

Analysis of Dynamical Behaviour of 2-D Adaptive Digital Filters

Maha Shadaydeh

A dissertation submitted in conformity with
the requirements for the degree of
Doctor in Electrical Communications Engineering

Department of Electrical Communications Engineering
Graduate School of Engineering
Tohoku University, JAPAN
January 1999

To my mother and in the memory of my father

Contents

Symbols	10
Acronyms	12
1 Introduction	13
1.1 Background and Motivations	13
1.2 Tour Map of the Thesis	17
2 Fundamental Study of 2-D Adaptive Digital Filters	20
2.1 Introduction	20
2.2 2-D Wiener Filters	21
2.3 2-D LMS Adaptive FIR Filters	23
2.4 2-D Adaptive IIR Filters	25
2.4.1 Output Error Formulation	26
2.4.2 Equation Error Formulation	27
2.5 2-D Adaptive Filtering Applications	29
2.5.1 2-D System Identification Configuration	29
2.5.2 2-D Adaptive Differential Pulse Code Modulation	30
2.5.3 2-D Adaptive Noise Cancellation	31

2.6	Performance Measures in 2-D Adaptive Filters	32
3	Bias Removal Algorithm for 2-D Equation Error Adaptive IIR	
	Filters	34
3.1	Introduction	34
3.2	2-D Backpropagation Adaptive Cascade IIR Filters	37
3.3	2-D Bias Removal Algorithm (2DBRA)	40
3.4	Convergence Analysis of the 2DBRA	43
	3.4.1 Convergence Analysis of the First Part of the Structure . .	45
	3.4.2 Convergence Analysis of the Second Part of the Structure .	50
3.5	Experimental Results and Discussion	53
3.6	Summary	66
4	Steady State Analysis of 2-D Doubly Indexed LMS Adaptive	
	Filters	67
4.1	Introduction	67
4.2	The 2-D Doubly Indexed LMS Algorithm	70
4.3	Steady State Analysis using the Independence Assumption	72
	4.3.1 MSE Calculation	73
	4.3.2 Weight-Error Covariance Matrix	74
4.4	Steady State MSE Analysis with White Gaussian Input Data . . .	82
4.5	Direct Averaging Method	85
4.6	Experimental Results and Discussion	88
4.7	Summary	91

5	2-D Doubly-Indexed Block LMS Adaptive Filters	96
5.1	Introduction	96
5.2	2-D Doubly-Indexed Block LMS Algorithm (2DDI-BLMS)	97
5.3	Convergence Analysis of the 2DDI-BLMS	100
5.3.1	Convergence of the Mean	101
5.3.2	Convergence of the BMSE	102
5.4	Steady State MSE Analysis with White Gaussian Input Data	106
5.4.1	Bounds on the Step Size Parameter μ	107
5.4.2	Adaptation Accuracy	108
5.5	Experimental Results and Discussion	109
5.6	Summary	111
6	Conclusions and Suggestions for Future Work	113
6.1	Conclusions	113
6.2	Suggestions for Future Work	115
A	2-D Signals and Systems Review	117
A.1	2-D Signals	117
A.2	2-D Digital Systems	120
A.2.1	Linear and Shift-Invariant Systems	120
A.2.2	Causal Systems	120
A.3	Difference Equation Representation of 2-D Systems	121
A.4	Transfer Function Representation of 2-D Systems	122
A.5	Stability of 2-D Systems	123

<i>CONTENTS</i>	5
A.6 2-D Separable Denominator Digital Filters	125
B Probability Review	126
B.1 Random Variables	126
B.2 Random Processes	128
Index	143

List of Figures

2.1	2-D Wiener filter.	22
2.2	2-D adaptive filter.	24
2.3	Indexing Schemes- a) rectangular indexing with short scan lines, b) rectangular indexing with image-length scan lines, c) diagonal indexing, and d) block diagonal indexing.	24
2.4	Output error formulation.	27
2.5	Equation error formulation.	28
2.6	2-D system identification configuration.	30
2.7	2-D ADPCM, encoder and decoder with quantizer Q.	31
2.8	2-D adaptive noise canceler.	32
3.1	1-D backpropagation adaptive cascade IIR filters.	35
3.2	Equation error formulation for separable denominator 2-D IIR filters.	36
3.3	Equation error formulation for separable denominator 2-D IIR fil- ters with the bias removal method.	41
3.4	The convergence of the mean squared output error of the bias removal algorithm for <i>Example 1</i> (average of 30 runs), $\mu_1 = \mu_2 =$ 0.0004 , $\alpha_1 = \alpha_2 = 0.5$	55

3.5	Plot of the scaling factor $\hat{\tau}_1(i)$	56
3.6	Plot of the scaling factor $\hat{\tau}_2(i)$	56
3.7	Original image “Lena”.	59
3.8	Noisy image “Lena”.	59
3.9	The processed image at the 41st pass using algorithm [15], $\mu_1 =$ $\mu_2 = 6 \times 10^{-9}$	60
3.10	The processed image at the 41st pass using the 2DBRA, $\mu_1 = \mu_2 =$ 6×10^{-9} , $\alpha_1 = \alpha_2 = 0.5$	60
3.11	The improvement of the SNR of the enhanced image through con- secutive passes in <i>Example 2</i>	63
3.12	The original image “Mandrill”.	64
3.13	The noisy image “Mandrill”.	64
3.14	The enhanced image at the 10th pass using algorithm [15], $\mu_1 =$ $\mu_2 = 7.6894 \times 10^{-8}$	65
3.15	The enhanced image at the 10th pass using the 2DBRA, $\mu_1 = \mu_2 =$ 7.6894×10^{-8} , $\alpha_1 = \alpha_2 = 0.7$	65
4.1	2-D adaptive filter.	71
4.2	Weight-error covariance coefficient γ_0 as a function of $\zeta = \mu\sigma_x^2$ for 2 by 2 adaptive FIR filter.	83
4.3	The root value ζ_∞ versus N	84

4.4	Comparison of the experimental results with the theoretical values for the misadjustment of the 2-D doubly indexed LMS in the simplified setting ($\lambda_0 = \lambda_1, \dots, \lambda_{N^2-1} = \sigma_x^2$, $f_h = f_v = 0.5$, and $\mu_h = \mu_v = \mu$), white Gaussian input.	93
4.5	Simulation results for correlated Gaussian input, (a) $\alpha = 0.2$, (b) $\alpha = 0.4$	94
4.6	Comparison of the experimental results with the theoretical values for the misadjustment of the 2-D doubly indexed LMS in the simplified setting ($\lambda_0 = \lambda_1, \dots, \lambda_{N^2-1} = \sigma_x^2$, $f_h = f_v = 0.5$, and $\mu_h = \mu_v = \mu$), i.i.d. binary input.	95
5.1	2-D adaptive FIR filter.	97
5.2	2-D block-wise processing scheme.	99
5.3	The value of the root ζ_∞ vs. filter order N	110
5.4	Learning curves for the 2DDI-BLMS and 2-D LMS (averages of 30 runs), step size $\mu = 0.05$	111
A.1	2-D discrete-space signals $x(m, n)$	118
A.2	Digital image of 256×256 pixels quantized at 8 bits/pixel.	119
A.3	Output of a 2-D IIR digital filter.	122
A.4	2-D FIR digital filter.	123
A.5	Cascade implementation of separable denominator IIR filter.	125

List of Tables

3.1	Parameter estimates for <i>Example 1</i> (30 runs).	54
3.2	Parameter estimates at the 41st pass, $\mu_1 = \mu_2 = 6 \times 10^{-9}$, $\alpha_1 =$ $\alpha_2 = 0.5$	58
3.3	The variance of the noise left in the processed images at different passes.	61
3.4	Parameter estimates for <i>Experiment 3</i> at the 10th pass.	62
4.1	Simulation results for Example 2, correlated Gaussian input . . .	89
4.2	Simulation results for Example 2, maximum step size value μ_{max} , Gaussian input, N=2.	90
5.1	Summary of experimental results	109

List of Symbols

Symbol	Definition
y	2-D output signal
x, u	2-D input signals
d	2-D desired signal
e	2-D error signal
e_o	2-D output error
e_e	2-D equation error
v	2-D measurement noise
τ_1, τ_2	scaling factors
μ_1, μ_2	step size parameters
μ_h, μ_v	step size parameters in the horizontal and vertical directions
d_1, d_2, y_1, y_2	intermediate signals in the cascade IIR filter
φ_i	regressor vector for part i of the cascade IIR filter
θ_i	true parameter vector for part i of the cascade IIR filter
$\hat{\theta}_i$	estimated parameter vector for part i of the adaptive IIR filter
$\tilde{\theta}_i$	$= \theta_i - \hat{\theta}_i$
E	statistical expectation operator
$N \times N$	order of 2-D filter
h_j	the impulse response of 1-D indexing scheme-based adaptive algorithm at iteration j
q_1^{-1}	spatial delay operator in the horizontal direction
q_2^{-1}	spatial delay operator in the vertical direction
$A(q_1^{-1}, q_2^{-1})$	transversal section for IIR filter
$B_1(q_1^{-1})$	denominator horizontal section of IIR filter
$B_2(q_2^{-1})$	denominator vertical section of IIR filter
$h_{m,n}$	the impulse response of 2-D indexing scheme-based adaptive algorithm at spatial indices (m, n)
\mathbf{H}_{opt}	optimum (Wiener) parameter vector for FIR filter
$\mathbf{X}_{m,n}$	input data vector at spatial indices (m, n)

$\mathbf{H}_{m,n}$	estimated parameter vector for FIR filter at spatial indices (m, n)
$\mathbf{G}_{m,n}$	the instantaneous gradient of the mean square error at spatial indices (m, n)
\mathbf{R}	input correlation matrix
λ_j	eigenvalues of the input correlation matrix \mathbf{R}
ε	measurement noise
$\mathbf{C}_{m,n}$	$= \mathbf{H}_{m,n} - \mathbf{H}_{opt}$
$K_{m_1, n_1; m_2, n_2}$	weight-error correlation matrix
$\mathbf{\Gamma}_{m_1, n_1; m_2, n_2}$	transformed weight-error correlation matrix
$\gamma_{m_1, n_1; m_2, n_2}^{i,j}$	the element of the matrix $\mathbf{\Gamma}_{m_1, n_1; m_2, n_2}$ at the i th row and j th column
\mathbf{Q}	orthogonal matrix
Λ	diagonal form of the input correlation matrix \mathbf{R}
σ_x^2	variance of the input signal x
σ_ε^2	variance of the noise ε
$Y_{i,j}$	$K^2 \times 1$ vector that consists of the element of the (i, j) th block of the output signal $y(m, n)$.
$D_{i,j}$	$K^2 \times 1$ vector that consists of the element of the (i, j) th block of the desired signal $d(m, n)$.
$\mathbf{H}_{i,j}$	the adaptive filter's weight-vector at the (i, j) th block
$E_{i,j}$	$D_{i,j} - Y_{i,j}$

List of Acronyms

Acronym	Definition
2DDI-BLMS	Two Dimensional Doubly-Indexed Block LMS
2-D ADPCM	2-D Adaptive Differential Pulse Code Modulation
2DBRA	Two Dimensional Bias Removal Algorithm
BMSE	Block Mean Square Error
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
F-M	Fornasini and Marchesini
IIR	Infinite Impulse Response
I/O	Input/Output
LMS	Least Mean Square
MSE	Mean Square Error
SNR	Signal to Noise Ratio
TDLMS	Two Dimensional LMS
WECM	Weight Error Correlation Matrix

Chapter 1

Introduction

This chapter states the motivations of this research work. A tour map of the thesis is also given.

1.1 Background and Motivations

With the extraordinary advances in the fields of digital computing, digital signal processing, and high speed integrated circuits, two dimensional digital filters have exhibited an impressive growth in the past three decades in terms of both theoretical development and applications [1]-[4]. However, it was only till the late 80th, and due to the increasing demand for real time image and video signal processing in telecommunications and multimedia technology, when researchers efforts started to be directed towards the development of multidimensional adaptive filters.

For image enhancement and restoration problems, the minimum mean square error (Wiener) non-adaptive filter, the classical stationary solution to these problems, is not adequate due to its low-pass characteristic, which gives rise to un-

acceptable blurring of line and edges in the images. To overcome such problem, some attempts have adopted non-stationary approaches in the filter design [6], [7], however, in many real time filtering applications, the spatial domain design of 2-D filter with an off-line method is not practical owing to the large number of filter parameters and the large size of data required in evaluating these parameters.

As an alternative approach, in the last decade, 2-D adaptive filters have received considerable research interest due to their ability to take into account the inherent nonstationary statistical properties of images [11]. A 2-D adaptive filter is a space varying filter relies for its operation on a recursive algorithm which is responsible for updating the filter's weights so that the filter can perform satisfactorily in an unknown environment and be able to track any spatial variation in the image statistics. For example, in the background region the filter's weights are adjusted to narrow the pass band of the filter to remove as much of the noise as possible. Within edge regions the filter's weights are adjusted to widen the filter's pass band and maintain the sharpness of the edges.

So far, several 1-D adaptive filtering techniques have been extended to two dimensional adaptive filtering applications, such as adaptive differential pulse code modulation, image restoration and wide band noise suppression [11]. However, such extension is not always straightforward and faced with some vital problems. (1) 2-D real time filtering applications, such as real time image and video signal processing, involve more data; thus, it is always desirable to improve the convergence speed and reduce the computational complexity of 2-D adaptive filtering

algorithms. (2) The mathematics for understanding the properties and characteristic of 2-D adaptive systems are less complete. As we know, any adaptive algorithm is a form of closed loop feedback system and there is always a potential for instability and divergence of the adaptive system. The issues of dynamical behaviour analysis of the adaptive system, stabilization, stability monitoring and testing are very difficult even in the 1-D case. For two dimensional adaptive filtering such issues are by far more difficult. (3) Images are highly nonstationary signals; hence high tracking ability in nonstationary environment becomes an essential goal in the development of 2-D adaptive filtering algorithms. (4) Images are highly correlated signals; unlike the 1-D case, the correlation of the image pixels exists in all directions. Accordingly the scanning scheme used to process the 2-D data has significant effect on the performance of the adaptive algorithms.

Considering these relevant issues, the purpose of this research work is to add few contributions on the line of the development of 2-D adaptive filtering algorithms in terms of both algorithm development and dynamical behaviour analysis. This thesis contains three main contributions to the area of 2-D adaptive filtering, and thus can be divided into three main parts summarized as follows.

The first part of this thesis focuses on the development of a bias removal algorithm for the 2-D equation error-based adaptive cascade IIR filters with separable denominator function [15]. As well known, equation error-based adaptive IIR filtering algorithms have the advantages of fast convergence and unimodal mean-square-error (MSE) surface. These advantages, however, come along with the drawback of biased parameter estimates in the presence of measurement noise.

The adaptive filter structure in the proposed algorithm is based on the concept of backpropagating the desired signal through a cascade of the denominator vertical and horizontal sections. To handle the bias problem, the proposed algorithm uses a scaled value of the output error of each of the cascaded sections as an estimate for the measurement noise embedded in the signal part of the coefficient-update procedure of that section. Thus, while maintaining the advantages of easy stability monitoring, fast convergence, and low computational load, the effect of the measurement noise is suppressed.

The second part of the thesis concerns the analysis of the convergence behaviour of the 2-D LMS adaptive FIR filtering algorithm [19] in which the filter's weights are updated along both the vertical and horizontal directions as a doubly-indexed dynamical system (so called Fornasini and Marchesini (F-M) state space model). In what follows we will refer to this algorithm as the 2-D doubly-indexed LMS. Updating the filter's weights in both directions enables efficient use of the 2-D correlation information of the image pixels in both vertical and horizontal directions and hence, provides better performance in nonstationary environments [19]. However, from convergence analysis aspect, this truly 2-D nature of the weights' update results in a new problem that is not encountered in the 1-D indexing scheme based adaptive methods. The MSE analysis for 1-D LMS, as well as 1-D indexing scheme based 2-D LMS methods, reduces mainly to the stability analysis of a set of first order coupled difference equations in the coefficients of the Weight Error Correlation Matrix (WECM). This set of difference equations maintains stability under a general condition imposed on the used step size value

[26], [27]. For the 2-D doubly-indexed LMS, as will be shown in the thesis, MSE analysis calls for stability analysis of a set of second-order coupled 2-D difference equations in the coefficients of the WECM, which is very difficult to handle mathematically. The second part of this thesis is devoted to solving this problem and to deriving the upper bounds on the step size parameters that insure the convergence of the 2-D doubly indexed LMS in the MSE sense.

The third part of this thesis focuses on the development of a new 2-D adaptive LMS FIR filtering algorithm by block-wise processing of data in order to gain the benefits of parallel computation and improved stability performance associated with block filtering scheme. In the proposed algorithm, the input signal is partitioned into non-overlapping blocks; the weights are then adjusted once per each block of the input signal. The filter weights update process is carried out along both the vertical and horizontal directions as a doubly-indexed dynamical system in accordance with the 2-D doubly indexed LMS. In addition to the improving the tracking ability in nonstationary environments, the 2-D doubly-indexed block LMS (2DDI-BLMS) is shown to be very suitable for parallel processing.

1.2 Tour Map of the Thesis

This thesis is organized as follows.

Chapter 1 explains the motivations and purposes of this research work. It also gives the thesis outline.

Chapter 2 first discusses the need for 2-D adaptive filters. Next, it presents principles of 2-D adaptive FIR filters, principles of 2-D adaptive IIR filters, and

2-D adaptive filtering applications. Then, performance measures in adaptive filtering are reviewed. This chapter furnishes the reader with the necessary background and information on the state-of-the-art.

Chapter 3 proposes a 2-D bias removal algorithm for 2-D adaptive cascade IIR filters with separable denominator transfer function. First, a 2-D backpropagation IIR filter structure with the filter denominator decomposed into cascade of vertical and horizontal sections is introduced. Next, the 2-D bias removal algorithm (2DBRA) is proposed. The Input-output (I/O) stability analysis for the 2DBRA is also discussed. The proposed algorithm is then applied to 2-D system identification and image enhancement experiments where the effectiveness and the superiority of the 2DBRA over other proposed methods are illustrated by numerical results.

Chapter 4 first reviews the 2-D doubly indexed LMS. Next, it addresses the problem that is encountered in evaluating the filter's weight-error covariance matrix (WECM) for doubly indexed LMS algorithm. The *independence assumption theory* is then reviewed and a method for the calculation of steady state value of the WECM is proposed using the *independence assumption*. The special case when the input signal is white Gaussian is further discussed, and the condition required to ensure the convergence in the MSE sense is derived. Finally, numerical experimental results that support the validity of the proposed analysis are presented.

Chapter 5 first develops a 2-D doubly-indexed block LMS (2DDI-BLMS) algorithm. The convergence behaviour of the 2DDI-BLMS is then analysed and the

upper bounds on the step size parameters that guarantee the convergence of the adaptive algorithm in the mean and the variance are also derived. Experimental results that support the validity of the obtained convergence analysis results are presented. The advantages of the proposed method over the other methods proposed in the literature are also discussed.

Chapter 6 concludes this thesis. The main contributions of this thesis to the area of 2-D adaptive filters are summarized. Suggestions for future work are also introduced.

Chapter 2

Fundamental Study of 2-D Adaptive Digital Filters

2.1 Introduction

The estimation of images is a fundamental problems which lies at the heart of two related areas of image processing: enhancement and data compressions. The enhancement problem is essentially one of the optimal filtering with respect to some error criterion. The most widely used criterion is the mean square error (MSE). The classical stationary solution to the problem, Wiener filter, has been used with limited success, because its design is based on the assumption that both the relevant signal and the noise are stationary signals and does not take into account the high frequency components in the image. This stationary assumption results in low pass characteristic that smears the edges of the images.

2-D Adaptive filter offers promising alternative solution due to its ability to change the filtering characteristic while scanning the image to match the image generating mechanism and track spatial variations of the image statistics.

This chapter reviews some of the fundamental principles and applications of

2-D adaptive filters.

2.2 2-D Wiener Filters

The basic structure of 2-D Wiener filter is shown in Fig. 2.1. In this figure, the desired image $d(m, n)$ and the input image $x(m, n)$ are of dimension M by M . The filter used is an N by N causal FIR filter. The filter weights are convolved with the input image $x(m, n)$ to produce the output image $y(m, n)$ as follows:

$$y(m, n) = \sum_{l=0}^{N-1} \sum_{k=0}^{N-1} h_j(l, k)x(m-l, n-k) \quad (2.1)$$

where j is some function of (m, n) specifying the indexing scheme used to process the 2-D data. At iteration j , the error signal $e(m, n)$ is given by

$$e(m, n) = d(m, n) - \sum_{l=0}^{N-1} \sum_{k=0}^{N-1} h_j(l, k)x(m-l, n-k) \quad (2.2)$$

The aim of Wiener filter is to obtain a set of weights such that the output error $e(m, n)$ given in Eq. (2.2) is minimized in the MSE sense. The MSE is defined as

$$\text{MSE} = E\{e^2(m, n)\} \quad (2.3)$$

where $E\{.\}$ denotes the mathematical expectation operator. Substituting Eq. (2.2) in Eq. (2.3), it can be shown after some mathematical manipulation that the weights $h_{opt}(l, k)$ that minimizes the MSE is given by the following 2-D Wiener-Hopf equation:

$$\mathbf{H}_{opt} = \mathbf{R}^{-1}\mathbf{P} \quad (2.4)$$

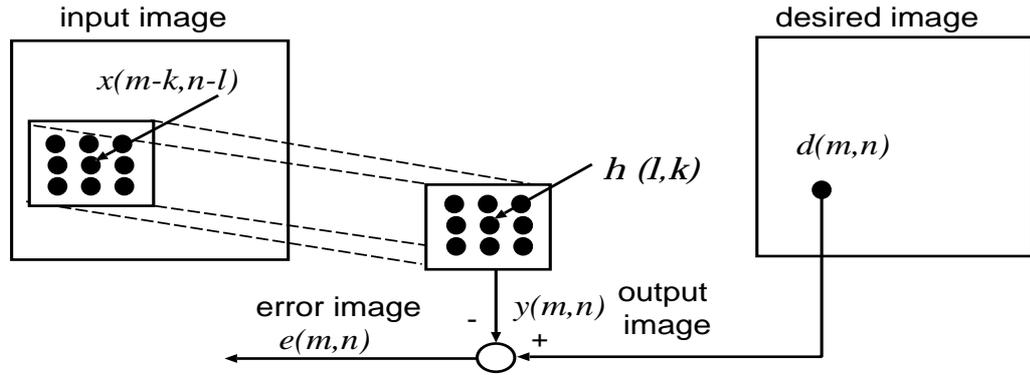


Figure 2.1: 2-D Wiener filter.

where

$$\mathbf{R} = E\{\mathbf{X}_j \mathbf{X}_j^t\} \quad (2.5)$$

$$\mathbf{P} = E\{\mathbf{X}_j d(m, n)\}. \quad (2.6)$$

$$\mathbf{H}_{opt} = [h_{opt}(0, 0), h_{opt}(1, N-1), \dots, h_{opt}(0, N-1), \\ h_{opt}(1, N-1), \dots, h_{opt}(N-1, N-1)]^t. \quad (2.7)$$

with

$$\mathbf{X}_j = [x(m, n), x(m+1, n), \dots, x(m-N+1, n), \\ x(m, n+1) \dots, x(m-N+1, n-N+1)]^t \quad (2.8)$$

However, no perfect solution for \mathbf{H}_{opt} can be obtained due to the fact that infinite data is required to calculate the exact correlation matrix \mathbf{R} and the cross correlation vector \mathbf{P} . Moreover, the statistics of the input signal $x(m, n)$ and the desired signal $d(m, n)$ may not be available, or could be space varying.

An alternative approach is to use adaptive filter. An adaptive filter relies for its operation on a recursive algorithm which is responsible for updating the filter's weights so that the error between the desired signal $d(m, n)$ and output signal $y(m, n)$ is minimized in the mean square error sense. The most commonly used adaptive algorithm is the least mean square (LMS) algorithm due to its simplicity in computation.

2.3 2-D LMS Adaptive FIR Filters

Consider the N by N , causal, 2-D adaptive FIR filter shown in Fig. 2.2. The filter's input $x(m, n)$ is a 2-D stationary signal of size $M \times M$. The filter output $y(m, n)$ is calculated by

$$\begin{aligned} y(m, n) &= \sum_{l=0}^{N-1} \sum_{k=0}^{N-1} h_j(l, k) x(m-l, n-k) \\ &= \mathbf{H}_j^t \mathbf{X}_j \end{aligned} \quad (2.9)$$

where j is some function of (m, n) specifying the indexing scheme used to process the 2-D data (cf. Fig. 2.3), and \mathbf{H}_j and \mathbf{X}_j are respectively the adaptive filter's weight-vector and the input data vector given at iteration j by

$$\begin{aligned} \mathbf{X}_j &= [x(m, n), \dots, x(m-N+1, n), \dots, x(m-N+1, n-N+1)]^t \\ \mathbf{H}_j &= [h_j(0, 0), \dots, h_j(0, N-1), \dots, h_j(N-1, N-1)]^t. \end{aligned} \quad (2.10)$$

The two-dimensional LMS (TDLMS) weights update can be obtained in straightforward manner using the steepest descent algorithm as follows [17]:

$$\mathbf{H}_{j+1} = \mathbf{H}_j - \mu \mathbf{G}_j \quad (2.11)$$

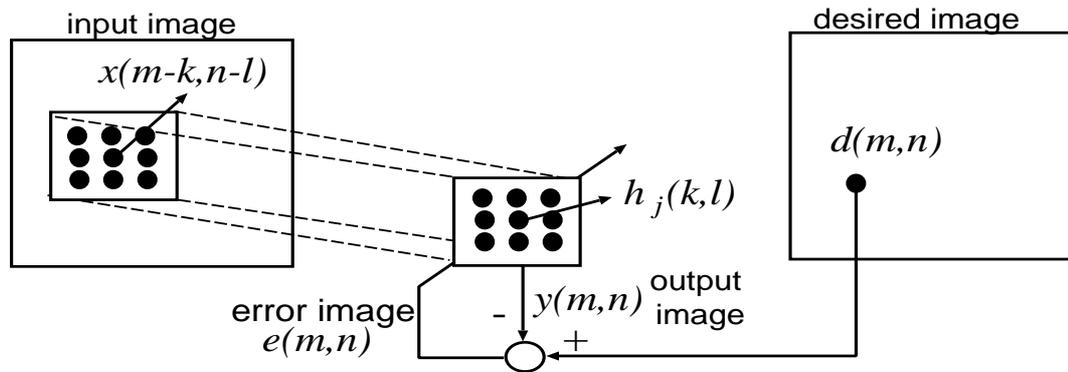


Figure 2.2: 2-D adaptive filter.

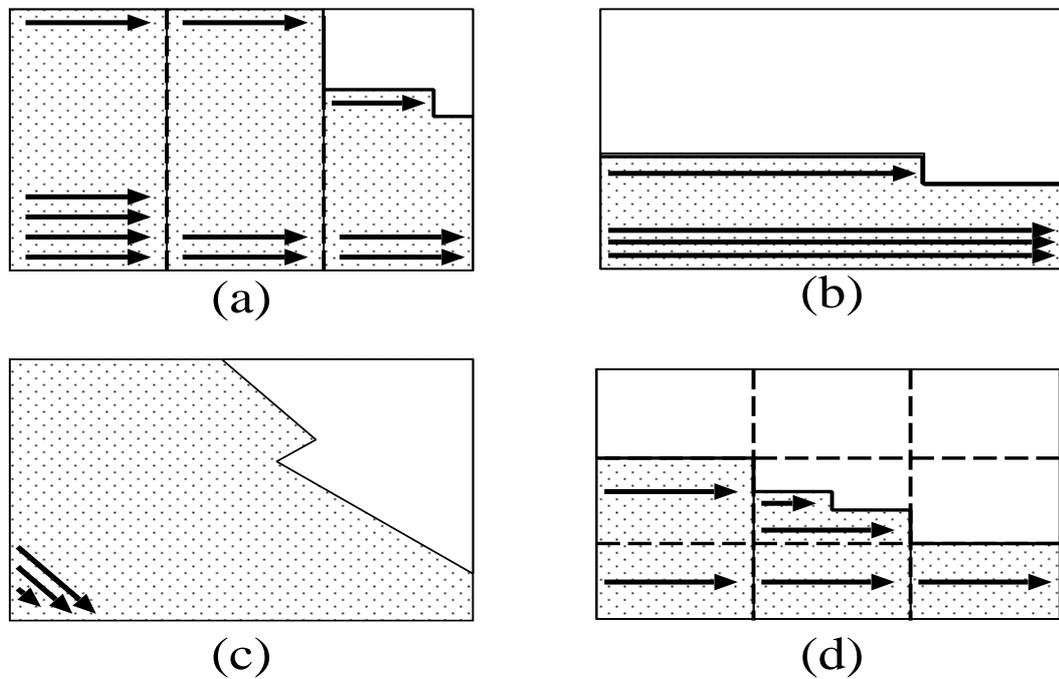


Figure 2.3: Indexing Schemes- a) rectangular indexing with short scan lines, b) rectangular indexing with image-length scan lines, c) diagonal indexing, and d) block diagonal indexing.

where

$$\mathbf{G}_j = \frac{\partial E[e^2(m, n)]}{\partial H_j} \approx \frac{\partial e^2(m, n)}{\partial H_j} \quad (2.12)$$

is the instantaneous gradient of the MSE. Accordingly, the weights update equation becomes:

$$\mathbf{H}_{j+1} = \mathbf{H}_j + \mu e(m, n) \mathbf{X}_{\mathbf{m}, \mathbf{n}} \quad (2.13)$$

To examine the performance of the TDLMS, it is common to consider first the convergence of the mean of the weight vector, and second, the convergence of the mean squared output error (MSE). The TDLMS is a direct extension of the well-known 1-D LMS proposed by Widrow [18]. And Eq. (2.9) is mathematically equivalent to an N^2 one dimensional convolution. Accordingly, the performance analysis of Eq. (2.13) is identical to that of the 1-D LMS [5], [18].

2.4 2-D Adaptive IIR Filters

Adaptive IIR filter can offer reduced computational complexity with reduced parameter set while achieving increased modeling flexibility provided by the recursive structure. However, there are several reason why the class of IIR adaptive filters has not received the same level of attention and success as the FIR class:

1. It is possible for the adaptive process to drive the poles of the adaptive filter outside the unit circle causing instability. Thus, there is a need for stability monitoring which is very difficult for 2-D IIR filters.
2. The error $e(m, n)$ is non-linear with respect to the filter parameters, consequently, the MSE is not quadratic with possible local minima causing the

adaptive process to converge to suboptimal solution.

3. Adaptive IIR filters have slow convergence due to the interaction between the movement of the poles and the movement of the zeros during the adaptive learning process. This means that although an IIR filter may have fewer coefficients than an equivalent FIR filter, the IIR filter may require more iterations to convergence.

In spite of these serious problems, the class of IIR filters remains of great interest for its potential to solve problems which require the synthesis of very long impulse responses. Two main approaches to adaptive IIR filters based on different error criteria have been considered so far [11]-[13]. The first one is called the output error formulation. The second approach is called the equation error formulation. In the following we introduce these two approaches to the IIR adaptive filtering.

