

**EFFICIENT MAXIMUM LIKELIHOOD PARAMETER
LEARNING: IMAGE and RADAR APPLICATIONS**

**Ph.D. Thesis by
Faruk SARI, M.Sc.**

Department : Electronics and Communication Engineering

Programme: Electronics and Communication Engineering

OCTOBER 2006

**EFFICIENT MAXIMUM LIKELIHOOD PARAMETER
LEARNING: IMAGE and RADAR APPLICATIONS**

**Ph.D. Thesis by
Faruk SARI, M.Sc.
494.D.001**

Date of submission : 22 September 2003

Date of defence examination: 10 October 2006

**Supervisor (Chairman): Assoc.Prof. Dr. Mehmet Ertuğrul
ÇELEBİ**
**Members of the Examining Committee Prof.Dr. Erdal PANAYIRCI (Kadir Has
University)
Prof.Dr. Ahmet Hamdi KAYRAN (İTÜ)
Prof.Dr. Cüneyt GÜZELİŞ (Dokuz Eylül
University)
Prof.Dr. Tayfun AKGÜL (İTÜ)**

OCTOBER 2006

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**ETKİN ENBÜYÜK OLABİLİRLİK PARAMETRE
ÖĞRENME: İMGE VE RADAR UYGULAMALARI**

DOKTORA TEZİ

Y. Müh. Faruk SARI

494.D.001

Tezin Enstitüye Verildiği Tarih : 22 Eylül 2003

Tezin Savunulduğu Tarih : 10 Ekim 2006

Tez Danışmanı: Doç.Dr. Mehmet Ertuğrul ÇELEBİ

**Diğer Jüri Üyeleri Prof.Dr. Erdal PANAYIRCI (Kadir Has
Üniversitesi)**

Prof.Dr. Ahmet Hamdi KAYRAN (İTÜ)

**Prof.Dr. Cüneyt GÜZELİŞ (Dokuz Eylül
Üniversitesi)**

Prof.Dr. Tayfun AKGÜL (İTÜ)

EKİM 2006

PREFACE

I ought to be immensely grateful to people who have helped me in this exciting venture. First of all, I would like to thank my supervisor, Assoc. Prof. Dr. M. Ertuğrul ÇELEBİ, for his interest in my work and for giving advice and direction to my studies. He introduced me to this subject for which I am really grateful.

I also wish to thank Prof. Dr. Lamine MILI from Virginia Tech, USA, for his encouragement and his faith in me to complete this thesis. Dr. MILI took a personal interest in me and made me feel that my work was important.

I would like to thank Prof. Dr. Simon HAYKIN for providing the sea clutter data from IPIX radar database.

In Ph.D. education phase of my life, some people inspired me directly or indirectly with hard working personality dedicated to their works and being loyal to the ethic rules. I appreciate all of them for their contribution to my progress.

Special thanks to my wife and my best friend, Nurşen Yıldız SARI, for her great support. My daughter, İrem, was a white noise of my studies. I thank her very much, she has also stimulated me greatly.

To all of you: I am grateful for your support and companionship.

OCTOBER 2006

Faruk SARI

P.S. We are now a family of four, welcoming our little son Asım Kerem.

CONTENTS

PREFACE	ii
CONTENTS.....	iii
LIST OF ABBREVIATIONS	vii
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF SYMBOLS	xiii
EFFICIENT MAXIMUM LIKELIHOOD PARAMETER LEARNING: IMAGE and RADAR APPLICATIONS.....	xvi
ETKİN ENBÜYÜK OLABİLİRLİK PARAMETRE ÖĞRENME: İMGE VE RADAR UYGULAMALARI	xix
1. INTRODUCTION.....	1
1.1. Motivations.....	1
1.2. Technical Contributions of This Thesis.....	3
1.2.1. FS Optimization in Direct ML Method.....	3
1.2.2. Trust Region FS Optimization in Direct ML Method.....	3
1.2.3. EM Based Learning for Recursive Processing.....	3
1.2.3.1. Closed Form Solution	4
1.2.3.2. Optimization Solution.....	4
1.2.4. EM Based Learning for Batch Processing.....	4
1.2.4.1. Closed Form Solution	4
1.2.4.2. Optimization Solution.....	4
1.2.4.3. Regularization.....	4
1.2.5. Radar Detection Application	5
1.3. Scope of the Thesis.....	5
2. MATHEMATICAL BACKGROUND and DEFINITIONS.....	8
2.1. Introduction.....	8

2.2.	Image and Observation Model	10
2.3.	State Space Image and Observation Models	12
2.3.1.	Hidden Markov Models and Dynamic Bayesian Networks	14
2.4.	Optimization Methods	16
2.4.1.	Line Search	19
2.4.1.1.	Gradient Descent Methods	19
2.4.1.2.	Newton Type Optimization	20
2.4.1.2.1.	<i>Quasi-Newton Methods</i>	21
2.4.1.2.2.	<i>Fisher Scoring Algorithm</i>	22
2.4.2.	Trust Region	23
2.4.2.1.	LM Method	24
2.4.3.	EM Algorithm	25
2.5.	Ill-Posed Problems	27
2.6.	Solving Equations in DFT Domain	28
2.7.	Concluding Remarks	31
3.	IMAGE RESTORATION AND IDENTIFICATION	32
3.1.	Measuring and Metrics	32
3.2.	Image Restoration Algorithms	33
3.2.1.	Inverse Filtering	35
3.2.2.	Least-Squares Filters	36
3.2.2.1.	The Wiener Filter	36
3.2.2.2.	Kalman Filter	37
3.2.3.	Regularized Image Restoration Approaches	38
3.2.3.1.	Deterministic Regularization	38
3.2.3.2.	Stochastic Regularization	39
3.2.4.	Blind Image Restoration	41
3.3.	Blur and Image Parameter Identification Problem	42
3.3.1.	Direct ML Blur and Image Parameter Identification	43
3.3.1.1.	Gradient Based Optimization	45
3.3.2.	Expectation Maximization (EM) Algorithm	47
3.3.2.1.	Convergence of the EM Algorithm	48
3.4.	Concluding Remarks	49
4.	NEWTON TYPE OPTIMIZATION of LOG-LIKELIHOOD FUNCTION	50
4.1.	The Method of Scoring for Blur and Image Identification	50
4.1.1.	Local Convergence Analysis of FS	52
4.2.	A New Trust Region Fisher Scoring Optimization	53
4.2.1.	Introduction	53

4.2.2.	Finding New Criteria	56
4.3.	Performance Evaluation Methodology	58
4.3.1.	Cramer-Rao Lower Bound	58
4.3.2.	Regularization Effects of the New RS-FS Method	59
4.3.3.	Computation Complexity	59
4.4.	Relations Between EM and Newton Type Method.....	60
4.5.	Experimental Results.....	61
4.5.1.	Experiments on Artificial Images.....	61
4.5.2.	Experiments on Real Images	64
4.6.	Concluding Remarks	70
5.	RECURSIVE LEARNING of IMAGE PARAMETERS and RESTORATION of IMAGES USING EM BASED LEARNING ALGORITHM 72	
5.1.	Introduction.....	72
5.2.	EM Learning on Linear Dynamic Systems	74
5.2.1.	E Step of the Algorithm.....	74
5.2.1.1.	Kalman Smoothing Algorithm and Restoration of Images	77
5.2.2.	M Step of the Algorithm	78
5.2.2.1.	Closed Form.....	78
5.2.2.2.	Optimization	81
5.3.	Experimental Results.....	81
5.3.1.	Experiments on Artificial Images.....	82
5.3.2.	Real Images	85
5.4.	Concluding Remarks	89
6.	A NEW ACCELERATED EM BASED LEARNING OF THE IMAGE PARAMETERS AND RESTORATION	90
6.1.	Introduction.....	90
6.2.	Learning of Image Parameters.....	91
6.2.1.	E-Step of the Algorithm	92
6.2.2.	M Step	94
6.2.2.1.	Closed Form.....	94
6.2.2.2.	Fast Optimization Solution	97
6.3.	Frequency Domain Solution	100
6.4.	Experimental Results.....	101
6.4.1.	Experiments on Artificial Images.....	101
6.4.2.	Real Images	105

6.5.	Concluding Remarks	113
7.	RADAR DETECTION APPLICATION: LEARNING OF THE SEA CLUTTER PARAMETERS FOR ADAPTIVE RADAR DETECTION.....	114
7.1.	Introduction.....	114
7.2.	Clutter Modeling.....	116
7.3.	ML Estimation of the Model Parameters.....	119
7.4.	Gradient Based Optimization	120
7.5.	FS Optimization	121
7.6.	The Probabilty of False Alarm	122
7.7.	Experimental Results.....	122
7.8.	Concluding Remarks	130
8.	CONCLUSIONS and SUGGESTIONS for FUTURE WORK.....	132
9.	REFERENCES.....	135
	APPENDIX A : DERIVATION OF (4.5-4.6)	156
	APPENDIX B : RTS RECURSION	158
	APPENDIX C: DERIVATION OF (5.14).....	159
	APPENDIX D: REQUIRED DERIVATIONS FOR (5.17-5.28).....	160
	APPENDIX E: THE DERIVATIVES OF $\Gamma(\theta)$ WITH RESPECT TO EACH UNKNOWN PARAMETERS.....	162
	APPENDIX F: REQUIRED DERIVATIONS FOR CHAPTER 6	163
	CURRICULUM VITAE.....	165

LIST OF ABBREVIATIONS

1-D	: One-Dimensional
2-D	: Two-Dimensional
3-D	: Three-Dimensional
AR	: Autoregressive
ARMA	: Autoregressive Moving Average
BFGS	: Broydon-Fletcher-Goldfarb-Shanno
BC	: Block Circulant
BCCB	: Block Circulant Circulant Block
BT	: Block Toeplitz
BTTB	: Block Toeplitz Toeplitz Block
CDF	: Cumulative Distribution Function
CFAR	: Constant False Alarm Rate
CG	: Conjugate Gradient
DA	: Deterministic Annealing
DAEM	: Deterministic Annealing Expectation Maximization
DBN	: Dynamic Bayesian Network
DFT	: Discrete Fourier Transform
DFP	: Davidon-Fletcher-Powell
EM	: Expectation Maximization
EKF	: Extended Kalman Filter
EKS	: Extended Kalman Smoothing
FIM	: Fisher Information Matrix
FS	: Fisher Scoring
GCV	: Generalized Cross Validation
GM	: Gaussian-Mixture
HMM	: Hidden Markov Model
I	: Inphase
IPIX	: Intelligent PIXel Processing Radar
KFM	: Kalman Filter Models
KL	: Kullback Leibler
KS	: Kolmogorov-Smirnov
LDS	: Linear Dynamical System
LM	: Levenberg Marquardt
LS	: Least Squares
LTI	: Linear Time Invariant
ML	: Maximum Likelihood
NR	: Newton Raphson
NSHP	: Non Symmetric Half Plane
pdf	: probability density function
PSF	: Point Spread Functions
Q	: Quadrature
QQ	: Quantile-Quantile

RADAR	: Radio Detection and Ranging
RLS	: Recursive Least Squares
RTS	: Rauch-Tung-Streibel
SA	: Simulated Annealing
SAR	: Synthetic Aperture Radar
SNR	: Signal-to-Noise Ratio
SNRimp	: SNR Improvement
SR	: Super Resolution
TR	: Trust-Region
VM	: Variable Metric
VQ	: Vector Quantizer.

LIST OF TABLES

	<u>Page Number</u>
Table 1.1 : Current and open problems in image and blur identification...	22
Table 4.1 : Identification of PSF coefficients, SNR=20 dB.....	82
Table 4.2 : Identification of blur (PSF) coefficients.....	84
Table 4.3 : Identification of AR image coefficients.....	85
Table 4.4 : Iteration Number and Running Time for 3 blur and 4 AR image parameters for 64x64 Lena image.....	85
Table 4.5 : CRB values of estimated blur and image parameters.....	86
Table 5.1 : Estimated blur parameters for different SNR values.....	103
Table 5.2 : Estimated blur parameters for different initial values	105
Table 5.3 : Identification of AR image coefficients.....	105
Table 6.1 : Estimated blur parameters for different initial values.....	122
Table 6.2 : Estimated blur parameters for different initial values.....	126
Table 6.3 : Identification of AR image coefficients.....	127
Table 7.1 : Estimated sea clutter parameters.....	144
Table 7.2 : KS test results and KL distance.....	147

LIST OF FIGURES

	<u>Page Number</u>
Figure 2.1 : First Order NSHP AR Model's Support Region.....	30
Figure 2.2 : Image Formation Model in The Spatial Domain	32
Figure 2.3 : State Support	34
Figure 2.4 : A Bayesian network representing of state space image model	35
Figure 2.5 : Periodically extended PSF	50
Figure 3.1 : A PSF example.....	54
Figure 3.2 : Block diagram of the restoration process.....	54
Figure 3.3 : Restored image by an Inverse Filter, SNRimp = -14.8 dB.....	56
Figure 3.4 : (a) Restored image by a deterministic regularization, $\alpha = 0.15$, SNRimp = 2.19 dB. (b) Restored image by a stochastic regularization $\alpha = 0.117$ SNRimp = 2.67 dB.....	61
Figure 4.1 : NSHP image support	81
Figure 4.2 : 2-D artificial random field.....	82
Figure 4.3 : (a) The original synthetic image; (b) Blurred and noisy version with SNR=20 dB.; (c) Restored image with learned parameters by gradient based and SNRimp = 2.18 dB.; (d) Restored image with learned parameters by FS and SNRimp = 2.07 dB.; (e) Restored image with learned parameters by RS- FS and SNRimp = 2.11 dB....	83
Figure 4.4 : (a) The original 64×64 Lena image; (b) Blurred and noisy version with SNR=40 dB.; (c) Restored image with learned parameters and SNRimp = 2.32 dB. ; (d) Blurred and noisy version with SNR=30 dB.;. (e) Restored image with learned parameters and SNRimp = 2.11 dB. ; (f) Blurred and noisy version with SNR=20 dB.;. (g) Restored image with learned parameters and SNRimp = 2.04 dB. ; (h) Blurred and noisy version with SNR=10 dB.;. (i) Restored image with learned parameters and SNRimp = 1.91 dB.....	87
Figure 4.5 : The variations of the condition number versus the iteration number.....	88
Figure 4.6 : MSE curves between $\underline{d}$ and $\hat{\underline{d}}$ for different SNR.....	89
Figure 4.7 : MSE curves between $\underline{a}$ and $\hat{\underline{a}}$, for different SNR.....	89
Figure 5.1 : SNR improvement values for ideal and estimated values.....	103
Figure 5.2 : (a) The original synthetic image; (b) Blurred and noisy version with SNR=30 dB.; (c) Restored image with estimated coefficients.....	104
Figure 5.3 : The MSE between $\underline{d}$ and $\hat{\underline{d}}$	104

Figure 5.4	: SNR improvement values for ideal and estimated values.....	106
Figure 5.5	: (a) The original 64×64 Lena image; (b) Blurred and noisy version with SNR=60 dB; (c) Restored image with learned parameters and SNRimp = 3.18 dB. ; (d) Blurred and noisy version with SNR=40 dB.; (e) Restored image with learned parameters and SNRimp = 2.32 dB. ; (f) Blurred and noisy version with SNR=30 dB.;. (g) Restored image with learned parameters and SNRimp = 1.81 dB. ; (h) Blurred and noisy version with SNR=20 dB.;. (i) Restored image with learned parameters and SNRimp = 1.29 dB. ; (j) Blurred and noisy version with SNR=10 dB.;. (k) Restored image with learned parameters and SNRimp = 0.61 dB.....	107
Figure 5.6	: The MSE between $\underline{d}$ and $\hat{\underline{d}}$ as a function of iteration.....	108
Figure 5.7	: The MSE between $\underline{a}$ and $\hat{\underline{a}}$ as a function of iteration.....	108
Figure 6.1	: SNR improvement values for ideal and estimated values.....	123
Figure 6.2	: (a) The original synthetic image; (b) Blurred and noisy version with SNR=30 dB.; (c) Restored image with estimated coefficients by closed form solution (SNRimp = 3.02 dB.) ; (d) Restored image with estimated coefficients by optimization solution (SNRimp = 2.85 dB.).....	124
Figure 6.3	: The MSE between $\underline{d}$ and $\hat{\underline{d}}$ for (a) Closed form solution , (b) Optimization solution.....	125
Figure 6.4	: SNR improvement values for ideal and estimated values.....	127
Figure 6.5	: (a) The original 64×64 Lena image; (b) Blurred and noisy version with SNR=60 dB; (c) Restored image with learned parameters and SNRimp = 3.36 dB. ; (d) Blurred and noisy version with SNR=40 dB.; (e) Restored image with learned parameters and SNRimp = 2.24 dB. ; (f) Blurred and noisy version with SNR=30 dB.;. (g) Restored image with learned parameters and SNRimp = 1.78 dB. ; (h) Blurred and noisy version with SNR=20 dB.;. (i) Restored image with learned parameters and SNRimp = 0.99 dB. ; (j) Blurred and noisy version with SNR=10 dB.;. (k) Restored image with learned parameters and SNRimp = 0.73 dB.....	128
Figure 6.6	: The MSE between $\underline{d}$ and $\hat{\underline{d}}$ as a function of iteration for closed form solution.(a) Case1 (b) Case2 (c) Case3.....	130
Figure 6.7	: The MSE between $\underline{d}$ and $\hat{\underline{d}}$ as a function of iteration for optimization solution. (a) Case1 (b) Case2 (c) Case3.....	131
Figure 6.8	: The MSE between $\underline{a}$ and $\hat{\underline{a}}$ as a function of iteration for (a) closed form solution. (b) optimization solution.....	132
Figure 7.1	: ML-CFAR structure.....	139
Figure 7.2	:QQ plot of sample data versus quantiles of standard normal distribution at low sea states.....	143
Figure 7.3	:QQ plot of sample data versus quantiles of standard normal distribution at high sea states.....	143

Figure 7.4	:Estimated and real pdfs at high and low sea states using Gradient descent method.....	145
Figure 7.5	:Estimated and real pdfs at high and low sea states using FS method.....	145
Figure 7.6	:Estimated and real CDFs at a high sea state using Gradient descent method.....	146
Figure 7.7	: Estimated and real CDFs using FS method using FS method.	146
Figure 7.8	: KS values at each iteration (a) with Gradient based iteration (b) with FS.....	148
Figure 7.9	: KL values at each iteration (a) with Gradient based iteration (b) with FS.....	149
Figure 7.10	: P_{fa} versus T using estimated parameters at high and low sea states.....	150

LIST OF SYMBOLS

$a(k, l)$	: 2-D AR model parameters
A	: $(N^2 \times N^2)$ image model coefficients matrix
$cond(A)$	: Condition number of matrix A.
C_r	: Cramer Rao lower bound value.
$C_{f g}$	: The conditional covariance of f given g .
$d(m, n)$	: Blurring operator.
D	: $(N^2 \times N^2)$ blur model coefficients matrix.
$E_\theta \{\bullet G\}$	: The conditional expectation given G .
f	: Original image in vector form.
$f(i, j)$	: Original image model.
$\hat{f}(i, j)$	: Restored (estimated) image.
f_0	: Initial state of the original image.
f_k	: the k^{th} state vector of the original image.
$\hat{f}_{k k-1}$	: Image prediction (time update).
$E_{x,y}$	: An elementary matrix.
$F(u, v)$	: Image model in DFT domain.
$\hat{F}(u, v)$	: Estimated image model in DFT domain.
$g(i, j)$	: Observed image.
g	: Observed image in vector form.
g_k	: The k^{th} observation vector.
$G(u, v)$	: Observed image in DFT domain.
$g(x)$	: Gradient of a function.
$h(i, j)$	: Appropriate restoration filters.
$H(x)$	: Hessian matrix of a function.
$h_{inv}(i, j)$	: Inverse filter.
$H_{inv}(u, v)$	: Inverse filter in DFT domain.
I	: Identity matrix.
$I_f(x)$	: Fisher Information Matrix.
$I_n(\bullet)$	: n^{th} order Modified Bessel function of the first kind of order.
$ J $	: Jacobian transformation.
J	: The component number of x associated with Inphase channel.
K	: The component number of y associated with Quadrature channel.
K_k	: Kalman gain matrix.

$L(\theta)$	: The (log)-likelihood function of θ
$M_{k k}$	: Filtered error covariance matrix.
$M_{k k-1}$	: Predicted error covariance matrix.
M_0	: Initial covariance matrix.
$M_{k N}$	: Conditional covariance of f_k given g_k .
$M_{k-1,k N}$	: The covariance across two time steps.
$N \times N$	: Size of image.
p	: The size of the f_k and g_k vectors.
P	: Covariance matrix of observed image.
$p(x)$	: The pdf of the Inphase channel.
$p(y)$	: The pdf of the Quadrature channel.
$p(f; A, Q)$	: The pdf of f , given the pdf of v and the image model.
$p(\{f\}, \{g\})$	: The joint pdf of f and g .
$p(f_0)$	: The pdf of the initial state.
$p(g f; D, R)$	: The pdf of g , given the pdf of the observation noise, observation model and the original image f .
$p(g_k f_k)$	: The pdf of g_k , given the pdf of f_k .
$p(g; \theta)$	: The pdf of the observed image given θ .
P_{fa}	: The probability of false alarm
Q	: Covariance matrix of modeling error
R	: Covariance matrix of observation noise.
r	: The output of the envelope detector.
R_{ff}	: Autocorrelation of the ideal image
R_{ww}	: Autocorrelation of the observation noise.
$SNR_{observed}$	: The signal-to-noise-ratio of the recorded image
$SNR_{estimated}$	: The signal-to-noise-ratio of the restored image
SNR_{imp}	: Improvement in signal-to-noise-ratio
$S_{ff}(u, v)$	: The power spectrum of the ideal image in DFT domain.
$S_{ww}(u, v)$	: The power spectrum of the noise in DFT domain.
T	: The threshold value.
$tr\{\bullet\}$	: The trace of a matrix.
$v(i, j)$	: Modeling error.
v	: Modeling error in vector form.
$w(i, j)$	: Observation noise.
w	: Observation noise in vector form.
W	: Upper triangular matrix.
Y	: Complete data.
$ \bullet $	: The determinant of a matrix.
$\ \bullet\ $	: The Euclidian norm.
α	: Regularization parameter.

α_j	: The mixture ratios associated with Inphase channel.
β_k	: The mixture ratios associated with Quadrature channel.
$\nabla F(x)$	: Gradient of a function.
$\nabla^2 F(x)$	: Hessian of a function.
$\Gamma(\theta)$	: The conditional expectation of the log-likelihood of complete data given the observed incomplete data.
$(\Gamma(\theta))$	: Element-by-element processing of matrix $(\Gamma(\theta))$
$\lambda_M(A)$	: The maximum eigenvalues of the A matrix.
$\lambda_m(A)$	: The minimum eigenvalues of the A matrix.
Λ	: Diagonal matrix comprising the 2-D DFT coefficients of $d(i, j)$.
Ω	: The matrix containing the components of the complex exponential basis functions of the 2-D DFT.
η	: Convergence parameter.
μ_0	: Initial mean.
$\mu_{k N}$	: The conditional mean of f_k given g_k .
$\mu_{f g}$	: The conditional mean of f given g .
$\hat{\theta}_{ML}$	: ML estimator of the parameter vector θ .
θ^*	: The optimal solution of θ .
σ_v^2	: The variance of modeling error.
σ_w^2	: The variance of modeling error.
$(\sigma_x^2)_j$	: The variance associated with Inphase channel.
$(\sigma_y^2)_k$	: The variance associated with Quadrature channel.

EFFICIENT MAXIMUM LIKELIHOOD PARAMETER LEARNING: IMAGE and RADAR APPLICATIONS

SUMMARY

Generally, in a real world problem, the corruption mechanism of any system is not known because only observed data is available. Learning the system parameters from observations is compulsory to execute processes on the system such as restoration, reconstruction and filtering. In realistic image restoration problems, the blurring effect is unknown or not exactly known. The first step for image restoration is the learning (identification) of degradation. This thesis presents four different new approaches to learn the blur and image parameters and to restore the images, taking the open problems in literature into consideration. We have also applied some of the above approaches to the radar detection problem where the sea clutter parameters are learned.

There are two main Maximum Likelihood (ML) approaches to identify the unknown blur and image parameters: Direct ML approach and Expectation Maximization (EM) based approach.

In direct method, ML estimation procedures are based on optimizing the probability density function (pdf) of the observed image with respect to unknown parameters. In this way the ML estimator finds the various values that most likely resulted in the blurred image observed. The only way to optimize the log-likelihood function is by means of mathematical programming. The blur identification and image restoration problem was formerly formulated as a constrained ML problem. Optimization of the log-likelihood function, which is nonlinear in nature, is solved by employing an iterative gradient based optimization procedure. Gradient based optimization techniques contain a convergence parameter, which plays an important role in determining the performance of the system. Gradient based methods have slow convergence, because of ill-posed (ill-conditioned) property of the image restoration problem. Therefore, second-order information is vital for fast convergence.

In the first approach, we have formulated the blur identification problem as a Maximum Likelihood (ML) problem, which has been solved by employing a Newton type optimization method based on the method of scoring. This method is called Fisher Scoring (FS) method. The gradient descent algorithm takes many steps to converge, while Newton's method reaches the minimum in a few steps. However Newton's method is more expensive, because it requires both the gradient and the Hessian to be evaluated. In the FS algorithm, it is possible to avoid complicated Hessian equations by using only the gradient values. Selecting a convenient step size, which enables fast convergence and provides stability, is very desirable in all optimization problems. We have solved the blur identification problem very

efficiently and rapidly using the FS method. Nevertheless, the determination of the convenient convergence parameter issues is still weaknesses of the method.

In the second approach, we try to find new criteria based on the convergence properties of the FS optimization. We attempt to combine the best properties of line search and Trust Region (TR) algorithms. We enforce line search conditions on each iteration to guarantee convergence. The new hybrid solution is fast and capable of handling ill-posed problems. We present a FS method that ensures the convergence in every iteration step. We call this method “Trust-Region Fisher Scoring Method”. This new algorithm is robust and efficient in practice. We present empirical results suggesting that our method regularizes the condition number of the Hessian. Regularized solution, which is necessary because of the ill-posed nature of the problem, is a side benefit of the proposed method. Thanks to the algorithm, convergence to a stationary point is guaranteed if the step in the “trust region”. We analyze the method for regularization of ill posed image problems and we present some results in Experimental Results section. In many examples, the direct optimization of the likelihood function using the gradient based techniques is not feasible. The EM algorithm is an iterative technique for ML estimation and is an alternative to gradient based methods. There are two steps in EM algorithm as E and M. In the E-step, the conditional expectation of the “hidden variables” is calculated. In the M-step, this expectation is maximized with respect to the parameters. It is highly preferable to have analytical formulas for the parameter update equations. However, in many problems, it is not possible to obtain a closed form solution. So, several alternatives to the EM algorithm have appeared. We have developed two different EM based methods.

In the third approach, we have worked on recursive image processing problem and we have obtained both a new closed form solution and an optimization solution based on gradient descent algorithm. For the image processing applications, recursive estimation techniques are important because of the huge size of the input data. Recursive processing allows dynamic processing with modest storage requirements. Learning of unknown image and blur parameters for batch image processing had been studied but recursive EM learning of unknown image and blur parameters has not been studied before. We introduce a new simultaneous recursive parameter learning and image restoration method. We present a new formulation which is given in a Dynamic Bayesian Network (DBN) framework. This technique incorporates optimal Kalman smoothing equations for ML parameter identification and state estimation. Because of the computationally heavy processing of smoothing, we use filtering approximation instead of Kalman smoothing. In addition, a restored image is obtained simultaneously at the output of the Kalman filter. The main advantage of our proposed method is that we learn the image parameters, which are compulsory for blind image restoration, and we obtain the restored images in a single structure. Although our proposed method deals with huge data sizes, because of its recursive structure, it does not need large size storage. Performance evaluation of the method is given based on simulations and experimental results are demonstrated by using both artificial and real images.

