INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Université d'Ottawa · University of Ottawa

Modeling and Estimation Using Maximum Entropy and Minimum Mean Squared Criteria Based On Partial and Noisy Observations

by

Afshin David

A thesis Submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirement for degree of

Doctoral of Philosophy

in Electrical and Computer Engineering

Ottawa-Carleton Institute of Electrical and Computer Engineering

School of Information Technology Engineering

Faculty of Engineering

University of Ottawa

Ottawa, Ontario, K1N 6N5

I hereby declare that I am the sole author of this thesis.

I authorize the University of Ottawa to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Afshin David

I further authorize the University of Ottawa to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Afshin David

The University of Ottawa requires the signatures of all persons using or photocopying this thesis. Please, sign below and give address and date.

Abstract

Modeling a process, based on fully or partially corrupted observations, is of great interest in areas of signal reconstruction and compression. It is particularly of interest to ensure that the quality of the model is solely based on observed data and no other implicit assumptions. To this end firstly, we propose a globally convergent algorithm, Modified Equation Error Output Error (MEEOE), for modeling Infinite Impulse Response (IIR) filters, where we have assumed full observation. Secondly, we offer another modeling scheme, Maximum Entropy Kalman Filter (MEKF), based on partial observations and a given set of constraints, that ensures the construction of the most appropriate model, *i.e.* solely based on observed data and the a priori constraints. In both cases we have assumed corrupted observations. The optimality of MEEOE under conditions of insufficient modeling and colored input is shown. Application of MEKF to image compression and reconstruction is also demonstrated.

Acknowledgments

I would like to thank my thesis supervisor, Dr. Tyseer Aboulnasr, for her infinite patience that is only matched by her immense passion for simplicity and elegance.

Table of Figures

List of Tables

List of Abbreviations

- DFT Discrete time Fourier Transform
- EEOE Equation Error Output Error
- FIR Finite Impulse Response
- IIR Infinite Impulse Response
- MAP Maximum A Posteriori
- ME Maximum Entropy
- MEEOE Modified EEOE
- MEKF Maximum Entropy Kalman Filter
- ML Maximum Likelihood
- MSE Mean Squared Error
- PLR Pseudo Linear Regression
- SPR Strictly Positive Real
- SRPE Simplified Recursive Predictive Error

Chapter 1 Introduction

In this thesis we address and offer solutions for two important issues in the field of Digital Signal Processing (DSP): modeling and reconstruction of linear processes based on fully or partially observed signals. These issues are crucial to applications where real-time decision making in the presence of limited bandwidth or incomplete observation is required. A typical example of such an application is digital image communication in web browsing applications where a highly complicated sequence, *i.e.* an image, must be transferred to the desired destination, over a band-limited network, in a timely manner. Solutions to this problem may be approached in two ways. Firstly, an image may be considered as the outcome of a system in response to a given input. The parameters of this system then may be transmitted in place of the original image. Secondly, only the relevant information within an image is sent across the network. In the former the hope is that the model will require a substantially lower bandwidth to be transmitted compared to the image. The image will be reconstructed on the receiver side as the output of the model to a given input. The latter approach transmits as much relevant information (based on certain criteria) as allowed by the application. On the receiver side, based on partial information received, an image is reconstructed to satisfy the same criteria as the one chosen by the transmitter.

1.1 Traditional methods

Current techniques, which allow for modeling of linear events, may be classified as Finite Impulse Response (FIR) or Infinite Impulse Response (IIR). It is well known that adaptive FIR filters have a unique stable optimum whereas, adaptive IIR filters most

often do not possess such properties. However, IIR filters are able to efficiently (in terms of computational and temporal complexity) model systems with sharp frequency peaks. In Chapter 3 we offer a novel adaptive IIR algorithm that, based on our simulations, ensures optimal solution regardless of the initial point.

Under the Bayesian framework, an optimal reconstruction may be achieved for a given cost function and joint probability distribution of the ideal signal and the observed one. By far the most widely used cost function is the Mean Squared Error (MSE). Other notable cost functions are the zero-one cost function, used in Maximum A Posteriori (MAP) estimation, and the entropy cost function. Depending on the assumed form of the joint probability distribution, different categories of approximation, such as Maximum Likelihood (ML), may be realized. A variety of numerical methods like Expectation Maximization (EM) or Newton-Raphson may be employed to construct appropriate numerical solutions. Aside from implementation difficulties associated with these methods notably stability, convergence, and uniqueness, the major obstacle in applying them readily is that a priori knowledge of the joint probability distribution is needed. Except for extremely trivial cases such knowledge may not be available theoretically or empirically, especially in case of images. In Chapter 4, we offer a modeling and reconstruction method based on the principle of Maximum Entropy using only partially observed data, with application to image compression, estimation and reconstruction.

1.2 Thesis Organization

Chapter 2 of this document provides some background on the traditional modeling methods and reconstruction criteria along with their respective strengths and weaknesses. In Chapter 3, we propose and give a detailed description of a novel adaptive IIR filtering

namely, Modified Equation Error and Output Error (MEEOE). Chapter 4 introduces a novel method for signal compression and reconstruction namely, Maximum Entropy Kalman Filter (MEKF). A detailed description of application and implementation of the MEKF in image compression and reconstruction is also provided in Chapter 4. Finally, the last chapter discusses the ongoing work and possible elaboration with respect to both the MEEOE and MEKF. Further possible improvements and applications will also be described in this chapter.

1.3 Main Contributions

The main contributions of this thesis are as follows:

1- Analysis and development of a combined Equation Error and Output Error IIR filter.

- Proof of global convergence for a single pole model
- Development of a fully adaptive algorithm
- Evaluation of limitations of other methods

2- Analysis and development of a signal compression and reconstruction algorithm, applicable to images, based on Maximum Entropy criterion

- Identification of the Kalman filter parameters, using Maximum Entropy criterion
- Extension of the solution to 2-D signals, with finite domain
- Development of a image compression algorithm using 2-D MEKF
- Development of a 1-D and 2-D signal reconstruction algorithm using MEKF

Chapter 2 System Identification

Our aim in this chapter is to provide sufficient background in system modeling. We will do so under the framework of linear structures assuming either a full set of observable data, for process modeling, or a partially observable data, for signal reconstruction.

2.1 Process modeling

The complexity and variety of many current DSP applications necessitate the development of generalized and powerful modeling methods. IIR filters are the most logical step in this direction, as compared to FIR filters. Their attractiveness is due to the efficient modeling, in terms of temporal and implementation complexity, of resonating systems. As with many nonlinear problems however, closed form solutions for IIR filters are either too complicated or undesirable, due to dynamics of the application. As such, adaptive methods offer the only viable solution in IIR based prediction and modeling applications. However, as we shall see proliferation of adaptive methods, and consequently IIR filters, are hampered by major obstacles such as stability monitoring, multiple optimal solutions and bias solutions.

2.1.1 Means and methods

The terminating criterion for adaptive processes may be utilized for two purposes. Firstly, it may be used to interpret the adaptation process as optimization and secondly, to obtain procedural means for the adaptation process. With respect to the former, any of the statistical methods such as Simulated Annealing [40], Golden Section Search [56], Expectation Maximization [10][31][30] and Steepest Descent [8][11][49] methods may be employed. However, only the latter approach lends itself readily to real-time

applications. Thus the procedural means is often aimed at following the gradient of the cost function. Two groups of algorithms may be considered in adaptive IIR filters. Firstly, there are those algorithms inspired by traditional FIR and linear regression methods, most notably Equation Error (EE) and Output Error (OE) [23] [47], and their composite variants [34][33][37][25][48]. Secondly, there are methods inspired by control and stability theory such as Instrumental Variable [49], Hyper Stability [8], Extended State Kalman Filter [49] and Unit Norm [42][45][26] approaches. Even though, the former is often the special case of the latter (at least theoretically) however, the practitioners of the field seem to heavily favor the former group, simply due to their ease of implementation. Hence, recently there has been an impetus to resolve many of disadvantages associated with EE and OE methods through composite approaches. In the next section we will describe in detail the EE, OE and some notable composite methods.

2.1.2 Linear Regression Methods

Figure 2.1 depicts a general adaptive filter setup, in the context of system identification. Figure 2.2 and Figure 2.3 respectively, show the setup for EE and OE algorithms. In what follows we shall assume that the unknown system itself is an IIR filter, with filter coefficients A_{ideal} and B_{ideal}.

2.1.3 Equation Error

As it can be seen from Figure 2.2, EE is a direct extension of FIR filter. Indeed, there is no feedback associated with the filter during the adaptation process. However, it should be noted that once the adaptation is terminated, the respective coefficients will be copied to an IIR filter. The MSE is the most frequently used measure in adaptive filtering. Eq.

(2.1) gives the error expression. Assuming a zero mean white additive noise, a mutually independent zero mean white input and sufficient modeling, Eq. (2.2) (MSE expression) results.

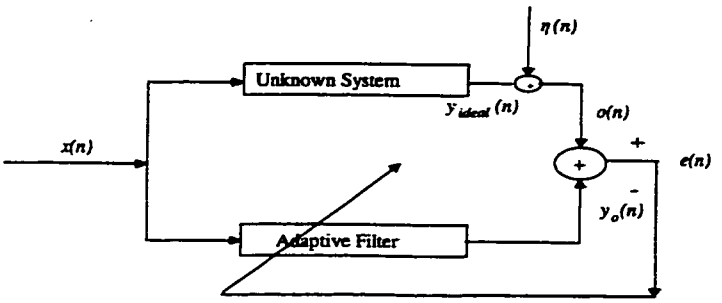


Figure 2.1 General Adaptive System Setup

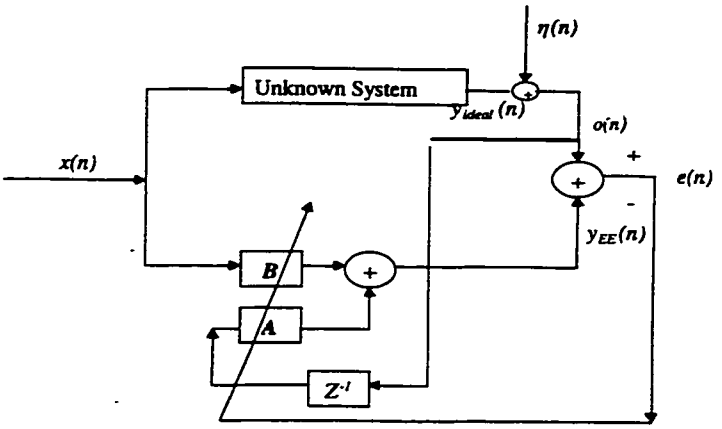


Figure 2.2 Equation Error Setup

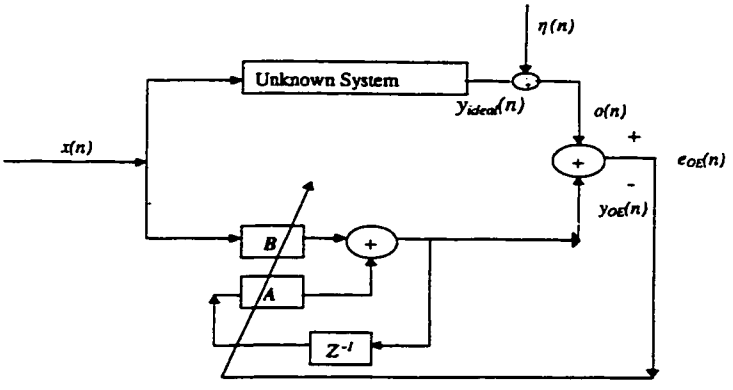


Figure 2.3 Output Error Setup

$$e_{EE}(n)=\left[A_{ideal}(q^{-1})-A_{EE}(n,q^{-1})\right]y_{ideal}(n)+\left[B_{ideal}(q^{-1})-B_{EE}(n,q^{-1})\right]x(n) \quad (2.1)$$
$$+\left[1-A_{EE}(n,q^{-1})\right]\eta(n)$$

where $A_{ideal}(q^{-1})=\sum_{i=1}^{N_A-1}A_{ideal_i}q^{-i}$, $B_{EE}(q^{-1})=\sum_{i=0}^{N_B-1}B_{EE_i}(n)q^{-i}$,*etc.*,

q^{-1} is the one step delay, N_A and N_B are the feed-back and feed-forward tap lengths, respectively. Taking the mean square of Eq. (2.1) Eq. (2.2) is obtained.

$$E\left[e_{EE}^2(n)\right]=\sum_{i=1}^{N_A-1}E\left[\tilde{A}_i^2\right]\sigma_y^2+\sum_{j=1}^{N_B-1}E\left[\tilde{B}_i^2\right]\sigma_x^2+\sum_{i=1}^{N_A-1}E\left[A_{EE_i}^2\right]\sigma_\eta^2+\sigma_\eta^2+$$
$$\sum_{v=1}^{N_A-1}E\left[\tilde{A}_v^2\right]\sum_{u=1,u\neq v}^{N_A-1}E\left[\tilde{A}_u^2\right]R_{yy}(v-u)+\sum_{q=1}^{N_A-1}E\left[\tilde{A}_q^2\right]\sum_{p=0}^{N_B-1}E\left[\tilde{B}_u^2\right]R_{xy}(q-p) \quad (2.2)$$

where $\tilde{A}_l=A_{ideal_l}-A_{EE_l}$. A_{ideal}, A_{EE} and B_{ideal}, B_{EE} are the feed-forward and feedback coefficients of the ideal system and adaptive system, respectively. σ_f is the standard deviation of the process f, $R_{fg}(.)$ is the cross-correlation matrix of processes f and g. $E[.]$ is the expectation operator.

It is noted that in the absence of additive noise and exact modeling, the right hand side of Eq.(2.2) is minimized, when the following holds.

$$A_{ideal}=A_{EE} \text{ and } B_{ideal}=B_{EE} \quad (2.3)$$

However, in the presence of the third term, the additive noise, minimization of Eq. (2.2) will lead to solutions other than Eq. (2.3).

2.1.4 Output Error

With respect to Figure 2.3 it is observed that, unlike EE formulation, due to feedback loop in OE adaptation the processes may become unstable and multiple minima may also

exist. This is readily observed by noting the error expression of Eq. (2.5), where we note that $e(n)$ is an ARMA process, which depends on the past values of $\boldsymbol{A_{OE}}$ and $\boldsymbol{B_{OE}}$.

$$\begin{aligned} e_{OE}(n) &= o(n) - y_{OE}(n) \\ &= o(n) - \sum_{i=1}^{N_A-1} A_{OE_i} y_{OE}(n-i) - \sum_{j=1}^{N_B-1} B_{OE_j} x(n-j) \\ &= o(n) + \sum_{i=1}^{N_A-1} A_{OE_i} e_{OE}(n-i) - \sum_{k=1}^{N_A-1} A_{OE_k} o(n-k) - \sum_{j=1}^{N_B-1} B_{OE_j} x(n-j) \\ &= e_{EE}(n) + \sum_{i=1}^{N_A-1} A_{OE_i} e_{OE}(n-i) \end{aligned} \tag{2.4}$$

where the last equality was obtained using Eq. (2.1). Assuming that the solution has converged to the global minimum, the expression for the MSE of OE is obtained as:

$$\begin{aligned} \sigma_{OE}^2 &= \sigma_{EE}^2 + \sum_{i=1}^{N_A-1} A_{OEi}^2 \sigma_{OE}^2 + 2\sum_{j=1}^{N_A-1} A_{OEj} E[e_{OE}(n-j) e_{EE}(n)] \\ &= \sigma_{EE}^2 + \sum_{i=1}^{N_A-1} A_{OEi}^2 \sigma_{OE}^2 + 2\sum_{j=1}^{N_A-1} A_{OEj} E\left[e_{OE}(n-j) \left[e_{OE}(n) - \sum_{k=1}^{N_A-1} A_{OEk} e_{OE}(n-k) \right] \right] \\ &= \sigma_{EE}^2 - \sum_{i=1}^{N_A-1} A_{OEi}^2 \sigma_{OE}^2 \\ \sigma_{OE}^2 &= \frac{\sigma_{EE}^2}{\left[1 + \sum_{i=1}^{N_A-1} A_{OEi}^2 \right]} \end{aligned} \tag{2.5}$$

where we have utilized Eq. (2.4) and assumed that $e_{OE}(.)$ at optimum is a white sequence. Assuming sufficient modeling and using Eq. (2.2) the following expression is obtained:

$$\begin{aligned} \sigma_{OE}^2 &= \frac{\left[1 + \sum_{i=1}^{N_A-1} A_{OEi}^2 \right] \sigma_\eta^2}{\left[1 + \sum_{i=1}^{N_A-1} A_{OEi}^2 \right]} \\ \sigma_{OE}^2 &= \sigma_\eta^2 \end{aligned} \tag{2.6}$$

Noting the sufficient modeling assumption and hence the uniqueness of the global optimum [2], Eq. (2.6) states that at optimum the only source of error, in MSE sense, is

the observation noise. This in turn implies that the estimation at global minimum is unbiased, in the MSE sense.

2.1.5 Combined methods

The motivation behind these methods is to combine the advantages of EE and OE methods and to overcome problems such as biased estimation and multiple minima. There are two main approaches in these methods. Firstly, there are those that combine the outputs of EE and OE and secondly, there are those methods that combine the error signals of EE and OE. Although, the latter can often be described in terms of the former, in this presentation we will treat them separately, by describing one representative from each class.

2.1.5.1 Combined output

As a representative of this group we describe the Composite Regressor Algorithm (CRA) [25]. As shown in Figure 2.4 the essence of CRA is that of combining the outputs of the unknown process and OE via a scalar weight.

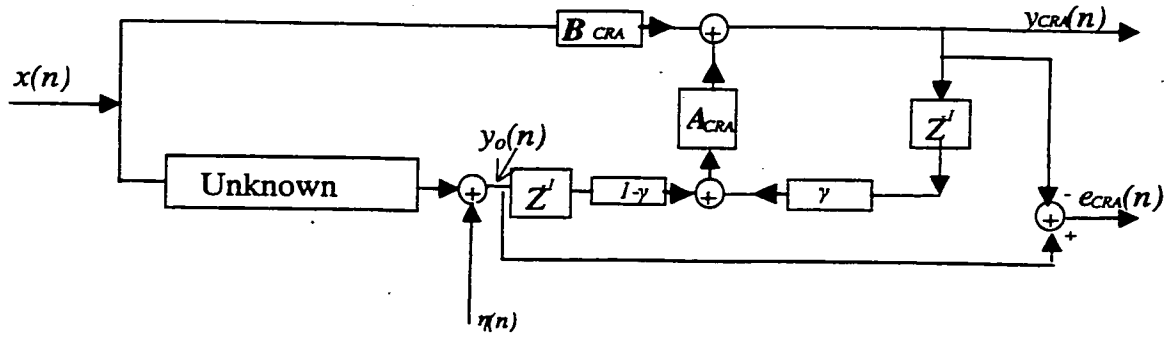


Figure 2.4 Composite Regressor Algorithm Setup

Furthermore, the coefficients A_{CRA} and B_{CRA} are updated through e_{CRA} signal. Note that for $\gamma=0$, CRA is identical to EE and for $\gamma=1$ it is identical to OE setup. The adaptation of the coefficients is given by Eq. (2.7) [25].

$$\begin{aligned}
&\theta(n)=\theta(n-1)+\frac{\mu(n)\varphi(n-1)e_{CRA}(n-1)}{1+\mu(n)\varphi^T(n-1)\varphi(n-1)}\\
&\theta(n)=\left[a_{CRA_1}(n),\cdots,a_{CRA_N}(n),b_{CRA_1}(n),\cdots,b_{CRA_M}(n)\right]^T\\
&\varphi(n)=\begin{bmatrix}\gamma[y_{CRA}(n-1),\cdots,y_{CRA}(n-N)]+\\(1-\gamma)[y_O(n-1),\cdots,y_O(n-N)]\\,x(n),\cdots,x(n-M)\end{bmatrix}^T\\
&e_{CRA}(n)=\zeta(n)+\gamma A_{CRA}(n)e_{CRA}(n-1)\\
&\zeta(n)=y_O(n)-\theta(n)\begin{bmatrix}y_O(n-1),\cdots,y_O(n-N),\\x(n),\cdots,x(n-M)\end{bmatrix}
\end{aligned} \tag{2.7}$$

Eq. (2.7) implies minimization of $E[e^2_{CRA}]$ as the error measure, where a linear relationship between e_{CRA} and adaptive coefficients is assumed. It is noted that for $\gamma=1$ the adaptation is identical to the Modified Output Error (MOE) [34] or Pseudo Linear Regression (PLR) algorithm [47], where the following approximation, Eq. (2.8), of output gradient with respect to adaptive coefficients is assumed.

$$\begin{aligned}
&\frac{\partial y_{PLR}(n)}{\partial a_i(n)}=y_{PLR}(n-i)\\
&\frac{\partial y_{PLR}(n)}{\partial b_j(n)}=x(n-j)
\end{aligned} \tag{2.8}$$

Although, CRA offers an intermediate performance between EE and OE formulations, it suffers from two disadvantages. Firstly, as described above and noted in [25], parameter γ does not ensure both unbiased and unimodal error surface, due to the recursive relationship of e_{CRA}. Secondly, there is no adaptive solution for varying γ optimally.

Indeed, due to the bias in $E[e^2_{CRA}]$, for any $\gamma \neq 1$, adaptations with respect to this measure will also lead to biased solutions.

2.1.5.2 Combined error

To partially overcome the difficulties of the CRA approach, in particular the coefficient bias (due to the recursive relationship in e_{CRA}), the Combined Error (CE) method based on combination of both outputs and errors of EE and OE has been suggested [37][34]. The update procedure for this method is shown in Eq. (2.9).

$$\begin{aligned}
&\theta(n) = \theta(n-1) + \mu(n)\varphi(n-1)e_{CE}(n-1) \\
&\theta(n) = \left[a_{CE_1}(n), \cdots, a_{CE_N}(n), b_{CE_1}(n), \cdots, b_{CE_M}(n)\right]^T \\
&\varphi(n) = \begin{bmatrix} \gamma[y_{OE}(n-1), \cdots, y_{OE}(n-N)] + \\ (1-\gamma)[y_{EE}(n-1), \cdots, y_{EE}(n-N)] \\ , x(n), \cdots, x(n-M) \end{bmatrix}^T \\
&e_{CE}(n-1) = \gamma e_{OE}(n-1) + (1-\gamma)e_{EE}(n-1)
\end{aligned} \tag{2.9}$$

It is noted that e_{CE} does not have a recursive relationship. As such, in adaptive methods that commence with $\gamma \neq 0$ and terminate with $\gamma \neq 1$ the initial bias, due to EE formulation, will not propagate. However, it is noted that for $\gamma \neq 1$ the formulation is identical to MOE or PLR. Hence, if $1\text{-}A_{ideal}$ is not Strictly Positive Real (SPR) the adaptive coefficients may converge to arbitrary points on the error surface. As noted in [47], this is a particularly common problem in system identification of oversampled systems, where the poles of such systems tend to be outside of SPR region.

2.2 Summary

As stated previously the potential gains in employing IIR filters are often offset by the undesirable characteristics of the MSE surface. We have briefly described some of these

obstacles along with traditional solutions. However, as noted in previous sections none of these solutions offer a robust convergence under different operating environments. In the next chapter we offer a globally convergent adaptive method with simple means of implementation and analysis.

2.3 Signal reconstruction

The Bayesian estimator, *i.e.* an estimator that assigns a cost for every pair of ideal and estimated signals, defines the best approximation of a signal as a signal with the minimum average cost [51]. It can be shown that, under the quadratic cost function assumption, the estimated signal is the conditional mean of a signal given an observed signal [13]. That is

$$\begin{aligned} \hat{x}_{MS} &= \int_{-\infty}^{+\infty} x f_{X|Z}(x \mid z)\,dx \\ &= \frac{1}{f_Z(z)} \int_{-\infty}^{+\infty} x f_{Z|X}(z \mid x) f_X(x)\,dx \end{aligned} \tag{2.10}$$

where f_g is the pdf of g,
Z is the observed signal, and
$\hat{x}_{MS}$ is the optimum MSE approximation of x, the ideal signal.

Depending on the form of the a priori density, f_x, three types of approximation are realized. If f_x is known a priori the approximation is referred to as Maximum A Posteriori (MAP). If no a priori knowledge is available, *i.e.* f_x is assumed uniform, the approximation is referred to as Maximum Likelihood (ML), and if the a priori density is a Gibbs distribution [54] then the approximation is referred to as Maximum Entropy (ME).

2.4 Implementation

Although Eq. (2.10) expresses the optimum approximation as a simple function of the conditional probability density $f_{X|Z}$ however, in reality determination of $f_{X|Z}$ is often impossible. Thus, simplifying assumptions are made to allow for a feasible implementation of Eq. (2.10). The following sections describe the traditional approaches as they pertain to digital image processing [19][43]. A common model that is often used, with regard to the degradation process, is as follows:

$$z = s(x) + n \tag{2.11}$$

where z is the observed image,
x is the ideal image,
s is an arbitrary degradation function, and
n is the additive noise.

2.4.1 Maximum A Posteriori (MAP)

Using Eq. (2.11) as the observation process and assuming f_X and f_N are multivariate Gaussian density functions, the a posteriori density, $f_{X|Z}$, is maximized. It can be shown [15] that this is equivalent to the minimization of the Eq. (2.12).

$$J = (x - \hat{x})^t C_{xx}^{-1} (x - \hat{x}) + (z - s(x))^t C_{nn}^{-1} (z - s(x)) \tag{2.12}$$

where C_{yy} is the covariance matrix of y and $\hat{x}$ is the estimate of x.

An iterative solution based on steepest descent, as shown in Eq. (2.13), may be devised to obtain a near optimum solution to Eq. (2.12).

$$\hat{x}_{k+1} = \hat{x}_k + \alpha \nabla J(\hat{x}_k) \tag{2.13}$$

where α is the step size, and $\nabla J(\hat{x}_k)$ is the gradient of the cost function with respect to $\hat{x}_k$.

The major difficulty in implementation of MAP based algorithm is the evaluation of $\nabla J(\hat{x}_k)$, especially when the degradation function, s, is nonlinear.

2.4.2 Maximum likelihood (ML)

In the absence of any a priori knowledge of f_X, the best course of action is to maximize the likelihood function $f_{Z|X}$. Under the same assumption as above, it can be shown [15] that the maximization of Eq. (2.10) is equivalent to the minimization of the following cost function.

$$J = (z - s(x))^t C_{nn}^{-1} (z - s(x)) \tag{2.14}$$

It is noted that if the degradation function, s, is linear this formulation is equivalent to the deterministic least-squares method [15]. Hence, any one of LMS or RLS algorithms may be used to identify the optimum approximation. However, for nonlinear degradation, algorithms based on Newton-Raphson or Expectation-Maximization (EM) methods may be employed [10][31][30].

2.4.3 Maximum Entropy (ME)

Although the notion of entropy is often associated with the information content or the uncertainty of a signal [7][5] however, entropy, in image processing, is often employed as a measure of the smoothness [15][57][24][36]. That is the following measure is maximized to smooth the effects of observation noise and degradation.

$$\begin{aligned} H &= H_x + \rho H_{n'} \\ n' &= n + B\big|_{n' \geq 0} \end{aligned} \tag{2.15}$$

where H_x is the entropy of the ideal image,
$H_{n'}$ is the entropy of the modified noise, such that it is always positive,
ρ is a weighting scalar, and entropy of an NxM image x is defined as

$$H_x = -\sum_{i=1}^{M}\sum_{j=1}^{N} x(i,j) \ln x(i,j) \tag{2.16}$$

subject to the following constraint, for a given x_{sum}.

$$x_{sum} = \sum_{i=1}^{M}\sum_{j=1}^{N} x(i,j) \tag{2.17}$$

It is easy to show [43] that, when Eq. (2.16) is maximized, the neighboring pixels tend to have a same intensity value. Newton-Raphson based algorithms are often employed to maximize Eq. (2.15).

2.5 Maximum Entropy Criterion as Information Measure

In this section we describe the ME criterion in the context of information measure, rather than smoothness measure of the previous section. To this end we define the following terms. Let *result* be the final status of a single trial, and *outcome* to refer to the whole experiment [21][44]. As an example consider the experiment of tossing a coin twice. The following table is obtained.

Outcome	Probability Distribution
HH	P(X=H)=1, P(X=T)=0
HT	P(X=H)=0.5, P(X=T)=0.5
TH	P(X=H)=0.5, P(X=T)=0.5
TT	P(X=H)=0, P(X=T)=1

H= Head, T= Tail

Each row refers to an outcome, and each Head or Tail refers to a result. Let us further assume that we wish to consider only the results where the first toss is a tail, that is the subclass *C*, the shaded area.

The ME estimation may be interpreted as that of a statistical inference, stated as follows: Given no other information, constraint, which probability distribution will best represent a typical result? The answer to this question can be quantified by the following theorem, Concentration theorem [21][44].