2.4.1 Output Error Formulation

Fig. 2.4 shows 2-D adaptive IIR filter in output-error formulation. The output error filter is described by the recursive difference equation

$$y(m, n) = \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} \hat{a}(i, j)u(m-i, n-j) + \sum_{\substack{i=0 \\ (i,j) \neq (0,0)}}^{M_1} \sum_{j=0}^{M_2} \hat{b}(i, j)y(m-i, n-j) \quad (2.14)$$

The objective of any adaptive algorithm is to minimize some performance criteria based on minimizing some error criteria. The most common cost function is the mean square error (MSE) given as:

$$\text{MSE} = E\{e^2(m, n)\} \quad (2.15)$$

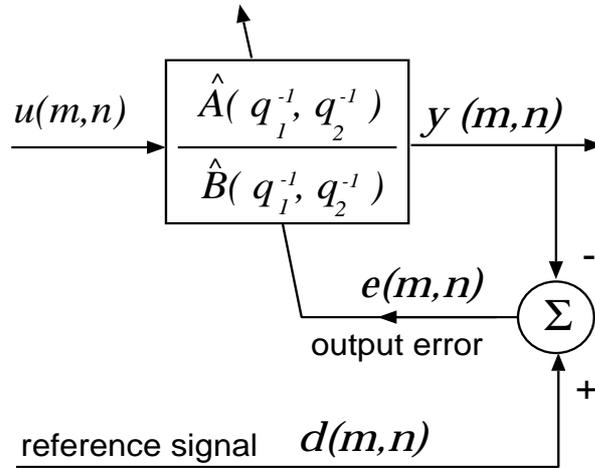


Figure 2.4: Output error formulation.

where $e(m,n)$ is called the output error, i.e., it is the instantaneous error between the adaptive filter output and the output of the unknown system which it attempts to match. It is known that the use of the output error in the formulation of the cost function prevents bias in the solution due to noise in the desired signal. However, since the output error is non-linear in terms of the filter parameters, the current filter parameters now depend upon previous filter coefficients, which are time varying. This leads to non-quadratic MSE surface with possible one or more local minima.

2.4.2 Equation Error Formulation

Fig. 2.5 shows 2-D adaptive IIR filter in equation-error formulation. Equation error adaptive filter improves upon the unsatisfactory performance of the output error IIR by filtering the output error with the denominator function and then

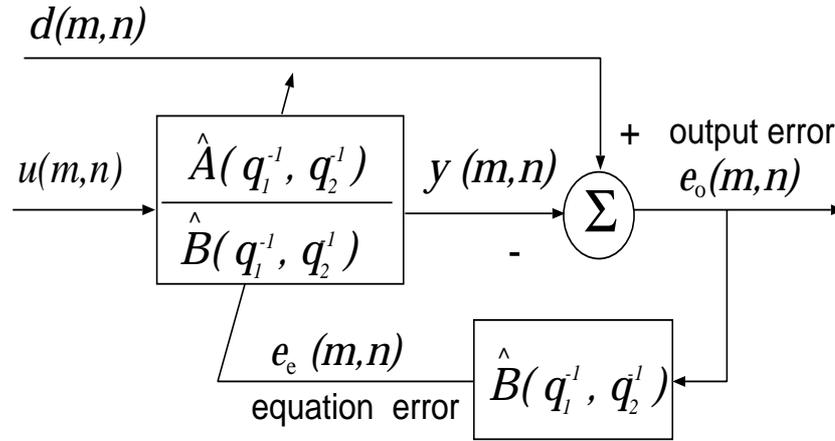


Figure 2.5: Equation error formulation.

minimizes the new error function. The new error function which is called equation error is given by

$$e_e(m, n) = \hat{B}(q_1^{-1}, q_2^{-1})d(m, n) - \hat{A}(q_1^{-1}, q_2^{-1})u(m, n) \quad (2.16)$$

The equation error is clearly linear in the filter's parameters. That is, the MSE surface is quadratic with one global minima. Moreover, equation-error adaptive IIR filter has fast convergence since it is equivalent to an FIR filter with two inputs namely $d(m, n)$ and $u(m, n)$. However, minimizing the equation error is not the same as minimizing the output error. That is, the solution given by LMS algorithm based on equation error criteria is not necessarily the same as the Wiener solution for the original problem. This is due to the fact that, by filtering the desired signal $d(m, n)$, the measurement noise buried in it is also filtered. And thus the adaptive filter will try to minimize the power of the noise that reaches the equation error while identifying the poles of the filter. These conflicting two

goals may lead to biased parameter estimates.

2.5 2-D Adaptive Filtering Applications

Two dimensional adaptive filters have found applications in the area of telecommunications, seismology, biomedicine, and image processing. Three applications of 2-D adaptive filters will be discussed in this section. 2-D system identification configuration, 2-D adaptive differential pulse code modulation and 2-D noise cancellation.

2.5.1 2-D System Identification Configuration

System identification configuration is a fundamental adaptive filtering concept that underlies many applications of adaptive filters. A 2-D adaptive filter is said to be used in system identification configuration when both the adaptive filter and the unknown system are excited by the same input signal $x(m, n)$ as shown in Fig. 2.6. Then the filter's parameters are iteratively adjusted to minimize some specified function of the error $e(m, n) = d(m, n) - y(m, n)$, where $y(m, n)$ is the output of the adaptive filter and $d(m, n)$ is the desired signal, which is the observed output of the unknown system. When the minimum of the cost function is achieved and the adaptive filter's parameters have converged to stable values, the adaptive filter provides a model of the unknown system in the sense that the adaptive process has formed the best approximation it can in the MSE sense using the structure imposed by the adaptive system. In order for the adaptive system to form a good model of the unknown system at all frequencies, it is important that the input signal has sufficiently rich spectral contents. A white

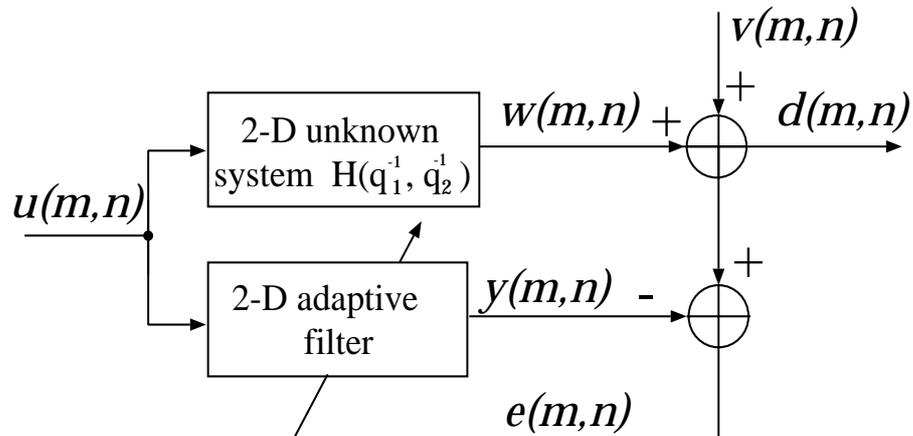
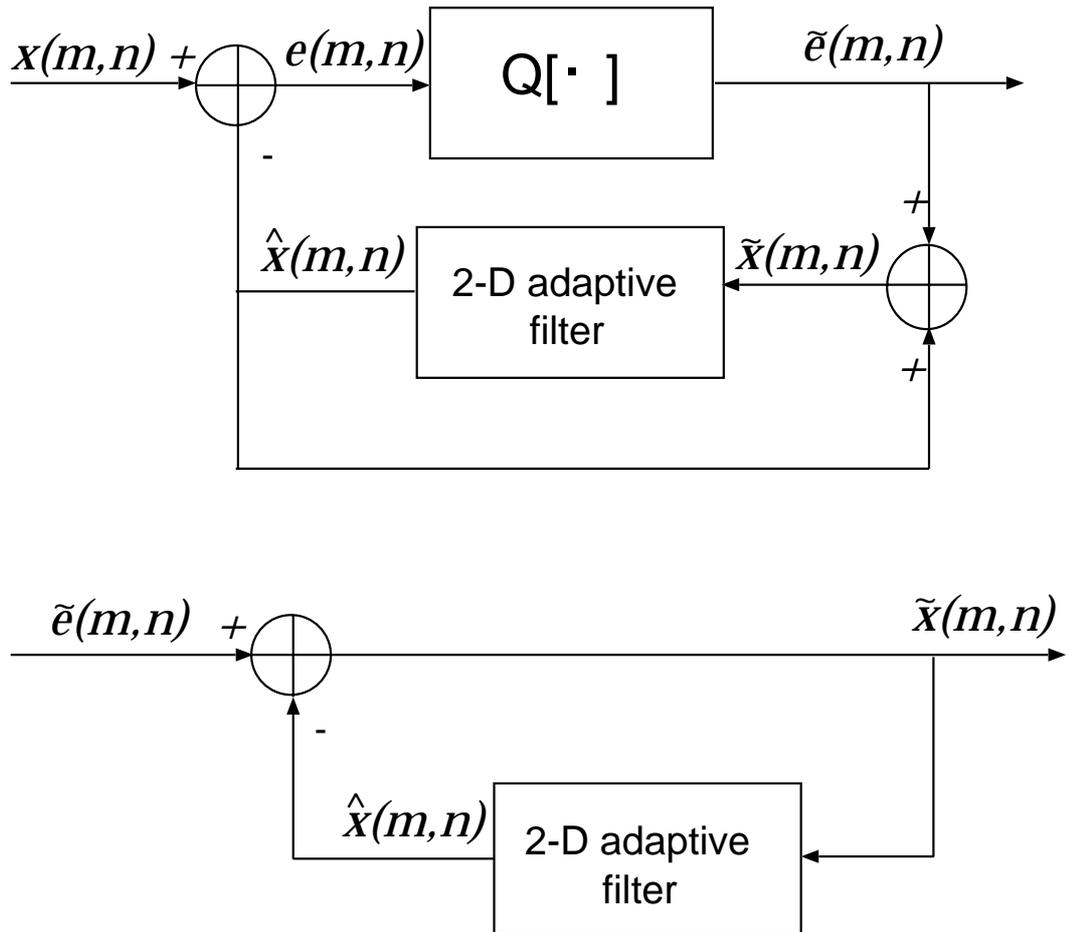


Figure 2.6: 2-D system identification configuration.

noise input signal is ideal because it excites all frequencies with equal power.

2.5.2 2-D Adaptive Differential Pulse Code Modulation

2-D adaptive filters can be used to update the predictor coefficients in adaptive differential pulse code modulation (DPCM) in image compression applications [11]. Fig. 2.7 shows the system identification configuration for the 2-D ADPCM. The predictor is an adaptive filter which accounts for the input signal's unknown and slowly changing statistics. The input signal can be a gray level image frame for which the error signal is encoded and transmitted. The error signal is calculated by subtracting the predicted signal from the input signal, and accordingly it has smaller variance and can be transmitted with fewer bits. For gray level images, experimental results [11] have shown that the reconstructed image using two-bit error word-length is visually indistinguishable from the original image.

Figure 2.7: 2-D ADPCM, encoder and decoder with quantizer Q .

2.5.3 2-D Adaptive Noise Cancellation

Fig. 2.8 shows 2-D adaptive noise canceler. As shown in this figure, there are two available signals: a signal which is contaminated with interference and a noise source which is in some way correlated with the interference. The objective is to produce the best possible estimate of the interference and subtract it from the

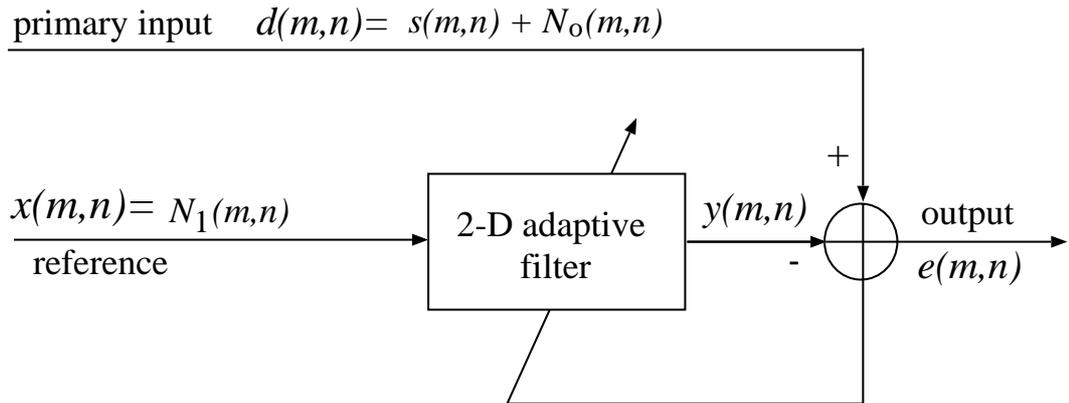


Figure 2.8: 2-D adaptive noise canceler.

contaminated signal.

2.6 Performance Measures in 2-D Adaptive Filters

In the development of adaptive systems, it is necessary to establish performance measures that provide comparative performance evaluations for different filter structures and adaptive algorithms. The choice of one algorithm over another is determined by one or more of the following factors [5], [11]:

1. **Stability.** Since any adaptive system is a form of closed loop feedback control system, there is always a potential for instability and divergence of the adaptive system, hence, the question of stability becomes of fundamental concern. Adaptive FIR filters are stable providing that the step size parameters are chosen small enough to satisfy general constraints. However, for adaptive IIR filters, the question of stability concerns as well the possible

immigration of the poles of the adaptive filters during the adaptive learning process. This problem is very essential in understanding the behaviour of adaptive IIR filtering algorithm because any reflection technique that can be used to keep the poles of the adaptive filter within the unit circle will slow the convergence speed of the adaptive algorithm.

2. **Convergence Rate.** This is defined as the number of iterations required for the adaptive algorithm with stationary input to converge close enough to the optimal Wiener solution.
3. **Misadjustment.** The misadjustment provides a measure of the amount by which the final value of the mean-squared error deviates from the minimum mean squared-error produced by the Wiener filter.
4. **Computational Complexity.** This issue includes, 1) number of multiplication, number of additions required to complete one iteration of the algorithm. 2) the size of memory locations required to store the data and 3) the investment required to program the algorithm on a computer.
5. **Robustness.** Robustness of the adaptive algorithms is very important measure and involves two important issues: 1) robustness with respect to external noise and 2) robustness with respect to algorithmic ill-conditioning and arithmetic quantization noise.

Chapter 3

Bias Removal Algorithm for 2-D Equation Error Adaptive IIR Filters

3.1 Introduction

Equation error adaptive IIR filters have the advantages of fast convergence and unimodal mean square error surface. However, their main drawback is that they converge to biased parameter estimates in the presence of measurement noise.

A 1-D equation error cascade IIR filtering algorithm has been proposed by Gao and Snelgrove [16]. This algorithm is based on the concept of backpropagating the desired signal $d(m)$ through the inverse of the all pole second order sections while the input signal $u(m)$ is passed through the adaptive filter transversal section, cf. Fig. 3.1, such that new intermediate errors $e_i(m)$, $i = 1, 2, \dots, m$ are generated; then the filter's parameters are adjusted to minimize those intermediate errors.

This cascade structure has the advantages of easy stability check and low parameter sensitivity. Moreover, minimizing the intermediate equation error functions instead of the output error offers significant reduction in the gradi-

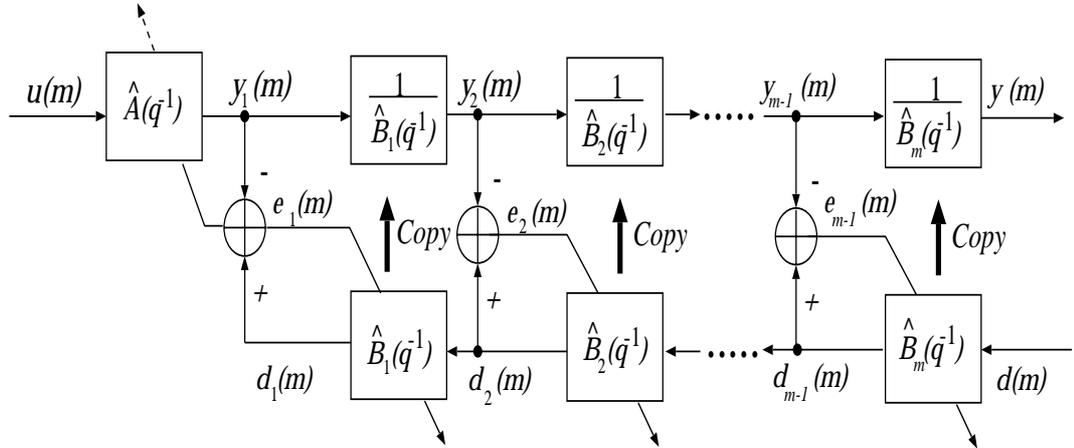


Figure 3.1: 1-D backpropagation adaptive cascade IIR filters.

ents' computational complexity [16]. Such advantages are of great interest in 2-D adaptive IIR filters. Thus, based on the extension of the backpropagation concept [16] to 2-D case, several 2-D cascade and parallel-cascade IIR filtering algorithms have been proposed in the literature [15], [33], [34]. The algorithm in [15] uses the backpropagation concept for 2-D IIR filters with separable denominator function. This class of 2-D IIR filters have several advantages. In addition to the simplicity of stability monitoring, separable denominator adaptive IIR filters offer significant reduction in the computational load when compared with direct form 2-D IIR filters. On the other hand, and if the numerator is nonseparable polynomial, separable denominator IIR filters can be efficiently used to approximate nonseparable 2-D IIR filters [2].

However, the main drawback of the algorithm [15], as well as [33] and [34], is that they are based on minimizing equation error functions; accordingly, it

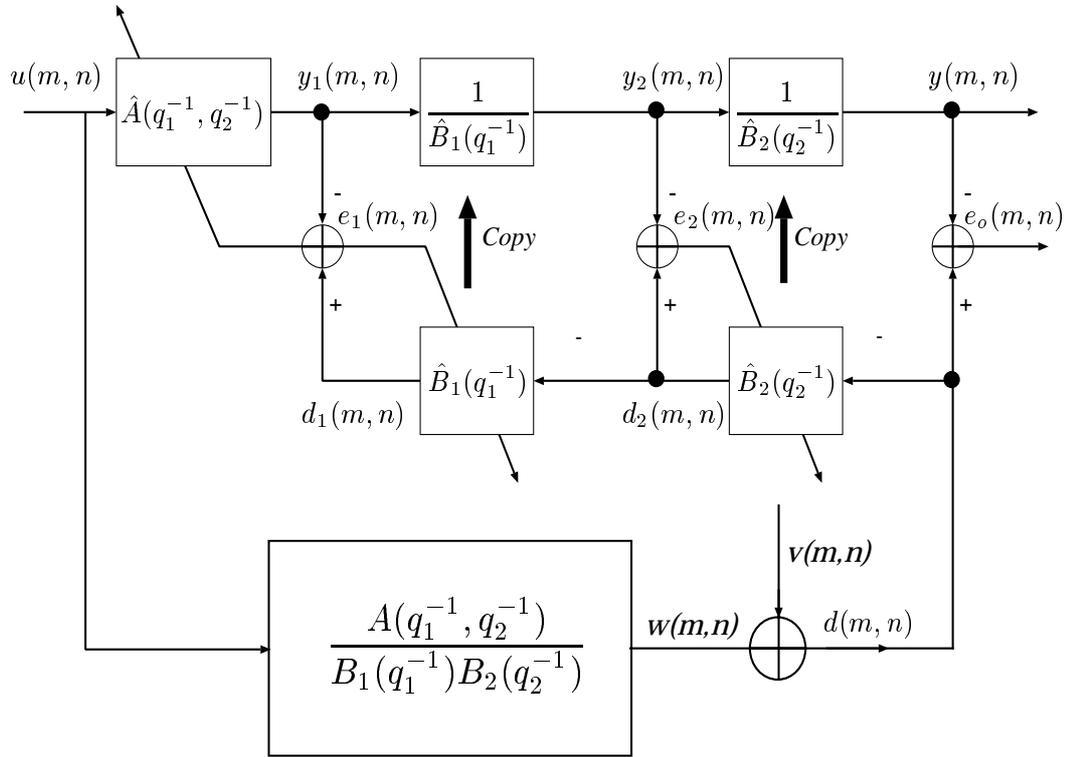


Figure 3.2: Equation error formulation for separable denominator 2-D IIR filters.

is expected that they converge to biased parameter estimates when the desired signal is contaminated with measurement noise. So far, stability analysis and the performance of the 1-D backpropagation cascade structure [16] as well as the 2-D backpropagation cascade structure [15]-[34] have not been considered.

This chapter presents a 2-D Bias Removal Algorithm (2DBRA) for the backpropagation cascade structure [15]. The 2DBRA presented here makes use of the idea of using a scaled value of the output-error as an estimate for the measurement noise which has been proposed in [36] for direct form 1-D adaptive IIR filters. This idea resembles in fact the error-feedback method used for round-

off noise reduction [37] in digital filters. However, for any adaptive algorithm that makes use of such idea, the scaling factor can not be fixed value as it is the case in roundoff noise reduction, and should be time varying in a way that the output-error feedback mechanism works only when the output-error becomes good estimates of the measurement noise.

3.2 2-D Backpropagation Adaptive Cascade IIR Filters

Consider the 2-D separable denominator IIR filter in a system identification configuration as shown in Figure 3.2. The observable output $d(m, n)$ of the unknown system is given by

$$d(m, n) = w(m, n) + v(m, n) \quad (3.1)$$

$$w(m, n) = \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} a(i, j)u(m-i, n-j) + \sum_{\substack{i=0 \\ (i,j) \neq (0,0)}}^{M_1} \sum_{j=0}^{M_2} b_1(i)b_2(j)w(m-i, n-j) \quad (3.2)$$

where $u(m, n)$, $m = 0, \dots, M$, $n = 0, \dots, N$, is the input signal, $w(m, n)$ is the noise free output signal, and $v(m, n)$ is a zero mean measurement noise assumed to be independent of $w(m, n)$ and $u(m, n)$.

In Figure 3.2, $A(q_1^{-1}, q_2^{-1})$ denotes the filter transversal section, $B_1(q_1^{-1})$ and

$B_2(q_2^{-1})$ are respectively the denominator horizontal and vertical parts, i.e.

$$\begin{aligned}
A(q_1^{-1}, q_2^{-1}) &= \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} a(i, j) q_1^{-i} q_2^{-j} \\
B_1(q_1^{-1}) &= 1 - \sum_{i=1}^{M_1} b_1(i) q_1^{-i} \\
B_2(q_2^{-1}) &= 1 - \sum_{j=1}^{M_2} b_2(j) q_2^{-j}
\end{aligned} \tag{3.3}$$

where q_1^{-1} and q_2^{-1} are used to denote spatial delay operators in the horizontal and vertical directions respectively.

For the separable denominator adaptive IIR filter shown in Figure 3.2, the desired signal $d(m, n)$ is backpropagated through the cascade of the adaptive filter's denominator vertical and horizontal sections in a way that two intermediate error functions, namely $e_1(m, n)$ and $e_2(m, n)$, can be generated as follows:

$$\begin{aligned}
e_1(m, n) &= d_1(m, n) - y_1(m, n) \\
&= d_2(m, n) - \sum_{i=1}^{M_1} \hat{b}_1(i) d_2(m - i, n) - \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} \hat{a}(i, j) u(m - i, n - j) \\
&= d_2(m, n) - \hat{\boldsymbol{\theta}}_1^T(k - 1) \boldsymbol{\varphi}_1(m, n)
\end{aligned} \tag{3.4}$$

$$\begin{aligned}
e_2(m, n) &= d_2(m, n) - y_2(m, n) \\
&= d(m, n) - \sum_{j=1}^{M_2} \hat{b}_2(j) d(m, n - j) - y_2(m, n) \\
&= \hat{\boldsymbol{\theta}}_2^T(k - 1) \boldsymbol{\varphi}_2(m, n) - y_2(m, n).
\end{aligned} \tag{3.5}$$

In the inner product notations, k denotes the iteration number used in updating the coefficients; the value of k is some function of (m, n) specifying the indexing

scheme, and

$$\hat{\boldsymbol{\theta}}_1(k) = [\hat{b}_1^k(1), \dots, \hat{b}_1^k(M_1), \hat{a}^k(0, 0), \dots, \hat{a}^k(N_1, N_2)]^T \quad (3.6)$$

$$\begin{aligned} \boldsymbol{\varphi}_1(m, n) &= [d_2(m-1, n), \dots, d_2(m-M_1, n), \\ &\quad u(m, n), \dots, u(m-N_1, n-N_2)]^T \end{aligned} \quad (3.7)$$

$$\hat{\boldsymbol{\theta}}_2(k) = [1, -\hat{b}_2^k(1), \dots, -\hat{b}_2^k(M_2)]^T \quad (3.8)$$

$$\boldsymbol{\varphi}_2(m, n) = [d(m, n), \dots, d(m, n-M_2)]^T. \quad (3.9)$$

Here, $\hat{\boldsymbol{\theta}}_1(k)$ denotes the parameter vector of the adaptive filter's transversal section and the denominator horizontal section; $\hat{\boldsymbol{\theta}}_2(k)$ denotes the parameter vector of the denominator vertical section; $\boldsymbol{\varphi}_1(m, n)$ and $\boldsymbol{\varphi}_2(m, n)$ are two regressor vectors associated with the adaptive filter.

The 2-D adaptive algorithm [15] updates the adaptive filter parameter vectors $\hat{\boldsymbol{\theta}}_1(k)$ and $\hat{\boldsymbol{\theta}}_2(k)$ subject to minimizing the intermediate error functions $e_1(m, n)$ and $e_2(m, n)$ respectively. The parameter update procedures are given by

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1^T(k) &= \hat{\boldsymbol{\theta}}_1^T(k-1) - \mu_1 \frac{1}{2} \frac{\partial e_1^2(m, n)}{\partial \hat{\boldsymbol{\theta}}_1(k-1)} \\ &= \hat{\boldsymbol{\theta}}_1^T(k-1) + \mu_1 e_1(m, n) \boldsymbol{\varphi}_1(m, n) \end{aligned} \quad (3.10)$$

$$\begin{aligned} \hat{\boldsymbol{\theta}}_2(k) &= \hat{\boldsymbol{\theta}}_2(k-1) - \boldsymbol{\mu}_2 \frac{1}{2} \frac{\partial e_2^2(m, n)}{\partial \hat{\boldsymbol{\theta}}_2(k-1)} \\ &= \hat{\boldsymbol{\theta}}_2(k-1) - \boldsymbol{\mu}_2 e_2(m, n) \boldsymbol{\varphi}_2(m, n) \end{aligned} \quad (3.11)$$

where, $\boldsymbol{\mu}_2 = \text{diag}[0, \mu_2, \dots, \mu_2]$ is a diagonal step size matrix.

Now, whether minimizing these intermediate error signals $e_1(m, n)$ and $e_2(m, n)$ will lead to minimizing the output error signal or not, requires further discussion. Indeed, experimental results have shown that, in analogy with the 1-D equation

error IIR adaptive filtering algorithms, the algorithm [15] performs significantly well in noise free case, however its performance degrades significantly when the desired signal $d(m, n)$ is contaminated with additive noise.

In the following section we present the 2-D Bias Removal Algorithm (2DBRA) 2DBRA which is derived from modifying the regressor vectors $\boldsymbol{\varphi}_1(m, n)$ and $\boldsymbol{\varphi}_2(m, n)$ in Eqs. (3.10) and (3.11) respectively. The performance of the algorithm [15] will be discussed in the following sections as a special case of the 2DBRA.

3.3 2-D Bias Removal Algorithm (2DBRA)

Using Eq. (3.1), the regressor vector $\boldsymbol{\varphi}_2(m, n)$ in Eq.(3.9) can be rewritten as

$$\boldsymbol{\varphi}(m, n) = \boldsymbol{\varphi}_{2_0}(m, n) + \mathbf{v}_2(m, n) \quad (3.12)$$

with

$$\boldsymbol{\varphi}_{2_0}(m, n) = [w(m, n), \dots, w(m, n - M_2)]^T \quad (3.13)$$

$$\mathbf{v}_2(m, n) = [v(m, n), \dots, v(m, n - M_2)]^T. \quad (3.14)$$

To handle the bias problem, the 2DBRA tries to counteract the effect of the noise vector $\mathbf{v}_2(m, n)$ embedded in the regressor vector $\boldsymbol{\varphi}_2(m, n)$ of the coefficients update equation (3.11), by using a scaled value of the output-error vector

$$\boldsymbol{\varepsilon}_o(m, n) = [e_o(m, n), e_o(m, n - 1), \dots, e_o(m, n - M_2)]^T \quad (3.15)$$

with

$$e_o(m, n) = d(m, n) - y(m, n) \quad (3.16)$$

to estimate the colored noise

$$v_1(m, n) = \hat{\boldsymbol{\theta}}_2^T(k-1)\mathbf{v}_2(m, n) \quad (3.19)$$

that reaches the intermediate signal $d_2(m, n)$ (see Figure 3.2). That is, the 2DBRA modifies Eq. (3.10) as follows:

$$\hat{\boldsymbol{\theta}}_1(k) = \hat{\boldsymbol{\theta}}_1(k-1) + \mu_1 e_1(m, n)[\boldsymbol{\varphi}_1(m, n) - \tau_1(m, n)\boldsymbol{\varepsilon}_2(m, n)] \quad (3.20)$$

where

$$\boldsymbol{\varepsilon}_2(m, n) = [e_2(m-1, n) \cdots e_2(m-M_1, n), \mathbf{0}]^T \quad (3.21)$$

and $\tau_1(m, n)$ is a scaling factor defined as

$$\tau_1(m, n) = \min\left(1, \alpha_1 \frac{\|\boldsymbol{\varphi}_1(m, n)\|}{\|\boldsymbol{\varepsilon}_2(m, n)\|}\right), \quad \alpha_1 \geq 0. \quad (3.22)$$

The time varying scaling factor $\tau_1(m, n)$ ($\tau_2(m, n)$) defined in Eq. (3.22) ((3.18)) is chosen to be inversely proportional to the variance of the error $e_2(m, n)$ ($e_o(m, n)$). At the beginning of the adaptive process, the variance of $e_2(m, n)$ ($e_o(m, n)$) is large, and the value of $\tau_1(m, n)$ ($\tau_2(m, n)$) is close to zero, i.e., the adaptive learning process works almost without the bias removal mechanism. As the variance of the error $e_2(m, n)$ ($e_o(m, n)$) decreases, $e_2(m, n)$ ($e_o(m, n)$) tends to be more accurate estimate of the measurement noise in $d_2(m, n)$ ($d(m, n)$), and the value of $\tau_1(m, n)$ ($\tau_2(m, n)$) increases gradually to reach a maximum value of unity. However, as the value of the scaling factor $\tau_1(m, n)$ ($\tau_2(m, n)$) increases, a larger portion of the output error $e_2(m, n)$ ($e_o(m, n)$) is installed in the 2DBRA. Consequently, the stability of the 2DBRA becomes more critical.

In the following section it is shown that the stability of the 2DBRA can be maintained under general conditions imposed on the step size parameters μ_1 and μ_2 and the constants α_1 and α_2 used in calculating the scaling factors $\tau_1(m, n)$ and $\tau_2(m, n)$ respectively.

3.4 Convergence Analysis of the 2DBRA

Rigorous convergence analysis of the 2-D adaptive cascade structure in Figure 3.3 as a whole dynamic is very complicated due to the interaction between the cascaded sections. And, Eqs. (3.4)((3.5)) when considered together will lead to highly nonlinear optimization problem.

In an attempt to analyze the convergence behaviour of the 2DBRA, as well as the algorithm [15], the cascade structure is divided into two parts. The first one consists of the transversal section and the denominator horizontal section. And the second part consists of the denominator vertical section. Then it is shown that the interconnection between the cascaded sections, can be replaced by a noise component in the desired signal of each part.

Now under the assumption that the adaptation process of the filter coefficients is slow, i.e. the used step size parameters are sufficiently small, stability analysis of each of the difference equation that describes the parameter-error vector of each part can be carried out as if these two part were independent using the stability robustness theory of perturbed linear system. On the other hand, and in order to further reduce the interconnection between the dynamic behavior of the cascaded sections, it is assumed that, at the update iteration k , the signals

$y_1(m, n)$, $d_2(m, n)$ and $d_1(m, n)$ are calculated from the estimated parameters at iteration $k - 1$, while the signal $y_2(m, n)$ and $y(m, n)$ are calculated from the estimated parameters at iteration $k - 2$. Accordingly, at iteration k , the intermediate signal $y_2(m, n)$ is independent of $\hat{\boldsymbol{\theta}}_1(k - 1)$, that is $e_2(m, n)$ is independent of $\hat{\boldsymbol{\theta}}_1(k - 1)$. In this view, the error function $e_1(m, n)(e_2(m, n))$ given in Eqs. (3.4)((3.5)) can be considered to be linear with respect to the updated parameter vector $\hat{\boldsymbol{\theta}}_1(k - 1)$ ($\hat{\boldsymbol{\theta}}_2(k - 1)$); and each section has its global minimum. However, there is no guarantee that in the presence of measurement noise, the convergence of each section to its global minimum will lead to global convergence of the output error of the whole structure.