In the fourth approach, we have developed a new analytical solution; furthermore we have built a new fast optimization method using FS method. In many problems, the EM algorithm is easy and stable, but there are two main drawbacks of the EM. One

of them is slow convergence and the other one is finding a closed form solution in the M step. We endeavor to solve both problems for blind image deconvolution as two different batch processing techniques. In the first technique, learning of unknown image and blur parameters is investigated in a closed form solution based on the EM algorithm. Cholesky factorization is used to find autoregressive image model parameters. In the second approach, we present a fast optimization method based on the accelerated EM algorithm. When it is impossible to carry out the M-step exactly, iterative techniques can be used. The fastest well-known algorithm would be Newton's method, which has quadratic convergence. We have introduced a new method based on the accelerated EM algorithm to learn the unknown image and blur parameters. The proposed method learns the image parameters and restores images simultaneously. Acceleration is provided using Fisher Scoring (FS) optimization in the M step. A small number of FS iterations satisfactorily perform well for each M step. The algorithm we propose reaches to the local minima in a few steps whereas conventional EM needs more iteration. Thanks to the FS optimization, it is possible to avoid complicated second derivative of the log-likelihood function by using only the gradient values. Also the regularization parameter is estimated in the same single structure. The restored image is obtained simultaneously using the regularized version of the Wiener filter after learning of the image, blur and noise parameters. Performance evaluation of the method is performed based on simulations and experimental results are demonstrated by using both artificial and real images.

As a radar application, we have also applied some of the above approaches to the radar detection problem where the sea clutter parameters are learned. Statistical characterization and modeling of radar clutter is of great importance for developing adaptive threshold algorithms. The clutter statistics vary widely and generally different type of clutter is modeled using different type of distribution. Design of adaptive radar detection algorithms requires that the parameters should be learned from the operational environment. A new and more suitable clutter model for changing environments is introduced. In order to model the sea clutter we have proposed to use a generalized distribution, which is valid for different background statistics like low and high sea states. A generalized Gaussian Mixture (GM) pdf with zero mean has been used in order to form a statistical model of the clutter for each of the inphase (I) and quadrature (Q) channels. New background pdf, which is the output of the envelope detector, resemble Rician Mixtures. The derivation of the envelope detector output, which has multi component GM inputs, and the ML estimation of this output are the issues, which were not addressed before. In order to detect targets using adaptive CFAR techniques, clutter background statistics should be known a priori. But in practice, this is typically not the case. So, we have to learn the sea clutter parameters, which are required to calculate the adaptive threshold correctly, from the environment. In addition, the estimation of mixture parameters required for calculating the detection threshold in CFAR has been investigated, for the first time. Parameters of the background statistics are formulated as a ML problem and; the log-likelihood minimization problem can be solved by employing an iterative optimization method. First, we have proposed to use the Gradient descent optimization method to estimate the unknown sea clutter parameters. Second, we have applied the FS optimization method to estimate the unknown sea clutter parameters, rapidly. Performance analysis has been presented using real sea clutter data, which was collected by IPIX radar at low and high sea states.

ETKİN ENBÜYÜK OLABİLİRLİK PARAMETRE ÖĞRENME: İMGE VE RADAR UYGULAMALARI

ÖZET

Genelde, gerçek hayat problemlerinde sadece gözlem verisinin elde bulunması nedeni ile sistemi bozan yapılar bilinmez. Sistem parametrelerinin gözlem üzerinden öğrenilmesi onarım, geriçatma ve süzgeçleme gibi süreçlerin işletilebilmesi için gereklidir. Gerçekçi görüntü onarım problemlerinde, bulanıklık etkisi ya tam olarak ya da kısmen bilinmez. Görüntü onarımının ilk adımı, bozucunun öğrenilmesi bir başka deyişle tanınmasıdır. Bu tez, görüntüyü onarmak amacı ile bulanıklık ve görüntü parametrelerinin öğrenilmesi için açık problemleri göz önüne alarak dört yeni yaklaşım sunmaktadır. Aynı zamanda, burada tanımlanan bazı yaklaşımlar deniz kargaşa parametrelerinin öğrenildiği radar sezim problemine uygulanmıştır.

Bilinmeyen bulanıklık ve görüntü parametrelerini tanınması için iki temel Enbüyük Olabilirlik (EO) yaklaşımı vardır: Doğrudan EO yaklaşımı ve Beklenti Enbüyüklemeye (BE) dayalı yaklaşım.

Doğrudan yaklaşımda, EO kestirim süreçleri, bilinmeyen parametrelere göre gözlem görüntüsünün olasılık yoğunluk işlevini eniyilemeye dayanır. Bu yolla, EO kestirici bulanık gözlem görüntüsünde en olası değerleri bulur. Logaritmik olabilirlik işlevini eniyilemenin tek yolu matematiksel programlama teknikleri kullanmaktır. Bulanıklık tanıma ve görüntü onarım problemi geçmişte zorlanmış EO problemi olarak tanımlanmıştır. Doğrusal olmayan logaritmik olabilirlik işlevinin eniyilenmesi dürümsel bayır temelli yöntemler kullanılarak çözülür. Bayır temelli yöntemler tekniğin performansına önemli derecede etki eden bir yakınsama parametresi içerirler. Aynı zamanda, bu yöntem görüntü onarım probleminin genelde kötü koşullu olma özelliği nedeni ile yavaş yakınsarlar. Bu sebeple hızlı yakınsama için ikinci türeve dayalı bilgileri kullanmak gerekli olmaktadır.

İlk yaklaşımda, bulanıklık tanımlama problemini bir EO problemi olarak tanımladık ve Newton temelli “Fisher Scoring” yöntemi ile çözdük. Bayır temelli yöntemler yakınsama için fazla sayıda adım ihtiyacı duyarken, Newton yöntemi birkaç adımda en iyi değere ulaşabilmektedir. Ancak, Newton yöntemi hem bayır hemde Hessian değerine ihtiyaç duması nedeni ile pahalı bir yöntemdir. FS yönteminde ise karmaşık Hessian ifadesi sadece bayır değerleri kullanılarak elde edilebilmektedir. Hızlı yakınsama ve kararlılığı sağlayacak uygun adım aralığının seçilmesi tüm eniyileme yöntemlerinde istenen bir özelliktir. FS yaklaşımını kullanarak bulanıklık tanıma problemi etkili ve hızlı bir şekilde çözülmüş olsa bile uygun yakınsama katsayısının seçimi açık problem olarak hala durmaktadır.

İkinci yaklaşımda, FS eniyilemenin yakınsama özelliklerine dayalı yeni bir kriter bularak yukarıda bahsedilen açık problemi çözmeye çalıştık. Hat arama ve güvenli bölge yordamlarının en iyi yanlarını birleştirmeye çalıştık. Bu amaçla hat arama koşullarını yakınsamayı garanti etmek için her durum adımında tekrar hesapladık. Bu yeni çözüm hızlı ve kötü koşullama probleminin üstesinden gelebilecek bir yöntemdir. Sunduğumuz FS yöntemi her durum adımında yakınsamayı sağlayacak bir yaklaşımdır. Biz gürbüz ve etkili olan bu yeni yönteme “Güvenli Bölge Fisher Scoring” yordamı adını verdik. Deneysel sonuçlar ile de gösterdik ki önerilen yöntem Hessian’ın koşul sayısını düzenliştirmektedir. Kötü koşullanmış problem için gereklilik olan düzenliştirilmiş çözümönerilen yöntemin bir diğer yararlı çıktısı olarak değerlendirilebilir. Yordam sayesinde eğer adım “güvenli bölge” içinde ise durağan bir noktaya yakınsama garanti edilir. Kötü koşullanmış görüntü problemlerinin düzenliştirilmesi için yöntem analiz edilmiş ve sonuçlar Deneysel Sonuçlar kısmında sunulmuştur. Birçok örnekte olabilirlik işlevinin bayır temelli teknikler kullanılarak doğrudan eniyilenmesi olurlu değildir. BE yordamı ise EO kestirimi için bayır temelli yöntemlere alternatif olabilecek dürümsel bir tekniktir. BE yordamının, B ve E olmak üzere iki temel adımı vardır. B adımında, saklı değerlerin koşullu beklentisi hesaplanır. E adımında ise, bu beklenti aranan parametrelere göre enbüyüklenir. Parametre güncelleme işlemleri için analitik çözümlerin elde edilebilmesi istenir. Bununla birlikte, birçok problemde kapalı formda bir çözüm elde etmek olası değildir. Bu yüzden BE yordamına alternatif olacak yöntemler önerilmiştir. Biz iki farklı BE temelli yöntem önerdik.

Üçüncü yaklaşım olarak, özyinelemeli görüntü işleme problemi ele alınmış ve hem kapalı form çözüm hem de bayır iniş temelli yordama dayalı eniyileme çözümü elde edilmiştir. Görüntü işleme uygulamaları için, özyinelemeli kestirim teknikleri büyük boyutlu veriler nedeni ile oldukça önemlidir. Özyinelemeli işlem kabul edilebilir depolama gereksinimi ile işlem yapma olanağı sağlar. Toplu görüntü işleme için bilinmeyen görüntü ve bulanıklık parametrelerinin öğrenilmesi çalışılmış, ancak, bilinmeyen görüntü ve bulanıklık parametrelerinin özyinelemeli olarak BE’ye dayalı öğrenilmesi daha önce çalışılmamıştır. Dinamik Bayesçi Ağ yapısında yeni bir eşzamanlı ve özyinelemeli parametre öğrenme ve görüntü onarım formülasyonu verilmiştir. Bu yöntem Enbüyük Olabilirlik (EO) parametre tanılama ve durum kestirimi için eniyi Kalman yumuşatma denklemlerini içermektedir. Yumuşatma işleminin yoğun hesap gerektiren yapısı nedeni ile Kalman yumuşatma yerine Kalman süzgeçleme yaklaşıklığı kullanılmıştır. Dahası, onarılmış görüntü eşzamanlı olarak Kalman süzgecin çıkışında elde edilmiştir. Önerdiğimiz yöntemin en önemli avantajı tek bir yapı içinde kör görüntü onarımı için bir zorunluluk olan görüntü parametrelerinin öğrenilmesi ve onarılmasıdır. Aynı zamanda özyinelemeli yapı sayesinde büyük boyutlu veriler ile işlem yapılabilmektedir. Önerilen yöntemin başarımı deneysel sonuçlar kısmında hem suni hem de gerçek görüntüler üzerinde verilmiştir.

Dördüncü yaklaşımda, Beklenti Enbüyükleme (BE) yordamına kapalı çözüm ve FS yöntemi kullanılarak hızlı eniyileme yöntemi sunulmuştur. Birçok problemde, BE yordamının uygulanması kolay ve kararlı olmakla birlikte BE yordamının iki temel noksanlığı vardır. Bunlardan ilki yakınsamanın yavaş olması diğeri ise E adımında kapalı yapıda bir çözüm bulmanın zor olmasıdır. Bu problemlerin çözümü için iki farklı toplu işlem yöntemi sunulmuştur. İlk yöntemde, bilinmeyen görüntü ve bulanıklık parametrelerinin Beklenti Enbüyükleme (BE) yordamına dayalı olarak öğrenilmesi için kapalı çözüm araştırılmıştır. Özbağlanımlı görüntü model

parametrelerinin bulunması için Cholesky ayrıştırma yöntemi kullanılmıştır. İkinci yöntemde ise, hızlandırılmış BE yordamına dayalı hızlı bir eniyileme yöntemi verilmiştir. E adımını analitik olarak çözmek olası olmadığı zaman dürümsel yöntemler kullanılabilir. Hızlı ve iyi bilinen bir yöntem karesel yakınsamaya sahip olan Newton yöntemidir. Bilinmeyen görüntü ve bulanıklık parametrelerini öğrenmek için hızlandırılmış BE'ye dayalı yeni bir yöntem verilmiştir. Önerilen yöntem eşzamanlı olarak görüntü parametrelerini öğrenirken görüntü onarımı yapabilir. Hızlandırma E adımında FS yöntemi kullanılarak sağlanır. Her E adımında az sayıda FS'nin uygulanması kabul edilebilir sonuçlar vermektedir. Önerilen yöntem yerel enküçük değere birkaç adımda yakınsamaktadır. FS eniyileme sayesinde logaritmik olabilirlik işlevinin karmaşık olan ikinci türevi yerine sadece birinci türev değerleri kullanılabilir. Ayrıca, düzenleme parametresi de aynı yapı içinde kestirilmekte ve düzenlenilmiş Wiener süzgeç yapısı kullanılarak eşzamanlı olarak görüntü onarımı yapılmaktadır. Önerilen yöntemin başarımı deneysel sonuçlar kısmında hem suni hem de gerçek görüntüler üzerinde verilmiştir.

Bir radar uygulaması olarak, yukarıda önerilen yaklaşımları deniz kargaşasının öğrenildiği radar sezim problemine uyguladık. Radar kargaşasının modellenmesi ve istatistiksel yapısı uyarlamalı eşik yordamlarının geliştirilmesi için çok önemlidir. Genelde, değişik tipteki kargaşa tipleri çeşitli istatistiksel dağılımlar ile modellenir. Uyarlamalı radar sezim yordamlarının tasarlanması için işlemsel ortamdan parametrelerin öğrenilmesi gereklidir. Değişken ortamlar için yeni ve daha uygun bir kargaşa modeli sunulmuştur. Deniz kargaşasını modellemek için düşük ve yüksek deniz durumları gibi farklı arkaplan istatistiklerinde geçerli olacak genelleştirilmiş bir dağılım kullanmayı önerdik. Zarf sezicinin I ve Q kanallarının giriş istatistikleri sıfır ortalamalı genelleştirilmiş Karma Gauss (KG) dağılımı olarak kabul edilmiştir. Zarf sezicinin çıkışı olan yeni arka plan olasılık yoğunluk işlevi Karma Rician (KR) dağılımına benzemektedir. Girişleri çoklu KG olan bir zarf sezicinin çıkışı ilk defa bu çalışmada hesaplanmıştır. Sabit sahte alarm Oranı ile uyarlamalı olarak hedef sezimi gerçekleştirmek için arka plan kargaşa istatistiklerinin bilinmesi gerekir. Ancak pratikte arka plan istatistikleri bilinmemektedir. Yine bu çalışmada sabit sahte alarm oranlı eşik hesaplanması için karışım parametrelerinin kestirimi ilk defa bu çalışmada ele alınmıştır. Zarf sezici çıkışında parametreler Enbüyük Olabilirlik (EO) yöntemi kullanarak kestirilmiş ve logaritmik olabilirlik işlevi dürümsel teknikler kullanılarak çözülmüştür. Bilinmeyen deniz kargaşa parametrelerinin kestirimi için dürümsel teknik olarak ilk olarak Bayır iniş yöntemi ve sonrasında hızlı eniyileme için FS yöntemi kullanılmıştır. Başarım analizi düşük ve yüksek deniz durumunda IPIX radarı ile toplanan gerçek deniz kargaşa verileri kullanılarak yapılmıştır.

1. INTRODUCTION

1.1. Motivations

In real life applications, it is necessary to handle with uncertain events and observations. Consequently, modeling uncertain relationships among many kinds of variables and learning such variables are important topic.

The image restoration problem can be defined as the general problem of estimating the ideal image from its blurred and noisy version. In the use of image restoration methods, generally, the characteristics of the degrading system are assumed to be known. Many classical image restoration techniques have been reported under the assumption that the blur operation is exactly known. In practical situations, however, one may not be able to obtain this information directly from the image formation process. The direct inversion of the blur transfer function usually has a large magnitude at high frequencies, therefore excessive amplification of noise results at those frequencies. Clearly, this is not an acceptable solution for noisy images. To overcome the noise sensitivity problem of the inverse filter, some filters have been developed based on the least-squares structure. The Wiener filter is based on batch processing which is usually implemented in the frequency domain. The Kalman filter is based on recursive processing which is usually implemented in the spatial domain. Both solutions only work when blur, image and noise parameters are known [1] . The learning (identification) of the blur and image parameters is a considerably important step in image restoration. In order to apply successful restoration techniques, these needed parameters must be learned from the given image data.

The combination of image restoration and blur identification is often referred to as blind image deconvolution. .Blind image deconvolution methods are very important in real image processing problems such as biomedical images, microscopic images and astronomical images. Although there are many methods to solve the blind

deconvolution problems but they currently have some troubles and open problems that are summarized in Table 1.1.

In many image processing problems we need to estimate the original complete data sets, generally from incomplete and most often, from degraded observations. Blind deconvolution refers to the problem of restoring the original image from a degraded observation and incomplete blur information. Learning the needed parameters from the given noisy and blurred image is still an area of active research. Still, the key problem of simultaneously estimating an unknown blur from the observed sequence and restoring of images remains to be undertaken. Numerous open problems in these areas are presented in [2] [3] [4] [5] [6] [7] [8] .

Table 1.1: Current and open problems in image and blur identification.

Current Situations	Open Problems
Present optimization methods have generally slow convergence rate. They need high computational power.	Acceleration of the methods.
Regularization techniques are required because of ill-posed structure of the image restoration problem but regularization parameter is chosen experimentally.	Choosing regularization problem.
In many problems, finding an analytical solution in the M step of the EM algorithm is not easy task.	Finding closed form solutions in EM.
Usually, these techniques apply to the batch data and they need huge memory storage requirements.	Recurisive and on-line learning.

In radar signal processing area, the understanding and modeling of radar clutter plays an essential role for radar system design and performance evaluation. Design of adaptive radar detection algorithms requires that the parameters should be learned from the operational environment. Learning of the image and blur parameters from observations are very similar with learning background statistics to detect target from

environment. We can apply the same idea to radar detection and clutter parameter learning problem, which has not been applied before.

1.2. Technical Contributions of This Thesis

The main contributions of the thesis are contained in Chapters 4, 5, 6 , 7. A challenging task, which is learning unknown parameters from the given observations, is the topic of this thesis. The main novel technical contributions of this thesis are as follows.

1.2.1. FS Optimization in Direct ML Method

The blur identification problem has been formulated as a ML problem and solved by employing a FS method. In the FS algorithm, it is possible to stay away from complicated Hessian equations by using only the gradient values. A comparative analysis of EM and Newton type optimization has been given.

1.2.2. Trust Region FS Optimization in Direct ML Method

A new criteria based on the convergence properties of the FS method has been considered. We have offered a FS method that ensures the convergence in each iteration step. We have called this method “Trust-Region Fisher Scoring Method”. Regularized solution is a side benefit of the proposed method. Thanks to the algorithm, global convergence to a stationary point is guaranteed if the step in the “trust region”.

1.2.3. EM Based Learning for Recursive Processing

The EM learning of unknown image and blur parameters for batch image processing had been studied before but recursive EM learning of unknown image and blur parameters has not been studied. We have introduced a new simultaneous recursive parameter learning and image restoration method. We have presented that state space image model can be represented as a DBN. We use Kalman filter structure in order to optimize the ML estimation of the unknown image and blur parameters. In addition, a restored image is obtained at the output of the Kalman filter simultaneously.

1.2.3.1.Closed Form Solution

Finding closed form solution in the M step of the EM, is not easy task. We have managed to solve the recursive EM learning problem for image and blur parameters in closed form.

1.2.3.2.Optimization Solution

In most cases, the solution to the M-step does not exist in closed form or it may not be feasible to attempt to find. We have solved the recursive EM learning problem by gradient based method as alternative to closed form solution.

1.2.4. EM Based Learning for Batch Processing

We have worked on batch image data and we have developed a new analytical solution and a new fast optimization method.

1.2.4.1. Closed Form Solution

Learning of unknown image and blur parameters has been investigated in a closed form based on the EM algorithm. Cholesky factorization has been used to find autoregressive image model parameters in the first time. Also the regularization parameter has been estimated in the same single structure. The restored image has been obtained simultaneously using the regularized version of the Wiener filter.

1.2.4.2. Optimization Solution

We have also presented an optimization method based on the accelerated EM algorithm. The proposed method learns the image parameters and restores images simultaneously. Acceleration have been provided using FS optimization in the M step. A small number FS iteration satisfactorily performs well for each M step.

1.2.4.3. Regularization

We have also estimated the regularization parameter so the regularized version of the Wiener has been applied to images.

1.2.5. Radar Detection Application

We have successfully applied our experiment to the radar detection problem for learning sea clutter parameters. We have introduced a new clutter model that is a more acceptable model for the varying environment. In order to model the sea clutter we have proposed to use a generalized distribution, which is applicable for different background statistics like low and high sea states. We learn the sea clutter parameters, which are required to calculate adaptive threshold correctly, from the environment. The derivation of the envelope detector output, which has multi component GM inputs, and the ML estimation of this output had not been addressed before. In addition, the estimation of mixture parameters required for calculating the detection threshold in CFAR was investigated, for the first time. First of all, we propose to use the Gradient descent optimization method to estimate the unknown sea clutter parameters. Secondly, we have applied the FS optimization method to estimate the unknown sea clutter parameters, rapidly.

1.3. Scope of the Thesis

In chapter 2, mathematical background and definitions, which are required the rest of thesis, are summarized. Theoretical descriptions of image formation, the basic concepts of image and observation models are discussed in this chapter with historical survey. Practical optimization techniques used in this thesis, convergence analysis of these optimization methods and some preliminary definitions are given in this chapter. Ill-posed structure of the image identification and restoration problem is cited and is mentioned about regularization methods to deal with ill-posed ness.

In Chapter 3, classical image restoration and identification methods are reviewed with some examples. Some metrics to measure the degradation of corruption and improvement in restoration are given. Previous studies on blur and image parameter identification problem is summarized. We mentioned about the optimization solution and EM based solution of ML approach.

In Chapter 4, we formulate the blur and image parameter identification problem from a ML point of view. A Newton type optimization method is described to optimize the log-likelihood function and this method applies to the problem at hand. We formulate

the blur identification problem as a ML problem and solve it by employing a Newton type optimization method based on FS. The convergence of FS method is examined using the classical line search optimization approaches to set up a new Trust Region method. A new criteria based on the convergence properties of the FS optimization is obtained. We also present a new TRFS algorithm that ensures the convergence in every iteration step to identify the image parameters. We give empirical results suggesting that our method regularizes the condition number of the Hessian. We analyze the method for regularization of ill posed image problems. First of all, experimental results obtained from a simulated two-dimensional (2-D) autoregressive (AR) image are given to confirm that algorithm is working properly. In the second part we use the reduced 32×32 Lena image and present some results.

In chapter 5, we introduce a new simultaneous recursive parameter learning and image restoration method based on the ML parameter identification and state estimation for images. Totally blind image restoration problem is solved using an EM based learning technique. This technique incorporates optimal Kalman smoothing equations for ML parameter identification and state estimation. Using Kalman filter instead of the Kalman smoothing is presented because of the computationally heavy processing of smoothing. We derived an analytical and optimization solution for this EM based recursive learning. The performance of the method is given in the experimental results section.

In chapter 6, the same problem in chapter 5 is considered as a batch image processing. Learning of unknown image and blur parameters is investigated in a closed form. Cholesky factorization is used to find autoregressive image model parameters. The regularization parameter is estimated and restored image is obtained using the regularized version of the Wiener filter. Some accelerated EM methods in the literature is reviewed and a new method based on the accelerated EM algorithm is introduced.

In chapter 7, the experiment on learning image and blur parameters is applied to radar detection and clutter parameter learning problem. The estimation of Mixtures parameters, which are the inputs of the envelope detector, were investigated in this chapter. Parameters of the background statistics are formulated as a ML problem that

is solved by employing an iterative optimization method. In the first technique, the Gradient descent optimization method is used and in the second method, we apply the FS optimization method to estimate the unknown sea clutter parameters. Performance analysis is given using real sea clutter data.

General conclusion about our new proposed methods is given in chapter 8. We have also presented some suggestions for further research in this chapter.

2. MATHEMATICAL BACKGROUND and DEFINITIONS

2.1. Introduction

Due to imperfections in imaging devices, the recorded image generally represents a degraded version of the original image. These degradations may have many causes. Typically, the types of degradation on recorded images are noise and blurring [9] [10] .

Noise may be introduced by the medium through which the image is created, by the recording medium, by measurement errors, and by quantization errors. We can summarize the type of noise in an imaging system as film grain noise, photo-electronic noise, thermal noise, and quantization noise.

Blurring is a form of bandwidth reduction of an ideal image owing to the imperfect image formation process. It can be caused by relative motion between the camera and the original scene, or by an optical system that is out of focus. In addition to these blurring effects, noise always corrupts any recorded image.

The image restoration problem can be defined as the general problem of estimating the ideal image from its blurred and noisy version. A fundamental issue in image restoration is blur removal in the presence of observation noise. The image restoration problem is in general ill-posed; a small perturbation on the given data produces large deviations in the solution [9] .

The direct inversion of the blur transfer function usually has a large magnitude at high frequencies, therefore excessive amplification of noise results at those frequencies. In the classical image restoration problems, the blurring function and the characteristics of noise are assumed known. Many methods have been reported for restoring the original image under the assumption that the blur operation is exactly known. In realistic image restoration problems, the blur function is unknown or not

exactly known. In this case we are faced with the difficult problem of “blind” image restoration. The first step for image restoration is the identification of degradation. In such cases we have to use some prior knowledge in order to somehow recover simultaneously both original image and blurring. These parameters have to be learned or identified from the observations [11] . Learning means estimating the parameters from observation data. This is often called system identification that has applied to widespread areas such as in control engineering, adaptive filtering, communications, geophysical engineering, radar signal processing and image processing [12] . Learning is a flexible and effective means of extracting the stochastic structure of the environment. Recent improvement of computers has made it possible to implement large-scale learning systems [13] [14] .

Learning techniques have been intensively used for practical applications, for example, image processing such as face recognition and finger print discrimination, and signal processing such as voice recognition, signal prediction, and independent component analysis such as blind separation and blind deconvolution [15] [16] [17] . Two different types of learning are used, namely batch learning and on-line learning [18] . Several researchers in the estimation, filtering and control fields have worked on the learning of parameters [19] [20] [21] [22] [23] [24] . Sari et al. [26] [27] [28] applied the statistical learning techniques to radar signal processing. In order to detect targets using adaptive Constant False Alarm Rate (CFAR) techniques, clutter background statistics should be known a priori. But in practice, background statistics are typically not known a priori. They learnt the parameters using the ML algorithm. They proposed a new generalized mixture distribution for modeling different types of sea clutter. Input statistics of the envelope detector were assumed to be a Gaussian-Mixture (GM). New background pdf was look like Rician Mixture. Thus, a joint estimation and detection process is employed whereby the background statistics are estimated before target detection. Therefore, in this study they learnt the sea clutter parameters, which are required to calculate adaptive threshold correctly, from the environment. They presented the results using real sea clutter data, which was collected by IPIX radar [29] at low and high sea states. Again, in [30] we have developed a fast learning technique to estimate the background statistics parameters from the output of the envelope detector, the inputs of which are multi-component Gaussian Mixture (GM) distributions. The parameters of the background statistics

have been learned using a ML algorithm with FS optimization, efficiently. Quite satisfied fits have been obtained using the ML method for estimating the parameters.

The goal of blur identification is to estimate the attributes of the imperfect imaging system from the observed degraded image itself prior to the restoration process. The combination of image restoration and blur identification is often referred to as blind image restoration [31] [32] .

In many image processing problems we need to estimate the original complete data sets, generally from incomplete and most often, from degraded observations [33] . This task of learning parameters from only corrupted data may be termed a “missing data” problem and is also the topic of this thesis.

2.2. Image and Observation Model

The original image $f(i, j)$, is modeled by the following 2-D autoregressive (AR) model:

$$f(i, j) = \sum_{k, l \in R_1} a(k, l) f(i - k, j - l) + v(i, j) \quad (2.1)$$

where (i, j) refers to the pixel in the i^{th} column and j^{th} row and R_1 represents the support region of the model. Some common model supports for various AR models can be found in [34] . A Non Symmetric Half Plane (NSHP) region of support is illustrated in Figure 2.1.

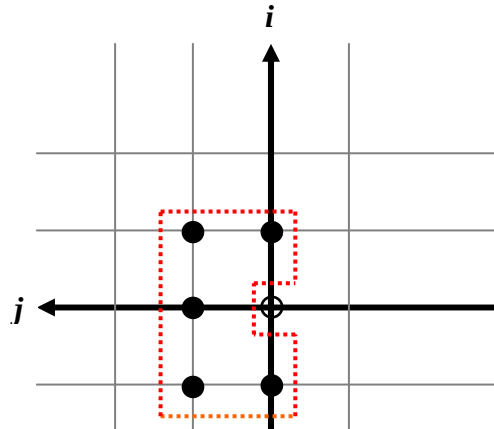


Figure 2-1: First order NSHP AR model's support region.

Causal AR model support is suitable for use in recursive image restoration algorithms and the NSHP causal support is the largest possible causal support.

In matrix form,

$$f = Af + v \quad (2.2)$$

where the vector f represents the unknown true image. The modeling error v is a homogeneous Gaussian distributed white noise process with zero mean and covariance matrix $Q = \sigma_v^2 I$, here σ_v^2 is positive, and covariance matrix is uncorrelated with f .

The observed image, $g(i, j)$, is modeled as

$$g(i, j) = \sum_{m, n \in R_2} d(m, n) f(i - m, j - n) + w(i, j) \quad (2.3)$$

where R_2 represents the support region of the blur, and $d(m, n)$ is the blurring operator. Here $w(i, j)$ is the noise that corrupts the blurred image.

