USCIPI Report 580

UNIVERSITY OF SOUTHERN CALIFORNIA

AN ALGORITHMIC IMAGE ESTIMATION METHOD APPLICABLE TO NONLINEAR OBSERVATIONS

by

Manouher Naraghi

June 1975

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

Sponsored by
Advanced Research Projects Agency
Contract No. F08606-72-C-0008
ARPA Order No. 1706

IPI

IMAGE PROCESSING INSTITUTE

AN ALGORITHMIC IMAGE ESTIMATION METHOD APPLICABLE TO NONLINEAR OBSERVATIONS

by

Manouher Naraghi

June 1975

Image Processing Institute
University of Southern California
University Park
Los Angeles, California 90007

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Eastern Test Range under Contract No. F08606-72-C-0008, ARPA Order No. 1706.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Image Processing Institute
University of Southern California, University Park
Los Angeles, California 90007

2a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

2b. GROUP

3. REPORT TITLE
AN ALGORITHMIC IMAGE ESTIMATION METHOD APPLICABLE TO NONLINEAR OBSERVATIONS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report, June 1975

5. AUTHOR(S) (First name, middle initial, last name)
Manouher Naraghi

6. REPORT DATE
June 1975

7a. TOTAL NO. OF PAGES
112

7b. NO. OF REFS
42

8a. CONTRACT OR GRANT NO.
F08606-7?-C-0008

b. PROJECT NO.
ARPA Order [redacted]-1706

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
USCIPI [redacted]-580

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
Approved for release: distribution unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

13. ABSTRACT

The principle of image estimation in the presence of linear and nonlinear observations is considered in this dissertation and a recursive estimation algorithm is developed. The development proceeds from the assumptions that the image is statistically characterized by its first two moments namely the mean and the autocorrelation while the observation is allowed to be a general function of the signal and noise. A two step recursive estimation procedure, compatible with the logical structure of the optimal minimum mean square estimator, is developed. The procedure consists of a linear one step prediction and a filtering operation.

In order to derive the linear predictor, the a priori mean and autocorrelation information is employed to obtain a linear finite order model of the two dimensional random process. This model is of an autoregressive form whose derivation requires only the numerical values of the mean and the correlation functions. At each step of the estimation, the autoregressive model is used in finding the best linear predicted value and its error variance as a function of past estimates and their error variances. Following the prediction process, the filtering operation proceeds to evaluate the estimate and its error variance as a function of the predicted value and the observation.

****

14. Key Words: Digital Image Processing, Image Estimation, Linear Estimation, Recursive Estimators, Minimum Mean Square.

391141

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

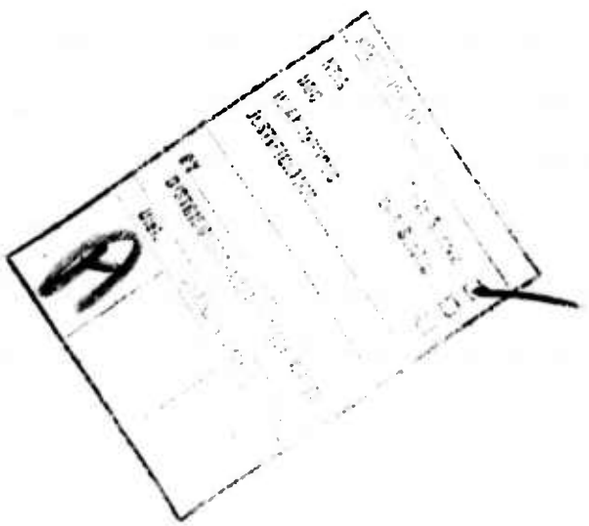
TO MARLENE

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor N. E. Nahi, Chairman of my Dissertation Committee, for the valuable guidance, criticism and countless number of discussions that he provided me in the course of this research. I am also indebted to him for the financial assistance which has made my graduate study possible.

I thank Professors R. S. Kashef and J. Pierce for the encouragement they have given me in my studies and their extensive constructive criticism of this work. My gratitude also goes to Professor W. K. Pratt and the staff of the Image Processing Institute for the support they have given me throughout this research.

I would like to thank my wife Marlene for her constant encouragement and understanding. I also thank her for the editorial assistance and typing of this dissertation.



ABSTRACT

The principle of image estimation in the presence of linear and nonlinear observations is considered in this dissertation and a recursive estimation algorithm is developed. The development proceeds from the assumptions that the image is statistically characterized by its first two moments namely the mean and the autocorrelation while the observation is allowed to be a general function of the signal and noise. A two step recursive estimation procedure, compatible with the logical structure of the optimal minimum mean square estimator, is developed. The procedure consists of a linear one step prediction and a filtering operation.

In order to derive the linear predictor, the a priori mean and autocorrelation information is employed to obtain a linear finite order model of the two dimensional random process. This model is of an autoregressive form whose derivation requires only the numerical values of the mean and the correlation functions. At each step of the estimation, the autoregressive model is used in finding the best linear predicted value and its error variance as a function of past estimates and their error variances. Following the prediction process, the filtering operation proceeds to evaluate the estimate and its error variance

as a function of the predicted value and the observation.

The estimation method is applied to a number of one and two dimensional problems and the appropriate estimators are developed for the cases where the observation contains additive and/or multiplicative noise term(s). The performance of the method is evaluated by applying the estimation procedure to two dimensional pictorial data corrupted by additive-Gaussian and multiplicative uniform noise.

The value of the method has been analyzed and discussed as to its application to practical problems and its optimality as an estimation technique.

CHAPTER 1

INTRODUCTION

Progress in the sophistication and the computational capability of digital computers has opened a field of research in the applied sciences dealing with the characterization, understanding and analysis of pictorial data. This field of image processing encompasses a variety of areas of study such as coding, recognition, enhancement, restoration, estimation, data compression and many more. A particular subject of interest among these is that of image estimation. This subject deals with the restoration of images containing degradations where only some statistical properties of both the image and the degrading phenomenon are known. In this respect, a picture is generally viewed as a two dimensional random process (field) [1], [6] and often characterized by its first two moments, namely the mean and the autocorrelation. Denoting the brightness function of the discrete image by b(i,j), with i and j as the row and column counters, the two moments are defined as

$$M(i,j) = Eb(i,j) \tag{1.1}$$

and

$$R(i,j,k,l) = E[b(i,j)-M(i,j)][b(k,l)-M(k,l)] \qquad (1.2)$$

where E is the mathematical expectation operator.

The list of degradation (noise) introducing sources in imaging systems is extensive and in particular includes inaccuracies in the sensing devices, the existence of air turbulence or cloud layer between the camera system and the scene, reflections from other objects in the scene, uncertainties in the transmission systems and film grain noise. The degraded image (observation), denoted by $y(i,j)$, specifies the functional relationship of the signal, $b(i,j)$ and the noise $\gamma(i,j)$. Symbolically

$$y(i,j) = f(b(i,j),\gamma(i,j)) \qquad (1.3)$$

where f may be nonlinear and $\gamma(i,j)$ may be vector valued (i.e. more than one noise term).

The values of $M(i,j)$, $y(i,j)$ and $R(i,j,k,l)$, for all i,j,k,l, the functional form of f and the density function of $\gamma(i,j)$ in (1.3) constitute the a priori information. This constitutes the total amount of information that the estimation procedure is to use in obtaining the improved image. An estimation procedure is the process of assigning a value to an unknown parameter based on the noise corrupted observations involving some function of the parameter. The assigned value is called the estimate and the system yielding the estimate is called the estimator. The

assignment of the estimate values, in general, is based on certain criterion known as the estimation criterion. One such criterion is that of minimizing the mean square error.

Optimum filtering of images under the general condition of (1.3) has received little attention, while a variety of procedures have been developed for the special linear case, where

$$y(i,j) = b(i,j)+\gamma(i,j) \tag{1.4}$$

with $\gamma(i,j)$ white and Gaussian [32]-[37]. Although (1.4) describes many natural forms of degradations [32]-[37], there are conceivably as many situations where this model does not apply. Examples of the above are the film grain noise and the taking of pictures through a nonhomogeneous layer of clouds, where the noise is a random attenuation factor. Hence the observation takes the form

$$y(i,j) = \gamma(i,j)b(i,j) \tag{1.5}$$

The majority of these linear estimation techniques require a rather specific analytical representation of the correlation function $R(i,j,k,l)$ and in order for their underlying estimators to become computationally efficient, the signal and the noise processes are required to be wide sense stationary. Due to these requirements, the above methods may not be suitable for certain practical problems.

Examples of which are the cases where the correlation function is specified numerically and/or partially (the correlation function is represented partially if $R(i,j,k,l)$ is defined only for $|i-k| \leq p$ and $|j-l| \leq q$, for some integers p and q). A property, though, of these procedures, which has great intuitive appeal and is crucial to the real time implementation of estimators, is the recursive nature of their underlying algorithms. This property arises from Kalman-Bucy [23]-[24] estimation theory reviewed briefly in Section 1.2. Section 1.3 contains a review of one dimensional nonlinear estimation and the extended Kalman-Bucy filtering methods. These techniques, as will be pointed out, deal with certain nonlinearities in estimation problems.

In this dissertation, a general estimation method will be developed having the following characteristics:

1. The method will be applicable to two as well as one dimensional estimation problems.
2. The estimation algorithm will only require specification of the numerical values of $R(i,j,k,l)$.
3. The procedure will be applicable to problems where only partial representation of the correlation function is available.
4. The method will be applicable to general linear and nonlinear observation systems of (1.3).

5. The procedure will be implementable; i.e. numerical in nature and computationally feasible.

Section 1.1 contains the definition of some notations and the description of a convention which unifies the one and two dimensional indexing. A brief review of one dimensional estimation techniques is presented in Sections 1.2 and 1.3.

The estimation method is derived in Chapters 2 through 5. In Chapter 2, the structure of the estimator is developed. It is shown that the general estimation technique consists of modeling, linear prediction and filtering steps.

The modeling problem is considered in Chapter 3. A general procedure is introduced which utilizes the a priori statistics and derives a linear autoregressive model of the process. Chapter 4 and 5 contain pertinent derivations of the linear one step predictor and the filtering steps, respectively.

The estimation process, as developed in Chapter 2 through 5, is applied to a number of linear and nonlinear problems in Chapter 6. This chapter also includes the derivation of the estimator for a few special cases. Included among these are the cases of additive-Gaussian and multiplicative uniform observation noise. In Chapter 7, the proposed estimation process is analyzed as to its

computational requirements. Discussion on the optimality of the estimator is presented in Chapter 8. Extensions, topics for further research and conclusions are also included in this chapter.

Appendix A contains a brief discussion on the error variance. Appendix B describes a fidelity measure for estimated images.

1.1 Notations

An image is viewed as an NxN matrix with elements b(i,j), where b(i,j) is the intensity of the image at pixel (i,j). To reduce the notational complexity, the pixels are indexed by $1,2,\ldots,N,N+1,\ldots,N^2$ consecutively from left to right and from top to bottom. This convention enables us to refer to the doubly indexed b(i,j) as b(k), symbolically. Hence (1.1) through (1.3) can be written as:

$$M(k) = E[b(k)] \tag{1.6}$$

$$R(k,l) = E[b(k)-M(k)][b(l)-M(l)] \tag{1.7}$$

$$y(k) = f(b(k),\gamma(k)) \tag{1.8}$$

Let us define the process x(k) as

$$x(k) = b(k)-M(k) \qquad k=1,2,\ldots,N^2 \tag{1.9}$$

Thus the problem of estimating b(k) reduces to estimating x(k).

1.2 A Survey of Discrete One Dimensional Estimation

At a given time k and for a given set of observations y(1),....,y(k), the minimum mean square (MMS) estimate of a random process x(k) is, by definition, the particular value of $x^e(k)$ which minimizes the quantity $e^2(k)$ defined as [see Appendix A]

$$e^2(k) = E[x(k)-x^e(k)]^2 | y(1),....,y(k) \tag{1.10}$$

Let us denote this quantity by $x^\sigma(k)$. Direct minimization of $e^2(k)$ with respect to $x^e(k)$ yields [22]

$$x^\sigma(k) = Ex(k) | y(1),....,y(k) \tag{1.11}$$

This is a general result, in that, regardless of the underlying probability density functions (PDF) of x(.) and y(.), the MMS estimate is given by (1.10).

When x(k) is a normal random process and processes x(.)and y(.) are jointly normal [22], then $x^\sigma(k)$ in (1.10) will be linear in y(1),....,y(k), having the form

$$x^\sigma(k) = \sum_{i=1}^{k} \alpha_i(k) y(i) \tag{1.12}$$

To determine the constants $\alpha_1(k),\ldots,\alpha_k(k)$, the right hand side of (1.11) is substituted in (1.9) and $e^2(k)$ is minimized with respect to $\alpha_1(k),\ldots,\alpha_k(k)$, resulting in k linear equations

$$E[x(k)-x^{\sigma}(k)]y(i) = 0 \qquad i=1,2,\ldots,k \tag{1.13}$$

collectively referred to as the orthogonality principle. This procedure of finding $x^{\sigma}(k)$, though clear and simple, is numerically inefficient since for each time k,a system of linear equations has to be solved where the size of the system of equation grows with k.

Kalman and Bucy [23]-[24] have shown that if the process x(k) can be generated by applying white noise to the input of a finite dimensional linear dynamical system, then the estimation process will be recursive yielding an implementable and computationally simple estimator. This is done if there exist a vector Z(k) such that

$$x(k) = C(k)Z(k) \tag{1.14}$$

with Z(k) satisfying a linear stochastic difference equation

$$Z(k+1) = A(k)Z(k)+B(k)u(k) \tag{1.15}$$

and the observation having the form

$$y(k) = C(k)Z(k)+D(k)v(k) \tag{1.16}$$

with

$$Eu(k) = Ev(k) = 0 \tag{1.17a}$$

$$Eu(i)u'(j) = K(i)\Delta(i-j) \tag{1.17b}$$

$$Ev(i)v'(j) = L(i)\Delta(i-j) \tag{1.17c}$$

where A(k), B(k), C(k), D(k), K(k), L(k) are nxn, nxr, sxn, sxq, rxr and qxq matrices, respectively. The term $\Delta(i-j)$ is the Kronecker delta function and the prime indicates matrix transposition.

The estimate of x(k) can be obtained from the estimate of Z(k) through (1.14). Denoting the MMS one step prediction value of Z(k) by $Z^{\sigma}(k)$, then

$$Z^{\sigma}(k) = EZ(k)|y(1),.....,y(k-1) \tag{1.18}$$

and the Kalman-Bucy linear estimator is given by [22]-[24]

$$Z^{\sigma}(k+1) = [A(k)-F(k)C(k)]Z^{\sigma}(k)+F(k)y(k) \tag{1.19}$$

where

$$F(k) = A(k)P(k)C'(k)[C(k)P(k)C'(k)+D(k)L(k)D'(k)]^{-1} \tag{1.20}$$

and

$$P(k+1) = [A(k)-F(k)C(k)]P(k)[A(k)-F(k)C(k)]' + B(k)K(k)B'(k)+F(k)D(k)L(k)D'(k)F'(k) \tag{1.21}$$

The matrix P(k+1) has the property that for each k

$$P(k+1) = E[Z(k+1)-Z^{\sigma}(k+1)][Z(k+1)-Z^{\sigma}(k+1)]' \qquad (1.22)$$

The above results have been presented in a summarized form (for more detail see [20], [22]) and are included in order to point out the recursive nature of the solution in (1.19) through (1.21). This property is conspicuous in (1.19) where the estimate at time k+1, $Z^{\sigma}(k+1)$, is only a function of the estimate at time k, $Z^{\sigma}(k)$ and the observation at time k, y(k). It is this attribute that makes the Kalman-Bucy linear estimator easily implementable on digital computers.

1.3 Nonlinear Estimation and Extended Kalman Filtering

The majority of the existing nonlinear estimation techniques are concerned with problems where the system and observation models (equations (1.14) and (1.15), respectively) are given as [22], [29]

$$Z(k+1) = f[Z(k),k]+B(k)u(k) \qquad (1.23)$$

$$y(k) = g[Z(k),k]+D(k)v(k) \qquad (1.24)$$

where f(.) and g(.) are general nonlinear functions.