In the following subsections, the convergence of the mean of the following parameter-error vectors is considered.

$$E\{\tilde{\boldsymbol{\theta}}_1(k)\} = E\{\boldsymbol{\theta}_1 - \hat{\boldsymbol{\theta}}_1(k)\} \quad (3.23)$$

$$E\{\tilde{\boldsymbol{\theta}}_2(k)\} = E\{\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2(k)\}. \quad (3.24)$$

Here, the tilde “ \sim ” is used to denote the error in the estimated entities, and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the ideal parameter vectors defined as

$$\boldsymbol{\theta}_1 = [b_1(1), \dots, b_1(M_1), a(0, 0), \dots, a(N_1, N_2)]^T \quad (3.25)$$

$$\boldsymbol{\theta}_2 = [1, -b_2(1), \dots, -b_2(M_2)]^T. \quad (3.26)$$

In the following convergence analysis, it is assumed that the parameter vector $\hat{\boldsymbol{\theta}}_1(k - 1)$ ($\hat{\boldsymbol{\theta}}_2(k - 1)$) is independent of the regressor vector $\boldsymbol{\varphi}_1(m, n)$ ($\boldsymbol{\varphi}_2(m, n)$). This assumption is similar to the well known independence assumption used in the analyzing the convergence behavior of LMS adaptive filtering algorithms [10].

3.4.1 Convergence Analysis of the First Part of the Structure

The parameters of the part of the adaptive filter that contains the transversal section and the denominator horizontal section are updated using Eq. (3.20). This part of the structure can be considered as a time-varying FIR filter with desired signal $d_2(m, n)$, c.f. Figure 3.3. Before going into the stability analysis of Eq. (3.23), we will first show that the desired signal $d_2(m, n)$ can be decomposed into three components as follows:

$$d_2(m, n) = d_{2_0}(m, n) + \zeta_2(m, n) + v_1(m, n) \quad (3.27)$$

with $d_{2_0}(m, n)$ represents the noise free desired signal for this part of the structure; $v_1(m, n)$ is the filtered version of the measurement noise $v(m, n)$ as defined in Eq. (3.19); $\zeta_2(m, n)$ is a perturbation component related to the fluctuation of the parameter-error vector $\tilde{\boldsymbol{\theta}}_2(k-1)$.

From Figure 3.3, and using Eqs. (3.12) and (3.19), we have

$$\begin{aligned} d_2(m, n) &= \hat{\boldsymbol{\theta}}_2^T(k-1)\boldsymbol{\varphi}_2(m, n) \\ &= \hat{\boldsymbol{\theta}}_2^T(k-1)^T\boldsymbol{\varphi}_{2_0} + v_1(m, n) \\ &= \boldsymbol{\theta}_2^T\boldsymbol{\varphi}_{2_0}(m, n) - \tilde{\boldsymbol{\theta}}_2^T(k-1)\boldsymbol{\varphi}_{2_0}(m, n) + v_1(m, n) \\ &= \boldsymbol{\theta}_2^T\boldsymbol{\varphi}_{2_0}(m, n) + \zeta_2(m, n) + v_1(m, n) \end{aligned} \quad (3.28)$$

where

$$\zeta_2(m, n) = -\tilde{\boldsymbol{\theta}}_2^T(k-1)\boldsymbol{\varphi}_{2_0}(m, n). \quad (3.29)$$

However, from Eq. (3.2) we can find that

$$\boldsymbol{\theta}_1^T \boldsymbol{\varphi}_{1_0}(m, n) = \boldsymbol{\theta}_2^T \boldsymbol{\varphi}_{2_0}(m, n) \quad (3.30)$$

where

$$\begin{aligned} \boldsymbol{\varphi}_{1_0}(m, n) &= [d_{2_0}(m-1, n), \dots, d_{2_0}(m-M_1, n), \\ &\quad u(m, n), \dots, u(m-N_1, n-N_2)]^T \end{aligned} \quad (3.31)$$

$$d_{2_0}(m, n) = d_{2_0}(m, n) = \boldsymbol{\theta}_2^T \boldsymbol{\varphi}_{2_0}(m, n). \quad (3.32)$$

Substituting Eq. (3.30) in Eq. (3.28) we arrive at Eq. (3.27) with $d_{2_0}(m, n) = \boldsymbol{\theta}_1^T \boldsymbol{\varphi}_{1_0}(m, n)$.

Now, substituting Eq. (3.20) in Eq. (3.23) and using Eqs. (3.4) and (3.27)

we get

$$\mathbb{E}\{\tilde{\boldsymbol{\theta}}_1(k)\} = [A_1 + B_1(m, n)]\mathbb{E}\{\tilde{\boldsymbol{\theta}}_1(k-1)\} + A_{v_1} + A_{\zeta_2} \quad (3.33)$$

where

$$A_1 = I - \mu_1 \mathbf{R}_{\varphi_1 \varphi_1} \quad (3.34)$$

$$B_1(m, n) = \mu_1 \mathbb{E}\{\tau_1(m, n) \boldsymbol{\varphi}_1(m, n) \boldsymbol{\varepsilon}_2^T(m, n)\} \quad (3.35)$$

$$A_{v_1} = \mu_1 \mathbb{E}\{[\mathbf{b}_1^T \mathbf{v}_1(m, n) - v_1(m, n)][\boldsymbol{\varphi}_1(m, n) - \tau_1(m, n) \boldsymbol{\varepsilon}_2(m, n)]\} \quad (3.36)$$

$$A_{\zeta_2} = -\mu_1 \mathbb{E}\{[1, -\mathbf{b}_1^T] \boldsymbol{\zeta}_2(m, n) [\boldsymbol{\varphi}_1(m, n) - \tau_1(m, n) \boldsymbol{\varepsilon}_2(m, n)]\} \quad (3.37)$$

with

$$\mathbf{R}_{\varphi_1 \varphi_1} = \mathbb{E}\{\boldsymbol{\varphi}_1(m, n) \boldsymbol{\varphi}_1^T(m, n)\} \quad (3.38)$$

$$\mathbf{b}_1 = [b_1(1), \dots, b_1(M_1)]^T \quad (3.39)$$

$$\mathbf{v}_1(m, n) = [v_1(m-1, n), \dots, v_1(m-M_1, n)]^T \quad (3.40)$$

$$\boldsymbol{\zeta}_2(m, n) = [\zeta_2(m, n), \zeta_2(m-1, n), \dots, \zeta_2(m-M_1, n)]^T. \quad (3.41)$$

Note that the difference equation (3.33) has two forcing terms. The first one A_{ζ_2} is due to the interaction between the cascaded sections. And the second A_{v_1} is due to the colored measurement noise $v_1(m, n)$. Hence, even in the absence of measurement noise, the forcing term A_{ζ_2} may also cause the adaptive algorithm to converge to biased parameter estimates.

Now, the difference equation (3.33) is stable if the autonomous perturbed system presented by the first term of Eq. (3.33) is asymptotically stable and the forcing terms A_{ζ_2} and A_{v_1} are bounded.

i) *Stability Analysis of the Autonomous Part of Eq. (3.33)*

The autonomous perturbed system

$$\mathbb{E}\{\tilde{\boldsymbol{\theta}}_1(k)\} = [A_1 + B_1(m, n)]\mathbb{E}\{\tilde{\boldsymbol{\theta}}_1(k-1)\} \quad (3.42)$$

is exponentially and asymptotically stable if the following two conditions hold [38].

1. *All the eigenvalues of the the stability matrix A_1 are within the unit circle.*

Such condition is satisfied if and only if

$$0 \leq \mu_1 \leq \frac{2}{\lambda_{1\max}} \quad (3.43)$$

where $\lambda_{1\max}$ denotes the maximum eigenvalue of the input autocorrelation matrix $\mathbf{R}_{\varphi_1\varphi_1}$. Condition (3.43) can be replaced with a more practical and strict one given by

$$\mu_1 = \frac{2\sigma_1}{\text{tr}(\mathbf{R}_{\varphi_1\varphi_1})} \leq \frac{2}{\lambda_{1\max}}, \quad 0 < \sigma_1 \leq 1. \quad (3.44)$$

Then we can define the state transition matrix

$$\Phi_1(k) = A_1^k, \quad k > 0 \quad (3.45)$$

that satisfies

$$\|\Phi_1(k)\| \leq c_1 \beta_1^k, \quad c_1 > 0, \quad \beta_1 \in [0, 1]. \quad (3.46)$$

2. *The perturbation term $B_1(m, n)$ satisfies*

$$0 \leq \beta_1 + c_1 \|B_1(m, n)\| \leq 1. \quad (3.47)$$

From Eq. (3.35) we have

$$\begin{aligned} \|B_1(m, n)\| &= \|\mathbb{E} \{ \tau_1(m, n) \mu_1 \boldsymbol{\varphi}_1(m, n) \boldsymbol{\varepsilon}_2^T(m, n) \} \| \\ &\leq \alpha_1 \mathbb{E} \left\{ \frac{\|\boldsymbol{\varphi}_1(m, n)\|}{\|\boldsymbol{\varepsilon}_2(m, n)\|} \frac{2\sigma_1 \|\boldsymbol{\varphi}_1(m, n)\|}{\text{tr}(\mathbf{R}_{\varphi_1 \varphi_1})} \|\boldsymbol{\varepsilon}_2(m, n)\| \right\} \\ &\leq 2\sigma_1 \alpha_1. \end{aligned} \quad (3.48)$$

Thus, the value of α_1 should satisfy

$$0 \leq \alpha_1 \leq \frac{1 - \beta_1}{2\sigma_1 c_1}. \quad (3.49)$$

ii) *The Boundedness of the Forcing Term A_{v_1}*

Making use of Eq. (3.27) in Eq. (3.36), and considering the assumption that the measurement noise $v(m, n)$ is independent of $u(m, n)$ and $w(m, n)$, we can find that

$$\begin{aligned} \|A_{v_1}\| &= \mu_1 \|\mathbb{E} \{ [\mathbf{b}_1^T \mathbf{v}_1(m, n) - v_1(m, n)] [\mathbf{v}_1(m, n) - \tau_1(m, n) \mathbf{v}_1(m, n)] \} \| \\ &= \mu_1 \|[1 - \tau_1(m, n)] [\mathbf{R}_{v_1 v_1} \mathbf{b}_1 - \mathbb{E} \{ v_1(m, n) \mathbf{v}_1(m, n) \}]\| \\ &\leq \mu_1 \|\mathbf{R}_{v_1 v_1} \mathbf{b}_1 - \mathbb{E} \{ v_1(m, n) \mathbf{v}_1(m, n) \} \| \end{aligned} \quad (3.50)$$

with $\mathbf{R}_{v_1 v_1} = \mathbb{E}\{\mathbf{v}_1(m, n)\mathbf{v}_1^T(m, n)\}$. Hence, the forcing term A_{v_1} is bounded, and $\|A_{v_1}\|$ will approach zero as $\tau_1(m, n) \rightarrow 1$.

Notice that, without the bias removing mechanism, i.e. $\tau_1(m, n) = 0$, the norm of the forcing term A_{v_1} is proportional to the power of the measurement noise $v_1(m, n)$. Thus unless the desired signal $d_2(m, n)$ is free from measurement noise, i.e. $v_1(m, n) \equiv 0$, the adaptive algorithm [15] converges to biased parameter estimates.

iii) *The Boundedness of the Forcing Term A_{ζ_2}*

From Eq. (3.37) we have

$$\begin{aligned} \|A_{\zeta_2}\| &= \mu_1 \|\mathbb{E}\{[1, -\mathbf{b}_1]\zeta_2(m, n)[\boldsymbol{\varphi}_1(m, n) - \tau_1(m, n)\boldsymbol{\varepsilon}_2(m, n)]\}\| \\ &\leq \mu_1 \left(\sum_{i=0}^{M_1} b_1^2(i) \right)^{1/2} \mathbb{E}\{\|\zeta_2(m, n)\| \|\boldsymbol{\varphi}_1(m, n) - \tau_1(m, n)\boldsymbol{\varepsilon}_2(m, n)\|\}. \end{aligned} \quad (3.51)$$

Equation (3.51) states that the norm of the forcing term A_{ζ_2} is bounded providing that the error signal $\zeta_2(m, n)$ has finite variance. Generally speaking, there is no guarantee that the variance of the error signal $\zeta_2(m, n)$ will approach zero. However, the boundedness of the error $\zeta_2(m, n)$ is guaranteed if the step size μ_2 , used in the update procedure of the denominator vertical section, satisfies the necessary condition for the convergence of the LMS FIR filter $\hat{B}_2(q_2^{-1})$ in the variance which is given by [10]

$$\mu_2 < \frac{2}{3\text{tr}(\mathbf{R}_{\varphi_2 \varphi_2})} \quad (3.52)$$

with $\mathbf{R}_{\varphi_2 \varphi_2} = \mathbb{E}\{\boldsymbol{\varphi}_2(m, n)\boldsymbol{\varphi}_2^T(m, n)\}$. In order to reduce the influence of the

forcing term A_{ζ_2} , a sufficiently small step size μ_2 that guarantees small variance for $\tilde{\boldsymbol{\theta}}_2(k)$, and hence for $\zeta_2(m, n)$, should be used.

3.4.2 Convergence Analysis of the Second Part of the Structure

In Figure 3.3, the intermediate signal $y_2(m, n)$ can be considered as the desired signal for the adaptive FIR filter $\hat{B}_2(q_2^{-1})$. It can be decomposed into two components as follows:

$$y_2(m, n) = \boldsymbol{\theta}_2^T \boldsymbol{\varphi}_{2_0}(m, n) + \zeta_1(m, n). \quad (3.53)$$

The first component $\boldsymbol{\theta}_2^T \boldsymbol{\varphi}_{2_0}(m, n)$ can be considered as the noise free stationary desired signal; $\zeta_1(m, n)$ is a time-varying fluctuation component related to the parameter-error vector $\tilde{\boldsymbol{\theta}}_1(k-2)$.

Substituting Eq. (3.53) in Eq. (3.5) and using Eq. (3.12) we find

$$\begin{aligned} e_2(m, n) &= \hat{\boldsymbol{\theta}}_2^T(k-1) \boldsymbol{\varphi}_2(m, n) - \boldsymbol{\theta}_2^T \boldsymbol{\varphi}_{2_0}(m, n) - \zeta_1(m, n) \\ &= -\tilde{\boldsymbol{\theta}}_2^T(k-1) \boldsymbol{\varphi}_2(m, n) + \boldsymbol{\theta}_2^T \mathbf{v}_2(m, n) - \zeta_1(m, n). \end{aligned} \quad (3.54)$$

Substituting Eq. (3.17) in Eq. (3.24) and using Eq. (3.54) we find

$$\mathbb{E}\{\tilde{\boldsymbol{\theta}}_2(k)\} = [A_2 + B_2(m, n)] \mathbb{E}\{\tilde{\boldsymbol{\theta}}_2(k-1)\} + A_{v_2} + A_{\zeta_1} \quad (3.55)$$

where

$$A_2 = I - \boldsymbol{\mu}_2 \mathbf{R}_{\varphi_2 \varphi_2} \quad (3.56)$$

$$B_2(m, n) = \mathbb{E}\{\tau_2(m, n) \boldsymbol{\mu}_2 \boldsymbol{\varphi}_2(m, n) \boldsymbol{\varepsilon}_o^T(m, n)\} \quad (3.57)$$

and

$$A_{v_2} = \boldsymbol{\mu}_2 \mathbb{E}\{\boldsymbol{\theta}_2^T \mathbf{v}_2(m, n)[\boldsymbol{\varphi}_2(m, n) - \tau_2(m, n)\boldsymbol{\varepsilon}_o(m, n)]\} \quad (3.58)$$

$$A_{\zeta_1} = -\boldsymbol{\mu}_2 \mathbb{E}\{\zeta_1(m, n)[\boldsymbol{\varphi}_2(m, n) - \tau_2(m, n)\boldsymbol{\varepsilon}_o(m, n)]\}. \quad (3.59)$$

i) *Stability Analysis for the Autonomous Part of Eq. (3.55)*

The autonomous perturbed system

$$\mathbb{E}\{\tilde{\boldsymbol{\theta}}_2(k)\} = [A_2 + B_2(m, n)]\mathbb{E}\{\tilde{\boldsymbol{\theta}}_2(k-1)\} \quad (3.60)$$

is asymptotically stable if the following two conditions hold:

1. *All the eigenvalues of the stability matrix A_2 are within the unit circle.* This condition holds if the step size μ_2 satisfies

$$0 \leq \mu_2 \leq \frac{2}{\lambda_{2\max}} \quad (3.61)$$

where $\lambda_{2\max}$ denotes the maximum eigenvalue of the autocorrelation matrix $\mathbf{R}_{\varphi_2\varphi_2}$.

Condition (3.61) can be replaced with a more practical one given by

$$\mu_2 = \frac{2\sigma_2}{\text{tr}(\mathbf{R}_{\varphi_2\varphi_2})} \leq \frac{2}{\lambda_{2\max}}, \quad 0 < \sigma_2 \leq 1. \quad (3.62)$$

Then, we can define the state transition matrix $\Phi_2(k) = A_2^k$ that satisfies

$$\|\Phi_2(k)\| \leq c_2 \beta_2^k, \quad c_2 > 0, \quad \beta_2 \in [0, 1]. \quad (3.63)$$

2. *The perturbation term $B_2(m, n)$ satisfies*

$$0 \leq \beta_2 + c_2 \|B_2(m, n)\| \leq 1. \quad (3.64)$$

Calculating $\|B_2(m, n)\|$, we can find that condition (3.64) holds if

$$0 \leq \alpha_2 \leq \frac{1 - \beta_2}{2\sigma_2 c_2}. \quad (3.65)$$

ii) *The Boundedness of the Forcing Term A_{v_2}*

Substituting Eq. (3.16) in Eq. (3.58), and using the assumption that $v(m, n)$ is independent of $u(m, n)$ and $d(m, n)$, we find

$$\begin{aligned} \|A_{v_2}\| &= \|\boldsymbol{\mu}_2 \mathbf{E}\{\boldsymbol{\theta}_2^T \mathbf{v}_2(m, n)[\mathbf{v}_2(m, n) - \tau_2(m, n)\mathbf{v}_2(m, n)]\}\| \\ &= \mu_2 \|[1 - \tau_2(m, n)][\mathbf{R}_{v_2 v_2} \boldsymbol{\theta}_2]\| \\ &\leq \mu_2 \|\mathbf{R}_{v_2 v_2} \boldsymbol{\theta}_2\| \end{aligned} \quad (3.66)$$

with $\mathbf{R}_{v_2 v_2} = \mathbf{E}\{\mathbf{v}_2(m, n)\mathbf{v}_2^T(m, n)\}$. Hence the forcing term A_{v_2} is bounded, and $\|A_{v_2}\| \rightarrow 0$ as $\tau_2(m, n) \rightarrow 1$.

iii) *The Boundedness of the Forcing Term A_{ζ_1}*

From Eq. (3.59) we find that

$$\begin{aligned} \|A_{\zeta_1}\| &= \|\boldsymbol{\mu}_2 \mathbf{E}\{\zeta_1(m, n)\boldsymbol{\varphi}_2(m, n)\}\| \\ &\leq \mu_2 \mathbf{E}\{\|\zeta_1(m, n)\| \|\boldsymbol{\varphi}_2(m, n) - \tau_2(m, n)\boldsymbol{\varepsilon}_o(m, n)\|\}. \end{aligned} \quad (3.67)$$

Equation (3.67) states that, the forcing term A_{ζ_1} is bounded providing that the error signal $\zeta_1(m, n)$ is bounded. Practically, if the denominator horizontal section, i.e. $1/B_1(q^{-1})$ is assured to be stable, then $y_2(m, n)$ is bounded, and consequently $\zeta_1(m, n)$ is bounded. In order to reduce the influence of the forcing term A_{ζ_1} , a step size parameter μ_1 that guarantees finite variance for the FIR filter with parameter vector $\hat{\boldsymbol{\theta}}_1(k)$, should be used. That is, μ_1 should satisfy the condition

[10]

$$\mu_1 < \frac{2}{3\text{tr}(\mathbf{R}_{\varphi_1\varphi_1})} \quad (3.68)$$

3.5 Experimental Results and Discussion

Example 1: Noisy desired signal (white measurement noise)

In this example, the 2-D adaptive algorithm [15] and the 2DBRA are applied to the system identification experiment. A 2-D zero mean white Gaussian signal of unit variance and size 256 by 256 is used for the input signal $u(m, n)$. And a zero mean, unit variance Gaussian noise which is independent of the input signal is used for the additive noise $v(m, n)$. The process $w(m, n)$ is generated by filtering the input signal $u(m, n)$ with the separable denominator 2-D IIR filter:

$$H(q_1^{-1}, q_2^{-1}) = \frac{1 + 0.8q_1^{-1} - 0.5q_2^{-1} - 0.4q_1^{-1}q_2^{-1}}{(1 - 1.2q_1^{-1} + 0.36q_1^{-2})(1 + 0.9q_2^{-1} + 0.2q_2^{-2})}. \quad (3.69)$$

Table 3.1 shows the obtained parameter estimates for $\mu_1 = 0.0012$, $\mu_2 = 0.0008$, and $\alpha_1 = \alpha_2 = 0.5$. Figure 3.4 shows the convergence of the output error $e_o(m, n)$ for both algorithms using the following 1-D error function:

$$\varepsilon(i) = \frac{1}{2(i+1)} \sum_{j=0}^i \{e_o(i, j)^2 + e_o(j, i)^2\}, \quad 0 \leq i \leq M-1. \quad (3.70)$$

In Eq. (3.70), the 2-D output error $e_o(m, n)$ is mapped to the 1-D error $\varepsilon(i)$ by averaging the squared output-error for the pixels that lie on the row segment $\{(i, 0), (i, i)\}$ and the column segment $\{(0, i), (i, i)\}$. Figs. 3.5 and 3.6 show how the value of each of the time varying scaling factors $\tau_k(m, n)$, $k = 1, 2$ changes from zero to the maximum value of one during the adaptive learning process. The

Table 3.1: Parameter estimates for *Example 1* (30 runs).

	$a(1, 0)$	$a(0, 1)$	$a(1, 1)$	$b_1(1)$	$b_1(2)$	$b_2(1)$	$b_2(2)$
True values	0.8	-0.5	-0.4	1.2	-0.36	-0.9	-0.2
Ref. [15]	1.3208	-0.4717	-0.6225	0.6844	0.0980	-0.9174	-0.2346
2DBRA	0.7967	-0.5028	-0.4063	1.2050	-0.3591	-0.8981	-0.1994

2-D values of the scaling factors are presented using the following 1-D functions:

$$\hat{\tau}_k(i) = \frac{1}{2(M-1-i)} \sum_{j=i}^{M-1} \tau_k^2(i, j) + \tau_k^2(j, i), \quad 0 \leq i \leq M-1; \quad k = 1, 2. \quad (3.71)$$

In agreement with the discussion given in previous section, experimental results have shown that the amount of the reduction in the bias caused by the interaction between the cascaded sections depends on the values of the used step size parameters μ_1 and μ_2 . For sufficiently small step size values, significant reduction in the bias caused by both the measurement noise and the cascade interaction can be obtained. This reduction in the bias is achieved, however, at the expense of moderate increase in the computational load of the 2DBRA over the algorithm [15]. This increase is merely due to the requirement of adjusting the time varying scaling factors $\tau_1(m, n)$ and $\tau_2(m, n)$ at each iteration.

Example 2: Noisy desired signal (colored measurement noise)

In this experiment we to compare the performances of the proposed 2DBRA with that of the 2-D LMS algorithm [15] and the family of hyperstable adaptive IIR filtering algorithms presented in [40]. In this experiment the proposed 2DBRA is applied to the system identification experiment described in [40] (Experiment 3 and Experiment 4). The following transfer function is used for the unknown

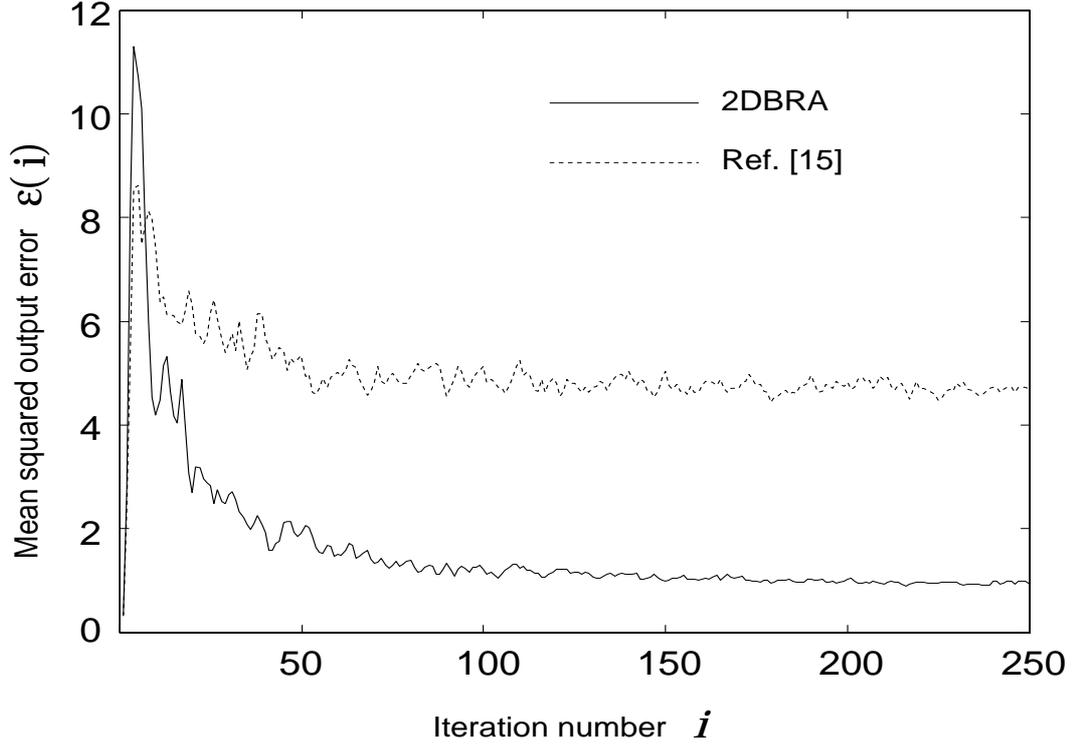
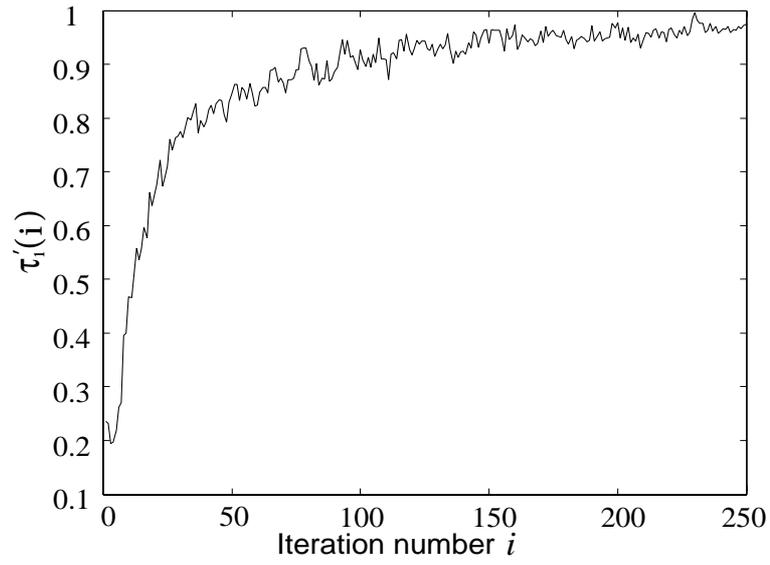
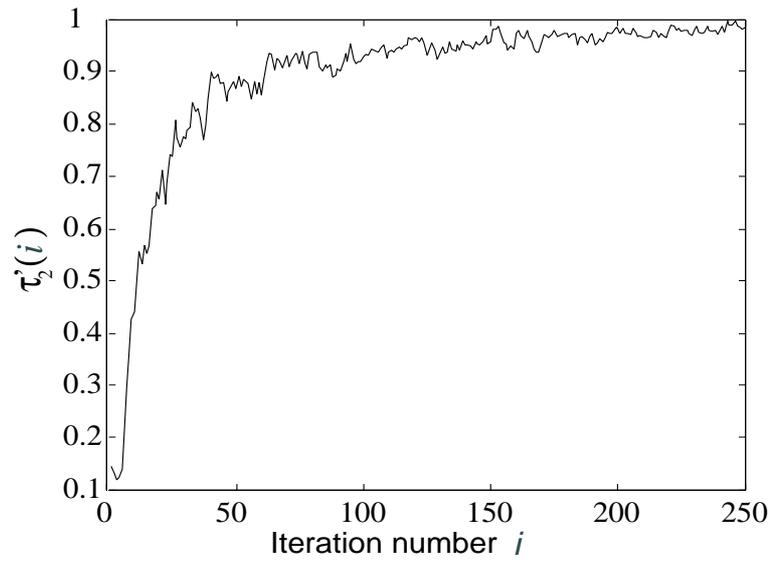


Figure 3.4: The convergence of the mean squared output error of the bias removal algorithm for *Example 1* (average of 30 runs), $\mu_1 = \mu_2 = 0.0004$, $\alpha_1 = \alpha_2 = 0.5$.

system:

$$H(q_1^{-1}, q_2^{-1}) = \frac{1 + q_1^{-1} + 2q_2^{-1} + 3q_1^{-1}q_2^{-1}}{1 - 0.25q_1^{-1} - 0.5q_2^{-1} + 0.125q_1^{-1}q_2^{-1}}. \quad (3.72)$$

A 2-D white Gaussian noise of variance 500.4 and zero mean is used for the input signal $u(m, n)$. And a 320 columns by 200 rows of the “Lena” image with variance 2216.6 and mean value 108.9 is used for the colored measurement noise $v(m, n)$. The variance of the noise $w(m, n)$ is initially 18125.14. The image is processed row by row repeatedly such that the values of the estimated parameters at the end of one pass are used as the parameter initial values at the beginning of the

Figure 3.5: Plot of the scaling factor $\hat{\tau}_1(i)$ Figure 3.6: Plot of the scaling factor $\hat{\tau}_2(i)$

next pass as performed in [40]. The error signal $e_0(m, n) = d(m, n) - y(m, n)$ gives the enhanced image. The variance of the noise left in the processed image is calculated by subtracting the noiseless image from it and then computing the variance.

Table 3.2 shows the obtained parameter estimates for the 2DBRA and the algorithm [15]. Table 3.3 shows the variance of the noise left in the processed image for different passes and for four different algorithms. The results for the 2-D modified HARF and 2-D SHARF are taken from [40] (Table 4 and Table 5 respectively). The image enhancement experiment results for the 2DBRA and the algorithm [15] are shown for different values of the step size parameter $\mu = \mu_1 = \mu_2$.

Figs. 3.7 and 3.8 show the original and the noisy image “Lena” respectively. Figs. 3.9 and 3.10 show the processed images at the 41st pass using algorithm [15] and the 2DBRA respectively.

From these results and other simulation examples, we found that the proposed 2DBRA algorithm performs much better than the algorithm [15] at the expense of very small increase in the computational load. On the other hand, the 2DBRA converges faster than the 2-D modified HARF and the 2-D SHARF. For sufficiently small step size parameters μ_1 and μ_2 the 2DBRA remains stable and no stability monitoring was required.

Table 3.2: Parameter estimates at the 41st pass, $\mu_1 = \mu_2 = 6 \times 10^{-9}$, $\alpha_1 = \alpha_2 = 0.5$.

Parameter	True Value	Ref. [15]	2DBRA
$a(1, 0)$	1	0.9409	0.9812
$a(0, 1)$	2	1.9544	1.9985
$a(1, 1)$	3	2.8432	2.9634
$b_1(1)$	0.25	0.2991	0.2520
$b_2(1)$	0.5	0.5556	0.5081



Figure 3.7: Original image "Lena".