Concentration Theorem:

Consider a fraction F of the outcomes in a class C, with entropy in the following range.

$$H_{max} - \Delta H \le H(f_1 \cdots f_n) \le H_{max} \tag{2.18}$$

where there are n possible results, f_i is the probability of the result i, H_{max} is the maximum entropy of the outcome in class C and N trials have been conducted. It can be shown that asymptotically, $2N\Delta H$ is distributed over class C as chi-squared, χ_k^2, with $k=n-m-1$ degrees of freedom, independent of the nature of the constraints. That is

$$2N\Delta H = \chi_k^2(1-F) \tag{2.19}$$

where m is the number of linearly independent constraints, not including the constraint that the sum of the probabilities should add up to unity.

It should be noted that Eq. (2.19) is combinatorial expressing the number of ways that the class C may be realized. In order for this expression to become a probability statement each outcome in class C is assumed to have the same probability.

To appreciate the implication of the Concentration theorem, we describe the Loaded Dice example of [44].

2.5.1 Loaded dice example

Consider the following experimental result. After 1000 tosses the average number of spots on a dice is 4.5, rather than 3.5. The question is, in absence of any other information, what is the best approximation of P, the probability distribution of the dice.

2.5.1.1 ME Solution

It can be shown that given the following constraints, *i.e. m=1*

$$\begin{aligned} \sum_{i=1}^{6} P(X=i) &= 1 \\ \sum_{i=1}^{6} i\,P(X=i) &= 4.5 \end{aligned} \tag{2.20}$$

the ME solution yields the following distribution [44].

$$\begin{aligned} P(X=1) &= 0.0543 \\ P(X=2) &= 0.0788 \\ P(X=3) &= 0.1142 \\ P(X=4) &= 0.1654 \\ P(X=5) &= 0.2398 \\ P(X=6) &= 0.3475 \end{aligned} \tag{2.21}$$

This distribution has the maximum entropy, $H_{max}=1.61358$, amongst all other distributions satisfying Eq. (2.20). Employing Eq. (2.19) and Eq. (2.18) for $F=0.95$ and $N=1000$ the following inequality is obtained.

$$1.609 \le H(f_1 \cdots f_6) \le H_{max} \tag{2.22}$$

Eq. (2.22) asserts that out of 1000 trials, 950 of them will have a probability distribution that satisfy Eq. (2.21) and Eq. (2.22).

Similarly, it can be shown that 99.99 percent of all allowable results have entropy in the following range.

$$1.602 \le H(f_1 \cdots f_6) \le H_{max} \tag{2.23}$$

This means that 999.9 of the results satisfy both Eq. (2.21) and Eq. (2.23).

With respect to these observations it is clear that probability distributions with greatest multiplicity, number of possible ways that they may be realized, are highly concentrated around the distribution with the highest entropy.

2.5.1.2 Alternative Solution

Now consider the following alternative solution, which satisfies the same constraints as Eq. (2.20).

$$P(X=i)=\binom{5}{i-1}0.7^{i-1}0.3^{6-i} \quad i=\{1\cdots6\} \tag{2.24}$$

The probability distribution of Eq. (2.24) has entropy $H=1.4136=H_{max}-0.200$. Using Eq. (2.19) it can be shown [44] that this probability distribution describes results with multiplicity, *i.e.* F, of less than 1 in 10^{83} , in 1000 tosses.

2.6 Closing remarks

It is clear, with respect to the discussions of the previous section, that inferences based on any criterion other than ME will result in discarding a vast majority of possible solutions. It should also be noted that Concentration theorem provides a vehicle to test hypotheses on the observed signal. That is, for a given number of assumed constraints, if the prediction by ME criterion tends to be greatly different from that of the observed signal then one can confidently state that there must exist one or more constraints with salient effect on the behavior of the observed signal. Otherwise, the current set of constraints provides a satisfactory description of the signal.

3 Adaptive IIR Filtering

An open issue in Adaptive IIR Filtering (AIF) is that of convergence to a global minimum in the presence of observation noise when the system is insufficiently modeled [14]. It is well known [23][47] that algorithms based on Equation Error (EE) formulation contain a single minimum that may be biased whereas, algorithms based on Output Error (OE) ensure the existence of an unbiased global minimum in presence of local minima. Recently, there have been several attempts to combine these formulations [34] in order to ensure the existence and uniqueness of an unbiased minimum. Works presented here, Equation Error Output Error (EEOE) and Modified EEOE (MEEOE) are such attempts in the context of system identification. Although the formulation of EEOE did not achieve the desired outcome and was later found out to be similar to that proposed in [25] however, the exploration of its limitations led to a superior algorithm namely, MEEOE.

3.1 Equation Error Output Error (EEOE)

With respect to Figure 3.1 the output of EEOE can be written as follows.

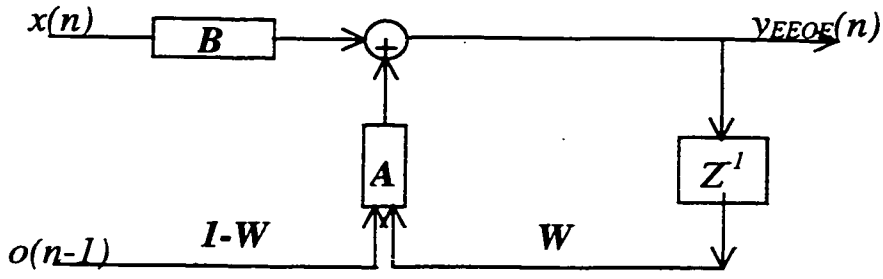


Figure 3.1 EEOE Structure

$$
\begin{aligned}
y_{EEOE}(n) &= \sum_{j=0}^{N_B-1} B_j(n)x(n-j) + \sum_{i=1}^{N_A-1} A_i(n)\begin{bmatrix} W_i(n)y_{EEOE}(n-i) + \\ (1-W_i(n))o(n-i) \end{bmatrix} \\
&= \sum_{j=0}^{N_B-1} B_j(n)x(n-j) + \sum_{i=1}^{N_A-1} A_i(n)W_i(n)[y_{EEOE}(n-i) - o(n-i)] + \\
&\qquad + \sum_{k=1}^{N_A-1} A_k(n)o(n-k) \\
y_{EEOE}(n) &= \sum_{j=0}^{N_B-1} B_j(n)x(n-j) - \sum_{i=1}^{N_A-1} A_i(n)W_i(n)e(n-i) + \sum_{k=1}^{N_A-1} A_k(n)o(n-k)
\end{aligned}
\tag{3.1}
$$

where, N_A and N_B are the number of taps in filters $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively,
A_j and B_j are the jth element of filters $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively,
$\boldsymbol{W}$ is the weighting filter vector with tap values W_j in the range 0-1,
o is the observed signal (noise induced ideal signal),
$x(n)$ is a zero mean stochastic sequence,
y_{EEOE} is the output of the Equation Error Output Error filter, and
$e(n) = o(n) - y_{EEOE}(n)$.

For simplicity it is assumed that filter $\boldsymbol{W}$ has the same length as the filter $\boldsymbol{A}$. With respect to Eq. (3.1) the following expression for MSE of EEOE may be obtained.

$$
\begin{aligned}
E[e^2(n)] &= E[o(n)^2] + E\left[\sum_{j=0}^{N_B-1}\sum_{j'=0}^{N_B-1} B_j(n)B_{j'}(n)x(n-j)x(n-j')\right] \\
&+ E\left[\sum_{i=1}^{N_A-1}\sum_{i'=1}^{N_A-1} A_i(n)A_{i'}(n)W_i(n)W_{i'}(n)e(n-i)e(n-i')\right] + E\left[\sum_{k=1}^{N_A-1}\sum_{k'=1}^{N_A-1} A_k(n)A_{k'}(n)o(n-k)o(n-k')\right] \\
&- 2E\left[\sum_{j=0}^{N_B-1}\sum_{i=1}^{N_A-1} B_j(n)A_i(n)W_i(n)x(n-j)e(n-i)\right] + 2E\left[\sum_{j=0}^{N_B-1}\sum_{k=1}^{N_A-1} B_j(n)A_k(n)x(n-j)o(n-k)\right] \\
&- 2E\left[\sum_{i=1}^{N_A-1}\sum_{k=1}^{N_A-1} A_i(n)W_i(n)A_k(n)o(n-k)e(n-i)\right] - 2E\left[\sum_{j=0}^{N_B-1} B_j(n)x(n-j)o(n)\right] \\
&+ 2E\left[\sum_{i=1}^{N_A-1} A_i(n)W_i(n)e(n-i)o(n)\right] - 2E\left[\sum_{k=1}^{N_A-1} A_k(n)o(n-k)o(n)\right]
\end{aligned}
\tag{3.2}
$$

where E is the expectation operator.
Assuming that coefficients of AIF vary much slower than the input signal $x(n)$ and $x(n)$ is stationary, Eq. (3.2) may be simplified as follows.

$$\sigma_e^2(n)=\sigma_o^2(n)+\sum_{j=0}^{N_B-1}\sum_{j'=0}^{N_B-1}B_j(n)B_{j'}(n)R_{xx}(j-j')+\sum_{i=1}^{N_A-1}\sum_{i'=1}^{N_A-1}A_i(n)A_{i'}(n)W_i(n)W_{i'}(n)R_{ee}(i-i')$$
$$+\sum_{k=1}^{N_A-1}\sum_{k'=1}^{N_A-1}A_k(n)A_{k'}(n)R_{oo}(k-k')-2\sum_{j=0}^{N_B-1}\sum_{i=1}^{N_A-1}B_j(n)A_i(n)W_i(n)R_{xe}(j-i)$$
$$+2\sum_{j=0}^{N_B-1}\sum_{k=1}^{N_A-1}B_j(n)A_k(n)R_{xo}(j-k)-2\sum_{i=1}^{N_A-1}\sum_{k=1}^{N_A-1}A_i(n)W_i(n)A_k(n)R_{oe}(i-k) \quad (3.3)$$
$$+2\sum_{i=1}^{N_A-1}A_i(n)W_i(n)R_{oe}(i)-2\sum_{j=0}^{N_B-1}B_j(n)R_{xo}(j)-2\sum_{k=1}^{N_A-1}A_k(n)R_{oo}(k)$$

where R_{kj} is the cross-correlation matrix of processes k and j.σ_x is the standard deviation of the process x.

With respect to Eq. (3.1) and the definition of *e(n)*, the sixth and seventh terms of Eq. (3.3) may be expressed as follows

$$2\sum_{j=0}^{N_B-1}\sum_{k=1}^{N_A-1}B_j(n)A_k(n)R_{xo}(j-k)-2\sum_{i=1}^{N_A-1}\sum_{k=1}^{N_A-1}A_i(n)W_i(n)A_k(n)R_{oe}(i-k)=$$
$$2E\left\{\sum_{k=1}^{N_A-1}A_k(n)o(n-k)\left[\sum_{j=0}^{N_B-1}B_j(n)x(n-j)-\sum_{i=1}^{N_A-1}A_i(n)W_i(n)e(n-i)\right]\right\}= \quad (3.4)$$
$$2\sum_{k=1}^{N_A-1}A_k(n)R_{y_{EEOE}o}(k)-2\sum_{j=1}^{N_A-1}\sum_{k=1}^{N_A-1}A_j(n)A_k(n)R_{oo}(j-k)$$

Similarly, the last three terms of Eq. (3.3) may be rewritten as follows.

$$2\sum_{i=1}^{N_A-1}A_i(n)W_i(n)R_{oe}(i)-2\sum_{j=0}^{N_B-1}B_j(n)R_{xo}(j)-2\sum_{k=1}^{N_A-1}A_k(n)R_{oo}(k)=-2R_{y_{EEOE}o}(0) \quad (3.5)$$

Using Eqs (3.4) and (3.5), assuming that at optimum the error is orthogonal to input [16] and the error sequence is white, Eq. (3.3) simplifies to Eq. (3.6).

$$\sigma_e^2=\sigma_o^2+\sum_{j=0}^{N_B-1}\sum_{j'=0}^{N_B-1}B_jB_{j'}R_{xx}(j-j')+\sum_{i=1}^{N_A-1}A_i^2W_i^2\sigma_e^2$$
$$-\sum_{k=1}^{N_A-1}\sum_{k'=1}^{N_A-1}A_k(n)A_{k'}(n)R_{oo}(k-k')+2\sum_{k=1}^{N_A-1}A_kR_{y_{EEOE}o}(k)-2R_{y_{EEOE}o}(0) \quad (3.6)$$

It is noted that Eq. (3.6), except for the fifth and sixth terms, is a sum of quadratic functions, with respect to coefficients of $\boldsymbol{A}$, $\boldsymbol{B}$ filters and $\boldsymbol{W}$ vector. Furthermore, due to Eq. (3.1), $R_{y_{EEOE^o}}(.)$ will always be non-zero for all values of $\boldsymbol{W}$, other than 1.

3.1.1 The EEOE Adaptive algorithm

In this section we describe an update algorithm for the EEOE structure that attempts to minimize Eq. (3.3) while satisfying the following constraints.

1. $W_i(.) \geq 0 \ \forall \ i$
2. $W_i(.) \leq 1 \ \forall \ i$
3. $W_i(n) \geq W_i(n-1) \ \forall \ n$

The first two constraints ensure that the solution at each time instance is a combination EE and OE formulation. The last constraint ensures that at each time instance the coefficients of the IIR filter move toward the OE solution, which is an unbiased solution.

3.1.1.1 IIR Coefficients Adaptation

Differentiating Eq. (3.3) with respect to coefficients of filters $\boldsymbol{A}$ and $\boldsymbol{B}$ the following is obtained.

$$\begin{aligned}
\frac{\partial e^2(n)}{\partial B_j(n)} &= 2e(n)\frac{\partial e(n)}{\partial B_j(n)} = 2e(n)\left[-x(n-j)+\sum_{i=1}^{N_A-1} A_i(n)W_i(n)\frac{\partial e(n-i)}{\partial B_j(n)}\right] \\
\frac{\partial e^2(n)}{\partial A_k(n)} &= 2e(n)\frac{\partial e(n)}{\partial A_k(n)} = 2e(n)\left[-o(n-k)+W_k(n)e(n-k)+\sum_{i=1}^{N_A-1} A_i(n)W_i(n)\frac{\partial e(n-i)}{\partial A_k(n)}\right]
\end{aligned} \tag{3.7}$$

Using the definition of $e(n)$, Eq. (3.7) can be rewritten as

$$\begin{aligned}
\frac{\partial e^2(n)}{\partial B_j(n)} &= -2e(n)\left[x(n-j)+\sum_{i=1}^{N_A-1} A_i(n)W_i(n)\frac{\partial y_{EEOE}(n-i)}{\partial B_j(n)}\right] \\
\frac{\partial e^2(n)}{\partial A_k(n)} &= -2e(n)\left[o(n-k)-W_k(n)e(n-k)+\sum_{i=1}^{N_A-1} A_i(n)W_i(n)\frac{\partial y_{EEOE}(n-i)}{\partial A_k(n)}\right]
\end{aligned} \tag{3.8}$$

Using the assumption of slow varying coefficients, similar to Simplified Recursive Predictive Error (SRPE) formulation [47], Eq. (3.8) may be rewritten as follows.

$$\begin{aligned}\frac{\partial e^2(n)}{\partial B_j(n)} &\approx -2e(n)\left[x(n-j)+\sum_{i=1}^{N_A-1} A_i W_i \frac{\partial y_{EEOE}(n-i)}{\partial B_j(n-i)}\right] \\ \frac{\partial e^2(n)}{\partial A_k(n)} &\approx -2e(n)\left[o(n-k)-W_k e(n-k)+\sum_{i=1}^{N_A-1} A_i W_i \frac{\partial y_{EEOE}(n-i)}{\partial A_k(n-i)}\right]\end{aligned} \tag{3.9}$$

Rewriting Eq. (3.9) in recursive form and moving in the opposite direction of gradient, the following update equations for coefficients of $\boldsymbol{A}$ and $\boldsymbol{B}$ are obtained.

$$\Delta \boldsymbol{B}(n)=e(n)\boldsymbol{x}_f \qquad \Delta \boldsymbol{A}(n)=e(n)\boldsymbol{y}_f \tag{3.10}$$

where $\boldsymbol{x}_f$ and $\boldsymbol{y}_f$ are the delayed-vector of Eq. (3.11) of appropriate size, with the first element given as follows.

$$\begin{aligned}x_f(n) &= x(n)+\sum_{i=1}^{N_A-1} A_i(n) W_i(n) x_f(n-i) \\ y_f(n-1) &= o(n-1)-W_1(n-1)e(n-1)+\sum_{i=1}^{N_A-1} A_i(n) W_i(n) y_f(n-i)\end{aligned} \tag{3.11}$$

It is noted that in Eq. (3.11) for $W_i(.)=0\ \forall\ i,n$ the updates degenerate to EE update and for $W_i(.)=1\ \forall\ i,n$ the updates simplify to OE formulation.

3.1.1.2 Weighting Filter Adaptation

With respect to constraints 1-3, the adaptation of $\boldsymbol{W}$ may be accomplished either deterministically or stochastically. In the former, starting from $\boldsymbol{W(0)=0}$, small predetermined increments are taken until for some N, $\boldsymbol{W(N)=1}$ is reached. Although the implementation of this approach is simple however, simulations have shown that in order to reach convergence the increments should be quite small, thus increasing the number of required iterations.

The stochastic approach attempts to minimize $e^2(.)$ while satisfying constraints 1-3. Employing the ALM (Augmented Lagrange Multiplier) method, the overall cost function may be written as follows.

$$C_W(n) = e^2(n,W) + \sum_{i=0}^{N_W-1} \lambda_i \psi_i + \mu_i \psi_i^2 + \beta_i \gamma_i + \theta_i \gamma_i^2 + \sigma_i \phi_i + \varepsilon_i \phi_i^2$$

$$C_W(n) = C_W^u(n) + C_W^c(n)$$

where

$$C_W^u(n) = e^2(n,W)$$

$$C_W^c(n) = \sum_{i=0}^{N_W-1} \lambda_i \psi_i + \mu_i \psi_i^2 + \beta_i \gamma_i + \theta_i \gamma_i^2 + \sigma_i \phi_i + \varepsilon_i \phi_i^2 \qquad (3.12)$$

$$\psi_i = \max\left(W_i(n) - 1, -\frac{\lambda_i}{2\mu_i}\right)$$

$$\gamma_i = \max\left(-W_i(n), -\frac{\beta_i}{2\theta_i}\right)$$

$$\phi_i = \max\left(W_i(n-1) - W_i(n), -\frac{\sigma_i}{2\varepsilon_i}\right)$$

Differentiating Eq. (3.12) with respect to coefficients of $\boldsymbol{W}$ the following expression is obtained.

$$\frac{\partial C_W(n)}{\partial W_k(n)} = \frac{\partial e^2(n)}{\partial W_k(n)} + \sum_{i=0}^{N_W-1} \left[\begin{array}{l} \lambda_i(n)\dfrac{\partial \psi_i(n)}{\partial W_k(n)} + 2\mu_i(n)\psi_i(n)\dfrac{\partial \psi_i(n)}{\partial W_k(n)} + \\ \beta_i(n)\dfrac{\partial \gamma_i(n)}{\partial W_k(n)} + 2\theta_i(n)\gamma_i(n)\dfrac{\partial \gamma_i(n)}{\partial W_k(n)} + \\ \sigma_i(n)\dfrac{\partial \phi_i(n)}{\partial W_k(n)} + 2\varepsilon_i(n)\phi_i(n)\dfrac{\partial \phi_i(n)}{\partial W_k(n)} \end{array} \right] \qquad (3.13)$$

Assuming similar conditions as Eq. (3.9), the first term of Eq. (3.13) (the unconstrained part) can be written as follows.

$$\frac{\partial C_W^u(n)}{\partial W_k(n)} = \frac{\partial e^2(n)}{\partial W_k(n)} \approx -2e(n)\left[-A_k e(n-k) + \sum_{i=1}^{N_A-1} A_i W_i \frac{\partial y_{EEOE}(n-i)}{\partial W_k(n-i)}\right] \qquad (3.14)$$

Noting that in our algorithm the update of $\boldsymbol{A}$ precedes that of $\boldsymbol{W}$, the recursive update of $\boldsymbol{W}$ is advanced by one time step. Thus the unconstraint update is formulated as shown.

$$\Delta W(n) = e(n) W_f$$
where
$$W_f(n) = e(n) + \sum_{i=1}^{N_A - 1} A_i(n) W_i(n) W_f(n-i) \tag{3.15}$$

where A_0=-1 has been employed. Using Eq. (3.12) the derivative of the constraint cost function, $C_W{}^c$, is written as follows.

$$\frac{\partial C_W{}^c(n)}{\partial W_i(n)} = \left\{ \begin{cases} \lambda_i(n) + 2\mu_i(n) \psi_i(n) & \text{if } W_i(n) > -\dfrac{\lambda_i(n)}{2\mu_i(n)_i} \\ 0 & \text{Otherwise} \end{cases} \right\} +$$
$$\left\{ \begin{cases} -\beta_i(n) - 2\theta_i(n) \gamma_i(n) & \text{if } -W_i(n) > -\dfrac{\beta_i(n)}{2\theta_i(n)} \\ 0 & \text{Otherwise} \end{cases} \right\} + \tag{3.16}$$
$$\left\{ \begin{cases} -\sigma_i(n) - 2\varepsilon_i(n) \phi_i(n) & \text{if } W_i(n-1) - W_i(n) > -\dfrac{\sigma_i(n)}{2\varepsilon_i(n)} \\ 0 & \text{Otherwise} \end{cases} \right\}$$

Thus the constrained update of W moves in the opposite direction of Eq. (3.16).

The overall update of W may be achieved by moving in the opposite direction of Eq. (3.13) or in opposite direction of Eq. (3.14) if the absolute sum of constraint violations is below some predefined threshold and in the opposite direction of Eq. (3.16) otherwise. It was found through a number of simulations that the latter has a faster rate of convergence.

3.1.1.2.1 EEOE Limitations

Using Eq. (3.1) the following expression for error is obtained.

$$e_{EEOE}(n) = o(n) + \sum_{i=1}^{N_A - 1} A_i(n) W_i(n) e_{EEOE}(n-i)$$
$$- \sum_{j=0}^{N_B - 1} B_j(n) x(n-j) - \sum_{k=1}^{N_A - 1} A_k(n) o(n-k) \tag{3.17}$$
$$e_{EEOE}(n) = e_{EE}(n) + \sum_{i=1}^{N_A - 1} A_i(n) W_i(n) e_{EEOE}(n-i)$$

where e_{EEOE} is the same as e in Eq. (3.1) and e_{EE} is the error with respect to EE structure, as described in Eq. (2.1). It is noted that Eq. (3.17) is an IIR filter with respect to e_{EEOE}, similar to the definition of e_{OE} in Eq. (2.4). Hence, not only for all values of W, except zero, the error surface may contain local minima, but also the initial coefficient bias estimations, *i.e.* when $W_i(.)$s are close to zero, will permanently affect the error measure. Indeed, as $W_i(.)$s grow larger the effect of the earlier $e_{EEOE}(.)$s also grow larger. Therefore, in adaptations based on EEOE formulation it is not unusual to find that the coefficients of adaptive filter remain biased even though, $W_i(.)$s have reached unity.

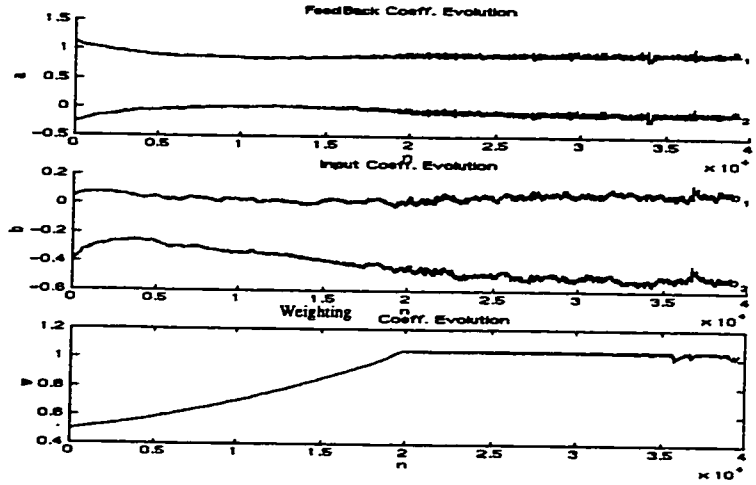


Figure 3.2 EEOE Divergence, Starting from Optimum

An example of this situation is shown in Figure 3.2. In this example the initial condition was the exact model of the unknown system with $W(0)$=0.5, Gaussian observation noise and SNR=1 . However, it is noted even when $W(.)$ reaches unity the coefficients of the adaptive filter are far from their desired values. In fact, the input coefficients are diverging. One should expect a similar behavior, regardless of update methods, as long as the recursive relationship of Eq. (3.17) holds.

3.2 Modified Equation Error Output Error (MEEOE)

To overcome the above problem, the recursive relationship in Eq. (3.17) must be abandoned. One way to achieve this objective is to modify Eq. (3.17) as follows.

$$e_{MEEOE}(n) = o(n) + \sum_{i=1}^{N_A-1} A_i(n) W_i(n) e_{OE}(n-i) - \sum_{j=0}^{N_B-1} B_j(n) x(n-j) - \sum_{i=1}^{N_A-1} A_i(n) o(n-i) \tag{3.18}$$

where $e_{OE}=o-y_{OE}$.

With respect to Eq. (3.18) the following equality is obtained.

$$e_{MEEOE}(n) = e_{EE}(n) + \sum_{i=1}^{N_A-1} A_i(n) W_i(n) e_{OE}(n-i) \tag{3.19}$$

where for $W=1$, $e_{MEEOE}= e_{OE}$, similar expression in [47] and [14] has been found.

Eq. (3.18) corresponds to a filter with the following output.

$$y_{MEEOE}(n) = \sum_{j=0}^{N_B-1} B_j(n) x(n-j) + \sum_{i=1}^{N_A-1} A_i(n) W_i(n) y_{OE}(n-i) + \sum_{k=1}^{N_A-1} A_k(n)(1-W_k(n)) o(n-k) \tag{3.20}$$

Figure 3.3 graphically depicts the operation of MEEOE.

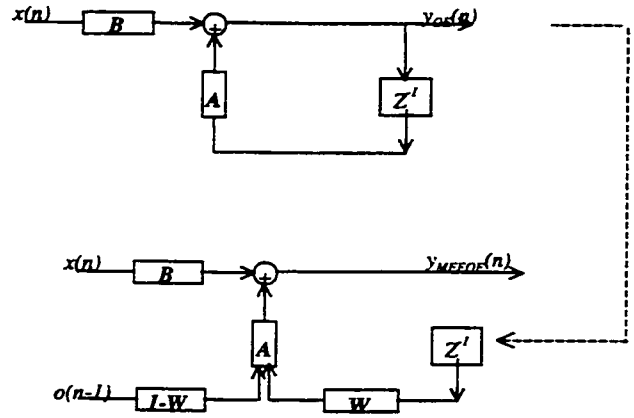


Figure 3.3 MEEOE Structure

Noting the definition of e_{OE} and based on similar derivations as in Section 1, the following expression, Eq. (3.21), for the MSE surface is obtained. It is noted that due to Eq. (3.20), unlike EEOE, the cross-correlation between the observed signal and filter output will be zero.

$$\sigma^2_{e_{MEEOE}} = \sigma^2_o + \sum_{j=0}^{N_B-1}\sum_{j'=0}^{N_B-1} B_j B_{j'} R_{xx}(j-j') + \sum_{i=1}^{N_A-1} A_i^2 W_i^2 \sigma^2_{e_{OE}} - \sum_{k=1}^{N_A-1}\sum_{k'=1}^{N_A-1} A_k(n) A_{k'}(n) R_{oo}(k-k') \quad (3.21)$$

Eq. (3.21) confirms the existence of a quadratic surface near the assumed optimum point however, it does not prove the uniqueness of such a point. This issue will be addressed in detail in Section 3.2.1 and Section 3.2.2.

The derivation of an adaptive algorithm for MEEOE is identical to EEOE, with consideration of the new error measure. The results are provided below.

$$\Delta B(n) = e_{MEEOE}(n) x_f(n) \qquad \Delta A(n) = e_{MEEOE}(n) y_f(n)$$

where

$$x_f(n) = x(n) + \sum_{i=1}^{N_A-1} A_i(n) W_i(n) x_f(n-i) \quad (3.22)$$

$$y_f(n) = o(n) - W_1(n) e_{OE}(n) + \sum_{i=1}^{N_A-1} A_i(n) W_i(n) y_f(n-i)$$

$$\Delta W(n) = e_{MEEOE}(n) W_f(n)$$

where (3.23)

$$W_f(n) = e_{OE}(n) + \sum_{i=1}^{N_A-1} A_i(n) W_i(n) W_f(n-i)$$

where the constrained adaptation are the same as Eq. (3.16).