An implementable nonlinear estimation approach, which uses linearization in obtaining a suitable procedure to estimate the states of the nonlinear system of (1.23), is

that of the extended Kalman method [22]. In this technique, relationships are obtained which describe the behavior of (1.23) and (1.24) in the vicinity of a nominal solution $Z^*(k)$. The dynamics of the difference $Z(k)-Z^*(k)$ is characterized by a set of linear equations. This characterization is achieved by assuming that $f(Z(k),k)$ and $g(Z(k),k)$ are twice differentiable in $Z(k)$ and defining the matrices

$$A(k) \triangleq \left[\begin{array}{ccc} \frac{\partial f_1}{\partial Z_1} & \frac{\partial f_1}{\partial Z_2} \cdots & \frac{\partial f_1}{\partial Z_n} \\ \frac{\partial f_2}{\partial Z_1} & \frac{\partial f_2}{\partial Z_2} & \\ \vdots & & \\ \frac{\partial f_n}{\partial Z_1} \cdots & & \frac{\partial f_n}{\partial Z_n} \end{array}\right]_{Z(k)=Z^*(k)} \quad (1.25)$$

$$C(k) \triangleq \left[\begin{array}{ccc} \frac{\partial g_1}{\partial Z_1} & \frac{\partial g_1}{\partial Z_2} \cdots & \frac{\partial g_1}{\partial Z_n} \\ \frac{\partial g_2}{\partial Z_1} & \frac{\partial g_2}{\partial Z_2} & \\ \vdots & & \\ \frac{\partial g_n}{\partial Z_1} \cdots & & \frac{\partial g_n}{\partial Z_n} \end{array}\right]_{Z(k)=Z^*(k)} \quad (1.26)$$

where A(k) and C(k) are used as coefficient matrices of a linearized representation of (1.23) and (1.24) in the neighborhood of $z^*(k)$.

It is shown in [22] that the application of the Kalman Bucy estimation technique along with the proper choice of $Z^*(k)$, results in a recursive nonlinear estimator of the form

$$\hat{z}(k+1) = f[\hat{z}(k),k]+F(k)[y(k)-g[\hat{z}(k),k]] \qquad (1.27)$$

where F(k) satisfies (1.20) and (1.21) with matrices A(k) and C(k) defined by (1.25) and (1.26).

Aside from the extended Kalman-Bucy technique, there are other methods that consider the models of (1.23) and (1.24) [28], [29]. These procedures, however, lack the ease of implementation inherent in the linear Kalman-Bucy and the extended Kalman-Bucy techniques.

CHAPTER 2

ESTIMATION METHOD

In this chapter, the structure of the general nonlinear estimator will be developed. Section 2.1 contains the derivation of the optimal minimum mean square estimator. Based on the properties of this estimator, the structure of a general implementable estimation procedure is developed in Section 2.2.

2.1 The Minimum Mean Square Estimation

For a given set of observation $y(1),....,y(k)$, the minimum mean square (MMS) estimate, $x^{\sigma}(k)$, of a process $x(.)$ at time k is obtained from [22]

$$x^{\sigma}(k) = Ex(k)|y(1),....,y(k) \tag{2.1}$$

Similarly the error variance of this estimate, $\sigma^{\sigma^2}(k)$, is defined as

$$\sigma^{\sigma^2}(k) = E[x(k)-x^{\sigma}(k)]^2|y(1),....,y(k) \tag{2.2}$$

Defining the set $Y(k)$ as

$$Y(k) = \{y(1),....,y(k-1)\} \tag{2.3}$$

the equivalent form of (2.1) is

$$x^{\sigma}(k) = \int x(k)\, p(x(k) \mid Y(k), y(k))\, dx(k) \tag{2.4}$$

with p(.) designating the probability density. By applying Baye's rule,

$$p(x(k) \mid Y(k), y(k)) = \frac{p(x(k), Y(k), y(k))}{p(Y(k), y(k))}$$

$$= \frac{p(y(k) \mid x(k), Y(k))\, p(x(k) \mid Y(k))}{p(y(k) \mid Y(k))} \tag{2.5}$$

Since y(k) is defined as only a function of x(k) and $\gamma(k)$ (equation (1.7)), where $\gamma(k)$ is independent of Y(k), then

$$p(y(k) \mid x(k), Y(k)) = p(y(k) \mid x(k)) \tag{2.6}$$

This simplifies (2.5) to

$$p(x(k) \mid Y(k), y(k)) = \frac{p(y(k) \mid x(k))\, p(x(k) \mid Y(k))}{p(y(k) \mid Y(k))} \tag{2.7}$$

The substitution of the above in (2.4) yields

$$x^{\sigma}(k) = \frac{1}{p(y(k) \mid Y(k))} \int x(k)\, p(y(k) \mid x(k))\, p(x(k) \mid Y(k))\, dx(k) \tag{2.8}$$

Furthermore

$$p(y(k)|Y(k))=\frac{p(y(k),Y(k))}{p(Y(k))}=\frac{\int p(y(k),Y(k),x(k))dx(k)}{p(Y(k))}$$

$$=\frac{1}{p(Y(k))}\int p(y(k)|x(k),Y(k))p(x(k)|Y(k))p(Y(k))dx(k)$$

$$=\int p(y(k)|x(k))p(x(k)|Y(k))dx(k) \qquad (2.9)$$

where again (2.6) has been used to obtain (2.9). Using (2.9) in (2.8) yields

$$x^{\sigma}(k)=\frac{\int x(k)p(y(k)|x(k))p(x(k)|Y(k))dx(k)}{\int p(y(k)|x(k))p(x(k)|Y(k))dx(k)} \qquad (2.10)$$

Similarly the error variance of (2.2) is given by

$$\sigma^{\sigma^2}(k)=\int [x(k)-x^{\sigma}(k)]^2 p(x(k)|Y(k),y(k))dx(k) \qquad (2.11)$$

Substituting (2.7) in the above, results in

$$\sigma^{\sigma^2}(k)=\frac{\int [x(k)-x^{\sigma}(k)]^2 p(y(k)|x(k))p(x(k)|Y(k))dx(k)}{p(y(k)|Y(k))} \qquad (2.12)$$

Finally the substitution of (2.9) in (2.12) yields the form of $\sigma^{\sigma^2}(k)$ to be

$$\sigma^{\sigma^2}(k)=\frac{\int [x(k)-x^{\sigma}(k)]^2 p(y(k)|x(k))p(x(k)|Y(k))dx(k)}{\int p(y(k)|x(k))p(x(k)|Y(k))dx(k)} \qquad (2.13)$$

Equations (2.10) and (2.13) suggest that the optimal estimation, at time k, is achieved first by finding $p(x(k)|Y(k))$ and then using it with $y(k)$ to arrive at $x^{\sigma}(k)$ and $\sigma^{\sigma^2}(k)$. Letting $x_p(k)$ and $\sigma_p^2(k)$ denote the optimal prediction value and its error variance of $x(k)$, respectively, then (see (1.17)

$$x_p(k) = Ex(k)|y(1),....,y(k-1) = Ex(k)|Y(k) \qquad (2.14)$$

$$\sigma_p^2(k) = E[x(k)-x_p(k)]^2|Y(k) \qquad (2.15)$$

But $x_p(k)$ and $\sigma_p^2(k)$ are the mean and the variance, respectively, of $p(x(k)|Y(k))$ in (2.10) and (2.13). Therefore, the optimal estimation at time k can be thought of as a two step procedure depicted in Fig. 2.1, where the blocks P and F may be identified as the prediction and filtering steps, respectively.

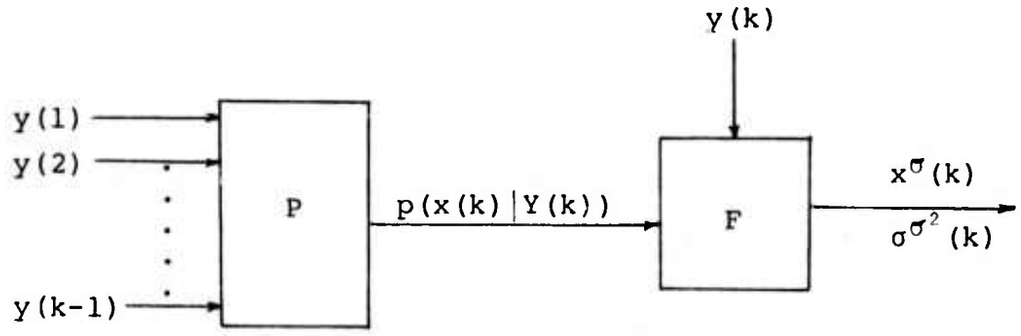


Fig. 2.1

In this system structure, $y(k)$ is isolated from the other random variables and if $p(x(k)|Y(k))$ is known, conceptually one can deal with its nonlinearities in block F. So if $p(x(k)|Y(k))$ is given, then the derivation of $x^{\sigma}(k)$ and $\sigma^{\sigma^2}(k)$ is accomplished by carrying out the integrations in (2.19) and (2.13). However, the derivation of this probability density for the general observation system of (1.13) does not lend itself to analytic methods and in addition, available numerical approaches are computationally unfeasible [22].

In the following section, an alternate procedure is considered, whereby an approximation to the probability density $p(x(k)|Y(k))$ is derived. The method is compatible with the logic of the optimal estimator in Fig. 2.1, in that this logic consists of the representation of past information (i.e. information due to a priori statistics and observations $y(1),\ldots,y(k-1)$) in the form of a probability density to be combined with present information (i.e. $Y(k)$) in block F.

2.2 Definition of the Proposed Estimator

In order to comply with the inherent logical structure of the optimum estimator and at the same time maintain the algorithmic implementability, the proposed estimator

is constructed according to the following restrictions.

a) Only the first two moments of any random variable are computed.

b) The prediction process is chosen to be linear.

c) The prediction is to be based on a selected small number of past estimates. This will impose a desired limited memory requirement for the estimator.

In the estimation process, the imposition of constraint (a) on the estimator results in the derivation of the estimate value and its error variance at each time k. The value of the variance represents a measure of uncertainty of the estimate's numerical value. This constraint alleviates the problem of deriving or approximating the probability density associated with the estimate. Due to the admissibility of general probability densities for the observation noise, the lack of this constraint will require estimation approaches similar to those described in [22, Chap. 7], which as mentioned before, are computationally unfeasible.

Since the prediction process is primarily a learning procedure based on the past information, then the linearity requirement of condition (b) does not violate the underlying logic of the optimal estimator. Although this requirement, in general, results in suboptimal processing,

it enhances the implementability of the overall procedure. This enhancement is due to the property of linear predictors being easily implementable.

Having chosen a specific form of the predictor, condition (c) requires the estimator to be recursive. This characteristic is much desired in estimation processes since it simplifies the implementation of the estimator. Basing the learning process (prediction) on the past estimates is justified since each estimate is obtained so that it represents the actual signal value with the least amount of uncertainty.

Letting the estimate and its error variance at time i be represented by $\hat{x}(i)$ and $\hat{\sigma}^2(i)$, respectively, then the block diagram of Fig. 2.2 represents the structure of the estimator.

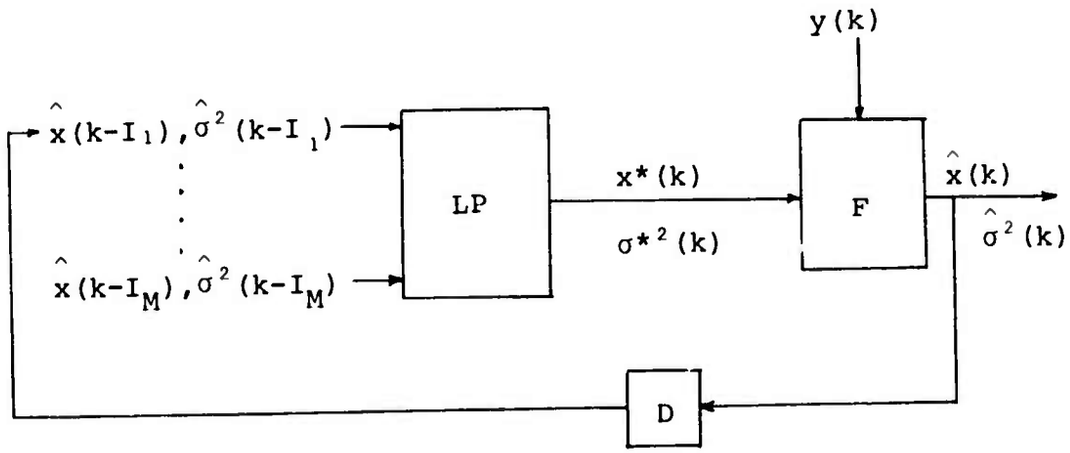


Fig. 2.2

In this figure, blocks LP, F and D signify linear prediction, filtering and one unit time delay operations, respectively. M is an indication of the size of the required memory and $x^*(k)$ and $\sigma^{*2}(k)$ represent the one step predicted value and its error variance, respectively. The set $\{k-I_1,\ldots,k-I_M\}$ is a set of two dimensional indices (time) each distinct and prior to k, i.e.

$$k-I_i \in \{1,2,\ldots,k-1\} \qquad i=1,2,\ldots,M$$

Figure 2.2 describes the structure of the proposed estimator whose operating logic is derived in Chapters 3, 4 and 5. In Chapter 3, a method is introduced to derive a linear M^{th} order model of the process x(k). This model is used in Chapter 4 to derive the desired linear predictor. Chapter 5 will describe the derivation of the filtering step.

CHAPTER 3

MODELING

In order to derive the linear predictor (block LP of Fig. 2.2), the a priori mean and the autocorrelation information of the random process x(k) are incorporated into a finite order linear model of the form

$$x(k) = \sum_{i=1}^{M} \beta_i x(k-I_i) + Bu(k) \qquad (3.1)$$

This model is used in Chapter 4 to complete the derivation of the prediction process. Since (3.1) is an autoregressive model [15], [16], [17], then finding such a model, in effect, solves the one step linear prediction problem in the degenerate case where the values of a sample function of x(i) is specified for i=1,2,....,k-1.

Section 3.1 contains a brief discussion of the form and the properties of the autoregressive models. In Section 3.2, a procedure is introduced that finds the autoregressive model associated with a given autocorrelation function. The derivation of the procedure is based on the a priori knowledge of the maximum allowable order of the model. The discussions and guidelines regarding the best choice of the maximum order is presented in Section

3.3. Finally, the properties of the modeling procedure are considered in Section 3.4.

3.1 Autoregressive Models

A discrete random process x(k) is represented by an autoregressive model of order M(k) if for each time k, x(k) satisfies a linear stochastic difference equation of the form [15]

$$x(k) = \sum_{i=1}^{M(k)} \beta_i(k)x(k-I_i)+B(k)u(k) \tag{3.2}$$

where $M(k)$, $\beta_1(k),\ldots,\beta_{M(k)}(k)$ are deterministic time constants and $u(1)$, $u(2),\ldots$ are a set of independent, identically distributed random variates with

$$Eu(i)u(j) = \Delta(i-j) = \begin{cases} 0 & \text{if } i\neq j \\ 1 & \text{if } i=j \end{cases} \tag{3.3}$$

The indices $k-I_i$, $i=1,2,\ldots,M(k)$ refer to pixels (time) previous to k, i.e.

$$k-I_i \in \{k-1,k-2,\ldots,1\} \quad \text{for all } i=1,2,\ldots,M(k)$$

For a zero mean random process x(k), when constants $M(k)$, $\beta_1(k),\ldots,\beta_{M(k)}(k)$ and B(k) are time invariant (i.e. independent of k) then x(k) will be wide sense stationary and (3.1) takes the form

$$x(k) = \sum_{i=1}^{M} \beta_i x(k-I_i) + Bu(k) \qquad (3.4)$$

and M will be called the order of the model.

3.2 Modeling Procedure

In the following derivations the process x(k) is assumed to be stationary. The extension of the procedure to modeling of nonstationary processes is presented in Section 3.4.

Let M be the given maximum order of the model and let S denote the set of all indices preceding k, so that

$$S = \{k-1, k-2, \ldots, 1\}$$

There are $2^{k-1}-1$ nonempty subsets of S. Let s_1, s_2, ... denote those subsets of S having M elements or less with $N(s_j)$ designating the number of elements in subset s_j. Note that $N(s_j) \leq M$ for all j For each s_j, let

$$x(k) = \sum_{i=1}^{N(s_j)} \beta_i^j x(k-I_i^j) + B_j u(k) \qquad (3.5)$$

where for $i=1,2,\ldots,N(s_j)$, the indices $k-I_i^j$ denote all elements of subset s_j.