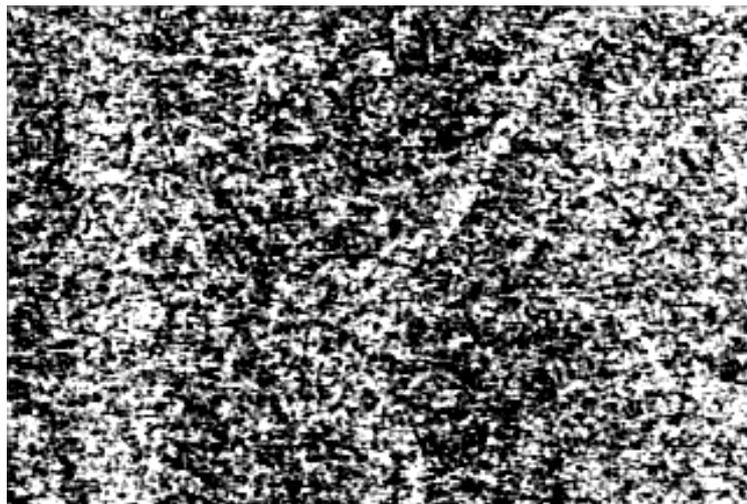


Figure 3.8: Noisy image "Lena".



Figure 3.9: The processed image at the 41st pass using algorithm [15], $\mu_1 = \mu_2 = 6 \times 10^{-9}$.



Figure 3.10: The processed image at the 41st pass using the 2DBRA, $\mu_1 = \mu_2 = 6 \times 10^{-9}$, $\alpha_1 = \alpha_2 = 0.5$.

Table 3.3: The variance of the noise left in the processed images at different passes.

Pass	2-D modified HARF Ref. [40]	2-D SHARF Ref. [40]	$\mu = 6 \times 10^{-9}$		$\mu = 12 \times 10^{-9}$	
			Ref. [15]	2DBRA	Ref. [15]	2DBRA
2	12986.05	12986.35	6499.09	6252.90	3790.08	3724.27
3	9963.90	9957.31	4480.32	4334.35	1934.58	1953.93
4	6864.88	6860.09	3163.56	3117.18	1059.03	1027.20
5	4358.73	4355.70	2273.30	2250.78	628.77	542.73
6	3097.05	3094.77	1660.42	1626.07	412.61	289.54
11	1083.04	1081.76	461.71	325.22	192.57	20.61
21	166.10	165.42	194.94	20.46	182.43	8.48
31	51.61	51.21	183.16	7.91	182.29	8.41
41	32.34	32.09	182.25	7.26	182.29	8.49

Example 3: Noisy desired signal (colored measurement noise)

In this example, we repeated the system identification experiment described in the previous example, however for this experiment we used 2-D white Gaussian noise of zero mean and 104 variance for the input signal $u(m, n)$. And we used the “Mandrill” image of mean 129.1378, variance 1749.8, and size 256 columns by 256 rows, for the measurement noise $v(m, n)$.

Table 3.4 shows the obtained parameter estimates for the 2DBRA and the algorithm [15]. Figure 3.11 shows the improvement in the the Signal to Noise Ratio (SNR) of the enhanced image through successive passes using the algorithm [15] and the 2DBRA. The SNR is calculated by

$$\text{SNR} = 10 \log \left(\frac{\sigma_v^2}{\sigma_n^2} \right) \quad (3.73)$$

where $\sigma_v^2 = 1749.8$ is the variance of the original image “Mandrill” and σ_n^2 is the

Table 3.4: Parameter estimates for *Experiment 3* at the 10th pass.

Parameter	True Value	Ref. [15]	2DBRA
$a(1,0)$	1	0.9158	0.96323
$a(0,1)$	2	1.9814	1.9862
$a(1,1)$	3	2.8038	2.9797
$b_1(1)$	0.25	0.3215	0.2629
$b_2(1)$	0.5	0.5811	0.5135

variance of the noise left in the enhanced image. The SNR value of the initial noisy image is -7.711 . As this figure indicates, for low SNR, the 2DBRA works almost without the bias removal mechanism as the algorithm [15]. As the SNR increases gradually, the effectiveness of the output-error-feedback becomes very clear. Figs. 3.12 and 3.13 show the original and the noisy image “Mandrill” respectively. Figs. 3.14 and 3.15 show the enhanced images at the 10th pass using algorithm [15] and the 2DBRA respectively.

In all the experiments presented here, stability monitoring was not required. It has been observed that, for sufficiently small step size parameters, whenever the poles of the adaptive filters start to immigrate outside the unit circle, the output error of each section increases suddenly and consequently the scaling factors $\tau_1(m, n)$ and $\tau_2(m, n)$ decreases; the adaptive algorithm works without output-error feedback and is able to draw the poles back to the unit circle.

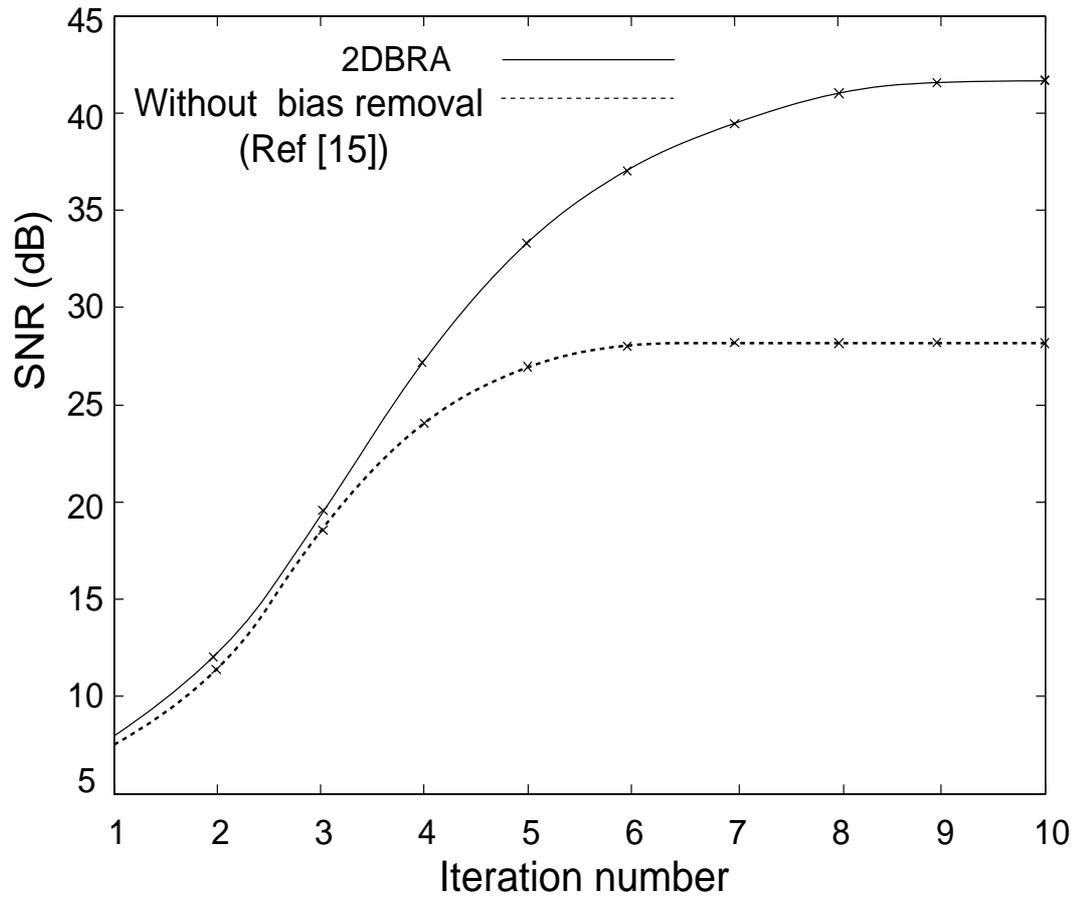


Figure 3.11: The improvement of the SNR of the enhanced image through consecutive passes in *Example 2*.

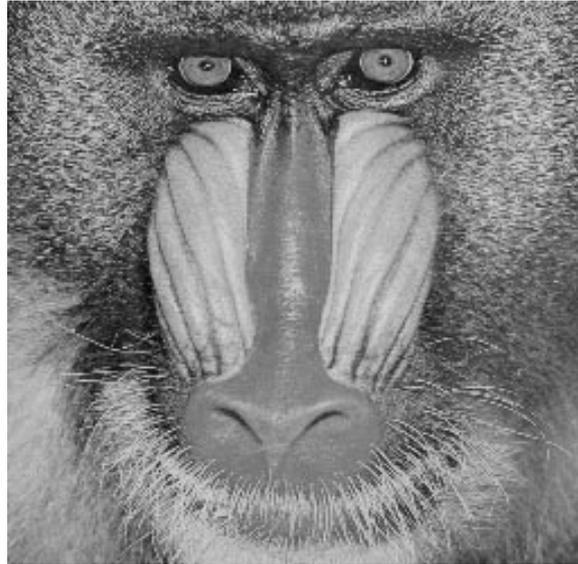


Figure 3.12: The original image “Mandrill”.

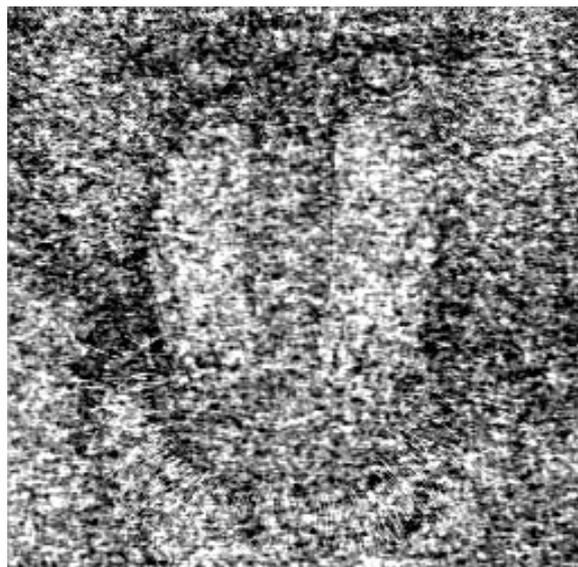


Figure 3.13: The noisy image “Mandrill”.

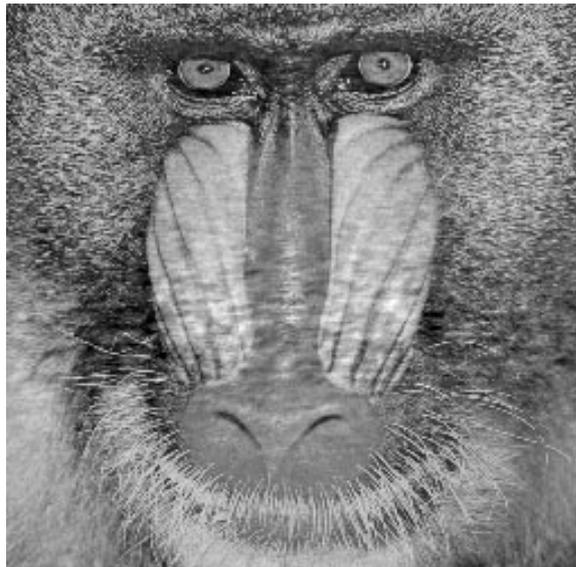


Figure 3.14: The enhanced image at the 10th pass using algorithm [15], $\mu_1 = \mu_2 = 7.6894 \times 10^{-8}$.

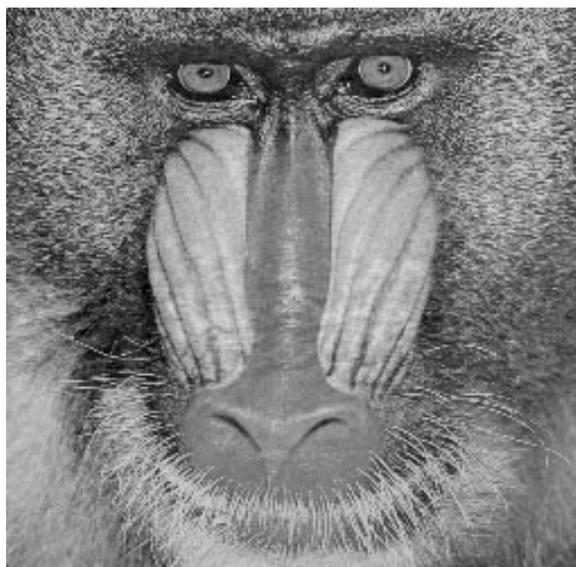


Figure 3.15: The enhanced image at the 10th pass using the 2DBRA, $\mu_1 = \mu_2 = 7.6894 \times 10^{-8}, \alpha_1 = \alpha_2 = 0.7$.

3.6 Summary

The bias removal algorithm for 2-D equation error adaptive IIR filters has been presented. The filter structure in the proposed algorithm is based on the concept of backpropagating the desired signal through a cascade of the denominator vertical and horizontal sections. The key idea in the proposed algorithm is to use a scaled value of the output error of each of the cascaded sections to counteract the effect of the measurement noise embedded in the regressor of the update procedure of that section. I/O stability analysis has been carried out. It has been shown that the proposed algorithm remains stable and the effect of the measurement noise can be significantly suppressed under general conditions imposed on the values of the used step sizes and scaling factors. Image enhancement and 2-D system identification experimental results have been presented to support the effectiveness of the proposed 2DBRA algorithm.

Chapter 4

Steady State Analysis of 2-D Doubly Indexed LMS Adaptive Filters

4.1 Introduction

Two dimensional Least Mean Square (LMS) type adaptive filters has received considerable research interest, mainly because of its simplicity in computation. Hadhoud and Thomas [17] have proposed a 2-D LMS algorithm, which is called TDLMS, by direct extension of the 1-D LMS algorithm [18]. In the TDLMS, the weights' update process is carried out using either vertical or horizontal 1-D indexing scheme. Accordingly, the authors of [17] have shown that the weights' update equation of TDLMS can be written in a form which is mathematically equivalent to the well known 1-D LMS [18]; hence, the analysis procedures and results of the 1-D LMS can be directly applied to the TDLMS. The drawback of the algorithm [17], however, is that it can only exploit the correlation information of the image pixels in the direction of the indexing scheme used to process the 2-D data. To overcome such problem, the authors of [19] have proposed a 2-D

LMS algorithm in which the filter's weights are updated along both the vertical and horizontal directions as a doubly-indexed dynamical system [20]. Such update mechanism enables efficient use of the 2-D correlation information of the image pixels in both vertical and horizontal directions and hence, provides better performance in nonstationary environments [19].

The convergence of the mean for the 2-D LMS [19] (in what follows, it will be referred to as 2-D doubly indexed LMS) has been investigated in [19] using stability theory of 2-D Fornasini and Marchesini (F-M) state space model [20]. Convergence of the mean does not, however, guarantee finite mean square error (MSE) for the adaptive algorithm.

In this chapter, we consider the MSE analysis of the 2-D LMS. The analysis presented here is the first attempt in the literature to investigate the steady state MSE analysis for a doubly-indexed 2-D LMS algorithm. The MSE analysis is carried out using the assumption that the successive input vectors are statistically independent, jointly Gaussian-distributed random variables. This assumption, generally referred to as the *independence assumption* [5], is widely used in the convergence analysis of 1-D LMS for two main reasons. The first is due to the simplification in analysis obtained under such assumption. The second is due to the good agreement between the analytical results obtained using the independence assumption and experimental results [10]-[27].

Though the 2-D MSE analysis will be significantly simplified when invoking the independence assumption, the use of 2-D indexing scheme in the weights' update equation of the 2-D doubly indexed LMS results in a new problem that

is not encountered in the 1-D case. For the 1-D LMS, as well as for the TDLMS, the adaptive filter's weight-vector update equation is a 1-D first order difference equation given by

$$\mathbf{H}_{j+1} = \mathbf{H}_j - \mu \mathbf{G}_j \quad (4.1)$$

where j is the iteration number; \mathbf{H}_j is the adaptive filter's weight-vector; μ is a scalar parameter that controls the convergence rate of the LMS algorithm, and \mathbf{G}_j is the instantaneous gradient of the MSE at iteration j . From Eq. (4.1), it follows that the weight-error covariance matrix is calculated by a set of 1-D first order difference equations. According to [26], [27], this set of difference equations maintains stability under a general condition imposed on the used step size parameter μ . For the 2-D doubly indexed LMS, however, the adaptive filter weight-vector update equation is described by the 2-D first order difference equation

$$\begin{aligned} \mathbf{H}_{m+1,n+1} = & f_h \mathbf{H}_{m,n+1} + f_v \mathbf{H}_{m+1,n} \\ & - \mu_h \mathbf{G}_{m,n+1} - \mu_v \mathbf{G}_{m+1,n} \end{aligned} \quad (4.2)$$

where m and n are two spatial indices in the vertical and horizontal direction respectively. $\mathbf{H}_{m,n}$ is the adaptive filter weight-vector at spatial indices (m, n) ; f_h , f_v , μ_h and μ_v are scalar parameters, and $\mathbf{G}_{m,n}$ is the instantaneous gradient of the MSE at spatial indices (m, n) . From Eq. (4.2), and as will be shown in the sequel, the weight-error covariance matrix for the 2-D doubly indexed LMS is calculated by a set of 2-D second order difference equations. Stability analysis for such set of equations is, however, very difficult to handle mathematically.

In Section 4.3 of this chapter, we show that for the steady state, this set of 2-D second order difference equations can be reduced to a set of linear simultaneous equations in the coefficients of weight-error correlation matrices at different spatial lags; however, the number of the unknowns in this set exceeds the number of equations. To solve this problem, we propose a method for the approximation of the coefficients of weight-error correlation matrices at large spatial lags. The approximation method is based on the extension of the direct averaging method [10] to 2-D case. It can also serve as an approximation method for the weight-error covariance matrix without invoking the independence assumption providing that the step size parameters are sufficiently small.

4.2 The 2-D Doubly Indexed LMS Algorithm

Consider the N by N , causal, 2-D adaptive FIR filter shown in Fig. 4.1. The filter's input $x(m, n)$ is a 2-D stationary signal of size $M_1 \times M_2$. The filter output $y(m, n)$ is calculated by

$$y(m, n) = \mathbf{H}_{m,n}^t \mathbf{X}_{m,n} \quad (4.3)$$

where $\mathbf{H}_{m,n}$ and $\mathbf{X}_{m,n}$ are respectively the adaptive filter's weight-vector and the input data vector given at spatial indices (m, n) by

$$\begin{aligned} \mathbf{X}_{m,n} &= [x(m, n), \dots, x(m - N + 1, n), \dots, x(m - N + 1, n - N + 1)]^t \\ \mathbf{H}_{m,n} &= [h_{m,n}(0, 0), \dots, h_{m,n}(0, N - 1), \dots, h_{m,n}(N - 1, N - 1)]^t. \end{aligned} \quad (4.4)$$

The 2-D doubly indexed LMS updates the filter's weight-vector along both the vertical and horizontal directions such that the error between the filter output

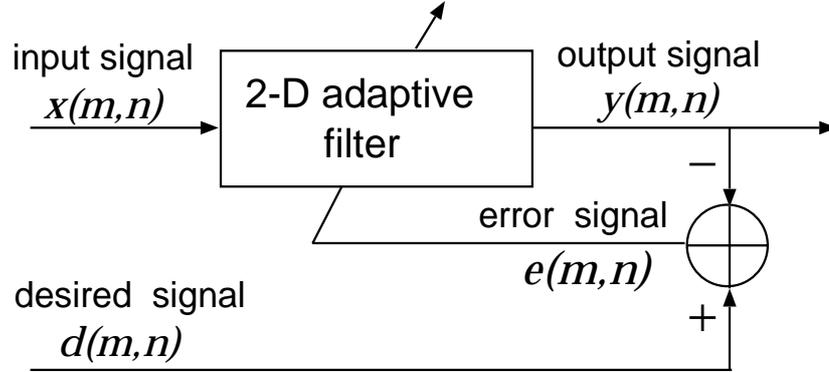


Figure 4.1: 2-D adaptive filter.

$y(m, n)$ and the desired signal $d(m, n)$ is minimized in the MSE sense. The MSE is defined as

$$\begin{aligned} \text{MSE} &= E\{e^2(m, n)\} \\ &= E\{(d(m, n) - \mathbf{H}_{m,n}^t \mathbf{X}_{m,n})^2\}. \end{aligned} \quad (4.5)$$

The update equation for the 2-D doubly indexed LMS is given by

$$\begin{aligned} \mathbf{H}_{m+1,n+1} &= f_h \mathbf{H}_{m,n+1} + f_v \mathbf{H}_{m+1,n} \\ &\quad + \mu_h e(m, n+1) \mathbf{X}_{m,n+1} + \mu_v e(m+1, n) \mathbf{X}_{m+1,n}; \\ \mathbf{H}_{m,0} &= \mathbf{0}, m = 0 \cdots M_1; \mathbf{H}_{0,n} = \mathbf{0}, n = 0 \cdots M_2; \\ f_h + f_v &= 1 \end{aligned} \quad (4.6)$$

where μ_h and μ_v denote the step size parameters in the horizontal and vertical directions respectively.

The optimal solution \mathbf{H}_{opt} that minimizes the MSE is given by the Wiener-

Hopf equation

$$\mathbf{H}_{opt} = \mathbf{R}^{-1}\mathbf{P} \quad (4.7)$$

where

$$\begin{aligned} \mathbf{R} &= \mathbf{E}\{\mathbf{X}_{m,n}\mathbf{X}_{m,n}^t\} \\ \mathbf{P} &= \mathbf{E}\{\mathbf{X}_{m,n}d(m,n)\}. \end{aligned} \quad (4.8)$$

In [19], it has been shown that the 2-D doubly indexed LMS converges to the optimal solution i.e., $\mathbf{E}\{\mathbf{H}_{m+1,n+1} - \mathbf{H}_{opt}\} \rightarrow 0$ as $m + n \rightarrow \infty$, if the following condition holds:

$$|f_h - \mu_h \lambda_i| + |f_v - \mu_v \lambda_i| < 1 \quad (4.9)$$

where $\lambda_i, i = 0, \dots, N^2 - 1$, are the eigenvalues of the input correlation matrix \mathbf{R} .

Condition (4.9) is, however, not sufficient to guarantee convergence of the 2-D doubly indexed LMS in the MSE sense. Moreover, convergence of the mean does not provide any information about the performance of the adaptive algorithm. In the following section we will present the steady state MSE analysis for the 2-D doubly indexed LMS.

4.3 Steady State Mean Square Error Analysis using the Independence Assumption

The MSE analysis will be carried out using the independence assumption [10], consisting of the following points:

A.1 The input vectors $\mathbf{X}_{0,0}, \mathbf{X}_{1,0}, \dots, \mathbf{X}_{m,n}$ are zero mean, statistically independent, Gaussian-distributed random variables.

A.2 The error

$$\varepsilon(m, n) = d(m, n) - \mathbf{H}_{opt}^t \mathbf{X}_{m,n} \quad (4.10)$$

is a zero mean, white Gaussian noise of variance σ_ε^2 , and is statistically independent of the input vector $\mathbf{X}_{m,n}$.

4.3.1 MSE Calculation

Let us define the adaptive filter weight-error vector

$$\mathbf{C}_{m,n} = \mathbf{H}_{m,n} - \mathbf{H}_{opt}. \quad (4.11)$$

Then, using Eqs. (4.10) and (4.11), the error signal $e(m, n)$ can be given by

$$\begin{aligned} e(m, n) &= d(m, n) - \mathbf{H}_{m,n}^t \mathbf{X}_{m,n} \\ &= \varepsilon(m, n) + \mathbf{H}_{opt}^t \mathbf{X}_{m,n} - \mathbf{H}_{m,n}^t \mathbf{X}_{m,n} \\ &= \varepsilon(m, n) - \mathbf{C}_{m,n}^t \mathbf{X}_{m,n}. \end{aligned} \quad (4.12)$$

Now if we substitute Eq. (4.12) in (4.5) and make use of assumptions A.1 and A.2, we can find that the steady state MSE is given by

$$\begin{aligned} \epsilon_\infty &= \lim_{m+n \rightarrow \infty} E\{e(m, n)^2\} \\ &= \sigma_\varepsilon^2 + \lim_{m+n \rightarrow \infty} E\{\mathbf{C}_{m,n}^t \mathbf{X}_{m,n} \mathbf{X}_{m,n}^t \mathbf{C}_{m,n}\} \end{aligned} \quad (4.13)$$

$$= \sigma_\varepsilon^2 + \lim_{m+n \rightarrow \infty} \text{tr}(\mathbf{R} \mathbf{K}_{m,n;m,n}) \quad (4.14)$$

where

$$\mathbf{K}_{m,n;m,n} = \mathbb{E}\{\mathbf{C}_{m,n}\mathbf{C}_{m,n}^t\} \quad (4.15)$$

is the weight-error covariance matrix.

Note that from A.1 it follows that the input vector $\mathbf{X}_{m,n}$ and the weight-error vector $\mathbf{C}_{m,n}$ are statistically independent. Accordingly, the expectation term in Eq. (4.13) can be treated as a product of two expectation terms. Strictly speaking, in adaptive filtering applications, these two vectors are dependent since the successive input vectors are statistically dependent. However, even when this statistical dependency is ignored, the independence assumption still preserve the correlation structure for $\mathbb{E}\{\mathbf{X}_{m,n}\mathbf{X}_{m,n}^t\}$ as well as for $\mathbb{E}\{\mathbf{C}_{m,n}^t\mathbf{C}_{m,n}\}$. Hence, the analysis under such assumption still retains enough information about the behaviour of the adaptive process even when the input signal is correlated, (see [10], [5] and references therein).

In the rest of this section we will consider the calculation of the weight-error covariance matrix. In this calculation, we assume that the condition (4.9), which is necessary for the convergence of the mean, holds.

4.3.2 Weight-Error Covariance Matrix

To calculate the weight-error covariance matrix we need first to derive the update equation for the weight-error vector. Indeed, if we subtract \mathbf{H}_{opt} from both sides

of Eq. (4.6) and make use of Eq. (4.12), we get

$$\begin{aligned}
 \mathbf{C}_{m+1,n+1} &= \mathbf{H}_{m+1,n+1} - \mathbf{H}_{opt} \\
 &= (f_h \mathbf{I} - \mu_h \mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t) \mathbf{C}_{m,n+1} \\
 &\quad + (f_v \mathbf{I} - \mu_v \mathbf{X}_{m+1,n} \mathbf{X}_{m+1,n}^t) \mathbf{C}_{m+1,n} \\
 &\quad + \mu_h \varepsilon(m, n+1) \mathbf{X}_{m,n+1} \\
 &\quad + \mu_v \varepsilon(m+1, n) \mathbf{X}_{m+1,n}.
 \end{aligned} \tag{4.16}$$

If Eq. (4.9) holds, the converges of the the mean of weight-error vector is guaranteed. That is:

$$\tilde{m} = E\{\mathbf{C}_{m,n}\} \rightarrow 0 \text{ as } m+n \rightarrow \infty. \tag{4.17}$$

The variance of $\mathbf{C}_{m+1,n+1}$ is defined as

$$\begin{aligned}
 \text{var}\{\mathbf{C}_{m+1,n+1}\} &= E\{\mathbf{C}_{m+1,n+1} \mathbf{C}_{m+1,n+1}^t\} - \tilde{m} \tilde{m}^t \\
 &= E\{\mathbf{C}_{m+1,n+1} \mathbf{C}_{m+1,n+1}^t\}.
 \end{aligned} \tag{4.18}$$

Multiplying each side of Eq. (4.16) by its transpose we obtain:

$$\mathbf{C}_{m+1,n+1} \mathbf{C}_{m+1,n+1}^t = T_1 + T_2 + T_3 + T_4 + T_5 + T_5 + T_6 \tag{4.19}$$

where t denotes the matrix transpose operator. And, for notational convenience, we have defined

$$\begin{aligned}
 T_1 &= (f_h I - \mu_h \mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t) \mathbf{C}_{m,n+1} \mathbf{C}_{m,n+1}^t (f_h I - \mu_h \mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t) \\
 T_2 &= (f_v I - \mu_v \mathbf{X}_{m+1,n} \mathbf{X}_{m+1,n}^t) \mathbf{C}_{m+1,n} \mathbf{C}_{m,n+1}^t (f_h I - \mu_h \mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t)^t \\
 T_3 &= (f_h I - \mu_h \mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t) \mathbf{C}_{m,n+1} \mathbf{C}_{m+1,n}^t (f_v I - \mu_v \mathbf{X}_{m+1,n} \mathbf{X}_{m+1,n}^t)^t \\
 T_4 &= (f_v I - \mu_v \mathbf{X}_{m+1,n} \mathbf{X}_{m+1,n}^t) \mathbf{C}_{m+1,n} \mathbf{C}_{m+1,n}^t (f_v I - \mu_v \mathbf{X}_{m+1,n} \mathbf{X}_{m+1,n}^t)^t \\
 T_5 &= ((f_h I - \mu_h \mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t) \mathbf{C}_{m,n+1} + (f_v I - \mu_v \mathbf{X}_{m+1,n} \mathbf{X}_{m+1,n}^t) \mathbf{C}_{m+1,n}) \\
 &\quad (\mu_h \mathbf{X}_{m,n+1} \varepsilon_{m,n+1} + \mu_v \mathbf{X}_{m+1,n} \varepsilon_{m+1,n})^t \\
 T_6 &= (\mu_h \mathbf{X}_{m,n+1} \varepsilon_{m,n+1} + \mu_v \mathbf{X}_{m+1,n} \varepsilon_{m+1,n}) (\mu_h \mathbf{X}_{m,n+1} \varepsilon_{m,n+1} + \mu_v \mathbf{X}_{m+1,n} \varepsilon_{m+1,n})^t
 \end{aligned} \tag{4.20}$$

Calculating the expectation values of each term of Eq. (4.20) we can write:

$$\begin{aligned}
 \mathbb{E}\{T_1\} &= f_h^2 \mathbb{E}\{\mathbf{C}_{m,n+1} \mathbf{C}_{m,n+1}^t\} - \mu_h f_h \mathbb{E}\{\mathbf{C}_{m,n+1} \mathbf{C}_{m,n+1}^t\} \mathbf{R} \\
 &\quad - \mu_h f_h \mathbf{R} \mathbb{E}\{\mathbf{C}_{m,n+1} \mathbf{C}_{m,n+1}^t\} + 2\mu_h^2 \mathbf{R} \mathbb{E}\{\mathbf{C}_{m,n+1} \mathbf{C}_{m,n+1}^t\} \mathbf{R} \\
 &\quad + \mu_h^2 \text{tr}(\mathbb{E}\{\mathbf{C}_{m,n+1} \mathbf{C}_{m,n+1}^t\} \mathbf{R}) \mathbf{R}.
 \end{aligned} \tag{4.21}$$

$$\begin{aligned}
 \mathbb{E}\{T_2\} &= \mathbb{E}\{(f_v I - \mu_v \mathbf{X}_{m+1,n} \mathbf{X}_{m+1,n}^t) \mathbf{C}_{m+1,n} \mathbf{C}_{m,n+1}^t (f_h I - \mu_h \mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t)^t\} \\
 &= f_h f_v \mathbb{E}\{\mathbf{C}_{m+1,n} \mathbf{C}_{m,n+1}^t\} - \mu_h f_v \mathbb{E}\{\mathbf{C}_{m+1,n} \mathbf{C}_{m,n+1}^t\} \mathbf{R} \\
 &\quad - \mu_v f_h \mathbf{R} \mathbb{E}\{\mathbf{C}_{m+1,n} \mathbf{C}_{m,n+1}^t\} + \mu_h \mu_v \mathbf{R} \mathbb{E}\{\mathbf{C}_{m+1,n} \mathbf{C}_{m,n+1}^t\} \mathbf{R}.
 \end{aligned} \tag{4.22}$$

$$\begin{aligned}
 E\{T_3\} &= f_h f_v E\{\mathbf{C}_{m,n+1} \mathbf{C}_{m+1,n}^t\} - \mu_v f_h E\{\mathbf{C}_{m,n+1} \mathbf{C}_{m+1,n}^t\} \mathbf{R} \\
 &\quad - \mu_h f_v \mathbf{R} E\{\mathbf{C}_{m,n+1} \mathbf{C}_{m+1,n}^t\} + \mu_h \mu_v \mathbf{R} E\{\mathbf{C}_{m,n+1} \mathbf{C}_{m+1,n}^t\} \mathbf{R} \quad (4.23)
 \end{aligned}$$

$$\begin{aligned}
 E\{T_4\} &= E\{(f_v I - \mu_v \mathbf{X}_{m+1,n} \mathbf{X}_{m+1,n}^t) \mathbf{C}_{m+1,n} \mathbf{C}_{m,n+1}^t (f_h I - \mu_h \mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t)\} \\
 &= f_h^2 E\{\mathbf{C}_{m+1,n} \mathbf{C}_{m+1,n}^t\} - \mu_v f_v E\{\mathbf{C}_{m+1,n} \mathbf{C}_{m+1,n}^t\} \mathbf{R} \\
 &\quad - \mu_h f_v \mathbf{R} E\{\mathbf{C}_{m+1,n} \mathbf{C}_{m+1,n}^t\} + 2\mu_h^2 \mathbf{R} E\{\mathbf{C}_{m+1,n} \mathbf{C}_{m+1,n}^t\} \mathbf{R} \\
 &\quad + \mu_h^2 \text{tr} (E\{\mathbf{C}_{m+1,n} \mathbf{C}_{m+1,n}^t\} \mathbf{R}) \mathbf{R}. \quad (4.24)
 \end{aligned}$$

$$E\{T_5\} = 0 \quad (4.25)$$

$$E\{T_6\} = (\mu_h^2 + \mu_v^2) \sigma_\varepsilon^2 \mathbf{R}. \quad (4.26)$$

Note that in the calculation of Eqs. 4.21-4.26 we have made use of the independence assumptions and the following property of zero mean Gaussian random variables [26]:

$$\begin{aligned}
 E\{\mathbf{X}_{m,n+1} \mathbf{X}_{m,n+1}^t (X_{m,n+1}^t \mathbf{C}_{m,n+1})^2\} &= 2\mathbf{R} E\{\mathbf{C}_{m,n+1} \mathbf{C}_{m,n+1}^t\} \mathbf{R} + \\
 &\quad \text{tr} (E\{\mathbf{C}_{m,n+1} \mathbf{C}_{m,n+1}^t\} \mathbf{R}) \mathbf{R}. \quad (4.27)
 \end{aligned}$$

Now before proceeding, we need to define some necessary notations. Since the input correlation matrix \mathbf{R} is symmetric, there exists an orthogonal matrix \mathbf{Q} such that

$$\begin{aligned}
 \mathbf{Q} \mathbf{R} \mathbf{Q}^t &= \mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N^2-1}) \\
 \mathbf{Q}^t &= \mathbf{Q}^{-1}. \quad (4.28)
 \end{aligned}$$

Thus, we can define the transformed matrix:

$$\begin{aligned}
 \mathbf{\Gamma}_{m_1, n_1; m_2, n_2} &= \mathbf{Q} \mathbf{E} \{ \mathbf{C}_{m_1, n_1} \mathbf{C}_{m_2, n_2}^t \} \mathbf{Q}^t \\
 &= [\gamma_{m_1, n_1; m_2, n_2}^{i, j}]; \\
 & \quad i, j = 0, \dots, N^2 - 1
 \end{aligned} \tag{4.29}$$

where the superscripts (i, j) in the notation $\gamma_{m_1, n_1; m_2, n_2}^{i, j}$ is used to point to the element at the i th row and j th column of the matrix $\mathbf{\Gamma}_{m_1, n_1; m_2, n_2}$.