The image formation process is typically modeled as a first kind of integral equation which, after discretization, results in a large scale linear system of the following type

$$g = Df + w \quad (2.4)$$

where g is the blurred noisy copy of f . D is a large ill-posed matrix representing the blurring, and w is modeling noise. The observation noise w in the image formation model is also assumed to be a homogeneous Gaussian distributed white noise process with zero mean and covariance matrix $R = \sigma_w^2 I$ ($\sigma_w^2 > 0$), being uncorrelated with v . f and g are the $(N^2 \times 1)$ original image vector and observation vector (blurred and noisy observed image) respectively. A is the $(N^2 \times N^2)$ image model coefficients matrix and D is the $(N^2 \times N^2)$ blur model

coefficients matrix [35] . Figure 2.2 shows the image formation model in the spatial domain

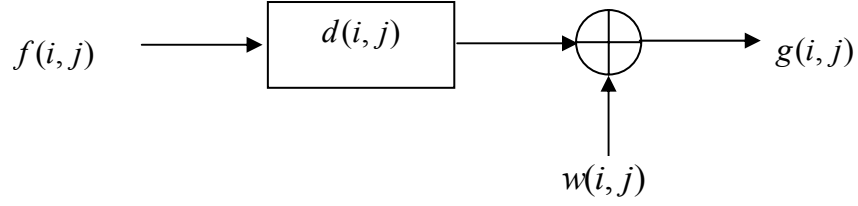


Figure 2-2 : Image formation model in the spatial domain.

Image restoration algorithms are based on an observation model that represents the relationship between the input and the output of the imaging system [9] . Clearly the objective of image restoration is to make an estimate $\hat{f}(i, j)$ of the ideal image $f(i, j)$, given only the degraded image $g(i, j)$, the blurring function $d(i, j)$ and some information about the statistical properties of the ideal image and the noise.

The pdf of f , given the pdf of v and the model (2.2), is then given by:

$$p(f; A, Q) = \sqrt{\frac{\det|I - A|^2}{2\pi^{N^2} \det|Q|}} \exp \left\{ -\frac{1}{2} f^T (I - A)^T Q^{-1} (I - A) f \right\} \quad (2.5)$$

The pdf of g , given the pdf of the observation noise, the model (2.4), and the original image f , can be expressed as:

$$p(g|f; D, R) = \frac{1}{\sqrt{2\pi^{N^2} \det|R|}} \exp \left\{ -\frac{1}{2} (g - Df)^T R^{-1} (g - Df) \right\} \quad (2.6)$$

2.3. State Space Image and Observation Models

The state of a system, which represents the minimum amount of information on the past behavior of the system, is sufficient to predict the future response of the system

[36] . Sequential data arises in many areas of science and engineering. We can assume that there is some basic hidden state of the world that generates the observations. The basic idea behind state-space models is that this hidden state progresses in time, possibly as a function of our inputs. In the state-space form the relationship between the input, noise and output signals is written as a system of first-order differential or difference equations [37] .

The 2-D image data can be ordered as the notion of past, present and future. The past, present and future pixels are defined with respect to the present pixel (i, j) . The 2-D image restoration problem is converted into an equivalent One-Dimensional (1-D) problem through raster scanning of the image [38] . By using the state vector definition $f(i, j)$, image and observation model can be expressed as the following state-space form:

$$f(i, j) = Af(i-1, j) + v(i, j) \quad (2.7)$$

$$g(i, j) = Df(i, j) + w(i, j) \quad (2.8)$$

where $f(i, j)$ is the image state vector at pixel (i, j) , $v(i, j)$ is the additive noise process, $g(i, j)$ is the observation vector of the blurred image, $w(i, j)$ is the observation noise vector. The state support is given in Figure 2.3. The state vector $f(i, j)$ can be defined as follows:

$$f(i, j) \triangleq [f(i, j), f(i-1, j), \dots, f(1, j); f(L, j-1), f(L-1, j-1), \dots, f(1, j-1); \dots; f(L, j-M_1), f(L-1, j-M_1), \dots, \dots, \dots, f(i-M_2, j-M_1)]^T \quad (2.9)$$

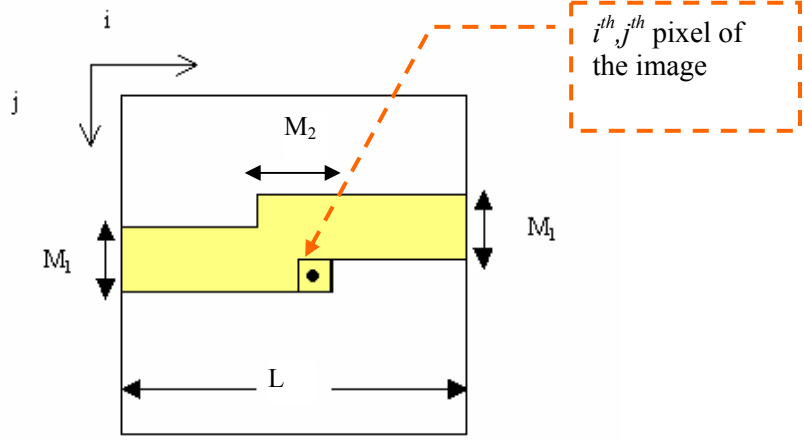


Figure 2-3: State support.

We can write the following equations in the matrix form:

$$f_k = Af_{k-1} + v_{k-1} \quad (2.10)$$

$$g_k = Df_k + w_k \quad (2.11)$$

where f_k and g_k are the $p \times 1$ state vector and the observation vector respectively. A is the state transition matrix and D is the observation matrix. v_k and w_k are zero mean white Gaussian sequences whose covariance matrices are given by $Q = \sigma_v^2 I$ and $R = \sigma_w^2 I$, respectively.

2.3.1. Hidden Markov Models and Dynamic Bayesian Networks

The Hidden Markov Model (HMM) is a finite set of *states*, each of which is associated with a probability distribution [39]. HMMs are used for the statistical modeling of non-stationary signal processes that can be defined as one whose statistical parameters vary over time as a probabilistic Markovian chain of N stationary sub-process. Also, a non-stationary process can be modeled as a double-layered stochastic process, with a hidden process [40].

A Bayesian network is simply a graphical model for representing conditional independencies between a set of random variables [41]. Each variable is represented by a node in the network. Many time series models, including the HMMs and

Kalman Filter Models (KFM) used in filtering and control applications, can be viewed as examples of Dynamic Bayesian Networks (DBNs) [42]. Since, a Bayesian Network is a graphical way to represent a particular factorization of joint distribution; we propose that state space image model, which is defined in (2.5) and (2.6), can be represented as a DBN. A Bayesian network representing of our state space image model is given in Figure 2.4.

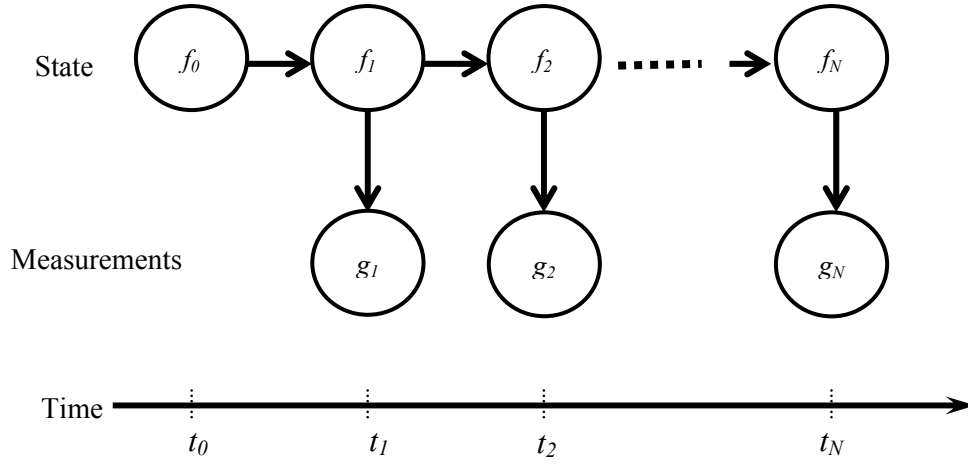


Figure 2-4: A Bayesian network representing of state space image model.

Because of the Markov structure of the state, the joint pdf can be factored into the product of conditional distributions of states,

$$p(\{f\}, \{g\}) = p(f_0) \prod_{k=1}^N p(f_k | f_{k-1}) \prod_{k=1}^N p(g_k | f_k) \quad (2.12)$$

Learning Bayesian networks from data has become an increasingly active area of research. Very recently, researchers have begun to deal with the problem of learning the structure of the network from incomplete data. Most of the research to date has relied on the assumption that data are complete; that is, the values of all variables are known for all cases in the database. This assumption is not very realistic, since most real world situations involve incomplete information. Myers et al. [43] described an evolutionary algorithm approach to learning Bayesian networks from incomplete data.

A major part of this research is that when information is missing, closed form expressions do not exist. This has led many researchers down the path of estimating the score using parametric approaches such as the expectation-maximization (EM) algorithm [44] .

Penny and Roberts [45] investigated the differences between dynamic linear models, Recursive Least Squares (RLS) and Steepest-Descent learning. All models can be viewed as special cases of linear dynamic systems and their learning rules as special cases of the Kalman filter. Dynamic linear models are adaptive learning rate algorithms for modeling stationary and non-stationary data [46] .

2.4. Optimization Methods

Image restoration and identification problems are typically posed as optimization problems in which a suitable objective function is minimized under certain constraints. The objective of this section is to present practical optimization techniques used in this thesis for image restoration and identification.

The fundamental problem of optimization is to obtain the best possible decision in any given set of circumstances. Many engineering problems, particularly in image processing, may be expressed in terms of nonlinear multivariate unconstrained optimization. Unconstrained optimization gives the best solution based on numerical minimization of a single, scalar-valued objective function or cost function. Unconstrained optimization problems have been intensively studied, and many algorithms and tools have been developed to solve them [48] [49] . The basic unconstrained optimization problem is to minimize a real-valued objective function $f(x)$ over all vectors $x \in \mathbb{R}^n$. Consider the following unconstrained optimization problem,

$$\min_{x \in \mathbb{R}^n} f(x) \tag{2.13}$$

Numerical optimization methods require some preliminary definitions.

Definition 2.1: Gradient (First Order Derivative)

We assume that $F(x)$ is a multivariate function and the gradient of $F(x)$ will be denoted by $\nabla F(x)$ or $g(x)$:

$$\nabla F(x) \equiv g(x) \equiv \begin{bmatrix} \frac{\partial F(x)}{\partial x_1} \\ \vdots \\ \frac{\partial F(x)}{\partial x_n} \end{bmatrix} \quad (2.14)$$

Definition 2.2: Hessian (Second Order Derivative)

The Hessian matrix of $F(x)$, which will be denoted by $\nabla^2 F(x)$, or $H(x)$, is defined:

$$\nabla^2 F(x) \equiv H(x) \equiv \begin{bmatrix} \frac{\partial^2 F(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 F(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 F(x)}{\partial x_n^2} \end{bmatrix} \quad (2.15)$$

Definition 2.3: Convergence

Suppose that a sequence $\{x^k\}$ converges to x^* as $k \rightarrow \infty$. The rate of the convergence of a sequence $\{x^k\}$ is measured in the following way.

- If there exists a constant C , $0 \leq C < 1$, such that for k sufficiently large,

$$\|x^{k+1} - x^*\| \leq C \|x^k - x^*\| \quad (2.16)$$

then it is said to be convergent linearly.

- If

$$\|x^{k+1} - x^*\| \leq C \|x^k - x^*\|^{1+\kappa} \quad (2.17)$$

with $0 < \kappa < 1$, then it is said to be convergent super-linearly.

- If

$$\|x^{k+1} - x^*\| \leq C \|x^k - x^*\|^2 \quad (2.18)$$

then it is said to be convergent quadratically.

Definition 2.4: Condition Number

The condition number describes the ill-condition or bad behavior of a matrix quantitatively [50]. The condition number of a nonsingular matrix A is defined as:

$$\text{cond}(A) = \frac{\lambda_M(A)}{\lambda_m(A)} \quad (2.19)$$

Here $\lambda_M(A)$ and $\lambda_m(A)$ are the maximum and minimum eigenvalues of the A matrix, respectively.

Suppose that f is twice continuously differentiable. As we know, a necessary condition for the point x^* be an optimal solution for (2.13) is:

$$\nabla f(x^*) = 0 \quad (2.20)$$

Iterative methods for optimization can be classified into two categories: line search methods and Trust Region (TR) methods [51]. The fundamental structure of local iterative techniques for solving unconstrained minimization problems is uncomplicated. A starting point is chosen; a direction of movement is set according to some algorithm, and a line search or TR approach is performed to determine an appropriate next step. The classical methods for optimization are line search algorithms, but TR algorithms are relatively new algorithms. Of course, only one of the two methods is needed for a given minimization algorithm. There has been no clear evidence for superiority of one class over another [52].

2.4.1. Line Search

In the line search method, the improved estimated point is achieved based only on function values. It is, however, more efficient to utilize the derivative of the function, in addition to the function value, because the first-order derivative (gradient) indicates the direction of change in the function value at the given point. In order to find the solution of a given problem, a general procedure includes an initial solution estimate followed by iterative updates of the solution directed toward minimizing the objective function. To improve the search direction, we can also utilize the second-order derivative. Gradient methods and Newton's methods fall into this category [53]. In the line search strategy, the algorithm chooses a direction d^k and searches along this direction, from the current iterate x^k for a new iterate. The general structure of the line search method is given as,

$$x^{k+1} = x^k + \eta d^k \quad (2.21)$$

where the positive scalar η is called the step length or convergence parameter.

2.4.1.1. Gradient Descent Methods

One of the first and well-known methods for unconstrained optimization is the gradient descent method in which the negative gradient directions are used to find local minimizers of a differentiable function. Standard gradient-based iterative optimization algorithm is formulated as follows:

$$x^{k+1} = x^k - \eta \nabla f(x^k) \quad (2.22)$$

where the initial point x^0 is given and is determined by a linear search procedure.

The difference in convergence rates due to the difference in magnitude of the eigenvalues is referred to as being due to the eigenvalue spread, which is defined as λ_M / λ_m . We know that a matrix is poorly conditioned if the eigenvalue spread is

large. Poorly conditioned matrix also has slow convergence properties when it is used in a gradient descent algorithm [54] .

For sufficiently large k , convergence of the gradient descent algorithm is given in the following structure.

$$\frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} \leq \|I + \eta H(x^*)\| < 1 \quad (2.23)$$

For the gradient descent method, the step size η must be chosen to ensure that (2.23). This requires a one dimensional line search or an optimization of η at each iteration, which requires extra computation. An alternative is to fix η to a very small value, which generally makes $\|I + \eta H(x^*)\|$ close to one and results in slow convergence.

2.4.1.2. Newton Type Optimization

Newton's method has been devised for minimization by utilizing the first and the second derivatives of the function. Newton's method is an iterative method for finding a solution of an equation of the form $f(x) = 0$. Newton's method is based on the idea of approximating $\nabla f(x)$ with its linear Taylor series expansion about a working value x^k .

Based on the idea that any valley of the function containing the minimum point at the bottom looks similar to a quadratic function, we can fit a quadratic function, $\nabla f(x)$, through x^k that matches its first and second derivatives with that of the function $f(x)$. The Newton's method is summarized as:

$$x^{k+1} = x^k - H^{-1}(x^k) \nabla f(x^k) \quad (2.24)$$

where $H(x^k)$ is the Hessian of $f(x)$. If the Hessian is positive definite, the extremum is a (local) minimum. The search direction is given as,

$$d^k = -H^{-1}(x^k) \nabla f(x^k) \quad (2.25)$$

Newton's method is powerful and simple to implement. Gradient descent algorithm takes many steps to converge, while Newton's method reaches the minimum in a few steps. However Newton's method is more expensive, because it requires both gradient and the Hessian to be evaluated. It converges to a fixed point from any sufficiently close starting value. However global convergence properties of Newton's method are poor [55]. The second-order derivative of a single-variable function, $f''(x)$, is equivalent to the Hessian of a multiple-variable function, $H(x)$. Based on this observation, every update by the Newton's direction does not guarantee a decrease in the function value. Moreover, even if the Hessian is positive definite, Newton's method may not be the best method for finding the minimum. This may occur if the starting point, x^0 , is far away from the solution.

2.4.1.2.1. Quasi-Newton Methods

One of the drawbacks of Newton's method is that it requires the analytic derivative $H(x)$ at each iteration. This is a problem if the derivative is very expensive or difficult to compute. In such cases it may be convenient to iterate according to

$$x^{k+1} = x^k - A^{-1}(x^k) \nabla f(x^k) \quad (2.26)$$

where $A(x^k)$ is an easily computed approximation to $H(x^k)$. Such an iteration is called a quasi-Newton method [56] [57]. If $A(x^k)$ is positive definite, as it usually is, an alternative name is variable metric (VM) method.

One positive advantage to using an approximation in place of $H(x)$ is that $A(x^k)$ can be chosen to be positive definite, ensuring that the step will not be attracted to a maximum of f when one wants a minimum. Another advantage is that

$A^{-1}(x^k)\nabla f(x^k)$ is a descent direction from x^k , allowing the use of line searches.

The best known quasi-Newton method in statistical contexts is Fisher's method of scoring, which is treated in more detail in the next section.

2.4.1.2.2. Fisher Scoring Algorithm

In some problems, the objective function and its gradient are available, but the Hessian matrix can not be computed or stored [58]. Similar problems are often solved by the FS Method (The Method of Scoring) [59]. FS algorithm is a Newton-based method and also known as VM approach for unconstrained optimization [60]. The FS algorithm is identical to the Newton-Raphson algorithm with one exception. The FS algorithm replaces $H(x)$ by its expectation $I_f(x)$, which is the expected Fisher Information Matrix (FIM). Since this matrix is the covariance matrix, it is always positive definite. It is expected that replacing the second derivative by its expected value will increase the stability of the iteration.

The FIM, $I_f(x)$ is defined by

$$I_f(x) = -E[H(x)] \quad (2.27)$$

where $E[\bullet]$ is the expected value operation. FS algorithm can be defined in the following form,

$$x^{k+1} = x^k + I_f^{-1}(x^k)\nabla f(x^k) \quad (2.28)$$

For many models, computation of the FIM is much simpler than $H(x)$. Furthermore, FIM is positive definite for any parameter value x for which the statistical model is not degenerate. FS is linearly convergent at a rate, which depends on the relative difference between observed and expected information.

Apart from these, there are some different techniques such as Conjugate Gradient (CG), Broydon-Fletcher-Goldfarb-Shanno (BFGS) algorithm, Davidon-Fletcher-Powell (DFP) algorithm and preconditioned based methods [61] [62] [63] [64].

2.4.2. Trust Region

Even though Newton's method usually requires the least iteration among all the methods, it has one disadvantage of using a step size of unity maybe not leading to converge. To overcome this drawback, two classes of methods exist to modify the pure Newton's method so that it is guaranteed to converge to a local minimum from an arbitrary starting point. The first of these, called Trust Region (TR) methods, minimize the quadratic approximation within an elliptical region, whose size is adjusted so that the objective improves at each iteration. The TR is the region, in which objective function is quadratic function and convex, guaranteed its local to be global optimum. On the other hand, the second method results in a slightly different minimum point at the cost of the function modification.

At each iteration of the TR method, the algorithm determines a step by finding the minimum of a quadratic function restricted to a given ball of radius s . The origin of TR Methods is found in the works of Levenberg [65] and Marquardt [66]. The idea of defining a region of trust for the search direction was suggested for nonlinear least squares problems. The application of the technique to general nonlinear problem was considered by Goldfeld *et al.* [67]. TR methods produce a trial step by minimizing a quadratic model of the objective function subject to a constraint on the length of the trial step. Because of this restriction, TR methods are sometimes known as restricted-step methods.

In principle, the implementation of a trust region requires the solution s^k of a quadratic constrained minimization problem,

$$\min_s Q^k(s) \text{ subject to } \|s\| \leq \Delta^k \quad (2.29)$$

Δ^k is a positive scalar called the TR radius, which is varied as the iteration proceeds. This yields iterates of the form,

$$x^{k+1} = x^k - [H(x^k) + \gamma^k I]^{-1} \nabla f(x^k) \quad (2.30)$$

where γ^k is zero if the constraint in (2.30) is inactive at iteration k , and γ^k is a positive Lagrange multiplier otherwise.

These methods have advantages over e.g., quasi-Newton methods [68]. The approximate model is only “trusted” in a region near the current iterate. This seems reasonable, because for nonlinear functions local approximate models can only fit the original function locally. However, their implementation can be more problematic than a line search implementation. TR algorithms are reliable and robust, they can be applied to ill-posed problems, and they have strong convergence properties [69]. TR methods tend to be more robust for ill-posed problems than a line search.

2.4.2.1. LM Method

Newton’s method requires a second derivative. For some problems, obtaining the Hessian can be difficult or impossible. So a variety of methods have been developed for estimating the Hessian may not be positive definite: some of the eigenvalues may be zero. The iterations cannot converge because of the zero eigenvalues. For these situations, Levenberg-Marquardt (LM) is used, which replaces the Hessian with a nearby matrix that is positive definite [70].

LM is a modified Newton method in which the objective Hessian is again approximated using the Jacobian matrix, but a TR approach is used in place of the line search in Gauss-Newton. The advantage of the TR technique is that it avoids problems involving rank deficiencies in the Jacobian matrix.

Newton’s algorithm is guaranteed to converge only when the Hessian matrix is positive definite. It was observed by Fletcher [71] that, even in the case of positive definite Hessians, convergence may not be assured. A robust algorithm to solve singular systems was developed independently by Levenberg and Marquardt and is known LM algorithm. Minimization of a quadratic function over a sphere can be done in a very efficient way by use of LM algorithm [72].

The LM algorithm has the same form as Eq. (2.30). In LM, $\gamma^k I$ is a diagonal matrix chosen to force $[H(x^k) + \gamma^k I]$ to be positive definite so that $[H(x^k) + \gamma^k I]^{-1}$ will always be computable. Choosing the initial values of γ^k , was discussed by Mendel [73]. The LM algorithm computes the log-likelihood function at each step. If $L(\theta_{ML}^{i+1}) < L(\theta_{ML}^i)$ condition is satisfied, he increased γ^k and tried again. Convergence properties of the LM algorithm have been investigated in [74] [75] [76] [77].

2.4.3. EM Algorithm

The EM algorithm is an iterative technique for ML estimation. The recent literature on probabilistic model has led to increased interest in EM as a possible alternative to gradient based methods for optimization. The EM algorithm, which is pioneered by Baum and colleagues [78] and later generalized by Dempster et al.[79], is a very popular and widely used algorithm for the computation of ML estimates. Hundreds of papers have been published about EM algorithm in many areas. The EM algorithm has been widely used in econometric, sociological studies, some applications of statistical methods and estimating parameters of mixture distributions [80]. For instance, Frenkel and Feder [81], applied the EM algorithm to multiple target tracking for a known number of targets. Three major algorithms were proposed based on different optimization criteria.

The EM algorithm is a framework for solving problems where it is difficult to obtain a direct ML estimate either because the data is incomplete or because the problem is difficult. The EM algorithm is specifically designed for problems with such many-to-one mappings. Let Y denote the sample space of the observations, and let $y \in R^m$ denote an observation from Y . Let X denote the underlying space and let $x \in R^n$ be an outcome from X , with $m < n$. The “complete data” x is not observed directly but is related to the observed data vector (“incomplete data”) with $y = h(x)$, where $h(\bullet)$ is a noninvertible transformation. The complete data x are assumed to have probability density $p_{X;\theta}(x; \theta)$, which is a function of a parameter vector θ as well as

of x . Maximization of the likelihood of the incomplete data, $p_{Y;\theta}(y;\theta)$, with respect to the parameter vector θ is often a difficult task, whereas maximization of the likelihood of the complete data $p_{X;\theta}(x;\theta)$ is relatively easy. Since the complete data is unavailable, the parameter estimate is obtained through maximization of the conditional expectation of the log-likelihood of the complete data defined as

$$E\left[\ln p_{X;\theta}(X;\theta)|y\right] = \int_X p_{X|Y;\theta}(x|y;\theta) \ln p_{X;\theta}(X;\theta) dx \quad (2.31)$$

The EM algorithm starts with an arbitrary initial guess. The following iteration cycle can then be described in two steps as E and M . In order to efficiently apply the EM algorithm, two requirements should be satisfied: *i)* In the E-step, we should be able to compute the conditional pdf of the “hidden variables” given the observation data. *ii)* In the M-step, it is highly preferable to have analytical formulas for the update equations of the parameters. On the other hand, in many problems, it is not possible to meet the above requirements and several variants of the basic EM algorithm have appeared.

In the E-step of the EM algorithm, the conditional expectation

$$Q(\theta, \theta^n) = E\left[\log p(x, \theta)|y, \theta^n\right] \quad (2.32)$$

is computed. Here θ^n is the current estimated value of θ .

In the M-step, the θ maximizing $Q(\theta, \theta^n)$ is found. This yields the new parameter estimate θ^{n+1} , and this two-step process is repeated until convergence. The essence of the EM algorithm is that increasing $Q(\theta, \theta^n)$ forces an increase in the log-likelihood $L(\theta)$ of the observed data.

2.5. Ill-Posed Problems

Inverse problems occur in a variety of important applications in science and industry. Unfortunately, a small amount of noise in the data can lead to massive errors in the estimates. This instability phenomenon is called ill-posed.

It is not practical to calculate the solution of ill-posed problems by the direct method. In the case of the ill-posed problem, approximate solutions are determined that are stable under small changes in the initial data based on the use of additional information. Mathematical techniques known as regularization methods have been developed to deal with ill-posedness [82]. Tikhonov regularization is the most well-known one [83].

Regularization is the most widely used method to efficiently solve practical optimization problems with one or more constraints. Major advantages of regularization can be summarized as:

- (i) simple and intuitive formulation of the cost function and
- (ii) flexibility in incorporating one or more constraints into the optimization process.

The term regularization refers to the transformation of an objective function with constraints into a different objective function, automatically reflecting constraints in the unconstrained minimization process. Because of its simplicity and efficiency, regularized optimization has many application areas, such as image restoration, image reconstruction, optical flow estimation, etc.[84].

As we have pointed out before, the image restoration problem is ill-posed. There exist several algorithms for the ill-posed image restoration problems based on regularization [85] [86] [87] [88]. Every regularization method uses a parameter to control the effect that the noise in the data has on the approximate solution. This parameter is called the “regularization parameter” and controls the trade-off between fidelity to the observation and smoothness of the estimated parameters [89]. According to Tikhonov, the regularization method consists of finding regularizing operators that operate on the data, and determining the regularization parameter.

Marroquin et al. [90] reviewed standard regularization theory, discussed its limitations, and presented new stochastic methods based on Bayesian techniques for their solution. Implementation of Tikhonov regularization is very expensive for large-scale applications [91].

2.6. Solving Equations in DFT Domain

In the image degradation process, D and A matrices are Block Toeplitz Toeplitz Block (BTTB) matrices. By padding g and f properly with zeros so that the results of linear and circular convolution are the same, D and A are Block Circulant Circulant Block (BCCB) matrices [92] [93]. Definition of the Toeplitz, Circulant, BTTB and BCCB matrices are given in the following definitions.

Definition 2.5: Toeplitz Matrix

A matrix is called Toeplitz if it is constant along diagonals. An $n \times n$ Toeplitz matrix T has the form

$$T = \begin{bmatrix} t_0 & t_{-1} & \cdots & t_{1-n} \\ t_1 & t_0 & \ddots & t_{2-n} \\ \vdots & \ddots & \ddots & \vdots \\ t_{n-1} & t_{n-2} & \cdots & t_0 \end{bmatrix} \quad (2.33)$$

Definition 2.6: Circulant Matrix

An $n \times n$ matrix C is called circulant if it is Toeplitz and its rows are circular right shifts of the elements of the preceding row. In this case we can write:

$$C = \begin{bmatrix} c_0 & c_{n-1} & \cdots & c_1 \\ c_1 & c_0 & \ddots & c_2 \\ \vdots & \ddots & \ddots & \vdots \\ c_{n-1} & c_{n-2} & \cdots & c_0 \end{bmatrix} \quad (2.34)$$

Definition 2.7: BTTB Matrix

An $n_x n_y \times n_x n_y$ matrix TB is called block Toeplitz with Toeplitz blocks (BTTB), if it has the block form

$$TB = \begin{bmatrix} T_0 & T_{-1} & \cdots & T_{1-n_y} \\ T_1 & T_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & T_{-1} \\ T_{n_y-1} & T_{n_y-2} & \cdots & T_0 \end{bmatrix} \quad (2.35)$$

where each block T_j is an $n_x \times n_x$ Toeplitz matrix.