The modeling criterion is chosen to be the minimiza-

tion of $E[B_j u(k)]^2$, thus the constants $\beta_1^j,\ldots,\beta_{N(s_j)}^j$ in (3.5) are chosen such that

$$E[B_j u(k)]^2 = E[x(k)-\sum_{i=1}^{N(s_j)} \beta_i^j x(k-I_i^j)]^2 \qquad (3.6)$$

is minimized. This criterion is the same as the minimizing the error variance of the one step predicted value since, $E[B_j u(k)]^2$ is the error variance associated with choosing the predicted value of x(k) to be

$$\sum_{i=1}^{N(s_j)} \beta_i^j x(k-I_i^j)$$

Equation (3.6) is minimized by differentiating its right hand side with respect to $\beta_1^j,\ldots,\beta_{N(s_j)}^j$ and setting the result equal to zero. This results in $N(s_j)$ linear equations of the form

$$E[x(k)-\sum_{i=1}^{N(s_j)} \beta_i^j x(k-I_i^j)]x(k-I_1^j) = 0 \quad 1=1,2,\ldots,N(s_j) \qquad (3.7)$$

Carrying the expectation through in (3.7), a system of linear equations of the form $A\beta=b$ is obtained where, the elements of matrix A and vector b will be in terms of the values of the correlation function R(m,n) (see example 3.1). Solving this system of equations defines the quantities β_i^j, $i=1,2,\ldots,N(s_j)$. These quantities, in turn,

define the values of $(B_j)^2$ given as

$$(B_j)^2 = E[B_j u(k)]^2 = E[x(k) - \sum_{i=1}^{N(s_j)} \beta_i^j x(k-I_i^j)]^2 \qquad (3.8)$$

Repeating this procedure for all subsets s_1, s_2,....., the quantities

$$(B_1)^2,\ (B_2)^2,\ldots..,\ (B_m)^2,.... \qquad (3.9)$$

are obtained. The model of the random process x(k) is chosen to be the autoregressive form associated with subset s_m such that $(B_m)^2$ is the minimum of the quantities in (3.9). If the minimum is not unique, then the model is chosen to be that of s_m where $(B_m)^2$ is a minimum of (3.9) and $N(s_m) \leq N(s_p)$ if $(B_p)^2$ is any other minimum. The following example is provided to clarify some of the above derivations.

Example 3.1:

Letting k correspond to two dimensional index (i,j) and

$$s_j = \{(i,j-1),(i-1,j),(i-1,j-1)\}$$

(In figure 3.1, O represents (i,j) and □ represents elements of s_j on a two dimensional grid) then (3.5) yields

$$x(i,j)=\beta_1^j x(i,j-1)+\beta_2^j x(i-1,j)+\beta_3^j x(i-1,j-1)+B_j u(i,j) \qquad (3.10)$$

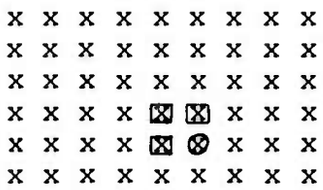


Fig. 3.1

Applying (3.7) and assuming (stationarity)

$$Ex(i,j)x(k,l) = R(|i-k|,|j-l|) \tag{3.11}$$

The system of linear equations for β_1^j, β_2^j, β_3^j will become,

$$\begin{bmatrix} R(0,0) & R(1,1) & R(1,0) \\ R(1,1) & R(0,0) & R(0,1) \\ R(1,0) & R(0,1) & R(0,0) \end{bmatrix} \begin{bmatrix} \beta_1^j \\ \beta_2^j \\ \beta_3^j \end{bmatrix} = \begin{bmatrix} R(0,1) \\ R(1,0) \\ R(1,1) \end{bmatrix} \tag{3.12}$$

Solving (3.12) results in the values for β_1^j, β_2^j, β_3^j and these values result in

$$\begin{aligned} (B_j)^2 &= E[B_j u(i,j)]^2 \\ &= E[x(i,j)-\beta_1^j x(i,j-1)-\beta_2^j x(i-1,j)-\beta_3^j x(i-1,j)] \\ &= R(0,0)-\beta_1^j R(0,1)-\beta_2^j R(1,0)-\beta_3^j R(1,1) \end{aligned}$$

For any other subset s_i, the above procedure is repeated and $(B_i)^2$ in (3.9) is obtained. Note that in finding $(B_j)^2$ from (3.13) and (3.12) only the numerical values of

of R(m,n) are required.

Having defined a general modeling procedure, the following theorem establishes its applicability to special cases.

Theorem 3.1: If a zero mean, normal random process x(k) satisfies an autoregressive model

$$x(k) = \sum_{i=1}^{M'} \alpha_i x(k-I_i) + Bu(k) \tag{3.14}$$

where $M' \leq M$, then the modeling procedure will yield the same model.

Proof: Let

$$s = \{(k-I_1),(k-I_2),\ldots,(k-I_M)\}$$

let

$$q = k-I_{M'+1}$$

where $k-I_{M'+1}$ is an index (time) preceding k and not included in s, i.e. $q \notin s$

let

$$s_1 = sUq = \{(k-I_1),(k-I_2),\ldots,(k-I_M),(k-I_{M'+1})\}$$

Using the modeling procedure to find the autoregressive form for s_1, results in

$$x(k) = \sum_{i=1}^{M'+1} \beta_i x(k-I_i) + B_1 u(k) \tag{3.15}$$

and $\beta_1, \ldots, \beta_{M'+1}$ are obtained such that

$$E[x(k)-x_p(k)]^2 \tag{3.16}$$

is minimized, where (see (3.6))

$$x_p(k) = \sum_{i=1}^{M'+1} \beta_i x(k-I_i) \tag{3.17}$$

But due to the Gaussian nature of $x(k)$, the quantity

$$Ex(k) \mid x(k-I_1), \ldots, x(k-I_{M'+1}) \tag{3.18}$$

is linear in $x(k-I_1), \ldots, x(k-I_{M'+1})$ and if substituted for $x_p(k)$, it minimizes (3.16) [22]. Therefore

$$Ex(k) \mid x(k-I_1), \ldots, x(k-I_{M'+1}) = \sum_{i=1}^{M'+1} \beta_i x(k-I_i)$$

Also, using the conditioning of (3.18) in (3.14) results in

$$Ex(k) \mid x(k-I_1), \ldots, x(k-I_{M'+1})$$

$$= \sum_{i=1}^{M'} \alpha_i Ex(k-I_i) \mid x(k-I_1), \ldots, x(k-I_{M'+1})$$

$$+BEu(k) \mid x(k-I_1), \ldots, x(k-I_{M'+1}) \tag{3.20}$$

But

$$Eu(k) \mid x(k-I_1), \ldots, x(k-I_{M'+1}) = Eu(k) = 0$$

and

$$Ex(k-I_i)\,|\,x(k-I_1),\ldots,x(k-I_i),\ldots,x(k-I_{M'+1})=x(k-I_i)$$

Therefore (3.20) becomes

$$Ex(k)\,|\,x(k-I_1),\ldots,x(k-I_{M'+1}) = \sum_{i=1}^{M'} \alpha_i x(k-I_i)\,, \qquad (3.21)$$

But the conditional expectation in (3.18) is a specific linear function of $x(k-I_1),\ldots,x(k-I_{M'+1})$, thus the comparison of (3.19) and (3.21) necessitates that in (3.19)

$$\beta_i = \begin{cases} \alpha_i & \text{if } i \leq M' \\ 0 & \text{if } i = M'+1 \end{cases} \qquad (3.22)$$

This is turn indicates that in (3.15)

$$B_m = B \qquad (3.23)$$

Now, by letting the set q be empty, then the same procedure will indicate that (3.15) becomes identical to (3.14) and by letting q contain m elements with $M'+m \leq M$, then (3.22) and (3.23) will become

$$\beta_i = \begin{cases} \alpha_i & \text{if } i \leq M' \\ 0 & \text{if } i > M' \end{cases} \qquad (3.24)$$

and

$$B_m = B \qquad (3.25)$$

This completes the proof.

3.3 Choice of the Model's Order

The modeling procedure of Section 3.2 is particularly useful in the cases where only a small number of correlation values are specified. In these cases, an upper bound for the value of M exists. If this upper bound is sufficiently small, then its value should be used to represent the order of the model.

Since the modeling criterion is taken to be the minimization of the model's noise variance then, in general, M should be selected on the basis of the rate of decrease of $(B_i)^2$ as a function of M. This idea is applied and explored further in example 3.2.

It will be shown in Chapters 4 and 5 that the complexity and the memory requirement of the proposed estimator will be a direct function of the value of M. A logical consideration in the choice of M, therefore, is the trade off between additional implementation complexity and the reduction of $(B_i)^2$.

Example 3.2:

Consider the stationary two diemnsional correlation function

$$R(i,j,k,l) = R(|i-k|,|j-l|) = Ex(i,j)x(k,l)$$

$$= \exp[-\sqrt{(i-k)^2+(j-l)^2}]$$

Application of the modeling procedure yields:

a) Best 2nd order model is

$$x(i,j)=0.3x(i,j-1)+0.3x(i-1,j)+0.883u(i,j)$$

b) Best 3rd order model is

$$x(i,j)=0.29x(i,j-1)+0.25x(i-1,j)+0.12x(i-1,j+1)$$
$$+0.8775u(i,j)$$

c) Best 4th order model is

$$x(i,j)=0.28x(i,j-1)+0.24x(i-1,j)+0.03x(i-1,j-1)$$
$$+0.12x(i-1,j+1)+0.8769u(i,j)$$

d) Best 5th order model is

$$x(i,j)=0.28x(i,j-1)+0.24x(i-1,j)+0.03x(i-1,j-1)$$
$$+0.11x(i-1,j+1)+0.02x(i-1,j+2)+0.8768u(i,j)$$

This indicates that additional complexity of going from the 3rd to the 5th order does not reduce $(B_i)^2$ appreciably. Hence, for example, to a third decimal place accuracy, 3rd order model is a sufficient approximation.

3.4 Properties of the Modeling Procedure

Derivation of the autoregressive model has been based on minimizing the uncertainty associated with pre-

dicting the present value of a sample function of a process by a linear combination of a finite number (M) of its past values [16]. This is in accord with the concept of Fig. 2.2. The modeling procedure is directly applicable to nonstationary problems where for these cases the procedure must be applied at each time k, resulting in one autoregressive form for each k.

A property of the procedure, which is of extreme practical value, is that the determination of the model is only a function of the numerical values of R(m,n) and, in fact, is independent of the analytical form of R(m,n). This, in turn, enhances the numerical and computational character of the estimation process.

The optimality of the modeling procedure, in the cases where the process x(k) has a corresponding autoregressive model of order less than M, is established by theorem 3.1. However, by applying this method a model of the form (3.1) can always be found even when a finite order autoregressive model does not precisely describe the correlation information, or if the exact model is of an order higher than the chosen M. Whatever the case, it should be noted, the correlation function generated by the model obtained from this procedure will be identical to R(m,n) at, at least, M+1 points. This is demonstrated for a simple case considered in Example 3.3.

Example 3.3:

Consider the stationary one dimensional random process $z(k)$ with

$$R(i,j) = Ez(i)z(j) = R(|i-j|) \tag{3.26}$$

Let M=2 and the model of $z(k)$ be given by

$$x(k) = \beta_1 x(k-1)+\beta_2 x(k-2)+Bu(k) \tag{3.27}$$

The system of linear equations that β_1 and β_2 must satisfy, becomes (see Example 3.1)

$$\begin{bmatrix} R(0) & R(1) \\ R(1) & R(0) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \end{bmatrix}$$

which results in

$$\beta_1 = \frac{R(0)R(1)-R(1)R(2)}{R^2(0)-R^2(1)} \tag{3.28}$$

$$\beta_2 = \frac{R(0)R(2)-R^2(1)}{R^2(0)-R^2(1)} \tag{3.29}$$

Accordingly

$$\begin{aligned} B^2 &= E[x(k)-\beta_1 x(k-1)-\beta_2 x(k-2)] \\ &= R(0)-\beta_1 R(1)-\beta_2 R(2) \end{aligned} \tag{3.30}$$

Letting

$$C(|i-j|) = Ex(i)x(j) \tag{3.31}$$

be the stationary correlation function generated by (3.27), then

$$C(0) = Ex^2(k) = E[\beta_1 x(k-1)+\beta_2 x(k-2)+Bu(k)]^2$$
$$= (\beta_1^2+\beta_2^2)C(0)+B^2+2\beta_1\beta_2 C(1) \tag{3.32}$$

$$C(1) = E(x(k)x(k-1) = E[\beta_1 x(k-1)+\beta_2 x(k-2)+Bu(k)]x(k-1)$$
$$= \beta_1 C(0)+\beta_2 C(1) \tag{3.33}$$

$$C(2) = Ex(k)x(k-2)$$
$$= E[\beta_1 x(k-1)+\beta_2 x(k-2)+Bu(k)]x(k-2)$$
$$= \beta_1 C(1)+\beta_2 C(0) \tag{3.34}$$

Substituting (3.28) through (3.30) in (3.32), (3.33) and (3.34) and solving for C(0), C(1), C(2) yields

$$C(0) = R(0)$$
$$C(1) = R(1) \tag{3.35}$$
$$C(2) = R(2)$$

This indicates that the correlation function generated by (3.27) matches R(i,j) at, at least, 3 points.

CHAPTER 4

LINEAR PREDICTION

In this chapter it is assumed that the model of the random process x(k) is derived to be of the form

$$x(k) = \sum_{i=1}^{M} \beta_i x(k-I_i) + Bu(k) \qquad (4.1)$$

and the process x(k) is assumed to be stationary. Derivation of the linear predictor for nonstationary processes is identical to that of stationary processes, hence is omitted. Section 4.1 contains a brief discussion on the optimal linear one step prediction and the difficulties associated with implementing such a procedure. The development and derivation of an implementable one step predictor compatible with the proposed system structure of Fig. 2.2 is presented in Section 4.2. Section 4.3 contains the derivation of the variance of the one step predicted value.

4.1 Optimal Linear One Step Prediction

Denoting the MMS one step prediction value of x(k) by $x^p(k)$, then

$$x^P(k) = Ex(k)|y(1),....,y(k-1) \tag{4.2}$$

Conditioning the right hand side of (4.1) on $y(1),...y(k-1)$ and taking the expectation results in

$$x^P(k)=\sum_{i=1}^{M} \beta_i Ex(k-I_i)|y(1),...,y(k-1)+BEu(k)|y(1),..,y(k-1) \tag{4.3}$$

But due to the statistical independence and the zero mean property of $u(k)$

$$Eu(k)|y(1),....,y(k-1) = Eu(k) = 0 \tag{4.4}$$

hence

$$x^P(k) = \sum_{i=1}^{M} \beta_i Ex(k-I_i)|y(1),....,y(k-1) \tag{4.5}$$

The difficulty in finding $x^P(k)$ in (4.5) is that at each pixel k, the M expectations

$$Ex(k-I_i)|y(1),....,y(k-1) \qquad i=1,2,....,M \tag{4.6}$$

have to be carried out. Performing this task involves interpolation of the random process $x(k)$, which in the simplest case is computationally unfeasible. To maintain implementability, a suboptimal recursive prediction procedure is introduced in the following section.

4.2 A Linear One Step Prediction

In accordance with the discussion of Chapter 2 and the system structure of Fig. 2.2, the linear predictor is to be based on past estimates and their error variances. In this respect, the one step predicted value $x^*(k)$ is given by

$$x^*(k) = \sum_{j=1}^{k-1} \alpha_j \hat{x}(k-j) \tag{4.7}$$

where $\alpha_1, \ldots, \alpha_{k-1}$ are to be chosen such that

$$E[x(k)-x^*(k)]^2 \tag{4.8}$$

is minimized. The above minimization is to be carried out based on the available information to the predictor. This information, in turn, consists of the values of $\hat{x}(i)$ and $\hat{\sigma}^2(i)$, $i \leq k-1$. Furthermore, each $\hat{x}(i)$ and $\hat{\sigma}^2(i)$ represents the mean and variance, respectively, of a posteriori probability density of $x(i)$ at time i. This interpretation is directly substantiated by the way each estimate $\hat{x}(i)$ at time i is obtained as a function of the previous estimates and the observation $y(i)$.

As a result of the above discussion, the expectation in (4.8) is well defined and operates on each random variable $x(i)$ such that

$$Ex(i) = \hat{x}(i) \qquad i=1,2,\ldots,k-1 \tag{4.9a}$$

and

$$E[x(i)-\hat{x}(i)]^2 = \hat{\sigma}^2(i) \qquad i=1,2,\ldots,k-1 \tag{4.9b}$$

Having (4.9a) and (4.9b) as the definition of the expectation operator, the following theorem establishes the optimal linear predictor.