3.2.1 Single Pole Filter

For a single pole filter we may ascertain a rigorous observation with respect to $\sigma^2_{MEEOE}(W)$ as a function of W. In particular we can show that $\frac{\partial \sigma^2_{MEEOE}(W)}{\partial W} < 0$, which

in turn indicates that $\sigma^2_{MEEOE}(W)$ is a decreasing function with respect to W, thus ensuring the existence of unique minimum, within the allowable range of W.

Using Eq. (3.18), Eq. (3.23) and Eq. (3.24) the derivative of y_{MEEOE} with respect to W, is shown in Eq. (3.25).

$$e_{MEEOE}(n) = e_{OE}(n) + \sum_{i=1}^{N_A-1} A_i(n)(W_i(n) - 1)e_{OE}(n-i) \tag{3.24}$$

$$\begin{aligned}\frac{\partial \sigma^2_{MEEOE}(n)}{\partial W(n)} &= -2E\left[e_{MEEOE}(n)\left[e_{OE}(n) + A(n)W(n)\frac{\partial y_{MEEOE}(n-1)}{\partial W(n-1)}\right]\right] \\ &= 2E\left[\left[-e_{OE}(n) + A(n)(1-W(n))e_{OE}(n-1)\right]\left[e_{OE}(n) + A(n)W(n)\frac{\partial y_{MEEOE}(n-1)}{\partial W(n-1)}\right]\right]\end{aligned} \tag{3.25}$$

Furthermore, linearly approximating the derivative of Eq. (3.20) with respect to W, as shown in Eq. (3.26), Eq. (3.27) is obtained

$$\frac{\partial y_{MEEOE}(n)}{\partial W(n)} \approx -A(n)e_{OE}(n-1) \tag{3.26}$$

$$\begin{aligned}\frac{\partial \sigma^2_{MEEOE}(n)}{\partial W(n)} &\approx 2E[-e_{OE}(n) + A(n)(1-W(n))e_{OE}(n-1)][e_{OE}(n) - A^2(n)W(n)e_{OE}(n-1)] \\ &\approx -2\sigma^2_{OE}(W)(1 + A^3W(1-W)) + 2R_{e_{OE}e_{OE}}(W,1)A(1-W+AW)\end{aligned} \tag{3.27}$$

where $R_{yx}(W_0,n)$ is the n^{th} lag auto-correlation of process x and y, when the current weight is W_0. With respect to Eq. (3.27), two cases may be considered. First, when the contribution of the second term is insignificant and second, when the second term has a significant contribution. In the former case, it is noted that the term enclosed in parenthesis is always positive, assuming stability bounds on A, hence ensuring that the error function is a monotonically decreasing function of W. In the latter case, noting that [39]

$$-\sigma_{OE}^2(W) \le R_{e_{OE}e_{OE}}(W,1) \le \sigma_{OE}^2(W) \tag{3.28}$$

Two extreme cases may be considered, by substituting Eq. (3.28) in Eq. (3.27).

$$R_{e_{OE}e_{OE}}(W,1) = \sigma_{OE}^2(W)$$

$$\frac{\partial \sigma_{MEEOE}^2(n)}{\partial W(n)} \approx 2\sigma_{OE}^2(W)\left[-1 + A^2W + (1-W)A(1-A^2W)\right] \tag{3.29}$$

$$R_{e_{OE}e_{OE}}(W,1) = -\sigma_{OE}^2(W)$$

$$\frac{\partial \sigma_{MEEOE}^2(n)}{\partial W(n)} \approx 2\sigma_{OE}^2(W)\left[-1 + A^2W - (1-W)A(1-A^2W)\right] \tag{3.30}$$

Noting that $|A|$ and W are both less than one, assuming stability, Eq. (3.29) and Eq. (3.30) may be further approximated as shown in Eq. (3.31) and Eq. (3.32). It is noted that with respect to domains of A and W, the derivatives in Eq. (3.31) and Eq. (3.32) are always negative. In fact, using exhaustive numerical substitution, it can be shown that the same conclusion holds for Eq. (3.29) and Eq. (3.30).

$$\frac{\partial \sigma_{MEEOE}^2(n)}{\partial W(n)} \approx 2\sigma_{OE}^2(W)\left[-1 + (1-W)A\right] \tag{3.31}$$

$$\frac{\partial \sigma_{MEEOE}^2(n)}{\partial W(n)} \approx 2\sigma_{OE}^2(W)\left[-1 - (1-W)A\right] \tag{3.32}$$

3.2.2 Multi-Pole Filter

In this section we attempt to gain an overall insight with respect to $\sigma_{MEEOE}^2(W)$ as a function of W, for a general IIR filter. Taking the mean square of Eq. (3.19) the following is obtained.

$$E\left[e^2_{MEEOE}(n)\right] = E\left[e^2_{EE}(n)\right] + E\left[\sum_{i=1}^{N_A-1} A_i(n) W_i(n) e_{OE}(n-i)\right]^2 + 2E\left[\sum_{i=1}^{N_A-1} A_i(n) W_i(n) e_{OE}(n-i) e_{EE}(n)\right] \tag{3.33}$$

Assuming that A and W adapt at much slower rate than the input signal, for a given $W=W_0$, Eq. (3.33) can be written as follows.

$$\sigma^2_{MEEOE}(W = W_0) = \sigma^2_{EE}(W = W_0) + \sum_{i=1}^{N_A-1}\sum_{j=1}^{N_A-1} A_i A_j W_{0_i} W_{0_j} \sigma^2_{OE}(W = W_0) + 2\sum_{j=1}^{N_A-1} A_j W_{0_j} R_{e_{OE}e_{EE}}(j) \tag{3.34}$$

Evaluating Eq. (3.34) for $W = W_0+\varepsilon$, where ε is a ball of sufficiently small radius ε [52], the following expression is obtained. (The nonlinear terms of the Taylor expansion have been omitted.)

$$\begin{aligned}
\sigma^2_{MEEOE}(W = W_0 + \varepsilon) &\approx \sigma^2_{MEEOE}(W = W_0) + 2\varepsilon E\left[\sum_{j=1}^{N_A-1} A_{W_{0_j}} W_{0\,j} e_{OE}(n-j) \sum_{k=1}^{N_A-1} A_{W_{0k}} e_{OE}(n-k)\right] \\
&\quad + 2\varepsilon E\left[\sum_{j=1}^{N_A-1} A_{W_{0_j}} e_{OE}(n-j) e_{EE}(n)\right] \\
\sigma^2_{MEEOE}(W = W_0 + \varepsilon) &\approx \sigma^2_{MEEOE}(W = W_0) \\
&\quad + 2\varepsilon E\left[\sum_{k=1}^{N_A-1} A_{W_{0k}} e_{OE}(n-k)\left[e_{EE}(n) + \sum_{j=1}^{N_A-1} A_{W_{0_j}} W_{0\,j} e_{OE}(n-j)\right]\right] \\
\sigma^2_{MEEOE}(W = W_0 + \varepsilon) &\approx \sigma^2_{MEEOE}(W = W_0) + 2\varepsilon \sum_{k=1}^{N_A-1} A_{W_{0k}} R_{e_{OE}e_{MEEOE}}(W = W_0, k)
\end{aligned} \tag{3.35}$$

where the last equality is obtained using Eq. (3.19). It is noted that unlike the Single-pole configuration, where it was proven that $\frac{\partial \sigma^2_{MEEOE}(W)}{\partial W} < 0$ (the second term of Eq. (3.35)), such claim may not be made with respect to Eq. (3.35), at least not to our knowledge. All

that may be established is that since the summation term is bounded, *i.e.* assuming BIBO stability, the contribution of the second term in Eq. (3.35) can be adjusted, using appropriate ε, such that the optimum of the $\sigma^2_{MEEOE}(W = W_0 + \varepsilon)$ is arbitrarily close to that of the $\sigma^2_{MEEOE}(W = W_0)$. With respect to Eq. (3.19) and Eq. (3.35) the following conjecture is made. At the beginning of adaptation, $W=0$, the error surface is characterized by a single biased optimum. As W increases other local minima may also appear, the existence of these local minima is due to the fact that as W increases e_{MEEOE} tends towards e_{OE}, which in turn is a non-linear function of A and B. However, assuming that for $W= W_0$ the adaptive coefficients have had sufficient time to reach their optimum values, albeit biased, for sufficiently small ε the optimum coefficients for $W= W_0+\varepsilon$ will be arbitrarily close to that of $W= W_0$. Hence, the new optimum coefficients are reachable from the initial points that are the optimum for $W= W_0$. Proceeding in the similar manner, our conjecture is that the global optimum for $W=I-\varepsilon$ will be arbitrarily close to that of the $W=I$, the unbiased solution, and hence reachable. The sufficient-time requirement can easily be achieved by adapting W at slower rate than A and B.

3.2.3 Simulation results

In this section four sets of simulations are presented, in order to verify the statements of the previous section.

3.2.3.1 Error surface evolution

In this section we verify the behavior of the MSE surface for a given W. The unknown system [47] has the following transfer function.

$$H(z) = \frac{0.05 - 0.4z^{-1}}{1 - 1.1314\,z^{-1} + 0.25\,z^{-2}} \tag{3.36}$$

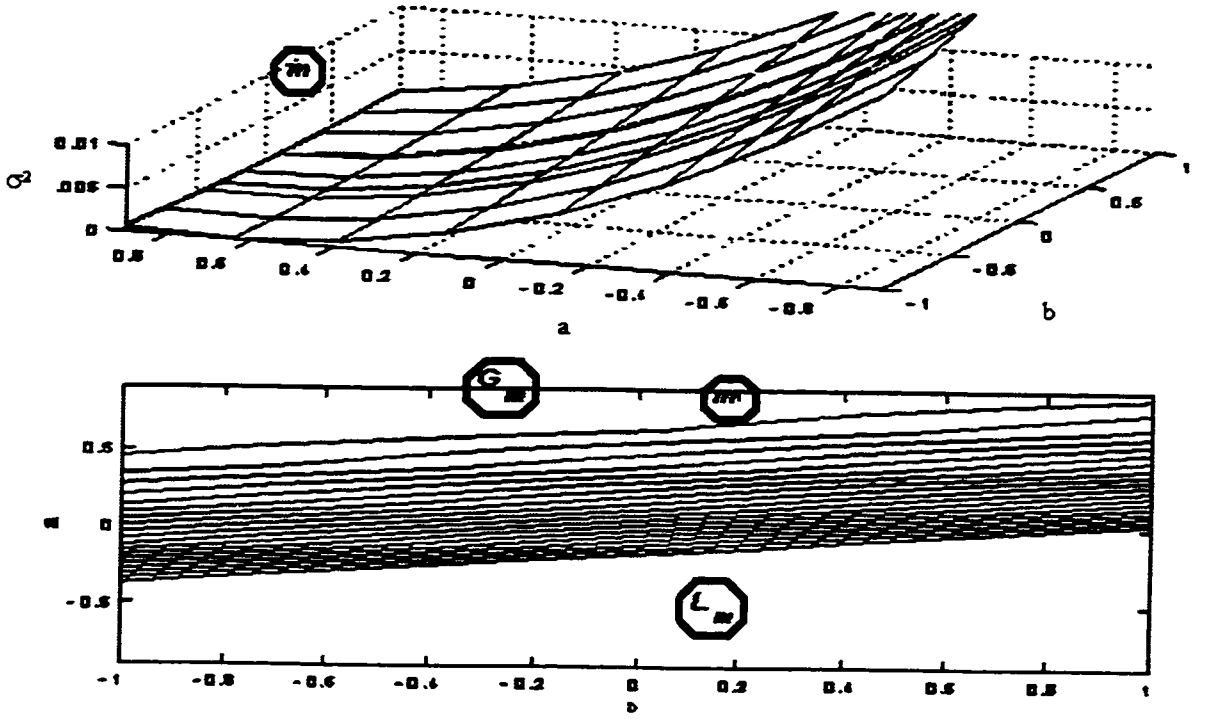
With the adaptive filter defined as

$$H(z)=\frac{b_1}{1-a_1 z^{-1}} \tag{3.37}$$

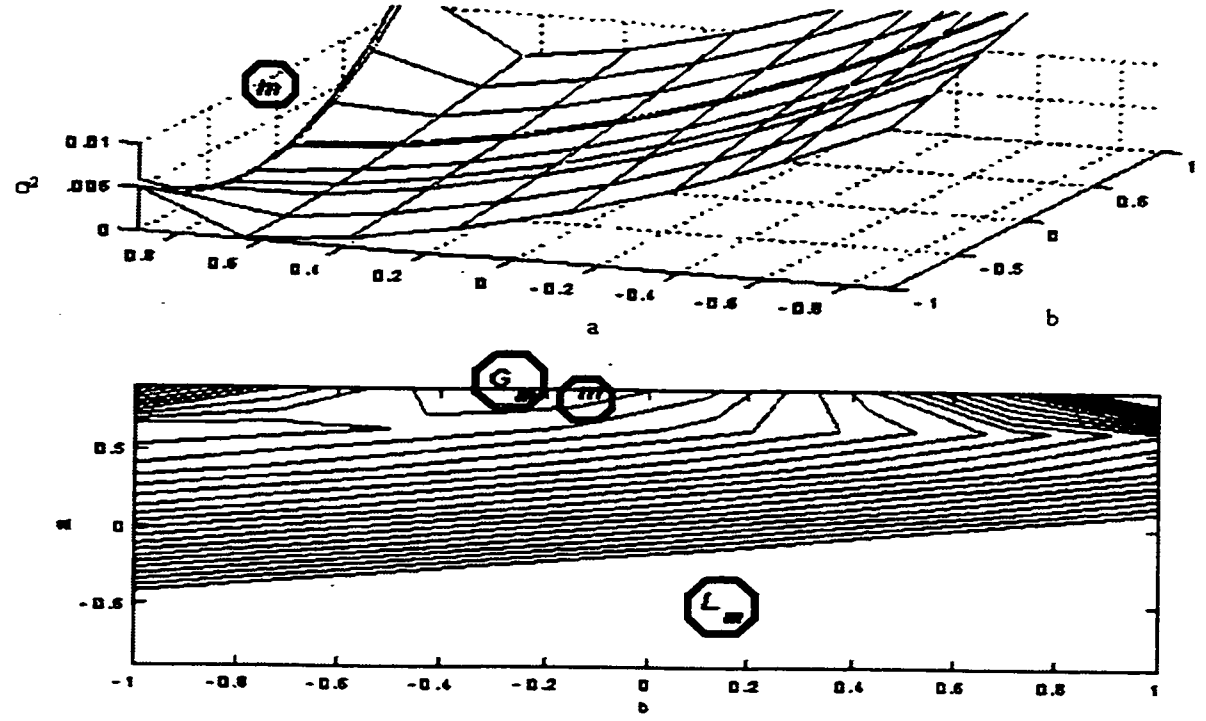
The first order approximation of this system has a local minimum with a pole at -0.519 and a numerator coefficient at 0.114, and a global minimum with a pole at 0.906 and a numerator coefficient at –0.311. The simulation result using a Gaussian white noise input with variance 1 and observation noise of variance 10 for different values of W is shown in Figure 3.4, where L_m and G_m are the theoretical local and global minimum [47] and m is the current minimum. The error surfaces have been normalized with respect to total energy of MSE. It is noted that as W increases the current minimum, m, moves toward the theoretical global minimum, G_m, and the local minimum appears at W=1.

3.2.3.2 MEEOE error surface

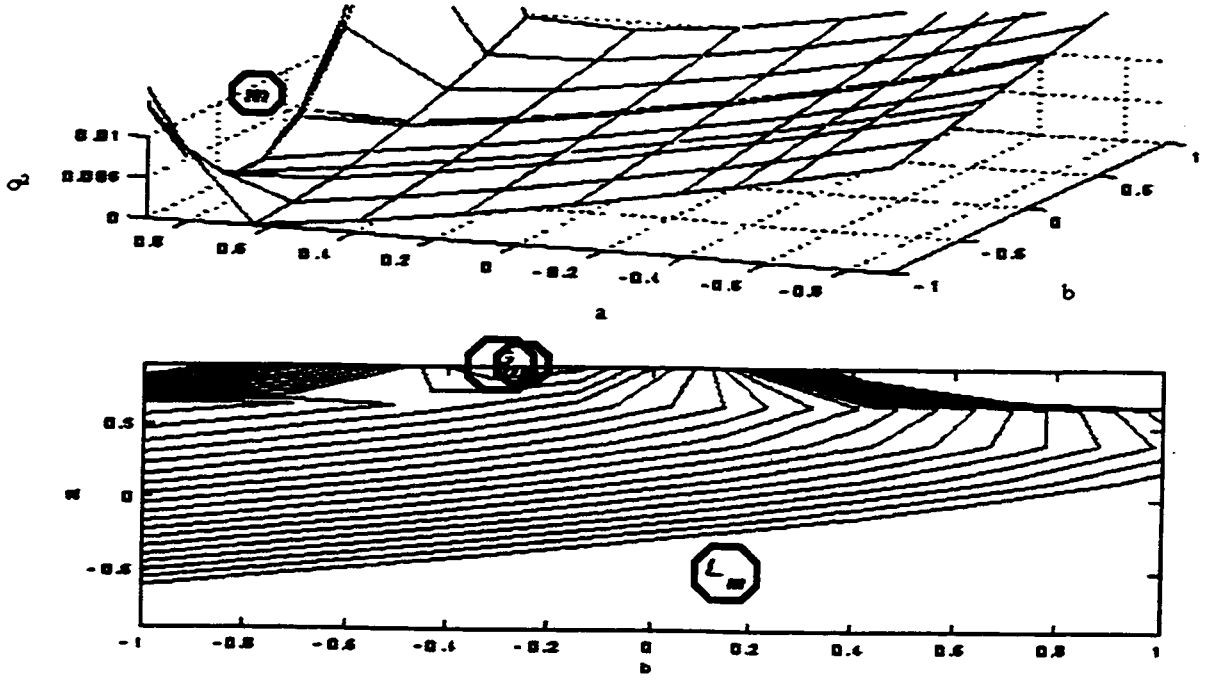
To demonstrate the uniqueness of the optimum, the above example with 20 different initial points around the region of interest has been simulated. The result is depicted in Figure 3.5. As shown in Figure 3.5 the adaptive filter reaches a unique optimum independent of the initial point, as W approaches unity. The IIR and Weighting filter step sizes are 5e-3 and 1e-4 respectively. Constraint tolerance is set at 1e-1.


m
0.01
.005
0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
a
b
0.6
0
-0.5
G
m
m
L
m
0.5
0
-0.5
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
b

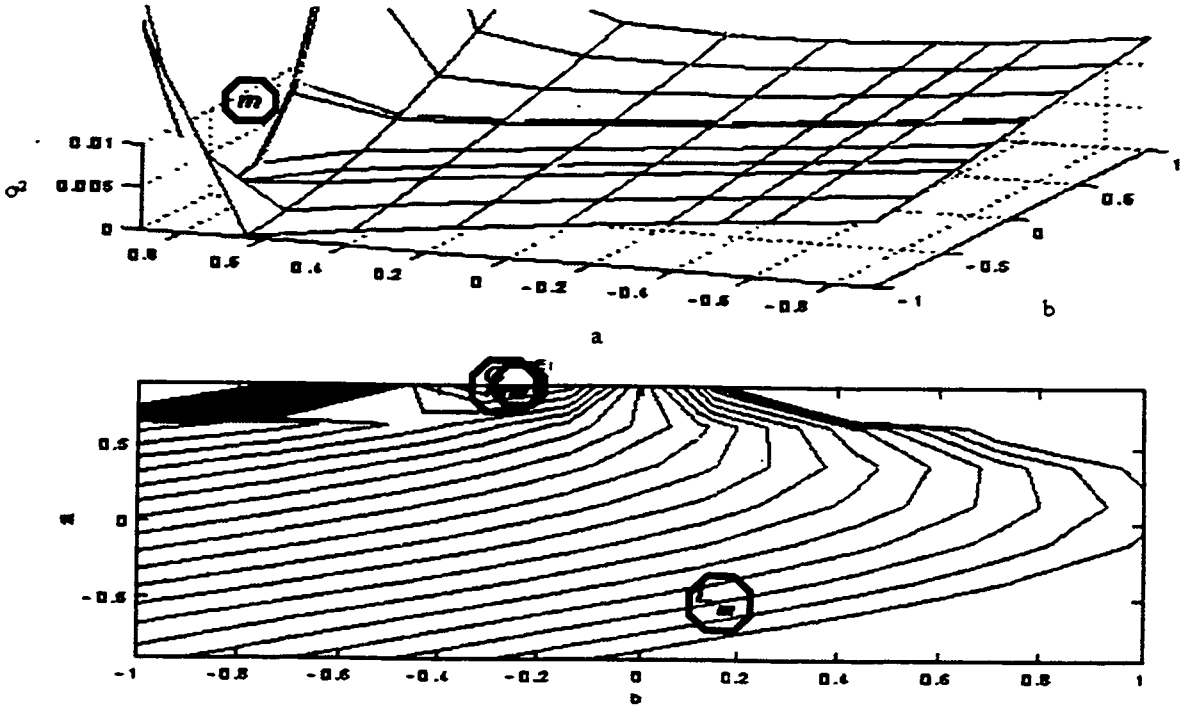
a) w=0


m
0.01
.005
0
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
a
b
0.6
0
-0.5
G
m
m
L
m
0.5
0
-0.5
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
b

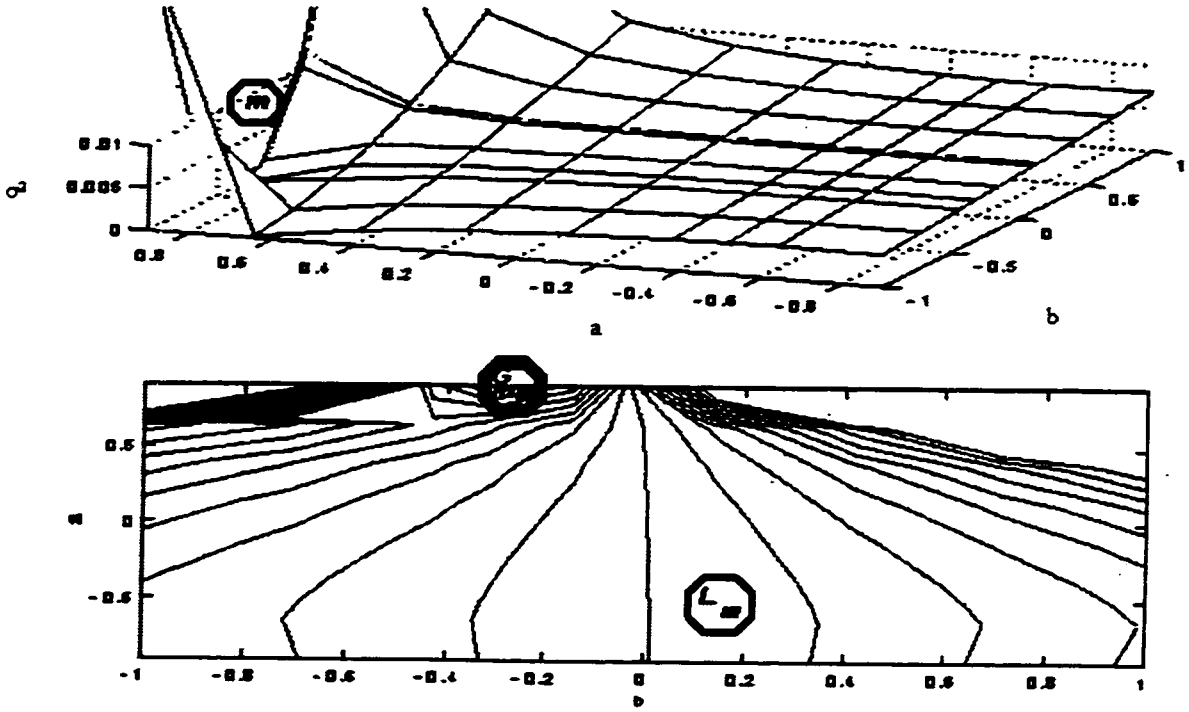
b) w=0.25



c) w=0.5



d) w=0.75



e) w=1

Figure 3.4 MSE Surface Evolution

3.2.3.3 Convergence of MEEOE from an initial point on a local minimum

In this example, [22], a single trial, starting from the theoretical local minimum with the similar step sizes and parameters as the previous section, is presented.

The unknown system has the following transfer function.

$$H(z) = \frac{0.05 - 0.4z^{-1}}{1 - 1.1314\,z^{-1} + 0.25\,z^{-2}} \tag{3.38}$$

With the adaptive filter defined as

$$H(z) = \frac{b_1}{1 - a_1 z^{-1}} \tag{3.39}$$

The first order approximation of this system has a local minimum with a pole at -0.519 and a numerator coefficient at 0.114, and a global minimum with a pole at 0.906 and a numerator coefficient at –0.311.

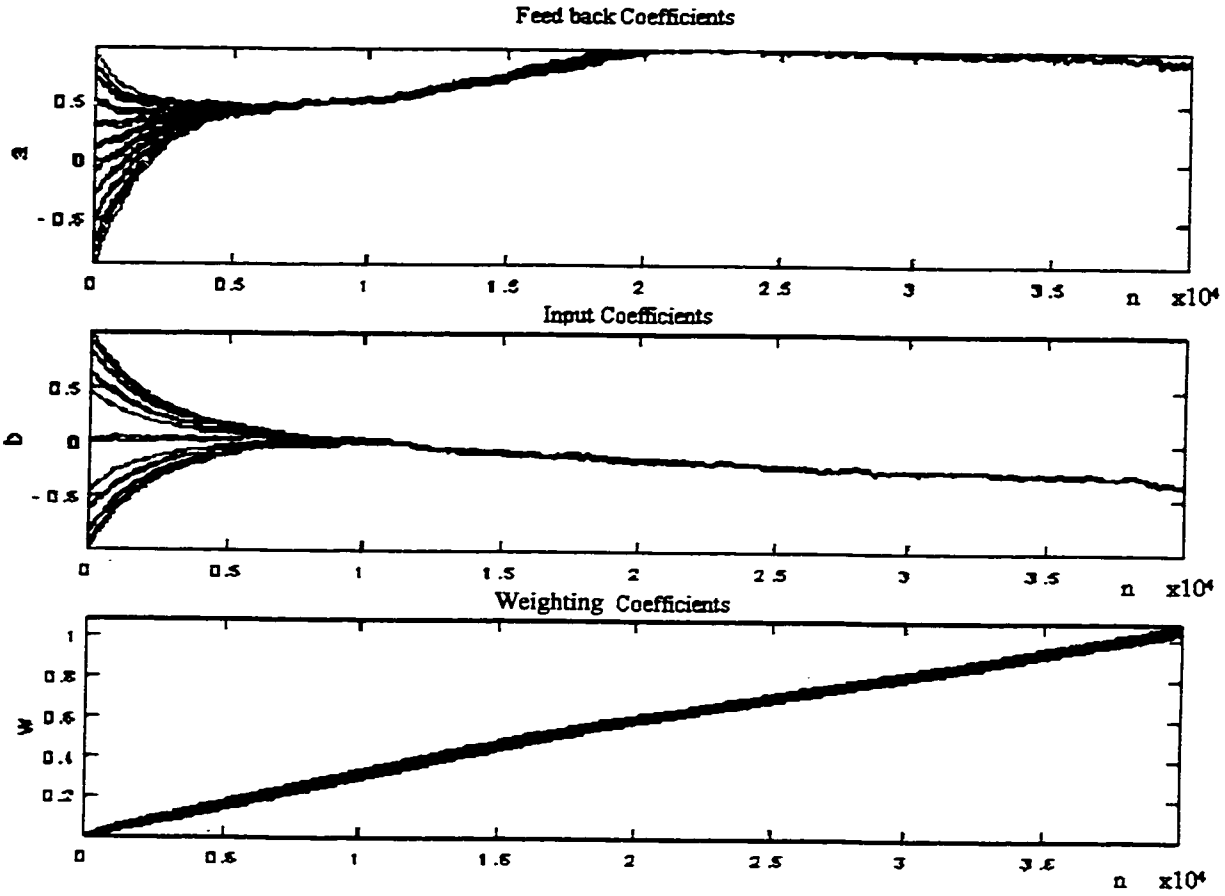


Figure 3.5 Convergence to Unique Optimum from Random Initial Points

The simulation result using a Gaussian white noise input with variance 1 and additive white noise with the same variance, *i.e.* *SNR*=1, is shown in

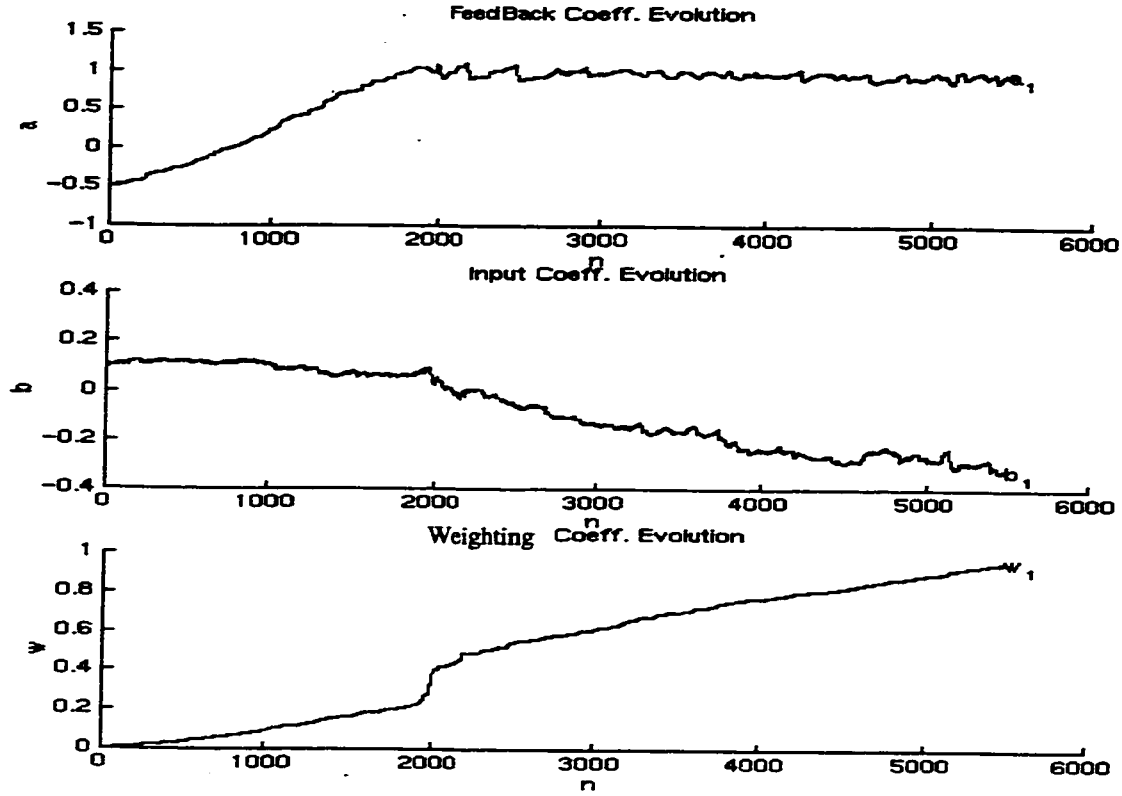


Figure 3.6. It is noted that even though the values of the initial coefficients were the same as those of the local minimum nonetheless, the algorithm managed to reach the global minimum in presence of the observation noise. The IIR and Weighting filter step sizes are 5e-3 and 1e-4 respectively. Constraint tolerance is set at 1e-1.