Definition 2.8: BCCB Matrix

A $n_x n_y \times n_x n_y$ matrix CB is block circulant with circulant blocks (BCCB), if (i) CB is BTTB; (ii) the $n_x \times n_x$ block rows of CB are all circulant right shifts of each other; and (iii) each block is a circulant matrix. In other words,

$$CB = \begin{bmatrix} C_0 & C_{n_y-1} & \cdots & C_1 \\ C_1 & C_0 & \ddots & C_2 \\ \vdots & \ddots & \ddots & \vdots \\ C_{n_y-1} & C_{n_y-2} & \cdots & C_0 \end{bmatrix} \quad (2.36)$$

We assume that the image is $n \times n$, where n is the number of pixels in each direction, and thus contains n^2 pixels, and accordingly D and A matrices are $n^2 \times n^2$. Typically n is chosen to be a power of 2, such as 256 or larger, then the number of unknowns grows to at least 65.536. Because of the large size of matrix, the direct methods should be avoided.

In order to utilize the special structure of the Toeplitz matrix D and A , each BTTB matrix is replaced a block-circulant-circulant-block (BCCB) matrix by padding appropriately with zeros. In Figure 2.5, padding process is given for a 3×3 blur operator.

BCCB matrices can be diagonalized with the use of the 2-D Discrete Fourier Transform (DFT) [94] [95]. This is because the eigenvalues of a BCCB matrix are the 2-D Discrete Fourier coefficients of the impulse response of the degradation system which is used in uniquely defining D , and the eigenvectors are the complex

exponential basis function if this transform. In the matrix form, this relationship can be expressed by

$$D = \Omega \Lambda \Omega^{-1} \quad (2.37)$$

where Λ is a diagonal matrix comprising the 2-D DFT coefficients of $d(i, j)$, and Ω is matrix containing the components of the complex exponential basis functions of the 2-D DFT. In this context, since D and A have circulant structure, minimization of the quadratic function can be dramatically simplified by using frequency-domain implementation. Using this diagonalization approach, it can be solved as a set of N^2 scalar problems.

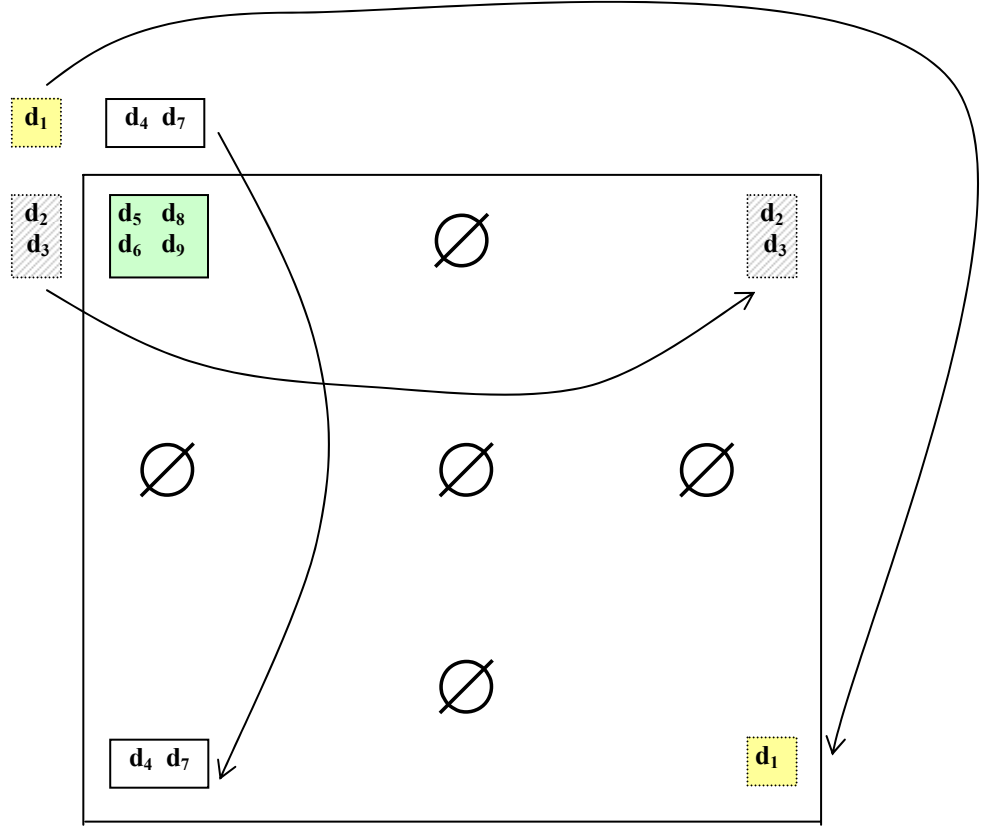


Figure 2-5 : Periodically extended PSF.

2.7. Concluding Remarks

In this chapter, we give some mathematical background and definitions, which are required the rest of thesis. Convergence analysis of the optimization methods and some preliminary definitions are given in this chapter. We reviewed ill-posed structure of the image identification and restoration problem and regularization methods.

3. IMAGE RESTORATION AND IDENTIFICATION

In this section, classical image restoration and identification methods are reviewed. First, some metrics to measure the degradation of corruption and improvement in image restoration are given.

3.1. Measuring and Metrics

The signal-to-noise-ratio (SNR) of the recorded image is defined as follows in decibels:

$$SNR_{observed} = 10 \log_{10} \left(\frac{\text{variance of the ideal image } f(i, j)}{\text{variance of the diffrence image } g(i, j) - f(i, j)} \right) dB \quad (3.1)$$

The signal-to-noise-ratio of the restored image is similarly defined as:

$$SNR_{estimated} = 10 \log_{10} \left(\frac{\text{variance of the ideal image } f(i, j)}{\text{variance of the diffrence image } \hat{f}(i, j) - f(i, j)} \right) dB \quad (3.2)$$

In image restoration, the improvement in quality of the restored image over the observed image is measured by the improvement in signal-to-noise-ratio (SNR_{imp}). The SNR improvement is defined as the difference between the SNRs of the images before and after restoration and is given by the following equation,

$$SNR_{imp} = SNR_{estimated} - SNR_{observation} \quad (3.3)$$

$$SNR_{imp} = 10 \log_{10} \frac{\sum_{i,j}^N (f(i, j) - g(i, j))^2}{\sum_{i,j}^N \left(f(i, j) - \hat{f}(i, j) \right)^2} \quad (3.4)$$

where $\hat{f}(i, j)$ denotes the restored image.

The improvement in SNR is basically a measure that expresses the reduction of difference with the ideal image when comparing the distorted and restored image. A positive SNR improvement indicates that the quality of distorted image is improved, while a negative one indicates deterioration. Note that all of the above signal-to-noise measures can only be computed in case the ideal image $f(i, j)$ is available. When applying restoration filters to real images of which the ideal image is not available, often only the visual decision of the restored image can be used. We are actually engaged in the blind deconvolution problem, where the original image cannot be used. But just for the experimental purposes, we can use the original image to measure the restoration performance.

The learning performance of the blur parameters is tested by calculating the mean square error (MSE) after each step. The MSE at the i^{th} step is given by:

$$MSE(i) = \frac{1}{n} \sum_{i=1}^n [\theta(i) - \hat{\theta}(i)]^2 \quad (3.5)$$

where $\theta(i)$ is the actual output and $\hat{\theta}(i)$ is the estimated output.

3.2. Image Restoration Algorithms

In this section, we review some widespread approaches to image restoration. The blurring of images is modeled in (2.4) with a 2-D PSF $d(i, j)$. An example of the PSF is shown in Figure 3.1.

The spatially continuous PSF of any blur satisfies three constraints:

1. $d(i, j)$ takes on non-negative values,
2. $d(i, j)$ is real-valued,
3. the PSF is constrained to satisfy $\sum_{i=0}^{N-1} \sum_{j=0}^{M-1} d(i, j) = 1$

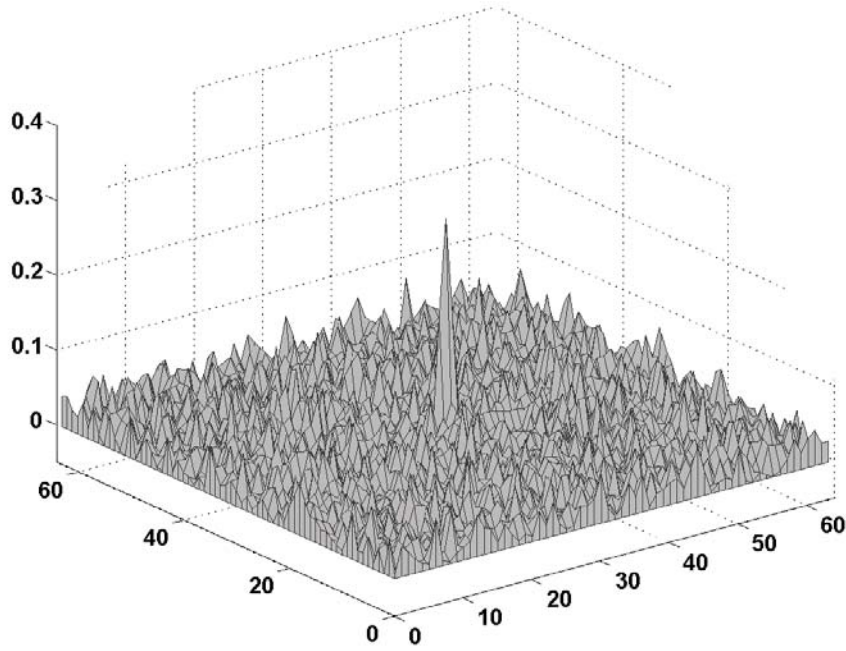


Figure 3-1:A PSF example.

In linear case, the restored image is given by

$$\hat{f}(i, j) = h(i, j) * g(i, j) = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{M-1} h(k_1, k_2) g(i - k_1, j - k_2) \quad (3.6)$$

or in the spectral domain by

$$\hat{F}(u, v) = H(u, v)G(u, v) \quad (3.7)$$

The objective of this section is to give details appropriate restoration filters $h(i, j)$ or $H(u, v)$. Block diagram of the restoration process is given in Figure 3.2.

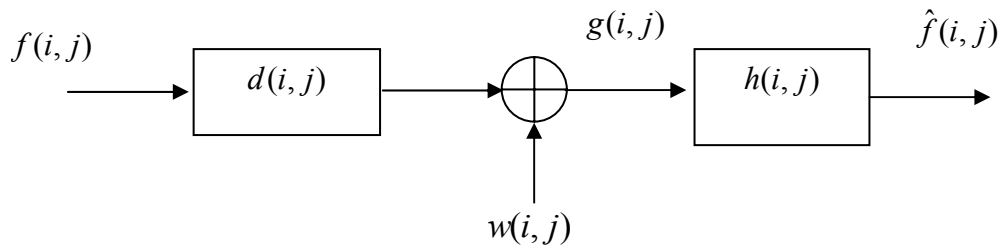


Figure 3-2 : Block diagram of the restoration process.

3.2.1. Inverse Filtering

An inverse filter is a linear filter whose PSF $h(i, j)$ is the inverse of the blurring function $d(i, j)$, in the sense that:

$$h_{inv}(i, j) * d(i, j) = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{M-1} h_{inv}(k_1, k_2) d(i - k_1, j - k_2) = \delta(i, j) \quad (3.8)$$

In frequency domain, we can write (3.8) as the following form.

$$H_{inv}(u, v) D(u, v) = 1 \quad (3.9)$$

The advantage of the inverse filter is that it requires only the blur PSF as a priori knowledge, and that it allows for perfect restoration in the case that noise is absent, as can easily be seen:

$$\hat{F}_{inv}(u, v) = H_{inv}(u, v) G(u, v) = F(u, v) + \frac{W(u, v)}{D(u, v)} \quad (3.10)$$

If the noise is absent, the second term in (3.10) disappears so that the restored image is identical to the ideal image.

In a mathematical sense, the problem of image restoration corresponds to the existence and uniqueness of an inverse transformation. In order to compute the inverse of a matrix, the matrix must be non-singular. One also will recall that a non-singular matrix has a non-zero determinant. It is desirable for the matrix, D , to be well-posed. Even if the inverse transformation exists and is unique, it may be ill-posed by which we mean that a trivial perturbation in g can produce non-trivial significant perturbations in the solution $\hat{f}$.

Since a noiseless image does not exist in real life, noise amplification will become very large. This effect is known as the inverse filtered noise. Inverse filtered images are therefore often dominated by excessively amplified noise. Figure 3.3 shows the effects of an inverse filter applied to the degraded image. Clearly, this is not an acceptable solution for noisy images. The 64×64 Lena image is used for the original

data. As the degradation operation, a 3×3 pixel blurring and 30 dB additive noise are used.

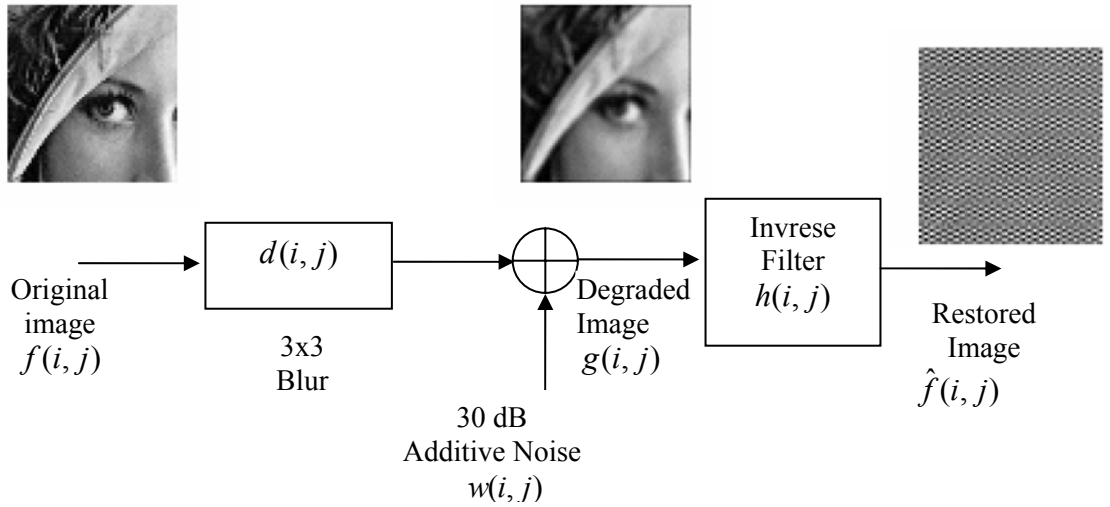


Figure 3-3: Restored image by an Inverse Filter, SNRimp = -14.8 dB.

When we directly solve the ill-posed problem, we usually get a very poorly estimated image.

3.2.2. Least-Squares Filters

To overcome the noise sensitivity problem of the inverse filter, some filters have been developed based on the least-squares structure. In this section, we will give brief information about the Wiener and Kalman filter. The Wiener filter is a batch processing which is usually implemented in the frequency domain. The Kalman filter is a recursive processing which is usually implemented in the spatial domain [96].

3.2.2.1. The Wiener Filter

The Wiener filter is designed to minimize the mean-squared error between the original image and the restoration result. The Wiener filter is defined in the spectral domain easily and can be written in the DFT domain by the following equation.

$$\hat{F}(u, v) = \frac{D^T(u, v)}{D^T(u, v)D(u, v) + \frac{S_{ww}(u, v)}{S_{ff}(u, v)}} G(u, v) \quad (3.11)$$

Here $S_{ff}(u, v)$ and $S_{ww}(u, v)$ are the power spectrum of the ideal image and the noise, respectively. The power spectrum is a measure for the average signal power per spatial frequency (u, v) carried by the image. In the noiseless case we have $S_{ww}(u, v) = 0$, so that the Wiener filter approximates the inverse filter.

3.2.2.2. Kalman Filter

The recursive equivalent of the Wiener filter is the discrete Kalman filter. A Kalman filter is a stochastic, recursive estimator, which estimates the state of a system based on the knowledge of the system input, the measurement of the system output, and a model of the relation between input and output [97]. The Kalman filter makes use of the AR image model and the causal support and gives the linear minimum variance estimate. State space image and observation models were given in (2.10) and (2.11). The standard Kalman algorithm recursively yields the estimates of the time-varying parameter vectors.

The prediction (time update) equations are given as follows:

$$\mu_{k|k-1} = \hat{f}_{k|k-1} = E[f_k | G] = A\hat{f}_{k-1} \quad (3.12)$$

$$M_{k|k-1} = AM_{k-1|k-1}A^T + Q \quad (3.13)$$

The filtering (measurement update) equations make a forward pass through the data to calculate $\hat{f}_k$ and $M_{k|k}$. The filtering equations are:

$$\hat{f}_k = \hat{f}_{k|k-1} + K_k [g_k - D\hat{f}_{k|k-1}] \quad (3.14)$$

$$K_k = M_{k|k-1} D^T \left[D M_{k|k-1} D^T + R \right]^{-1} \quad (3.15)$$

$$M_{k|k} = M_{k|k-1} - M_{k|k-1} D^T \left[D M_{k|k-1} D^T + R \right]^{-1} D M_{k|k-1} \quad (3.16)$$

Here $\hat{f}_k$ denotes the estimate of f_k and K_k is the Kalman gain matrix. $M_{k|k-1}$ and $M_{k|k}$ are the predicted and filtered error covariance matrices, respectively and are defined as

$$M_{k|k-1} = E \left\{ \left(f_k - \hat{f}_{k|k-1} \right) \left(f_k - \hat{f}_{k|k-1} \right)^T \right\} \quad (3.17)$$

$$M_{k|k} = E \left\{ \left(f_k - \hat{f}_k \right) \left(f_k - \hat{f}_k \right)^T \right\} \quad (3.18)$$

3.2.3. Regularized Image Restoration Approaches

It is well known that image restoration problem is an ill-posed inverse problem. Due to the ill-posed nature of the underlying deconvolution problem the discrete system should be regularized. There are several ways this can be done [98] [99] . Ill-posed problems and their regularization were analyzed and characterized for vision systems in [100] . Regularization is an effective method for obtaining satisfactory solutions to problems that involve inversion of ill-posed operators. Regularized image restoration methods can be examined in two groups as deterministic and stochastic regularization.

3.2.3.1. Deterministic Regularization

According to the regularization approach the solution of (2.4) is replaced by the minimization of

$$J_\lambda(f) = \| g - Df \|^2 + \lambda \| \xi f \|^2 \quad (3.19)$$

where ξ is the regularization operator . Galatsanos et al. [98] examined the problem of choosing the regularization parameter and estimating the noise variance. The regularization parameter λ represents the trade-off between fidelity to the data and smoothness of the estimate $\hat{f}$. The value of the regularization parameter depends on the variance of the noise in the data and the properties of D , ξ and f [101] .

If we assume that λ is equal to zero, we can easily see that the solution for (3.19) is identical to inverse solution by ignoring noise w . However, matrix $D^T D$ is better posed than D because the energy of diagonal elements increases by multiplying the matrix by itself. Typical choice of ξ is the 2-D Laplacian operator, given by,

$$\xi = \begin{pmatrix} 0 & -\frac{1}{4} & 0 \\ -\frac{1}{4} & -1 & -\frac{1}{4} \\ 0 & -\frac{1}{4} & 0 \end{pmatrix} \quad (3.20)$$

The minimization of Equation (3.19) leads the following equation;

$$\hat{f} = (D^T D + \lambda \xi^T \xi)^{-1} D^T g \quad (3.21)$$

Sarı and Çelebi [102] applied the deterministic regularization technique to restore Three Dimensional (3-D) microscopic images and they choose the regularization parameter experimentally.

3.2.3.2. Stochastic Regularization

Stochastic regularization can lead to the choice of a linear filtering approach that computes the estimate, $\hat{f}$, according to

$$\min E \left\{ \left\| f - \hat{f} \right\|^2 \right\} \quad (3.22)$$

Using a stochastic model for f and n requires some prior knowledge of the statistics of the data, which are then used to regularize the problem. The linear estimate, which minimizes (3.22) is given by

$$\hat{f} = R_{ff} D^T (D R_{ff} D^T + R_{ww})^{-1} g \quad (3.23)$$

Equation (3.23) is the classical formulation of the Wiener filter [103]. R_{ff} is the autocorrelation of the ideal image f and R_{ww} is the autocorrelation of the observation noise and $R_{ww} = R = \sigma_w^2 I$. Since the image model is an AR process, the image autocorrelation matrix is given by,

$$R_{ff} = E[ff^T] = (I - A)^{-1} Q (I - A)^{-T} = \sigma_v^2 (I - A)^{-1} (I - A)^{-T} \quad (3.24)$$

$$R_{ff}^{-1} = (I - A)^T Q^{-1} (I - A) = \frac{1}{\sigma_v^2} (I - A)^T (I - A) \quad (3.25)$$

If we rewrite the (3.23), we can reach the Regularized Wiener filter,

$$\hat{f} = \left(D^T D + \frac{\sigma_w^2}{\sigma_v^2} (I - A)^T (I - A) \right)^{-1} D^T g \quad (3.26)$$

where $\alpha = \sigma_w^2 / \sigma_v^2$ is called as regularization parameter [104].

The 64×64 Lena image is blurred by a 3×3 PSF and added 18 dB additive noise. Figure 3.4 shows the deterministic and stochastic regularization filter results.

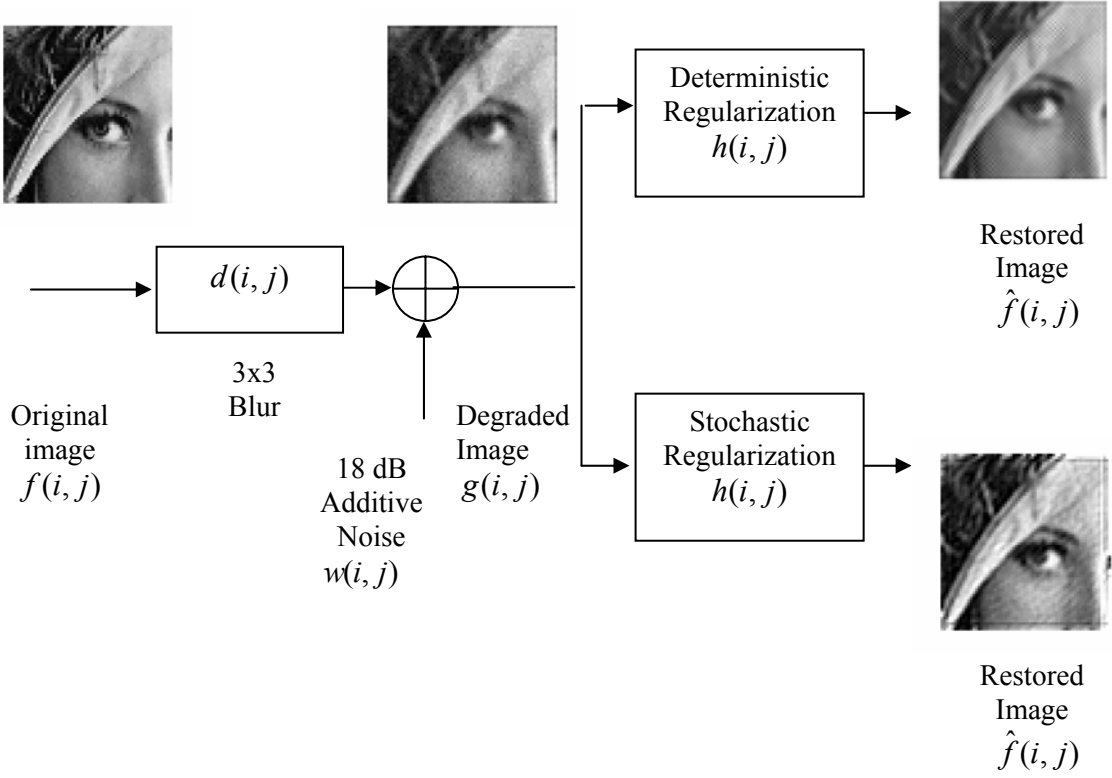


Figure 3-4: (a) Restored image by a deterministic regularization, $\alpha = 0.15$, $\text{SNR}_{\text{imp}} = 2.19$ dB. (b) Restored image by a stochastic regularization $\alpha = 0.117$ $\text{SNR}_{\text{imp}} = 2.67$ dB.

The regularization parameter α plays an essential role in restoring the image. If α is too small, the resulting image cannot avoid noise amplification and cannot satisfactorily reflect *a priori* smoothness limitations. Conversely if α is too large, the resulting image becomes too smooth and it cannot keep detail information of the original image.

3.2.4. Blind Image Restoration

In the previous image restoration methods, it was assumed that the PSF $d(n_1, n_2)$ of the blur was sufficiently known. But in real image restoration problems, blur function, noise characteristics and regularization parameter are not known exactly. In many practical situations, it is difficult to measure the degradation. In this case, we are faced to blind image restoration (deconvolution) problem. For instance, in remote sensing, space imaging and Synthetic Aperture Radar (SAR) applications, fluctuations in the PSF are difficult to characterize as a random process [4] [105]

[106] [107] [108] . In blind deconvolution problem, the system is unknown and its input is inaccessible; hence, precise knowledge of the input signal is not available. Blind deconvolution can be viewed as a self-organized learning process [109] .

Kundur and Hatzinakos [32] developed a blind deconvolution technique for the restoration of linearly degraded images. Explicit knowledge of either the original image or PSF was not required. The only information required for restoration was the non-negativity of the true image and support size of the original object. The blind image deconvolution problem has been solved using Bayesian framework and an EM algorithm based on the variational approximation methodology has been presented to find the parameters of the proposed Bayesian model [5] [110] . Recently, Nakagaki and Katsaggelos [111] proposed a learning-based algorithm for image restoration and blind image restoration based on Vector Quantizer (VQ). In a paper by Bronstein et al. [112] , a novel learning approach of optimal sparse representation was presented for blind deconvolution of images using Quasi-Maximum Likelihood (QML) framework. A blind deconvolution algorithm based on the relative Newton method, which brought further acceleration, was presented in [113] . They utilized special Hessian structure to derive a fast version of the algorithm. Their algorithm constructed an expanding restoration filter, permitting the estimation of long restoration kernels by using cheap iterations.

3.3. Blur and Image Parameter Identification Problem

The first step for image restoration is the identification of degradation [114] . In many practical cases the actual restoration process has to be preceded by the identification of PSF. The initial works on blur identification focused on identifying PSF that have zeros only on the unit bi-circle. Since these zeros can also be located in the spectrum of the blurred image, spectral or cepstral techniques can be used to identify PSF from the distance between spectral zeros of the blurred image [115] [116] .

Blurred and noisy images can often be represented as non-stationary two-dimensional stochastic processes that can be modeled by a set of linear space-varying state equations, or by an Autoregressive Moving Average (ARMA) input-output

equation with space varying coefficients. Tekalp et al. [117] derived ML estimates for these ARMA parameters. Sari and Kayran [118] proposed an identification method based on 2-D ARMA Hybrid Lattice structures. Again, Sari *et al.* [119] applied their methods to identify the blur and image parameters in noisy environment. Reeves and Mersereau [120] presented Generalized Cross-Validation (GCV) method to address the blur identification problem. The GCV criterion identified model parameters for the blur, the image, and the regularization parameter, providing all the information necessary to restore the image. A few studies have been recently investigated for estimation of unknown blur and restoration of the super-resolution (SR) image, simultaneously [7]. Initial results in this area of blind SR have been given in [121].

ML approaches to image/blur identification problems are proposed by Lagendijk et al. [122], Lay and Katsaggelos [123]. The main idea here is to find the parameter values that have most likely resulted in the observed image. Different implementations of ML image and blur identification are discussed by Lagendijk, Tekalp and Bieomond [124]. The blur identification and image restoration problem was formulated as a constrained ML problem. Katsaggelos et al. [125] formulated the same problem as least-squares problem.

3.3.1. Direct ML Blur and Image Parameter Identification

ML estimation is a well-known technique for parameter estimation in situations where no stochastic knowledge is available about the parameters to be estimated [126] [127].