Now, substituting Eqs. (4.21)-(4.26) in Eq. (4.18) and making use of the orthogonal transform \mathbf{Q} we arrive at

$$\begin{aligned}
 \mathbf{\Gamma}_{m+1, n+1; m+1, n+1} &= f_h^2 \mathbf{\Gamma}_{m, n+1; m, n+1} \\
 & \quad - \mu_h f_h \mathbf{\Gamma}_{m, n+1; m, n+1} \mathbf{\Lambda} - \mu_h f_h \mathbf{\Lambda} \mathbf{\Gamma}_{m, n+1; m, n+1} \\
 & \quad + 2\mu_h^2 \mathbf{\Lambda} \mathbf{\Gamma}_{m, n+1; m, n+1} \mathbf{\Lambda} + \mu_h^2 \text{tr}(\mathbf{\Gamma}_{m, n+1; m, n+1} \mathbf{\Lambda}) \mathbf{\Lambda} \\
 & \quad + f_h f_v \mathbf{\Gamma}_{m+1, n; m, n+1} - \mu_h f_v \mathbf{\Gamma}_{m+1, n; m, n+1} \mathbf{\Lambda} \\
 & \quad - \mu_v f_h \mathbf{\Lambda} \mathbf{\Gamma}_{m+1, n; m, n+1} + \mu_h \mu_v \mathbf{\Lambda} \mathbf{\Gamma}_{m+1, n; m, n+1} \mathbf{\Lambda} \\
 & \quad + f_h f_v \mathbf{\Gamma}_{m, n+1; m+1, n} - \mu_v f_h \mathbf{\Gamma}_{m, n+1; m+1, n} \mathbf{\Lambda} \\
 & \quad - \mu_h f_v \mathbf{\Lambda} \mathbf{\Gamma}_{m, n+1; m+1, n} + \mu_h \mu_v \mathbf{\Lambda} \mathbf{\Gamma}_{m, n+1; m+1, n} \mathbf{\Lambda} \\
 & \quad + f_v^2 \mathbf{\Gamma}_{m+1, n; m+1, n} - \mu_v f_v \mathbf{\Gamma}_{m+1, n; m+1, n} \mathbf{\Lambda} \\
 & \quad - \mu_v f_v \mathbf{\Lambda} \mathbf{\Gamma}_{m+1, n; m+1, n} + 2\mu_v^2 \mathbf{\Lambda} \mathbf{\Gamma}_{m+1, n; m+1, n} \mathbf{\Lambda} \\
 & \quad + \mu_v^2 \text{tr}(\mathbf{\Gamma}_{m+1, n; m+1, n} \mathbf{\Lambda}) \mathbf{\Lambda} + (\mu_h^2 + \mu_v^2) \sigma_\varepsilon^2 \mathbf{\Lambda}.
 \end{aligned} \tag{4.30}$$

Analysing the stability of the set of second-order coupled 2-D difference equations (4.30) is a very complicated task. Thus, we propose to simplify the analysis by making use of the following two facts.

1. For the transformed weight-error correlation matrix defined in Eq. (4.29), it follows from Schwartz' inequality [42] that

$$(\gamma_{m_1, n_1; m_2, n_2}^{i, j})^2 \leq \gamma_{m_1, n_1; m_1, n_1}^{i, i} \cdot \gamma_{m_2, n_2; m_2, n_2}^{j, j}. \quad (4.31)$$

That is to say, the boundedness of the diagonal terms of the weight-error correlation matrices ensures the boundedness of the off-diagonal ones. Hence, it is sufficient to analyze the stability of the diagonal terms of the matrix equation (4.30). Note that, as for the MSE evaluation (see Eq. (4.14)), we are only interested in the diagonal terms since

$$\begin{aligned} \text{tr}(\mathbf{RK}_{m, n; m, n}) &= \text{tr}(\mathbf{\Lambda}\mathbf{\Gamma}_{m, n; m, n}) \\ &= \sum_{j=0}^{N^2-1} \gamma_{m, n; m, n}^{j, j} \lambda_j. \end{aligned} \quad (4.32)$$

2. Let, for notational convenience, $\gamma_k^{i, i}, k = 0, 1, \dots$, denote the steady state values of the weight error correlation coefficients at spatial lag $(k, -k)$. That is

$$\gamma_k^{i, i} = \lim_{m+n \rightarrow \infty} \gamma_{m+1, n+1-k; m+1-k, n+1}^{i, i}. \quad (4.33)$$

Now, if the adaptive algorithm reaches the steady state, the signal $\mathbf{C}_{m, n}$ becomes stationary random signal. Consequently, if the weight-error covariance coefficient $\gamma_{m+1, n+1; m+1, n+1}^{i, i}, i = 0, \dots, N^2 - 1$, has a steady state value, say $\gamma_0^{i, i}$, then the following equality should holds:

$$\begin{aligned} \lim_{m+n \rightarrow \infty} \gamma_{m+1, n; m+1, n}^{i, i} &= \lim_{m+n \rightarrow \infty} \gamma_{m, n+1; m, n+1}^{i, i} \\ &= \gamma_0^{i, i}. \end{aligned} \quad (4.34)$$

Similarly, if the weight-error correlation coefficient $\gamma_{m+1,n;m,n+1}^{i,i}$, $i = 0, \dots, N^2 - 1$, has a steady state value, say $\gamma_1^{i,i}$, then the following equality should hold:

$$\begin{aligned} \lim_{m+n \rightarrow \infty} \gamma_{m+1,n;m,n+1}^{i,i} &= \lim_{m+n \rightarrow \infty} \gamma_{m,n+1;m+1,n}^{i,i} \\ &= \gamma_1^{i,i}. \end{aligned} \quad (4.35)$$

Consequently, for the steady state, the N^2 diagonal coefficients of the Eq. (4.30) should obey the equality

$$\begin{aligned} \gamma_0^{i,i} &= (f_h^2 + f_v^2 - 2(\mu_h f_h + \mu_v f_v)\lambda_i + 2(\mu_h^2 + \mu_v^2)\lambda_i^2) \gamma_0^{i,i} \\ &\quad + 2(f_h f_v - (\mu_h f_v + \mu_v f_h)\lambda_i + \mu_h \mu_v \lambda_i^2) \gamma_1^{i,i} \\ &\quad + (\mu_h^2 + \mu_v^2)\lambda_i \sum_{j=0}^{N^2-1} \gamma_0^{j,j} \lambda_j + (\mu_h^2 + \mu_v^2)\sigma_\varepsilon^2 \lambda_i. \end{aligned} \quad (4.36)$$

There is a need for another set of equations in the unknowns $\gamma_0^{i,i}$ and $\gamma_1^{i,i}$.

If we apply the same way of analysis to evaluate the matrix

$$\lim_{m+n \rightarrow \infty} \mathbf{\Gamma}_{m+1,n;m,n+1} = [\gamma_1^{i,j}], \quad i, j = 0, \dots, N^2 - 1$$

we can find that for the steady state, i.e. $m + n \rightarrow \infty$, the diagonal terms of the correlation matrix $\mathbf{\Gamma}_{m+1,n;m,n+1}$ should obey the equality

$$\begin{aligned} \gamma_1^{i,i} &= (f_h f_v - (\mu_h f_v + \mu_v f_h)\lambda_i + 2\mu_h \mu_v \lambda_i^2) \gamma_0^{i,i} \\ &\quad + (f_h^2 + f_v^2 - 2(\mu_h f_h + \mu_v f_v)\lambda_i + (\mu_h^2 + \mu_v^2)\lambda_i^2) \gamma_1^{i,i} \\ &\quad + (f_h f_v - (\mu_h f_v + \mu_v f_h)\lambda_i + \mu_h \mu_v \lambda_i^2) \gamma_2^{i,i} \\ &\quad + \mu_h \mu_v \lambda_i \sum_{j=0}^{N^2-1} \gamma_0^{j,j} \lambda_j + \mu_h \mu_v \lambda_i \sigma_\varepsilon^2 \end{aligned} \quad (4.37)$$

where

$$\gamma_2^{i,i} = \lim_{m+n \rightarrow \infty} \gamma_{m+1,n-1;m-1,n+1}^{i,i}. \quad (4.38)$$

If we continue in similar way evaluating the weight-error correlation matrices for higher spatial lags, i. e.

$$\lim_{m+n \rightarrow \infty} \mathbf{\Gamma}_{m+1,n+1-k;m+1-k,n+1} = [\gamma_k^{i,j}]; \quad k = 2, 3, \dots,$$

at each stage $k \in 0, 1, \dots$, we will have a set of $(k+1) \times N^2$ equations in $(k+2) \times N^2$ unknowns, namely, $\gamma_j^{i,i}; 0 \leq j \leq k+1, 0 \leq i \leq N^2 - 1$. To solve this problem, we propose two methods. The first method, presented in Section 4.5, makes use of the direct averaging method [10]. It approximates the stochastic difference Eq. (4.16) of the weight-error vector with a simpler time-invariant averaged system. The proposed direct averaging-based analysis can be used to derive an approximation of the weight-error correlation matrix $\mathbf{\Gamma}_{m,n-k;m-k,n}$ for an arbitrary integer k without invoking the independence assumption given by A.1 and A.2. For example, we can use this method to obtain an approximation for the weight error correlation coefficients, $\gamma_2^{i,i} = \lim_{m+n \rightarrow \infty} \gamma_{m+1,n-1;m-1,n+1}^{i,i}, i = 1, \dots, N^2$; using this approximation in Eq. (4.37), the two sets of equations (4.36) and (4.37) can then be solved for $\gamma_0^{i,i}$ and $\gamma_1^{i,i}$.

The second alternative method is to state that, under the white Gaussian assumption for the input vector $\mathbf{X}_{m,n}$ and the error signal $\varepsilon(m,n)$, the weight-error correlation coefficients $\gamma_{k+1}^{i,i}$ for $k \gg 1$ can be approximated with zero. Thus, the available $(k+1) \times N^2$ equations can be solved for the $(k+1) \times N^2$ unknowns to obtain the weight-error covariance coefficients $\gamma_0^{i,i}, i = 0, \dots, N^2 - 1$. The solution

of these $(k+1) \times N^2$ simultaneous equations can be obtained using mathematical tool box for the general case. In the Section 4.5 we show that the error that results from setting the weight-error correlation coefficients $\gamma_{k+1}^{i,i}$, $k \gg 1$ decreases as the spatial lag k increases.

In the following section we will discuss in more details the steady state analysis for the simple case when the input signal is white Gaussian noise.

4.4 Steady State MSE Analysis with White Gaussian Input Data

In this section we deal with the steady state analysis for the case when the input signal is white Gaussian noise with variance σ_x^2 ; the correlation coefficients $\gamma_2^{i,i}$, $i = 0, \dots, N^2 - 1$ are set to zero; $f_h = f_v$, and $\mu_h = \mu_v = \mu$. We choose to work with this case merely to make the solution of the equations traceable. Similar kind of analysis can be applied to any other case within which A.1 and A.2 hold.

For the white Gaussian input case, $\lambda_0 = \lambda_1, \dots, \lambda_{N^2-1} = \sigma_x^2$. Accordingly, $\gamma_0^{0,0} = \gamma_0^{1,1} = \gamma_0^{i,i} = \gamma_0$, $i = 0, \dots, N^2 - 1$. Hence, solving Eqs. (4.36) and (4.37) for γ_0 we get

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{\sigma_x^2} \frac{\zeta^2}{(0.25 - \zeta + (2+p)\zeta^2)} \times \frac{0.375 - \zeta + (0.5 + 1.5p)\zeta^2 + (6 + 2p)\zeta^3 - (4 + 2p)\zeta^4}{0.125 + 3\zeta - (2.5 + 1.5p)\zeta^2 - (6 + 2p)\zeta^3 + (4 + 2p)\zeta^4} \quad (4.39)$$

where, for notational convenience, we have defined $p = N^2$, and

$$\zeta = \mu\sigma_x^2. \quad (4.40)$$

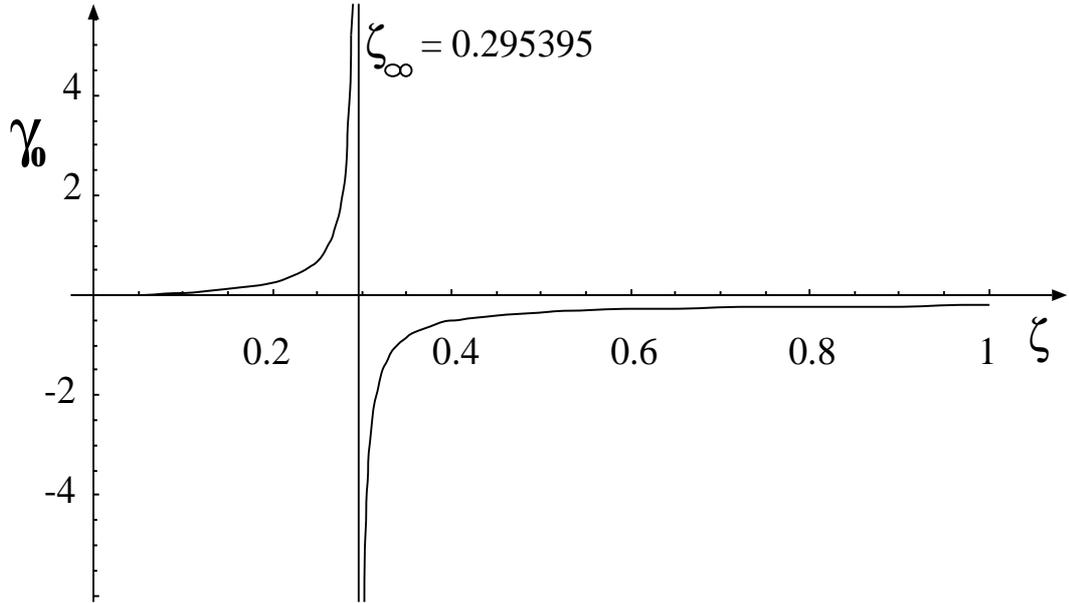
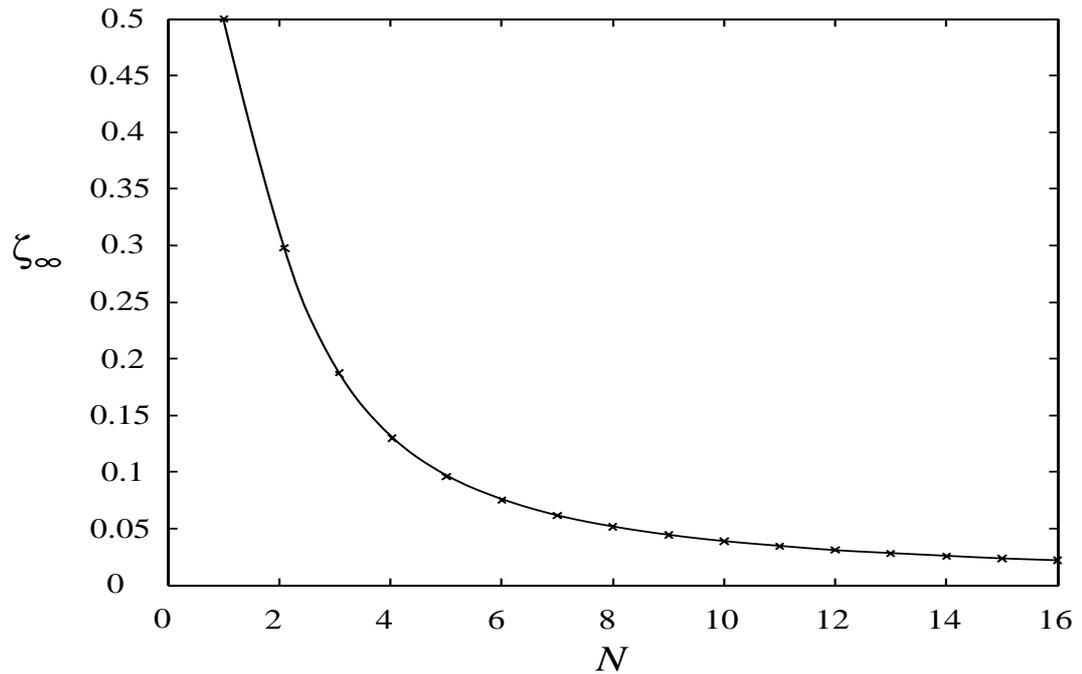


Figure 4.2: Weight-error covariance coefficient γ_0 as a function of $\zeta = \mu\sigma_x^2$ for 2 by 2 adaptive FIR filter.

Now, since the weight-error covariance coefficient $\gamma_0^{i,i}, i = 0, 1, \dots, N^2 - 1$ should be positive and finite, the range of the step size μ that ensures the convergence of the 2-D doubly indexed LMS in the MSE sense can be determined by the following condition

$$0 \leq \gamma_0^{i,i} < \infty, \quad i = 0, 1, \dots, N^2 - 1. \quad (4.41)$$

For this simplified case, analysis of Eq. (4.39) reveals that in this equation, the first term and the numerator of the second term are always positive for $0 \leq \zeta < 1$, and that for any value of $N^2 \geq 1$, the polynomial in the denominator of the second term has only one real positive root, say ζ_{∞} , in the range $0 \leq \zeta < 1$ where the

Figure 4.3: The root value ζ_∞ versus N

sign of this polynomial changes from positive to negative. Thus, we can deduce that the upper bound of the step size value that ensures finite variance is given by

$$0 \leq \mu < \frac{\zeta_\infty}{\sigma_x^2}. \quad (4.42)$$

Fig. 4.2 shows γ_0 as a function of ζ with $\sigma_\varepsilon^2/\sigma_x^2 = 1$ for a 2 by 2, 2-D adaptive FIR filter. Fig. 4.3 shows the values of the root ζ_∞ for different values of N . From Fig. 4.3, it is clear that for any filter order, $\zeta_\infty < 1$. Accordingly, the condition required for the convergence in the MSE sense, as given in Eq. (4.42), decreases significantly the convergence region of the 2-D doubly indexed LMS algorithm

when comparing to the condition necessary for the convergence of the mean

$$0 \leq \mu < \frac{1}{\sigma_x^2} \quad (4.43)$$

given by Eq. (4.9).

4.5 Direct Averaging Method for the Approximation of the Weight-Error Correlation Matrix

Providing that the step sizes μ_h and μ_v are small, and based on the direct averaging method [10], the solution of the stochastic difference Eq. (4.16) can be approximated with that of the following averaged system:

$$\mathbf{C}_{m+1,n+1} = \mathbf{A}_h \mathbf{C}_{m,n+1} + \mathbf{A}_v \mathbf{C}_{m+1,n} + \mu_h \varepsilon_{m,n+1} \mathbf{X}_{m,n+1} + \mu_v \varepsilon_{m+1,n} \mathbf{X}_{m+1,n} \quad (4.44)$$

where

$$\mathbf{A}_h = f_h \mathbf{I} - \mu_h \mathbf{R}$$

$$\mathbf{A}_v = f_v \mathbf{I} - \mu_v \mathbf{R}.$$

Eq. (4.44) is a 2-D F-M state space model with local state space vector $\mathbf{C}_{m,n}$ and input vector $\varepsilon_{m,n} \mathbf{X}_{m,n}$. This 2-D F-M model is exponentially stable if and only if [20]

$$\det(\mathbf{I} - z_1^{-1} \mathbf{A}_h - z_2^{-1} \mathbf{A}_v) \neq 0 \quad (4.45)$$

in the region

$$U_2 = \{(z_1, z_2) \mid |z_1| \geq 1, |z_2| \geq 1\}.$$

Note that the condition (4.45) is the same condition required for the convergence of the mean which was reduced in [19] to the condition (4.9).

Now, the transfer function between the input $\varepsilon_{m,n}\mathbf{X}_{m,n}$ and the state space vector $\mathbf{C}_{m,n}$ is given by

$$\begin{aligned} H(z_1, z_2) &= (I - \mathbf{A}_h z_1^{-1} - \mathbf{A}_v z_2^{-1})^{-1} (\mu_h z_1^{-1} + \mu_v z_2^{-1}) \\ &= (\mu_h z_1^{-1} + \mu_v z_2^{-1}) \sum_{k=0}^{\infty} (\mathbf{A}_h z_1^{-1} + \mathbf{A}_v z_2^{-1})^k \\ &= (\mu_h z_1^{-1} + \mu_v z_2^{-1}) \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \mathbf{A}^{i,j} z_1^{-i} z_2^{-j} \end{aligned} \quad (4.46)$$

where the series expansion is absolutely convergent in the region U_2 [20], and

$$\begin{aligned} \mathbf{A}^{0,0} &= \mathbf{I} \\ \mathbf{A}^{i,j} &= \mathbf{A}_h \mathbf{A}^{i-1,j} + \mathbf{A}_v \mathbf{A}^{i,j-1}, \text{ for } i + j > 0 \\ \mathbf{A}^{i,j} &= \mathbf{0}, \text{ for } i < 0 \text{ or } j < 0. \end{aligned} \quad (4.47)$$

Hence, from Eq. (4.46), the weight-error vector $\mathbf{C}_{m,n}$ can be calculated by

$$\mathbf{C}_{m,n} = \sum_{i=0}^m \sum_{j=0}^n H(i, j) \varepsilon_{m-i, n-j} \mathbf{X}_{m-i, n-j} \quad (4.48)$$

with

$$H(i, j) = \mu_h \mathbf{A}^{i-1, j} + \mu_v \mathbf{A}^{i, j-1}.$$

From Eq. (4.48), the weight-error correlation matrix $\mathbf{K}_{m,n-k;m-k,n}$ can be calculated for any spatial lag k as follows:

$$\begin{aligned} \mathbf{K}_{m,n-k;m-k,n} &= \mathbf{E}\{\mathbf{C}_{m,n-k}\mathbf{C}_{m-k,n}^t\} \\ &= \sum_{i=0}^m \sum_{j=0}^n \sum_{p=0}^m \sum_{q=0}^n H(i, j-k) \mathbf{E}\{\mathbf{V}_{m-i,n-k-j}\mathbf{V}_{m-p-k,n-q}^t\} H(p-k, q) \end{aligned} \quad (4.49)$$

where, for notational convenience, we have defined:

$$\mathbf{V}_{m,n} = \varepsilon_{m,n} \mathbf{X}_{m,n}. \quad (4.50)$$

If the probability distribution of the input signal $x(m, n)$ and the measurement noise $\varepsilon_{m,n}$ are available, Eq. (4.49) can be used to obtain the weight-error correlation matrix $\mathbf{K}_{m,n-k;m-k,n}$.

For the special case when the measurement noise $\varepsilon(m, n)$ is white Gaussian noise and independent of $x(m, n)$, Eq. (4.49) is reduced to:

$$\mathbf{K}_{m,n-k;m-k,n} = \sigma_\varepsilon^2 \sum_{i=k}^m \sum_{j=k}^n H(i, j-k) \mathbf{R} H(i-k, j). \quad (4.51)$$

Stability condition (4.9) guarantees that the spectral norm of each of the matrices \mathbf{A}_h , \mathbf{A}_v , and $\mathbf{A}^{i,j}$ are less than unity. And since these matrices are symmetric, it is straightforward to show that, $\lim_{i,j \rightarrow \infty} H(i, j) = 0$. Thus, we can deduce that the error that results from using the approximation (4.51) decreases as the spatial lag k increases. For sufficiently large k , the correlation matrix $\mathbf{K}_{m,n;m,n}$ can be approximated with zero as it has been suggested in Subsection 3.2. For $k = 0$, Eq. (4.51) can be used as an approximation of the weight error covariance matrix $\mathbf{K}_{m,n;m,n}$.

4.6 Experimental Results and Discussion

Example 1:

In this example we aim to test the accuracy of the obtained analytical results for the simplified setting ($\lambda_0 = \lambda_1, \dots, \lambda_{N^2-1} = \sigma_x^2$, $f_h = f_v = 0.5$, and $\mu_h = \mu_v = \mu$). We performed system identification experiment for the following 2-D FIR filter:

$$\begin{aligned} d(m, n) &= x(m, n) + 0.5x(m-1, n) + 0.5x(m, n-1) \\ &\quad + 0.125x(m-1, n-1) + \varepsilon(m, n). \end{aligned} \quad (4.52)$$

We used two independent, 2-D white Gaussian sequences with variances $\sigma_x^2 = 1$, and $\sigma_\varepsilon^2 = 1$ for the input signal $x(m, n)$ and the additive noise $\varepsilon(m, n)$ respectively.

As a measure for the performance of the 2-D doubly indexed LMS we used the misadjustment M which is defined as

$$\begin{aligned} M &= \frac{\epsilon_\infty - \sigma_\varepsilon^2}{\sigma_\varepsilon^2} \\ &= \frac{1}{\sigma_\varepsilon^2} \lim_{m+n \rightarrow \infty} \text{tr}(\mathbf{RK}_{m,n;m,n}) \\ &= \frac{1}{\sigma_\varepsilon^2} \sum_{j=0}^{N^2-1} \gamma_0^{j,j} \lambda_j. \end{aligned} \quad (4.53)$$

Fig. 4.4 shows a comparison between experimental results and the misadjustment obtained using two different methods. In the first method (referred to as the independent assumption method in Fig. 4.4), the coefficients of the WECM in Eq. (4.53) were calculated using Eq. (4.39). And in the second (referred to as the the direct averaging method in Fig. 4.4), the WECM in Eq. (4.53) were calculated using the direct averaging method presented in Section 4.5 with k set

Table 4.1: Simulation results for Example 2, correlated Gaussian input

step size μ	$\alpha = 0.2$ $\beta = 0.19$		$\alpha = 0.3$ $\beta = 0.16$		$\alpha = 0.4$ $\beta = 0.11$	
	Indep.	Exper.	Indep.	Exper.	Indep.	Exper.
$0.3 \times \beta$	0.0929	0.0929	0.0821	0.0821	0.0544	0.0545
$0.4 \times \beta$	0.1522	0.1522	0.1339	0.1341	0.0877	0.0901
$0.5 \times \beta$	0.2288	0.2289	0.2001	0.2010	0.1283	0.1287
$0.6 \times \beta$	0.3297	0.3316	0.2856	0.2871	0.1776	0.1864
$0.7 \times \beta$	0.4661	0.5011	0.3989	0.4783	0.2381	0.3180
$0.8 \times \beta$	0.6600	0.8611	0.5550	1.1034	0.3130	0.6101

to zero in Eq. (4.51). The experimental misadjustment is calculated by averaging the results of 30 independent runs. For each run the misadjustment is calculated by averaging 40000 iteration in the steady state.

From Fig. 4.4, we can observe that the MSE analysis using both the independent assumption and the direct averaging method gives satisfactory results for small step size value. However, as the step size μ increases, the error in the estimated MSE increases. On the other hand, we can notice that the performance of the 2-D doubly indexed LMS was well preserved using the independence assumption based analysis, whereas, the direct averaging method fails completely for large step size values.

Example 2:

In this example we performed the system identification experiment described in Example 1, however, with correlated input signal. The correlated input $x(m, n)$ was generated by filtering a 2-D white Gaussian noise $u(m, n)$ of zero mean and

Table 4.2: Simulation results for Example 2, maximum step size value μ_{max} , Gaussian input, $N=2$.

α	0	0.1	0.2	0.3	0.4	0.5
Ref. [19]	1	0.81	0.67	0.55	0.46	0.38
Indep.	0.29	0.26	0.22	0.19	0.16	0.14

unit variance with the following 2-D, 2×2 filter:

$$x(m, n) = u(m, n) + \alpha u(m - 1, n) + \alpha u(m, n - 1). \quad (4.54)$$

Accordingly, the 4×4 input correlation matrix \mathbf{R} is given by:

$$\mathbf{R} = \begin{bmatrix} 1 + \alpha^2 & \alpha & \alpha & 0 \\ \alpha & 1 + \alpha^2 & \alpha^2 & \alpha \\ \alpha & \alpha^2 & 1 + \alpha^2 & \alpha \\ 0 & \alpha & \alpha & 1 + \alpha^2 \end{bmatrix}. \quad (4.55)$$

We repeated the same experiment for different value of α , ($\alpha = 0.2, 0.3, 0.4$), to test the accuracy of the obtained analytical results for different levels of input correlation. In this example, an approximation of the weight-error correlation coefficients $\gamma_2^{i,i}$ was calculated using the proposed direct averaging method (Eq. (4.51) with $k = 2$). The weight-error correlation coefficients $\gamma_0^{i,i}$, $i = 0, \dots, N^2 - 1$ were then obtained by solving Eqs. (4.36) and (4.37).