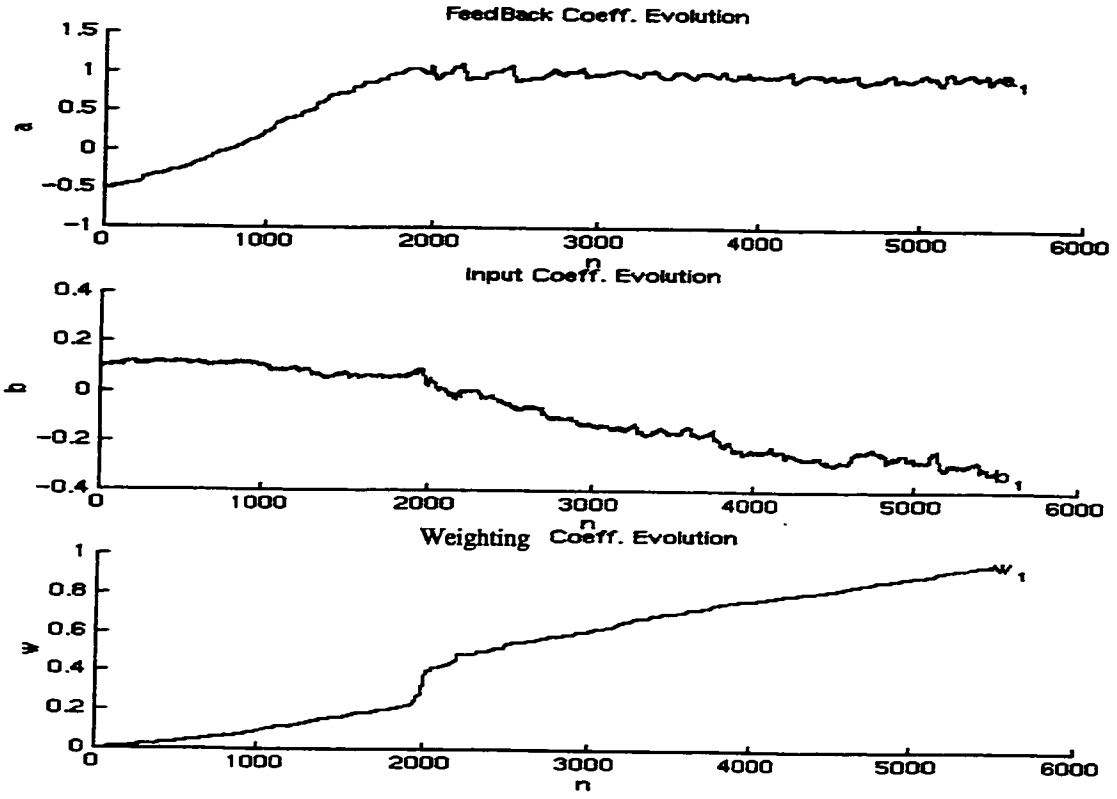


Figure 3.6 MEEOE Convergence Starting on Local Minimum

3.2.3.4 Convergence of MEEOE for Colored Input

In this section the performance of MEEOE under conditions of insufficient modeling and colored excitation source is examined [12]. This setup is shown in Figure 3.7.

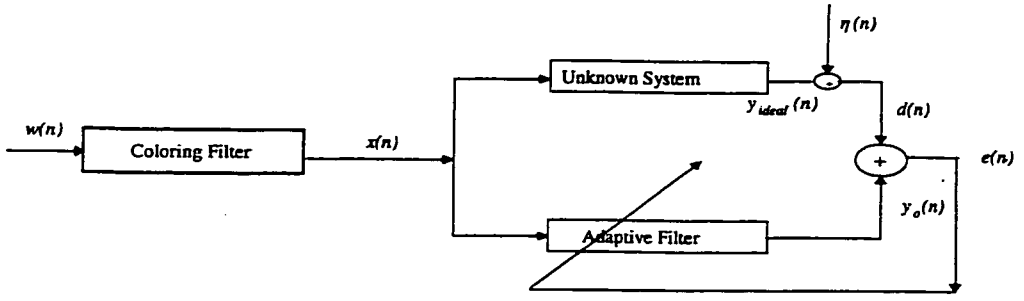


Figure 3.7 MEEOE System with Colored Input

3.2.3.4.1 Insufficient modeling

The unknown system has the following transfer function

$$H(z)=\frac{1}{1-1.84z^{-1}+1.08z^{-2}-0.216z^{-3}} \tag{3.40}$$

with the adaptive filter defined as

$$H(z)=\frac{b_1}{1+a_1z^{-1}+a_2z^{-2}} \tag{3.41}$$

and the following coloring filter

$$H(z)=\left(1-0.6z^{-1}\right)^2\left(1+0.6z^{-1}\right)^2 \tag{3.42}$$

The second order approximation of this system has a local minimum with the first pole at –0.9, the second one at -0.35 and a numerator coefficient at –0.33, and a global minimum with the first pole at 1.6, the second one at –0.7, and a numerator coefficient at 1.14. The IIR and Weighting filter step sizes are 1e-4 and 5e-6 respectively. Constraint tolerance is set at 1e-1.

The simulation result using a white Gaussian noise input with variance 1 and additive white noise with the same variance, *i.e.* SNR=1, is shown in Figure 3.8. It should be noted that the weighting factors w_1 and w_2 are increasing at a very slow rate, towards unity.

Again it is noted that even though the values of the initial coefficients were the same as those of the local minimum nonetheless, the algorithm managed to reach the global minimum in presence of the observation noise. The square-error of coefficients averaged over the number of adaptive taps is 0.0041627.

To investigate the effect of a single weighting factor, the above example using the same setup and a single weight was executed. The result is shown in Figure 3.9. The square-error of coefficients averaged over the number of adaptive taps is 0.0073466.

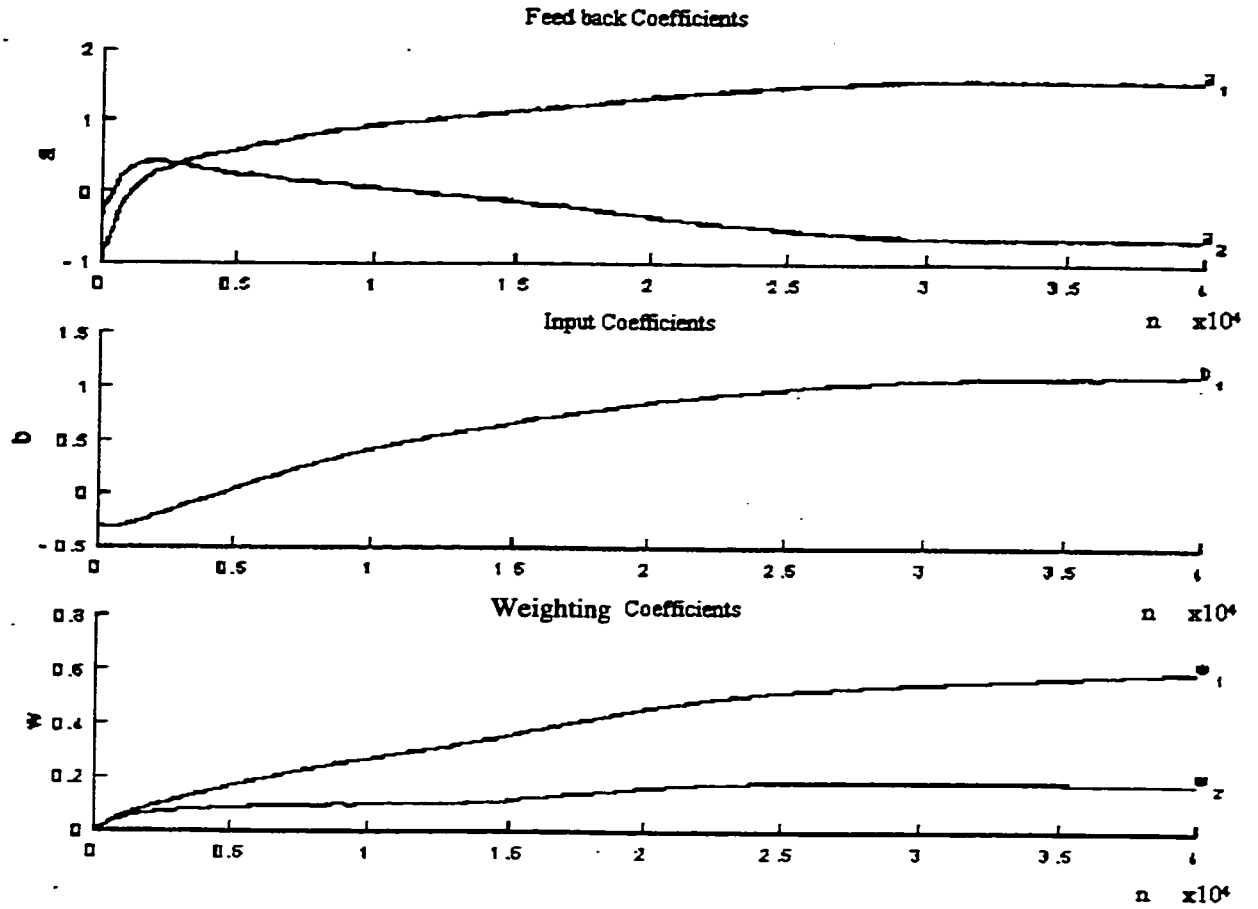


Figure 3.8 MEEOE Performance for Insufficient Modeling and Colored Noise

2.6.1.1.1 Sufficient Modeling

The unknown system has the following transfer function

$$H(z) = \frac{1}{1 - 1.4z^{-1} + 0.49z^{-2}} \tag{3.43}$$

with the adaptive filter defined as

$$H(z) = \frac{b_1}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{3.44}$$

and the following coloring filter

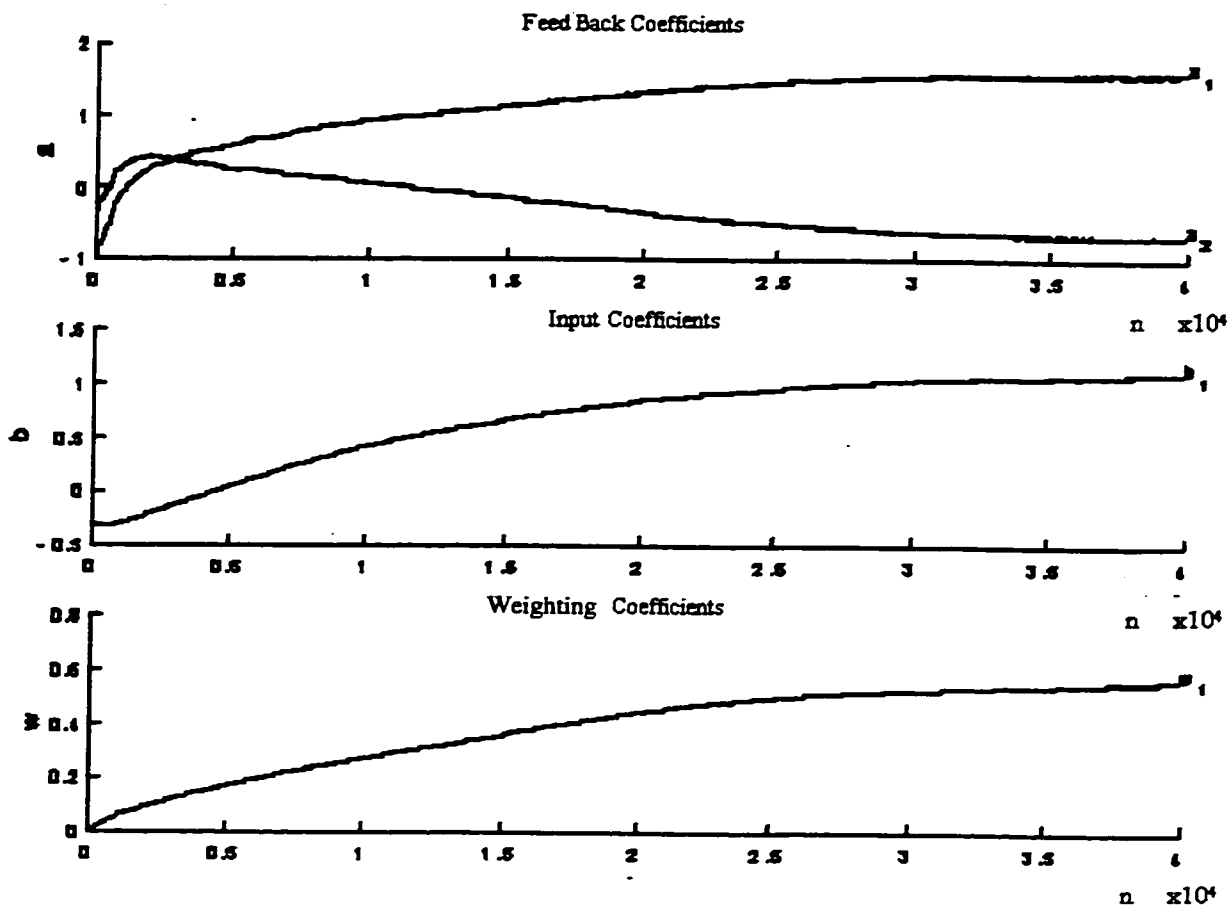


Figure 3.9 MEEOE Performance for Insufficient Modeling and Colored Noise Single Weight

and the following coloring filter

$$H(z) = (1 - 0.7z^{-1})^2 (1 + 0.7z^{-1})^2 \tag{3.45}$$

The second order approximation of this system has a local minimum with the first pole at -1.35, the second one at -0.49 and a numerator coefficient at –0.22, and a global minimum with the first pole at 1.4, the second one at –0.41, and a numerator coefficient at 1.

The IIR and Weighting filter step sizes are 1e-3 and 3e-5 respectively. Constraint tolerance is set at 1e-1. The simulation result using a white Gaussian noise input with variance 1 and additive white noise with the same variance, *i.e.* *SNR*=1, is shown in

Figure 3.10. The square-error of coefficients averaged over the number of adaptive taps is 0.0035417.

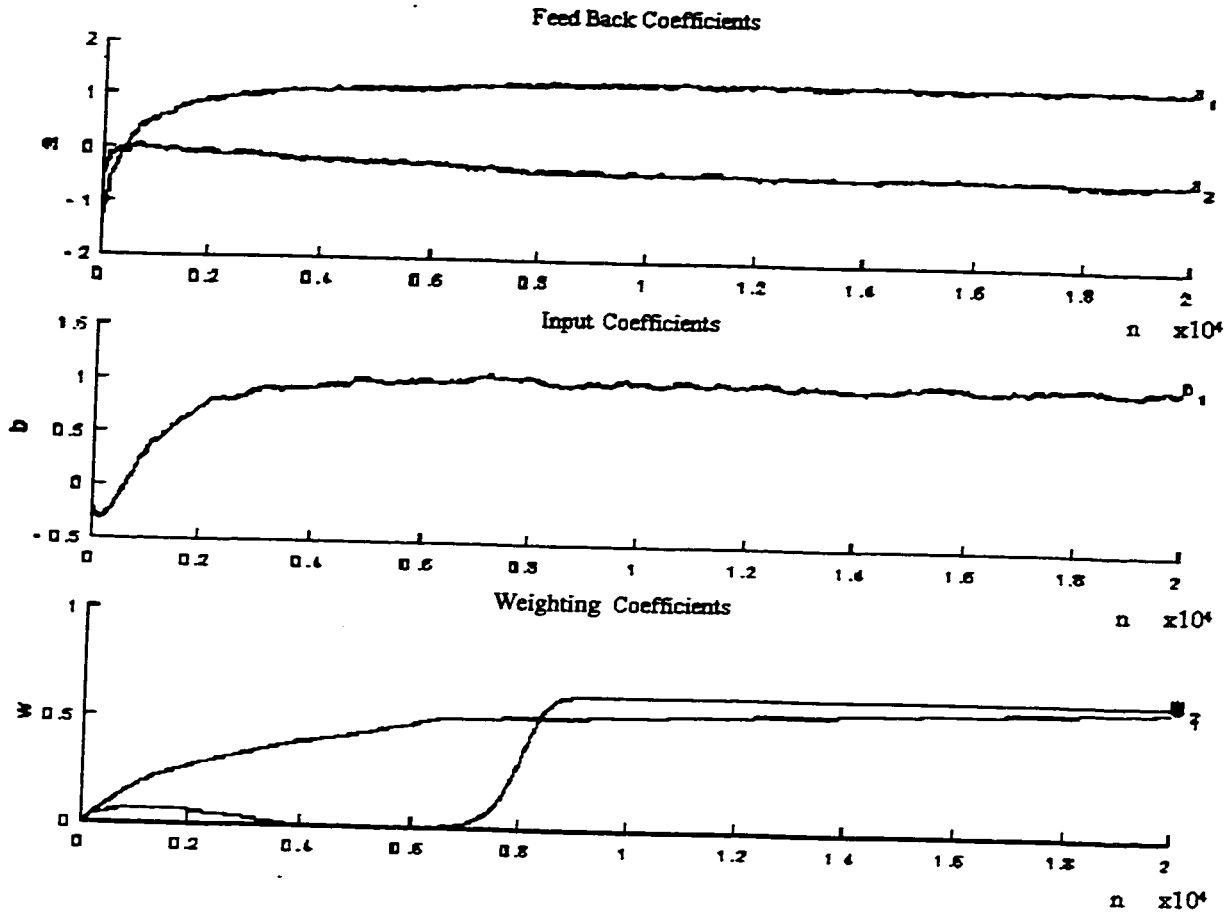


Figure 3.10 MEEOE Performance for Sufficient Modeling and Colored Noise

To investigate the effect of a single weighting factor, the above example using the same setup and a single weight was executed. The result is shown in Figure 3.11. The square-error of coefficients averaged over the number of adaptive taps is 0.00054549. However, in this example it is noted that the weighting factor has reached its steady state in a much longer time, as compared to Figure 3.10.

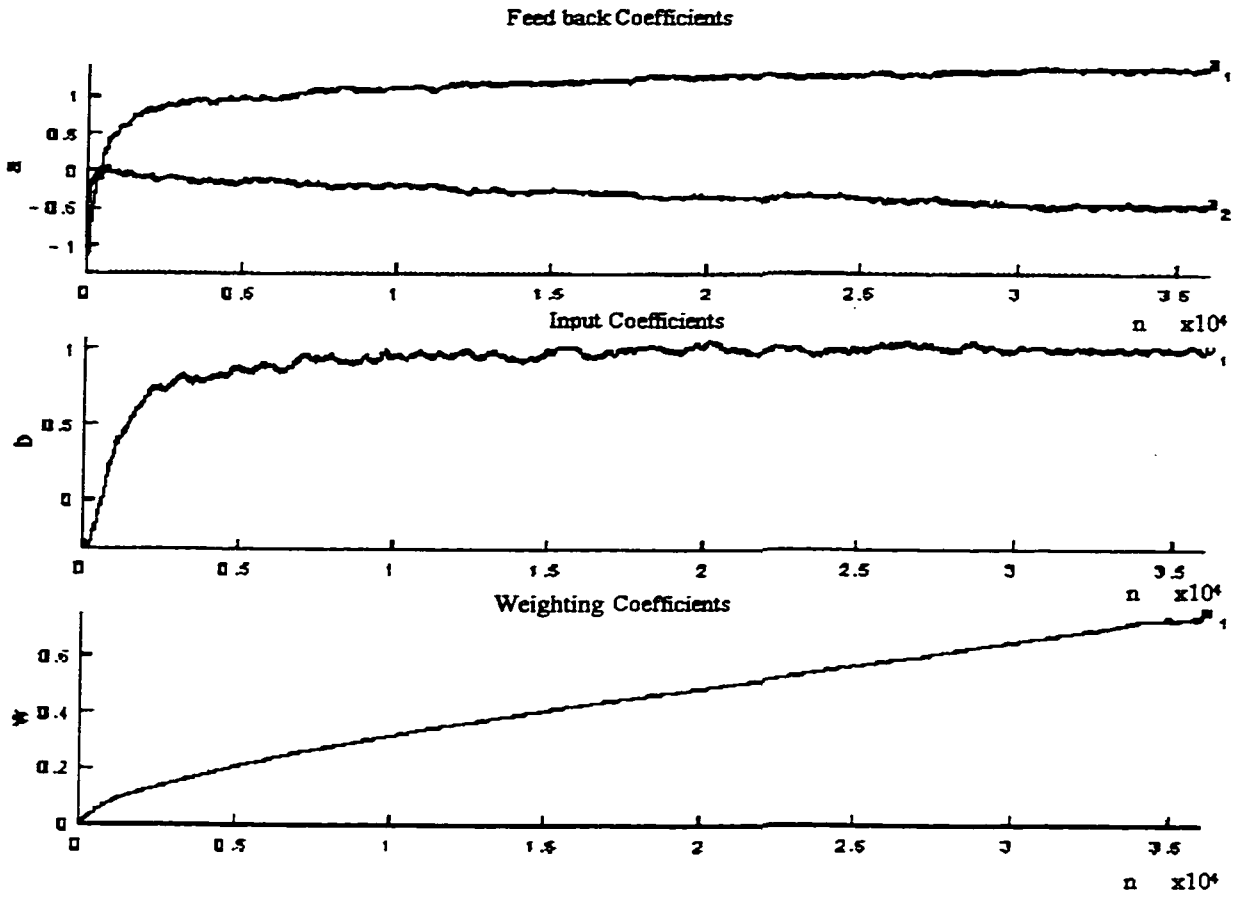


Figure 3.11 MEEOE Performance for Sufficient Modeling and Colored Noise Single Weight

3.2.3.4.2 Fast Adapting Weight

So far, in all stated examples, the weighting factor step size has been about one 10^{th} of that of the respective adaptive coefficients, which is in line with our stated requirement of Section 3.2. To verify the validity of this requirement, in this example, we allow the weighting factors, $\boldsymbol{W}$, to adapt at the same rate as the coefficients $\boldsymbol{A}$ and $\boldsymbol{B}$, using the same setup as that of the Section 3.2.3.3. Figure 3.12 is obtained. It is noted, even though the

weighting factor has reached unity, the adaptive coefficients have returned to their initial point, *i.e.* the local minimum of OE formulation.

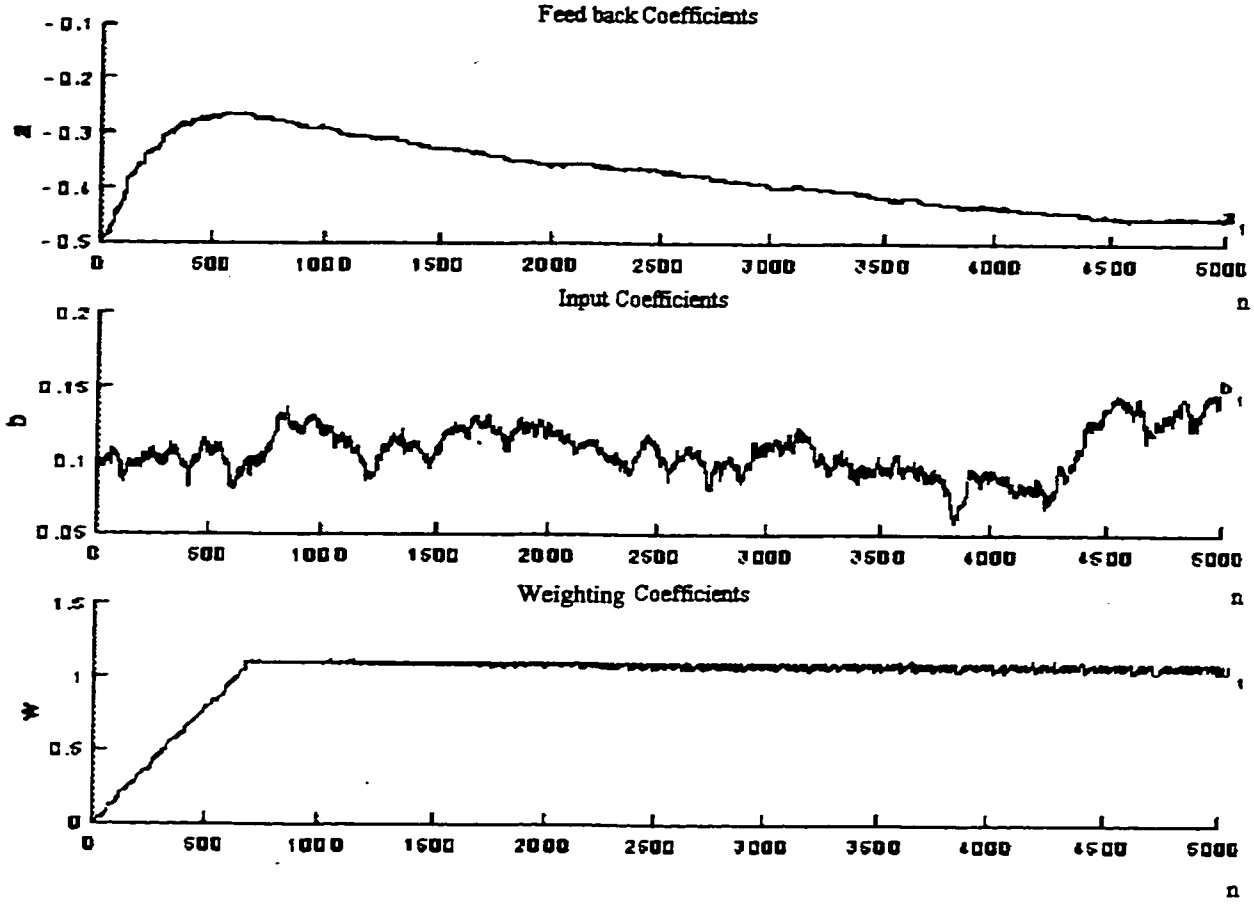


Figure 3.12 MEEOE Performance for Insufficient Adaptation Time

3.3 Conclusion and summary

In this chapter a novel method for AIF was presented namely, MEEOE. Based on our simulations, it was shown that for weight factor step size less than that of the coefficients of AIF, typically one tenth, global convergence to a correct optimum for insufficiently modeled system in presence of observation noise and colored input may be achieved.