In image and blur identification problem, ML estimation procedures are based on optimizing the pdf of the observed image g with respect to unknown parameters. In this way the ML estimator finds the various values in θ which most likely resulted in the blurred image observed. The unknown parameters are denoted by the vector θ and the ML estimator of the parameter vector θ is defined by [128],

$$\hat{\theta}_{ML} = \arg \left\{ \max_{\theta \in \Theta} L(\theta) \right\} = \arg \left\{ \max_{\theta \in \Theta} \log p(g; \theta) \right\} \quad (3.27)$$

Here $L(\theta)$ denotes the (log)-likelihood function of θ and $p(g; \theta)$ denotes the pdf of the observed image given θ , and Θ specifies the range of the parameters θ . Image and blur models can be specified in terms of the deterministic parameters of the parameter vector θ by,

$$\theta = [\theta_1, \theta_2, \dots, \theta_M]^T = [\dots d(m, n) \dots a(k, l) \dots \sigma_v^2, \sigma_w^2]^T = [\underline{d}^T, \underline{a}^T, \sigma_v^2, \sigma_w^2]^T \quad (3.28)$$

where $\underline{d}$ is the $(L \times 1)$ unknown blur parameter vector and $\underline{a}$ is the $(K \times 1)$ unknown image parameter vector.

The image identification and restoration problem is the estimation of these parameters and the determination of an image as close as possible to the original one. The ML image and blur identification problem can be expressed as a minimization problem for the negative of the $L(\theta)$. If image model is replaced into the observation model, we can obtain the following equation.

$$g = Df + w = D(I - A)^{-1}v + w \quad (3.29)$$

The ML image identification problem is given as,

$$\hat{\theta}_{ML} = \arg \left\{ \min_{\theta \in \Theta} L(\theta) \right\} = \arg \min_{\theta \in \Theta} \left\{ \log(\det|P|) + g^T P^{-1} g \right\} \quad (3.30)$$

Pdf of the observed image is a Gaussian process with zero mean covariance matrix P . Covariance matrix P is given by:

$$\begin{aligned} P = \text{Cov}(g; \theta) &= E \left\{ (D(I - A)^{-1}v + w)(D(I - A)^{-1}v + w)^T \right\} \\ &= \sigma_v^2 D(I - A)^{-1}(I - A)^{-T} D^T + \sigma_w^2 I \end{aligned} \quad (3.31)$$

It is not possible to compose a simple closed-form, because $L(\theta)$ is a complicated nonlinear behavior. The only way to maximize the log-likelihood function is by

means of mathematical programming. The use of ML estimation in practice leads to difficult nonlinear programming problems.

The most common method to optimize $L(\theta)$ is the gradient-based method, which requires the computation of the log-likelihood gradient [129]. Standard gradient-based iterative optimization algorithm was formulated in Chapter 2. The gradient of the log-likelihood function should be calculated with respect to each unknown parameter. The partial derivative of $L(\theta)$ with respect to blur, image, model noise and observation noise can be calculated easily.

3.3.1.1. Gradient Based Optimization

Standard gradient-based iterative optimization algorithm,

$$\hat{\theta}^{i+1} = \hat{\theta}^i - \eta \frac{\partial L(\theta)}{\partial \theta^i} \quad (3.32)$$

where η is the convergence parameter which must be in the interval $0 < \eta < 2/\lambda_M$. Here λ_M is the largest eigenvalue of the Hessian of the log-likelihood function.

For the image identification problem, the partial derivative of $L(\theta)$ with respect to one of the elements of θ is given by:

$$\frac{\partial L(\theta)}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} [\log \det |P|] + \frac{\partial}{\partial \theta_i} [g^T P^{-1} g] \quad (3.33)$$

$$\frac{\partial L(\theta)}{\partial \theta_i} = \text{tr} \left\{ \frac{\partial P}{\partial \theta_i} P^{-1} \right\} - g^T P^{-1} \frac{\partial P}{\partial \theta_i} P^{-1} g \quad (3.34)$$

If the matrices D and A have a BC structure, and P has a BC. Furthermore, the covariance matrix P is positive definite. Since there is always noise present in a blurred image, this condition is satisfied. As a consequence the inverse of P and the logarithm of $\det |P|$ always exist. The likelihood function $L(\theta)$ can be evaluated very

efficiently in the frequency domain because of BC structure of the D and A matrices. In DFT domain,

$$\frac{\partial L(\theta)}{\partial \theta_i} = \sum_{u,v} \left\{ \left[\frac{1}{P(u,v)} - \frac{|G(u,v)|^2}{P(u,v)^2} \right] \frac{\partial P(u,v)}{\partial \theta_i} \right\} \quad (3.35)$$

provided that the matrices A and D have a BC structure. $\frac{\partial P}{\partial \theta_i}$ can be written for blur and image parameters in the following form,

$$\frac{\partial P}{\partial d(m,n)} = \sigma_v^2 (I - A)^{-1} (I - A)^{-T} \{ E_{m,n} D^T + E_{m,n}^T D \} \quad (3.36)$$

$$\frac{\partial P}{\partial a(k,l)} = \sigma_v^2 D (I - A)^{-1} (I - A)^{-T} D^T \{ E_{k,l} (I - A)^{-1} + E_{k,l}^T (I - A)^{-T} \} \quad (3.37)$$

where $E_{x,y}$, is an elementary matrix, which is BT matrix defined as all zero except for a 1 at location (x, y) in each block. In the DFT domain, the above equations can be written as,

$$\begin{aligned} \frac{\partial P(u,v)}{\partial d(m,n)} = \sigma_v^2 \frac{1}{|1 - A(u,v)|^2} & \left\{ D(-u, -v) \exp \left[-\frac{2\pi j}{N} (um + vn) \right] \right. \\ & \left. + D(u, v) \exp \left[\frac{2\pi j}{N} (um + vn) \right] \right\} \end{aligned} \quad (3.38)$$

$$\begin{aligned} \frac{\partial P(u,v)}{\partial a(k,l)} = \sigma_v^2 \frac{|D(u,v)|^2}{|1 - A(u,v)|^2} & \left\{ \frac{1}{1 - A(u,v)} \exp \left[-\frac{2\pi j}{N} (uk + vl) \right] \right. \\ & \left. + \frac{1}{1 - A(u,v)} \exp \left[\frac{2\pi j}{N} (uk + vl) \right] \right\} \end{aligned} \quad (3.39)$$

The likelihood function is often optimized using a gradient approach starting from some a priori values. It is very common to find situations where the likelihood surface has multiple maxima, saddle-points, discontinuities, and singular Hessian in parameter space. Good starting points are important to ensure convergence to the absolute maximum. Nonlinear programming methods, including random search

methods, can be used to improve the probability of convergence and of locating the absolute maximum of the likelihood function.

3.3.2. Expectation Maximization (EM) Algorithm

In image and blur identification problem, the direct optimization of the likelihood function using the gradient based techniques is not feasible, because of its highly nonlinear character, in a relatively large number of unknowns. Lagendijk [130] proposed an image identification algorithm that is based on the EM algorithm to compute ML parameter estimates. The advantage of this method is such that it avoids operating directly on the nonlinear likelihood function.

In the E step of the algorithm, the conditional expectation of the log likelihood of complete data, given the observed incomplete data and an initial estimate of the parameters, is determined. In the next step (the M step), the parameter values that maximize this conditional expectation are found. A restored version of the noisy and blurred image is determined in the E-step of the algorithm, whereas image and blur identification is performed in the M-step. The steps of the EM algorithm are repeated until convergence.

$$\hat{\theta}_{ML,Y} = \arg \left\{ \max_{\theta \in \Theta} L_Y(\theta) \right\} \quad (3.40)$$

In the EM algorithm, “complete data” vector Y is related to the observed data vector (“incomplete data”) G with the following transformation:

$$G = H(Y) \quad (3.41)$$

where $H(.)$ is a noninvertible transformation. Let $p_Y(y;\theta)$ and $p_G(g;\theta)$ denote the pdfs of the vector Y and vector G , respectively. In this case, the new likelihood function can be optimized by using the following scheme:

$$\hat{\theta}_{ML,Y} = \arg \left\{ \max_{\theta \in \Theta} L_Y(\theta) \right\} \quad (3.42)$$

In many situations, it is simpler to solve (3.42) than to solve (3.27). However, the complete data is not available and we only have the observed data via the non-invertible transformation defined above. The EM algorithm finds the conditional expectation of the log-likelihood of complete data, denoted by $\Gamma(\theta)$, given the observed incomplete data [79] :

$$\Gamma(\theta) = E_{\theta} \left\{ \log p_Y(y; \theta) \middle| G = g; \hat{\theta} \right\} = E_{\theta} \left\{ L_Y(\theta) \middle| G = g; \hat{\theta} \right\} \quad (3.43)$$

where $E_{\theta} \{ \bullet | G \}$ denotes the conditional expectation given G , computed at a pre-specified θ and $G = [g_1 \ g_2 \ \dots \ g_N]^T$. (3.43) is also known as the E-step of the EM algorithm. A restored image is obtained in each E-step of the iterations. The choice of the incomplete data set as $\{g\}$ and the complete data set as $[f \ g]^T$ yields simultaneous identification and restoration. The choice of complete data is not unique and the other choices are given by Katseggelos [131]. The choice of the incomplete data set as $\{g\}$ and the complete data set as $\{f, g\}$ yields simultaneous identification and restoration. In the M-step, log-likelihood function $\Gamma(\theta)$ is maximized with respect to the parameters θ . For instance, the image model coefficients $a(k, l)$ and the PSF coefficients were calculated from a set of linear equations in [10].

3.3.2.1. Convergence of the EM Algorithm

At each iteration of the EM algorithm, the estimated parameter provides an increase in the likelihood function increases until a local maximum is achieved. It has been established that under mild conditions EM is guaranteed to convergence to a local maximum of the $\Gamma(\theta)$. However, there is no guarantee that the convergence will be to a global maximum [132]. For likelihood functions with multiple maxima, convergence will be to a local maximum, which depends on the initial starting point. Likelihood surfaces are often multi-modal and thus have many local maxima. Poor initialization of the algorithm leads to converge to a local maximum, which may be

far from the global maximum. Convergence rate is usually slower than the quadratic convergence typically available with a Newton's type- method. Further information about the convergence of EM appears in [133] [134] [135] .

Demoment [136] showed the relation between EM and regularization. The important practical problem of the choice of regularization parameter, and more generally of the hyperparameters, remains open problem [137] [138] [139] .

3.4. Concluding Remarks

In this chapter, we reviewed common classical image restoration and identification methods. And, we give some metrics to measure the degradation of corruption and improvement in image restoration.

4. NEWTON TYPE OPTIMIZATION of LOG-LIKELIHOOD FUNCTION

Gradient descent methods are inefficient particularly when the number of design variables is large. In addition, the application of gradient descent methods leads to slow convergence rate [140] . A large deviation in the eigenvalues of the Hessian is well known to cause slowness of convergence towards the minimum in gradient-descent methods [71] . The EM algorithm appears as a possible alternative to gradient-based methods for optimization. In many examples, either the E-step or the M-step proves to be intractable. A method to overcome this difficulty is the Newton method where the Hessian has to be computed [54] . Therefore, second-order information is vital for fast convergence. Pure form of the Newton's method has several drawbacks; therefore the modification is required to convert it to a reliable algorithm [141] . It is well known that computing second derivatives for a large class of optimization is not feasible but relatively inexpensive [142] . Generally in many practical situations, gradient is available, but the Hessian matrix cannot be computed or stored. When faced with such a problem, most statisticians turn to other algorithms such as “scoring”. Superlinear (quasi-Newton, method of scoring) and second order methods have been preferred to EM [143] .

4.1. The Method of Scoring for Blur and Image Identification

According to FS method, it is assumed that, $\left. \frac{\partial^2 L(\theta)}{\partial \theta^2} \right|_{\theta=\theta^i} \approx -I_f(\theta^i)$. To minimize $L(\theta)$ with respect to θ , the iterative method of scoring algorithm was defined as,

$$\hat{\theta}^{k+1} = \hat{\theta}^k + \eta I_f^{-1}(\theta^k) \frac{\partial L(\theta)}{\partial \theta^k} \quad (4.1)$$

Here, η is the convergence parameter $I_f(\theta^k)$ is the FIM at the k^{th} estimate of θ .

The parameter vectors $\underline{d}, \underline{a}$ and noise variances σ_v^2, σ_w^2 are independent from each other, so the FIM will have a block diagonal form as follows:

$$I_f(\theta) = \begin{bmatrix} [I_f(d)]_{(L \times L)} & \emptyset & \emptyset & \emptyset \\ \emptyset & [I_f(a)]_{(K \times K)} & \emptyset & \emptyset \\ \emptyset & \emptyset & I_f(\sigma_v^2) & \emptyset \\ \emptyset & \emptyset & \emptyset & I_f(\sigma_w^2) \end{bmatrix} \quad (4.2)$$

The matrices $I_f(\underline{d}), I_f(\underline{a}), I_f(\sigma_v^2)$ and $I_f(\sigma_w^2)$ are the corresponding FIM's and are defined as:

$$I_f(\underline{\theta}) = \begin{bmatrix} [I_f(\theta)]_{11} & \cdot & \dots & [I_f(\theta)]_{1j} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ [I_f(\theta)]_{i1} & \cdot & \dots & [I_f(\theta)]_{ij} \end{bmatrix} \quad (4.3)$$

The FIM, $[I_f(\theta)]_{i,j}$ is defined by

$$[I_f(\theta)]_{i,j} = E \left[\frac{\partial L(\theta)}{\partial \theta_i} \frac{\partial L(\theta)}{\partial \theta_j} \right] \quad (4.4)$$

where $E[\bullet]$ is the expected value operation. For image and blur identification problems, we obtain the FIM as the following form;

$$[I_f(\theta)]_{i,j} = 2tr \left(P^{-1} \frac{\partial P}{\partial \theta_i} P^{-1} \frac{\partial P}{\partial \theta_j} \right) \quad (4.5)$$

In DFT domain, we can write the FIM as;

$$[I_f(\theta)]_{i,j} = 2 \sum_{u,v} \left\{ \frac{1}{P^2(u,v)} \frac{\partial P(u,v)}{\partial \theta_i} \frac{\partial P(u,v)}{\partial \theta_j} \right\} \quad (4.6)$$

The detailed formulations of (4.5) and (4.6) are given in Appendix A.

4.1.1. Local Convergence Analysis of FS

If the initial guess is close to the global minimum, second order approaches reach the exact solution in the ML estimation. If it is not, convergence may not be attained, or is directed toward to a local minimum only. In this section, we will examine the convergence of FS method.

Newton-like methods for nonlinear equations in convergence theory are described in [144]. Galántai [145] reviewed the theoretical results on Newton's method concerning the convergence properties, the error estimates, the numerical stability and the computational complexity of the algorithm.

It is required to analyze the convergence of the FS algorithm by using the classical line search optimization approaches. Note that FS is an ascending algorithm. For the gradient ascent algorithm, the step size η must be chosen to ensure that,

$$\frac{\|\theta^{i+1} - \theta^*\|}{\|\theta^i - \theta^*\|} \leq \|I + \eta H(\theta^*)\| < 1 \quad (4.7)$$

where $H(\theta)$ is the Hessian matrix, $\|\bullet\|$ is the Euclidian norm and θ^* is the optimal solution of θ [146]. If (4.7) is not satisfied, the algorithm cannot be stable. For the FS algorithm, using a similar approximation, we can state the following condition.

$$\|I + \eta I_f(\theta)^{-1} H(\theta^*)\| < 1 \quad (4.8)$$

The choice of a suitable convergence parameter η remains as an essential problem.

4.2. A New Trust Region Fisher Scoring Optimization

4.2.1. Introduction

In some cases, the line search method does not yield a significant decrease in the objective function. This difficulty is due to ill-posedness of the Hessian matrix. To overcome these difficulties a TR method can be used. TR sub-problems arise in optimization in the context of globalization strategies for locally convergent methods [147] . In literature, TR methods have been applied for well-posed problems, successfully. In particular the use of TR modifications for FS has been very rare, one example being Osborne [148] . Since TR algorithms are reliable, robust and have strong convergence properties, they can be applied to ill-posed problems [69] [149] [150] [151] . TR algorithms are popular for their strong convergence properties and have some applications in regularization of ill-posed problems. Recently, Wang and Yuan [152] applied TR methods for solving nonlinear ill-posed inverse problems and they showed that the TR method had regularization property.

Although Newton's method usually requires a small number of iterations, it has a disadvantage of using a step size, which may not lead to convergence. Selecting a convenient step size which enable fast convergence, and possessing the stability property is very desirable in all optimization problems. In this part of the thesis, we try to find new criteria based on the convergence properties of the FS optimization. In our new proposed method, we attempt to combine the best properties of line search and TR algorithms. We enforce line search conditions on every iteration to guarantee convergence. The new hybrid solution is fast and capable of handling ill-posed problems.

There were many studies about convergence of TR methods. Lin and Moré [153] analyzed a TR Newton's method for the linearly constrained optimization problem based on the convergence theory of Moré [154] . A survey on convergence problems of unconstrained optimization algorithms was presented including line search algorithms and TR algorithms [155] . Some well-known convergence problems of VM methods were also presented. El-Alem [156] proved global convergence to a stationary point for a general TR based algorithm for equality constrained

optimization. This general algorithm is based on appropriate choices of TR sub-problems and seems particularly suitable for large problems. Dennis and Vicente [151] showed global convergence to a point satisfying the second-order necessary optimality conditions for the same general TR based algorithm. The results given here can be seen as a generalization of the convergence results for TR unconstrained optimization. The behavior of the trust radius and the local rate of convergence are analyzed. Yamashita [76] investigated the rate of convergence of LM method with unit step size and discussed its global convergence with Armijo step size. Aitkin and Foxall [157] presented an FS algorithm, which is easy to implement, but starting values were still important.

To understand the new proposed method, it is very useful to examine the relationship between line search and TR methods. The classical TR method for unconstrained minimization can be augmented with a line search that finds a point that satisfies the Wolfe conditions [158]. Gertz [159] developed new algorithms that apply a line search to the TR trial step. He developed the convergence theory of a method that performs a line search on the direction proposed by the solution of a TR sub-problem. He used this new method to define an algorithm that simultaneously satisfies the quasi-Newton condition at each iteration. This method, combined with a BFGS update and a Wolfe condition-based line search, produces an effective quasi-Newton method. Vicente [160] forced the TR technique to act like a line search he accomplished this by choosing always the step along the Quasi-Newton direction. He obtained global convergence to a stationary point as long as the condition number of the second-order approximation was uniformly bounded.

Some TR studies were applied to image restoration problem, recently. Wang et al. [161] solved large scale ill-posed problems, particularly the image restoration problem in atmospheric imaging sciences, by a TR-CG algorithm. TR method has been used in parameter identification problem and image restoration problem [162]. Rojas and Steihaug [163] presented a new method for large-scale nonnegative regularization and some numerical results have been given on image restoration problems.

Although, the blur and image learning problem is solved very efficiently and rapidly as a ML problem by using a FS method, converging to the local minima and the

determination of the convenient convergence parameter are still drawbacks [164] . Especially, when the initialization values are not close to the solution, converging to a local minimum is not robust. FS methods must be modified to enforce convergence to a local minima even the initialization values are not good. Selecting a convenient step size, which enables fast convergence and provides stability, is very desirable in all optimization problems.

In this section, we present a new FS method that ensures the convergence in every iteration step to learn the blur and image parameters. We also try to find new criteria based on the convergence properties of the FS optimization by the Levenberg-Marquardt strategy. We analyze the TR version of FS and we describe a new optimization criterion. We call this method Trust-Region Fisher Scoring (TR-FS). Sarı and Çelebi [165] have recently introduced a new algorithm, which has more robust convergence properties for ill-posed image identification problems than line search methods. In our proposed method, we attempt to combine the best properties of line search and TR algorithms. We enforce line search conditions on every iteration to guarantee convergence. The new hybrid solution is fast and capable of handling ill-posed problems. Thanks to the method, convergence to a stationary point is guaranteed if the step is in the “trust region”. This new algorithm is robust and efficient in practice. Regularized solution, which is necessary because of the ill-posed nature of the image restoration problem, is a side benefit of the proposed method. Restored images are obtained by Regularized Wiener filter after learning of the image, blur and noise parameters. Performance evaluation of the method was performed based on simulations and experimental results were demonstrated on artificial and real images. Since all of the required parameters will be learned, regularized Wiener filter can be applied to restore images.

We use the FS method with a simple Levenberg-Marquardt (LM) restricted step strategy. LM method can be considered as a special case of the TR method [74] . Replacing the information matrix with the $I_f(\theta^k) + \lambda I$ in the FS scheme, gives the RS version of the FS algorithm. The new RS-FS algorithm is given as

$$\hat{\theta}^{k+1} = \hat{\theta}^k + \eta \left(I_f(\theta^k) + \lambda I \right)^{-1} \frac{\partial L(\theta)}{\partial \theta^k} \quad (4.9)$$

where λ is a nonnegative parameter and I is the identity matrix. The diagonal matrix is modified during the computation. The λ parameter is increased as required during the iteration to avoid ill-conditionedness and to still provide convergence [166]. A compromise between gradient ascent and RS-FS direction can be offered in the following items:

- If λ is close to zero, then $I_f(\theta^k) + \lambda I$ is close to $I_f(\theta^k)$. So, the method goes towards the classical FS method.
- If λ is very big, then $I_f(\theta^k) + \lambda I$ is close to λI . So, the new method goes towards classical gradient ascent with $\frac{\eta}{\lambda}$ being the convergence parameter.

4.2.2. Finding New Criteria

FS methods can also handle the indefinite Hessian matrix problem, but for non-quadratic nonlinear optimization, FS requires a step size η [167]. In accordance with the FS approach, we can use FIM approximation instead of Hessian matrix. Therefore, we can write the following relation.

$$\left\| I + \eta \left(I_f(\theta) + \lambda I \right)^{-1} I_f(\theta) \right\| < 1 \quad (4.10)$$

Let us assume that $\Gamma(\theta) = \left(I_f(\theta) + \lambda I \right)^{-1} I_f(\theta)$ and therefore we can define the following expression using the inner product defined in [168].

$$\left\| I + \eta \Gamma(\theta) \right\| = \left\{ \text{tr} \left[\left(I + \eta \Gamma(\theta) \right)^T \left(I + \eta \Gamma(\theta) \right) \right] \right\}^{\frac{1}{2}} \quad (4.11)$$

The new condition can be expressed as follows.

$$tr(\Gamma^T(\theta)\Gamma(\theta))\eta^2 + 2tr(\Gamma(\theta))\eta + (n-1) < 0 \quad (4.12)$$

Here, n denotes the number of diagonal elements. To obtain real roots, the following condition should be satisfied.

$$[tr(\Gamma(\theta))]^2 \geq (n-1)tr(\Gamma^T(\theta)\Gamma(\theta)) \quad (4.13)$$

or

$$[tr(\Gamma(\theta))]^2 \geq (n-1) \left(\sum_i \sum_j ((\Gamma(\theta))_{ij})^2 \right) \quad (4.14)$$

Here $(\Gamma(\theta))_{ij}$ denotes an element-by-element processing of matrix $(\Gamma(\theta))$ and i, j denote the row and column of the matrix $(\Gamma(\theta))$, respectively.

The values between the real roots of the Equation (4.12) are the desired convergence values for our algorithm. We can find the real roots by using the following equation:

$$\eta_{1,2} = \frac{-tr(\Gamma(\theta)) \mp [tr^2(\Gamma(\theta)) - (n-1)tr(\Gamma^T(\theta)\Gamma(\theta))]^{1/2}}{tr(\Gamma^T(\theta)\Gamma(\theta))} \quad (4.15)$$

Choosing the greater root means faster convergence. We chose the convergence parameter as $\eta = \frac{1}{2}(\eta_1 + \eta_2)$.

In the new FS algorithm, we propose that λ is to be increased as required during the iteration. In addition, a λ value must be found to satisfy the condition in (4.13) or (4.14); that means there is an upper limit for λ . If we satisfy the above condition, it is guaranteed that each iteration will improve the likelihood of our estimates.

4.3. Performance Evaluation Methodology

Performance evaluation of the method was carried out by Cramer-Rao Bound (CRB) and the condition number of the Hessian matrix. The MSE for any estimate of a nonrandom parameter has a lower bound, which is known as the CRB [169]. A measure for the level of computational difficulty is the condition number of the Hessian. For deterministic inverse problems, the condition number can be preferred as an alternative measure [170] [171]. The condition number describes the ill-condition or bad behavior of a matrix quantitatively [50]. We have also calculated the distance between the original and identified parameter vectors by the use of L_1 and L_2 vector norms [172]. For image restoration, we use the SNR improvement factor. Finally, we discuss the computational complexity of the method, in this section.

4.3.1. Cramer-Rao Lower Bound

In any estimator, the main indicator is the error occurred in the estimates relative to the true parameters. However, these true parameters are unknown in our blind deconvolution problem. CRB can be used to predict the performance of a given modeled system. By minimizing the CRB, the optimum system design can be found [173]. The actual performance cannot be better than the CRB. The estimator $\hat{\theta}$ is some unbiased estimator of the parameter θ . Then the error covariance matrix for $\hat{\theta}$ is bounded as follows,

$$C_r = E\left\{\left(\hat{\theta} - \theta\right)\left(\hat{\theta} - \theta\right)^T\right\} \geq I_f^{-1}(\theta) \quad (4.16)$$

provided that I_f is positive definite. The bound $\left(I_f^{-1}(\theta)\right)_{kk}$ represents a variance bound on the estimator $\hat{\theta}_k$ as,

$$\text{var}\left(\hat{\theta}_k\right) \geq \left(I_f^{-1}(\theta)\right)_{kk} \quad (4.17)$$

where $I_f(\theta)$ is the FIM of the parameter vector θ and $(I_f^{-1}(\theta))_{kk}$ is the k^{th} diagonal element of $I_f^{-1}(\theta)$.

4.3.2. Regularization Effects of the New RS-FS Method

TR methods are robust and can be applied to ill-posed problems. Apart from the widely known robustness property of TR methods, they also have regularization effect. TR techniques to determine the regularization parameter are introduced in [69]. Vicente [160] claimed that the TR techniques have a built-in appropriate regularization of the ill-posed second-order approximation. The convergence of a so-called regularizing LM method for ill-posed problems, has been proven by Hanke [174]. Chan [175] examined the regularization effects of the LM learning and showed that LM learning allows other forms of regularization operators by some simple modifications. The applicability of the conventional LM method for ill-posed problems is still an open research issue [176].

Our proposed TR or RS algorithm can be seen as a regularization method, but it differs from the traditional regularization method in that no penalty term is needed. Thus, the determining of the so-called regularization parameter in a standard regularization method is avoided [177]. In our method, the λI value added to the diagonal of I_f can be interpreted as a regularization parameter.

To show the resulting regularization effect of our method, we calculate the condition number of the Hessian matrix (in our problem of $-I_f(\theta)$) at each iteration. We present some results suggesting that our method regularizes the condition number of the Hessian in experimental results section.

4.3.3. Computation Complexity

Newton's method requires gradient and Hessian computation, as well as $O(n^3)$ multiplications at each step [178]. Similarly, the computation of the FS search direction $I_f^{-1}(\theta^k) \frac{\partial L(\theta)}{\partial \theta^k}$ requires $O(n^3)$ flops. Computing I_f requires three

multiplications of square matrices with size n . This operation has also $O(n^3)$ complexity [179]. Our proposed method requires only slightly more computation per iteration than the FS method.

4.4. Relations Between EM and Newton Type Method

Salakhutdinov *et al.* [180] showed a close relationship between the EM algorithm and direct optimization algorithms such as gradient-based methods for parameter learning. They presented an analysis of the bound and direct optimization algorithms, and developed the connection between those two classes of optimizer. Titterton [181] also showed that the EM algorithm is related to the method of Scoring.

EM is guaranteed to increase the likelihood function,

$$L(\theta^{i+1}) \geq L(\theta^i) \quad (4.18)$$

This monotonic convergence of the likelihood is achieved without step-size parameters or line searches. Other gradient-based optimization techniques, including the gradient descent method, quasi-Newton and Newton's method, do not provide such a simple theoretical guarantee [182]. For Newton's method to reach a solution, the iterative process is usually required to start at some point close to the solution, otherwise the Hessian may be indefinite and the iteration may not converge [183].

EM and Newton's method are the same for simple quadratic problems. For general non-quadratic optimization problems, Newton's method yields fast convergence but deviation from the solution is possible. Unlike other optimization techniques, EM algorithm does not require the computation of gradients or Hessian. LM, FS and our new proposed method handle the indefinite Hessian matrix problem, at the expense of increased computational cost. In the EM algorithm, in some situations, either the E-step or the M-step proves to be intractable. The other main drawback of the EM algorithm is slow convergence. The convergence of gradient descent and EM algorithms both depend on the shape of the log likelihood function as measured by the condition number. When condition number is close to one, the configuration is quite regular, and the update direction points directly to the solution yielding fast convergence. When condition number is very large, the search along the update direction is a zigzag path, making convergence very slow.