Theorem 4.1: When the random process x(k) satisfies a model of the form (4.1), then the (optimal) choice of $\alpha_1,\alpha_2,\ldots,\alpha_{k-1}$ in (4.7) which minimizes (4.8), is given by

$$\alpha_j = \begin{cases} \beta_i \text{ if } k-j=k-I_i \text{ for some } i=1,2,\ldots,M \\ 0 \quad \text{otherwise} \end{cases}$$

Proof: Substitution of (4.1) and (4.7) in (4.8) yields

$$E[x(k)-x^*(k)]^2 = E[Bu(k)+\sum_{i=1}^{M}\beta_i x(k-I_i)-\sum_{j=1}^{k-1}\alpha_j \hat{x}(k-j)]^2 \tag{4.10}$$

Since $k-I_i$, i=1,2,....,M is a set of two dimensional indices and their particular values are immaterial to this proof, assume

$$I_i = i \tag{4.11}$$

in order to reduce notational complexity. With this,

(4.10) then becomes

$$E[x(k)-x^*(k)]^2 = E[Bu(k) + \sum_{i=1}^{M} \beta_i x(k-i) - \sum_{j=1}^{k-1} \alpha_j \hat{x}(k-j)]^2$$

$$= E[Bu(k)]^2 + E[\sum_{i=1}^{M} \beta_i x(k-i) - \sum_{j=1}^{k-1} \alpha_j \hat{x}(k-j)]^2$$

$$+ 2E\{[Bu(k)][\sum_{i=1}^{M} \beta_i x(k-i) - \sum_{j=1}^{k-1} \alpha_j \hat{x}(k-j)]\} \qquad (4.12)$$

But (4.9) and the statistical independence of u(k) imply that

$$E[Bu(k)]^2 = B^2 \qquad (4.13)$$

$$E\{[Bu(k)][\sum_{i=1}^{M} \beta_i x(k-i) - \sum_{j=1}^{k-1} \alpha_j \hat{x}(k-j)]\} = 0 \qquad (4.14)$$

Substitution of (4.13) and (4.14) in (4.12) yields

$$E[x(k)-x^*(k)]^2 = B^2 + E[\sum_{i=1}^{M} \beta_i x(k-i) - \sum_{j=1}^{k-1} \alpha_j \hat{x}(k-j)]^2$$

$$= B^2 + E[\sum_{i=1}^{M} \beta_i x(k-i) - \sum_{j=1}^{M} \alpha_{k-j} \hat{x}(k-j) - \sum_{j=1}^{k-M-1} \alpha_j \hat{x}(j)]^2$$

(4.15)

Since

$$\sum_{i=1}^{M} \alpha_{k-i} \hat{x}(k-i) = \sum_{i=1}^{M} [\beta_i + \alpha_{k-i} - \beta_i] \hat{x}(k-i)$$

$$= \sum_{i=1}^{M} \beta_i \hat{x}(k-i) + \sum_{i=1}^{M} (\alpha_{k-i} - \beta_i) \hat{x}(k-i)$$

(4.16)

then, substitution of (4.16) in (4.15) and the expansion of the square term results in

$$E[x(k)-x^*(k)]^2$$

$$= B^2 + E\{ \sum_{i=1}^{M} \beta_i [x(k-i)-\hat{x}(k-i)] - \sum_{i=1}^{M} (\alpha_{k-i}-\beta_i)\hat{x}(k-i) - \sum_{i=1}^{k-M-1} \alpha_i \hat{x}(i) \}^2$$

$$= B^2 + E\{ \sum_{i=1}^{M} \beta_i [x(k-i)-\hat{x}(k-i)] \}^2 + E[\sum_{i=1}^{M} (\alpha_{k-i}-\beta_i)\hat{x}(k-i)]^2$$

$$+ E[\sum_{i=1}^{k-M-1} \alpha_i \hat{x}(i)]^2 - 2E[\sum_{i=1}^{M} \beta_i (x(k-i)-\hat{x}(k-i))][\sum_{i=1}^{k-M-1} \alpha_i \hat{x}(i)]$$

$$- 2E[\sum_{i=1}^{M} \beta_i (x(k-i)-\hat{x}(k-i))][\sum_{i=1}^{M} (\alpha_{k-1}-\beta_i)\hat{x}(k-i)$$

$$+ 2E[\sum_{i=1}^{M} (\alpha_{k-i}-\beta_i)\hat{x}(k-i)][\sum_{i=1}^{k-M-1} \alpha_i \hat{x}(i)]$$

(4.17)

But in (4.17) $\hat{x}(i)$, i=1,2,....,k-1 are a set of nonrandom

quantities. This along with (4.9a) indicates that

$$E[\sum_{i=1}^{M}\beta_i(x(k-i)-\hat{x}(k-i))][\sum_{i=1}^{M}(\alpha_{k-i}-\beta_i)\hat{x}(k-i)]$$

$$=[\sum_{i=1}^{M}\beta_i(Ex(k-i)-\hat{x}(k-i))][\sum_{i=1}^{M}(\alpha_{k-i}-\beta_i)\hat{x}(k-i)]$$

$$=[\sum_{i=1}^{M}\beta_i(\hat{x}(k-i)-\hat{x}(k-i))][\sum_{i=1}^{M}(\alpha_{k-i}-\beta_i)\hat{x}(k-i)] = 0 \tag{4.18}$$

and similarly

$$E[\sum_{i=1}^{M}\beta_i(x(k-i)-\hat{x}(k-i))][\sum_{i=1}^{k-M-1}\alpha_i\hat{x}(i)] = 0 \tag{4.19}$$

Furthermore, realizing that the third, fourth and seventh terms of (4.17) are nonrandom, use of (4.18) and (4.19) reduces (4.17) to

$$E[x(k)-x^*(k)]^2$$

$$=B^2+E[\sum_{i=1}^{M}\beta_i(x(k-i)-\hat{x}(k-i))]+[\sum_{i=1}^{M}(\alpha_{k-i}-\beta_i)\hat{x}(k-i)]^2$$

$$+[\sum_{i=1}^{k-M-1}\alpha_i\hat{x}(i)]^2+2[\sum_{i=1}^{M}(\alpha_{k-i}-\beta_i)\hat{x}(k-i)][\sum_{i=1}^{k-M-1}\alpha_i\hat{x}(i)]$$

or

$$E[x(k)-x^*(k)]^2$$

$$=B^2+E\Big[\sum_{i=1}^{M}\beta_i(x(k-i)-\hat{x}(k-i))\Big]^2$$

$$+\Big[\sum_{i=1}^{M}(\alpha_{k-i}-\beta_i)\hat{x}(k-i)+\sum_{i=1}^{k-M-1}\alpha_i\hat{x}(i)\Big]^2 \qquad (4.20)$$

The first two terms of (4.20) are independent of α_i, $i=1,2,\ldots,k-1$. The third term is a complete square and its minimum is zero. Therefore, the minimum of (4.20) is achieved if

$$\alpha_i=\begin{cases}\beta_i & \text{for } i=k-1,k-2,\ldots,k-M\\ 0 & \text{for } i=1,2,\ldots,k-M-1\end{cases} \qquad (4.21)$$

This completes the proof of Theorem 4.1.

The foregoing theorem establishes that the best linear predictor is given by

$$x^*(k)=\sum_{i=1}^{M}\beta_i\hat{x}(k-I_i) \qquad (4.22)$$

which yields itself to on line implementation and satisfies the finite memory requirement of Chapter 2.

4.3 Error Variance of Predicted Value

Letting $\sigma_p^2(k)$ denote the error variance of the predicted value $x^*(k)$ at time k, then

$$\sigma_p^2(k) = E[x(k)-x^*(k)]^2 \tag{4.23}$$

Substituting for x(k) and $x^*(k)$ in (4.23) from (4.1) and (4.22), respectively, yields

$$\sigma_p^2(k) = E[Bu(k) + \sum_{i=1}^{M} \beta_i x(k-I_i) - \sum_{i=1}^{M} \beta_i \hat{x}(k-I_i)]^2$$

$$= B^2 + E\{\sum_{i=1}^{M} \beta_i [x(k-I_i)-\hat{x}(k-I_i)]\}^2 \tag{4.24}$$

where in deriving (4.24), use of the statistical independence and the zero mean property of u(k) is made. Let

$$e(k-I_i) = x(k-I_i)-\hat{x}(k-I_i) \tag{4.25}$$

then at each time $k-I_i$, $e(k-I_i)$ is a random variable whose first two moments are (see (4.9a) and (4.9b)

$$Ee(k-I_i) = 0 \tag{4.26}$$

$$Ee^2(k-I_i) = \hat{\sigma}^2(k-I_i) \tag{4.27}$$

Expansion of (4.24) in terms of e(.) as defined in (4.25)

results in

$$\sigma_p^2(k)=B^2+E\Big[\sum_{i=1}^{M}\beta_i e(k-I_i)\Big]^2$$

$$=B^2+\beta_1^2Ee^2(k-I_1)+.....+\beta_M^2Ee^2(k-I_M)$$

$$+2\beta_1\beta_2Ee(k-I_1)e(k-I_2)+...+2\beta_1\beta_MEe(k-I_1)e(k-I_M)$$

$$\vdots$$

$$+2\beta_{M-1}\beta_MEe(k-I_{M-1})e(k-I_M) \qquad (4.28)$$

This relation shows that the evaluation of $\sigma_p^2(k)$ requires the knowledge of cross covariances of the random variables $e(k-I_i)$, $i=1,2,....,M$. To avoid the numerical difficulties associated with the evaluation of these cross covariances at each time k, the following upper bound of $\sigma_p^2(k)$, denoted by $\sigma^{*2}(k)$, is derived and used in the filtering step of the next chapter. The reason for finding an upper bound, as opposed to a lower bound, is due to the fact that $\sigma_p^2(k)$ is a measure of uncertainty of the value $x^*(k)$, thus by assigning value $\sigma^{*2}(k)$ to $x^*(k)$ the uncertainty associated with $x^*(k)$ is increased. This causes the estimation process to remain suboptimal and is discussed in more detail in Chapter 8.

Taking the absolute values of the right hand side of

(4.28) and using (4.27) results in

$$\sigma_p^2(k) \leq B^2 + |\beta_1|^2 \hat{\sigma}^2(k-I_1) + \ldots\ldots + |\beta_M|^2 \hat{\sigma}^2(k-I_M)$$

$$+2|\beta_1||\beta_2||Ee(k-I_1)e(k-I_2)| + \ldots\ldots$$

$$+2|\beta_2||\beta_3||Ee(k-I_2)e(k-I_3)| + \ldots\ldots$$

$$\vdots$$

$$+2|\beta_{M-1}||\beta_M||Ee(k-I_{M-1})e(k-I_M)| \qquad (4.29)$$

But for each pair of random variables $e(k-I_i)$ and $e(k-I_j)$, Cauchy-Shwartz inequality [30] establishes that

$$|Ee(k-I_i)e(k-I_j)| \leq [Ee^2(k-I_i)Ee^2(k-I_j)]^{1/2} = \hat{\sigma}(k-I_i)\hat{\sigma}(k-I_j) \qquad (4.30)$$

so

$$\sigma_p^2(k) \leq B^2 + |\beta_1|^2 \hat{\sigma}^2(k-I_1) + \ldots\ldots + |\beta_M|^2 \hat{\sigma}^2(k-I_M)$$

$$+2|\beta_1||\beta_2|\hat{\sigma}(k-I_1)\hat{\sigma}(k-I_2) + \ldots\ldots$$

$$\vdots$$

$$+2|\beta_{M-1}||\beta_M|\hat{\sigma}(k-I_{M-1})\hat{\sigma}(k-I_M) \qquad (4.31)$$

where (4.31) is equivalently written as

$$\sigma_p^2(k) \leq B^2 + \left[\sum_{i=1}^{M} |\beta_i| \hat{\sigma}(k-I_i)\right]^2 \qquad (4.32)$$

Hence, the upper bound $\sigma^{*2}(k)$ is given by

$$\sigma^{*2}(k)=B^2+\left[\sum_{i=1}^{M}|\beta_i|\hat{\sigma}(k-I_i)\right]^2 \tag{4.33}$$

It should be noted that $\sigma^{*2}(k)$ can easily be evaluated at each time k, since quantities $\hat{\sigma}^2(k-I_i)$, i=1,2,....,M are available to the predictor for each k.

CHAPTER 5

FILTERING

This chapter develops the derivation of the final step of the proposed estimation technique, where block F of Fig. 2.2 will be analyzed and defined. In Section 5.1, the information obtained from the linear predictor is incorporated into an approximate probability density function (PDF) for x(k). This density is then used in Section 5.2 as an a priori statistic for the observation y(k) to derive the estimate $\hat{x}(k)$ and its error variance $\hat{\sigma}^2(k)$.

5.1 Choice of A Posteriori PDF for x(k)

Assuming that the actual prediction variance $\sigma_p^2(k)$ is computed and available then the predicted value $x^*(k)$ and its variance $\sigma_p^2(k)$ represent the mean and the variance of the a posteriori PDF on x(k). This density represents the available knowledge of the random variable x(k) prior to the reception of the observation y(k). Since, for a given mean and variance the normal distribution represents the maximum uncertainty (entropy) [30], this density function is assumed to be normal. Further uncertainty is associated with x(k) if $\sigma^{*2}(k)$ is used in place of $\sigma_p^2(k)$.

Consequently, an approximate and a rather conservative choice of an a posteriori probability density for x(k) is

$$p(x(k)) = \frac{1}{\sigma^*(k)\sqrt{2\pi}} \exp\{-\frac{[x(k)-x^*(k)]^2}{2\sigma^{*2}(k)}\} \qquad (5.1)$$

5.2 Evaluation of the Estimate and its Variance

The density p(x(k)) in (5.1) is used as an a priori statistic for y(k) in order to obtain the Baye's estimate, $\hat{x}(k)$, of x(k) as follows

$$\hat{x}(k) = Ex(k)|y(k) = \int x(k)p(x(k)|y(k))dx(k) \qquad (5.2)$$

Using the Baye's rule

$$p(x(k)|y(k)) = \frac{p(x(k),y(k))}{p(y(k))} = \frac{p(y(k)|x(k))p(x(k))}{p(y(k))} \qquad (5.3)$$

Therefore

$$\hat{x}(k) = \frac{1}{p(y(k))}\int x(k)p(y(k)|x(k))dx(k) \qquad (5.4)$$

Using the same procedure as outlined in Section 2.1, it follows that

$$p(y(k))=\int p(y(k),x(k))dx(k)=\int p(y(k)|x(k))p(x(k))dx(k) \qquad (5.5)$$

Therefore, the estimate $\hat{x}(k)$ is given by

$$\hat{x}(k) = \frac{\int x(k)p(y(k)|x(k))p(x(k))dx(k)}{\int p(y(k)|x(k))p(x(k))dx(k)} \tag{5.6}$$

Similarly, the error variance, $\hat{\sigma}^2(k)$, is obtained by

$$\hat{\sigma}^2(k) = E[x(k)-\hat{x}(k)]^2|y(k)$$

$$= [x(k)-\hat{x}(k)]^2 p(x(k)|y(k))dx(k) \tag{5.7}$$

Again by using the identity of (5.3), (5.7) becomes

$$\hat{\sigma}^2(k) = \frac{1}{p(y(k))} \int [x(k)-\hat{x}(k)]^2 p(y(k)|x(k))p(x(k))dx(k) \tag{5.8}$$

The substitution of (5.5) in (5.8) results in

$$\hat{\sigma}^2(k) = \frac{\int [x(k)-\hat{x}(k)]^2 p(y(k)|x(k))p(x(k))dx(k)}{\int p(y(k)|x(k))p(x(k))dx(k)} \tag{5.9}$$

In (5.6) and (5.9), $p(x(k))$ is given by (5.1) and $p(y(k)|x(k))$ is obtained from that part of the a priori information which describes the observation system structure and the probability density of the observation noise (equation (1.7)).

The comparison of equations (5.6) and (5.9) with those of the optimal filter given in (2.10) and (2.13) indicates that the proposed estimator exhibits the same logic as the

optimal estimator. This logic consists of each estimate and its error variance being a specific function of the past and present information where $p(x(k))$ and $p(y(k)|x(k))$ represent these two quantities, respectively. In fact, had the prediction been done optimally, the proposed procedure would have been optimal.

The integrals involved in evaluating $\hat{x}(k)$ and $\hat{\sigma}^2(k)$ in (5.6) and (5.9) may or may not have analytic solutions. If such solutions exist, then the computational requirement of the procedure is reduced tremendously. If such solutions do not exist then these integrals can be evaluated numerically. This in turn allows the procedure to be applicable to a broad class of observation systems including nonlinear forms of $y(k)$. Chapter 6 contains examples demonstrating and substantiating these properties.

The estimator developed so far is both feasible and implementable. Its feasibility is due to the structure of Fig. 2.2, which leads to (5.6) and (5.9) being scalar operations.

CHAPTER 6

DERIVATION OF A FEW SPECIFIC ESTIMATORS AND APPLICATIONS

6.1 Introduction

Based on the general estimation procedure as developed in Chapters 2 through 5, a number of specific estimators for various observation systems are derived in this chapter. Each estimator then is utilized in estimating a number of noisy images as examples of the applicability of the procedure. The procedure is almost exclusively applied to two dimensional pictorial data and its applicability to similar one dimensional problems is implied implicitly. Section 6.2 contains a discussion on the methods of finding and specifying two dimensional a priori correlation functions. In Section 6.3, a linear estimator is derived by applying the estimation procedure to the case of additive-Gaussian observation noise. The properties of this linear estimator are outlined and discussed in more detail in Section 6.4. Section 6.5 contains the derivation of the estimator for observations having bounded multiplicative noise. This section also includes the application of the procedure to two dimen-

sional pictorial data corrupted by uniform multiplicative noise. Finally in Section 6.6, the application of the procedure to observations containing multiplicative and additive noise terms is considered.