We have shown the deficiencies of Combined Output methods [25] and how MEEOE resolves them. Furthermore, it is also noted that MEEOE belongs to the family of

Combined Error algorithms, as mentioned in Chapter 2. In fact if we were to use a single weighting factor, allow for the deterministic update of W and employ PLR method for the update of A and B, MEEOE would be identical to the methods represented by [33]. However, we have shown that more flexibility and faster convergence may be obtained by multiple weighting factors. Furthermore, allowing for non-deterministic update of W allows for faster convergence, independent of dynamics of the process. Lastly, unlike PLR, SRPE ensures convergence to an optimum point on MSE surface. Thus unlike Combined Error algorithms convergence to an optimum can be assumed.

Chapter 4 Maximum Entropy Kalman Filter

Adaptive processes often attempt to minimize the MSE to filter a partially observed digital signal. While mathematically tractable, the MSE criterion causes over-smoothing of the filtered signal. In this chapter, we propose using Maximum Entropy (ME) as the optimization criterion to avoid over-smoothing of the signal. This criterion is motivated by the fact that ME methods make no assumptions regarding the unobserved data. The Maximum Entropy Kalman Filter (MEKF) presented in this chapter employs ME as its optimization criterion to identify the appropriate parameters of the standard Kalman filter.

4.1 Introduction

Intuitive justification of the ME principle may be established by considering the following example. Let $x(n)$ be a finite length time series (critically sampled version of $x(t)$) with $X(0)...X(N-1)$ corresponding to its DFT. Furthermore, assume due to bandwidth limitations, half of $x(n)$ samples were dropped. Let the down-sampled sequence be called $x'(n)$ with its respective DFT $X'(0)...X'(N/2-1)$. We wish to estimate $X(0)...X(N-1)$ given $X'(0)...X'(N/2-1)$. Given N unknowns and $N/2$ equations, in the absence of any a priori information, there is an infinite number of possible solutions to this problem. The premise of the ME method is to assume a priori that the entropy, *i.e.* uncertainty or information content, of the reconstructed DFT sequence is maximized.

Theoretical justifications for ME criterion is based on Concentration theorem [21], as reviewed in Chapter 2. Briefly, the theory states that among all probability distributions,

which satisfy a given number of constraints, those with higher entropy have a much higher multiplicity.

When filtering a partially observed stochastic signal, one must define i) a filtering criterion, ii) a cost function, and iii) an optimization method. Within the Bayesian framework Maximum a posteriori (MAP), Maximum Likelihood (ML), and Maximum Entropy are the most commonly used filtering criteria. The main advantage of the ME over MAP and ML is that it makes no assumption regarding the unobserved data [21]. The cost function, which ultimately defines the mode of convergence, should preferably allow for the existence of a unique optimum. Finally, an efficient search method is needed to find the optimum. Kalman filter defines the trace of the error covariance matrix as its cost function, and gradient search as its optimization method. Furthermore, the proposed MEKF defines the process model so as to maximize the entropy of the filtered signal.

This chapter is organized as follows. In Section 2, we develop a method for modeling digital signals using the ME principle. In Section 3, application of MEKF in image compression is discussed. In Section 4, a more general framework of estimation using the ME principle is developed. In Section 5, application of MEKF in image reconstruction is discussed. Finally, future work and possible improvements are discussed.

4.2 1-D Maximum Entropy Kalman Filter (MEKF) for modeling

In this section we seek to develop plant and observation dynamics such that for a given auto-correlation function the output of the Kalman filter will be a ME sequence. It is noted that, unlike the more general problem of estimation, both of these dynamics may be

chosen freely. A typical Kalman filter is described by the process model and the observation model [29][6] given by:

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k w_{k+1} \\ y_k &= C_k x_k + v_k \end{aligned} \tag{4.1}$$

where, $x_k(Nx1)$ and $y_k(Mx1)$ are the state of the process and observation, respectively, at time index k,
A_k (NxN) is the state transition matrix,
B_k $(Nx1)$ is the coloring filter,
C_k (MxN) is the observation matrix, and
$w_k \sim (0,Q)$ and $v_k \sim (0,R)$ are uncorrelated, zero mean white Gaussian noise processes with variances Q, R and dimensions $(1x1)$, $(Mx1)$, respectively.

Assuming a time invariant, stationary and auto-regressive (AR) process, with an observation matrix C_k, Eq. (4.1) can be rewritten as:

$$\begin{aligned} x_{k+1} &= A_k x_k + [0 \ . \ . \ 1]^T w_{k+1} \\ y_k &= \begin{bmatrix} c_{11} & . & . & c_{1N} \\ . & . & . & . \\ c_{M1} & . & . & c_{MN} \end{bmatrix}_k x_k + v_k \end{aligned} \tag{4.2}$$

Note that the output of a Kalman filter may be expressed as [29]:

$$\hat{x}_{k+1} = (I - K_{k+1} C_{k+1}) A_k \hat{x}_k + K_{k+1} y_{k+1} \tag{4.3}$$

where K_k is the Kalman gain at time index k.
Eq. (4.3) may be rewritten as follows:

$$\begin{aligned} \hat{x}_{k+1} &= A_k \hat{x}_k + K_{k+1}(y_{k+1} - C_{k+1} A_k \hat{x}_k) \\ &= A_k \hat{x}_k + K_{k+1}(y_{k+1} - C_{k+1} \hat{x}_{k+1}^-) \\ &= A_k \hat{x}_k + K_{k+1} \xi_{k+1} \\ &= \hat{x}_{k+1}^- + K_{k+1} \xi_{k+1} \end{aligned} \tag{4.4}$$

where $\hat{x}_{k+1}^-$ is the one-step prediction of $\hat{x}_{k+1}$, *i.e.* time update, and ξ_{k+1} is the innovation process at time index $k+1$. Assuming the dynamics are stationary and the transients have vanished, ξ_{k+1} can be characterized by a white Gaussian noise $(0, \sigma_\xi)$ [16]. The MEKF requires A_k and K_{k+1} to be such that $\hat{x}$ is a ME sequence, where σ_ξ may be expressed as follows.

$$\begin{aligned}\sigma_\xi &= E\left[(y_{k+1} - C_{k+1}A_k\hat{x}_k)(y_{k+1} - C_{k+1}A_k\hat{x}_k)^T\right] \\ &= E\left[(C_{k+1}x_{k+1} + v_{k+1})(C_{k+1}x_{k+1} + v_{k+1})^T\right] + E\left[(C_{k+1}A_k\hat{x}_k)(C_{k+1}A_k\hat{x}_k)^T\right] \\ &= C_{k+1}R_xC_{k+1}^T + R + C_{k+1}A_kR_{\hat{x}}A_k^TC_{k+1}^T \\ &= C_{k+1}R_xC_{k+1}^T + R + C_{k+1}A_kR_xA_k^TC_{k+1}^T\end{aligned} \tag{4.5}$$

where the last equality follows from the basic assumption that the output sequence should have a similar auto-correlation function as the desired sequence.

In order for a sequence, x^{ME}, to be a maximum entropy sequence, for a given auto-correlation function r, x^{ME} must be an AR process, as described in Eq. (4.6) [7][5]:

$$x_{k+1}^{ME} = -\sum_{i=0}^{N-1} a_i^{ME} x_{k-i}^{ME} + z_{k+1} \tag{4.6}$$

satisfying the following equalities:

$$\begin{aligned}R^{ME}a^{ME} &= r \\ Q^{ME} &= r(0) - a^{ME^T}r\end{aligned} \tag{4.7}$$

where $z_k \sim (0, Q^{ME})$ and R^{ME} (NxN) is the auto-correlation matrix corresponding to the given auto-correlation function r.
Eq. (4.6) may be written in state form as follows:

$$\begin{aligned}x_{k+1}^{ME} &= \begin{bmatrix} 0 & 1 & 0 & . & . \\ 0 & 0 & 1 & 0 & . \\ . & . & . & . & . \\ . & . & . & . & . \\ -a_1^{ME} & -a_2^{ME} & . & . & -a_N^{ME} \end{bmatrix} x_k^{ME} \\ &\quad + [0 \ . \ . \ 1]^T z_{k+1} \\ x_{k+1}^{ME} &= A^{ME}x_k^{ME} + [0 \ . \ . \ 1]^T z_{k+1}\end{aligned} \tag{4.8}$$

where the a^{ME} coefficients are determined using the Levinson-Durbin algorithm.

In order for $\hat{x}$ to be a ME sequence, comparing a time-invariant dynamics of Eq. (4.4) and (4.8), the equalities in Eq. (4.9) should hold.

$$A_{k+1} = A^{ME}$$
$$K_{k+1} = [0 \quad . \quad . \quad . \quad \sigma_{ME}]^T \tag{4.9}$$

for all k, where

$$\sigma_{ME} = \sqrt{\frac{Q^{ME}}{\sigma_\xi}} \tag{4.10}$$

Furthermore, it is assumed that C_k is $(1xN)$.
Noting that Kalman gain may be expressed as follows [29]:

$$K_{k+1} = P_{k+1} C_{k+1}^T R^{-1} \tag{4.11}$$

where P_{k+1} is the posteriori error covariance matrix. Equating the Kalman gain definition in Eq. (4.9) and Eq. (4.11), the following is obtained.

$$\begin{aligned} &P_{k+1} C_{k+1}^T R^{-1} = [0 \cdots \sigma_{ME}]^T \\ &\left[\left(P^-_{k+1}\right)^{-1} + C_{k+1}^T R^{-1} C_{k+1}\right]^{-1} C_{k+1}^T R^{-1} = [0 \cdots \sigma_{ME}]^T \\ &C_{k+1}^T = \left[\left(P^-_{k+1}\right)^{-1} + C_{k+1}^T R^{-1} C_{k+1}\right] [0 \cdots \sigma_{ME} R^{-1}]^T \\ &C_{k+1}^T - C_{k+1}^T C_{k+1} [0 \cdots \sigma_{ME}]^T = \left(P^-_{k+1}\right)^{-1} [0 \cdots \sigma_{ME} R^{-1}]^T \end{aligned} \tag{4.12}$$

where P^-_k is the a-priori error covariance matrix. Eq. (4.12) may be simplified as follows.

$$[C_1 - \sigma_{ME} C_1 C_N, C_2 - \sigma_{ME} C_2 C_N, \cdots C_N - \sigma_{ME} C_N C_N]^T_{k+1} = [\alpha_1 \cdots \alpha_N]_{k+1}^T \tag{4.13}$$

where the α vector corresponds to the right hand side of Eq. (4.12). Solving for C_N, the following is obtained.

$$\begin{aligned} &C_N^2 - \frac{1}{\sigma_{ME}} C_N + \frac{\alpha_N}{\sigma_{ME}} = 0 \\ &C_N = \frac{\frac{1}{\sigma_{ME}} \pm \sqrt{\frac{1}{\sigma_{ME}^2} - 4\frac{\alpha_N}{\sigma_{ME}}}}{2} \end{aligned} \tag{4.14}$$

where the time indices are implicitly assumed.

The first issue in determining C_N is to ensure that it is a real value. If at some time step i the term under the square root becomes negative, R will be adjusted according to following.

$$\frac{1}{\sigma_{ME}^2} - 4\frac{\alpha_N}{\sigma_{ME}} = \chi \tag{4.15}$$

where χ is a real number ≥ 0. Substituting for α_N, according to Eq. (4.12), in Eq. (4.15) the following is obtained.

$$\frac{1}{\sigma_{ME}^2} - 4\frac{q_{NN}\sigma_{ME}R}{\sigma_{ME}} = \chi \tag{4.16}$$

where q_{NN} is the *(N,N)* element of $(\boldsymbol{P}^-)^{-1}$.

Substituting Eq. (4.10) in Eq. (4.16) the following is obtained.

$$\frac{\sigma_\xi}{Q^{ME}} = \chi + 4q_{NN}R \tag{4.17}$$

Using the expression of Eq. (4.5) in the above, the following results.

$$\begin{aligned} &\boldsymbol{C}_{k+1}\boldsymbol{R}_x\boldsymbol{C}_{k+1}^T + R + \boldsymbol{C}_{k+1}\boldsymbol{A}_k\boldsymbol{R}_x\boldsymbol{A}_k^T\boldsymbol{C}_{k+1}^T = Q^{ME}\chi + 4Q^{ME}q_{NN}R \\ &R(1-4q_{NN}Q^{ME}) = \chi Q^{ME} - \boldsymbol{C}_{k+1}\boldsymbol{R}_x\boldsymbol{C}_{k+1}^T - \boldsymbol{C}_{k+1}\boldsymbol{A}_k\boldsymbol{R}_x\boldsymbol{A}_k^T\boldsymbol{C}_{k+1}^T \\ &R = \frac{\chi Q^{ME} - \boldsymbol{C}_{k+1}\boldsymbol{R}_x\boldsymbol{C}_{k+1}^T - \boldsymbol{C}_{k+1}\boldsymbol{A}_k\boldsymbol{R}_x\boldsymbol{A}_k^T\boldsymbol{C}_{k+1}^T}{1-4q_{NN}Q^{ME}} \end{aligned} \tag{4.18}$$

The second issue is to determine which of the two possible solutions of Eq. (4.14) leads to a desirable solution. Possible solutions to Eq. (4.14), at time index k, are as follows.

$$C_N(k+1) = \frac{\frac{1}{\sigma_{ME}} + \sqrt{\frac{1}{\sigma_{ME}^2} - 4\frac{\alpha_N(k+1)}{\sigma_{ME}}}}{2} \tag{4.19-a}$$

$$C_N(k+1) = \frac{\frac{1}{\sigma_{ME}} - \sqrt{\frac{1}{\sigma_{ME}^2} - 4\frac{\alpha_N(k+1)}{\sigma_{ME}}}}{2} \tag{4.19-b}$$

To identify the suitable solution for C_N, the following simplifying assumption is made, which is in line with the requirement that the observation matrix $\boldsymbol{C}$ be real at every time index i.

$$\sigma_{ME} << 1 \tag{4.20}$$

Employing the above assumption in Eq. (4.19), the following is obtained.

$$C_N(k+1) \propto \frac{1}{\sigma_{ME}} \tag{4.21-a}$$

$$C_N(k+1) \propto \eta \tag{4.21-b}$$

where η is some small value > 0.

The following observation with respect to Eq. (4.5) is made.

$$\sigma_\xi \propto C_N^2(k) \tag{4.22}$$

It is also noted that with respect to Eq. (4.10), the following holds.

$$\sigma_{ME} \propto \sigma_\xi^{\frac{1}{2}} \tag{4.23}$$

From Eq. (4.22) and Eq. (4.23), Eq. (4.24) is obtained.

$$\sigma_{ME} \propto C_N^{-1} \tag{4.24}$$

Combining the expressions in Eq. (4.21-a) and Eq. (4.24) the following relationship is obtained.

$$\begin{aligned} C_N(k+1) &\propto \frac{1}{\sigma_{ME}} \propto C_N(k) \\ C_N(k+1) &= \beta(k) C_N(k) \end{aligned} \tag{4.25}$$

It is noted that the above yields to a stable solution if $|\beta|<1$, for all time steps k. To avoid unstable solutions, update based on Eq. (4.19-b) is employed. In fact our simulations have shown that updates based on Eq. (4.19-a) yield to unstable solutions for C_N.

Employing both Eq. (4.12) and Eq. (4.19-b), other observation coefficients are obtained as follows, at time index $k+1$.

$$C_i = \frac{\alpha_i}{(1-\sigma_{ME} C_N)} \quad i = 1 \cdots N-1 \tag{4.26}$$

4.2.1 1-D MEKF modeling algorithm

The following pseudo-code summarizes one possible implementation of MEKF, where we have assumed that the data points are generated by a stationary process.

1. Initialization
   - Randomly initialize C_0 *(1xN)*
   - Choose a small initial R
   - Choose a small positive χ
   - For the given r *(1xN)*, calculate the transitional matrix A *(NxN)*
   - Set the diffusion matrix $B=[0 \ldots 1]^T$ *(Nx1)*
2. Analysis - encode all data points
   - Using Equations (4.26), (4.19-b), (4.7), (4.5) and if necessary Eq. (4.18) calculate a new observation matrix C_k, Q^{ME} and R
   - Generate encoded observation sequence y_k
3. Synthesis - decode all observation points
   - Generate Kalman filter approximation with observation sequence y_k, parameters A, B, C, R and $Q=Q^{ME}$

4.2.2 MEKF based modeling results

Two sets of AR and ARMA process simulations are described here.

4.2.2.1 MEKF modeling of an AR signal

In the first set, the actual model, as given by Eq. (4.2), is described by an AR model $a_{Ideal}=[0.5,-0.1,0.2,-0.3]$, $w_k \sim (0,1)$, Figure 4.1 shows 1:4 analysis-synthesis using DCT

(the first ¼ DCT samples are kept). Figure 4.2 shows the simulation results using MEKF, with observation vector, C, of size *(1x4)*. Results are summarized in Table 4.1.

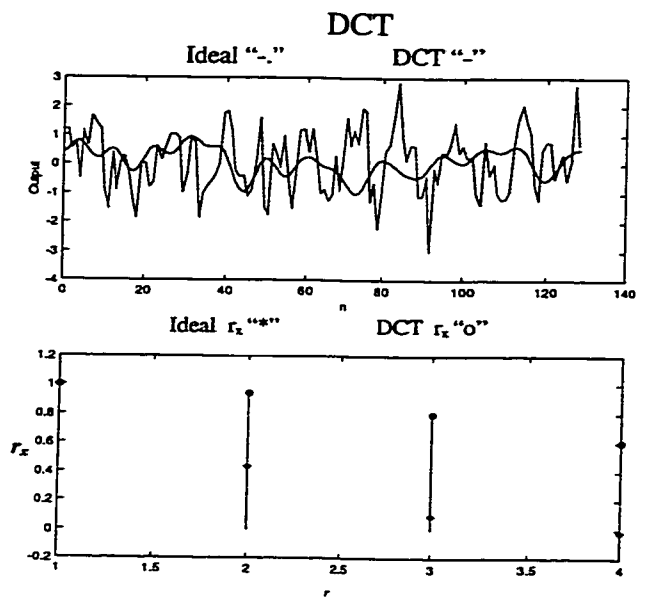


Figure 4.1 AR Signal DCT Reconstruction

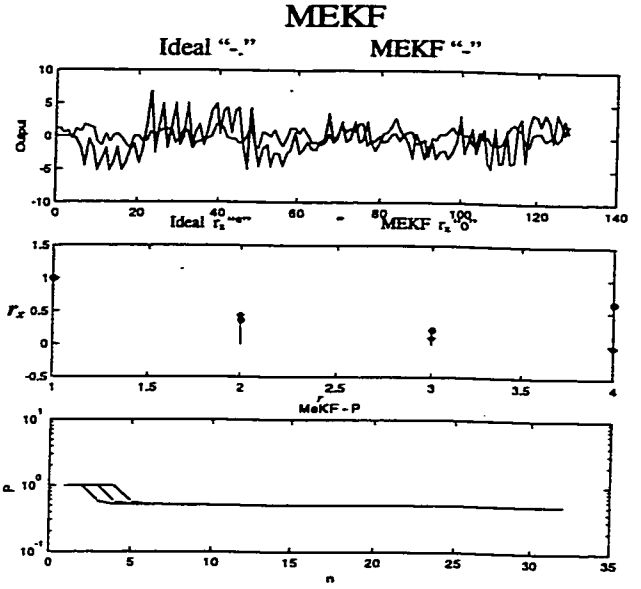


Figure 4.2 AR Signal MEKF Reconstruction

Method / Errors	DCT	MEKF
Signal MSE	0.0091	0.0488
Auto-Correlation MSE	0.2877	0.1192

Table 4.1 MEKF modeling of an AR signal

It is noted although the DCT transformation has a lower MSE, however, its auto-correlation function, r_x, poorly approximates that of the ideal signal. Furthermore, it should also be noted that DCT has an unfair advantage over MEKF in that it is a non-causal analysis-synthesis reconstruction, *i.e.* it is block-based transform. The a posteriori error covariance, P, in Figure 4.2 shows that P is nearly Toeplitz. This indicates that the

error signal is stationary, which in turn implies, as expected, the MEKF reconstruction is stationary.

4.2.2.2 MEKF modeling of an ARMA signal

The second set of simulations was performed on a ARMA sequence where the actual model is described by $a_{Ideal}=[0.5,-0.1]$, $B_{ideal}=[0.2,-0.2]$, $w_k \sim(0,1)$, Figure 4.3 shows 1:4 analysis-synthesis using DCT while application of the MEKF algorithm is depicted in Figure 4.4. Results are summarized in Table 4.2.

Evidently, MEKF performs poorly, as expected, when the dynamics of the process are ARMA, which has lower entropy with respect to a comparable AR process. Based on this observation, in the next section we offer a hybrid analysis-synthesis method for images, where depending on the local entropy of the image either ZOH or MEKF reconstruction is employed.

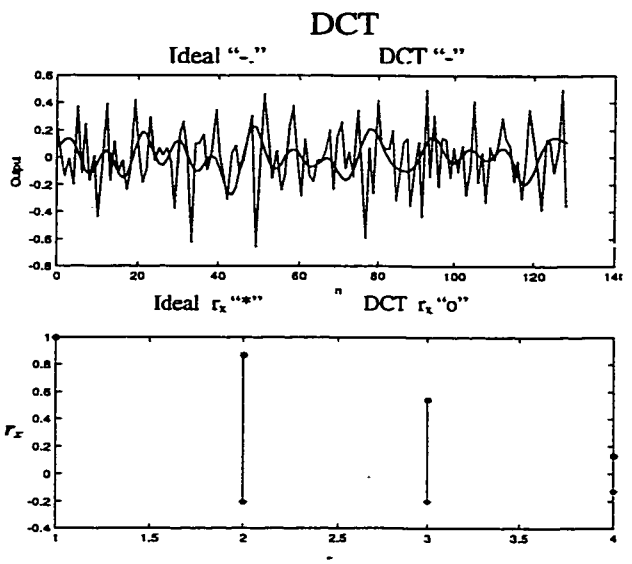


Figure 4.3 ARMA Signal DCT Reconstruction

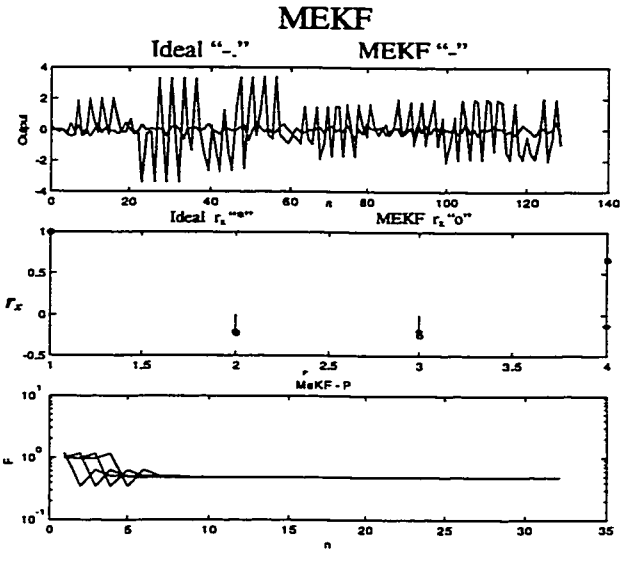


Figure 4.4 ARMA Signal MEKF Reconstruction

Error \ Method	DCT	MEKF
Signal MSE	0.0089	0.3867
Auto-Correlation MSE	0.4477	0.1631

Table 4.2 MEKF modeling of an ARMA signal

4.3 2-D Maximum Entropy Kalman Filter (MEKF) for modeling

To apply MEKF to 2-D signals, particularly images, two issues must be addressed firstly, dimensions of the filter and secondly, the range of approximated signal. The most convenient way to address the former is to vectorize all segments within a signal and treat them as a 1-D signal, in which case the 2-D correlation of the signal is lost. A convenient way of addressing this issue is to expand a 1-D filter to a 2-D one assuming separability of the filter and auto-correlation function (which in general, one does not imply the other [19][43]). Assume that $\boldsymbol{a}$ is the optimal, linear strictly-causal, prediction filter over the signal $\boldsymbol{U}$. That is,

$$\overline{U}(x,y) = a(x-N_x, y-N_y)U(x-N_x, y-N_y) + \ldots + a(x-1, y-1)U(x-1, y-1) \quad (4.27)$$

where x and y are the two spatial coordinates, $\overline{U}$ is the predicted signal, and N_x and N_y are the segment sizes.
And further assuming stationarity within all segments, Eq. (4.27) can be written as follows:

$$\overline{U}(x,y) \approx a(N_x, N_y)U(x-N_x, y-N_y) + \ldots + a(1,1)U(x-1, y-1) \quad (4.28)$$

Finally, assuming a sub-optimal separable prediction filter, Eq. (4.28) may be written as the following:

$$\overline{U}(x,y) \approx a^R(N_x)a^C(N_y)U(x-N_x, y-N_y) + \ldots + a^R(1)a^C(1)U(x-1, y-1) \quad (4.29)$$

where a^R and a^C are the prediction filters in row and column directions respectively.

This in turn can be expressed as follows:

$$\overline{U}(x,y) \approx (a^R \otimes a^C)U \tag{4.30}$$

where $\otimes$ is the Kronecker multiplication operation.

The question is whether such a separable expansion can result in a ME sequence. Assuming a separable Kalman filter where a^R and a^C are the ME prediction filters in row and column direction respectively and similarly R^R and R^C are the correlation matrices in row and column direction, the ME condition [7] may be written as follows:

$$\begin{aligned} &Ra = r \\ &(R^R \otimes R^C)(a^R \otimes a^C) = r \\ &(R^R a^R) \otimes (R^C a^C) = r \end{aligned} \tag{4.31}$$

where r is the auto-correlation function.

The last equality shows that, under the assumptions of filter and auto-correlation function separability, a ME sequence can be achieved.

The second point of interest in 2-D applications, especially images, is that the predicted pixels must lie within a given range, typically between 0 and 255. These constraints may be imposed on the Kalman gain. Although, this formulation is by no means unique however, it is convenient. Thus application of Augmented Lagrange Multiplier (ALM) [4][55] over the original cost function, one obtains the following cost function.

$$\begin{aligned} G_k &= tr(P_k) + \sum_{j=1}^{N} \lambda_{jk} \psi_{jk} + \alpha_{jk} \psi_{jk}^2 + \sum_{i=1}^{N} \mu_{ik} \varphi_{ik} + \beta_{ik} \varphi_{ik}^2 \\ \psi_{jk} &= max(x_{jk} - x_{MAX}, -\frac{\lambda_{jk}}{2\alpha_{jk}}) \\ \varphi_{ik} &= max(-x_{ik} + x_{MIN}, -\frac{\mu_{jk}}{2\beta_{jk}}) \end{aligned} \tag{4.32}$$

where, $tr(\boldsymbol{P})$ is the trace of the error covariance matrix,
λ and μ are the Lagrange multipliers,
ψ and ϕ are the cost functions,
α and β are the cost function step sizes, and
x_{MAX} and x_{MIN} are the constraints.