4.5. Experimental Results

The performance of the proposed algorithm is tested using two different types of images, a 64×64 artificial image and the 64×64 real “Lena” image. The results are demonstrated for the gradient descent method, pure FS algorithm and the proposed RS-FS algorithm. The imaging system does not absorb or generate energy, so: $\sum_{m,n \in W_D} d(m,n) = 1$. It can easily be seen that the computation density is very high.

For instance, we assume that the image is $N \times N$, where N is the number of pixels in each direction, and thus contains N^2 pixels, and accordingly D is $N^2 \times N^2$. Typically N is chosen to be a power of 2, such as 32, 64, 128, 256 or larger, then the number of unknowns grows to at least 1024, 4096, 16384, 65536, respectively. In order to reduce the computational density, we assume that the unknown PSF is real and symmetric, therefore a number of PSF coefficients have equal values. To cope with the computational mass, one of the other techniques is recursive processing in time domain. The other one is the frequency domain solution. The AR model for the original image is assumed not to absorb or generate energy: $\sum_{k,l \in W_A} a(k,l) = 1.0$.

4.5.1. Experiments on Artificial Images

The artificial image is produced using an AR model with Non Symmetric Half Plane (NSHP) support [184]. Independent white Gaussian noise is added to this artificial image. We use the following AR parameters to produce the synthetic image. The NSHP support region is given in Figure 4.1.

$a_{11} = -0.449$	$a_{01} = 0.630$	$a_{-11} = 0.127$
$a_{1,0} = 0.671$		

Figure 4-1: NSHP image support.

We produced a 100×100 2-D random field, and we only used lower right corner of this artificial data, which is 64×64 in size. Because of the symmetry properties of

the PSF, we have only three parameters d_5, d_6, d_9 to estimate for PSF. The chosen region is given in Figure 4.2.

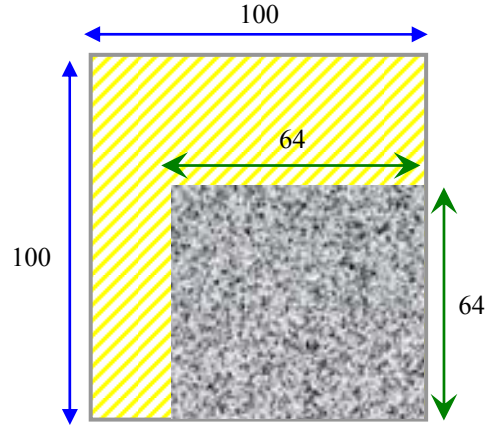


Figure 4-2: 2-D artificial random field.

The artificial image is blurred by using the following (3 x 3) blur PSF.

$$\begin{bmatrix} d(-1,-1) & d(-1,0) & d(-1,1) \\ d(0,-1) & d(0,0) & d(0,1) \\ d(1,-1) & d(1,0) & d(1,1) \end{bmatrix} = \begin{bmatrix} d_1 & d_2 & d_3 \\ d_4 & d_5 & d_6 \\ d_7 & d_8 & d_9 \end{bmatrix} = \begin{bmatrix} 0.075 & 0.12 & 0.075 \\ 0.12 & 0.22 & 0.12 \\ 0.075 & 0.12 & 0.075 \end{bmatrix} \quad (4.19)$$

In this experiment, the AR image model coefficients are assumed as known. Table 4.1 summarizes the blur identification results on the noisy and blurred images with the number of iterations listed for each algorithm.

Table 4.1: Identification of the PSF coefficients, SNR=20 dB.

	D₅=0.22	d₆=0.12	d₉=0.075	Iteration #	Execution Time for each step
Initialization	0.2	0.15	0.05		
Gradient Descent	0.2213	0.1117	0.0713	97	186.43
FS	0.2369	0.1108	0.0698	4	214.41
Restricted Step FS	0.2314	0.1104	0.0707	5	249.15

Since Gradient Descent and FS optimization methods converge just for the suitable initial values, we need to choose good initial values. We experimentally chose the convergence parameter of Gradient descent and FS method as 2×10^{-7} and 5.4×10^{-8} , respectively. Restoration is performed using the regularized Wiener filter. The original synthetic image, its blurred and noisy version and the restored images are given in Figure 4.3.

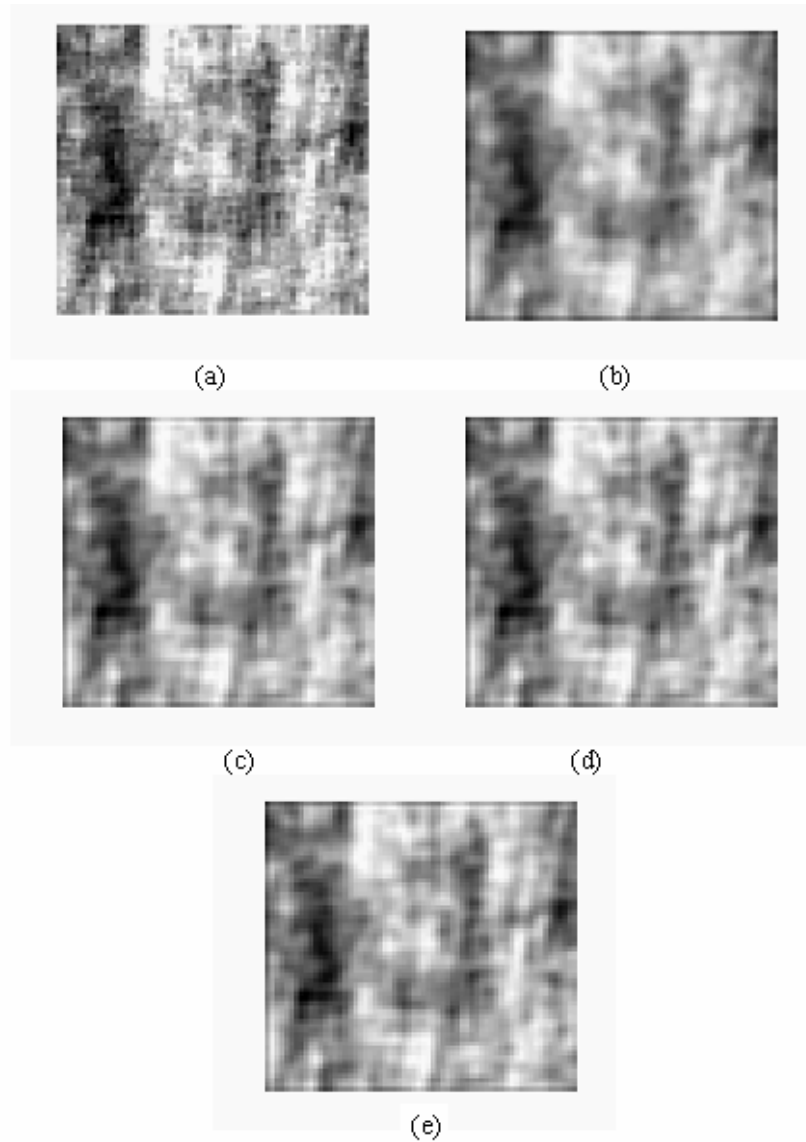


Figure 4-3: (a) The original synthetic image; (b) Blurred and noisy version with SNR=20 dB.; (c) Restored image with learned parameters by gradient based and SNRimp = 2.18 dB. ; (d) Restored image with learned parameters by FS and SNRimp = 2.07 dB.; (e) Restored image with learned parameters by RS-FS and SNRimp = 2.11 dB.

4.5.2. Experiments on Real Images

We also apply the RS-FS algorithm to the real images. The 64×64 Lena image was synthetically blurred by a 3×3 blur PSF defined in (4.19) and noise has been added. We need to initialize the image model and blur parameters in order to identify the noisy and blurred image. As initialization parameters, we use the image model coefficients obtained via the LS method from the noisy and blurred image. The NSHP image model coefficients were computed from the original 64×64 Lena image by the Least Squares (LS) from the observed image. For blur parameter identification, each value of the PSF function is chosen to be equivalent to the initial values. Table 4.2 summarizes the blur identification results on the observed image when the conventional FS and Newton methods do not converge for the given initial values. Table 4.3 demonstrates the AR image parameter identification results on the noisy and blurred images.

Table 4.2 :Identification of blur (PSF) coefficients.

		d₅	d₆	d₉
Original PSF		0.22	0.12	0.075
Initial		0.111	0.111	0.111
Estimated parameters	40 dB	0,2256	0,1197	0,0739
	30 dB	0,2424	0,1118	0,0776
	20 dB	0,2084	0,1123	0,0856
	10 dB	0,1764	0,0998	0,1061

Table 4.3 : Identification of AR image coefficients.

		$a_{1,1}$	$a_{1,0}$	$a_{-1,1}$	$a_{0,1}$
Estimated from original image by LS		-0.1434	0.1395	0.5215	0.4782
Initialization parameters from observed image by LS		-0.5224	0.2961	0.3475	0.8767
Estimated Parameters	40 dB	-0.1823	0.1581	0.4272	0.5876
	30 dB	-0.2137	0.1665	0.3992	0.6104
	20 dB	-0.2912	0.1738	0.3631	0.6940
	10 dB	-0.2912	0.1946	0.3299	0.7191

A typical running time for performing 3 blur and 4 AR image parameters on a 64x64 sized image is given in Table 4.4 using our Pentium IV PC at 1.8.

Table 4.4 : Iteration Number and Running Time for 3 blur and 4 AR image parameters for 64x64 Lena image.

SNR	Iteration Number	Execution Time on each iteration
40 dB	7	287.33
30 dB	11	293.67
20 dB	21	308.86
10 dB	23	312.44

To gain more insight into the performance of the proposed algorithm, we evaluate the CRB on the error variance in estimating the blur and image model parameters. The CRB on the parameters of the proposed method are listed in Table 4.5.

Table 4.5 : CRB values of estimated blur and image parameters

	Blur Parameters			Image Model Parameters			
SNR	d_5	d_6	d_9	$a_{1,1}$	$a_{1,0}$	$a_{-1,1}$	$a_{0,1}$
40	0,00062465	0,0011001	0,00062943	0,000710	0,000973	0,000637	0,000986
30	0,00088535	0,0014983	0,00099194	0,000978	0,001772	0,001142	0,001802
20	0,0011573	0,0017624	0,0013471	0,001959	0,002021	0,001891	0,002135
10	0,0030018	0,0037385	0,0036149	0,004123	0,004216	0,003978	0,004761

Pretty good identification and restoration quality is obtained for SNR starting from 40 to 10 dB. Figure 4.4 shows the original, blurred and noisy and restored images using the identified values at SNR of 10, 20 30 and 40 dB.



(a)



(b)



(c)



(d)



(e)

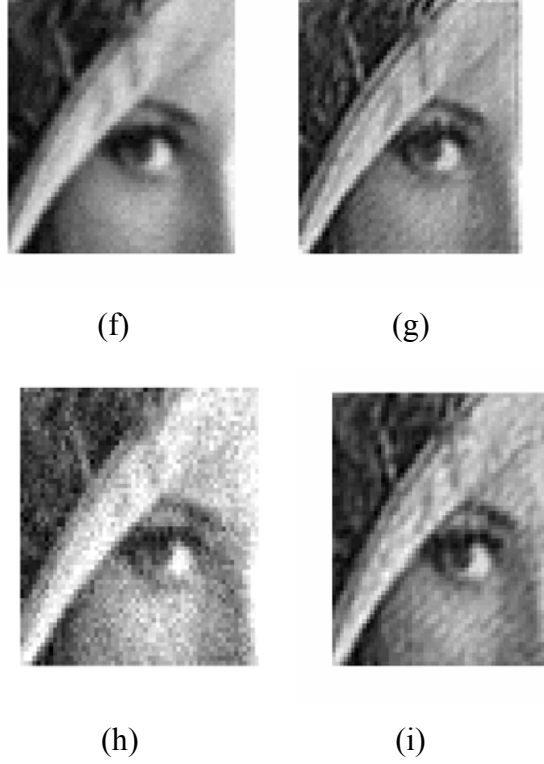


Figure 4-4: (a) The original 64×64 Lena image; (b) Blurred and noisy version with SNR=40 dB.; (c) Restored image with learned parameters and SNRimp = 2.32 dB. ; (d) Blurred and noisy version with SNR=30 dB.; (e) Restored image with learned parameters and SNRimp = 2.11 dB. ; (f) Blurred and noisy version with SNR=20 dB.; (g) Restored image with learned parameters and SNRimp = 2.04 dB. ; (h) Blurred and noisy version with SNR=10 dB.; (i) Restored image with learned parameters and SNRimp = 1.91 dB.

At each simulation step, we calculate the condition number of the FIM. Figure 4.5 shows the variations of the condition number versus the iteration number. It can be seen that our proposed method regularizes the ill-posed structure. As the number of parameters increases, the condition number of the FIM increases dramatically. In that case, our proposed method will be come more important.

The learning performance of the blur parameters is tested by calculating the mean square errors after each step.

MSEs between $\underline{d}$ and $\hat{\underline{d}}$, and also between $\underline{a}$ and $\hat{\underline{a}}$, are computed as a function of iteration for different SNR values are illustrated in Figure 4.6 and Figure 4.7, respectively.

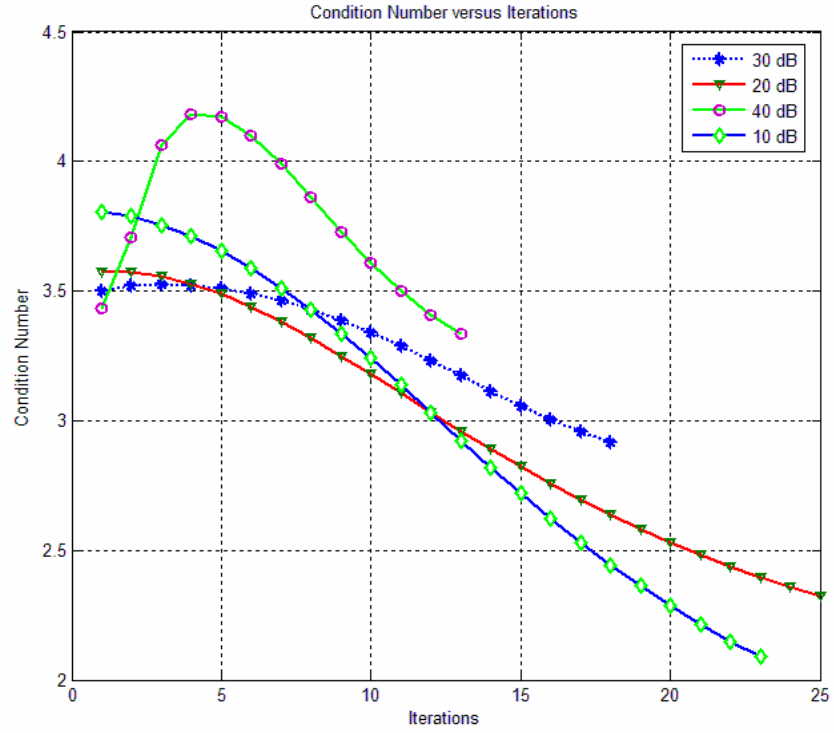


Figure 4-5 : The variations of the condition number versus the iteration number.

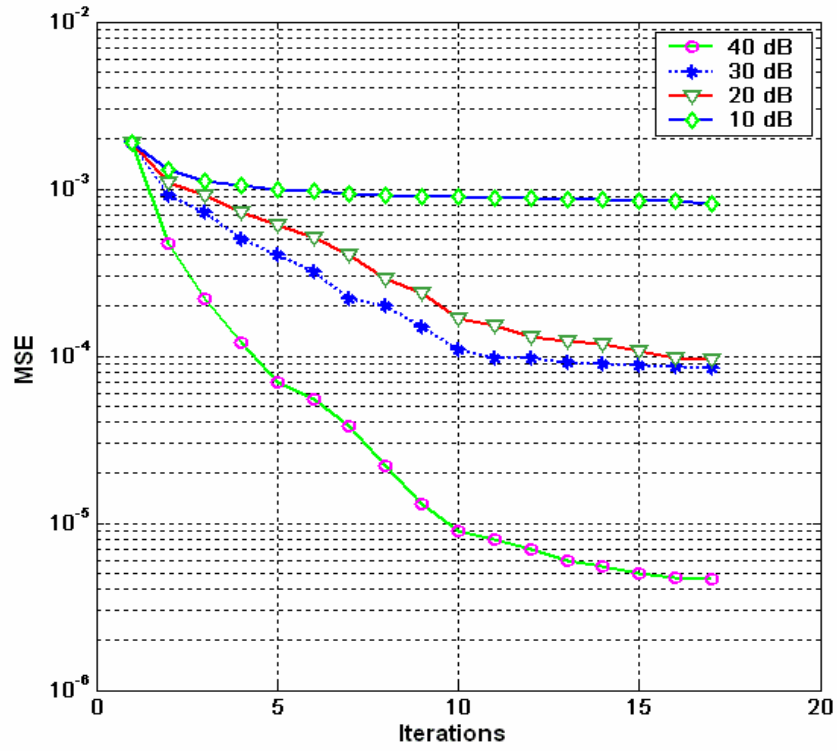


Figure 4-6 : MSE curves between $\underline{d}$ and $\hat{\underline{d}}$ for different SNR.

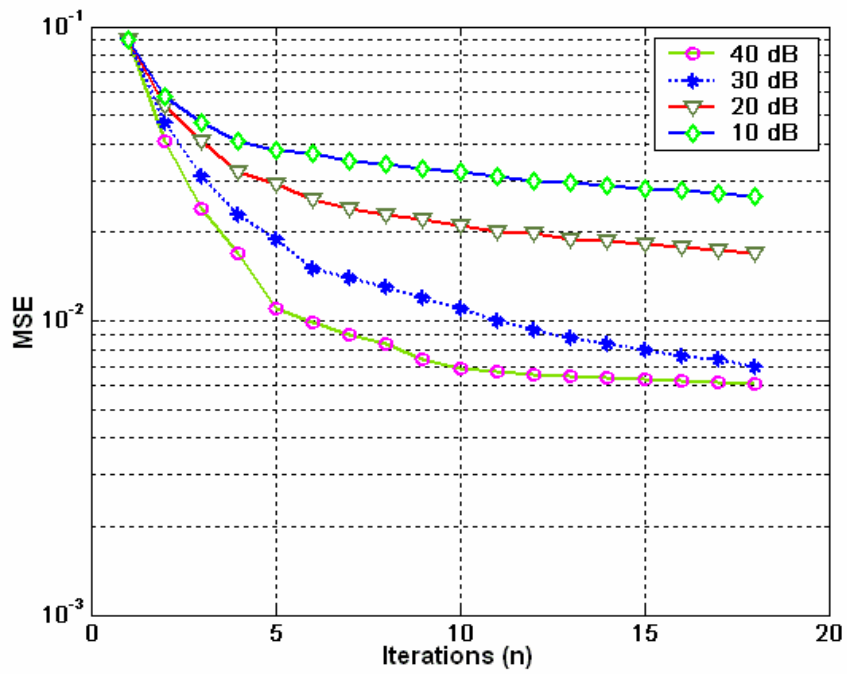


Figure 4-7: MSE curves between $\underline{a}$ and $\hat{\underline{a}}$, for different SNR.

As an empirical rule of thumb, we stop the iterations when the ratio of the estimated parameter during the previous iteration and the current iteration is less than threshold ($\epsilon=0.0001$).

4.6. Concluding Remarks

A new approach was presented to learn the blur and image parameters and to restore the images by taking the fast and robust optimization desires into consideration. The blur and image identification problem was solved by applying a FS method, which is a Newton type optimization. Despite the proposed method reaches the solution in a few steps, it is sensitive to initial values. The choice of the convergence parameter directly affects the performance of the method. In order to obtain more powerful and robust technique, convergence parameter should be chosen conveniently. We modified the method to enforce convergence even the initialization values are not good. We attempt to establish new criteria to choose convergence parameter based on the convergence properties of the FS method under the Levenberg-Marquardt strategy.

This new hybrid solution is fast, capable of handling ill-posed problems and has strong convergence properties. It is also robust and efficient in practice than conventional FS and Newton methods. The proposed algorithm gives a good compromise between the speed of the FS algorithm and the stability of the Steepest Descent method, and also it provides a skillful transition between these methods. This method incurs extra computational costs at each iteration. To cope with the computational mass, one of the techniques is recursive processing in time domain, which is considered in [185].

Apart from the speed and convergence properties, it has a regularization effect on the ill-posed problems. Regularization effect of the method, which is needed because of the ill-posed structure of the image restoration problem, was shown in terms of the condition number of the Hessian matrix. Condition number of the Hessian matrix decreases noticeably especially at the low SNR values. As the number of parameters increases, the condition number of the FIM increases dramatically. In that case, our proposed method will be come more important. Another advantage of the method, it

is possible to stay away from Hessian equations, which may be impossible to obtain in many examples, by using only the gradient values. After learning of the image, blur and noise parameters, the Regularized Wiener filter was used to restore images.

5. RECURSIVE LEARNING of IMAGE PARAMETERS and RESTORATION of IMAGES USING EM BASED LEARNING ALGORITHM

5.1. Introduction

EM based blur identification and restoration methods have been developed for batch processing that needs large memory size or frequency domain solution. For the image processing applications, recursive estimation techniques are important because of the huge size of the input data, and therefore various types of the Kalman filter have been adopted in [185] [186] [187] [188] [189] [190] [191] [192] [193] [194] [195] . Recursive processing allows dynamic processing with modest storage requirements. Many linear and nonlinear recursive restoration algorithms are presented [196] .

Direct ML and EM based approach in linear dynamical systems has been considered in the past. Mendel [73] [96] developed a formula for the log-likelihood function for the basic state-variable model. Gupta and Mehra [197] discussed the numerical aspects of computing ML estimates for linear dynamical systems in state-vector form. Different gradient-based nonlinear programming methods were discussed in a unified framework. Bergboer et al.[198] presented a numerically efficient implementation of ML identification of Linear Time Invariant (LTI) state space models by using local gradient search. A method for evaluating the gradient, Hessian, and FIM of linear dynamic stochastic systems is presented by Segal et al. [199] . The method incorporated the optimal Kalman equations and, was therefore ideal for simultaneous state and parameter estimation. Charalambous [200] derived explicit filters for computing ML parameter estimates via the Newton Raphson (NR) method and the scoring method. Dethlefsen [201] applied the iterated Kalman smoother using the Gaussian Markov random field model with parameters obtained

from ML estimation for image restoration with focus on finding the edges in the image.

EM algorithm in linear dynamical systems has been considered in the past. Chen [202] applied the EM algorithm to linear state space models. Watson and Engle [203] proposed to use the EM algorithm, in conjunction with the method of scoring, for the estimation of linear dynamic factor. Recursive EM algorithm along with stochastic approximation was employed to estimate unknown time-invariant/variant parameters [204]. Shumway and Stoffer [205] showed how the EM algorithm could be used for parameter estimation and discussed its advantages and disadvantages over previous nonlinear optimization methods. They derived an EM algorithm in which the parameters of the model (state transition matrix, state noise vector and observation noise vector) can be learned. Only the observation matrix was assumed known. Ghahramani and Hinton [206] have recently extended the algorithm to allow the observation matrix to be learned as well. However, these learning algorithms are batch learning algorithms rather than recursive update procedures. Therefore they are not suitable for “on-line” learning. A modified version the Shumway-Stoffer algorithm was presented that permits several frequently needed linear restrictions on the model parameters [207]. Although the EM learning was applied to learning of unknown image and blur parameters based on batch image processing before, recursive EM learning of unknown image and blur parameters has not been studied as much as necessary.

In this chapter, we introduce a new simultaneous recursive parameter learning and image restoration method based on the ML parameter identification and state estimation for images. We propose that state space image model can be represented as a DBN. Although our proposed method deals with huge data sizes, because of recursive structure it does not need large size storage. We use Kalman filter structure in order to optimize the ML estimation of the unknown image and blur parameters. We present the learning algorithm for the output-only case and the problem can be defined as an unsupervised learning problem. We focus on the problem of learning the parameters of a Bayesian network. In addition, a restored image is obtained at the

output of the Kalman filter simultaneously. The performance of the method is given in the experimental results section.

5.2. EM Learning on Linear Dynamic Systems

Only the output of the system is observed, the state and all the noise variables are hidden. The set of unknown parameters of the system generally consists of components of D , A , Q and R matrices.

5.2.1. E Step of the Algorithm

In the E step of the EM algorithm, we have to calculate the

$$E_{\theta} \left\{ \log p_Y(y; \theta) \mid G = g; \hat{\theta} \right\} \quad (5.1)$$

We choose the complete data Y as the following form.

$$Y = \begin{pmatrix} f_0^T & f_1^T & \dots & f_N^T & g_1^T & g_2^T & \dots & g_N^T \end{pmatrix}^T = \begin{pmatrix} \{f\} \\ \{g\} \end{pmatrix} \quad (5.2)$$

The linear Dynamical System (LDS) can be seen as a continuous-state analogue of the HMM, which is a tool for representing probability distributions over sequences of observations [39] [208] [209] .

From basic probability theory we know that we can factor the joint probability as a product of conditional probabilities. This factorization of the joint probability can be drawn graphically. This graph, known as a Bayesian Network, shows the dependencies between the variables. Only the output of the system is observed, the state and all the noise variables are hidden. Using Markov property and invoking the Bayes rule, under the regularity conditions, the joint pdf can be factored as a product of conditional probabilities for N measurements [169] [210] ;

$$p_Y(y; \theta) = p(f_0; \theta) \prod_{k=1}^N p(f_k \mid f_{k-1}; \theta) \prod_{k=1}^N p(g_k \mid f_k; \theta) \quad (5.3)$$

We can write the log-likelihood function as follows,

$$L_Y(\theta) = \log p_Y(y; \theta) = \log p(f_0; \theta) + \log p(f_k | f_{k-1}; \theta) + \log p(g_k | f_k; \theta) \quad (5.4)$$

The initial state f_0 is taken to have mean μ_0 and nonsingular covariance matrix M_0 .

The pdf of the initial state is given by:

$$p(f_0) = \frac{1}{[(2\pi)^p |M_0|]^{1/2}} \exp \left\{ -\frac{1}{2} (f_0 - \mu_0)^T M_0^{-1} (f_0 - \mu_0) \right\} \quad (5.5)$$

The conditional densities for the state and output can be written as follows:

$$p(g_k | f_k; \theta) = \frac{1}{\sqrt{(2\pi)^p \det |R|}} \exp \left\{ -\frac{1}{2} (g_k - Df_k)^T R^{-1} (g_k - Df_k) \right\} \quad (5.6)$$

$$p(f_k | f_{k-1}; \theta) = \frac{1}{\sqrt{(2\pi)^p \det |Q|}} \exp \left\{ -\frac{1}{2} (f_k - Af_{k-1})^T Q^{-1} (f_k - Af_{k-1}) \right\} \quad (5.7)$$

Here, the size of the f_k and g_k vectors are $p \times 1$. The log-likelihood function $L_Y(\theta)$ can be obtained as a sum of quadratic terms as follows;

$$\begin{aligned} L_Y(\theta) = c - \frac{1}{2} \Big\{ \log |M_0| + (f_0 - \mu_0)^T M_0^{-1} (f_0 - \mu_0) \Big\} \\ - \frac{1}{2} \sum_{k=1}^N \Big\{ \log |Q| + (f_k - Af_{k-1})^T Q^{-1} (f_k - Af_{k-1}) \Big\} \\ - \frac{1}{2} \sum_{k=1}^N \Big\{ \log |R| + (g_k - Df_k)^T R^{-1} (g_k - Df_k) \Big\} \end{aligned} \quad (5.8)$$

where c is a constant independent of θ and $|\bullet|$ denotes the determinant of a matrix.