6.2 Two Dimensional A Priori Statistics

In applying the foregoing estimation method to the noise corrupted two dimensional pictorial data, the knowledge of the a priori mean and at least a few values of the autocorrelation function of the image is required. These two quantities are defined by (1.1) and (1.2). If the image is a member of a stationary two dimensional random process, then they are defined to be

$$M(i,j) = M = Eb(i,j) \tag{6.1}$$

$$R(i,j,k,l,) = R(|i-k|,|j-l|) = E[b(i,j)-M][b(k,l)-M] \tag{6.2}$$

where E represents the ensemble averaging.

Experimental results indicate that random fields with exponential autocorrelation functions are realistic models for a variety of pictorial data [2] - [5]. Two widely used forms of these functions for stationary processes are

$$R(|i-k|,|j-l|) = \sigma_s^2 \exp[-\alpha_1|i-k|-\alpha_2|j-l|] \tag{6.3}$$

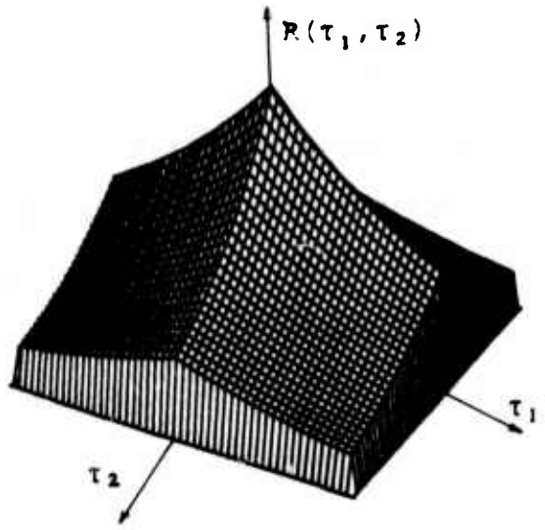
and

$$R(|i-k|,|j-l|) = \sigma_s^2 \exp[-\sqrt{\alpha_1^2(i-k)^2 + \alpha_2^2(j-l)^2}] \qquad (6.4)$$

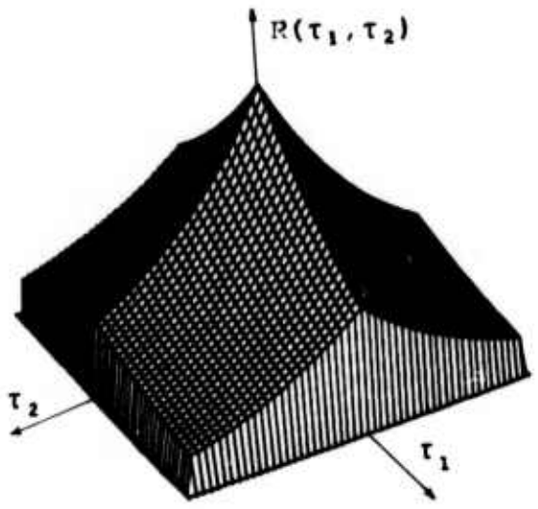
where σ_s^2 is the signal power and $|i-k|$ and $|j-l|$ are the increments in the verticle and horizontal directions, respectively. Although these correlation functions are both of the exponential decaying type, they exhibit dissimilar characteristics in that the separable correlation function of (6.3) assumes more correlation in the horizontal and vertical directions while the nonseparable function in (6.4) indicates a smooth and rotationally invariant correlation in all directions. Figures 6.1 and 6.2 represent two views of the three dimensional graphs of these functions. It is suggested in [1] that pictures of the natural scenes exhibit nonseparable and rotationally invariant correlations while the images of man-made objects correspond to separable autocorrelation functions.

The complete definition of the correlation function in (6.3) or (6.4) requires the specification of the three quantities σ_s, α_1 and α_2. Note that if any three correlation values (for example R(0,0), R(0,1), R(1,0)) are known then these quantities can be obtained.

For the processed images in this dissertation, the approximate values for the mean M and correlations R(0,0), R(0,1) and R(1,0) are obtained, wherever necessary, from



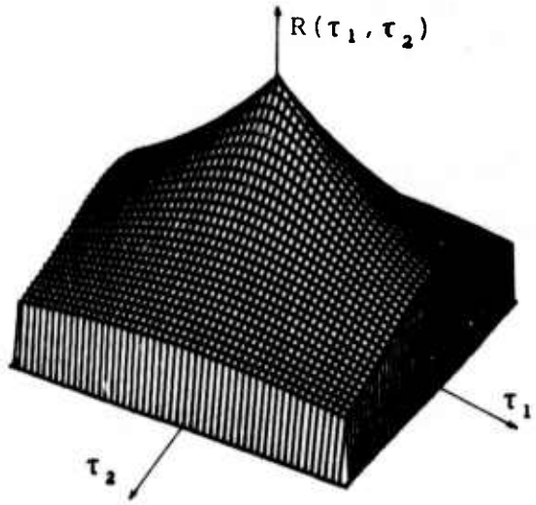
(a)



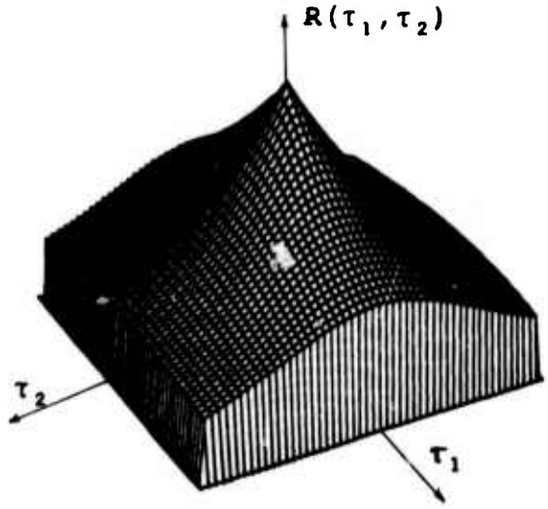
(b)

Fig. 6.1 Two Views of the Separable Correlation Function

$$R(\tau_1, \tau_2) = \sigma_s^2 \exp[-\alpha_1|\tau_1| - \alpha_2|\tau_2|]$$



(a)



(b)

Fig. 6.2 Two Views of the Nonseparable Correlation Function

$$R(\tau_1, \tau_2) = \sigma_s^2 \exp[-\sqrt{\alpha_1^2 \tau_1^2 + \alpha_2^2 \tau_2^2}]$$

$$M \simeq \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} b(i,j) \tag{6.5}$$

$$R(p,q) \simeq \frac{1}{N^2} \sum_{r=1}^{N-p} \sum_{s=1}^{N-q} [b(r,s)-M][b(r+p,q+s)-M] \tag{6.6}$$

where $b(i,j)$ is the intensity value of the original image at pixel (i,j) and N represents the size of the image.

6.3 Additive-Gaussian Observation Noise

In the case where the observation is given by

$$y(k) = x(k)+\gamma(k) \tag{6.7}$$

with $\gamma(k)$ normal and

$$E\gamma(k) = 0 \tag{6.8}$$

$$E\gamma(i)\gamma(j) = \begin{cases} 0 & \text{if } i \neq j \\ \sigma_\gamma^2(i) & \text{if } i = j \end{cases} \tag{6.9}$$

then

$$p(y(k)\,|\,x(k)) = \frac{1}{\sqrt{2\pi}\sigma_\gamma(k)} \exp\left[-\frac{(y(k)-x(k))^2}{2\sigma_\gamma^2(k)}\right] \tag{6.10}$$

Lemma 6.1: The estimate $\hat{x}(k)$ and its error variance $\hat{\sigma}^2(k)$, for this case of additive-Gaussian observation noise

are given by

$$\hat{x}(k) = \frac{\sigma^{*2}(k)}{\sigma^{*2}(k)+\sigma_\gamma^2(k)}\, y(k)+\frac{\sigma_\gamma^2(k)}{\sigma^{*2}(k)+\sigma_\gamma^2(k)}x^*(k) \qquad (6.11)$$

$$\hat{\sigma}^2(k) = \frac{\sigma^{*2}(k)\,\sigma_\gamma^2(k)}{\sigma^{*2}(k)+\sigma_\gamma^2(k)} \qquad (6.12)$$

Proof: Substituting (6.10) in (5.6) and (5.9) and dropping the time dependence k from all variables, then the estimate x and its error variance $\hat{\sigma}^2$ at time k are given by

$$\hat{x} = \frac{\int_{-\infty}^{\infty} x \exp[-\frac{(y-x)^2}{2\sigma_\gamma^2} - \frac{(x-x^*)^2}{2\sigma^{*2}}]dx}{\int_{-\infty}^{\infty} \exp[-\frac{(y-x)^2}{2\sigma_\gamma^2} - \frac{(x-x^*)^2}{2\sigma^{*2}}]dx} \qquad (6.13)$$

$$\hat{\sigma}^2 = - \frac{\int_{-\infty}^{\infty} (x-\hat{x})^2 \exp[-\frac{(y-x)^2}{2\sigma_\gamma^2} - \frac{(x-x^*)^2}{2\sigma^{*2}}]dx}{\int_{-\infty}^{\infty} \exp[-\frac{(y-x)^2}{2\sigma_\gamma^2} - \frac{(x-x^*)^2}{2\sigma^{*2}}]dx} \qquad (6.14)$$

But

$$\frac{(y-x)^2}{\sigma_\gamma^2}+\frac{(x-x^*)^2}{\sigma^{*2}}=\frac{\sigma^{*2}y^2+\sigma_\gamma^2x^{*2}+(\sigma^{*2}+\sigma_\gamma^2)x^2-2(\sigma^{*2}y+\sigma_\gamma^2x^*)x}{\sigma_\gamma^2\sigma^{*2}}$$

$$=\frac{\sigma_\gamma^2+\sigma^{*2}}{\sigma_\gamma^2\sigma^{*2}}\left[x^2-2\left(\frac{\sigma^{*2}}{\sigma^{*2}+\sigma_\gamma^2}y+\frac{\sigma_\gamma^2}{\sigma^{*2}+\sigma_\gamma^2}x^*\right)\right]+\frac{1}{\sigma_\gamma^2}y^2+\frac{1}{\sigma^{*2}}x^{*2} \tag{6.15}$$

Let

$$\alpha_1=\frac{\sigma^{*2}}{\sigma^{*2}+\sigma_\gamma^2}$$

$$\alpha_2=\frac{\sigma_\gamma^2}{\sigma^{*2}+\sigma_\gamma^2}$$

$$\xi^2=\frac{\sigma_\gamma^2\,\sigma^{*2}}{\sigma^{*2}+\sigma_\gamma^2} \tag{6.16}$$

then

$$\frac{(y-x)^2}{\sigma_\gamma^2}+\frac{(x-x^*)^2}{\sigma^{*2}}$$

$$=\frac{1}{\xi^2}[x-(\alpha_1y+\alpha_2x^*)]^2-\frac{1}{\xi^2}(\alpha_1y+\alpha_2x^*)^2+\frac{1}{\sigma_\gamma^2}y^2+\frac{1}{\sigma^{*2}}x^{*2} \tag{6.17}$$

letting

$$q = \frac{1}{\sigma_\gamma^2} y^2 + \frac{1}{\sigma^{*2}} x^{*2} - \frac{1}{\xi^2} (\alpha_1 y + \alpha_2 x^*)^2$$

then (6.17) becomes

$$\frac{(y-x)^2}{\sigma_\gamma^2} \frac{(x-x^*)^2}{\sigma^{*2}} = \frac{1}{\xi^2} [x-(\alpha_1 y+\alpha_2 x^*)]^2 + q \qquad (6.18)$$

Substituting (6.18) in (6.13) and (6.14) and noting that α_1, α_2, ξ^2 and q are independent of x, yields

$$\hat{x} = \frac{e^{-q/2} \int_{-\infty}^{\infty} x \exp\left[-\frac{[x-(\alpha_1 y+\alpha_2 x^*)]^2}{2\xi^2}\right] dx}{e^{-q/2} \int_{-\infty}^{\infty} \exp\left[-\frac{[x-(\alpha_1 y+\alpha_2 x^*)]^2}{2\xi^2}\right] dx} \qquad (6.19)$$

$$\hat{\sigma}^2 = \frac{e^{-q/2} \int_{-\infty}^{\infty} (x-\hat{x})^2 \exp\left[-\frac{[x(\alpha_1 y+\alpha_2 x^*)]^2}{2\xi^2}\right] dx}{e^{-q/2} \int_{-\infty}^{\infty} \exp\left[-\frac{[x-(\alpha_1 y+\alpha_2 x^*)]^2}{2\xi^2}\right] dx} \qquad (6.20)$$

Canceling the common terms from the numerator and the denomenator of (6.19) and (6.20) and realizing that

$$\frac{1}{\xi\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{[x-(\alpha_1 y+\alpha_2 x^*)]^2}{2\xi^2}\right] dx = 1$$

then

$$\hat{x} = \frac{1}{\xi\sqrt{2\pi}}\int_{-\infty}^{\infty} x \exp[-\frac{[x-(\alpha_1 y+\alpha_2 x^*)]^2}{2\xi^2}]dx \tag{6.21}$$

$$\hat{\sigma}^2 = \frac{1}{\xi\sqrt{2\pi}}\int_{-\infty}^{\infty} (x-\hat{x})^2 \exp[-\frac{[x-(\alpha_1 y+\alpha_2 x^*)]^2}{2\xi^2}]dx \tag{6.22}$$

But these two relationships indicate that $\hat{x}$ and $\hat{\sigma}^2$ are the mean and variance of the Gaussian density function

$$p(x) = \frac{1}{\xi\sqrt{2\pi}} \exp[-\frac{[x-(\alpha_1 y+\alpha_2 x^*)]^2}{2\xi^2}] \tag{6.23}$$

therefore

$$\hat{x} = \alpha_1 y+\alpha_2 x^* = \frac{\sigma^{*2}}{\sigma^{*2}+\sigma_\gamma^2} y + \frac{\sigma_\gamma^2}{\sigma^{*2}+\sigma_\gamma^2} x^* \tag{6.24}$$

and

$$\hat{\sigma}^2 = \xi^2 = \frac{\sigma_\gamma^2 \sigma^{*2}}{\sigma^{*2}+\sigma_\gamma^2} \tag{6.25}$$

This completes the proof.

This lemma indicates that the estimation process, in this case of additive-Gaussian noise, involves obtaining $x^*(k)$ and $\sigma^{*2}(k)$ from (4.22) and (4.33) and inserting them in (6.11) and (6.12) to find the estimate and its error variance at each time k.

Figures 6.3 and 6.4 describe the application of the foregoing linear estimation method to two dimensional pictorial data. In these figures, a stationary nonseparable correlation of the form

$$R(|i-k|,|j-1|) = Ex(i,j)x(k,l)$$

$$= \sigma^2 \exp[-\sqrt{\alpha_1^2(i-k)^2+\alpha_2^2(j-1)^2}] \qquad (6.26)$$

was assumed for the originals in Fig. 6.3(a) and Fig. 6.4(a). From (6.6), the three correlation values which are used in finding σ , α_1, and α_2, were found to be

$$R(0,0) = 1816$$

$$R(0,1) = 1807$$

$$R(1,0) = 1797$$

Using the modeling procedure the best 7th order autoregressive model was obtained as

$$\begin{aligned} x(i,j) = {} & 0.87x(i,j-1)+0.02x(i,j-2)-0.03x(i-1,j-2) \\ & +0.01x(i-1,j-1)+0.03x(i-1,j)+0.03x(i-1,j+1) \\ & +0.13x(i-1,j+2)+0.29u(i,j) \end{aligned} \qquad (6.27)$$

Note that with the above model, the estimator requires only the storage of the current and the previously estimated line of the image at each time k. Based on the availability of two lines of image, at each time k, and the use of the guidelines of the modeling procedure, the

(a) Original

(b) Noisy, SNR=1

(c) Estimate

Fig. 6.3 Additive-Gaussian Observation Noise

(a) Original

(b) Noisy, SNR=1/2

(c) Estimate

Fig. 6.4 Additive-Gaussian Observation Noise

7th order model of (6.27) was found to be adequate.

Figures 6.3(b) and 6.4(b) represent noisy images, where the noise is white additive-Gaussian with the signal to noise ratios of one and one-half respectively. The estimated image of Fig. 6.3(c) represents a 7.54 and Fig. 6.4(c) represents an 8.4 db improvement (see Appendix B for the definition of db improvement).