Differentiating Eq. (4.32) with respect to the Kalman gain $\boldsymbol{K_k}$ the following is obtained.

$$\frac{dG_k}{d\boldsymbol{K}_k}=\frac{dtr(\boldsymbol{P}_k)}{d\boldsymbol{K}_k}+\sum_{j=1}^{N}\lambda_{jk}\frac{d\psi_{jk}}{d\boldsymbol{K}_k}+\alpha_{jk}\frac{\psi_{jk}^2}{d\boldsymbol{K}_k}+\sum_{i=1}^{N}\mu_{ik}\frac{\varphi_{ik}}{d\boldsymbol{K}_k}+\beta_{ik}\frac{\varphi_{ik}^2}{d\boldsymbol{K}_k} \tag{4.33}$$

The derivative of a scalar with respect to a matrix is defined as

$$\frac{dG}{d\boldsymbol{K}}=\begin{bmatrix}\frac{dG}{dK_{11}} & \cdots & \frac{dG}{dK_{1M}}\\ \vdots & \cdots & \vdots\\ \frac{dG}{dK_{N1}} & \cdots & \frac{dG}{dK_{NM}}\end{bmatrix} \tag{4.34}$$

Noting the following relationships

$$\left(\frac{d\psi_j}{dK_{nm}}\right)_k=\begin{cases}0 & x_j-x_{MAX}<-\dfrac{\lambda_j}{2\alpha_j}\\ \left(\dfrac{dx_j}{dK_{nm}}\right)_k & Otherwise\end{cases} \qquad m\in\{1\cdots M\} \tag{4.35}$$

$$x_{jk+1}=x_{jk+1}^{-}+\sum_{i=1}^{M}K_{ji}y_{ik+1}-\sum_{l=1}^{N}x_{lk+1}^{-}\sum_{m=1}^{M}K_{jm}C_{ml} \qquad j\in\{1\cdots N\}$$

where k is the time index, n and j are process state indices, and m is the observation state index, it follows that,

$$\left(\frac{dx_j}{dK_{nm}}\right)_k=\begin{Bmatrix}0 & j\neq n\\ y_m-\sum_{l=1}^{N}C_{ml}x_l^{-} & Otherwise\end{Bmatrix}_k \tag{4.36}$$

which is independent of process state index j. Hence the following equality is obtained.

$$\sum_{j=1}^{N}\lambda_j\frac{d\psi_{jk}}{d\boldsymbol{K}_k}=(\boldsymbol{J}\lambda\boldsymbol{e})_k \tag{4.37}$$

where λ, $\boldsymbol{J}$ and $\boldsymbol{e}$ are defined as follows:

$$J = \begin{Bmatrix} J_{ij} = 0 & i \neq j \\ J_{ii} = 0 & x_i - x_{MAX} < -\dfrac{\lambda_i}{2\alpha_i} \\ J_{ii} = 1 & Otherwise \end{Bmatrix}_{(NxN)}$$

$$e = \left(y - Cx^{-}\right)^T \qquad (4.38)$$

$$\lambda = [\lambda_1 \cdots \lambda_N]$$

Similar sets of equations may be found for other terms of Eq. (4.33), as follows:

$$\sum_{j=1}^{N} \alpha_j \frac{d\psi_{jk}^2}{dK_k} = 2Kxe - 2x_{MAX} J\alpha e$$

$$\sum_{i=1}^{N} \mu_i \frac{d\varphi_{ik}^2}{dK_k} = -L\mu e \qquad (4.39)$$

$$\sum_{i=1}^{N} \beta_i \frac{d\varphi_{ik}^2}{dK_k} = 2Mxe - 2x_{MIN} L\beta e$$

where L, M and K are defined as follows:

$$L = \begin{Bmatrix} L_{ij} = 0 & i \neq j \\ L_{ii} = 0 & -x_i + x_{MIN} < -\dfrac{\mu_i}{2\beta_i} \\ L_{ii} = 1 & Otherwise \end{Bmatrix}_{(NxN)} \quad K = \begin{Bmatrix} K_{ij} = 0 & i \neq j \\ K_{ii} = 0 & x_i - x_{MAX} < -\dfrac{\lambda_i}{2\alpha_i} \\ K_{ii} = \alpha_i & Otherwise \end{Bmatrix}_{(NxN)}$$

$$M = \begin{Bmatrix} M_{ij} = 0 & i \neq j \\ M_{ii} = 0 & -x_i + x_{MIN} < -\dfrac{\mu_i}{2\beta_i} \\ M_{ii} = \beta_i & Otherwise \end{Bmatrix}_{(NxN)} \qquad (4.40)$$

Using Eq. (4.37) through Eq. (4.40) and equating Eq. (4.33) to zeros, the following expression is obtained:

$$K_k = K_k^U + \frac{0.5J\lambda - 0.5L\mu - x_{MAX} J\alpha - x_{MIN} L\beta + NK_k e^T + Nx^{-}}{CP^{-}C^T + R} e \qquad (4.41)$$

where K_k^U is the unconstrained Kalman gain,
N is $M+K$,
x^{-} is the prior estimate of the states, and
P^{-} is the prior of the error covariance matrix.

It is noted, in Eq. (4.41), that K_k appears on both sides of the equality. Hence an iterative solution, using K_{k-1} instead of K_k on the right hand side, is used to obtain a solution for K_k. Furthermore, with respect to Eq. (4.41) it is seen that the unconstrained solution, the initial guess, is the correct one if all the constraints are satisfied.

4.3.1 2-D MEKF modeling algorithm

The following pseudo-code summarizes one possible implementation of 2-D MEKF for image modeling.

1. Initialization
   - Set N, analysis-synthesis ratio
   - Segment the signal ($\sqrt{N} \times \sqrt{N}$)
   - Proceed with the same initialization steps as Section 4.2.1
2. Analysis – encode all segments
   - Obtain row and column ordered vectors, based on the current segment
   - Obtain the a^R, a^C, Q^R and Q^C, for each vector above, using Levinson-Durbin algorithm
   - If $Q^R Q^C$ is less than a predetermined threshold, transmit the average value and perform ZOH to recover the segment, as discussed in Section 2, otherwise,
   - Update C as described by equations (4.26) and (4.19-b)
   - Calculate A as the state description of Eq. (4.30)
   - Generate observation sequence y_k
3. Synthesis – decode all observation points
   - Apply standard Kalman filter, with parameters A,B, C and observation points y_k, to reconstruct row ordered image vector
   - If the reconstructed points violate the constraints, use Eq. (4.41)

4.3.2 MEKF modeling of 2-D signals

In this section we present the simulation results of MEKF, DCT, and MEKF-ZOH modeling method using *512x512x8* images of *Baboon*, *Barbara*, and *Lenna* with 1:16 analysis-synthesis ratio. The original images are shown in Figure 4.5, Figure 4.6 and Figure 4.7. For each image we will provide an image of $Q^R Q^C$ to identify a proper threshold.

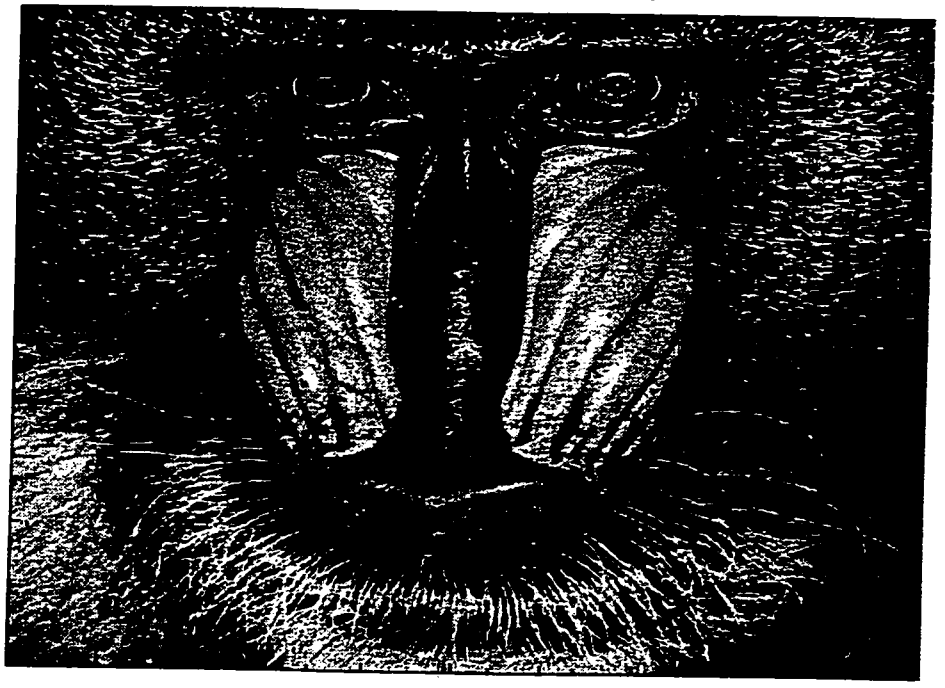


Figure 4.5 Original Baboon Image

In Figure 4.8, Figure 4.9 and Figure 4.10 the dark areas correspond to sections of the image where the entropy is minimum hence, the use of ZOH, using the local average, is appropriate. However, regions with high entropy, light areas, are more appropriately modeled using MEKF. The threshold value of *0.08*, segment size of *4x4* for MEKF and MEKF-ZOH and *8x8* for DCT was found to provide a satisfactory result.

Figure 4.6 Original Barbara Image

Figure 4.7 Original Lenna Image

For comparison purposes analysis-synthesis based purely on MEKF and DCT are also provided. Based on these variance images, we expect that MEKF-ZOH to provide a superior performance for *Baboon* and *Barbara* images and not for *Lenna* image.

It is observed, from Figure 4.11 through Figure 4.19, that in *Baboon* image, use of MEKF-ZOH has prevented the blockiness artifact, as compared with the DCT reconstruction. It has also preserved the texture in *Barbara* image whereas, no clear benefit is observed in *Lenna* image. It is noted that in the former two, MSE does not present an appropriate subjective measure.