Table 4.1 shows the values of the misadjustment calculated both experimentally and using the proposed independence assumption based analysis. From the table entries which are illustrated also in Fig. 4.5, it is seen that the independence assumption based analysis provides accurate results for small step size values. However, for large step size values, the error in estimating the misadjustment of the adaptive filter increases as the level of the input correlation

increases.

The upper bound on the step size parameter μ that ensures the convergence in the MSE, say μ_{max} , were calculated for each particular value of α from condition (4.41). Table 4.2 shows the obtained numerical results in comparison with the upper bound on the step size parameter μ that ensures convergence of the mean as given by Eq. (4.9) [19]. It is seen that the maximum step size values that ensures convergence of the MSE is significantly smaller than those that ensures the convergence of the mean.

Example 3:

In this example we aim to test the obtained analytical results for non Gaussian input. We performed a system identification experiment similar to that presented in Example 1, however with uncorrelated binary input signal of unit variance. Fig. 4.6 shows the values of the misadjustment obtained both experimentally and using the independence assumption analysis. It can be seen that the independence assumption-based analysis can serve to give good insight to the behavior of the adaptive process even when the Gaussian assumption does not hold.

4.7 Summary

We have considered the steady state MSE analysis for 2-D doubly indexed LMS algorithm using the independence assumption. We have shown that the evaluation of the weight-error covariance matrix for doubly-indexed 2-D LMS algorithm requires approximation of the weight error correlation coefficients at large spatial lags. Then, we have proposed a method to solve this problem. We have

shown that the convergence in the MSE sense occurs for step size range that is significantly smaller than the one necessary for the convergence of the mean. Simulation example was presented to support the analytical results and to show that the analysis using the independence assumption does provide good insight to the performance of the 2-D doubly indexed LMS algorithm.

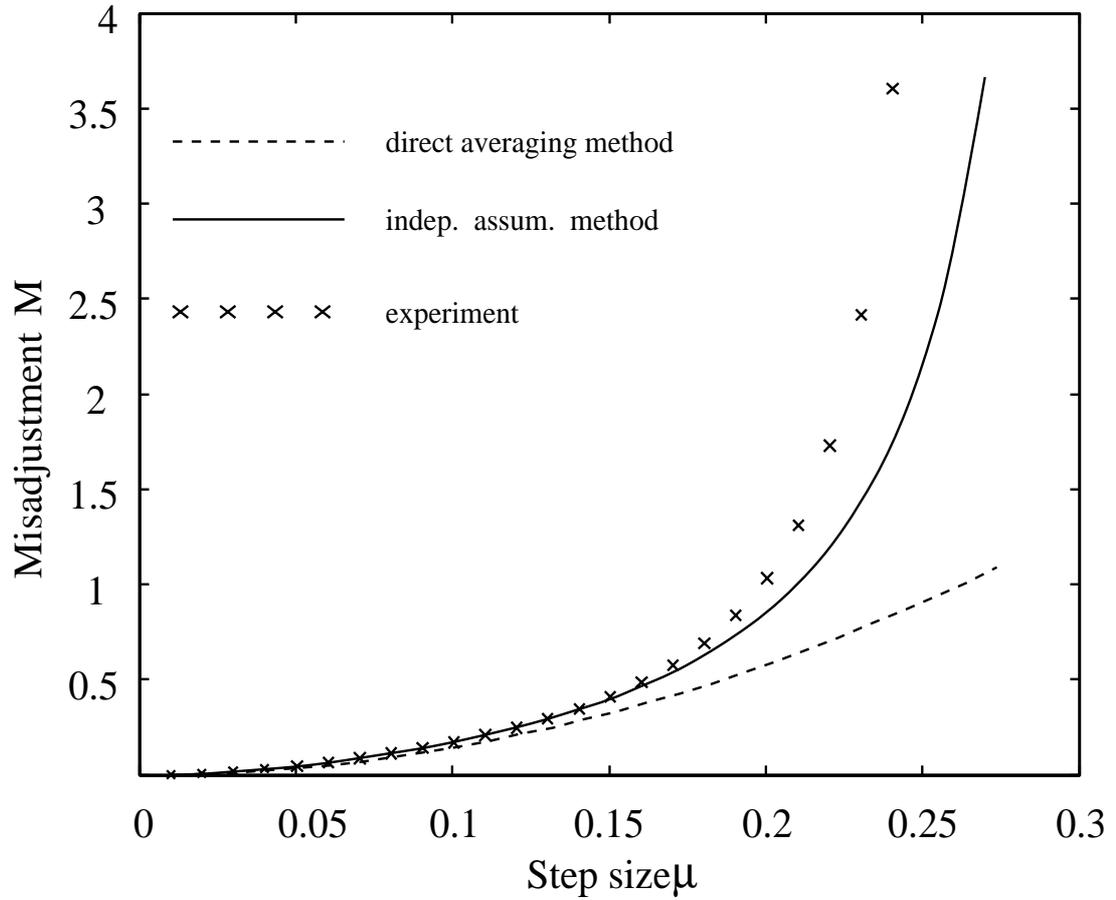
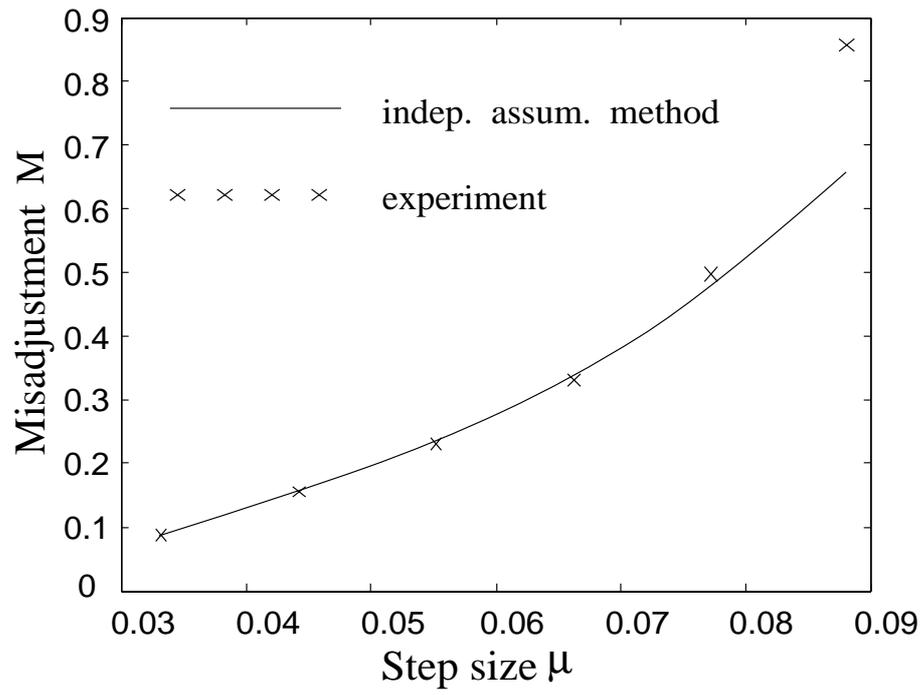
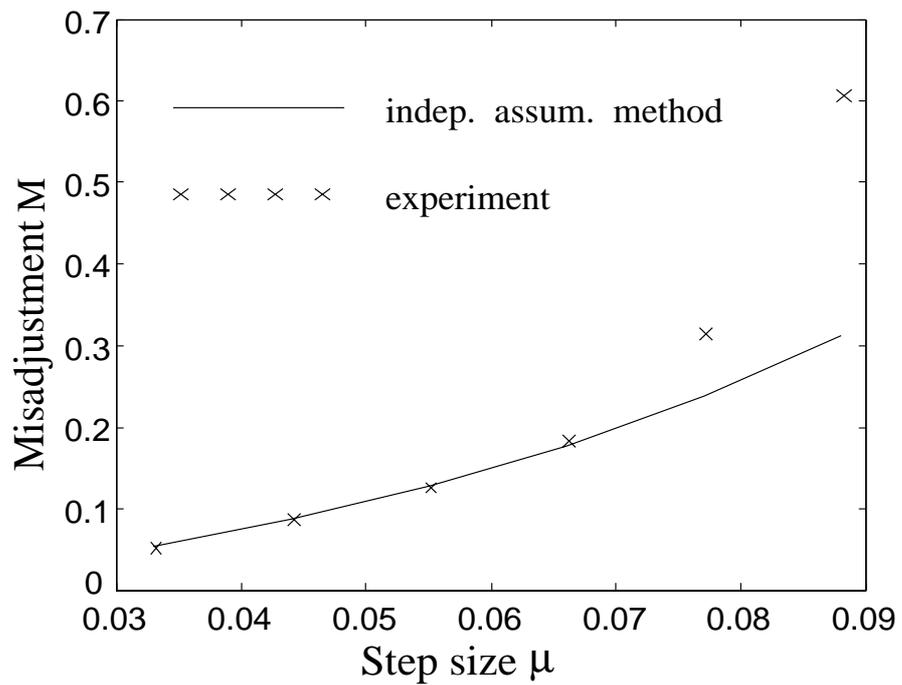


Figure 4.4: Comparison of the experimental results with the theoretical values for the misadjustment of the 2-D doubly indexed LMS in the simplified setting ($\lambda_0 = \lambda_1, \dots, \lambda_{N^2-1} = \sigma_x^2$, $f_h = f_v = 0.5$, and $\mu_h = \mu_v = \mu$), white Gaussian input.



(a)



(b)

Figure 4.5: Simulation results for correlated Gaussian input, (a) $\alpha = 0.2$, (b) $\alpha = 0.4$.

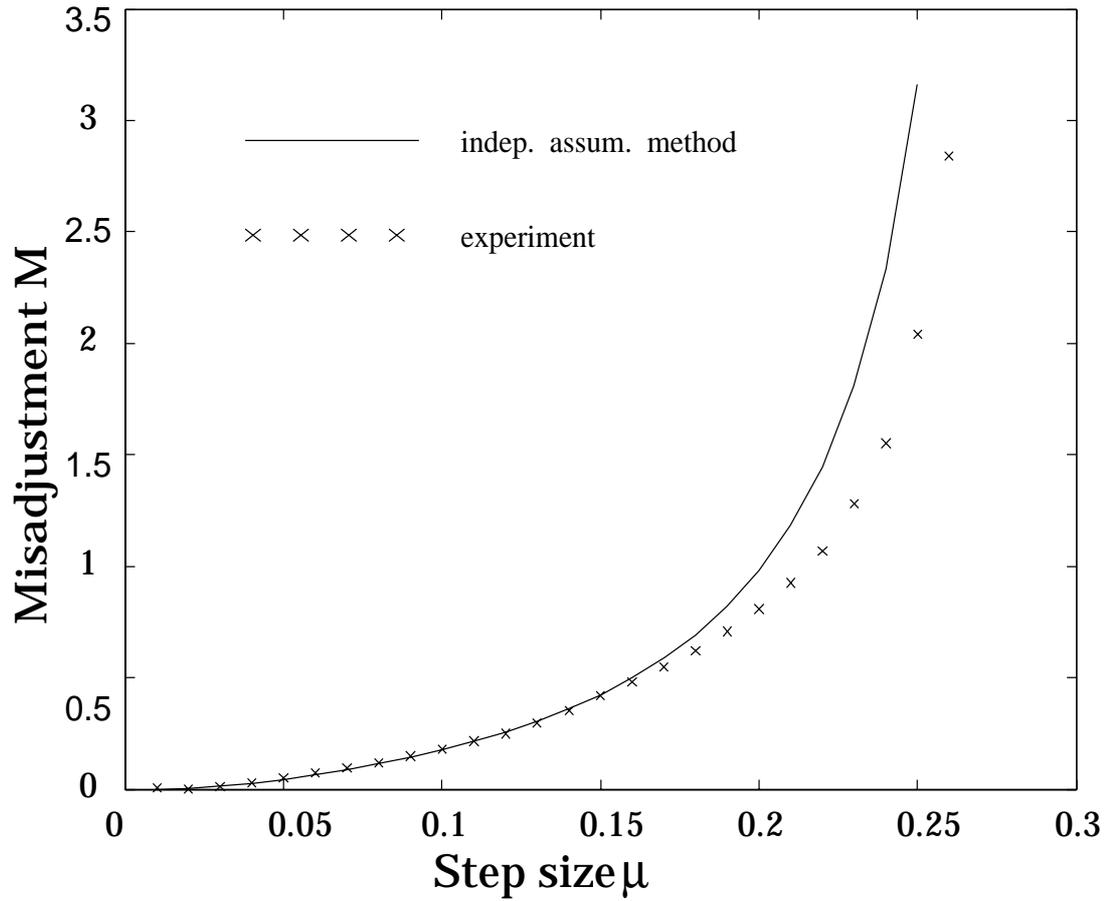


Figure 4.6: Comparison of the experimental results with the theoretical values for the misadjustment of the 2-D doubly indexed LMS in the simplified setting ($\lambda_0 = \lambda_1, \dots, \lambda_{N^2-1} = \sigma_x^2$, $f_h = f_v = 0.5$, and $\mu_h = \mu_v = \mu$), i.i.d. binary input.

Chapter 5

2-D Doubly-Indexed Block LMS Adaptive Filters

5.1 Introduction

Block adaptive filtering allows efficient implementation of parallel processors and Fast Fourier Transform (FFT) for convolution [5], [21]. Images are in particular very suitable for block-wise processing scheme. Therefore, several 2-D block LMS adaptive FIR filtering algorithms, with application to image enhancement and 2-D system identification, have been proposed [22]-[24]. In [22], a 2-D block LMS algorithm was introduced by direct extension of the 1-D block LMS [21]. However, in order to preserve the local correlation information of the image pixels, a block diagonal indexing scheme was employed. On the other hand, the authors in [23] and [24] have worked on the development of an optimal space varying convergence parameter that is adjusted once per each data block in order to improve the convergence behavior in nonstationary environment.

This chapter focuses on the development of a new 2-D adaptive LMS FIR filtering algorithm by block-wise processing of data in order to gain the benefits

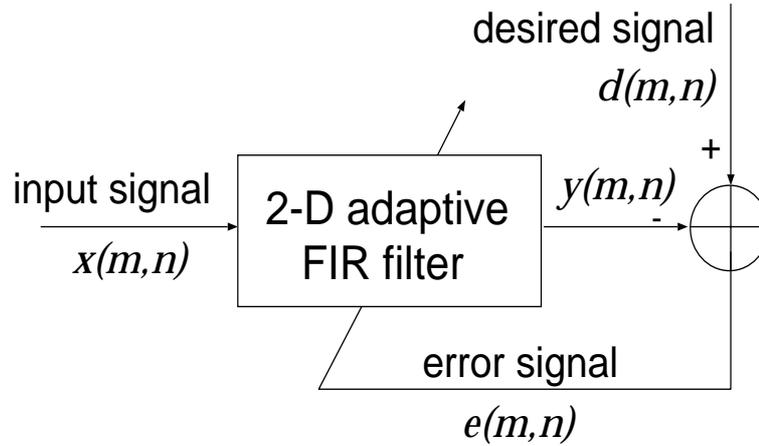


Figure 5.1: 2-D adaptive FIR filter.

of parallel computation and improved stability performance associated with block filtering scheme. In the proposed algorithm, the input signal is partitioned into non-overlapping blocks; the weights are then adjusted once per each block of the input signal. The filter weights update process is carried out along both the vertical and horizontal directions as a doubly-indexed dynamical system in accordance with the 2-D LMS [19].

5.2 2-D Doubly-Indexed Block LMS Algorithm (2DDI-BLMS)

Consider the 2-D, $N \times N$ -tap adaptive FIR filter shown in Fig. 5.1. The adaptive filter's input $x(m,n)$ is a 2-D signal of size $M \times M$ partitioned into non-overlapping blocks; each block is of size $K \times K$ as shown in Fig. 5.2. The (i,j) th block of the

output signal $y(m, n)$ is calculated by

$$Y_{i,j} = \mathcal{X}_{i,j} \mathbf{H}_{i,j} \quad (5.1)$$

where $Y_{i,j}$ is the $K^2 \times 1$ vector that consists of the (i, j) th output block elements arranged in row order as follows:

$$Y_{i,j} = [\mathbf{y}_{iK}, \mathbf{y}_{iK+1}, \dots, \mathbf{y}_{iK-K+1}] \quad (5.2)$$

with

$$\mathbf{y}_p = [y(p, jK), \dots, y(p, jK + K - 1)]. \quad (5.3)$$

$\mathbf{H}_{i,j}$ is the adaptive filter's weight-vector at the (i, j) th block given by

$$\mathbf{H}_{i,j} = [h_{i,j}(0, 0), \dots, h_{i,j}(0, N - 1), \dots, h_{i,j}(N - 1, N - 1)]^t \quad (5.4)$$

and $\mathcal{X}_{i,j}$ is the (i, j) th block input matrix of size $K^2 \times N^2$, defined as

$$\begin{aligned} \mathcal{X}_{i,j} = & [\mathbf{X}_{iK,jK}, \mathbf{X}_{iK,jK+1}, \dots, \mathbf{X}_{iK,jK+K-1}, \mathbf{X}_{iK+1,jK}, \\ & \mathbf{X}_{iK+1,jK+1}, \dots, \mathbf{X}_{iK+1,jK+K-1}, \dots, \mathbf{X}_{iK+K-1,jK+K-1}]^t \end{aligned} \quad (5.5)$$

with

$$\mathbf{X}_{m,n} = [x(m, n), \dots, x(m, n - N + 1), \dots, x(m - N + 1, n - N + 1)] \quad (5.6)$$

denoting the adaptive filter's input vector at spatial indices (m, n) .

The adaptive filter's weight-vector is adjusted subject to minimizing the 2-D Block Mean Square Error :

$$\text{BMSE} = \frac{1}{K^2} \text{E}\{E_{i,j}^t E_{i,j}\} \quad (5.7)$$

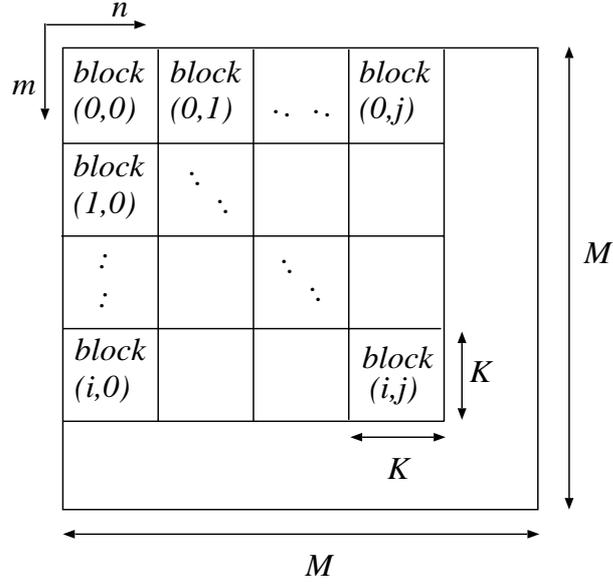


Figure 5.2: 2-D block-wise processing scheme.

where E denotes the expectation operator, and

$$E_{i,j} = D_{i,j} - Y_{i,j} \quad (5.8)$$

with $D_{i,j}$ denoting the $K^2 \times 1$ vector that consists of the elements of the (i, j) th block of the desired signal $d(m, n)$, defined in a way similar to $Y_{i,j}$. The weight-vector update equation for the 2DDI-BLMS is given by

$$\begin{aligned} \mathbf{H}_{i+1,j+1} &= f_h \mathbf{H}_{i,j+1} + f_v \mathbf{H}_{i+1,j} \\ &+ \frac{\mu_h}{K^2} \mathcal{X}_{i,j+1}^t E_{i,j+1} + \frac{\mu_v}{K^2} \mathcal{X}_{i+1,j}^t E_{i+1,j}; \\ \mathbf{H}_{i,0} &= \mathbf{0}, i = 0, \dots, i_{max}; \quad \mathbf{H}_{0,j} = \mathbf{0}, j = 0, \dots, j_{max}; \\ f_h + f_v &= 1 \end{aligned} \quad (5.9)$$

where μ_h and μ_v denote the step size parameters in the horizontal and vertical

directions respectively.

In Eq. (5.9), the weight-vector is updated in both directions as a 2-D state space model. This update method enables efficient use of the 2-D correlation information of the image pixels in both vertical and horizontal directions and hence, provides better performance in nonstationary environments. Moreover, using such update method, it is possible to adjust simultaneously all the adaptive filter's weight-vectors $\mathbf{H}_{i,j}$'s that lie on the same diagonal. This property makes the 2DDI-BLMS very suitable for parallel processing.

In the following section, we analyse the convergence behavior of the 2DDI-BLMS.

5.3 Convergence Analysis of the 2DDI-BLMS

The convergence analysis of the 2DDI-BLMS is carried out using the following assumptions:

A.1 The input vectors $\mathbf{X}_{0,0}, \mathbf{X}_{1,0}, \dots, \mathbf{X}_{m,n}$ are zero mean, statistically independent, Gaussian-distributed random variables.

A.2 There exists a true weight-vector \mathbf{H}_{opt} such that the error signal

$$\varepsilon(m, n) = d(m, n) - \mathbf{H}_{opt}^t \mathbf{X}_{m,n} \quad (5.10)$$

is a zero mean, white Gaussian noise of variance σ_ε^2 , and is statistically independent of the input vector $\mathbf{X}_{m,n}$.

To analyze the convergence behavior of the adaptive algorithm in Eq. (5.9), we should first derive the update equation for the weight-error vector

$$\mathbf{C}_{i+1,j+1} = \mathbf{H}_{i+1,j+1} - \mathbf{H}_{opt}. \quad (5.11)$$

Now, using Eqs. (5.10) and (5.11), the output error vector $E(i, j)$ in Eq. (5.8) can be calculated by

$$E(i, j) = \mathcal{E}_{i,j} - \mathcal{X}_{i,j} \mathbf{C}_{i,j}. \quad (5.12)$$

where $\mathcal{E}_{i,j}$ is the $K^2 \times 1$ vector that consists of the elements of the (i, j) th block of the error signal $\varepsilon(m, n)$.

Then, if we subtract \mathbf{H}_{opt} from both sides of Eq. (5.9) and make use of Eq. (5.12), we get

$$\begin{aligned} \mathbf{C}_{i+1,j+1} &= \mathbf{H}_{i+1,j+1} - \mathbf{H}_{opt} \\ &= (f_h \mathbf{I} - \frac{\mu_h}{K^2} \mathcal{X}_{i,j+1}^t \mathcal{X}_{i,j+1}) \mathbf{C}_{i,j+1} \\ &\quad + (f_v \mathbf{I} - \frac{\mu_v}{K^2} \mathcal{X}_{i+1,j}^t \mathcal{X}_{i+1,j}) \mathbf{C}_{i+1,j} \\ &\quad + \frac{\mu_h}{K^2} \mathcal{X}_{i,j+1}^t \mathcal{E}_{i,j+1} + \frac{\mu_v}{K^2} \mathcal{X}_{i+1,j}^t \mathcal{E}_{i+1,j}. \end{aligned} \quad (5.13)$$

We can now study the convergence properties of the 2DDI-BLMS. We do this by calculating the mean of the weight-error vector $\mathbf{C}_{i+1,j+1}$ and the steady state value of the BMSE.

5.3.1 Convergence of the Mean

Taking the expectation of both sides of Eq. (5.13) and using the assumptions A.1 and A.2 we get

$$\mathbf{E}\{\mathbf{C}_{i+1,j+1}\} = (f_h \mathbf{I} - \mu_h \mathbf{R}) \mathbf{E}\{\mathbf{C}_{i,j+1}\} + (f_v \mathbf{I} - \mu_v \mathbf{R}) \mathbf{E}\{\mathbf{C}_{i+1,j}\}. \quad (5.14)$$

Eq. (5.14) is the same 2-D state space model derived for the mean of weight-error vector of the 2-D LMS. Accordingly, proceeding the mean analysis as discussed in [19] it can be shown that $E\{\mathbf{C}_{i,j}\} \rightarrow 0$ as $i+j \rightarrow \infty$, if the following condition holds:

$$|f_h - \mu_h \lambda_p| + |f_v - \mu_v \lambda_p| < 1 \quad (5.15)$$

where $\lambda_p, p = 0, \dots, N^2 - 1$, are the eigenvalues of the input correlation matrix $\mathbf{R} = E\{\mathbf{X}_{m,n} \mathbf{X}_{m,n}^t\}$.

Note that condition (5.15) is exactly the same condition necessary for the convergence of the mean of the 2-D LMS.

5.3.2 Convergence of the BMSE

Calculation of the steady state BMSE

Making use of Eq. (5.12), and in the light of assumptions *A.1* and *A.2*, we can find that the steady state BMSE is given by

$$\begin{aligned} \Xi_\infty &= \lim_{i+j \rightarrow \infty} \frac{1}{K^2} E\{E_{i,j} E_{i,j}^t\} \\ &\quad \sigma_\varepsilon^2 + \lim_{i+j \rightarrow \infty} \text{tr}(\mathbf{R} K_{i,j;i,j}) \end{aligned} \quad (5.16)$$

where

$$K_{i,j;i,j} = E\{\mathbf{C}_{i,j} \mathbf{C}_{i,j}^t\} \quad (5.17)$$

is the weight-error covariance matrix. In the rest of this section we will consider the calculation of the weight-error covariance matrix. In this calculation, we assume that the condition (5.15), which is necessary for the convergence of the mean, holds.

Calculation of the weight-error covariance matrix

For the calculation of the weight-error covariance matrix $\mathbf{K}_{i,j;i,j}$, it will be useful to use a unitary transformation \mathbf{Q} such that

$$\mathbf{Q}\mathbf{R}\mathbf{Q}^t = \mathbf{\Lambda} = \text{diag}(\lambda_p), \quad p = 0, \dots, N^2 - 1. \quad (5.18)$$

Then, we can define the following transformed variance matrix:

$$\begin{aligned} \mathbf{\Gamma}_{i_1,j_1;i_2,j_2} &= \mathbf{Q}\mathbf{E}\{\mathbf{C}_{i_1,j_1}\mathbf{C}_{i_2,j_2}^t\}\mathbf{Q}^t \\ &= [\gamma_{i_1,j_1;i_2,j_2}^{p,q}]; p, q = 0, \dots, N^2 - 1. \end{aligned} \quad (5.19)$$

Now, multiplying each side of Eq. (5.13) with its transpose, taking the expected values, and making use of the orthogonal transform \mathbf{Q} we arrive at

$$\begin{aligned} \mathbf{\Gamma}_{i+1,j+1;i+1,j+1} &= f_h^2 \mathbf{\Gamma}_{i,j+1;i,j+1} - \mu_h f_h (\mathbf{\Gamma}_{i,j+1;i,j+1} \mathbf{\Lambda} \\ &\quad + \mathbf{\Lambda} \mathbf{\Gamma}_{i,j+1;i,j+1}) + \frac{\mu_h^2}{K^2} (2(K^2 + 1) \mathbf{\Lambda} \mathbf{\Gamma}_{i,j+1;i,j+1} \mathbf{\Lambda} + \\ &\quad + \text{tr}(\mathbf{\Gamma}_{i,j+1;i,j+1} \mathbf{\Lambda}) \mathbf{\Lambda}) + f_h f_v \mathbf{\Gamma}_{i+1,j;i,j+1} \\ &\quad - \mu_h f_v \mathbf{\Gamma}_{i+1,j;i,j+1} \mathbf{\Lambda} - \mu_v f_h \mathbf{\Lambda} \mathbf{\Gamma}_{i+1,j;i,j+1} \\ &\quad + \frac{\mu_h \mu_v}{K^2} \mathbf{\Lambda} \mathbf{\Gamma}_{i+1,j;i,j+1} \mathbf{\Lambda} + f_h f_v \mathbf{\Gamma}_{i,j+1;i+1,j} \\ &\quad - \mu_v f_h \mathbf{\Gamma}_{i,j+1;i+1,j} \mathbf{\Lambda} - \mu_h f_v \mathbf{\Lambda} \mathbf{\Gamma}_{i,j+1;i+1,j} \\ &\quad + \frac{\mu_h \mu_v}{K^2} \mathbf{\Lambda} \mathbf{\Gamma}_{i,j+1;i+1,j} \mathbf{\Lambda} + f_v^2 \mathbf{\Gamma}_{i+1,j;i+1,j} \\ &\quad - \mu_v f_v (\mathbf{\Gamma}_{i+1,j;i+1,j} \mathbf{\Lambda} + \mathbf{\Lambda} \mathbf{\Gamma}_{i+1,j;i+1,j}) \\ &\quad + \frac{\mu_v^2}{K^2} (2(K^2 + 1) \mathbf{\Lambda} \mathbf{\Gamma}_{i+1,j;i+1,j} \mathbf{\Lambda} \\ &\quad + \text{tr}(\mathbf{\Gamma}_{i+1,j;i+1,j} \mathbf{\Lambda}) \mathbf{\Lambda}) + \frac{\mu_h^2 + \mu_v^2}{K^2} \sigma_\varepsilon^2 \mathbf{\Lambda}. \end{aligned} \quad (5.20)$$

Analysing the stability of the set of second-order coupled 2-D difference equations (5.20) is a very difficult to handle mathematically. Thus, we propose to simplify the analysis as follows.

1. For the transformed weight-error correlation matrix defined in Eq. (5.19) we have

$$(\gamma_{i_1, j_1; i_2, j_2}^{p, q})^2 \leq \gamma_{i_1, j_1; i_1, j_1}^{p, p} \cdot \gamma_{i_2, j_2; i_2, j_2}^{q, q}. \quad (5.21)$$

Therefore, it is sufficient to analyse the stability of the diagonal terms of the matrix equation (5.20). Note that, as for the BMSE evaluation (see Eq. (5.16)), we are only interested in the diagonal terms since

$$\text{tr}(\mathbf{R}K_{i, j; i, j}) = \text{tr}(\mathbf{\Lambda}\mathbf{\Gamma}_{i, j; i, j}) = \sum_{p=0}^{N^2-1} \gamma_{i, j; i, j}^{p, p} \lambda_p. \quad (5.22)$$

2. If $\gamma_{i+1, j+1; i+1, j+1}^{p, p}$ will reach a limit value, say $\gamma_0^{p, p}$, as $i + j \rightarrow \infty$, then this implies that

$$\lim_{i+j \rightarrow \infty} \gamma_{i+1, j; i+1, j}^{p, p} = \lim_{i+j \rightarrow \infty} \gamma_{i, j+1; i, j+1}^{p, p} = \gamma_0^{p, p}. \quad (5.23)$$

3. Similarly, if $\gamma_{i+1, j; i, j+1}^{p, p}$ will reach a limit value, say $\gamma_1^{p, p}$, as $i + j \rightarrow \infty$, then

$$\lim_{i+j \rightarrow \infty} \gamma_{i+1, j; i, j+1}^{p, p} = \lim_{i+j \rightarrow \infty} \gamma_{i, j+1; i+1, j}^{p, p} = \gamma_1^{p, p}. \quad (5.24)$$

Accordingly, for the steady state, the diagonal terms of the matrix equation (5.20) should obey the equality

$$\begin{aligned}
& 2 \left(f_h f_v + (\mu_h f_h + \mu_v f_v) \lambda_p - \frac{(K^2 + 1)(\mu_h^2 + \mu_v^2)}{2K^2} \lambda_p^2 \right) \gamma_0^{p,p} \\
& - 2 \left(f_h f_v - (\mu_h f_v + \mu_v f_h) \lambda_p + \frac{\mu_h \mu_v}{K^2} \lambda_p^2 \right) \gamma_1^{p,p} \\
& - \frac{(\mu_h^2 + \mu_v^2) \lambda_p}{K^2} \left(\sum_{u=0}^{N^2-1} \gamma_0^{u,u} \lambda_u + \sigma_\varepsilon^2 \right) = 0; \\
& p = 0, \dots, N^2 - 1.
\end{aligned} \tag{5.25}$$

There is a need for another set of equations in the unknowns $\gamma_0^{p,p}$ and $\gamma_1^{p,p}$. If we apply the same way of analysis to evaluate the matrix $\mathbf{\Gamma}_{i+1,j;i,j+1}$, we can find that for the steady state, i.e. $i + j \rightarrow \infty$, the diagonal terms of the correlation matrix $\mathbf{\Gamma}_{i+1,j;i,j+1}$ should obey the equality

$$\begin{aligned}
& \left(f_h f_v - (\mu_h f_v + \mu_v f_h) \lambda_p + \frac{\mu_h \mu_v (K^2 + 1)}{K^2} \lambda_p^2 \right) \gamma_0^{p,p} \\
& + \left(-2f_h f_v - 2(\mu_h f_h + \mu_v f_v) \lambda_p + \frac{\mu_h^2 + \mu_v^2}{K^2} \lambda_p^2 \right) \gamma_1^{p,p} \\
& + \left(f_h f_v - (\mu_h f_v + \mu_v f_h) \lambda_p + \frac{\mu_h \mu_v}{K^2} \lambda_p^2 \right) \gamma_2^{p,p} \\
& + \frac{\mu_h \mu_v}{K^2} \lambda_p \left(\sum_{u=0}^{N^2-1} \gamma_0^{u,u} \lambda_u + \sigma_\varepsilon^2 \right) = 0; \\
& p = 0, \dots, N^2 - 1
\end{aligned} \tag{5.26}$$

where

$$\gamma_2^{p,p} = \lim_{i+j \rightarrow \infty} \gamma_{i+1,j-1;i-1,j+1}^{p,p}.$$

If we continue in similar way evaluating the steady state values of the weight-error correlation matrices for higher diagonal spatial lags, i. e.

$$\lim_{i+j \rightarrow \infty} \mathbf{\Gamma}_{i+1,j+1-k;i+1-k,j+1} = [\gamma_k^{p,q}]; k = 3, 4, \dots,$$

at each stage k ($k = 0, 1, \dots$), we will have a set of $(k + 1) \times N^2$ equations in $(k + 2) \times N^2$ unknowns, namely, $\gamma_l^{p,p}; 0 \leq l \leq k + 1, 0 \leq p \leq N^2 - 1$. However, under the white Gaussian assumption for the input vector $\mathbf{X}_{m,n}$ and the error signal $\varepsilon(m, n)$, we can assume that for sufficiently large spatial lags, i.e. $k \gg 1$, the correlation coefficients $\gamma_{k+1}^{p,p}$ can be approximated with zero; therefore the available $(k + 1) \times N^2$ equations can be solved for the $(k + 1) \times N^2$ unknowns to obtain $\gamma_0^{p,p}, p = 0, \dots, N^2 - 1$. The solution of these $(k + 1) \times N^2$ simultaneous equations can be obtained for the general case using software package for algebraic computation such as Mathematica. In the following section we will discuss in more detail the steady state analysis for the simple case when the input signal is white Gaussian noise.