6.4 Properties of the Linear Estimator

For an autoregressive model of the form

$$x(k) = \sum_{i=1}^{M} \beta_i x(k-I_i) + Bu(k) \tag{6.28}$$

when the observation noise is additive-Gaussian, then the estimate and its error variance, at each time k are given by

$$\hat{x}(k) = \frac{\sigma_\gamma^2(k)}{\sigma^{*2}(k)+\sigma_\gamma^2(k)} x^*(k) + \frac{\sigma^{*2}(k)}{\sigma^{*2}(k)+\sigma_\gamma^2(k)} y(k) \tag{6.29}$$

$$\hat{\sigma}^2(k) = \frac{\sigma_\gamma^2(k)\sigma^{*2}(k)}{\sigma^{*2}(k)+\sigma_\gamma^2(k)} \tag{6.30}$$

where

$$x^*(k) = \sum_{i=1}^{M} \beta_i \hat{x}(k-I_i) \tag{6.31}$$

$$\sigma^{*2}(k) = B^2 + [\sum_{i=1}^{M} |\beta_i| \hat{\sigma}(k-I_i)]^2 \tag{6.32}$$

and $\sigma_\gamma^2(k)$ is the observation noise variance at time k.

Relations (6.29) through (6.32) indicate that implementation of the linear estimator is quite simple leading to on line processing of either images or one dimensional signals. An encouraging property of this estimator is that the error variance $\hat{\sigma}^2(k)$ satisfies the set of recursive equations (6.30) and (6.32) which are independent of the value of the observation y(k). This enables the implementation of (6.30) and (6.31) on a digital computer, prior to reception of any observations, and the investigation of the steady state behavior of the error variance. Since this estimation method is, in general, suboptimal, as will be discussed in Chapter 8, the above a priori analysis could be used as a basis for deciding for or against the use of the foregoing estimation technique even when the optimal solution exists. Of course, the advantage of the use of this procedure when the optimal solution exists is the computational and implementation simplicity of this estimation algorithm. This is shown in more detail in the following example.

Example 6.1: Consider the one dimensional random process $x(k)$ satisfying the following autoregressive model.

$$x(k) = 0.248x(k-1)+0.014x(k-2)+0.969u(k) \tag{6.33}$$

For the observation

$$y(k) = x(k)+\gamma(k) \tag{6.34}$$

with

$$\begin{aligned} E\gamma(k) &= 0 \\ E\gamma^2(k) &= 0.5 \end{aligned} \tag{6.35}$$

the error variance of the one step predicted value can be obtained from the two recursive relationships

$$\sigma^{*2}(k) = (0.969)^2+[0.248\hat{\sigma}(k-1)+0.014\hat{\sigma}(k-2)]^2 \tag{6.36}$$

and

$$\hat{\sigma}^2(k) = \frac{0.5\sigma^{*2}(k)}{0.5+\sigma^{*2}(k)} \tag{6.37}$$

The converging value of $\sigma^{*2}(k)$, denoted by σ^{*2}, in (6.36) is found to be

$$\sigma^{*2} = 0.959 \tag{6.38}$$

The estimation problem as defined by (6.33) through (6.35) can also be done optimally by defining a random vector $Z(k)$ [22] as

$$Z(k) = \begin{bmatrix} Z_1(k) \\ Z_2(k) \end{bmatrix} \tag{6.39}$$

with

$$Z_1(k) = x(k-1) \tag{6.40}$$

and

$$Z_2(k) = Z_1(k-1) = x(k-2) \tag{6.41}$$

Using (6.33), (6.40) and (6.41), the elements of Z(k+1) can be written as

$$Z_1(k+1) = x(k) = 0.248Z_1(k)+0.014Z_2(k)+0.969u(k) \tag{6.42}$$

$$Z_2(k+1) = Z_1(k) \tag{6.43}$$

or with the use of (6.39) in a vector form as

$$Z(k+1) = \begin{bmatrix} 0.248 & 0.014 \\ 1 & 0 \end{bmatrix} Z(k) + \begin{bmatrix} 0.969 \\ 0 \end{bmatrix} u(k) \tag{6.44}$$

Similarly the observation at time k-1 can be written as

$$y(k-1) = [1 \quad 0]Z(k)+\gamma(k-1) \tag{6.45}$$

Equations (6.44) and (6.45) have the same form as (1.14) and (1.15), thus the Kalman filtering technique is applicable and the optimal estimates can be obtained from (1.18) through (1.20) with

$$A(k) = \begin{bmatrix} 0.248 & 0.014 \\ 1 & 0 \end{bmatrix}$$

$$B(k) = \begin{bmatrix} 0.969 \\ 0 \end{bmatrix} .$$

$$C(k) = [1 \quad 0]$$

$$D(k)L(k)D'(k) = 1$$

$$K(k) = 1$$

Denoting the convergent value of the variance of the optimal one step prediction value, obtained from implementing (1.19) and (1.20), by σ_K^2, this value was found to be

$$\sigma_K^2 = 0.957 \tag{6.46}$$

Comparison of (6.38) and (6.46) reveals that, if the third decimal place accuracy is negligible, then the proposed method should be used for estimation since this procedure is easily implementable and does not require matrix operations. It should be noted that a second order model was considered in this example, in order to outline the desired properties with the least amount of notational complexity. The computational simplicity of the proposed

estimator will become exceedingly attractive when the autoregressive model is of a much higher order, in which case the optimal procedure would require operations on large dimensional matrices.

6.5 Bounded Multiplicative Noise

When

$$Y(k) = \gamma(k)[x(k)+M(k)] \tag{6.47}$$

and $\gamma(k)$ has a density function $p_\gamma(\gamma(k))$ bounded between $\gamma_1(k)$ and $\gamma_2(k)$ with $0<\gamma_1(k)<\gamma_2(k)$, then [26]

$$p(y(k)\,|\,x(k)) = \begin{cases} \dfrac{p_\gamma\left[\dfrac{y(k)}{x(k)+M(k)}\right]}{|x(k)+M(k)|} & \text{if } \gamma_1(k) \leq \dfrac{y(k)}{x(k)+M(k)} \leq \gamma_2(k) \\ 0 & \text{Otherwise} \end{cases} \tag{6.48}$$

For images, the quantity $x(k)+M(k)$ designates the intensity of the original image at pixel k and hence it is always positive. This is used in reducing condition

$$\gamma_1(k) \leq \frac{y(k)}{x(k)+M(k)} \leq \gamma_2(k)$$

in (6.48) to a condition on $x(k)$ as

$$\frac{y(k)}{\gamma_2(k)} - M(k) \leq x(k) \leq \frac{y(k)}{\gamma_1(k)} - M(k)$$

Thus, (6.48) can equivalently be written as

$$p(y(k)|x(k)) = \begin{cases} \dfrac{p_\gamma\left[\dfrac{y(k)}{x(k)+M(k)}\right]}{x(k)+M(k)} & \text{if } \dfrac{y(k)}{\gamma_2(k)} - M(k) \leq x(k) \leq \dfrac{y(k)}{\gamma_1(k)} - M(k) \\ 0 & \text{Otherwise} \end{cases} \tag{6.49}$$

Substituting (6.49) in (5.6) and (5.9) and dropping subscript k from all variables, then $\hat{x}$ and $\hat{\sigma}^2$ for each k are obtained from

$$\hat{x} = \frac{1}{G}\int_{b_1}^{b_2} \frac{x}{x+M} p_\gamma\left(\frac{y}{x+M}\right) \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right] dx \tag{6.50}$$

$$\hat{\sigma}^2 = \frac{1}{G}\int_{b_1}^{b_2} \frac{(x-\hat{x})^2}{x+M} p_\gamma\left(\frac{y}{x+M}\right) \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right] dx \tag{6.51}$$

where

$$G = \int_{b_1}^{b_2} \frac{1}{x+M} p_\gamma\left(\frac{y}{x+M}\right) \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right] dx \tag{6.52}$$

$$b_1 = \frac{y}{\gamma_2} - M$$

$$b_2 = \frac{y}{\gamma_1} - M \tag{6.53}$$

Thus, computation of $\hat{x}$ and $\hat{\sigma}^2$ involves evaluation of the above definite integrals.

When $\gamma(k)$ is uniform then

$$p(\gamma(k)) = \begin{cases} \dfrac{1}{\gamma_2(k)-\gamma_1(k)} & \text{if } \gamma_1(k) \leq \gamma(k) \leq \gamma_2(k) \\ 0 & \text{Otherwise} \end{cases} \tag{6.54}$$

and (6.50) through (6.52) become

$$\hat{x} = \frac{1}{H}\int_{b_1}^{b_2} \frac{x}{x+M} \exp\left[-\frac{(x-x^*)}{2\sigma^{*2}}\right]dx \tag{6.55}$$

$$\hat{\sigma}^2 = \frac{1}{H}\int_{b_1}^{b_2} \frac{(x-\hat{x})^2}{x+M} \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx \tag{6.56}$$

$$H = \int_{b_1}^{b_2} \frac{1}{x+M} \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx \tag{6.57}$$

The numerical technique of evaluating (6.55) through (6.57) is presented in more detail in Chapter 7.

Figures 6.5 through 6.8 represent the application of the foregoing estimation process to images containing uniform multiplicative noise. Fig. 6.5(a) represents the same image as in Fig. 6.3(a), therefore its autoregressive model was chosen to be the same as in (6.27).

The binary square picture of Fig. 6.6(a) and Fig.

6.7(a) are 32x32 images with a background intensity of 10 and a foreground intensity of 20. The shaded square picture of Fig. 6.8(a) is again a 32x32 image with intensity values of 10, 15 and 20. The three correlation values of the square picture were found to be R(0,0)=18.75, R(0,1)= 16.99 and R(1,0)=16.99 and those of the shaded square were R(o,o)=12.1, R(0,1)=11.08 and R(1,0)=11.08.

A stationary nonseparable correlation function of the form (6.4) was assumed for these images and the following models were obtained:

a) For the square picture the best 4th order model was found to be

$$x(i,j)=0.48x(i,j-1)+0.27x(i-1,j)+0.18x(i-1,j+1)+0.07x(i-1,j+2)+0.46u(i,j) \quad (6.58)$$

b) For the shaded square the best 4th order model was given by

$$x(i,j)=0.48x(i,j-1)+0.27x(i-1,j)+0.18x(i-1,j+1)+0.07x(i-1,j+2)+0.32u(i,j) \quad (6.59)$$

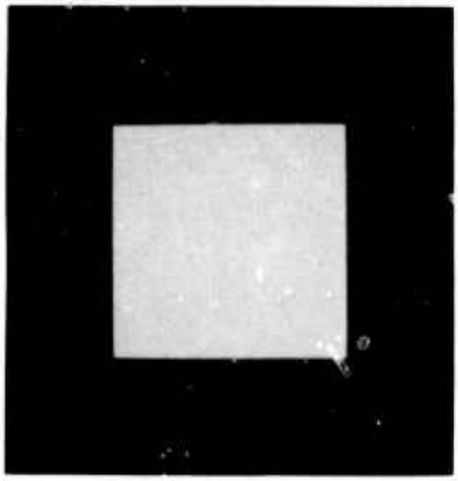
Table 6.1 summarizes the result of the application of the estimation procedure.

(a) Original

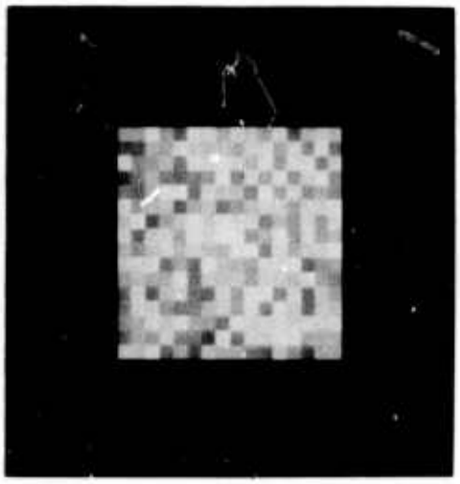
(b) Noisy, noise=0.7-1.0

(c) Estimate

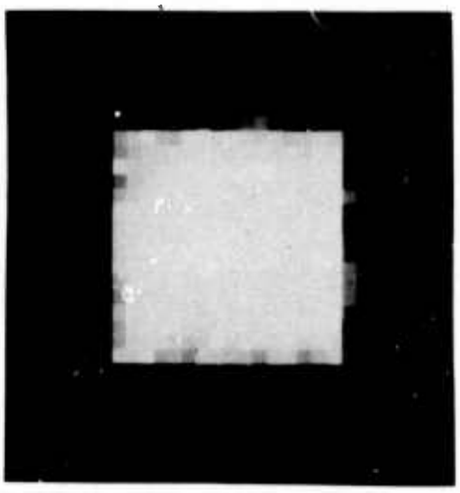
Fig. 6.5 Uniform Multiplicative Noise



(a) Original

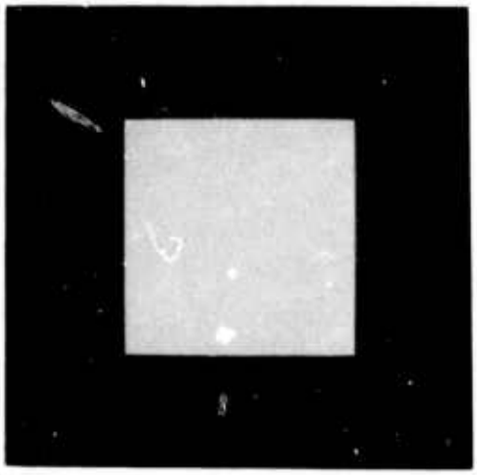


(b) Noisy, noise=0.7-1.0

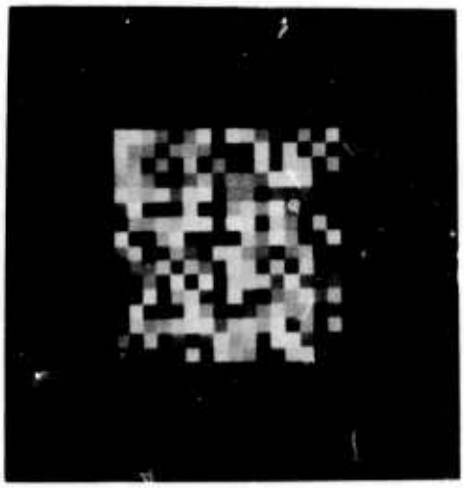


(c) Estimate

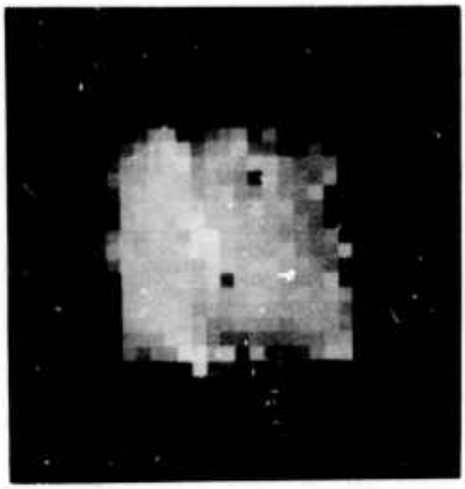
Fig. 6.6 Uniform Multiplicative Noise



(a) Original

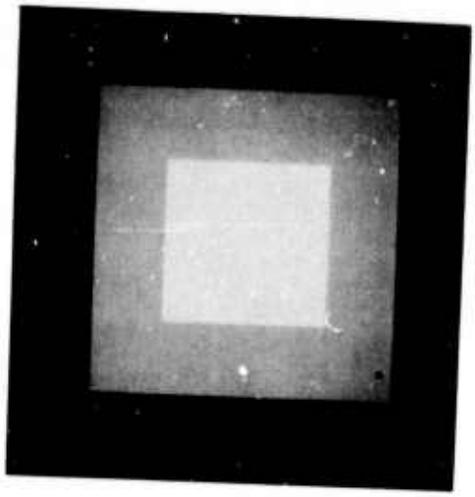


(b) Noisy, noise=0.4-1.0

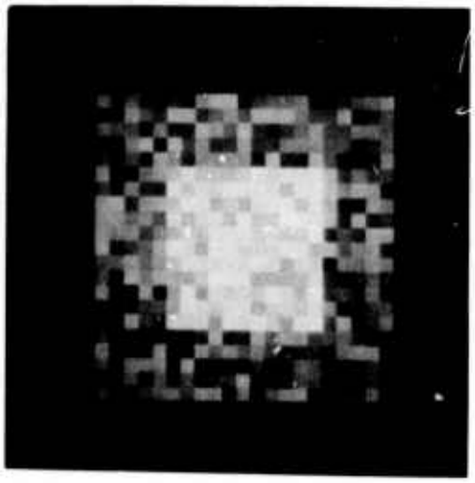


(c) Estimate

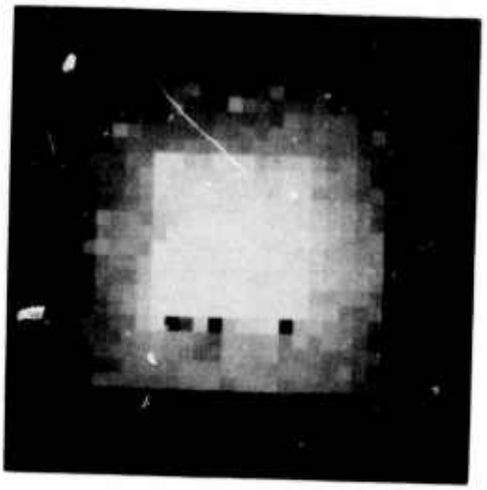
Fig. 6.7 Uniform Multiplicative Noise



(a) Original



(b) Noisy, noise=0.6-1.0



(c) Estimate

Fig. 6.8 Uniform Multiplicative Noise

IMAGE	FIGURE	NOISE BOUNDS	db IMPROVEMENT
Girl	6.5	0.7-1.0	5.48
Square	6.6	0.7-1.0	7.58
Square	6.7	0.4-1.0	7.72
Shaded Square	6.8	0.6-1.0	7.70

Table 6.1

Aside from the quantitative improvement, as indicated in Table 6.1, note the preservation of edges in the estimated images of Fig. 6.5(c) through Fig. 6.8(c), which is a measure of subjective improvement. The responsiveness of the estimator to abrupt pixel to pixel intensity changes is due to the estimator structure of Fig. 2.2, since it is this structure that allows the estimator to respond to observation nonlinearities.