In (5.8) μ_0 and M_0 are the prior mean and state covariance matrix, respectively. The ML image and blur identification problem can be expressed as a minimization problem if equation (5.8) is multiplied by (-1) . $\Gamma(\theta)$ can be defined as,

$$\begin{aligned}
\Gamma(\theta) = & -\left\{ c - \frac{1}{2} \log |M_0| - \frac{1}{2} \text{tr} \left[M_0^{-1} \left(M_{0|N} + \mu_{0|N} \mu_{0|N}^T \right) \right] + \mu_{0|N}^T M_0^{-1} \mu_0 \right. \\
& - \frac{1}{2} \mu_0^T M_0^{-1} \mu_0 - \frac{1}{2} \sum_{k=1}^N \log |\mathcal{Q}| - \frac{1}{2} \text{tr} \left[\sum_{k=1}^N \mathcal{Q}^{-1} \left(M_{k|N} + \mu_{k|N} \mu_{k|N}^T \right) \right] \\
& + \text{tr} \left[\sum_{k=1}^N \mathcal{Q}^{-1} A \left(M_{k-1,k|N} + \mu_{k-1|N} \mu_{k|N}^T \right) \right] \\
& - \frac{1}{2} \text{tr} \left[\sum_{k=1}^N A^T \mathcal{Q}^{-1} A \left(M_{k-1|N} + \mu_{k-1|N} \mu_{k-1|N}^T \right) \right] \\
& - \frac{1}{2} \sum_{k=1}^N \log |R| - \frac{1}{2} \sum_{k=1}^N \mathbf{g}_k^T R^{-1} \mathbf{g}_k + \sum_{k=1}^N \mathbf{g}_k^T R^{-1} D \mu_{k|N} \\
& \left. - \frac{1}{2} \text{tr} \left[\sum_{k=1}^N D^T R^{-1} D \left(M_{k|N} + \mu_{k|N} \mu_{k|N}^T \right) \right] \right\}
\end{aligned} \tag{5.9}$$

where $\text{tr}\{\bullet\}$ denotes the trace of a matrix. Here $\mu_{k|N}$ and $M_{k|N}$ are the conditional mean and conditional covariance of f_k given $\mathbf{g}_k$, respectively. $M_{k-1,k|N}$ is the covariance across two time steps. They are defined with the following relations,

$$\hat{f}_k = \mu_{k|N} = E_\theta \{ f_k | \mathbf{g} \} \tag{5.10}$$

$$M_{k|N} = E_\theta \left\{ \left(f_k - \mu_{k|N} \right) \left(f_k - \mu_{k|N} \right)^T \middle| \mathbf{g} \right\} \tag{5.11}$$

$$M_{k-1,k|N} = E_\theta \left\{ \left(f_{k-1} - \mu_{k-1|N} \right) \left(f_k - \mu_{k|N} \right)^T \middle| \mathbf{g} \right\} \tag{5.12}$$

It can be seen easily that,

$$E_\theta \left\{ f_k f_k^T \middle| \mathbf{g} \right\} = M_{k|N} + \mu_{k|N} \mu_{k|N}^T \tag{5.13}$$

The computation of $\mu_{k|N}$, $M_{k|N}$ and $M_{k-1,k|N}$ may be carried out using any one of a variety of methods for optimal linear smoothing [211] [212] [213]. Efficient smoothing algorithms are obtained by iteration of common linear Kalman filtering and smoothing.

5.2.1.1. Kalman Smoothing Algorithm and Restoration of Images

The computation of $\mu_{k|N}$, $M_{k|N}$, $M_{k-1,k|N}$ can be performed using Kalman smoothing methods in two parts: a forward recursion, which uses the observations from g_1 to g_k , and a backward recursion, which uses the observations from g_N to g_{k+1} . The forward part of the forward-backward algorithm is computed by the well-known Kalman filter and the backward part of the filter is computed by using Rauch's recursion [214]. The forward and backward recursions together are also known Rauch-Tung-Streibel (RTS) smoother [215]. These equations are computed using a forward Kalman filtering recursion followed by a Kalman smoothing recursion. The RTS recursion structure is given in Appendix B.

Applying the Kalman smoother is not feasible because of its complicated form. In order to reduce the computations involved, it has been suggested to replace the full smoothing operation with fixed-lag smoothing or even just filtering. These modifications will yield a sub-optimal algorithm but it is not necessary to apply the smoothing equations any more. The resulting algorithm still maintains its nice monotonic convergence behavior [216]. Instead of using $\mu_{k|N}$ and $M_{k|N}$, here we use Kalman filter output $\hat{f}_k$ which gives the restored image and $M_{k|k}$. After these modifications, we obtain the following equation for $M_{k-1,k|N}$.

$$M_{k-1,k|N} = M_{k-1|k-1} (I - K_k D)^T A^T \quad (5.14)$$

Here, I is the identity matrix and K_k is the Kalman gain. The detailed formulation of (5.14) is given in Appendix C.

5.2.2. M Step of the Algorithm

5.2.2.1. Closed Form

Each of the unknown parameters is calculated by taking the corresponding partial derivative of the $\Gamma(\theta)$, setting to zero, and solving. Differentiating $\Gamma(\theta)$ with respect to the components of θ_i yields,

$$\frac{\partial \Gamma(\theta)}{\partial \theta_i} = E_{\theta} \left\{ -\frac{\partial}{\partial \theta} \log p_Y(y; \theta) \middle| G = g; \hat{\theta} \right\} \quad (5.15)$$

$$\begin{aligned} \frac{\partial \Gamma(\theta)}{\partial \theta_i} &= E_{\theta} \left\{ -\frac{\partial}{\partial \theta_i} \log p_Y(y; \theta) \middle| g; \hat{\theta} \right\} \\ &= \frac{1}{2} \text{tr} \left[M_0^{-1} \frac{\partial M_0}{\partial \theta_i} + \left(M_{0|N} + \mu_{0|N} \mu_{0|N}^T \right) \frac{\partial M_0^{-1}}{\partial \theta_i} \right] - \mu_{0|N}^T \frac{\partial}{\partial \theta_i} (M_0^{-1} \mu_0) \\ &\quad + \frac{1}{2} \frac{\partial}{\partial \theta_i} (\mu_0^T M_0^{-1} \mu_0) + \frac{1}{2} \text{tr} \left[\sum_{k=1}^N Q^{-1} \frac{\partial Q}{\partial \theta_i} + \left(M_{k|N} + \mu_{k|N} \mu_{k|N}^T \right) \frac{\partial Q^{-1}}{\partial \theta_i} \right] \\ &\quad - \text{tr} \left[\sum_{k=1}^N \left(M_{k-1, k|N} + \mu_{k-1|N} \mu_{k|N}^T \right) \frac{\partial}{\partial \theta_i} (Q^{-1} A) \right] \\ &\quad + \frac{1}{2} \text{tr} \left[\sum_{k=1}^N \left(M_{k-1|N} + \mu_{k-1|N} \mu_{k-1|N}^T \right) \frac{\partial}{\partial \theta_i} (A^T Q^{-1} A) \right] \\ &\quad + \frac{1}{2} \text{tr} \left[\sum_{k=1}^N R^{-1} \frac{\partial R}{\partial \theta_i} \right] + \frac{1}{2} \sum_{k=1}^N g_k^T \frac{\partial R^{-1}}{\partial \theta_i} g_k - \sum_{k=1}^N g_k^T \frac{\partial}{\partial \theta_i} (R^{-1} D) \mu_{k|N} \\ &\quad + \frac{1}{2} \text{tr} \left[\sum_{k=1}^N \left(M_{k|N} + \mu_{k|N} \mu_{k|N}^T \right) \frac{\partial}{\partial \theta_i} (D^T R^{-1} D) \right] \end{aligned} \quad (5.16)$$

The partial derivatives of $\Gamma(\theta)$ with respect to blur parameters, image model parameters, model noise and observation noise are given by the following equations.

For image parameters, the update equation for A is derived as follows.

$$\begin{aligned} \frac{\partial \Gamma(\theta)}{\partial A} = & -\frac{\partial \text{tr} \left[\sum_{k=1}^N (M_{k-1,k|N} + \mu_{k-1|N} \mu_{k|N}^T) (Q^{-1} A) \right]}{\partial A} \\ & + \frac{1}{2} \frac{\partial \text{tr} \left[\sum_{k=1}^N (M_{k-1|N} + \mu_{k-1|N} \mu_{k-1|N}^T) (A^T Q^{-1} A) \right]}{\partial A} \end{aligned} \quad (5.17)$$

$$A^{new} = \left(\sum_{k=1}^N E_{\theta} \left\{ f_{k-1} f_k^T \middle| g \right\} \right) \left[\sum_{k=1}^N E_{\theta} \left\{ f_{k-1} f_{k-1}^T \middle| g \right\} \right]^{-1} \quad (5.18)$$

For blur parameters, the update equation for D is derived as follows.

$$\frac{\partial \Gamma(\theta)}{\partial D} = -\sum_{k=1}^N g_k^T R^{-1} \mu_{k|N} + \frac{1}{2} \frac{\partial \text{tr} \left[\sum_{k=1}^N (M_{k|N} + \mu_{k|N} \mu_{k|N}^T) (D^T R^{-1} D) \right]}{\partial D} \quad (5.19)$$

$$D^{new} = \left(\sum_{k=1}^N g_k^T \mu_{k|N} \right) \left[\sum_{k=1}^N E_{\theta} \left\{ f_k f_k^T \middle| g \right\} \right]^{-1} \quad (5.20)$$

For covariance of output noise, the update equation for R is derived as follows.

$$\begin{aligned} \frac{\partial \Gamma(\theta)}{\partial R^{-1}} = & \frac{1}{2} \text{tr} \left[\sum_{k=1}^N R^{-1} \frac{\partial R}{\partial R^{-1}} \right] + \frac{1}{2} \sum_{k=1}^N g_k^T \frac{\partial R^{-1}}{\partial R^{-1}} g_k - \sum_{k=1}^N g_k^T \frac{\partial R^{-1}}{\partial R^{-1}} D \mu_{k|N} \\ & + \frac{1}{2} \frac{\partial \text{tr} \left[\sum_{k=1}^N (M_{k|N} + \mu_{k|N} \mu_{k|N}^T) (D^T R^{-1} D) \right]}{\partial R^{-1}} \end{aligned} \quad (5.21)$$

$$R^{new} = \frac{1}{N} \sum_{k=1}^N \left[g_k g_k^T - 2 D^{new} \mu_{k|N} g_k^T + D^{new} E_{\theta} \left\{ f_k f_k^T \middle| g \right\} (D^{new})^T \right] \quad (5.22)$$

For covariance of state (model) noise, the update equation for Q is derived as follows.

$$\begin{aligned}
\frac{\partial \Gamma(\theta)}{\partial Q^{-1}} = & \frac{1}{2} \text{tr} \left[\sum_{k=1}^N Q^{-1} \frac{\partial Q}{\partial Q^{-1}} + (M_{k|N} + \mu_{k|N} \mu_{k|N}^T) \frac{\partial Q^{-1}}{\partial Q^{-1}} \right] \\
& - \text{tr} \left[\sum_{k=1}^N (M_{k-1|N} + \mu_{k-1|N} \mu_{k|N}^T) \frac{\partial Q^{-1}}{\partial Q^{-1}} A^{new} \right] \\
& + \frac{1}{2} \frac{\partial \text{tr} \left[\sum_{k=1}^N (M_{k-1|N} + \mu_{k-1|N} \mu_{k-1|N}^T) (A^T Q^{-1} A) \right]}{\partial Q^{-1}}
\end{aligned} \tag{5.23}$$

$$Q^{new} = \frac{1}{N} \sum_{k=1}^N \left[E_{\theta} \left\{ f_k f_k^T \mid \mathcal{g} \right\} - 2 A^{new} E_{\theta} \left\{ f_{k-1} f_k^T \mid \mathcal{g} \right\} + A^{new} E_{\theta} \left\{ f_{k-1} f_{k-1}^T \mid \mathcal{g} \right\} (A^{new})^T \right] \tag{5.24}$$

Initial state mean is given as below.

$$\frac{\partial \Gamma(\theta)}{\partial \mu_0} = -\mu_{0|N}^T \frac{\partial}{\partial \mu_0} (M_0^{-1} \mu_0) + \frac{1}{2} \frac{\partial}{\partial \mu_0} (\mu_0^T M_0^{-1} \mu_0) \tag{5.25}$$

$$\mu_0 = \mu_{0|N} = \hat{f}_0 \tag{5.26}$$

Initial state covariance is given as below.

$$\begin{aligned}
\frac{\partial \Gamma(\theta)}{\partial M_0^{-1}} = & \frac{1}{2} \text{tr} \left[M_0^{-1} \frac{\partial M_0}{\partial \theta_i} + (M_{0|N} + \mu_{0|N} \mu_{0|N}^T) \frac{\partial M_0^{-1}}{\partial \theta_i} \right] \\
& - \mu_{0|N}^T \frac{\partial}{\partial \theta_i} (M_0^{-1} \mu_0) + \frac{1}{2} \frac{\partial}{\partial \theta_i} (\mu_0^T M_0^{-1} \mu_0)
\end{aligned} \tag{5.27}$$

$$M_0 = E_{\theta} \left\{ f_0 f_0^T \mid \mathcal{g} \right\} - \mu_{k|N}^T \mu_{k|N} = M_{0|N} \tag{5.28}$$

All required derivatives are given in Appendix D and more information can be taken from [168] [217] [218] [219] [220] .

In some circumstances, to find a closed form might be inefficient, if not impossible. The most common method to optimize $\Gamma(\theta)$ is the gradient based method that requires the computation of the log-likelihood gradient [49] .

5.2.2.2.Optimization

In some cases, the solution to the M-step does not exist in closed form or it may not be feasible to attempt to find the value of θ that globally maximizes the function $\Gamma(\theta)$. The most common method to optimize $\Gamma(\theta)$ is the gradient based method that requires the computation of the log-likelihood gradient. In Chapter 2, Gradient-based iterative optimization algorithm was formulated.

The gradient of the log-likelihood function should be calculated with respect to each unknown parameter. The partial derivative of $\Gamma(\theta)$ with respect to blur parameters, image model parameters, model noise and observation noise can be calculated easily.

5.3. Experimental Results

Experimental results with blurred and noisy images are demonstrated using two different types of images, a 64×64 artificial image and the 64×64 real “Lena” image. The images were synthetically blurred by a 3×3 causal PSF and Gaussian noise was added. In order to reduce the computational density, we assume that the unknown PSF is real and symmetric, therefore a number of PSF coefficients have equal values as given in the following blur matrix d.

$$\begin{bmatrix} d(0,0) & d(1,0) & d(2,0) \\ d(0,1) & d(1,1) & d(2,1) \\ d(0,2) & d(1,2) & d(2,2) \end{bmatrix} = \begin{bmatrix} d_1 & d_2 & d_3 \\ d_4 & d_5 & d_6 \\ d_7 & d_8 & d_9 \end{bmatrix} = \begin{bmatrix} 0.075 & 0.12 & 0.075 \\ 0.12 & 0.22 & 0.12 \\ 0.075 & 0.12 & 0.075 \end{bmatrix} \quad (5.29)$$

Blurred image is formulated using the blur coefficients as follows.

$$\begin{aligned}
g(i, j) = & d_1 f(i, j) + d_2 f(i-1, j) + d_3 f(i-2, j) + d_4 f(i, j-1) + d_5 f(i-1, j-1) \\
& + d_6 f(i-2, j-1) + d_7 f(i, j-2) + d_8 f(i-1, j-2) + d_9 f(i-2, j-2) \quad (5.30) \\
& + w(i, j)
\end{aligned}$$

It is well known that the state support region must include both image and blur model support regions. In our case, the minimum state vector size has to be 130×1 .

5.3.1. Experiments on Artificial Images

The artificial image is produced using an AR model with non-symmetric half plane (NSHP) support [184]. For the AR model of the image, $\sum_{k,l \in W_A} a(k,l) \leq 1$ is assumed.

Independent white Gaussian noise is added to this artificial image. NSHP AR image model is given,

$$f(i, j) = a_{11} f(i-1, j-1) + a_{01} f(i, j-1) + a_{-11} f(i+1, j-1) + a_{10} f(i-1, j) + v(i, j) \quad (5.31)$$

Here a_{kl} are the NSHP AR model coefficients. We chose the NSHP AR parameters as, $a_{11} = -0.449$; $a_{01} = 0.630$; $a_{-11} = 0.127$; $a_{1,0} = 0.671$. We produced a (100×100) 2-D random field and didn't use the first values of the region because of the substandard values. The last 64×64 part of the region was chosen. Since the symmetry properties of the PSF and the AR image model coefficients are assumed to be known, we have only three parameters d_5, d_6, d_9 to estimate in this experiment. In this part of experiment, we assumed that the AR parameters are known. The EM-algorithm needs the initialization parameters to estimate the unknown blur parameters. Choosing the blur coefficients requires special attention. The low-pass properties of the blurring function can be used. Additionally, the imaging system does not absorb or generate energy, so it is assumed that $\sum_{m,n \in W_D} d(m,n) = 1$.

All initial PSF coefficients were chosen equally as $d_5 = 0.111$, $d_6 = 0.111$, $d_9 = 0.111$. The estimated blur parameters are given in Table 5.1. for different SNRs.

Table 5.1 : Estimated blur parameters for different SNR values.

Original Parameters		$d_5 = 0.22$	$d_6 = 0.12$	$d_6 = 0.075$
Initialization parameters		0.111	0.111	0.111
Estimated parameters		$\hat{d}_5$	$\hat{d}_6$	$\hat{d}_9$
SNRs	60 dB	0.2193	0.1216	0.0736
	40 dB	0.2221	0.1218	0.0727
	30 dB	0.2247	0.1221	0.0717
	20 dB	0.2437	0.1229	0.0662
	10 dB	0.2611	0.1233	0.0614

SNR improvement values are shown in Figure 5.1. for the ideal and estimated coefficients. The original synthetic image, its blurred and noisy version and the restored images are demonstrated in Figure 5.2 for 30 dB SNR and SNR improvement is 2.82 dB.

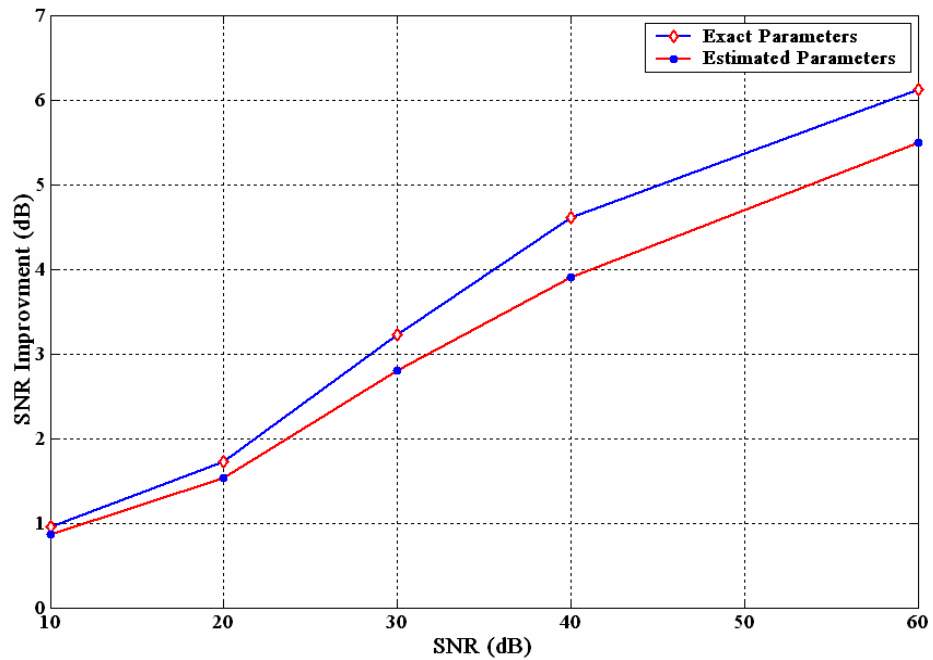


Figure 5-1: SNR improvement values for ideal and estimated values.

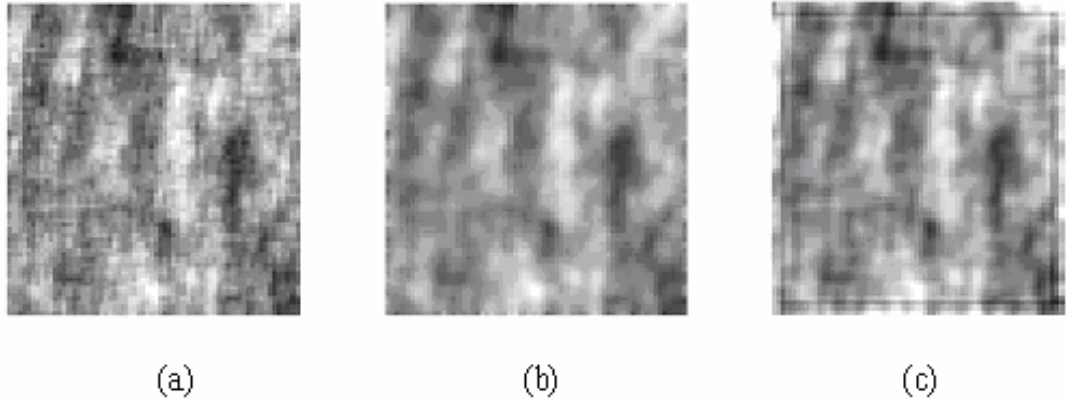


Figure 5.2: (a) The original synthetic image; (b) Blurred and noisy version with SNR=30 dB.; (c) Restored image with estimated coefficients.

The MSE between $\underline{d}$ and $\hat{\underline{d}}$ was computed for comparison purposes and plotted as a function of iteration for 30 dB SNR in Figure 5.3.

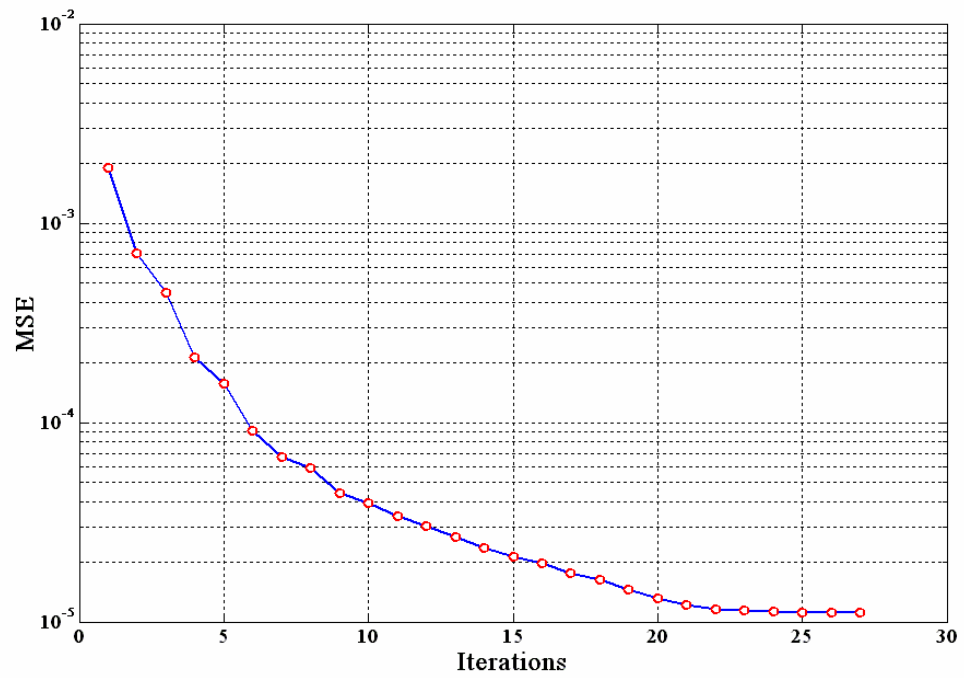


Figure 5.3: The MSE between $\underline{d}$ and $\hat{\underline{d}}$.

5.3.2. Real Images

We also applied the method to the real 64×64 blurred and noisy “Lena” images using the same blur parameters defined in Equation (5.29), additive Gaussian noise and first order NSHP AR model. For choosing the initialization parameters, we used the above properties of blurring. The initial image model coefficients were derived from the blurred and noisy version of the image by the Least Squares (LS) method. The identified parameters, which are obtained by the EM-algorithm, are listed in Table 5.2 and Table 5.3 for blur and image model parameters, respectively.

Table 5.2 : Estimated blur parameters for different initial values.

Original Parameters		$d_5 = 0.22$	$d_6 = 0.12$	$d_6 = 0.075$
Initialization parameters		0.111	0.111	0.111
Estimated parameters		$\hat{d}_5$	$\hat{d}_6$	$\hat{d}_9$
SNRs	60 dB	0.2214	0.1217	0.0729
	40 dB	0.2231	0.1219	0.0723
	30 dB	0.2252	0.1223	0.0714
	20 dB	0.2464	0.1231	0.0653
	10 dB	0.2632	0.1235	0.0607

Table 5.3 : Identification of AR image coefficients.

		$a_{1,1}$	$a_{1,0}$	$a_{-1,1}$	$a_{0,1}$
Estimated from original image by LS		-0.1434	0.1395	0.5215	0.4782
Initialization parameters from observed image by LS		-0.5224	0.2961	0.3475	0.8767
SNRs	60 dB	-0.1705	0.1641	0.3974	0.5342
	40 dB	-0.1839	0.1679	0.4123	0.5734
	30 dB	-0.2163	0.1711	0.4213	0.6274
	20 dB	-0.2345	0.1807	0.4392	0.6843
	10 dB	-0.2934	0.1985	0.4621	0.7115

SNR improvement values for different SNR's were calculated using the ideal and estimated coefficients and are shown in Figure 5.4. Figure 5.5. shows the original, blurred and noisy and restored images using the identified values obtained at SNRs of 10, 20 30, 40 and 60 dB.

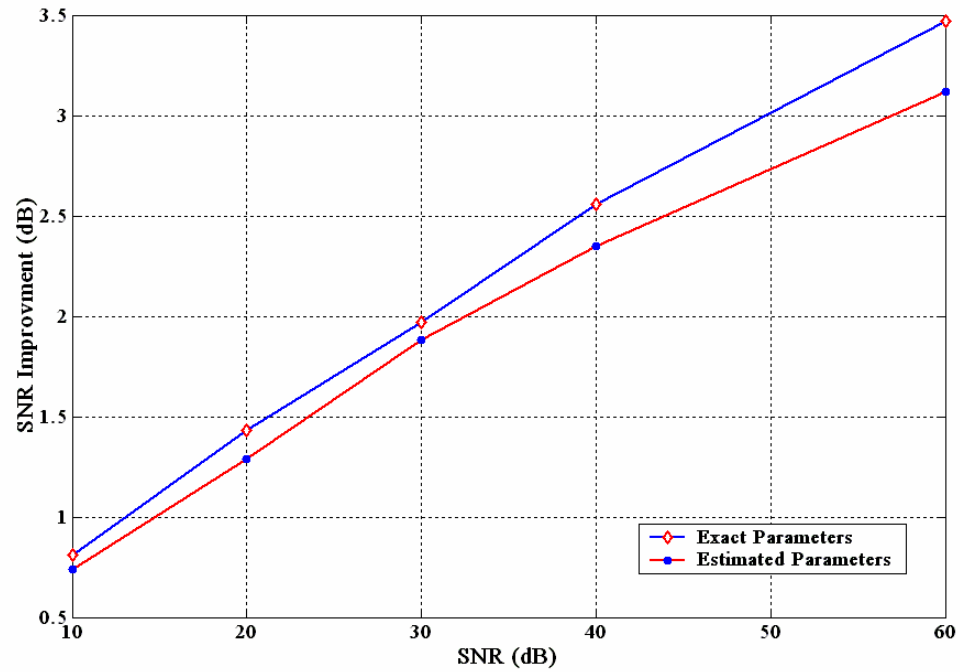


Figure 5-4: SNR improvement values for ideal and estimated values.