5.4 Steady State MSE Analysis with White Gaussian Input Data

For the mathematically simplest case where the input signal is white Gaussian with variance σ_x^2 , $f_h = f_v = 0.5$, and $\mu_h = \mu_v = \mu$, we have $\gamma_0^{0,0} = \gamma_0^{1,1} = \dots, \gamma_0^{N^2-1, N^2-1} = \gamma_0$. Therefore, if we set $\gamma_2^{p,p}, p = 0, \dots, N^2 - 1$ to zero in Eq.(5.26), and solve the two sets of linear equations (5.25) and (5.26) for the weight-error covariance coefficient γ_0 , we obtain Eq. (5.27), which appears at the bottom of this page,

$$\gamma_0 = \left[\frac{\sigma_\varepsilon^2 \zeta^2}{\sigma_x^2 (0.25K^2 - K^2\zeta + (1 + K^2 + N^2)\zeta^2)} \right] \times \left[\frac{0.375K^4 - K^4\zeta + K^2(1 + 1.5N^2 - 0.5K^2)\zeta^2 + K^2(4 + 2N^2 + 2K^2)\zeta^3 - 2(1 + K^2 + N^2)\zeta^4}{0.125K^4 + 3K^4\zeta - K^2(3 + 1.5N^2 - 0.5K^2)\zeta^2 - K^2(4 + 2N^2 + 2K^2)\zeta^3 + 2(1 + N^2 + K^2)\zeta^4} \right] \quad (5.27)$$

with

$$\zeta = \mu\sigma_x^2. \quad (5.28)$$

Note that, for block size $K=1$, Eq. (5.27) reduces to the weight-error covariance coefficient γ_0 for the 2-D LMS [30].

5.4.1 Bounds on the Step Size Parameter μ

Now, since the weight-error covariance coefficient γ_0 should be positive and finite, the range of the step size μ that ensures the convergence of the BMSE can be determined by the following condition

$$0 \leq \gamma_0 < \infty. \quad (5.29)$$

For this simplified case, analysis of Eq. (5.27) reveals that in this equation, the first term is always positive for $0 \leq \zeta < 1$ (the range of the step size μ that ensures the convergence of the mean), and that for any value of $N \geq 1$, the denominator of the second term has only one real positive root, say ζ_∞ , in the range $0 \leq \zeta < 1$, where the sign of γ_0 changes from positive to negative. Thus, we can deduce that the upper bound on the step size value that ensures finite variance is given by

$$0 \leq \mu\sigma_x^2 < \zeta_\infty. \quad (5.30)$$

Fig. 5.3 shows the values of the root ζ_∞ for different values of the filter order N with block size $K = N$. The values of ζ_∞ for the 2-D LMS [30] are also illustrated in the same figure. From this figure, it is clear that $\zeta_\infty < 1$ for any filter order

$N > 1$. Hence, the condition required for the convergence in the BMSE sense, as given in Eq. (5.30), decreases significantly the convergence region of the 2DDI-BLMS algorithm when comparing to the condition necessary for the convergence of the mean, $0 \leq \mu\sigma_x^2 < 1$, given by Eq. (5.15). On the other hand, we can observe that the choice of the step size parameter μ for the 2DDI-BLMS is less critical than for the 2-D LMS. This indicates that the 2DDI-BLMS is more stable than the 2-D LMS. Such result is expected since the 2DDI-BLMS uses more accurate estimate of the gradient of the mean squared output error.

5.4.2 Adaptation Accuracy

A commonly used measure of the adaptation accuracy is the misadjustment, defined as

$$M_B = \frac{1}{\sigma_\varepsilon^2} \lim_{i+j \rightarrow \infty} \text{tr}(\mathbf{R}\mathbf{K}_{i,j;i,j}) = \frac{1}{\sigma_\varepsilon^2} \sum_{p=0}^{N^2-1} \gamma_0^{p,p} \lambda_p. \quad (5.31)$$

A comparison between the misadjustment of the 2DDI-BLMS, M_B , and the misadjustment of the 2-D LMS, M can be achieved numerically since Eq. (5.27) is too complicated to derive a simple relation between the misadjustment of those two algorithms. For the case where the block size $K = N$, numerical results have shown that

$$\frac{M}{M_B} \approx K^2. \quad (5.32)$$

That is, the 2DDI-BLMS offers improvement in the misadjustment upon the 2-D LMS proportional to the used block size. However, this improvement comes along with slower convergence rate. In the following section we give one representative example that supports the convergence results presented in this section.

Table 5.1: Summary of experimental results

Algorithm	2-D LMS [19]	2DDI-BLMS
Exper. misadjustment	0.1661	0.0152
Theor. misadjustment	0.1545	0.0148

5.5 Experimental Results and Discussion

In this experiment, both the 2DDI-BLMS and the 2-D LMS were realized in 2-D system identification form. The following 3 by 3, 2-D FIR filter was used for the unknown system:

$$\begin{aligned}
 H(z_1^{-1}, z_2^{-1}) = & 0.914 + 0.926z_1^{-1} + 0.338z_1^{-2} + 0.494z_2^{-1} + 0.677z_1^{-1}z_2^{-1} \\
 & - 0.843z_1^{-1}z_2^{-2} + 0.894z_1^{-2} - 0.517z_1^{-2}z_2^{-1} - 0.777z_1^{-2}z_2^{-2}.
 \end{aligned}
 \tag{5.33}$$

A 2-D white Gaussian sequence with variance $\sigma_x^2 = 1$ was used for the input signal $x(m, n)$, and a 2-D white Gaussian sequence with variance $\sigma_\varepsilon^2 = 0.49$ was used for the error signal $\varepsilon(m, n)$. The step size parameter μ was set to 0.05 for both algorithms, and we used block size $K = N$ for the 2DDI-BLMS.

Table 5.1 shows the theoretical and experimental misadjustment of the 2DDI-BLMS and the 2-D LMS. The theoretical misadjustment was calculated using Eqs. (5.27) and (5.31).

Fig. 5.4 shows 1-D learning curves of the 2DDI-BLMS and the 2-D LMS

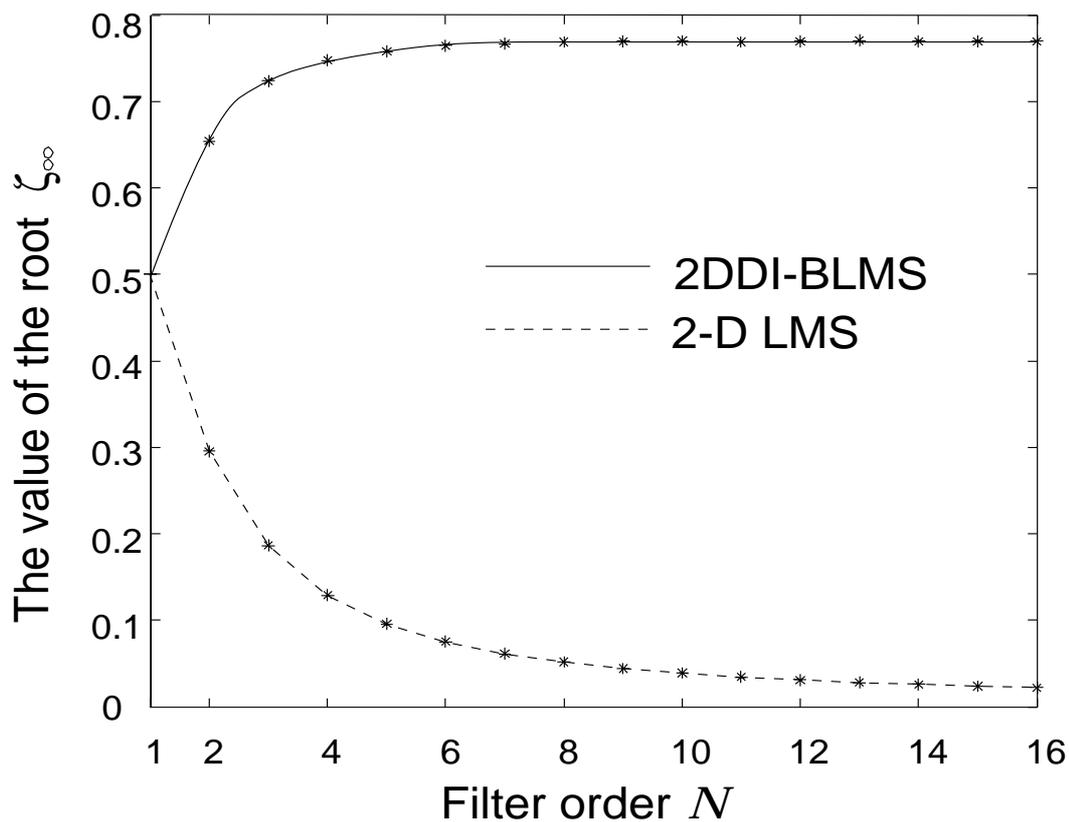


Figure 5.3: The value of the root ζ_∞ vs. filter order N .

which were obtained using the following 1-D error function

$$\text{err}(k) = \frac{1}{2(k+1)} \sum_{l=0}^k \{e(l, k)^2 + e(k, l)^2\}, 0 \leq k \leq M-1. \quad (5.34)$$

These experimental results are in good agreement with the theoretical analysis presented in the previous section.

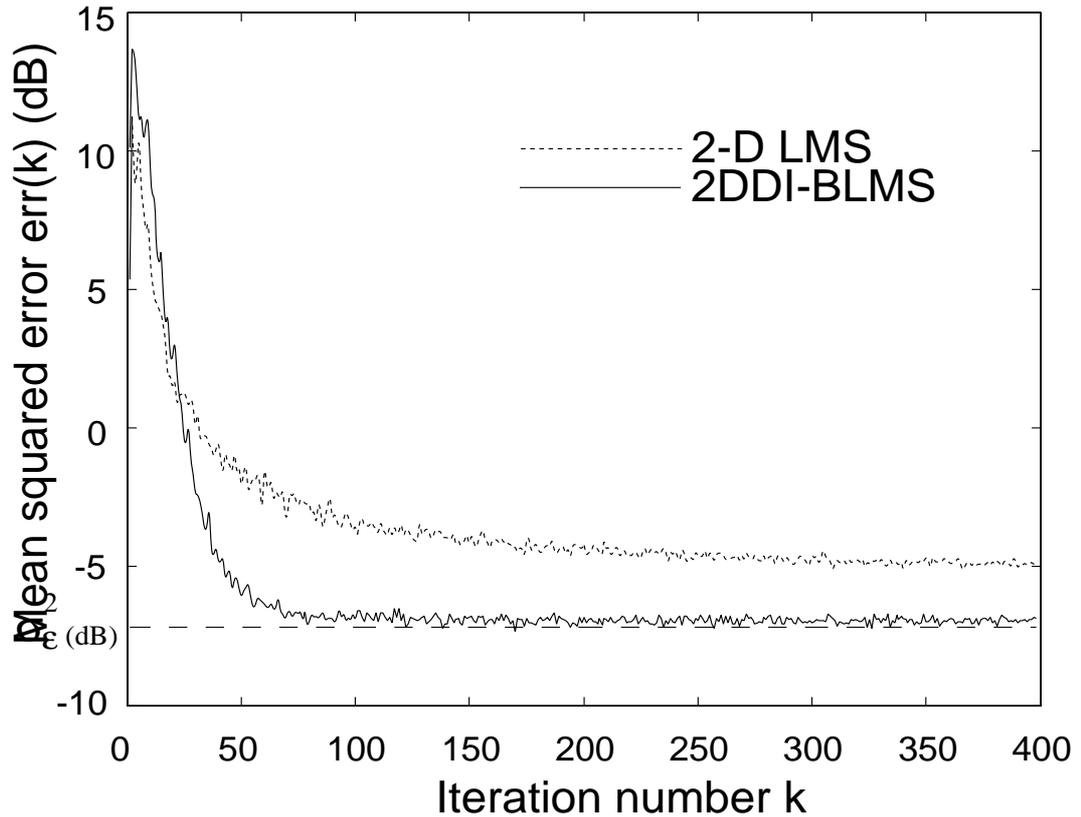


Figure 5.4: Learning curves for the 2DDI-BLMS and 2-D LMS (averages of 30 runs), step size $\mu = 0.05$.

5.6 Summary

In this chapter we have proposed the 2-D doubly indexed block LMS adaptive filtering algorithm. The weights update process is carried out along both the vertical and horizontal directions as a doubly-indexed dynamical system in accordance with the 2-D LMS[19] such that the correlation information of the image pixel can be efficiently exploited in both directions. Convergence analysis

has been presented, and bounds upon the convergence parameters that guarantee convergence of the proposed algorithm in the mean and the variance have been also derived. Theoretical comparison with 2-D LMS [19], has revealed that both algorithms require similar bounds on the step size parameters to guarantee convergence of the mean; however, the 2DDI-BLMS has improved variance performance over the 2-D LMS. Experimental results that support the validity of the obtained convergence analysis results have been also presented.

Chapter 6

Conclusions and Suggestions for Future Work

6.1 Conclusions

The main goal of this thesis is the development and analysis of the convergence behaviour of some 2-D adaptive filtering algorithms. The contributions of this thesis can be summarized as follows.

In Chapter 3, we have proposed the bias removal algorithm for 2-D equation error adaptive IIR filters with separable denominator function. The filter structure in the proposed algorithm is based on the concept of backpropagating the desired signal through a cascade of the denominator vertical and horizontal sections. To handle the bias problem, the proposed algorithm uses a scaled value of the output error of each of the cascade sections to counteract the effect of the measurement noise embedded in the regressor of the update procedure of that section. I/O stability analysis has been carried out. It has been shown that the proposed algorithm remains stable and the effect of the measurement noise can be significantly suppressed under general conditions imposed on the values of the

used step sizes and scaling factors. The performance of the proposed algorithm was compared with that of the family of hyperstable adaptive IIR Filtering algorithms [40] and the algorithm [15]. Image enhancement experiment results have been presented to show the faster convergence rate and better performance of the proposed 2DBRA algorithm.

In Chapter 4, we have discussed the the steady state MSE analysis for 2-D LMS algorithm in which the filter's weights are updated as doubly indexed dynamical system. The MSE analysis have been studied using the independence assumption. We have shown that the MSE analysis, essentially, the calculation of the weight-error covariance matrix for doubly-indexed 2-D LMS algorithm calls for stability analysis of a set of second-order coupled 2-D difference equations in the coefficients of the WECM, which is very difficult to handle mathematically. Then we have shown that for the steady state, this problem can be transformed to the problem of solving sets of linear equations in the weight-error correlation coefficients at different spatial lags. As a results of our analysis, we have shown that the convergence in the MSE sense occurs for step size range that is significantly smaller than the one necessary for the convergence of the mean. Simulation examples were also presented to support the analytical results and to show that the analysis using the independence assumption does provide good insight to the performance of the 2-D doubly indexed LMS algorithm even when the input signal is colored.

In Chapter 5, we have developed the block LMS adaptive filtering algorithm in which the weights update process is carried out along both the vertical and

horizontal directions as a doubly-indexed dynamical system in accordance with the 2-D LMS [19] such that the correlation information of the image pixel can be efficiently exploited in both directions. We have then considered the convergence analysis for this algorithm following a way of analysis similar to the one adopted in Chapter 3 for the doubly indexed LMS.

Theoretical comparison with 2-D doubly indexed LMS has revealed that both algorithms require similar bounds on the step size parameters to guarantee convergence of the mean; however, the 2DDI-BLMS offer improvement in the misadjustment proportional to the used block size. Experimental results that support the validity of the obtained convergence analysis results have been also presented.

6.2 Suggestions for Future Work

1. For the bias removal algorithm presented in Chapter 3, the stability of the cascade structure can be improved if the scaling factor of the output error of each of the cascaded section can be selected such that it takes into account not only the power of the output error of the corresponding section but also the power of the output error of all other sections in the cascade structure.
2. The steady state MSE analysis for 2-D LMS adaptive FIR filters has been considered in this thesis using the independence assumption. However, as it has been shown in Chapter 4, for large step size values as well as for highly colored input signals, the error in estimating the misadjustment of the adaptive algorithm is very large. For future work it is worth to work on the extension of the iterative steady state analysis approach proposed

by Butterweck [5] to the 2-D case for the doubly indexed LMS. The approximation method for the weight-error correlation matrix that has been proposed in Chapter 4 can be used as initial solution.

3. For the 2-D doubly indexed block LMS adaptive filtering algorithm proposed in Chapter 5, one may try to search for an optimal choice of the scalar parameters f_h and f_v and the step size parameters μ_h and μ_v based on the value of vertical and horizontal variances of each of the input blocks.

Appendix A

2-D Signals and Systems Review

In this appendix we review some of the fundamental concepts of 2-D discrete signals and 2-D digital filters. The materials given in this appendix is adopted mostly from reference [1].

A.1 2-D Signals

Most signals can be classified into three categories [1]:

1. Analog signals: are continuous in both space and amplitude. Examples of such signals are: images, seismic, radar, and speech signals.

2. Discrete-space signals: are discrete in space and continuous in amplitude. A common way to generate discrete-space signals is by sampling analog signals. All the signals used in this thesis are discrete space signals.

A 2-D discrete-space signal will be represented by a function whose two arguments are integers. For example, a 2-D discrete-space signal $x(m, n)$ denotes a sequence which is defined for all integer values of m and n that

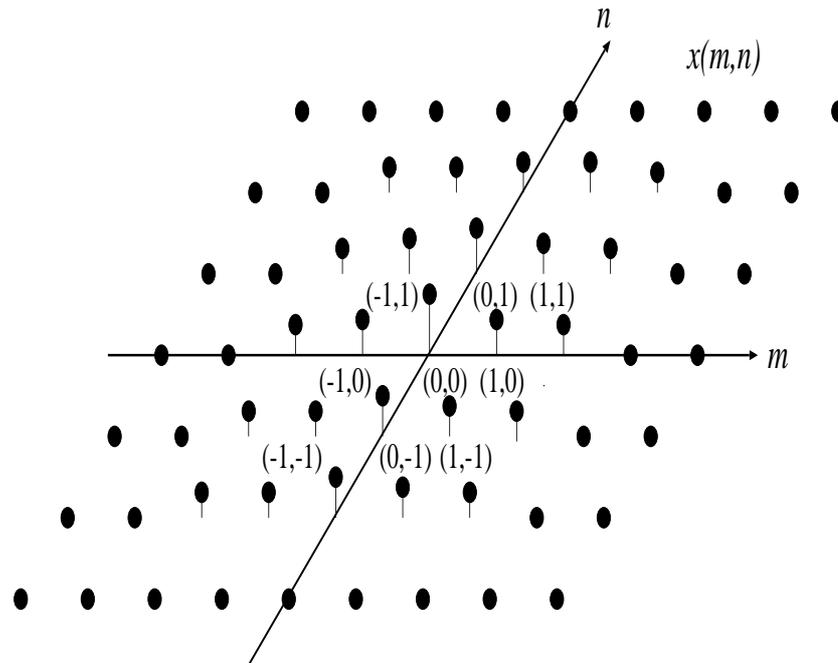


Figure A.1: 2-D discrete-space signals $x(m, n)$.

propagate in the horizontal and vertical directions, respectively. The sequence $x(m, n)$ is sketched in Figure A.1. In the figure, the height at (m, n) represents the amplitude at (m, n) .

3. Digital signals: are discrete in both space and amplitude. One way in which digital signals are created is by amplitude quantization of discrete-space signals. An example of 2-D digital signals is digital images.

A digital image, which can be denoted by $x(m, n)$ is typically obtained by sampling an analog image, for example, an image on film. The amplitude of digital images are often quantized to 256 levels (which can be represented by

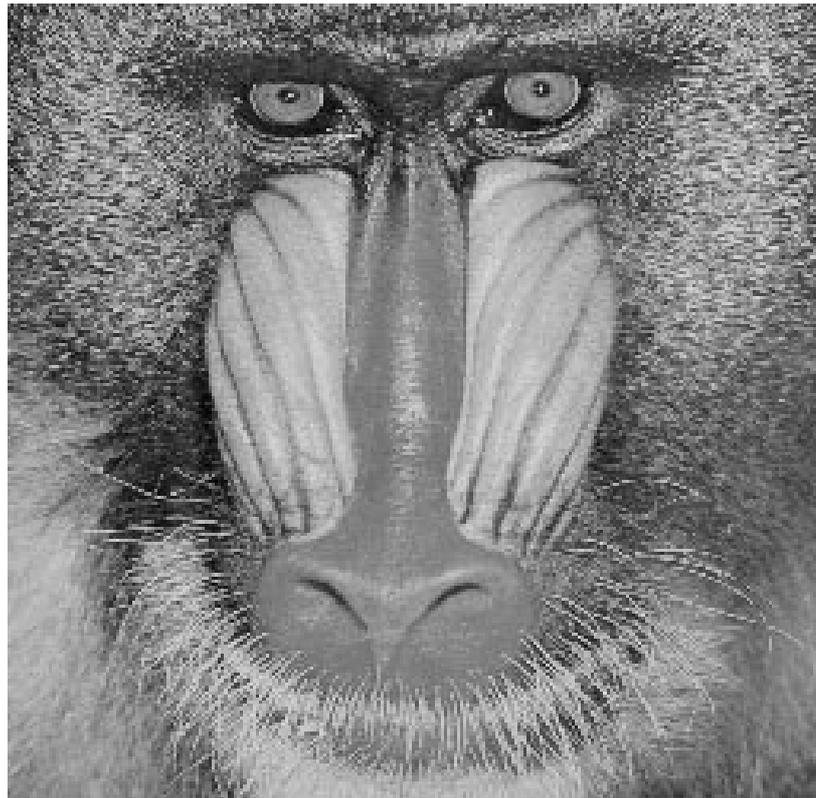


Figure A.2: Digital image of 256×256 pixels quantized at 8 bits/pixel.

eight bits). Each level is commonly denoted by integer, with 0 corresponding the darkest level and 255 to the highest. Each point (m, n) is called a pixel. Fig. A.2 shows a digital image of 256×256 pixels with each pixel represented by eight bits.

A.2 2-D Digital Systems

A 2-D system $T[\cdot]$ can be defined as an operator that transforms an input discrete sequence $x(m, n)$ to an output sequence $y(m, n)$. That is,

$$y(m, n) = T[x(m, n)]. \quad (\text{A.1})$$

where m and n are two integer indices

A.2.1 Linear and Shift-Invariant Systems

A 2-D system $T[\cdot]$ is said to be *linear* if it obeys the principle of superposition, that is

$$T[ax_1(m, n) + bx_2(m, n)] = ay_1(m, n) + by_2(m, n) \quad (\text{A.2})$$

where $y_1(m, n) = T[x_1(m, n)]$, $y_2(m, n) = T[x_2(m, n)]$, and a, b , are arbitrary constants.

A 2-D system $T[\cdot]$ is shift-invariant if it satisfies the following condition

$$T[x(i - p_1, j - p_2)] = y(i - p_1, j - p_2), \quad (\text{A.3})$$

where p_1 and p_2 are any integer values.

A.2.2 Causal Systems

If the impulse response $h(m, n)$ satisfies the condition

$$h(m, n) = 0 \quad \text{for } i < 0 \text{ or } j < 0, \quad (\text{A.4})$$

the 2-D digital system is said to be a causal system. If some independent variable of the signal to be processed corresponds to time, the causality is an important

constraint which should be imposed on 2-D digital system for real-time signal processing.

The relation between input $x(m, n)$ and output $y(m, n)$ of a causal 2-D digital filter can be simplified as

$$y(m, n) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} h(k_1, k_2)x(m - k_1, n - k_2) \quad (\text{A.5})$$

A.3 Difference Equation Representation of 2-D Systems

A linear shift invariant 2-D digital filter can be represented by the constant coefficient difference equation as follows:

$$y(m, n) = \sum_{\substack{i=0 \\ (i,j) \neq (0,0)}}^{N_1} \sum_{j=0}^{N_2} b(i, j)y(m - i, n - j) + \sum_{i=0}^P \sum_{j=0}^Q a(i, j)x(m - i, n - j) \quad (\text{A.6})$$

where N_1 and N_2 denote the horizontal and vertical filter order, respectively, $a(i, j)$ and $b(i, j)$ are constant coefficients.

If $b(0, 0) \neq 0$, we can normalize the coefficient array a and b by dividing both sides of equation (A.6) by $b(0, 0)$. This defines new coefficients arrays a and b , allowing us to assume that $b(0, 0) = 1$ without loss of generality.

For the general case where $N_1 \geq 1, N_2 \geq 1$, the difference equation (A.6) corresponds to the class of 2-D infinite impulse response (IIR) filters. The difference equation represents an algorithm for computing the sample of y at (m, n) under the assumptions that the required samples of the input are available and that those samples of y which appear on the right hand side of Eq. (A.6) have

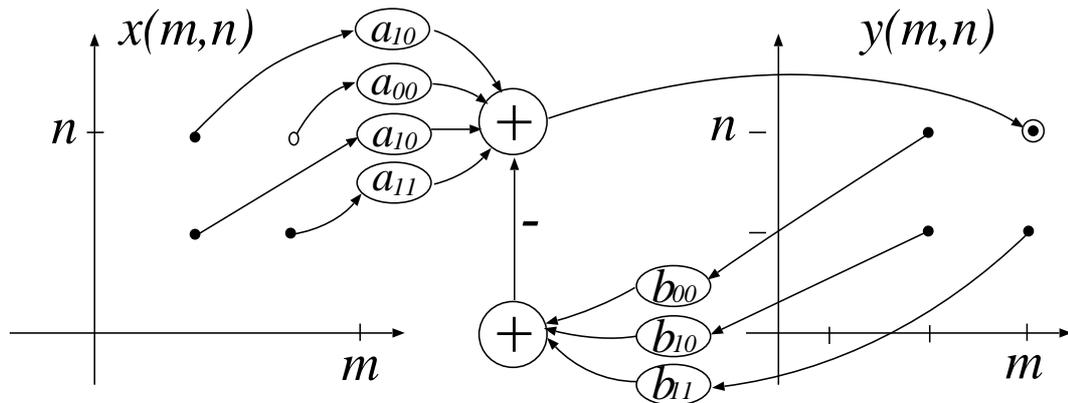


Figure A.3: Output of a 2-D IIR digital filter.

either been previously computed or have been specified as initial conditions. This equation is interpreted graphically in Fig. A.3.

The class of difference equation for which $b(i, j) = 1$ for $i = 0, j = 0$ and $b(i, j) = 0$ otherwise, is the special important case that corresponds to 2-D finite impulse response (FIR) filters. For this case, the difference equation (A.6) reduces to

$$y(m, n) = \sum_{i=0}^P \sum_{j=0}^Q a(i, j)x(m-i, n-j) \quad (\text{A.7})$$

Equation is interpreted graphically in Fig. A.4.

A.4 Transfer Function Representation of 2-D Systems

The transfer function of a system specified by a difference equation is the ratio of the z-transforms of the coefficient arrays $a(i, j)$ and $b(i, j)$. Since each of these

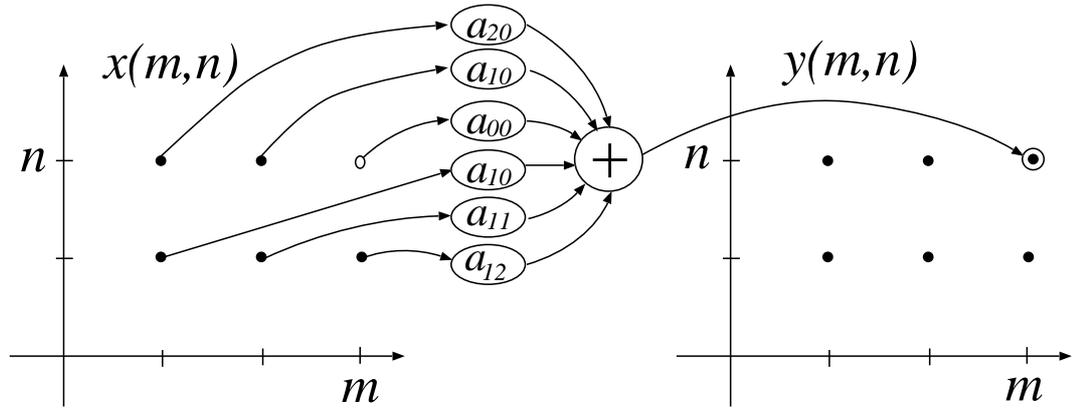


Figure A.4: 2-D FIR digital filter.

arrays has a finite area of support, their z-transforms are polynomials. For a system given by the difference equation (A.6), the transfer function is given by

$$H(z_1, z_2) = \frac{Y(z_1, z_2)}{U(z_1, z_2)} \quad (\text{A.8})$$

$$= \frac{\sum_{i=0}^P \sum_{j=0}^Q a(i, j) z_1^{-i} z_2^{-j}}{1 - \sum_{\substack{i=0 \\ (i,j) \neq (0,0)}}^P \sum_{j=0}^Q b(i, j) z_1^{-i} z_2^{-j}}. \quad (\text{A.9})$$

A.5 Stability of 2-D Systems

A system is considered to be stable in the bounded-input-bounded-output sense if and only if a bounded input always leads to bounded output. For linear shift invariant system, a necessary and sufficient condition for the system to be stable

is that its impulse response $h(l, k)$ be absolutely summable:

$$\sum_{l=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |h(l, k)| \leq \infty. \quad (\text{A.10})$$

Although this condition is a straightforward extension of the 1-D results, the 2-D case differs significantly from the 1-D case with regards to stability. In a typical 1-D stability problem we are given the system function $H(z) = A(z)/B(z)$. Since a 1-D polynomial $B(z)$ can be always factorized as a product of first-order polynomials, we can easily determine the poles of $H(z)$. The stability of the causal system is equivalent to having all the poles inside the unit circle.