6.6 Observations Containing Additive and Multiplicative Noise Terms

In the case that the observation is of the form

$$y(k) = \gamma(k)[x(k)+M(k)]+v(k) \tag{6.60}$$

then the application of the estimation method requires the

derivation of the density function $p(y(k)|x(k))$. This density function, then can be substituted in (5.6) and (5.9) to determine the estimate and its error variance at each time k.

Assuming that $\gamma(k)$ and $v(k)$ in (6.60) are independent, the conditional density $p(y(k)|x(k))$ can be obtained in terms of the convolution of the probability density functions of $\gamma(k)$ and $v(k)$ [26]. This is achieved by noting that conditioning of the right hand side of (6.60) on $x(k)$ will make the quantity $x(k)+M(k)$ nonrandom, hence [26]

$$p(y(k)|x(k))=\frac{1}{|x(k)+M(k)|}\int_{-\infty}^{\infty} p_{\gamma}\left(\frac{y(k)-\xi}{x(k)+M(k)}\right)p_{v}(\xi)\,d\xi \tag{6.61}$$

In order to outline the procedure for solving (6.61), let us assume that $\gamma(k)$ and $v(k)$ are both uniformly distributed with

$$p_{\gamma}(\gamma(k))=\begin{cases}\dfrac{1}{\gamma_2(k)-\gamma_1(k)} & \text{if } \ 0<\gamma_1(k)\leq\gamma(k)\leq\gamma_2(k)\\ 0 & \text{Otherwise}\end{cases} \tag{6.62}$$

and

$$p_v(v(k)) = \begin{cases} \dfrac{1}{v_2(k)-v_1(k)} & \text{if} \quad v_1(k) \leq v(k) \leq v_2(k) \\ 0 & \text{Otherwise} \end{cases} \tag{6.63}$$

Dropping the time dependence k from all variables and again assuming that $x+M>0$ at each pixel (which is the case for images) then,

$$p_\gamma\left(\frac{y-\xi}{x+M}\right) = \begin{cases} \dfrac{1}{\gamma_2-\gamma_1} & \text{if} \quad \gamma_1 \leq \dfrac{y-\xi}{x+M} \leq \gamma_2 \\ 0 & \text{Otherwise} \end{cases} \tag{6.64}$$

or

$$p_\gamma\left(\frac{y-\xi}{x+M}\right) = \begin{cases} \dfrac{1}{\gamma_2-\gamma_1} & \text{if} \quad y-\gamma_2(x+M) \leq \xi \leq y-\gamma_1(x+M) \\ 0 & \text{Otherwise} \end{cases} \tag{6.65}$$

and

$$p_v(\xi) = \begin{cases} \dfrac{1}{v_2-v_1} & \text{if} \quad v_1 \leq \xi \leq v_2 \\ 0 & \text{Otherwise} \end{cases} \tag{6.66}$$

In order to find the product of $p_\gamma\left(\frac{y-\xi}{x+M}\right)$ and $p_v(\xi)$, it is

required to know whether

$$v_2 - v_1 \leq (x+M)(\gamma_2 - \gamma_1) \tag{6.67}$$

or

$$v_2 - v_1 \geq (x+M)(\gamma_2 - \gamma_1) \tag{6.68}$$

The necessity of this requirement can easily be substantiated by trying to find the density function of the random variable z, where

$$z = a+b$$

where a and b are uniformly distributed between a_1, a_2 and b_1, b_2, respectively.

Assuming v is the dominant noise term in (6.60), i.e. (6.68) hold at each pixel, and by inserting (6.65) and (6.66) in (6.61) and carrying out the integration in terms of ξ, then $p(y|x)$ is given as

$$p(y|x) = \begin{cases} \dfrac{1}{\Delta}\left[\dfrac{v_2 - y}{x+M} - \gamma_2\right] & \text{if } \dfrac{y-v_2}{\gamma_2} - M \leq x \leq \dfrac{y-v_2}{\gamma_1} - M \\ \dfrac{1}{v_2 - v_1} & \text{if } \dfrac{y-v_2}{\gamma_1} - M \leq x \leq \dfrac{y-v_1}{\gamma_2} - M \\ \dfrac{1}{\Delta}\left[\dfrac{y-v}{x+M} - \gamma_1\right] & \text{if } \dfrac{y-v}{\gamma_2} - M \leq x \leq \dfrac{y-v_1}{\gamma_1} - M \\ 0 & \text{Otherwise} \end{cases} \tag{6.69}$$

where

$$\Delta = (\gamma_2 - \gamma_1)(v_2 - v_1)$$

Relation (6.69) can be substituted in (5.6) and (5.9) to obtain the pertinent filtering equations.

CHAPTER 7

COMPUTATIONAL ASPECTS

The estimation procedure developed in Chapters 2 through 5 has been mainly motivated by the ease of the implementation and computational considerations. Since the method consists of the three parts, namely modeling, prediction and filtering, the computational requirements of each will be discussed separately.

To find the model of the random process x(k), a series of systems of linear equations must be solved. This does not hamper the speed or running time of the estimation process, since the modeling procedure is implemented prior to the reception of any observations. The numerical methods, used in this work for solving each system of linear equation, is just one of the many standard available methods [40]-[42].

In the actual operation of the estimator, the prediction scheme requires a minimal amount of computation. This only involves arithmetic operations to find $x^*(k)$ and the upper bound of the variance $\sigma^{*2}(k)$, equations (4.22) and (4.33). Depending on whether or not the observation noise is additive-Gaussian, the filtering step may represent the bulk of the computational requirements of the entire esti-

mation procedure. In the linear case, again, finding $\hat{x}(k)$ and $\hat{\sigma}^2(k)$ involves simple arithmetic operations and the computational aspects of the general case is considered below.

7.1 Nonlinear Filtering

Dropping the time dependence k (to reduce notational complexity) from all variables, the pertinent filtering equations will become (equations (5.6) and (5.9)):

$$\hat{x} = \frac{\int x p(y|x) \exp[-\frac{(x-x^*)^2}{2\sigma^{*2}}]dx}{\int p(y|x) \exp[-\frac{(x-x^*)^2}{2\sigma^{*2}}]dx} \tag{7.1}$$

$$\hat{\sigma}^2 = \frac{\int (x-\hat{x})^2 p(y|x) \exp[-\frac{(x-x^*)^2}{2\sigma^{*2}}]dx}{\int p(y|x) \exp[-\frac{(x-x^*)^2}{2\sigma^{*2}}]dx} \tag{7.2}$$

In the general case, where (7.1) and (7.2) do not have a closed analytic form, the integrals should be evaluated numerically. Although (7.1) and (7.2) suggest that three numerical integrations are required at each time k,

the following expansion of (7.2) can help in reducing this number to two. An example of this is the case of the uniform multiplicative observation noise of Chapter 6, derived in Section 7.2.

$$\hat{\sigma}^2=\frac{1}{\int P(y|x)\exp[-\frac{(x-x^*)^2}{2\sigma^{*2}}]}\{\int[x^2-2x\hat{x}+\hat{x}^2]p(y|x)\exp[-\frac{(x-x^*)^2}{2\sigma^{*2}}]dx\}$$

$$=\frac{\int x^2 p(y|x)\exp[-\frac{(x-x^*)^2}{2\sigma^{*2}}]dx}{\int p(y|x)\exp[-\frac{(x-x^*)^2}{2\sigma^{*2}}]dx}-\hat{x}^2 \qquad (7.3)$$

7.2 Uniform Multiplicative Noise

Having developed the estimator for the case of bounded multiplicative noise, the detailed expansion and the methods of evaluation of the integrals in (7.1) and (7.2) for the particular case of uniform multiplicative observation noise is presented in this section. Therefore, when

$$p(y|x) = \frac{1}{(\gamma_2-\gamma_1)(x+M)}$$

then (7.1) and (7.2) become

$$\hat{x} = \frac{\int_a^b \frac{x}{x+M} \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right] dx}{\int_a^b \frac{1}{x-M} \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right] dx} \tag{7.4}$$

$$\hat{\sigma}^2 = \frac{\int_a^b \frac{x^2}{x+M} \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right] dx}{\int_a^b \frac{1}{x+M} \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right] dx} - \hat{x}^2 \tag{7.5}$$

where

$$a = \frac{y}{\gamma_2} - M$$

$$b = \frac{y}{\gamma_1} - M \tag{7.6}$$

Let

$$I_1 = \int_a^b \frac{1}{x+M} \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right] dx \tag{7.7}$$

Letting

$$z = x+M \tag{7.8}$$

results in

$$dz = dx \tag{7.9}$$

so

$$\int_a^b \frac{x}{x+M}\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx = \int_{a+M}^{b+M} \frac{z-M}{z}\exp\left[-\frac{(z-M-x^*)^2}{2\sigma^{*2}}\right]dz$$

$$=\int_a^b \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx - M\int_a^b \frac{1}{x+M}\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx \tag{7.10}$$

and

$$\int_a^b \frac{x^2}{x+M}\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx = \int_{a+M}^{b+M} \left[z-2M+\frac{M^2}{z}\right]\exp\left[-\frac{(z-M-x^*)^2}{2\sigma^{*2}}\right]dz$$

$$=\int_a^b (x+M)\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx - 2M\int_a^b \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx$$

$$+M^2\int_a^b \frac{1}{x+M}\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx \tag{7.11}$$

Let

$$I_2 = \int_a^b \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx \tag{7.12}$$

Using (7.7) and (7.12), then relations (7.10) and (7.11) become

$$\int_a^b \frac{x}{x+M}\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx = I_2 - MI_1 \tag{7.13}$$

$$\int_a^b \frac{x^2}{x+M}\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx = \int_a^b x\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx - MI_2 + M^2 I_1 \tag{7.14}$$

Furthermore, with a change of the variable

$$z = x - x^* \qquad dz = dx \tag{7.15}$$

then

$$\int_a^b x\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx$$

$$= \int_{a-x^*}^{b-x^*} z\exp\left[-\frac{z^2}{2\sigma^{*2}}\right]dz + x^*\int_a^b \exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx \tag{7.16}$$

But

$$\int_{a-x^*}^{b-x^*} z\exp\left[-\frac{z^2}{2\sigma^{*2}}\right]dz = \sigma^{*2}\left[\exp\left[-\frac{(a-x^*)^2}{2\sigma^{*2}}\right] - \exp\left[-\frac{(b-x^*)^2}{2\sigma^{*2}}\right]\right] \tag{7.17}$$

Letting

$$Q = \sigma^{*2}\left\{\exp\left[-\frac{(a-x^*)^2}{2\sigma^{*2}}\right] - \exp\left[-\frac{(b-x^*)^2}{2\sigma^{*2}}\right]\right\} \tag{7.18}$$

then using (7.16) relation (7.14) can be written as

$$\int_a^b \frac{x^2}{x+M}\exp\left[-\frac{(x-x^*)^2}{2\sigma^{*2}}\right]dx = Q+(x^*-M)I_2+M^2I_1 \qquad (7.19)$$

Substituting (7.13) and (7.19) in (7.4) and (7.5) results in

$$\hat{x} = \frac{I_2}{I_1}-M \qquad (7.20)$$

$$\hat{\sigma}^2 = \frac{Q+(x^*-M)I_2}{I_1}+M^2-\hat{x}^2 \qquad (7.21)$$

where I_1, I_2 and Q are given by (7.7), (7.12) and (7.18), respectively.

Thus, finding $\hat{x}$ and $\hat{\sigma}^2$ requires carrying out the two integrations I_1 and I_2. Since, the functions involved in these integrations are continuous and well behaved, the Romberg's method of numerical integration was used [42] to determine the estimates of 32x32 images in Fig. 6.6(c) through Fig. 6.8(c). This integration method is reasonably fast and requires minimal coding. For the images in Fig. 6.6 through Fig. 6.8, the total CPU time of setting up the image, the observation and obtaining the estimated image, was about 2.5 minutes on PDP-10 computer.

Although the CPU time of 2.5 minutes for the 32x32 images of Figs. 6.6 through 6.8 is reasonable, this time

increases linearly with the increase in the size of the image. Such an increase, in general, may not be tolerable in practical situations. Therefore, for images of the size 256x256 (such as those in Fig. 6.5) or higher, the application of the method may warrant some approximation of the integrals I_1 and I_2. For small values of the prediction variance σ^{*2}, a crude approximation on the exponential form of p(x(k)) can be made by expanding the exponential in its series equivalent and retaining the first two terms. This results in representing p(x(k)) at each time k by

$$p(x) \simeq \frac{2\sigma^*}{\sqrt{2\pi}}\left[\frac{1}{2\sigma^{*2}+(x-x^*)^2}\right] \tag{7.22}$$

With this it can be shown that $\hat{x}$ and $\hat{\sigma}^2$ are obtained from

$$\hat{x} = \frac{B}{A} \tag{7.23}$$

$$\hat{\sigma}^2 = \frac{1}{A}[Q-2B\hat{x}+A\hat{x}^2] \tag{7.24}$$

where

$$A = \int_a^b p(y|x)p(x)dx$$

$$= \left[\frac{1}{2q_1}\ln\frac{(x+M)^2}{2\sigma^{*2}+(x-x^*)^2} - \frac{q_2}{q_1\sqrt{q}}\tan^{-1}\frac{2(x+M)+q_2}{\sqrt{q}}\right]_{x=a}^{x=b} \tag{7.25}$$

$$B = \int_a^b xp(y|x)p(x)dx = \left[\frac{2}{\sqrt{q}}\tan^{-1}\frac{2(x+M)+q_2}{\sqrt{q}}\right]_{x=a}^{x=b} - MA \tag{7.26}$$

$$Q = \int_a^b x^2 p(y|x)p(x)dx$$

$$= M^2A + \left[\frac{1}{2}\ln(2\sigma^{*2}+(x-x^*)^2) - (2M+\frac{q_2}{2})(\frac{2}{\sqrt{q}})\tan^{-1}\frac{2x+2M+q_2}{\sqrt{q}}\right]_{x=a}^{x=b} \tag{7.27}$$

with

$$q_1 = 2\sigma^{*2}+(M+x^*)^2$$

$$q_2 = -2(M+x^*)$$

$$q = 4[2\sigma^{*2}+(M+x^*)^2]-4(M+x^*)^2$$

and

$$a = \frac{y}{\gamma_2}-M$$

$$b = \frac{y}{\gamma_1}-M$$

CHAPTER 8

A DISCUSSION ON THE OPTIMALITY OF THE PROCEDURE, EXTENSIONS AND CONCLUSIONS

8.1 Discussion on Optimality

Due to the various restrictions imposed on the estimator in Chapter 2, the estimation procedure as developed in Chapters 2 through 5 is, in general, suboptimal. An exception to this is the case considered in the following theorem.

Theorem 8.1: If the observation noise is additive-Gaussian and the process $x(k)$ is a first order normal Markov process, then for given initial conditions, $\hat{x}(0)$ and $\hat{\sigma}^2(0)$, the procedure is optimal.

