEMBRES /: Dr.A.KEZOUH, Maitre de Conference , U.S.T.H.B

Dr.K.HARICHE, Chargé de Recherche, I.N.E.L.E.C