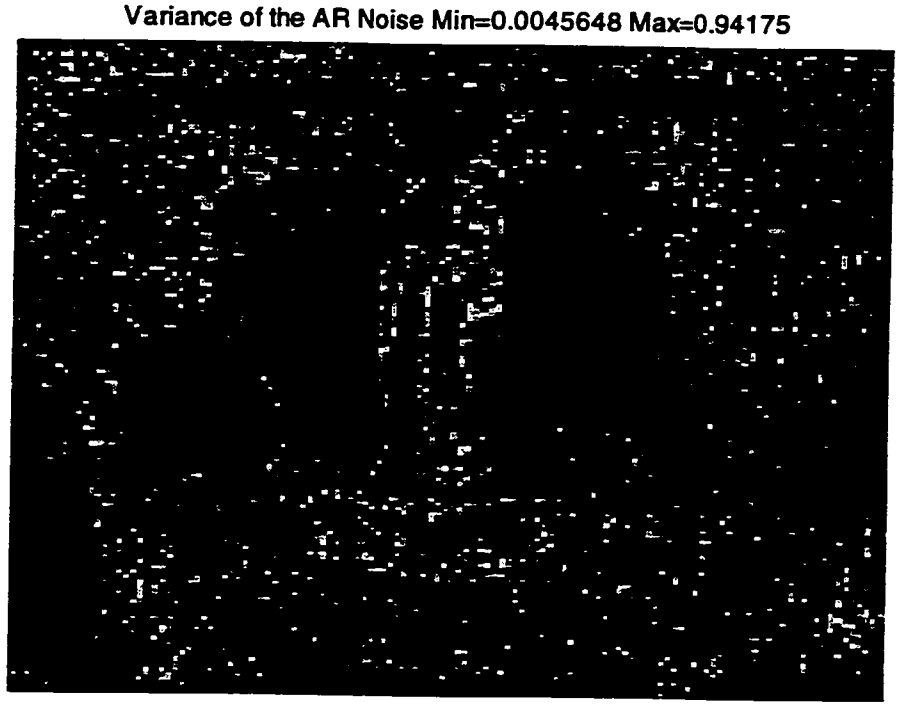


Figure 4.8 Baboon AR Variance Distribution

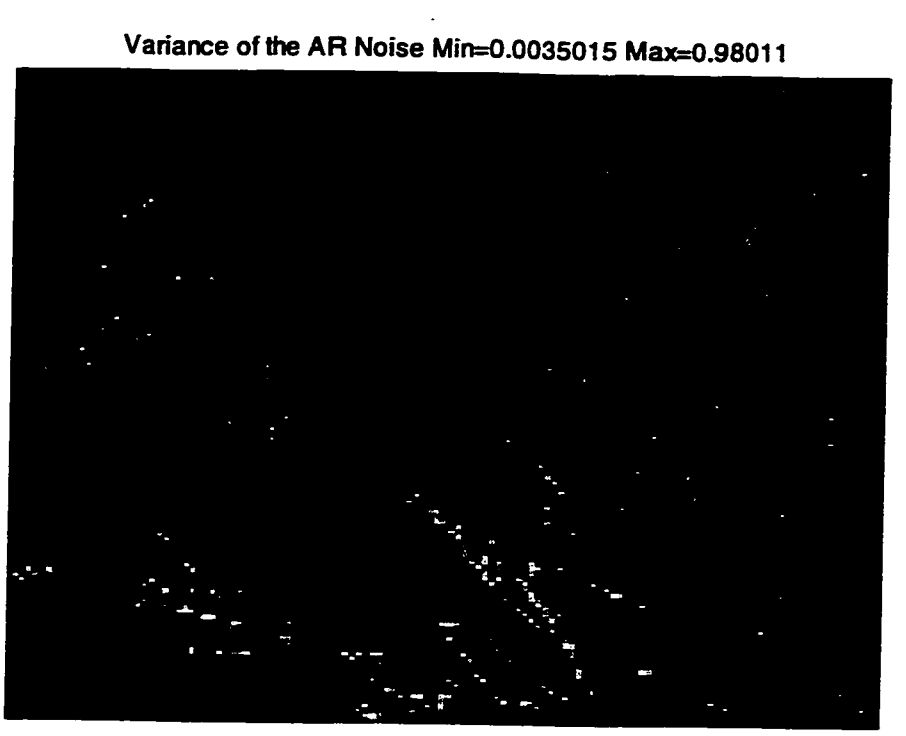


Figure 4.9 Barbara AR Variance Distribution

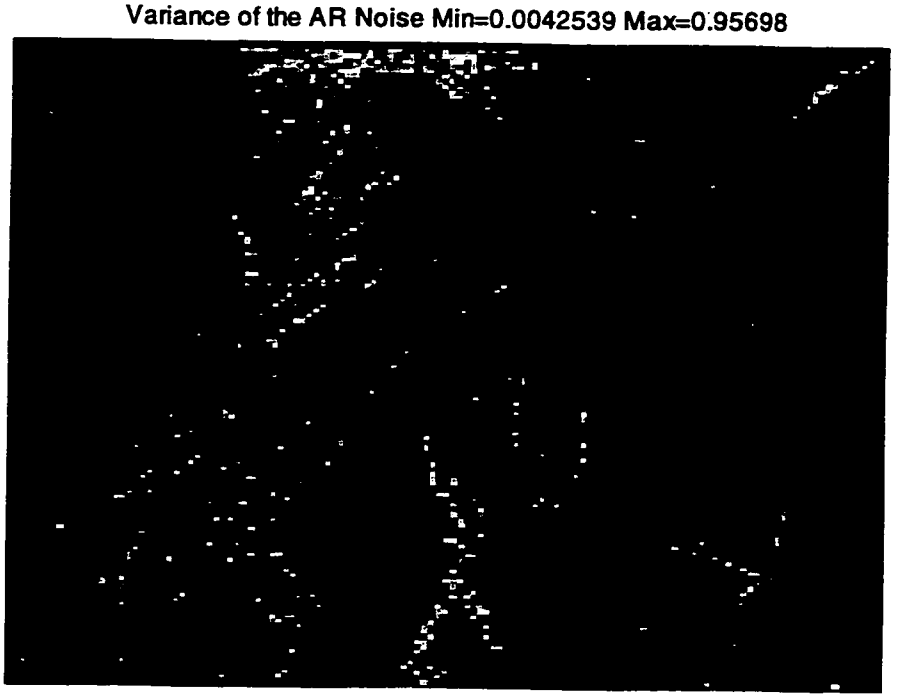


Figure 4.10 Lenna AR Variance Distribution

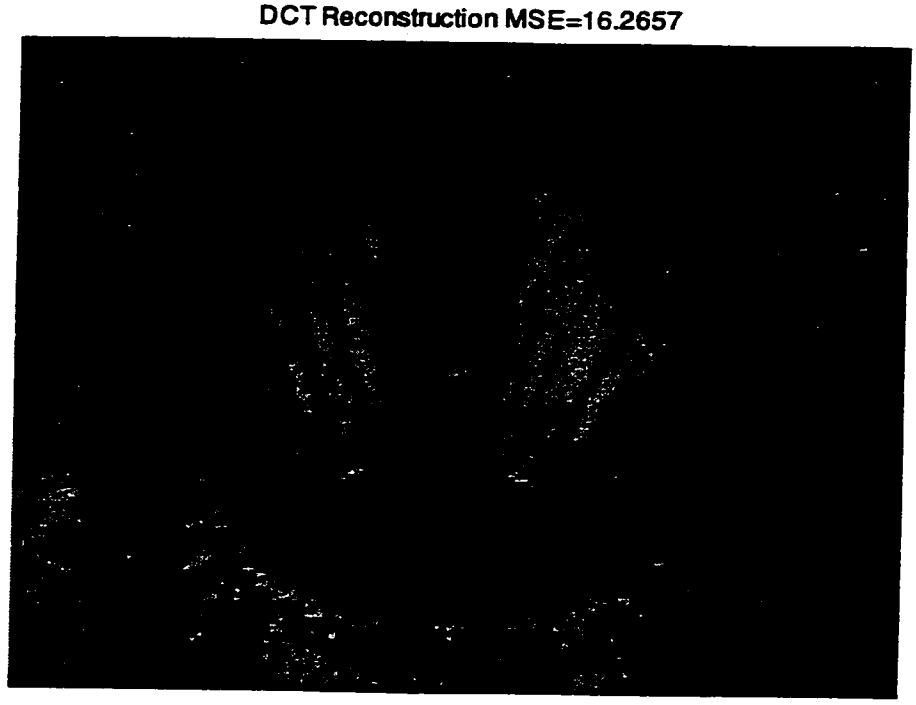


Figure 4.11 Baboon Reconstruction Based on DCT

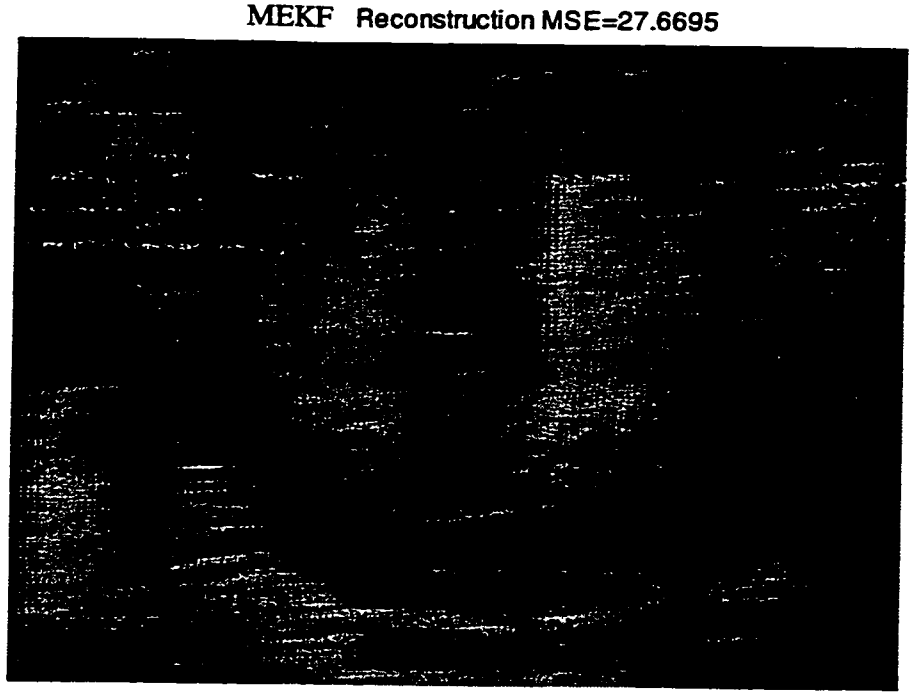


Figure 4.12 Baboon Reconstruction Based on MEKF

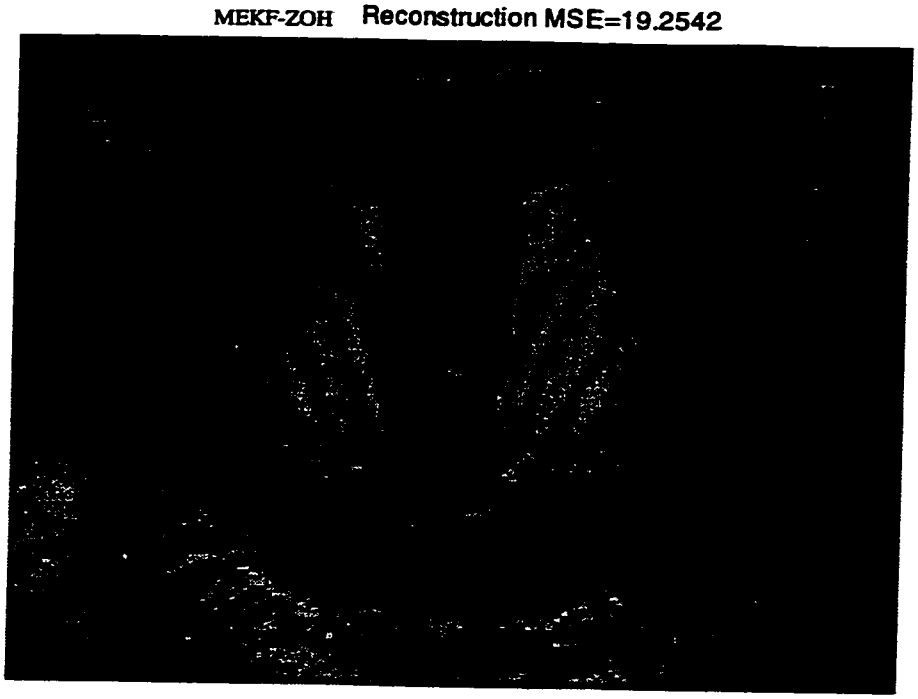


Figure 4.13 Baboon Reconstruction Based on MEKF-ZOH

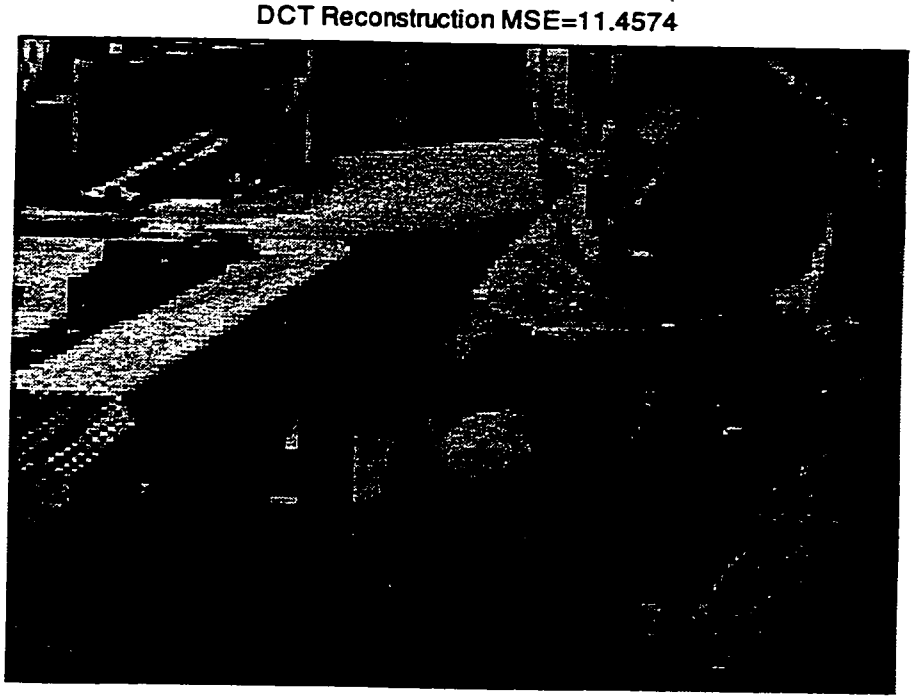


Figure 4.14 Barbara Reconstruction Based on DCT

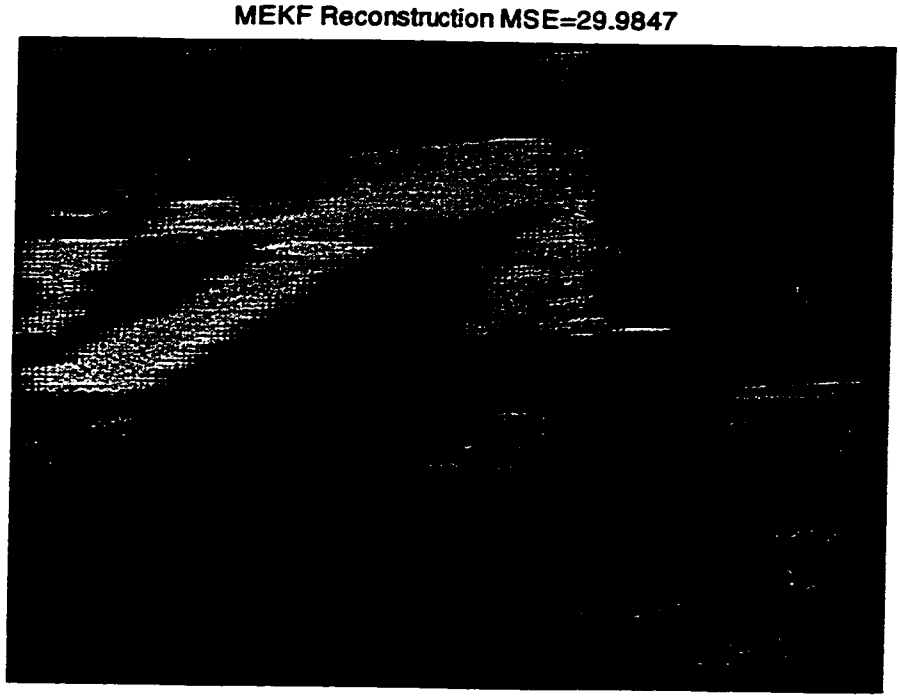


Figure 4.15 Barbara Reconstruction Based on MEKF

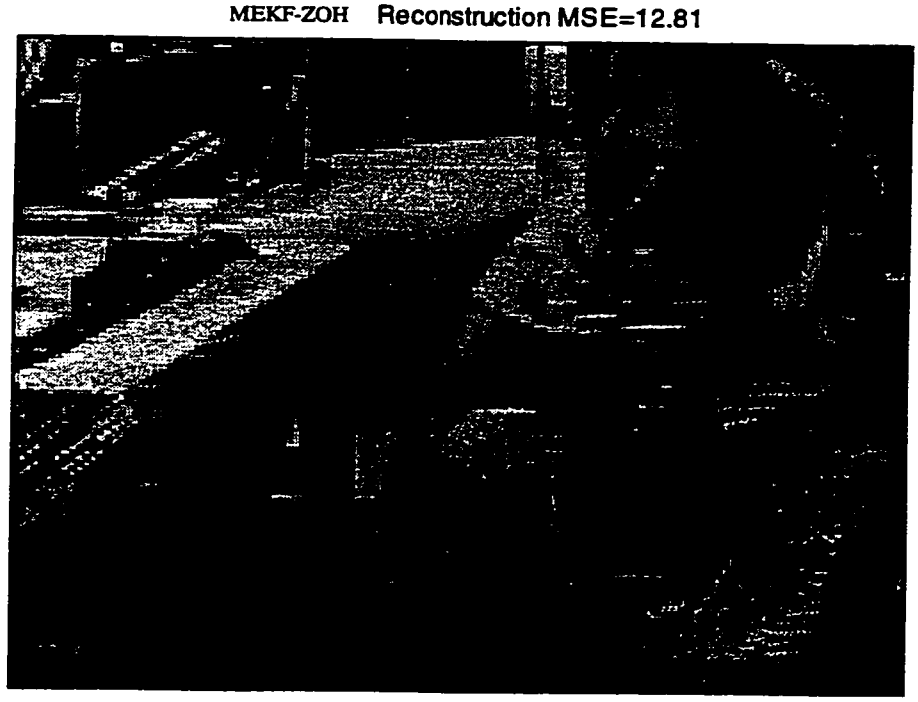


Figure 4.16 Barbara Reconstruction Based on MEKF-ZOH

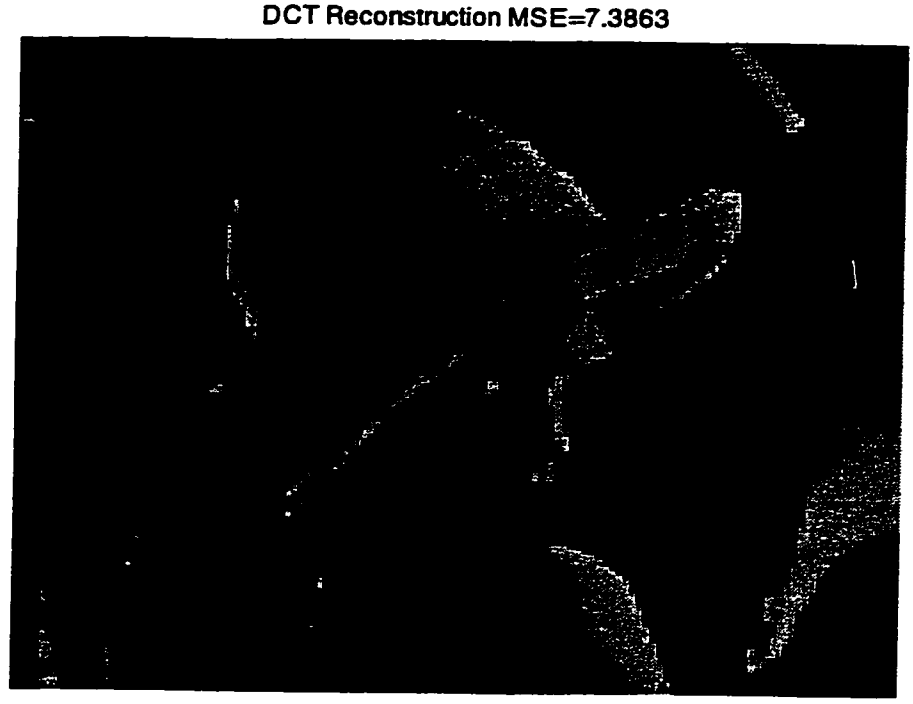


Figure 4.17 Lenna Reconstruction Based on DCT

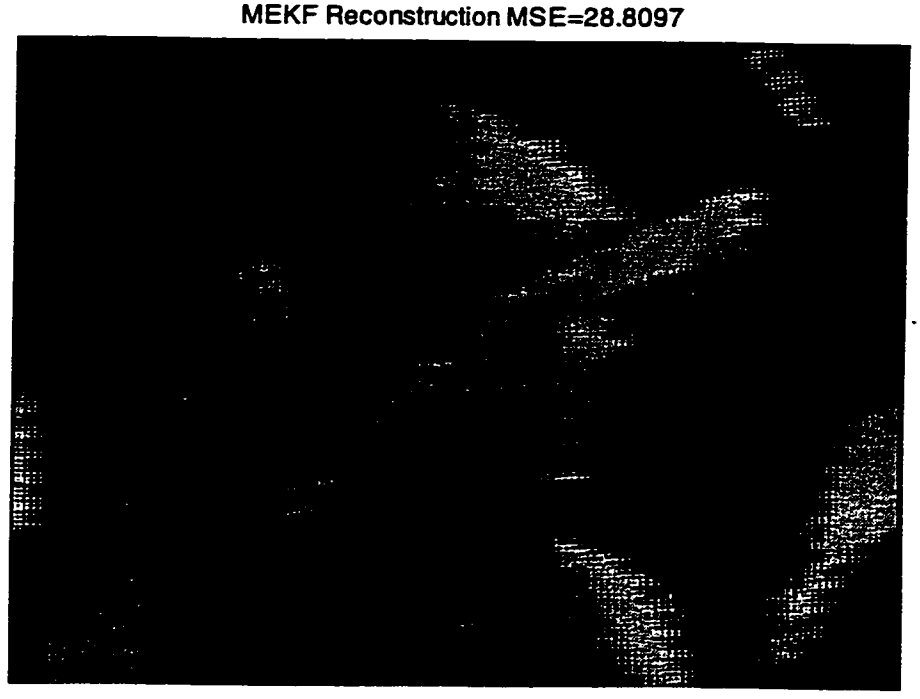


Figure 4.18 Lenna Reconstruction Based on MEKF

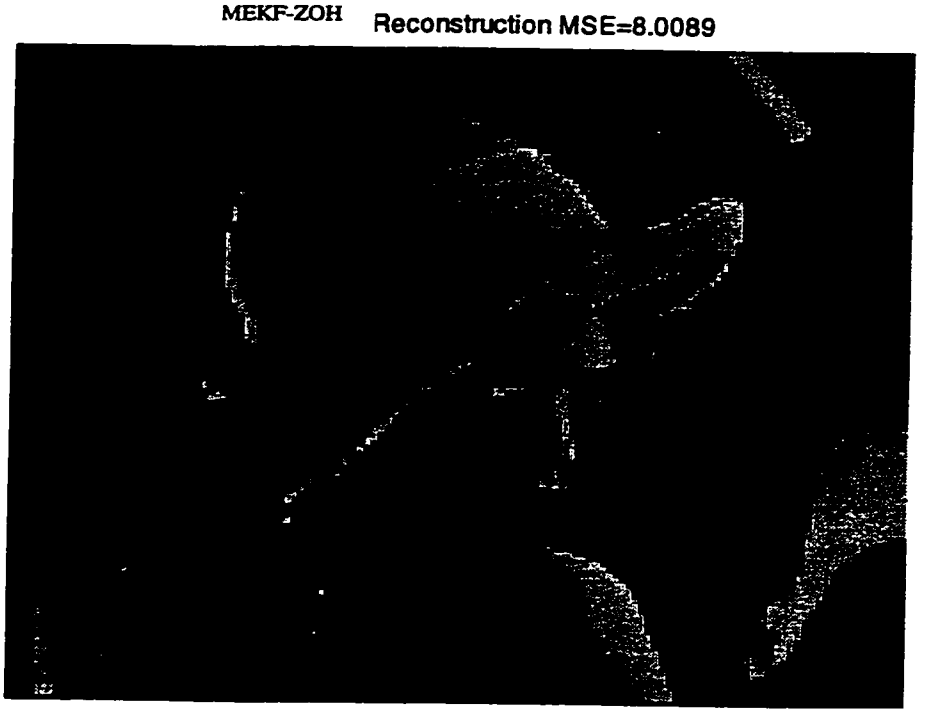


Figure 4.19 Lenna Reconstruction Based on MEKF-ZOH

4.3.3 2-D Maximum Entropy Kalman Filter (MEKF) for compression

In the previous section we described a modeling approach based on ME criterion. In this section we propose a compression scheme based on MEKF-ZOH. This method separates the signal into two parts. The first part is the observation sequence y_k, as described by Eqs. (4.26), (4.19-*b*), (4.7), (4.5) and if necessary Eq. (4.18), and the second part is the error signal. The error signal is defined as the difference between the reconstructed signal and the original signal, only at segments where $Q^R Q^C$ is below a prescribed threshold, and zero otherwise. These signals are then uniformly quantized, coded and decoded using JPEG standard [46]. Details of implementation are described in the following pseudo-code.

Initialization

- The same as Section 4.3.1

2. Compression
   - For each segment obtain y_k, as in Section 4.3.1, and pass it through a μ-law compressor [19]
   - Uniformly quantize the resulting sequence using 8 bit accuracy
   - Reconstruct the signal using the decoded observations
   - Generate the error signal, as described above, using the reconstructed signal
   - Code the error signal by passing through a μ-law compressor, uniform quantizer and finally through a JPEG coder

Reconstruction

- Pass both y_k and the error signal through μ-law expander, with appropriate parameters
- If $Q^R Q^C$ is above a given threshold use the method of Section 4.3.1 to reconstruct the segment, otherwise
- Reconstruct using JPEG decoder

For comparison purposes, the reconstructed signal using MEKF-ZOH and JPEG standard are presented, Figure 4.21 to Figure 4.28. For the JPEG method, the quality factor of 25 was used, where the quality factor 100 corresponds to no compression. MEKF-ZOH parameters were as follows:

- Segment size of 8x8
- 1:64 ratio
- μ-law compandor with parameter p 20 and 5 for error and y_k sequences respectively, and ν 120. See Eq. (4.42)

- JPEG quality factor of 5 for the error signal
- $Q^R Q^C$ threshold value of 8e-2
- The μ-law compandor is defined in Eq. (4.42).

$$\mu_{in} = \frac{v\,ln\left(1+\frac{p}{v}|x|\right)}{ln(1+p)} sgn(x)$$
$$\mu_{out} = \frac{v\left[e^{|\mu_{in}|\frac{ln(1+p)}{v}} - 1\right]}{p} sgn(\mu_{in}) \quad (4.42)$$

where x is the input,
v and p are the μ-law compandor parameters,
μ_{in} is the compressor output and
μ_{out} is the expander output
The original images are shown in The original images are shown in Figure 4.5, Figure 4.6, Figure 4.7 and Figure 4.20.

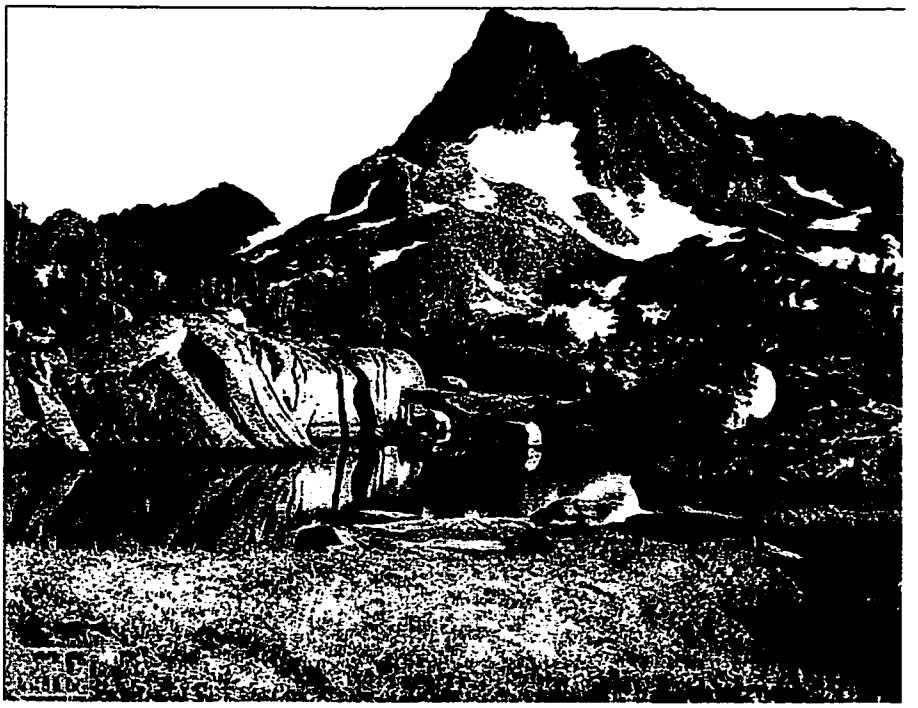


Figure 4.20 Original Mountain Image

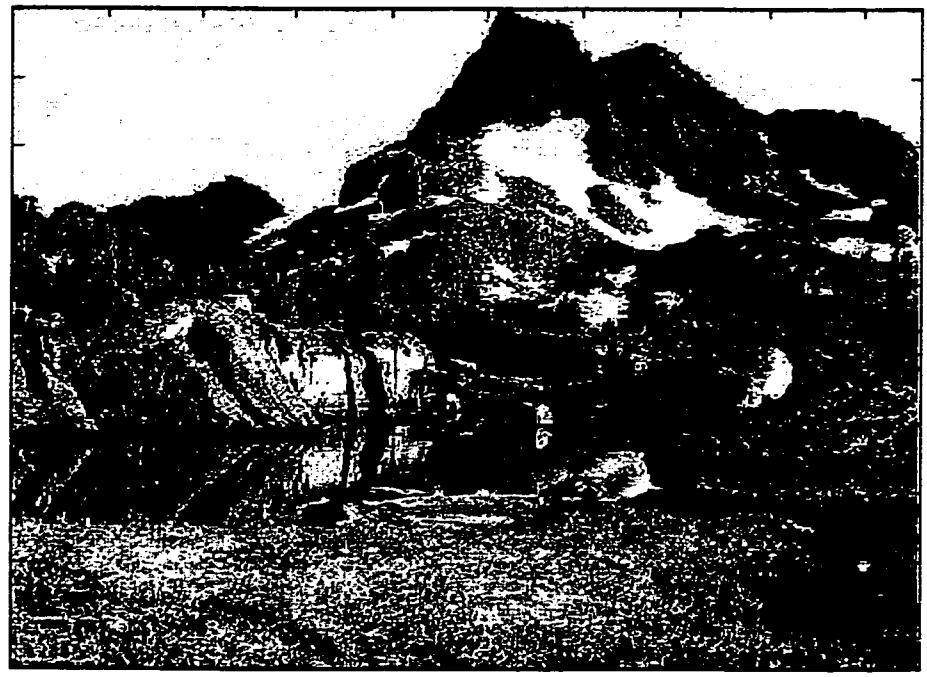


Figure 4.21 MEKF-ZOH Compression of Mountain Image

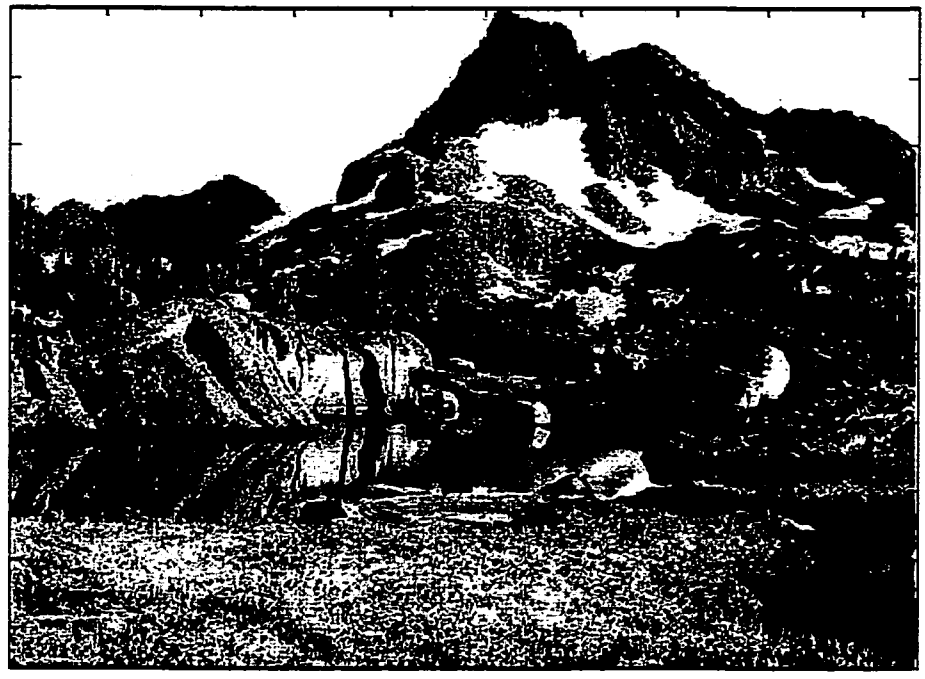


Figure 4.22JPEG Compression Compression of Mountain Image

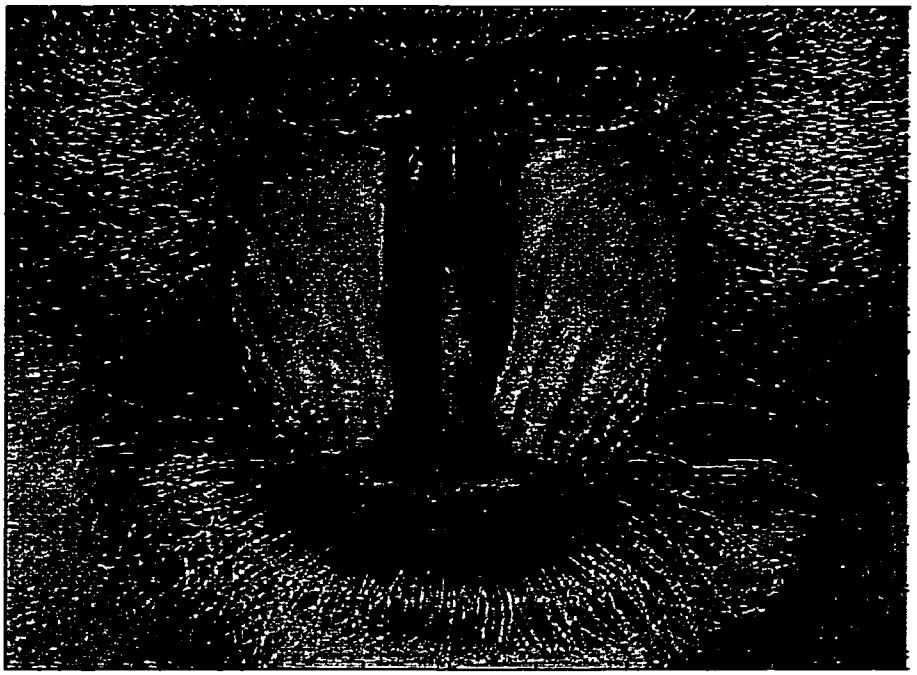


Figure 4.23MEKF-ZOH Compression Compression of Babbon Image

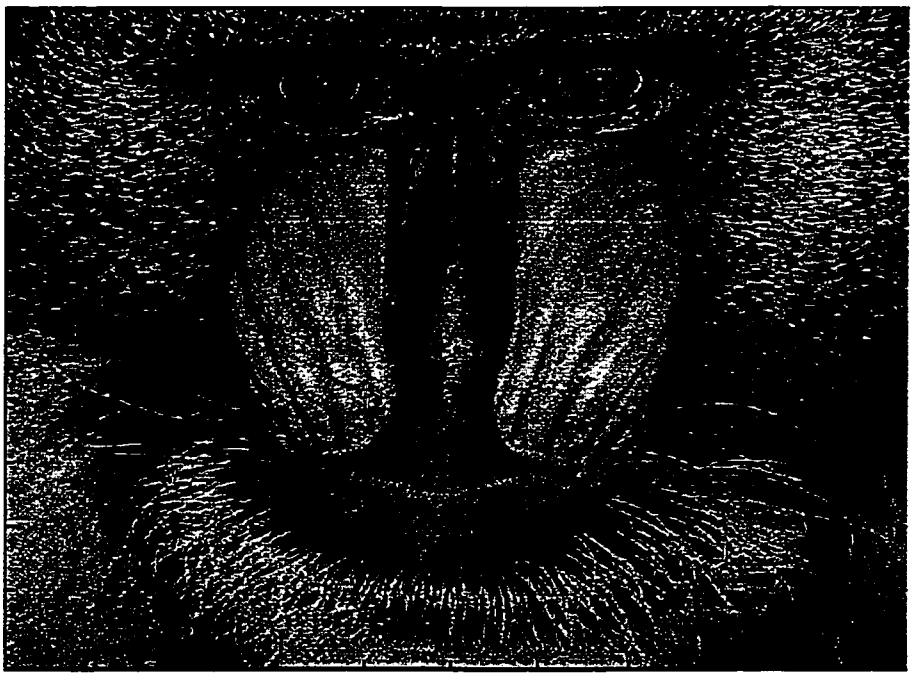


Figure 4.24JPEG Compression Compression of Babbon Image

Figure 4.25MEKF-ZOH Compression Compression of Barbara Image

Figure 4.26JPEG Compression Compression of Barbara Image

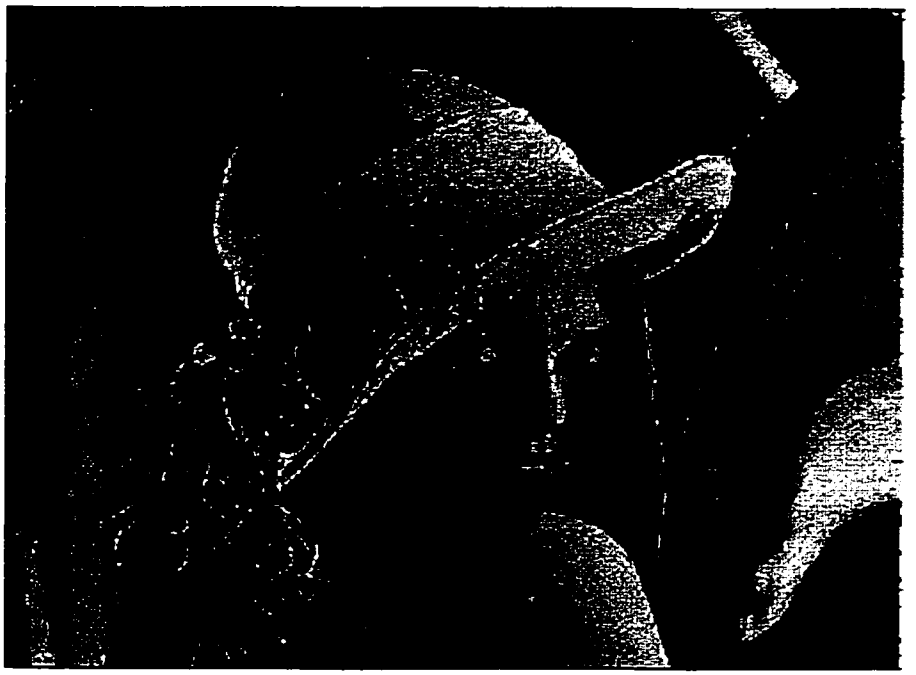


Figure 4.27MEKF-ZOH Compression Compression of Lenna Image

Figure 4.28JPEG Compression Compression of Lenna Image

4.3.3.1 Discussion and conclusion

Image \ Method	JPEG	MEKF-ZOH
Mountain	0.1439	0.1289
Baboon	0.1106	0.0840
Barbara	0.0783	0.0781
Lenna	0.0569	0.0513

Table 4.3 Compression ratio for MEKF-ZOH and JPEG test images

Image \ Method	JPEG	MEKF-ZOH
Mountain	13.95	19.22
Baboon	9.471	15.15
Barbera	5.800	9.222
Lenna	4.358	6.725

Table 4.4 MSE for MEKF-ZOH and JPEG test images

With respect to Table 4.3, it is noted that for compression ratios less than that of JPEG, MEKF-ZOH has retained the sharpness of the image in textured regions and edges. As expected, noting Table 4.4, the MSE of an equivalent JPEG reproduction is lower than MEKF-ZOH. However, it should be noted that neither the quantizer table nor the entropy-coder table were optimized for MEKF-ZOH usage. Certainly, lower compression ratios at lower MSE can be obtained if the zero regions of the error signal would have been coded separately and a smoothing Kalman filter was employed, versus the strictly causal one used for these simulations.

4.4 1-D Maximum Entropy Kalman Filter (MEKF) for estimation

In this section, we develop an estimation scheme based on the principle of ME. We differentiate between modeling and estimation on basis of controllability of the operating environment. Unlike modeling, in estimation, the observation matrix, C, is dictated externally. As such we do not expect to obtain results as good as those in modeling case. Indeed, for a given C, a ME sequence may not be realizable. Thus for practical purposes certain assumptions must be made. The theoretical conditions are developed. Justification and implication of implicit assumptions are also discussed.

In order for $\hat{x}$ to be an ME sequence, comparing Eqs (4.3) and (4.8), it is noted that firstly, the term $K_{k+1}\, y_{k+1}$ must be whitened and secondly, A should be chosen such that the first terms of the two equations become equal.

Thus pre-multiplying Eq. (4.3) by S (a whitening filter), the following is obtained:

$$\begin{aligned} S\hat{x}_{k+1} &= S(I - K_{k+1}C)A_k\hat{x}_k \\ &\quad + SK_{k+1}y_{k+1} \\ \bar{\hat{x}}_{k+1} &= S(I - K_{k+1}C)A_kS^{-1}\bar{\hat{x}}_k \\ &\quad + SK_{k+1}y_{k+1} \end{aligned} \tag{4.43}$$

Furthermore, equating the first terms of Eqs (4.8) and (4.43), the following expression for A, the state transition matrix, is obtained:

$$A_{k+1} = (I - K_{k+1}C)^{-1}S^{-1}A^{ME}S \tag{4.44}$$

where A^{ME} is defined in Eq. (4.8).

An expression for S may be obtained as follows [29]:

$$
\begin{aligned}
&S^T S = V^{-1} \\
&V = E\left[K_k y_k y_k^T K_k^T\right] \\
&V = E\left[K_k (Cx_k + v_k)(Cx_k + v_k)^T K_k^T\right] \\
&\quad = E\left[(K_k C_k x_k + K_k v_k)(x_k^T C_k^T K_K^T + v_k^T K_K^T)\right] \\
&\quad = E\begin{bmatrix} K_k C_k x_k x_k^T C_k^T K_K^T + K_k C_k x_k v_k^T K_K^T \\ + K_k v_k x_k^T C_k^T K_K^T + K_k v_k v_k^T K_K^T \end{bmatrix}
\end{aligned}
\tag{4.45}
$$

Assuming stationary signal, Eq. (4.45) can be written as follows:

$$
\begin{aligned}
V = KCR_x C^T K^T + KC\,E\left[x_k v_k^T\right]K^T \\
+ K\,E\left[v_k x_k^T\right]C^T K^T + KR_v K^T
\end{aligned}
\tag{4.46}
$$

where R_x is the auto-correlation matrix of the process x_k, and
R_v is the auto-correlation matrix of the observation noise v_k.

Assuming uncorrelated process dynamics and observation noise the second and third terms of (4.46) are eliminated. Hence the following expression for S is obtained:

$$
\begin{aligned}
&V = K\left(CR_x C^T + R_v\right)K^T \\
&V = KR_y K^T \\
&V^{-1} = \left(KR_y K^T\right)^{-1} \\
&S = \left(KR_y K^T\right)^{-\frac{1}{2}}
\end{aligned}
\tag{4.47}
$$

where R_y is the auto-correlation matrix of the observation process y_k. To ensure that the variance of $SK_{k+1}\,y_{k+1}$ is the same as the desired ME sequence of Eq. (4.6), the following modification to Eq. (4.47) is applied.

$$
S = \left(Q^{ME} KR_y K^T\right)^{-\frac{1}{2}} \tag{4.48}
$$

where Q^{ME} is defined in Eq. (4.7).

Thus, application of Eq. (4.43) and Eq. (4.44) ensures that the output of the Kalman filter, $\bar{\hat{x}}$, is indeed an ME sequence.

It is noted that in Eq. (4.47) for the inverse to exist, in the usual sense, $\boldsymbol{K}$ must be full rank. A necessary, but not sufficient, condition is that $\boldsymbol{K}$ must be NxN, which can only occur if $\boldsymbol{C}$ is also NxN. Noting that $\boldsymbol{K}$ may be expressed as follows [29].

$$\boldsymbol{K}_{k+1} = \boldsymbol{P}_{k+1}^{-}\boldsymbol{C}^{T}\left(\boldsymbol{C}\boldsymbol{P}_{k+1}^{-}\boldsymbol{C}^{T} + \boldsymbol{R}\right) \tag{4.49}$$

where $\boldsymbol{R}$ is defined in Eq. (4.1).
Furthermore, assuming noiseless observation, it follows that

$$\begin{aligned} \boldsymbol{C}\boldsymbol{K}_{k+1} &= \boldsymbol{C}\boldsymbol{P}_{k+1}^{-}\boldsymbol{C}^{T}\left(\boldsymbol{C}\boldsymbol{P}_{k+1}^{-}\boldsymbol{C}^{T}\right) \\ \boldsymbol{C}\boldsymbol{K}_{k+1} &= \boldsymbol{I} \end{aligned} \tag{4.50}$$

If $\boldsymbol{C}$ and $\boldsymbol{K}$ are both full ranks it follows from Eq. (4.50) that

$$\boldsymbol{K}_{k+1}\boldsymbol{C} = \boldsymbol{C}\boldsymbol{K}_{k+1} = \boldsymbol{I} \tag{4.51}$$

Under these assumptions, Eq. (4.47) simplifies to the following:

$$\begin{aligned} \boldsymbol{V} &= \boldsymbol{K}\left(\boldsymbol{C}\boldsymbol{R}_{x}\boldsymbol{C}^{T}\right)\boldsymbol{K}^{T} \\ \boldsymbol{V} &= \boldsymbol{R}_{x} \\ \boldsymbol{V}^{-1} &= \boldsymbol{R}_{x}^{-1} \\ \boldsymbol{S} &= \boldsymbol{R}_{x}^{\frac{1}{2}} \end{aligned} \tag{4.52}$$

In cases where the observation matrix $\boldsymbol{C}$ is not full rank we use Eq. (4.52) as an approximation of Eq. (4.47).

4.4.1 Implementation of the proposed MEKF

In this section, we describe an algorithm implementing MEKF, for the purpose of estimation. It is noted that in Eq. (4.44) and Eq. (4.47), the Kalman gain $\boldsymbol{K}$ must be known ahead of time. Although this may be achievable via a two-pass filter, however, it would be of little practical value. To overcome this problem, the following algorithm is suggested.

1. Initialization
   - Set N, number of states
   - Let $\boldsymbol{K}$ be a vector of small random number

- Segment the signal

2. Iteration for all segments

- Calculate S as given in Eq. (4.47) or Eq. (4.52), where K is equal to the Kalman gain of the previous segment.
- Calculate A based on Eq. (4.44), where K is equal to the Kalman gain of the previous segment.

2.1 Iterate in the current segment

- Apply standard Kalman filter with K_{k+1} y_{k+1} whitened by S and A as the transitional matrix.
- Update A as described by Eq. (4.44), where the current K is used for the update.

4.4.2 MEKF based estimation results

Two sets of simulations are described here. Both AR and ARMA processes have been simulated. In the following examples, no observation noise is assumed. To further evaluate the performance of MEKF, reconstruction resulting from the direct substitution of A^{ME} of Eq. (4.8) in Eq. (4.1) is also provided.

4.4.2.1 MEKF estimation of an AR signal

In the first set, the actual model, as given by Eq. (4.2), is described by an AR model a_{Ideal}=[0.5,-0.1,0.2,-0.3], $w_k \sim (0,1)$, and C=[0.25,0.25,0.25,0.25]. Such observation vector, C, effectively implies analysis-synthesis ratio of 1:4. Figure 4.29 shows the simulation results, where A^{ME} is directly substituted for A. Figure 4.30 depicts

application of the Section 4.4.1 algorithm in the same model. With respect to Table 4.5, the superior performance of MEKF is evident.