Proof: In this case, the model of the process $x(k)$ and that of observation are given by

$$x(k) = \beta x(k)+Bu(k) \tag{8.1}$$

$$y(k) = x(k)+\gamma(k) \tag{8.2}$$

In order to show the optimality of the procedure in this particular case, let us assume that the estimate $\hat{x}(k-1)$

and its error variance $\hat{\sigma}^2(k-1)$ at time k-1 have been found optimally, i.e.

$$\hat{x}(k-1) = Ex(k-1)|y(1),....,y(k-1) \tag{8.3}$$

$$\hat{\sigma}^2(k-1) = E[x(k-1)-\hat{x}(k-1)]^2|y(1),...,y(k-1) \tag{8.4}$$

The predicted value $x^*(k)$ and its error variance $\sigma^{*2}(k)$ are obtained from

$$x^*(k) = \beta\hat{x}(k-1) \tag{8.5}$$

$$\sigma^{*2}(k) = B^2+\beta^2\hat{\sigma}^2(k-1) \tag{8.6}$$

But from (8.1)

$$\begin{aligned} Ex(k)|y(1),...,y(k-1) &= \beta Ex(k-1)|y(1),...,y(k-1) \\ &\quad + BEu(k)|y(1),...,y(k-1) \\ &= \beta Ex(k-1)|y(1),...,y(k-1) \end{aligned} \tag{8.7}$$

where by using (8.3) and (8.5), it follows that

$$Ex(k)|y(1),...,y(k-1) = \beta\hat{x}(k-1) = x^*(k) \tag{8.8}$$

and similarly

$$\begin{aligned} &E[x(k)-Ex(k)|y(1),...,y(k-1)]^2|y(1),...,y(k-1) \\ &= B^2+\beta^2\hat{\sigma}^2(k-1) \end{aligned} \tag{8.9}$$

But $x^*(k)$ and $\sigma^{*2}(k)$ being optimal quantities makes the chosen a posteriori density of x(k) in Section 5.1 to be

$$p(x(k)) = p(x(k)\,|\,y(1),\ldots,y(k-1)) \tag{8.10}$$

and the substitution of (8.10) in (5.6) and (5.9) results in

$$\hat{x}(k)=\frac{\int x(k)p(y(k)\,|\,x(k))p(x(k)\,|\,y(1),\ldots,y(k-1))dx(k)}{\int p(y(k)\,|\,x(k))p(x(k)\,|\,y(1),\ldots,y(k-1))dx(k)} \tag{8.11}$$

$$\hat{\sigma}^2(k)=\frac{\int [x(k)-\hat{x}(k)]^2 p(y(k)\,|\,x(k))p(x(k)\,|\,y(1),..,y(k-1))dx(k)}{\int p(y(k)\,|\,x(k))p(x(k)\,|\,y(1),..,y(k-1))dx(k)} \tag{8.12}$$

A comparison of (8.11) and (8.12) with (2.10) and (2.13) indicates that $\hat{x}(k)$ and $\hat{\sigma}^2(k)$ are optimal quantities.

The proof of the theorem follows by applying the above argument for k=1,2,....

In the derivation of the general estimation procedure, the value of the prediction error variance was approximated by its upper bound in Section 4.3. As stated previously, this approximation was introduced in order to maintain computational simplicity of the overall algorithm, but since the error variance is an indication of the uncertainty of the value of each estimate then the effect of such an approximation on the values of the subsequent error variances should be investigated. In the following, the effect of such an approximation on the error variance of the linear estimator of Section 6.3 is analyzed and it is

shown that the introduction of the upper bound causes all future computed error variances to be larger than the actual variances.

At each time k, the actual unavailable prediction error variance $\sigma_p^2(k)$ is approximated by its upper bound $\sigma^{*2}(k)$ where (equations (4.32) and (4.33))

$$\sigma^{*2}(k) = B^2 + \left[\sum_{i=1}^{M} |\beta_i| \hat{\sigma}(k-I_i)\right]^2 \tag{8.13}$$

with

$$\sigma_p^2(k) \leq \sigma^{*2}(k) \tag{8.14}$$

The computed error variance of the filtered quantity is given by

$$\hat{\sigma}^2(k) = \frac{\sigma^{*2}(k)\sigma_\gamma^2(k)}{\sigma^{*2}(k)+\sigma_\gamma^2(k)} \tag{8.15}$$

Assuming $\sigma_p^2(k)$ is available at time k then the actual variance $\sigma_a^2(k)$ would be

$$\sigma_a^2(k) = \frac{\sigma_p^2(k)\sigma_\gamma^2(k)}{\sigma_p^2(k)+\sigma_\gamma^2(k)} \tag{8.16}$$

But due to (8.14) and $\sigma_\gamma^2(k)$ being the observation noise variance, then it follows from (8.15) and (8.16) that

$$\sigma_a^2(k) \leq \hat{\sigma}^2(k) \tag{8.17}$$

Recursively, the substitution of (8.17) in (8.13) shows that the computed prediction error variance, $\sigma^{*2}(k)$, is always larger than the actual variance. Therefore, for the linear estimator of Section 6.3, all computed variances are the upper bounds for the actual variances.

8.2 Discussion of Nonlinear Case

In the light of the discussions in the previous section, it is expected that the computed and actual variances of the nonlinear estimator behave similar to those of the linear case. The difficulty in showing this is in having $\hat{\sigma}^2(k)$ satisfy an integral relation of (5.9). It is suspected, however, that the relationship between the computed and the actual variances of the linear case does not hold true for all nonlinear observations while for a certain class of nonlinearities the same results may exist.

This subject requires a more exact and rigorous analysis and is an excellent candidate for topic of future investigation.

8.3 Extensions and Topics for Further Research

The estimation method of Chapters 2 through 5 has been derived based on the restriction of the linear prediction.

But if a nonlinear model of the random process x(k) exists then the prediction procedure could be replaced by the appropriate nonlinear one. In general, if the form of nonlinearity is known then the modeling procedure can be modified in order to determine a nonlinear model of the process x(k). This modification is conciptually simple and would require a priori decision on the order and the degree of the nonlinearity. For example, for a given order M and the 2nd degree polynomial nonlinearity the modeling procedure can be used in obtaining a model of the form

$$x(k) = \sum_{i=1}^{M} \beta_i x^2(k-I_i) + \sum_{i=1}^{M} \beta_{M+i} x(k-I_{M+i})$$

$$+ \beta_{2M+1}\beta_{2M+2} x(k-I_{2M+1})x(k-I_{2M+2}) + \ldots + Bu(k) \qquad (8.18)$$

Of course, (8.18) is a particular form of nonlinearity and it should be noted that in order to apply the modeling procedure the a priori statistics must contain up to and including the 3rd moment of the random process x(k).

8.4 Conclusion

This dissertation has examined and expanded the subject of general image estimation. An estimation procedure

has been developed with a particular emphasis on the multiplicative and non-Gaussian observation noise. An analysis of the optimal discrete filter has been presented to show that the principle of the estimation at each time k consists of a one step prediction and filtering operations. Conceptually these operations are shown to closely resemble a learning procedure based on the past information and the optimal use of the present information. Accordingly, a recursive estimation procedure is derived such that the logic of the estimation principle is maintained and, at the same time, the procedure is implementable.

Although the derivation and application of the method has primarily been presented in terms of the two dimensional processes, the procedure is directly applicable to one dimensional problems. A particularly important and practical feature of the estimation method is the method's independence of the analytic representation of the a priori correlation function. An equally significant value of the procedure, in this respect, is its applicability to problems where only partial values of the correlation function are specified.

The estimation method is demonstrated to be applicable to a broad class of observation systems and, in fact, the degree of ease or difficulty in applying the method to general nonlinear systems is directly related to the ease

APPENDIX A

A DISCUSSION ON ERROR VARIANCE

For a given set of observation $y(1),\ldots,y(k)$, the estimate $\hat{x}(k)$ of the process $x(k)$ at time k is, in general, some function f of the observations which can be written as

$$\hat{x}(k) = f(y(1),\ldots,y(k)) \tag{A.1}$$

Accordingly, the two error variances, $\hat{\sigma}_1^2(k)$ and $\hat{\sigma}_2^2(k)$, of $\hat{x}(k)$ can be defined as

$$\hat{\sigma}_1^2(k) = E\{[x(k)-\hat{x}(k)]^2\} \tag{A.2}$$

$$\hat{\sigma}_2^2(k) = E\{[x(k)-\hat{x}(k)]^2 | y(1),\ldots,y(k)\} \tag{A.3}$$

From (A.1) and (A.2), it follows that

$$\hat{\sigma}_1^2(k) = E_Y\hat{\sigma}_2^2(k) \tag{A.4}$$

where E_Y represents the expectation with respect to $y(1),\ldots,y(k)$. Letting

$$e(k) = x(k)-\hat{x}(k) \tag{A.5}$$

then $e(k)$ is a random variable and (A.2) and (A.3) can equivalently be written as

$$\hat{\sigma}_1^2(k) = \tag{A.6}$$

$$\int \cdots \int e^2(k)p(e(k),y(1),\ldots,y(k))de(k)dy(1),\ldots,dy(k)$$

$$\hat{\sigma}_2^2(k) = \int e^2(k)p(e(k)|Y(1),\ldots,y(k))de(k) \tag{A.7}$$

In the linear optimal case of Section 1.3, the existence of the orthogonality principle (equation (1.12)) along with the zero mean and the Gaussian nature of the random variables result in the statistical independence of e(k) and y(1),....,y(k) [26] which reduces (A.6) and (A.7) to

$$\hat{\sigma}_1^2(k) = \hat{\sigma}_2^2(k) = \int e^2(k)p(e(k))de(k) \tag{A.8}$$

but in general

$$\hat{\sigma}_1^2(k) \neq \hat{\sigma}_2^2(k) \tag{A.9}$$

For a given sample function of the observation, $\hat{\sigma}_2^2(k)$ in (A.3) specifies the amount of variation (uncertainty) associated with choosing $\hat{x}(k)$ as the estimate of x(k) at time k. The quantity $\hat{\sigma}_1^2(k)$, on the other hand, represents the ensemble average of the variance when finding $\hat{x}(k)$ from (A.1) and for all possible values of y(1),....,y(k). Since in this dissertation, our interest is in a particular sample function of the observation (the degraded image), quantity (A.3) is taken to represent the variance.

APPENDIX B

AN IMPROVEMENT MEASURE FOR ESTIMATED IMAGES

Letting $b(i,j)$, $y(i,j)$ and $\hat{b}(i,j)$ denote, in order the intensities of the original image, the noise corrupted image and the estimated image at pixel (i,j), then the two quantities σ_n^2 and σ_e^2 are computed as

$$\sigma_n^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} [b(i,j)-y(i,j)]^2 \qquad (B.1)$$

$$\sigma_e^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} [b(i,j)-\hat{b}(i,j)]^2 \qquad (B.2)$$

where N is the size of the image.

Viewing σ_n^2 and σ_e^2 as the average error variance associated with the observation and the estimate, respectively, the amount of improvement in db is obtained from

$$\text{db improvement} = 10 \log_{10} \left[\frac{\sigma_n^2}{\sigma_e^2} \right] \qquad (B.3)$$

REFERENCES

[1] W. K. Pratt, Digital Image Processing, to be published.

[2] L. E. Franks, "A Model for the Random Video Process," Bell Syst. Tech. J., pp. 609-630, April 1966.

[3] T. S. Huang, "Subjective Effect of Two Dimensional Pictorial Noise," IEEE Trans. Infor. Theory, Vol. IT-11, pp. 43-54, Jan. 1965.

[4] E. R. Kretzmer, "Statistics of Television Signals," Bell Syst. Tech. J., pp. 751-763, July 1952.

[5] V. M. Moroz, "Description of Images of Visual Objects Through Correlation Functions," Automatika, Vol. 14, pp. 80-82, Oct. 1969.

[6] E. Wong, "Two Dimensional Random Fields and Representation of Images," SIAM J. Appl. Math., Vol. 16, pp. 756-770, July 1968.

[7] J. W. Woods, "Two Dimensional Discrete Markovian Fields," IEEE Tran. Infor. Theory, Vol. IT-18, pp. 232-240, March 1972.

[8] A. Habibi and P. A. Wintz, "Image Coding by Linear Transformation and Block Quantization," IEEE Trans. Commun. Technol., Vol. COM-19, pp. 50-62, Feb. 1971.

[9] D. C. Fraser and J. E. Potter, "The Optimum Linear Smoother as a Combination of Two Linear Filters," IEEE Trans. Automat. Control, Vol. AC-14, pp. 387-390, Aug. 1969.

[10] C. W. Helstrom, "Image Restoration by the Method of Least Squares," J. Opt. Soc. Amer., Vol. 57, pp. 297-303, March 1967.

[11] B. R. Frieden, "Optimum Nonlinear Processing of Noisy Images," J. Opt. Soc. Amer., Vol. 58, pp. 1272-1275, Sept. 1968.

[12] D. Slepian, "Linear Least Square Filtering of Distorted Images," J. Opt. Soc. Amer., Vol. 57, pp. 918-922, July 1967.

[13] A. Habibi, "Linear Transformation for Encoding Two Dimensional Sources," School of Elec. Eng. Purdue Univ., Tech. Rep. TR-EE 70-2, Appendix E, June 1970.

[14] M. Marden, The Geometry of the Zeros of a Polynomial in a Complex Variable, American Mathematical Soc., N. Y., 1949.

[15] K. S. Miller, "A Note on Stochastic Difference Equations," Ann. Math. Statist., Vol. 39, No. 1, pp. 270-271, 1968.

[16] H. Akaike, "Fitting Autoregressive Models for Prediction," Ann. Inst. Statist. Math., Vol. 21, pp. 243-247, 1969.

[17] H. Akaike, "Power Spectrum Estimation Through Autoregressive Model Fitting," Ann. Inst. Statist. Math., Vol. 21, pp. 407-419, 1969.

[18] M. Marcus and H. Minc, A Survey of Matrix Theory and Matrix Inequalities, Allyn and Bacon, Boston, 1964.

[19] R. V. Hogg and A. T. Craig, Introduction to Mathematical Statistics, MacMillan Co., Second Edition, N. Y., 1965.

[20] A. H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, N. Y., 1970.

[21] P. B. Liebelt, An Introduction to Optimal Estimation, Addison-Wesley, Reading, Mass., 1967.

[22] N. E. Nahi, Estimation Theory and Application, J. Wiley and Sons, N. Y., 1969.

[23] R. E. Kalman and R. C. Bucy, "New Results in Linear Filtering and Prediction Theory," ASME Trans. (J. Basic Eng.), Vol. 83D, pp. 95-108, March 1961.

[24] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problem," ASME Trans. (J. Basic Eng.), Vol. 82D, pp. 35-45, March 1960.

[25] W. K. Pratt, "Generalized Wiener Filtering Computation Techniques," IEEE Trans. Computers, Vol. C-21, No. 7, pp. 636-641, July 1972.

[26] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, N. Y., 1965.

[27] B. R. Hunt, "The Application of Constrained Least Squares Estimation to Image Restoration by Digital Computers," IEEE Trans. Computers, Vol. C-22, No. 9, pp. 805-812, Sept. 1973.

[28] R. S. Bucy, M. J. Merritt and D. S. Miller, "Hybrid Computer Synthesis of Optimal Discrete Nonlinear Filter," Univ. South. Calif., Tech. Rep. No. 71-38, Sept. 1971.

[29] J. Ting-Ho Lo, "Finite Dimensional Sensor Orbits and Optimal Nonlinear Filtering," IEEE Trans. Infor. Theory, Vol. IT-18, No. 5, Sept. 1972.

[30] C. R. Rao, Linear Statistical Inference and Its Application, J. Wiley and Sons, N. Y., 1965.

[31] A. V. Oppenheim, R. W. Schafer and T. G. Stockham Jr., "Nonlinear Filtering of Multiplied and Convolved Signals," Proceedings IEEE, Vol. 56, No. 8, pp. 1264-1291, Aug. 1968.

[32] A. Habibi, "Two Dimensional Bayesian Estimate of Images," Proceedings IEEE, Vol. 60, No. 7, pp. 878-883, July 1972.

[33] N. E- Nahi and T. Assefi, "Bayesian Recursive Image Estimation," IEEE Trans. Computers, Vol. C-12, No. 7, pp. 734-738, July 1972.

[34] N. E. Nahi and C. Franco, "Recursive Image Enhancement Vector Processing," IEEE Trans. Commun., Vol. COM-21, No. 4, pp. 305-311, April 1973.

[35] N. E. Nahi, "Role of Recursive Estimation in Statistical Image Enhancement," Proceedings IEEE, Vol. 60, No. 7, pp. 872-877, July 1972.

[36] S. R. Powell and L. M. Silverman, "Modeling of Two Dimensional Random Fields with Application to Image Restoration," IEEE Trans. Automat. Control, Vol. AC-19, No. 1, pp. 9-13, Feb. 1974.

[37] A. K. Jain and E. Angel, "Image Restoration, Modeling and Reduction of Dimensionality," IEEE Trans. Computers, Vol. C-23, No. 5, pp. 470-477, May 1974.

[38] D. Middleton, An Introduction to Statistical Communication Theory, McGraw-Hill, N. Y., 1960.

[39] J. B. Thomas, An Introduction to Statistical Communication Theory, J. Wiley and Sons, N. Y., 1969.

[40] R. S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, N. J., 1962.

[41] E. K. Blum, Numerical Analysis and Computation: Theory and Practice, Addison-Wesley, Reading, Mass., 1972.

[42] T. R. McCalla, Introduction to Numerical Methods and Fortran Programming, J. Wiley and Sons, N. Y., 1967.