The above approach can not be used in testing the stability of 2-D system. That approach requires the specific location of all poles to be determined. Partly because a 2-D polynomials $B(z_1, z_2)$ cannot in general be factored as a product of lower-order polynomials, its extremely difficult to determine all the pole surfaces of $H(z_1, z_2)$ and the approach based on explicit determination of all pole surfaces has not led to successful practical procedures for testing the system stability. Here, we present one representative stability theorem for 2-D systems which is called Shank 's theorem.

Shank's Theorem. Let $H(z_1, z_2) = 1/B(z_1, z_2)$ be a first quadrant recursive filter. This filter is stable if and only if $B(z_1, z_2) \neq 0$ for every point (z_1, z_2) such that $|z_1| \geq 1$ and $|z_2| \geq 1$.

A.6 2-D Separable Denominator Digital Filters

Consider a stable and causal 2-D digital filter with the following transfer function:

$$H(z_1, z_2) = \frac{A(z_1, z_2)}{B_1(z_1)B_2(z_2)} \quad (\text{A.11})$$

$$= \frac{\sum_{k_1=0}^P \sum_{k_2=0}^Q a_{k_1 k_2} z_1^{-k_1} z_2^{-k_2}}{(1 - \sum_{k_1=1}^P b_1^{(k_1)} z_1^{-k_1})(1 - \sum_{k_2=1}^Q b_2^{(k_2)} z_2^{-k_2})} \quad (\text{A.12})$$

The transfer function $H(z_1, z_2)$ can be viewed as the cascade of 2-D nonseparable FIR filter $A(z_1, z_2)$ and a 2-D separable all-pole IIR filter $1/B_1(z_1)B_2(z_2)$. This class of separable denominator IIR filters retains much of the flexibility of nonseparable filters and yet offer implementation advantages of separable IIR filters. If we neglect the numerator, which may be implemented separately as an FIR filter, then the remaining part of the filter is a separable filter. The implementation of this part of the filter can be structured as a set of 1-D convolutions on the rows of the signal array followed by another set of 1-D convolution on the columns of the resulting signal array. The entire implementation of the separable denominator filter is shown in Fig. A.5.

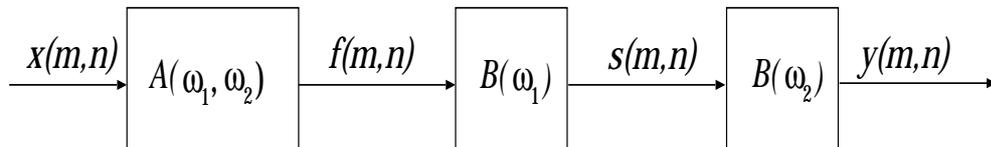


Figure A.5: Cascade implementation of separable denominator IIR filter.

Appendix B

Probability Review

This appendix summarizes the fundamentals of random process which are necessary in understanding the convergence analysis of the adaptive algorithms presented in this thesis. The materials given in this appendix is adopted mostly from reference [1].

B.1 Random Variables

A real random variable x is a variable that takes on real values at random, for instance, from the outcome of flipping a coin. It is completely characterized by its probability density function $p_x(x_0)$. The subscript x in $p_x(x_0)$ denotes the random variable x , and x_0 is a dummy variable that denotes a specific value of x . The probability that x will lie between a and b is given by

$$Prob[a \leq x \leq b] = \int_{x_0=a}^b p_x(x_0) dx_0. \quad (\text{B.1})$$

Since an event that is certain to occur is assumed to have a probability of 1,

$$Prob[-\infty \leq x \leq \infty] = \int_{x_0=-\infty}^{\infty} p_x(x_0) dx_0 = 1. \quad (\text{B.2})$$

The expectation of a function of a random variable x , $E\{f(x)\}$, is defined by

$$E\{f(x)\} = \int_{x_0=-\infty}^{\infty} f(x_0)p_x(x_0)dx_0. \quad (\text{B.3})$$

The expectation defined above is a linear operator and satisfies

$$E\{f(x) + g(x)\} = E\{f(x)\} + E\{g(x)\} \quad (\text{B.4})$$

$$E\{cf(x)\} = c E\{f(x)\} \quad (\text{B.5})$$

where c is a scalar constant.

The n th moment of a random variable x , $E\{x^n\}$, is defined by

$$E\{x^n\} = \int_{x_0=-\infty}^{\infty} x_0^n p_x(x_0)dx_0. \quad (\text{B.6})$$

The first moment of x is called the mean or average of x . From (B.6),

$$E\{x\} = \int_{x_0=-\infty}^{\infty} x_0 p_x(x_0)dx_0. \quad (\text{B.7})$$

The variance of X , $\text{var}\{x\}$, is defined by

$$\text{var}\{x\} = E\{(x - [E\{x\}]^2)\} = E\{x^2\} - E^2\{x\}. \quad (\text{B.8})$$

The standard deviation of x , σ_x is defined by

$$\sigma_x = (\text{var}\{x\})^{1/2} \quad (\text{B.9})$$

Two random variables x and y , are completely characterized by their joint probability density function $p_{x,y}(x_0, y_0)$. They are said to be *statistically independent* if they satisfy

$$p_{x,y}(x_0, y_0) = p_x(x_0)p_y(y_0) \quad \text{for all}(x_0, y_0). \quad (\text{B.10})$$

The expectation of a function of two random variables, $E\{f(x, y)\}$, is defined by

$$E\{f(x, y)\} = \int_{x_0=-\infty}^{\infty} \int_{y_0=-\infty}^{\infty} f(x_0, y_0)p_x(x_0)p_y(y_0)dx_0dy_0. \quad (\text{B.11})$$

Two random variables x and y , are said to be *linearly independent* if

$$E\{x y\} = E\{x\}E\{y\}. \quad (\text{B.12})$$

Statistical independence implies linear independence, but linear independence does not imply statistical independence.

The probability density function of a random variable x given (conditioned on) another random variable y is denoted by $p_{x|y}(x_0|y_0)$, and is defined by

$$p_{x|y}(x_0|y_0) = p_{x,y}(x_0, y_0)|p_y(y_0) \quad (\text{B.13})$$

If x and y are statistically independent, knowing y does not tell us anything about x , and $p_{x|y}(x_0|y_0)$ reduces to $p_x(x_0)$.

B.2 Random Processes

A collection of an infinite number of random variables is called a random process.

If the random variables are real, the collection is called a *real random process*.

Let us denote an infinite number of real random variables by $x(n_1, n_2)$, where $x(n_1, n_2)$, for a particular (n_1, n_2) is a real random variable. The random process $x(n_1, n_2)$ is completely characterized by the joint probability density function of all the random variable. If we obtain one sample, or realization, of the random process $x(n_1, n_2)$, the result will be a 2-D sequence. We will refer to this 2-D sequence as a random signal and we will denote it also by $x(n_1, n_2)$.

The *auto-correlation* function, or for short, the *correlation* function of the random process $x(n_1, n_2)$, $R_x(n_1, n_2; k_1, k_2)$ is defined by

$$R_x(n_1, n_2; k_1, k_2) = E\{x(n_1, n_2)x(k_1, k_2)\}. \quad (\text{B.14})$$

The correlation is the expectation of the product of two random variables, $x(n_1, n_2)$ and $x^*(k_1, k_2)$. The *auto-covariance function*, or the *covariance function*, for short, of $x(n_1, n_2)$, $\gamma_x(n_1, n_2; k_1, k_2)$, is defined by

$$\begin{aligned} \gamma_x(n_1, n_2; k_1, k_2) &= E\{[x(n_1, n_2) - E(x(n_1, n_2))][x(k_1, k_2) - E(x(k_1, k_2))]\} \\ &= E\{x(n_1, n_2)x(k_1, k_2)\} - E\{x(k_1, k_2)\}E\{x(n_1, n_2)\} \quad (\text{B.15}) \end{aligned}$$

$$= R_x(n_1, n_2; k_1, k_2) - E\{x(n_1, n_2)\}E\{x(k_1, k_2)\}. \quad (\text{B.16})$$

A random process $x(n_1, n_2)$ is called a zero-mean process if

$$E\{x(n_1, n_2)\} = 0 \quad \text{for all } (n_1, n_2). \quad (\text{B.17})$$

For a zero-mean random process,

$$R_x(n_1, n_2; k_1, k_2) = \gamma_x(n_1, n_2; k_1, k_2). \quad (\text{B.18})$$

A random process $x(n_1, n_2)$ with nonzero mean can always be transformed to a zero-mean random process by subtracting $E\{x(n_1, n_2)\}$ from $x(n_1, n_2)$. Unless specified otherwise, we will assume that $x(n_1, n_2)$ is a zero-mean process and (B.18) is valid.

A random process $x(n_1, n_2)$ is said to be *stationary* or *homogeneous in the strict sense* if the joint probability density function does not depend on the origin

of the index (n_1, n_2) :

$$P_{x(n'_1, n'_2), x(n''_1, n''_2), \dots}(x_1, x_2, \dots) = P_{x(n'_1+k+1, n'_2+k_2), x(n''_1+k_1, n''_2+k_2), \dots}(x_1, x_2, \dots) \quad (\text{B.19})$$

For any fixed k_1 and k_2 . For a stationary random process $x(n_1, n_2)$, $E\{x(n_1, n_2)\}$ is a constant independent of n_1 and n_2 , and $R_x(n_1, n_2; k_1, k_2)$ is a function of only $n_1 - k_1$ and $n_2 - k_2$:

$$E\{x(n_1, n_2)\} = m_x \quad \text{for all } (n_1, n_2) \quad (\text{B.20})$$

$$\begin{aligned} R_x(n_1, n_2; k_1, k_2) &= R_x(n_1 - k_1, n_2 - k_2; 0, 0) \\ &= E\{x(n_1 - k_1, n_2 - k_2)x(0, 0)\}. \end{aligned} \quad (\text{B.21})$$

Rewriting (B.21), we obtain

$$R_x(n_1, n_2) = E\{x(k_1, k_2)x(k_1 - n_1, k_2 - n_2)\} \quad \text{for all } (k_1, k_2) \quad (\text{B.22})$$

Note that the arguments n_1 and n_2 in $R_x(n_1, n_2)$ in (B.22) are $k_1 - n_1$ subtracted from k_1 and $k_2 - n_2$ subtracted from k_2 .

A stationary random process $x(n_1, n_2)$ is said to be *ergodic* if the time (or space) average equals the ensemble average. Suppose we wish to estimate $m_x = E[x(n_1, n_2)]$ from realization, or samples, of a stationary $x(n_1, n_2)$. Since m_x represents an ensemble average, we need an ensemble (an entire collection of all possible outcomes) of $x(n_1, n_2)$ for any particular (n_1, n_2) . If the random process is ergodic, then m_x can be computed from one realization of $x(n_1, n_2)$ by

$$m_x = E\{x(n_1, n_2)\} = \lim_{N \rightarrow \infty} \frac{1}{(2N + 1)^2} \sum_{n_1=-N}^N \sum_{n_2=-N}^N x(n_1, n_2). \quad (\text{B.23})$$

Similarly, for an ergodic process,

$$\begin{aligned} R_x(n_1, n_2) &= E\{x(k_1, k_2)x(k_1 - n_1, k_2 - n_2)\} \\ &= \lim_{N \rightarrow \infty} \frac{1}{(2N + 1)^2} \sum_{k_1 = -N}^N \sum_{k_2 = -N}^N x(k_1, k_2)x(k_1 - n_1, k_2 - n_2). \end{aligned} \quad (\text{B.24})$$

Equations (B.23) and (B.24) allow us to determine m_x or $R_x(n_1, n_2)$ from one realization of $x(n_1, n_2)$. Note that ergodicity implies stationary (in the wide sense), but stationarity does not imply ergodicity.

The *power spectrum* of a stationary random process $x(n_1, n_2)$, $P_x(n_1, n_2)$, is defined by

$$P_x(\omega_1, \omega_2) = F[R_x(n_1, n_2)] = \sum_{n_1 = -\infty}^{\infty} \sum_{n_2 = -\infty}^{\infty} R_x(n_1, n_2) e^{-j\omega_1 n_1} e^{-j\omega_2 n_2}. \quad (\text{B.25})$$

From (B.22) and (B.25),

$$\begin{aligned} R_x(0, 0) &= \sigma_x^2 = E[x(n_1, n_2)x(n_1, n_2)] \\ &= \frac{1}{(2\pi)^2} \int_{\omega_1 = -\pi}^{\pi} \int_{\omega_2 = -\pi}^{\pi} P_x(\omega_1, \omega_2) d\omega_2 d\omega_1. \end{aligned} \quad (\text{B.26})$$

It can also be shown that

$$R_x(0, 0) \geq |R_x(n_1, n_2)| \quad \text{for all } (n_1, n_2). \quad (\text{B.27})$$

The value σ_x^2 is called the average power of the random process $x(n_1, n_2)$.

A random process is called a *white noise process* if

$$\begin{aligned} R_x(n_1, n_2; k_1, k_2) &= E\{x(n_1, n_2)x(k_1, k_2)\} \\ &= \begin{cases} \sigma_x^2(n_1, n_2), & n_1 = k_1, n_2 = k_2 \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (\text{B.28})$$

For a stationary white noise process, then,

$$\begin{aligned} R_x(n_1, n_2) &= E\{x(k_1, k_2)x(k_1 - n_1, k_2 - n_2)\} \\ &= \sigma_x^2 \delta(n_1, n_2) \end{aligned} \quad (\text{B.29})$$

From (B.25) and (B.29), the power spectrum of a stationary white noise process is given by

$$P_x(\omega_1, \omega_2) = \sigma_x^2 \quad \text{for all } (\omega_1, \omega_2). \quad (\text{B.30})$$

The power spectrum is constant for all frequencies; hence the term “white.”

For stationary process $x(n_1, n_2)$ and $y(n_1, n_2)$,

$$R_{xy}(n_1, n_2) = E\{x(k_1, k_2)y(k_1 - n_1, k_2 - n_2)\} \quad \text{independent of } (k_1, k_2). \quad (\text{B.31})$$

For ergodic processes $x(n_1, n_2)$ and $y(n_1, n_2)$,

$$\begin{aligned} R_{xy}(n_1, n_2) &= E\{x(k_1, k_2)y(k_1 - n_1, k_2 - n_2)\} \\ &= \lim_{N \rightarrow \infty} \frac{1}{(2N + 1)^2} \sum_{k_1=-N}^N \sum_{k_2=-N}^N x(k_1, k_2)y(k_1 - n_1, k_2 - n_2). \end{aligned} \quad (\text{B.32})$$

Bibliography

- [1] J. S. Lim, Two-Dimensional Signal and Image Processing, Prentice-Hall, New Jersey, 1990.
- [2] D. E. Dudgeon, R. Mersereau, Multidimensional Digital Signal Processing, Prentice Hall 1994.
- [3] W. S. Lu, A. Antoniou, Two-Dimensional Digital Filters, Marcel Dekker, New York, 1992.
- [4] M. Kawamata, T. Higuchi, Multidimensional Digital Signal Processing, Asakura Shoten, Tokyo, 1995.
- [5] S. Haykin, Adaptive Filter Theory, NJ: Prentice-Hall, Third edition, 1996.
- [6] J. Abramatic, L. M. Silveram, "Nonlinear Restoration of Noisy Images," IEEE Trans. Acoust., Speech, Signal Processing, PAMI-4, no. 2, pp. 141-149, 1982.
- [7] H. E. Knutsson, R. Wilson, G. H. Granlund, "Anisotropic Nonstationary Image Estimation and It Applications:Part I - Restoration of Noisy Images," IEEE Trans. Communications, vol. Com-31, no. 3, pp. 141-149, 1983.
- [8] K. R. Castleman, Digital Image Processing, Prentice Hall, 1996.

- [9] T. Hinamoto, N. Hamada, M. Kawamata, A. Taguchi, T. Muraoka, Two-Dimensional Signal and Image Processing, SICE, 1996.
- [10] V. Solo, X. Kong, Adaptive Signal Processing Algorithms, NJ: Prentice-Hall, 1995.
- [11] W. K. Jenkins, A. W. Hull, J. C. Strait, B. A. Schnaufer, and X. Li, Advanced Concepts in Adaptive Signal Processing, Kluwer Academic Publishers, 1996.
- [12] J. N. Lin, R. Unbehauen, "Two-Dimensional LMS Adaptive Filter Incorporating a Local-Mean Estimator for Image Processing," IEEE Trans. Circuits Syst. II, pp. 417-428, July. 1993.
- [13] J. J. Shynk, "Adaptive IIR Filtering," IEEE ASSP Mag., pp. 4-21, April 1989.
- [14] C. R. Johnson, JR, "Adaptive IIR Filtering: Current Results and Open Issues," IEEE Trans. Inform. Theory, vol. IT-30, pp. 237-250, Mar. 1984.
- [15] K. Toshima, M. Ohki, X. Zhou, H. Hashiguchi, "2-D Backpropagation Cascade Adaptive Recursive Filter with the Separable Denominator Function," Third International Symposium on Consumer Electronics, vol. 2, pp. 441-446, Nov. 1994.
- [16] F. X. Y. Gao, W. M. Snelgrove, "An Adaptive Backpropagation Cascade IIR Filter," IEEE Trans. Circuits Syst. II, vol. CAS-39, no. 9, pp. 606-610, Sept. 1992.

- [17] M. M. Hadhoud D. W. Thomas, "The Two-Dimensional Adaptive LMS (TDLMS) Algorithm," *IEEE Trans. Circuits Syst.*, pp. 485-494, May. 1988.
- [18] B. Widrow, S. D. Stearn, *Adaptive Signal Processing*, NJ:Prentice-Hall, 1985.
- [19] M. Ohki, S. Hashiguchi, "Two-Dimensional LMS Adaptive FIR Filters," *IEEE Trans. Consumer Electronics*, vol. 37, no. 1, pp. 66-73, Feb. 1991.
- [20] E. Fornasini, G. Marchesini, "Stability Analysis of 2-D Systems," *IEEE Trans. Circuits Syst.*, vol. CAS-27, pp. 2085-2088, Dec. 1980.
- [21] G. A. Clark, S. Mitra, S. R. Parker, "Block Implementation of Adaptive Digital Filters," *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-29, no. 12, pp. 744-752, 1986.
- [22] M. R. Azimi-Sadjadi, H. Pan, "Two-Dimensional Block Diagonal LMS Adaptive Filtering," *IEEE Trans. Signal Processing*, vol. 42, no. 9, pp. 2420-2429, 1994.
- [23] W. Mikhael, S. M. Gosh, "Two-Dimensional Block Adaptive Filtering Algorithm," in *Proc. IEEE Int. Symp. Circuits and Syst.*, San Diego, CA, pp. 1219-1222, May 1992.
- [24] T. Wang, C. L. Wang, "A New Two-Dimensional Block FIR Filtering Algorithm," in *Proc. IEEE Int. Symp. Circuits and Syst.*, San Francisco, CA, pp. 1219-1222, May 1992.
- [25] B. Widrow, J. M. McCool, M.G. Larimore, C. R. Johnson, "Stationary and

- Nonstationary Learning Characteristic of the LMS Adaptive Filter,” Proceedings of the IEEE, vol. 64, no. 8, pp. 1151-1162, Aug. 1976.
- [26] L. L. Horowitz and K. D. Senne, “Performance Advantage of Complex LMS for Controlling Narrow-band Adaptive Array,” IEEE Trans. Circuits Syst., vol. CAS-28, pp. 562-576, Jun. 1981.
- [27] M. A. Feuer, E. Weinstein, “Convergence Analysis of LMS Filters with Uncorrelated Gaussian Data,” IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-33, pp. 222-230, Feb. 1985.
- [28] S. Florian, A. Feuer, “Performance Analysis of the LMS Algorithm with a Tapped Delay Line (Two Dimensional Case),” IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 1542-1549, Dec. 1986.
- [29] S. D. Douglas, W. Pan, “Exact Expectation Analysis of the LMS Adaptive Filter,” IEEE Trans. Signal Processing, vol. 43, no. 12, pp. 2863-2871, Dec. 1995.
- [30] M. Shadaydeh, M. Kawamata, “Convergence Analysis of Two-Dimensional LMS FIR Filters,” Proc. of the Thirty First Asilomar Conference on Signals, Systems, and Computers, California, pp. 348-353, Nov. 1997,
- [31] M. Shadaydeh, M. Kawamata, “Bias Removal Algorithm for 2-D Equation Error Adaptive IIR Filters,” IEEE International Workshop on Intelligent Signal Processing and Communication Systems (ISPACS'97), Malaysia, pp. 16.3.1-16.3.6, Nov. 1997.

- [32] M. Shadaydeh, M. Kawamata, "Convergence Analysis of 2-D Doubly-Indexed Block LMS Adaptive Filters," IEEE International Workshop on Intelligent Signal Processing and Communication Systems (ISPACS'98), Melbourne, pp. 782-786, Nov. 1998.
- [33] M. Muneyasu, E. Uemoto, N. Ishizaki, T. Hinamoto, "A Two-Dimensional Cascade IIR Adaptive Filter Using Backpropagation Method," Proceedings of the 1994 IEICE Spring Conference, p. D-167, Mar. 1994.
- [34] M. Muneyasu, E. Uemoto, and T. Hinamoto, "A Novel Type of 2-D Cascade-Parallel Adaptive IIR Filters," Journal of Franklin Institute, vol. 333 (B), no. 4, pp. 453-465, 1996.
- [35] M. Shadaydeh, M. Kawamata "Bias Removal Algorithm for 2-D Equation Error Adaptive IIR Filters," Proc. 5th IEEE International Workshop on Intelligent Signal Processing and Communication Systems, Malaysia, Nov. 1997.
- [36] J. N. Lin, R. Unbehauen, "Bias-Remedy Least Mean Square Equation Error Algorithm for IIR Parameter Recursive Estimation," IEEE Trans. Signal Processing, vol. 40, no. 1, pp. 62-69, Jan. 1992.
- [37] S. Mitra, J. F. Kaiser, Handbook for Digital Signal Processing, Wiley-Interscience Publication, 1993.
- [38] Z. Gajić, M. T. J. Qureshi, Lyapunov Matrix Equation in System Stability and Control, Academic Press, 1995.

- [39] F. X. Y. Gao, Adaptive Filtering, Cascade IIR Filters, Linearization Techniques, and Nonlinear Recursive State Space Filters, Ph.D thesis, University of Toronto, 1991.
- [40] A. C. Tan, S. T. Chen, S. Basu, "Image Filtering Using Hyperstable Adaptive Algorithms," IEEE Trans. Circuits Syst. II, vol. CAS-44, no. 5, pp. 358-370, May. 1997.
- [41] W. S. Lu, A. Antoniou, Two-Dimensional Digital Filters, Marcel Dekker, New York, 1992.
- [42] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, 1984.
- [43] G. Zelniker, F. J. Taylor, Advanced Digital Signal Processing: Theory and Applications, Dekker, 1994.
- [44] E. C. Zachmanoglou, D. W. Thoe, Introduction to Partial Differential Equations with Applications, Dover, 1986.
- [45] J. W. Dettman, Mathematical Methods in Physics and Engineering, Dover, 1969.

List of Publications

Journal Papers

- (1) Y. Xiao, M. Shadaydeh, and Y. Tadokoro, "Overdetermined $C(q,k)$ formula using the third and fourth order cumulants," *Electronics Letters (UK)*, vol. 32, no. 6, pp 601-603, Mar. 1996.
- (2) M. Shadaydeh, Y. Xiao, and Y. Tadokoro, "Identification of nonminimum phase FIR systems via fourth order cumulants and genetic algorithm," *Signal Processing*, vol. 60, no. 3, pp. 339-347, Aug. 1997.
- (3) M. Shadaydeh and M. Kawamata, "Steady state analysis of two-dimensional LMS adaptive filters using the independence assumption," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, to be published.
- (4) M. Shadaydeh and M. Kawamata, "Bias Removal Algorithm for 2-D Equation Error Adaptive IIR Filters," *Multidimensional Systems and Signal Processing*, to be published.

International Conferences

- (1) M. Shadaydeh and M. Kawamata, "Convergence Analysis of Two-Dimensional LMS FIR Filters," *Thirty First Asilomar Annual Conference on Signals, Systems, and Computers*, Nov. 1997, California, pp. 348-353.

- (2) M. Shadaydeh and M. Kawamata, "Bias Removal Algorithm for 2-D Equation Error Adaptive IIR Filters," IEEE International Workshop on Intelligent Signal Processing and Communication Systems (ISPACS'97), Nov. 1997, Malaysia, pp. 16.3.1-16.3.6..
- (3) M. Shadaydeh and M. Kawamata, "Convergence Analysis of 2-D Doubly-Indexed Block LMS Adaptive Filters," IEEE International Workshop on Intelligent Signal Processing and Communication Systems (ISPACS'98), Nov. 1998, Melbourne, pp. 782-786.

Domestic Conferences

- (1) M. Shadaydeh, Y. Xiao, and Y. Tadokoro, "Moving Average System Identification Using Overdetermined $C(q,k)$ Formula," Toukai-section Joint Convention Record of the Six Institutes of Electrical and Related Engineers, Oct. 1994, Japan, p. 264.
- (2) Y. Xiao, M. Shadaydeh and Y. Tadokoro, "Moving Average system identification using fourth-order cumulants", Presented at IEICE Spring Conference, Mar. 1995, p.168, Japan.
- (3) M. Shadaydeh, Y. Xiao, and Y. Tadokoro, "Moving Average System Identification Using Fourth-order Cumulants," Technical Reports of IEICE, DSP95-111, Oct. 1995, Japan, pp. 55-62.
- (4) M. Shadaydeh and M. Kawamata, "On the Convergence of 2-D LMS Algorithm for Backpropagation Adaptive IIR Filters," Technical Reports of

- IEICE, CS96-138, Dec. 1996, Japan, pp. 61-66.
- (5) M. Shadaydeh and M. Kawamata, "Steady State Mean Square Error Analysis for 2-D LMS FIR Filters," Presented at IEICE Spring Conference, Mar. 1998, p.161, Japan.
- (6) M. Shadaydeh and M. Kawamata, "Steady state analysis of two-dimensional LMS adaptive filters using the independence assumption," Proceedings of the 11th Workshop on Circuits and Systems (Karuizawa), April 1998, Japan, pp. 309-314.
- (7) M. Shadaydeh and M. Kawamata, "2-D Doubly-Indexed Block LMS Adaptive Filters," Presented at IEICE Autumn Conference, Sept. 1998, p.59, Japan.

Acknowledgments

I would like to express my heartfelt appreciation to Professor Masayuki Kawamata of the Department of Electronic Engineering for his capable supervision in this research and unforgettable kind help over the course of my doctoral research in Tohoku University.

I am also indebted to Professor Tatsuo Higuchi of the Graduate School of Information Sciences and Professor Kenichi Abe of the Department of Electrical Communications Engineering, who have served as members of my doctoral dissertation examination committee, for their highly valuable comments and enlightening discussions.

I am specially grateful to Professor Yoshiaki Tadokoro at Toyohashi University of Technology, for recommending me for doctoral course of Tohoku University. His unforgettable encouragement and continuous care about my study life in Japan are deeply appreciated.

In addition, I take this opportunity to express my sincere appreciation to Associate Prof. Yegui Xiao at Saga University for his continuous kind help and unfailing encouragement.

I also would like to thank Associate Prof. George Stoyanov at Technical University of Sofia for his frank advice and useful discussions during his stay in Kawamata laboratory.

Finally, I wish to express my sincere appreciation to all members of Kawamata Laboratory for their close collaborations and kind assistances.

Index

Symbol	
2-D block mean square error	97
1-D block LMS	95
1-D indexing scheme	15
1-D LMS	15
2-D adaptive differential pulse code modulation	28
2-D adaptive filters	13
2-D block LMS	95
2-D correlation information	15
2-D doubly-indexed LMS	16
2-D finite impulse response (FIR) fil- ter	121
2-D infinite impulse response (IIR) filter	120
2-D LMS	15
2-D noise cancellation	28
2-D system identification	28
2-D Wiener filter	20
2DBRA	17, 35, 39
2DDI-BLMS	17, 99
A	
adaptation accuracy	107
averaged system	84
B	
backpropagation concept	34
backpropagation IIR filter	17
bias removal algorithm	14
biased parameter estimates	35
binary input signal	90
biomedicine	28
block LMS	17
block size	107
block-wise processing	16
blurring	13
bounded-input	122
bounded-output	122
C	
cascade IIR filter	17

- cascade structure.....33
 causal filter 69
 computational complexity 13
 convergence analysis 99
 convergence behaviour 15, 42
 convergence in the MSE.....17
 convergence of the mean 24, 71
 convergence of the variance.....18
 convergence rate 68
 convergence speed.....13
 correlation 14
 correlation matrix.....76
 coupled difference equations 15
- D
- data compression.....19
 denominator horizontal section ... 38
 denominator vertical section.....42
 difference equation 15, 42
 digital signal processing 12
 direct averaging method.....84
 doubly-indexed dynamical system 15
 dynamical behaviour 14
- E
- equation error33
 equation error formulation 25
 error-feedback.....35
 external noise 32
- F
- F-M state space model 15
 Fast Fourier Transform.....95
 feedback control system 31
 FFT 95
 filter design 13
 first order difference equation 68
 Fornasini and Marchesini (F-M) .. 67
- G
- Gaussian noise 99
 Gaussian-distribution 72
 global minima.....27
 gray level image.....29
- H
- high speed integrated circuits 12
 hyperstable adaptive IIR Filters.113
- I
- I/O stability 17
 image enhancement 12
 image restoration 13

- image statistics 19
 immigration of the poles 32
 independence assumption.....17, 67
 Input-output (I/O).....17
 instability.....31
 instantaneous gradient 68
 intermediate errors.....33

J
 jointly Gaussian.....67

L
 learning curve.....108
 least mean square (LMS).....22
 linear simultaneous equations 69
 linear system.....119
 local minima 26
 low pass characteristic 19

M
 measurement noise 35
 minimum mean square error12
 misadjustment 87
 MSE analysis 16
 multidimensional adaptive filters . 12
 multimedia technology 12

N
- non-quadratic MSE surface.....26
 nonlinear optimization 42
 nonstationary signal.....14
 number of additions.....32
 number of multiplication 32

O
 off-line method.....13
 optimal filtering.....19
 orthogonal matrix.....76
 orthogonal transform.....77
 output error formulation 25
 output-error 35
 output-error feedback 36

P
 parallel computation 16
 parallel-cascade IIR filtering.....34
 parameter sensitivity.....33
 parameter update 38
 parameter vector38
 parameter-error vector 42
 performance criteria.....25
 performance measure.....31
 probability distribution 86

R

- random variables 72
 regressor vector 38
 robustness 32
 roundoff noise 36

S
 scaling factor 36
 scanning scheme 14
 Schwartz' inequality 78
 second order difference equations 68
 seismology 28
 separable denominator IIR filter 36,
 124
 separable denominator transfer func-
 tion 17
 series expansion 85
 Shank 's theorem 123
 shift-invariant systems 119
 spatial delay operator 37
 spatial indices 68
 stability analysis 42
 stability monitoring 14
 stability performance 16
 stability theory 67
 state space model 67
 stationary random signal 78

 statistically independent random vari-
 ables 72
 steady state analysis 66
 steady state BMSE 101
 steady state MSE 67
 steepest descent algorithm 22
 step size parameter 42

T
 TDLMS 66
 telecommunications 28
 tracking ability 14
 transfer function 85
 transformed weight-error correlation
 matrix 78
 transversal section 38
 two dimensional digital filters 12

U
 unimodal mean square error surface
 33
 upper bounds on the step size param-
 eters 18

W
 WECM 15, 17
 weight error correlation matrix ... 15

weight-error covariance matrix . . .	69
white Gaussian noise	72
Wiener-Hopf equation	20
word-length	29