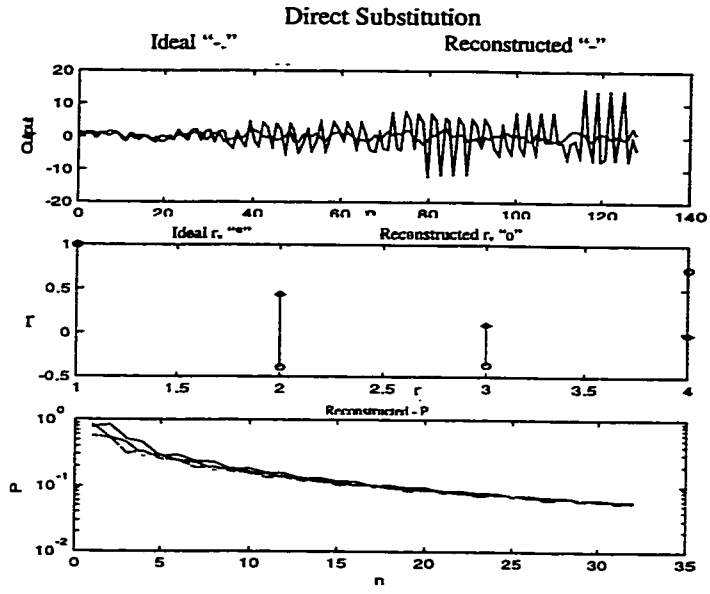


Figure 4.29 Estimation of an AR Signal using Direct Substitution in the Kalman Filter

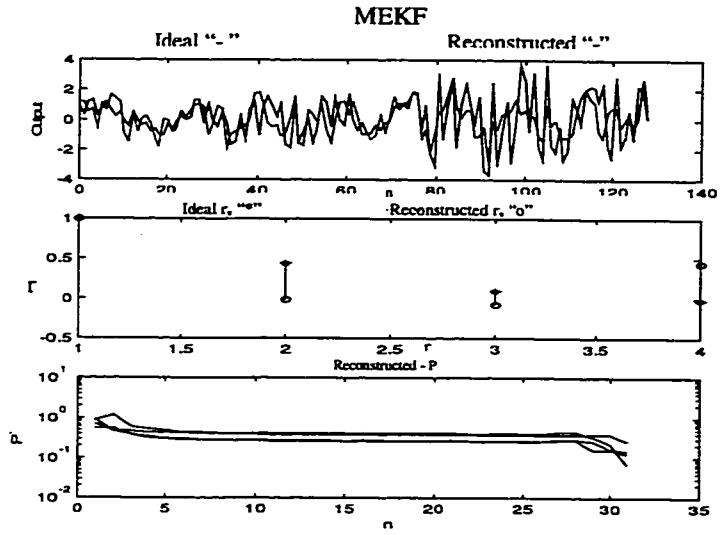


Figure 4.30 Estimation of an AR Signal using MEKF

Errors \ Method	Direct Substitution	MEKF
Signal MSE	0.1712	0.0136
Auto-Correlation MSE	0.3631	0.1096

Table 4.5 MEKF estimation of an AR signal

4.4.2.2 MEKF estimation of an ARMA signal

The second set of simulations was performed on a ARMA sequence where the actual model is described by $a_{Ideal}=[0.5,-0.1]$, $B_{ideal}=[0.2,-0.2]$, $w_k \sim (0,1)$, and $C=[0.25,0.25,0.25,0.25]$. Figure 4.31 shows using direct substitution while application of the MEKF algorithm is depicted in Figure 4.32. Results are summarized in Table 4.6.

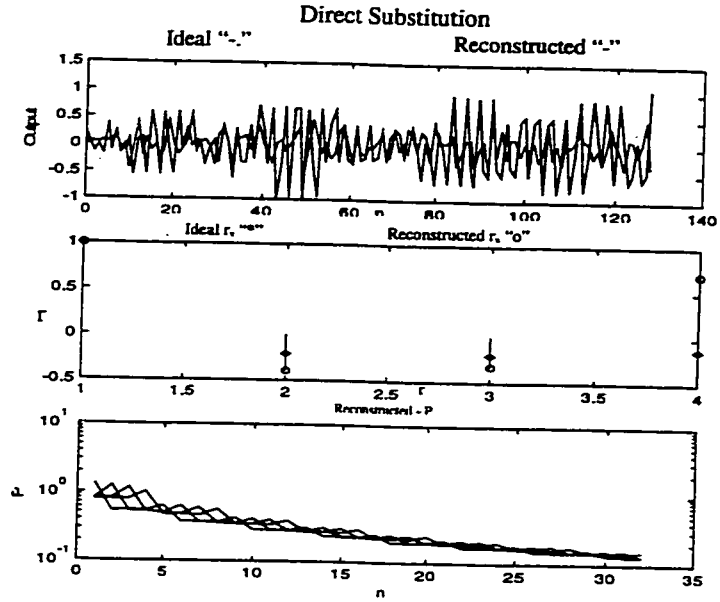


Figure 4.31 Estimation of an ARMA Signal using Direct Substitution in the Kalman Filter

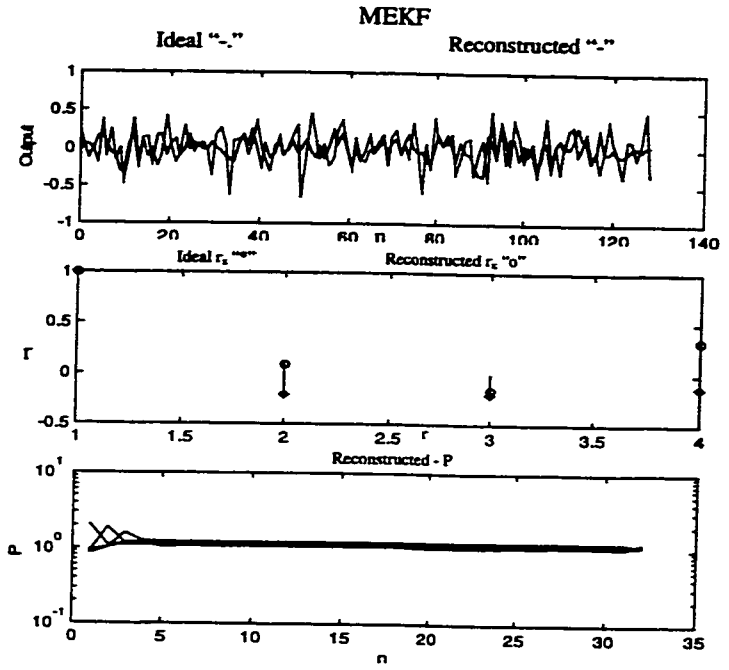


Figure 4.32 Estimation of an ARMA Signal using MEKF

Errors \ Method	Direct Substitution	MEKF
Signal MSE	0.0427	0.0089
Auto-Correlation MSE	0.1860	0.0784

Table 4.6 MEKF estimation of an ARMA signal

Comparing the above results with those of Section 2, It may seem that MEKF with a fixed C vector has a similar performance however, this is not the case in general. In fact the performance of the latter is highly dependent to the observation vector, and typically worse than those shown in Section 2. The reason for choosing this particular observation vector was to obtain a fair comparison between transform based methods such as DCT and MEKF.

4.5 2-D Maximum Entropy Kalman Filter (MEKF) for estimation

A similar algorithm to that of the Section 3 has been implemented. As with the implementation of Section 3, an alternative reconstruction method should be considered when the segment of interest is not best represented by a ME model. Since most observation degradations, C, are low pass filters, we have found that reconstruction using ZOH to be satisfactory. The pseudo-code of this algorithm is given below.

1. Initialization
   - Set N, number of states
   - Segment the signal
2. Iteration for all segments
   - Obtain row and column ordered vectors based on the current segment
   - Obtain for each vectors above, the a^R ,a^C , Q^R and Q^C using Levinson-Durbin algorithm
   - If $Q^R Q^C$ is less than a predetermine threshold, perform ZOH to recover the segment otherwise,
   - Calculate S as given in Eq. (4.47) or Eq. (4.52), where K is equal to the Kalman gain of the previous segment.
   - Calculate A based on Eq. (4.44), where K is equal to the Kalman gain of the previous segment.

   2.1 Iterate in the current segment
   - Apply standard Kalman filter with K_{k+1} y_{k+1} whitened by S and A as the transitional matrix.
   - If the constraints are not satisfied use Eq. (4.41)

- Update A as described by Eq. (4.44), where the current K is used for the update.

4.5.1 MEKF reconstruction for 2-D signals

A similar set of images to those of the Section 3 has been used with the observation matrix C=[*0.0625...0.0625*], analysis-synthesis ratio 1:16 and threshold of *0.05*.

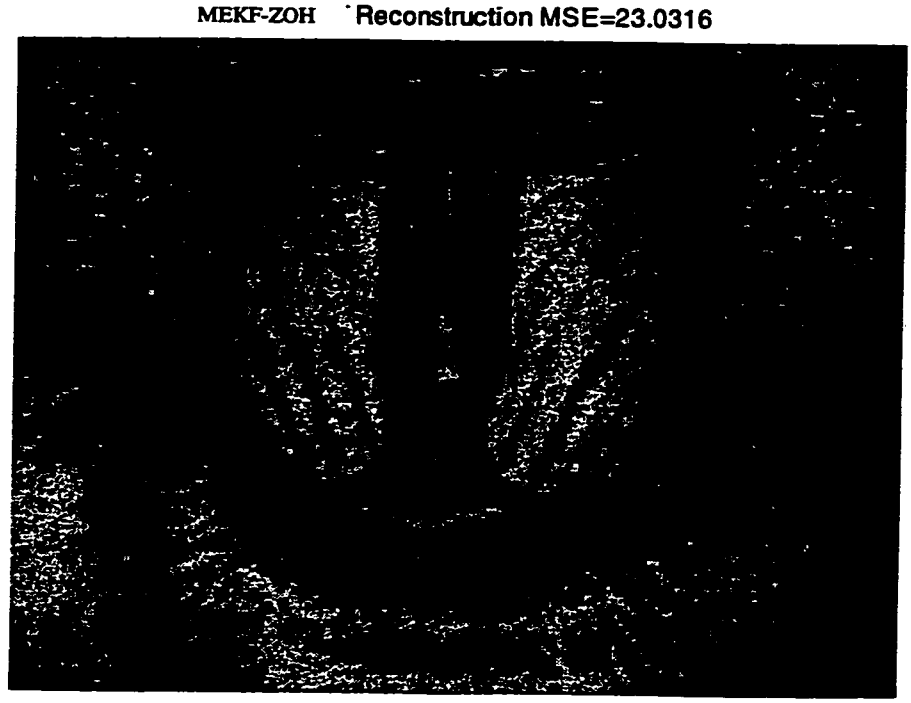


Figure 4.33 Baboon MEKF-ZOH Reconstruction

With respect to Figure 4.33, Figure 4.34 and Figure 4.35 once again, as with analysis-synthesis, the best results are obtained when an image is mainly made up segments of high entropy. It is also noted that due to restrictions on the observation matrix C, the results obtained in estimation are not as good as those obtained in Section 3. Results are summarized in Table 4.7.

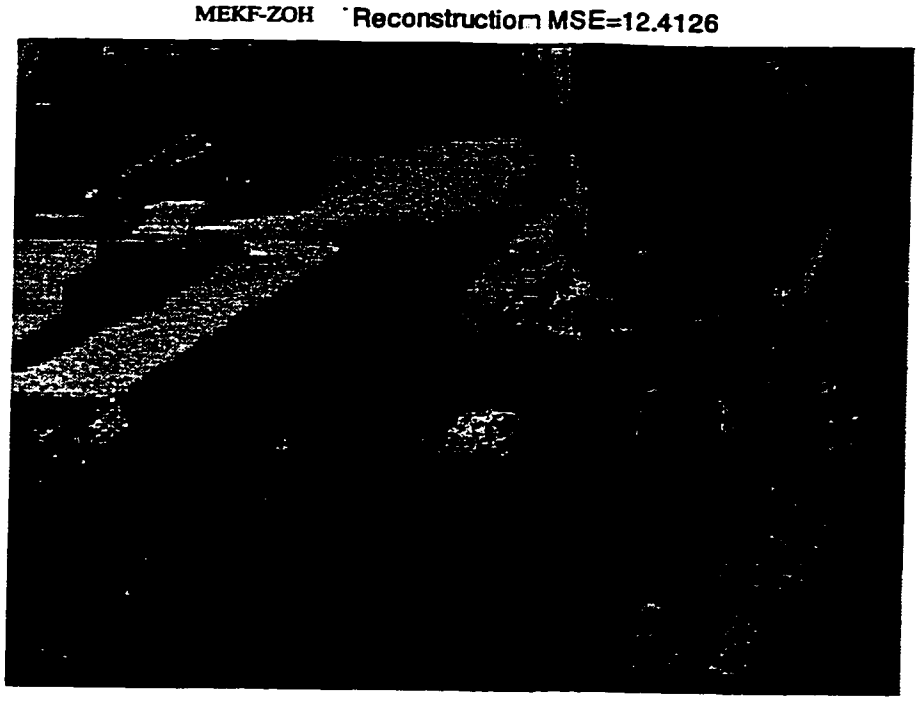


Figure 4.34 Barbara MEKF-ZOH Reconstruction

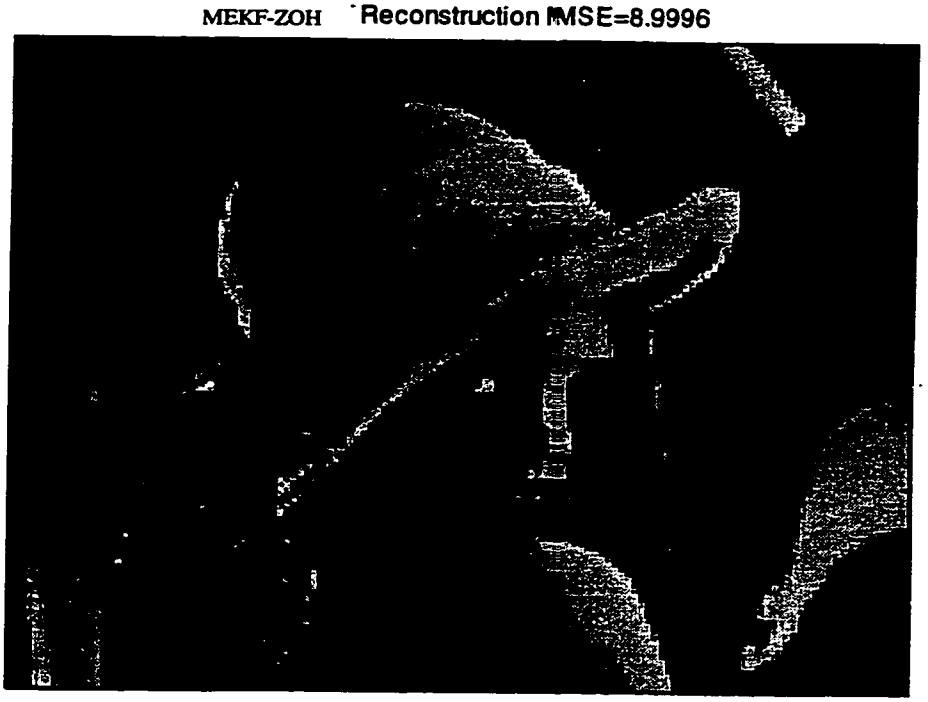


Figure 4.35 Lenna MEKF-ZOH Reconstruction

Image \ Method	MEKF-ZOH
Baboon	23.03
Barbara	12.41
Lenna	8.999

Table 4.7 Reconstruction MSE of MEKF for test images

4.5.2 Summary and future work

We have demonstrated the suitability of the ME criterion in signal modeling, compression and reconstruction. Furthermore, methods using the classical Kalman filter have been developed to this end. A simple application of MEKF as a method for modeling, compression and reconstruction of degraded signals, both 1-D and 2-D, was demonstrated. Future work will include the smoothing operation (using both past and future information) in the Kalman filter and non-separable expansion to achieve a better signal reproduction. Although, at first sight Linear Predictive Coding (LPC) and (MEKF) may seem similar however, it should be noted that unlike LPC, MEKF assures the maximum entropy dynamics on the reconstructed signal and not on the assumed dynamics of the plant process. The advantage of the latter has been shown in simulations of Section 4.4.2.1 and Section 4.4.2.2.

Chapter 5 Summary and future work

In this thesis, we have described and proposed solutions to various aspects of the problem of system identification. In particular, the problems of local minima in IIR adaptive signal processing and of modeling based on partially observed data have been addressed. With respect to the former, we observed that global convergence to a correct optimum for insufficiently modeled system in presence of observation noise and colored input may be achieved. Furthermore, we were able to show the uniqueness of the global minimum for a single-pole approximation. Although our simulations have shown the same to be true for multi-pole systems however, the main task of future work is to *derive* the same result for multi-pole systems.

We have shown the deficiencies of Combined Output methods [25], particularly that of the biased coefficients and local minima, and how MEEOE resolves them. It was noted that our method, *i.e.* Modified Equation Error Output Error (MEEOE), belongs to the family of Combined Error algorithms. As mentioned in Chapter 3, if we were to use a single weighting factor, allow for the deterministic update of $\boldsymbol{W}$ and employ Pseudo Linear Regression (PLR) method for the update of $\boldsymbol{A}$ and $\boldsymbol{B}$, MEEOE would be identical to the methods represented by [33]. However, we have shown that more flexibility and faster convergence may be obtained by multiple weighting factors. Furthermore, allowing for non-deterministic update of $\boldsymbol{W}$ allows for faster convergence, independent of dynamics of the process. Lastly, unlike PLR, Simplified Recursive Predictive Error (SRPE), our method of update, has lead to convergence to an optimum point on MSE surface, in all our simulations.

For the problem of modeling, based on partial observation, we have demonstrated the suitability of the maximum entropy criterion in signal modeling, compression and reconstruction of 1-D and 2-D signals. We observed its advantages as compared to MSE criterion, *when a complete set of data is unavailable.* Furthermore, methods using the classical Kalman filter have been developed to this end. Future work may include the smoothing operation (using both past and future information) in the Kalman filter and non-separable expansion to achieve a better signal reproduction, in image processing applications.

Although, it may be tempting to draw direct analogy between Linear Predictive Coding (LPC) and our method, Maximum Entropy Kalman Filter (MEKF) however, it should be noted that unlike LPC, MEKF assures the maximum entropy dynamics on the reconstructed signal and not on the assumed dynamics of the plant process. The advantage of the latter has been shown in our simulations.

We have attempted to resolve some of the outstanding problems in adaptive signal processing that are based on methods well understood within this field. As with many other such attempts, the theoretical coherence is inversely proportional to empirical diversity. In fact, as seen in Chapter 3, closed-form description of the evolution of the error surface, except for a single-pole system, becomes an intractable task. It is therefore rewarding to note that the application of the principle of maximum entropy, in image processing, does indeed result in what was expected.

However, it is our opinion that the intrinsic tradeoff between practical utility and mathematical rigor is inherent in application of linear methods to non-linear problems [1]. Therefore, the main contribution of the future works will be on application of well-

established non-linear methods to the traditional adaptive signal processing problems. It should be noted that a vast majority of current problems can be resolved by application of simple non-linear methods that are easily derived from the linear ones [9].

References

[1] Henry D. I. Adarbanel, Reggie Brown, John J. Sidorowich and Lev Sh. Tsimring, "The Analysis of Observed Chaotic Data in Physical Systems," *Review of Modern Physics*, Vol. 65, No. 4, pp. 1331-1392, Oct. 1993.

[2] Karl J. Astrom and Torsten Soderstrom, "Uniqueness of the Maximum Likelihood Estimates of the Parameters of an ARMA Model," *IEEE Transactions on Automatic Control*, Vol. AC-19, No. 6, pp. 769-773, Dec.-1974.

[3] C S. Berger,"Linear splines with adaptive mesh sizes for modelling nonlinear dynamic systems," *IEE Proceedings: Control Theory & Applications*, Vol. 141, No. 5, pp. 277-284, Sept 1994.

[4] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[5] Richard E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley Publishing Company, New York, 1987.

[6] R. Grover Brown and Patrick Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley & Sons, New York, 1997.

[7] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[8] C.F.N. Cowan and P. M. Grant, *Adaptive Filters*, Prentice-Hall, New Jersey, 1985.

[9] Afshin David and Tyseer Aboulnasr, "Adaptive Non-linear Modeling," *1998 IEEE Symposium on Advances in Digital Filtering and Signal Processing*, pp. 126-130, June 1998.

[10] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Series B, Vol. 39, pp. 1-38, 1977.

[11] Paulo S. R. Diniz, *Adaptive Filtering Algorithms and Practical Implementation*, Kluwer Academic Publishers, Boston, 1997.

[12] H. Fan and W. K. Jenkins, "A New Adaptive IIR Filter," *IEEE Trans. Circuit and Systems*, Vol. CAS-33, No. 10, pp. 939-947, Oct. 1986.

[13] Arthur Gelb, Joseph F. Kasper, Jr., Raymond A. Nash, Jr., Charles F. Price and Arthur A. Sutherland, Jr., *Applied Optimal Estimation*, THE M.I.T. PRESS Massachusetts, 1974.

[14] Graham C. Goodwin and Kwai Sang Sin, *Adaptive Filtering Prediction and Control*, Prentice Hall Inc., NJ, 1984.

[15] Image Recovery: *Theory and Application Ed. Henry Stark*, Chapter 3, Kenneth M. Hanson, "Bayesian and Related Methods in Image Reconstruction from Incomplete Data," Academic Press Inc. New York, 1987.

[16] Simon Haykin, *Adaptive Filter Theory*, Second Ed., Prentice Hall Inc., NJ, 1991.

[17] Simon Haykin, *Neural Networks A Comprehensive Foundation*, Macmillan College Publishing Company, New York, 1994.

[18] Masaru Hoshiya and Osamu Maruyama, "Identification of Running Load and Beam System," *Journal of Engineering Mechanics-ASCE*, Vol. 113, No. 6, pp. 813-824, 1987.

[19] Anil K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall Inc., NJ, 1989.

[20] W. Jakoby and M Pandit, "A Prediction-error-method for Recursive Identification of Nonlinear Systems," Automatica, Vol. 23, No. 4, pp. 491-496, 1987.

[21] E.T. Jaynes, "On the rational of maximum entropy methods," *Proc. IEEE*, Vol. 70, pp. 939-953, Sept. 1982.

[22] C. Richard Johnson and M. G. Larimore, "Comments on and additions to 'An adaptive recursive LMS filter,'," *Proc. IEEE*, Vol. 65, No. 9, pp. 1399-1402, Sep. 1977.

[23] C. Richard Johnson, "Adaptive IIR Filtering: Current Results and Open Issues," *IEEE Trans. on Info. Theory*, Vol. IT-30, No. 2, pp. 237-250, March 1984.

[24] Guy Jumarie, "Entropy of Markovian Processes: Application to Image Entropy in Computer Vision," Elsevier Science Ltd., Great Britain, Vol. 335B, No. 7, pp. 1327-1338, 1998.

[25] John B. Kenny and Charles E. Rohrs, "The Composite Regressor Algorithm for IIR Adaptive Systems," *IEEE Trans. on Signal Processing*, Vol. 41, No. 2, pp. 617-628, Feb. 1993.

[26] Hyoung-Nam Kim and Woo-Jin Song, "Adaptive IIR Filtering with Monic Normalization: Reduced-Order Approximation," *IEEE Trans. On Signal Processing Letters*, Vol. 7, No. 3, pp. 54-56, March-2000

[27] S. Kock and H. Kaufman, "Image Restoration Using Extended Kalman Filters," *IFAC Adaptive Systems in Control and Signal Processing*, Published by Pergamon Press Inc. NY, USA, pp. 477-480, 1993.

[28] Shlomo Kock and Howard Kaufman, "Restoration of spatially varying images using multiple model extended Kalman filters," *Proceedings of the IEEE Conference on*

Decision and Control. Published by IEEE, IEEE Service Center, Piscataway, NJ, USA, 93CH3307-6. Vol. 2, pp. 1216-1221, 1993.

[29] Frank L. Lewis, *Optimal Estimation With an Introduction to Stochastic Control Theory*, John Wiley & Sons, New York, 1986.

[30] Geoffery J. McLachlan and Thriyambakam Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, Inc., New York, 1997.

[31] Todd K. Moon, "The Expectation-Maximization Algorithm," *IEEE Signal Processing Magazine*, pp. 47-60, 1996.

[32] F G B De. Natale, G S. Desoli and D D. Giusto, "Adaptive least-squares bilinear interpolation (ALSBI): a new approach to image-data compression," *Electronics Letters*, Vol. 29, No. 18, pp. 1638-1640, Sep 2 1993.

[33] Sergio L. Netto and Paulo S. R. Diniz, "Composite Algorithms for Adaptive IIR Filtering," *Electronics Letters*, Vol. 28, No. 9, pp. 886-888, April. 1992.

[34] Sergio L. Netto, Paulo S. R. Diniz and Panajotis Agathoklis, "Adaptive IIR Algorithms for System Identification: A General Framework," *IEEE Trans. on Education*, Vol. 38 No. 1, pp. 54-66, Feb. 1995.

[35] Alan V. Oppenheim and Ronald W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[36] Rallis C. Papademetriou, "Data Processing Using Information Theory Functionals," Kybernetes, MCB University Press, Vol. 27, No. 3, pp. 264-272, 1998.

[37] Min-Soo Park and Woo-Jin Song, "Adaptive IIR Filtering with Combined Repressor and Combined Error," *Signal Processing*, Vol. 56, pp. 191-197, 1997.

[38] Dimitris Nicolas Politis, "Moving Average Processes and Maximum Entropy," *IEEE Trans. on Info. Theory*, Vol. 38, No. 3, pp. 1174-1177, May 1992.

[39] L.R. Rabbiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall Inc., New York, 1978.

[40] Jan Radecki, Januss Konrad and Eric Dubois, "Design of Finite Word Length IIR Filters With Prescribed Magnitude, Group Delay and Stability Properties Using Simulated Annealing," Proceedings - *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 1637-1640, May 1991.

[41] G. Ramponi and G. L. Sicuranza, "Decision-Directed Nonlinear Filter for Image Processing," *Electronics Letters*, Vol. 23, No. 23, pp. 1218-1219, 1987.

[42] Philip A. Regalia, "An Unbiased Equation Error Identifier and Reduced-Order Approximations," *IEEE Trans. on Signal Processing*, Vol. 42, No. 6, June 1994.

[43] Azriel Rosenfeld and Avinash C. Kak, Digital Picture Processing, 2nd Ed., Vol. 1, Academic Press Inc. New York, 1982.

[44] R.D. Rosenkrantz Editor, E.T. Jaynes: *Papers on Probability, Statistics and Statistical Physics*, D. Reidel Publishing Company, Boston, 1983.

[45] Sundar G. Sankaran and A. A. Louis Beex, "Hypersphrical Parameterization for Unit-Norm Based Adaptive IIR Filtering," *IEEE Signal Processing Letters*, Vol. 6, No. 12 Dec. 1999.

[46] Khalid Sayood, *Introduction to Data Compression*, Morgan Kaufmann Publishers Inc., 1996.

[47] John J. Shynk, "Adaptive IIR Filtering," *IEEE ASSP Magazine*, pp. 4-20, April-1989.

[48] Gyula Simon and Gabor Peceli, "A New Composite Gradient Algorithm to Achieve Global Convergence," *IEEE Trans. on CAS-II*, Vol. 42, No. 10, pp. 681-684, 1995.

[49] Victor Solo and Xuan Kong, *Adaptive Signal Processing Algorithms Stability and Performance*, Prentice-Hall, New Jersey, 1995.

[50] David A. Sprecher, "On the Structure of Continuos Functions of Several Variables," *Trans. Math Soc.*,Vol. 115, pp. 340-355, March 1965.

[51] Henry Stark and John W. Wood, *Probability, Random Processes*, and Estimation Theory for Engineers, 2nd Ed., Prentice Hall Inc., New York, 1994.

[52] Steven H. Strogatz, *Nonlinear Dynamics and Chaos with Applications to Physics, Biology, Chemistry, and Engineering*, Addison-Wesley Publishing Company, New York, 1994.

[53] Jeffrey D. Taft and N. K. Bose, "Quadratic-Linear Filters for Signal Detection," *IEEE Transactions on Signal Processing*, Vol. 39, No. 11, pp. 2557-2559, 1991.

[54] H. J. Trussell, "The Relationship Between Image Restoration by the Maximum A Posteriori Method and a Maximum Entropy Method," *IEEE Transactions on Acoustics, Speech , and Signal Processing*, Vol. ASSP-28, No. 1, pp. 114-117, Feb. 1980.

[55] Garret N. Vanderplaasts, *Numerical Optimization Techniques for Engineering Design With Applications*, McGraw-Hill Book Company, New York, 1984.

[56] Wayne L. Winston, *Introduction to Mathematical Programming*, Applications and Algorithms, Duxbury Press, Belmont California, 1995.

[57] Image Recovery: *Theory and Application Ed. Henry Stark*, Chapter 5, Xinhua Zhuang, Einar Ostevold, and Robert M. Haralick, "The Principle of Maximum Entropy in Image Recovery," Academic Press Inc. New York, 1987