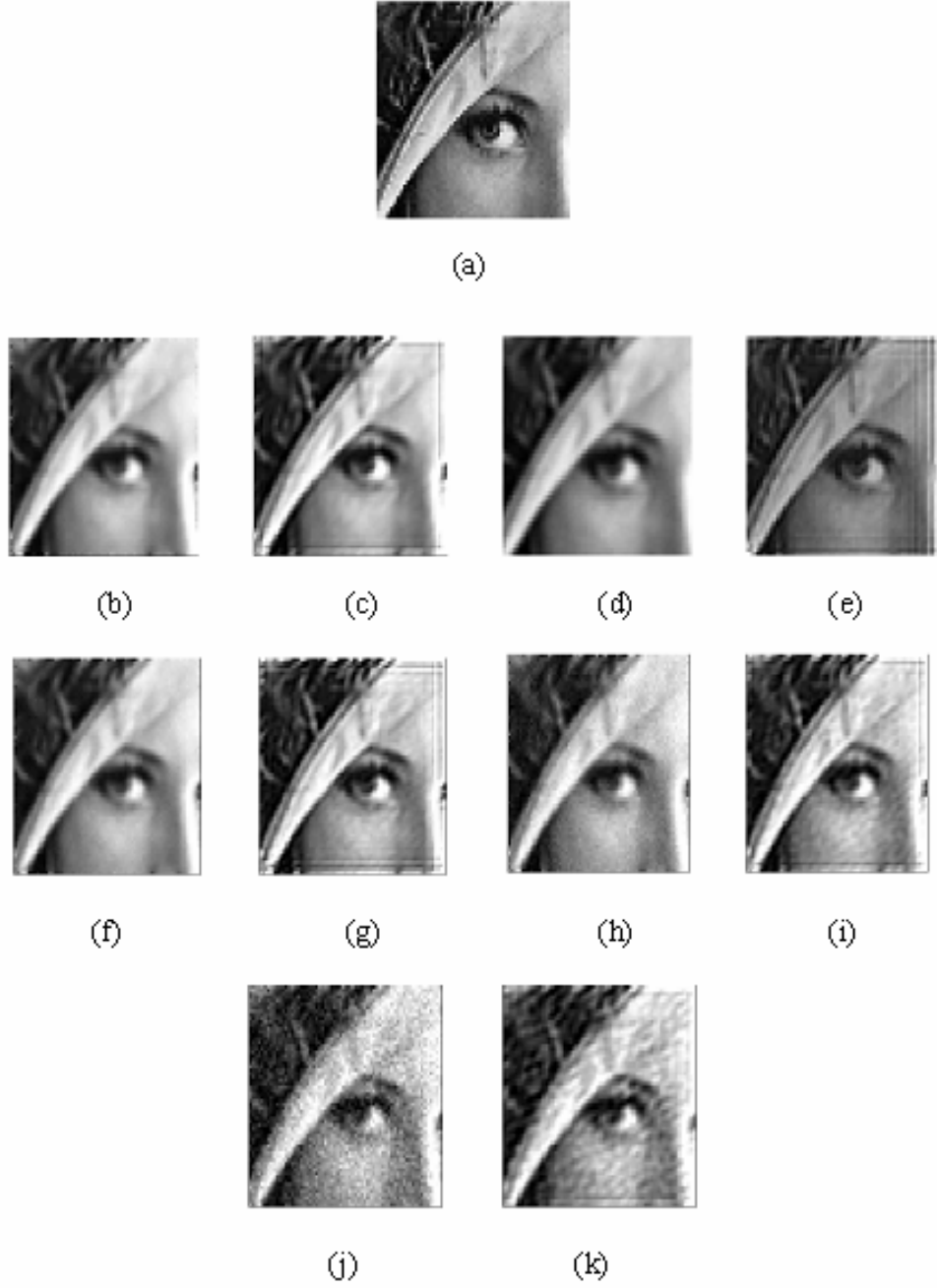


Figure 5-5: (a) The original 64×64 Lena image; (b) Blurred and noisy version with SNR=60 dB; (c) Restored image with learned parameters and SNR_{imp} = 3.18 dB. ; (d) Blurred and noisy version with SNR=40 dB.; (e) Restored image with learned parameters and SNR_{imp} = 2.32 dB. ; (f) Blurred and noisy version with SNR=30 dB.; (g) Restored image with learned parameters and SNR_{imp} = 1.81 dB. ; (h) Blurred and noisy version with SNR=20 dB.; (i) Restored image with learned parameters and SNR_{imp} = 1.29 dB. ; (j) Blurred and noisy version with SNR=10 dB.; (k) Restored image with learned parameters and SNR_{imp} = 0.61 dB.

The MSE between $\underline{d}$ and $\hat{\underline{d}}$ was computed as a function of iteration and is illustrated in Figure 5.6.

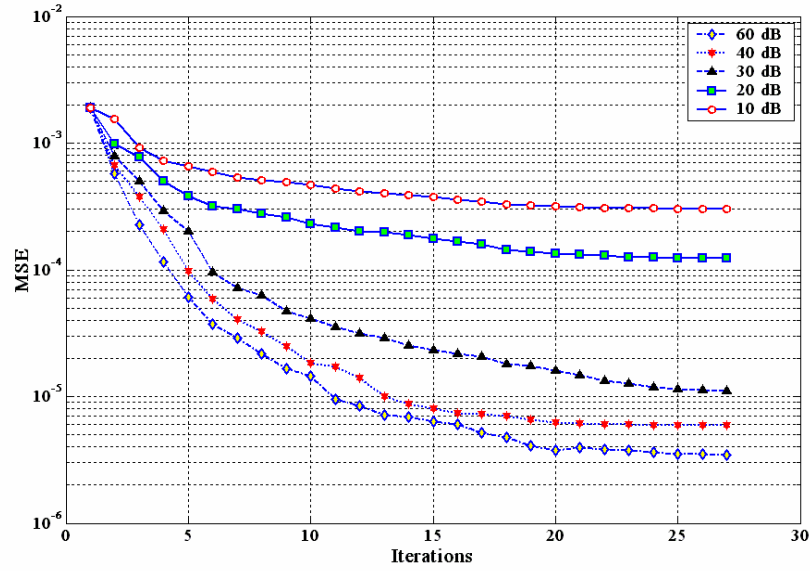


Figure 5-6: The MSE between $\underline{d}$ and $\hat{\underline{d}}$ as a function of iteration.

The MSE between $\underline{a}$ and $\hat{\underline{a}}$ was computed as a function of iteration and is illustrated in Figure 5.7.

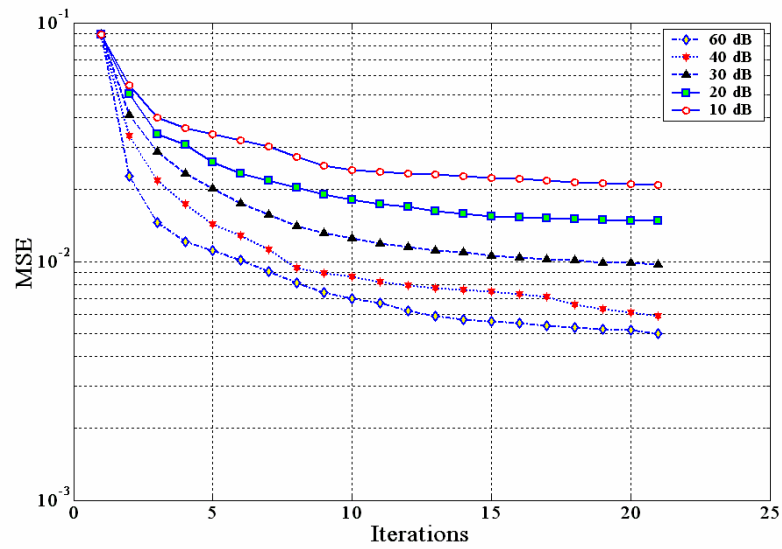


Figure 5-7: The MSE between $\underline{a}$ and $\hat{\underline{a}}$ as a function of iteration.

The steps of the EM algorithm are repeated until convergence. As an empirical rule of thumb, we stop the iterations when the ratio of the estimated parameter during the previous iteration and the current iteration is less than a threshold ($\epsilon=0.0001$).

5.4. Concluding Remarks

In this chapter, we introduced a new simultaneous recursive parameter learning and image restoration method based on the ML parameter identification and state estimation for images. This technique incorporates optimal Kalman smoothing equations for ML parameter identification and state estimation. The use of Kalman filtering instead of Kalman smoothing is presented because of the computationally heavy processing of smoothing. We derived an analytical and optimization solution for this EM based recursive learning. Although our proposed method deals with huge data sizes, because of recursive structure it does not need large size storage. We present the learning algorithm for the output-only case and the problem can be defined as an unsupervised learning problem. We focus on the problem of learning the parameters of a Bayesian network. In addition, a restored image is obtained simultaneously at the output of the Kalman filter. The algorithm requires extensive computation, which can be reduced considerably by using the reduced update Kalman filter [188] or the reduced order model Kalman filter [221] .

6. A NEW ACCELERATED EM BASED LEARNING OF THE IMAGE PARAMETERS AND RESTORATION

6.1. Introduction

Applying the EM algorithm is easy and stable in many problems, but there are two main drawbacks of the EM. One of them is slow convergence and the other one is finding closed form solution in the M step of EM. The EM algorithm becomes more attractive if its maximization step can be formulated analytically. But in many practical problems, the solution to the M step may not exist in closed form or may not be feasible in some practical problems [222] . For such situations, an optimization technique is used. Generalized EM (GEM) approach, which requires gradient or second-order ascent techniques for the M-step, is a typical example. The GEM algorithm uses the exact posterior in the E-step, but just partially maximizes the joint probability in the M-step. Of course, due to the iterative optimization in the M step, EM convergence speed can be extremely slow. Lagendijk [130] proposed an image identification algorithm that is based on the EM algorithm. Thanks to the EM algorithm, highly nonlinear operations of the ML estimation could be avoided. But, Lagendijk's solution was based on the optimization in the M step.

Many methods have been proposed to enhance the convergence speed of the EM algorithm, mostly based on the conventional optimization theory [223] [224] [225] [226] [227] . The procedures are usually not easy and need significant calculations [228] [229] [230] . Some of them are based on partitioning techniques of the data space such as the ordered-subset EM algorithm and its variants [231] . Liu and Sun [232] implemented Expectation-Conditional Maximization Either (ECME) to accelerate the EM algorithm for mixture distributions.

This chapter presents new approaches to learn the blur and image parameters and to restore the images by taking the above problems into account. We present a closed form solution in the M step for batch processing of images. In the case of non-closed form solutions, we introduce a new accelerated EM algorithm based on the optimization method in the M step to learn the unknown image parameters and to restore the images. The proposed methods simultaneously learn the image parameters and restore the images.

In our first solution, learning of unknown image and blur parameters is investigated in a closed form based on the EM algorithm. Cholesky factorization is used to find autoregressive (AR) image model parameters. Because of the Cholesky solution of AR image parameters, finding the other parameters is achievable in closed form. Regularization techniques are required because of the ill-posed structure of the image restoration problem but regularization parameter is generally chosen experimentally. In this work, also the regularization parameter is estimated in the same single structure. The restored image is obtained simultaneously using the regularized version of the Wiener filter.

In our second solution, acceleration is provided using FS optimization in the M step. A small number FS iterations satisfactorily perform well for each M step. Thanks to the EM, this method avoids operating directly on the nonlinear likelihood function. Additionally, it is possible to avoid complicated Hessian equations by using only the gradient values obtained by FS [233]. We also estimate the regularization parameter, so the regularized version of the Wiener is applied to images. Our proposed algorithm reaches to the local minima in few steps whereas conventional EM needs more iteration in M step. We also estimate the regularization parameter so the regularized version of the Wiener is applied to images.

6.2. Learning of Image Parameters

The unknown parameter vector θ is defined as,

$$\theta = [\theta_1, \theta_2, \dots, \theta_M]^T = [\dots d(m, n) \dots, \dots a(k, l) \dots, \sigma_v^2, \sigma_w^2]^T = [\underline{d}^T, \underline{a}^T, \sigma_v^2, \sigma_w^2]^T \quad (6.1)$$

where $\underline{d}$ is the $(L \times 1)$ unknown blur parameter vector and $\underline{a}$ is the $(K \times 1)$ unknown image parameter vector.

The complete data Y is chosen as the following form,

$$Y = [f \ g]^T \quad (6.2)$$

A restored version of the noisy and blurred image is determined in the E-step of the algorithm, whereas image and blur identification is performed in the M-step.

6.2.1. E-Step of the Algorithm

We can write the log-likelihood function of Y by factorization the joint probability as a product of conditional probabilities in the following form,

$$L_Y(\theta) = \log p_Y(y; \theta) = \log p_F(f; \theta) + \log p_{G|F}(g|f; \theta) \quad (6.3)$$

The log-likelihood function can be obtained as

$$\begin{aligned} L_Y(\theta) = c + \frac{1}{2} \{ \log |I - A|^2 - \log |Q| - f^T (I - A)^T Q^{-1} (I - A) f \} \\ - \frac{1}{2} \{ \log |R| + (g - Df)^T R^{-1} (g - Df) \} \end{aligned} \quad (6.4)$$

where c is a constant independent of θ and $|\bullet|$ denotes the determinant of a matrix.

The ML image and blur identification problem can be expressed as a minimization problem for the negative of the expression (6.4). $\Gamma(\theta)$ is defined as,

$$\Gamma(\theta) = -\left\{ c + \frac{1}{2} \log |I - A|^2 - \frac{1}{2} \log |Q| - \frac{1}{2} \text{tr} \left\{ \left[(I - A)^T Q^{-1} (I - A) \right] E(f f^T | g) \right\} \right. \\ \left. - \frac{1}{2} \log |R| - \frac{1}{2} g^T R^{-1} g + g^T R^{-1} D \mu_{f|g} - \frac{1}{2} \text{tr} \left\{ (D^T R^{-1} D) E(f f^T | g) \right\} \right\} \quad (6.5)$$

where $E(f f^T | g)$ is defined as;

$$E(f f^T | g) = C_{f|g} + \mu_{f|g} \mu_{f|g}^T \quad (6.6)$$

here $\mu_{f|g}$ and $C_{f|g}$ are the conditional mean and conditional covariance of f given g and θ_i respectively. $\mu_{f|g}$ and $C_{f|g}$ are defined by the following equations:

$$\mu_{f|g} = E_{\theta} \{ f | g \} = \hat{f} \quad (6.7)$$

$$C_{f|g} = E_{\theta} \left\{ (f - \mu_{f|g}) (f - \mu_{f|g})^T | g \right\} \quad (6.8)$$

The following equations can be derived for the conditional mean and conditional covariance,

$$\mu_{f|g} = \hat{f} = R_{ff} D^T (D R_{ff} D^T + R_{ww})^{-1} g \quad (6.9)$$

$$C_{f|g} = R_{ff} - R_{ff} D^T (D R_{ff} D^T + R_{ww})^{-1} D R_{ff} \quad (6.10)$$

Here, $\hat{f}$ is the restored image as Wiener filter output and R_{ff} is the autocorrelation of the ideal image f . Since the image model is an AR process, the image autocorrelation matrix is given by,

$$R_{ff} = E[ff^T] = (I - A)^{-1} Q (I - A)^{-T} = \sigma_v^2 (I - A)^{-1} (I - A)^{-T} \quad (6.11)$$

$$R_{ff}^{-1} = (I - A)^T Q^{-1} (I - A) = \frac{1}{\sigma_v^2} (I - A)^T (I - A) \quad (6.12)$$

R_{ww} is the autocorrelation of the observation noise and $R_{ww} = R = \sigma_w^2 I$.

If we rewrite the Equation (6.9), we can conclude with the Regularized Wiener filter given as follows,

$$\hat{f} = \left(D^T D + \frac{\sigma_w^2}{\sigma_v^2} (I - A)^T (I - A) \right)^{-1} D^T g \quad (6.13)$$

where $\alpha = \sigma_w^2 / \sigma_v^2$ is called as the regularization parameter [104]. The regularization parameter α plays an essential role in restoring the image. If α is too small, the resulting image cannot avoid noise amplification and cannot satisfactorily reflect a priori smoothness limitations. On the contrary, if α is too large, the resulting image becomes too smooth and it cannot keep detail information of the original image. After learning of the original image AR parameters, blur parameters and variances, we can use the regularized Wiener filter to restore the observed image.

6.2.2. M Step

6.2.2.1. Closed Form

Each of the unknown parameters is calculated by taking the corresponding partial derivative of the $\Gamma(\theta)$, setting to zero, and solving. Differentiating with respect to the components of θ_1 yields,

$$\begin{aligned}
\frac{\partial \Gamma(\theta)}{\partial \theta_i} = & -\frac{1}{2} \frac{\partial}{\partial \theta_i} \log |I - A|^2 + \frac{1}{2} \frac{\partial}{\partial \theta_i} \log |Q| \\
& + \frac{1}{2} \frac{\partial}{\partial \theta_i} \text{tr} \left\{ \left[(I - A)^T Q^{-1} (I - A) \right] E(f f^T | g) \right\} \\
& + \frac{\partial}{\partial \theta_i} \frac{1}{2} \log |R| + \frac{1}{2} \frac{\partial}{\partial \theta_i} (g^T R^{-1} g) - \frac{\partial}{\partial \theta_i} (g^T R^{-1} D \mu_{f|g}) \\
& + \frac{1}{2} \frac{\partial}{\partial \theta_i} \text{tr} \left\{ (D^T R^{-1} D) E(f f^T | g) \right\}
\end{aligned} \tag{6.14}$$

For blur parameters, closed form solution can be found easily as,

$$\frac{\partial \Gamma(\theta)}{\partial D} = -R^T g \hat{f}^T + R^T D E(f f^T | g) = 0 \tag{6.15}$$

$$D^{new} = g \hat{f}^T \left[E(f f^T | g) \right]^{-1} \tag{6.16}$$

Finding AR image model parameters is not as easy as finding the blur parameters.

Let us determine the derivate of $\Gamma(\theta)$ with respect to $(I - A)$ and set it to zero.

$$\frac{\partial \Gamma(\theta)}{\partial (I - A)} = -(I - A)^{-T} + Q^{-1} (I - A) E(f f^T | g) = 0 \tag{6.17}$$

We can reach the following equation,

$$(I - A)^T Q^{-1} (I - A) = \left[E(f f^T | g) \right]^{-1} \tag{6.18}$$

where $E(f f^T | g)$ is a positive definite matrix. The matrix $(I - A)$ is a lower triangular matrix because of the causal image model supports and disregard of the image model boundary effects. Q is a diagonal matrix, which is defined as $\sigma_v^2 I$. It can be seen easily that Q is a strictly positive definite matrix. Since the diagonal of Q is strictly positive, the above equation can be written as:

$$E(f f^T | g) = \left[(I - A)^{-1} Q^{\frac{1}{2}} \right] \left[Q^{\frac{1}{2}} (I - A)^{-T} \right] = \left[(I - A)^{-1} Q^{\frac{1}{2}} \right] \left[(I - A)^{-1} Q^{\frac{1}{2}} \right]^T \quad (6.19)$$

Let us define W as,

$$W^T = (I - A)^{-1} Q^{\frac{1}{2}} \quad (6.20)$$

So,

$$E(f f^T | g) = W^T W \quad (6.21)$$

here W is a nonsingular and upper triangular matrix and called as the Cholesky factor or “square root” of $E(f f^T | g)$. We can solve the above equation by using the Cholesky Factorization (Decomposition) [54] [168] [234] :

$$A^{new} = I - Q^{1/2} W^{-T} \quad (6.22)$$

For the covariance matrix of the model noise,

$$\frac{\partial \Gamma(\theta)}{\partial Q^{-1}} = -\frac{1}{2} Q + \frac{1}{2} (I - A) E^T(f f^T | g) (I - A)^T = 0 \quad (6.23)$$

$$Q^{new} = (I - A^{new})^T \left[E(f f^T | g) \right]^T (I - A^{new}) \quad (6.24)$$

For the covariance matrix of the observation noise,

$$\frac{\partial \Gamma(\theta)}{\partial R^{-1}} = -\frac{1}{2} R + \frac{1}{2} g g^T - g^T D \hat{f}^T + \frac{1}{2} D \left[E(f f^T | g) \right]^T D^T = 0 \quad (6.25)$$

$$R^{new} = g g^T - 2 g^T D^{new} \hat{f} + D^{new} \left[E(f f^T | g) \right]^T (D^{new})^T \quad (6.26)$$

The following steps summarize the closed form solution of the proposed EM algorithm:

1) Initialization :

- Initialize the unknown parameters $\theta^{(0)} = \left[\underline{d}^{T^{(0)}}, \underline{a}^{T^{(0)}}, (\sigma_v^2)^{(0)}, (\sigma_w^2)^{(0)} \right]^T, i=0.$

2) E-Step:

- Determine conditional mean (the restored image) $\hat{f}$, by using the regularized Wiener filter given by Equation (6.13).
- Determine $E(f f^T | g)$ by using conditional mean (6.13) and conditional covariance (6.10).

3) M-Step:

- Calculate the new estimate of θ^{new} to learn unknown parameters.
 - o For blur parameters: Compute (6.16).
 - o For image model parameters: Determine W^T from $E(f f^T | g)$ by using Cholesky factorization, then compute (6.22).
 - o For model noise variance: Compute (6.24).
 - o For observation noise variance: Compute (6.26).

4) Stopping criterion:

- Continue until $\|\theta^{i+1} - \theta^i\| < \varepsilon$.

6.2.2.2.Fast Optimization Solution

In some circumstances, to find a closed form might be inefficient if not impossible. When it is difficult to carry out the M-step exactly, iterative techniques can be used. The most popular method is the Newton-Raphson method, which has quadratic convergence. But one of the problems with this method is that it can be very sensitive to the choice of starting values and therefore can often fail to converge to a solution [235].

The other drawback of Newton's method is that it requires the analytic derivative of Hessian matrix per iteration. But computing second derivatives for a large class of

optimization is not feasible [236] . We prefer to use the FS optimization because of fast convergence and computational effectiveness.

The implementation of the FS algorithm requires evaluation of the log-likelihood gradient and the FIM. The parameter vectors $\underline{d}, \underline{a}$ and noise variances σ_v^2, σ_w^2 are independent of each other, so the FIM will have a block diagonal form as follows:

$$I_f(\theta) = \begin{bmatrix} [I_f(\underline{d})]_{(L \times L)} & \emptyset & \emptyset & \emptyset \\ \emptyset & [I_f(\underline{a})]_{(K \times K)} & \emptyset & \emptyset \\ \emptyset & \emptyset & I_f(\sigma_v^2) & \emptyset \\ \emptyset & \emptyset & \emptyset & I_f(\sigma_w^2) \end{bmatrix} \quad (6.27)$$

The matrices $I_f(\underline{d})$ and $I_f(\underline{a})$ are the corresponding FIM's which are defined as follows:

$$I_f(\underline{d}) = \begin{bmatrix} [I_f(d)]_{11} & \cdot & \dots & [I_f(d)]_{1l} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ [I_f(d)]_{l1} & \cdot & \dots & [I_f(d)]_{ll} \end{bmatrix} \quad (6.28)$$

$$I_f(\underline{a}) = \begin{bmatrix} [I_f(a)]_{11} & \cdot & \dots & [I_f(a)]_{1k} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ [I_f(a)]_{k1} & \cdot & \dots & [I_f(a)]_{kk} \end{bmatrix} \quad (6.29)$$

The i, j th component of the each corresponding FIM, $[I_f(\theta)]_{i,j}$ is defined by,

$$[I_f(\theta)]_{i,j} = E \left[\frac{\partial \Gamma(\theta)}{\partial \theta_i} \frac{\partial \Gamma(\theta)}{\partial \theta_j} \right] \quad (6.30)$$

$\frac{\partial \Gamma(\theta)}{\partial \theta_i}$ can be written as,

$$\frac{\partial \Gamma(\theta)}{\partial \theta_i} = \frac{1}{2} \text{tr} \{ \Lambda_i \} + \frac{1}{2} \mathbf{g}^T \mathcal{G}_i \quad (6.31)$$

where

$$\begin{aligned} \Lambda_i = & \left\{ Q^{-1} \frac{\partial Q}{\partial \theta_i} + R^{-1} \frac{\partial R}{\partial \theta_i} - \left((I - A)(I - A)^T \right)^{-1} \frac{\partial}{\partial \theta_i} \left((I - A)(I - A)^T \right) \right. \\ & \left. + E \left(f f^T | g \right) \left[\frac{\partial}{\partial \theta_i} \left((I - A)^T Q^{-1} (I - A) \right) + \frac{\partial}{\partial \theta_i} \left(D^T R^{-1} D \right) \right] \right\} \end{aligned} \quad (6.32)$$

$$\mathcal{G}_i = \frac{\partial R^{-1}}{\partial \theta_i} \mathbf{g} - 2 \frac{\partial}{\partial \theta_i} (R^{-1} D) \mu_{f|g} \quad (6.33)$$

For the FIM, we obtain the following equation:

$$\left[I_f(\theta) \right]_{i,j} = \frac{1}{4} E \left\{ \text{tr} \left(\Lambda_i \otimes \Lambda_j \right) \right\} + \frac{1}{4} E \left\{ \text{tr} \left(\Lambda_i \right) \mathcal{G}_j^T \mathbf{g} \right\} + \frac{1}{4} E \left\{ \mathbf{g}^T \mathcal{G}_i \text{tr} \left(\Lambda_j \right) \right\} + \frac{1}{4} E \left\{ \mathbf{g}^T \mathcal{G}_i \mathcal{G}_j^T \mathbf{g} \right\} \quad (6.34)$$

In Equation (6.34), since all expectations contain real valued constants, we can write the FIM as the following form.

$$\left[I_f(\theta) \right]_{i,j} = \frac{1}{4} \left\{ \text{tr} \left(\Lambda_i \otimes \Lambda_j \right) + \text{tr} \left(\Lambda_i \right) \mathcal{G}_j^T \mathbf{g} + \mathbf{g}^T \mathcal{G}_i \text{tr} \left(\Lambda_j \right) + \mathbf{g}^T \mathcal{G}_i \mathcal{G}_j^T \mathbf{g} \right\} \quad (6.35)$$

We use the following properties,

$$\text{tr}(A) \text{tr}(B) = \text{tr}(A \otimes B) \quad (6.36)$$

here $\otimes$, denotes Kronecker product operator [237]. The derivatives of $\Gamma(\theta)$ with respect to each unknown parameter are given in Appendix E. All detailed derivations are given in Appendix F.

The following steps summarize the iterative optimization solution of the proposed EM algorithm:

1) Initialization :

- Initialize the unknown parameters $\theta^{(0)} = \left[d^{T(0)}, a^{T(0)}, (\sigma_v^2)^{(0)}, (\sigma_w^2)^{(0)} \right]^T$, $i=0$.

2) E-Step:

- Determine conditional mean (the restored image) $\hat{f}$, by using the regularized Wiener filter given by Equation (6.13).
- Determine $E(f f^T | g)$ by using conditional mean (6.13) and conditional covariance (6.10).

3) M-Step:

- Calculate the new estimate of θ^{new} to learn unknown parameters.

For $k=1,2,\dots$

- Determine (6.27) by using (6.35) for each unknown parameters
- Compute FS optimization equation.

If $\|\theta^{k+1} - \theta^k\| < \delta$, stop. Otherwise, set $k \leftarrow k + 1$.

4) Stopping criterion:

- Continue until $\|\theta^{i+1} - \theta^i\| < \varepsilon$.

6.3. Frequency Domain Solution

Implementation of the method in the spatial domain is impractical due to the inversion of the huge size matrices. The matrices D and A are BTTB structures. A BT matrix is often approximated by a BC matrix, because these two matrix types are structurally closely related as explained in section 2.6. The advantage of having a circular structure is that BC matrices can be efficiently evaluated using 2-D DFT. For large matrices, the errors introduced by approximations are usually small. Using the diagonalization properties, the method is computed in the DFT domain as a set of scalar problems and implementation is possible even for large images. Wiener equations in DFT domain were given in Section 3.9.

6.4. Experimental Results

The performance of the proposed algorithm is tested using two different types of images, a 64×64 artificial image and the 64×64 real “Lena” image. We assume that the image is $N \times N$, where N is the number of pixels in each direction, and thus contains N^2 pixels, and accordingly D is $N^2 \times N^2$. Typically N is a power of 2, such as 32, 64, 128, 256 or larger, then the number of unknowns grows to at least 1024, 4096, 16384, 65536, respectively. In order to reduce the computational density, we assume that the unknown PSF is real and symmetric, therefore a number of PSF coefficients have equal values. The results are demonstrated for both Closed Form and Fast Optimization methods.

6.4.1. Experiments on Artificial Images

The artificial image is produced using an AR model with NSHP support. For the AR model of the image, $\sum_{k,l \in W_A} a(k,l) \leq 1$ is assumed. Independent white Gaussian noise is added to this artificial image. We chose the following NSHP AR parameters as, $a_{11} = -0.449$; $a_{01} = 0.630$; $a_{-11} = 0.127$; $a_{1,0} = 0.671$. We produced a (100x100) 2-D random field, and we only used lower right corner of this artificial data.

The artificial image is blurred by using the following (3 x 3) blur PSF.

$$\begin{bmatrix} d(-1,-1) & d(-1,0) & d(-1,1) \\ d(0,-1) & d(0,0) & d(0,1) \\ d(1,-1) & d(1,0) & d(1,1) \end{bmatrix} = \begin{bmatrix} d_1 & d_2 & d_3 \\ d_4 & d_5 & d_6 \\ d_7 & d_8 & d_9 \end{bmatrix} = \begin{bmatrix} 0.075 & 0.12 & 0.075 \\ 0.12 & 0.22 & 0.12 \\ 0.075 & 0.12 & 0.075 \end{bmatrix} \quad (6.37)$$

Since the symmetry properties of the PSF and the AR image model coefficients are assumed to be known, we have only three parameters d_5, d_6, d_9 to estimate in this experiment. The EM-algorithm needs the initialization parameters. Especially, choosing the blur coefficients requires special attention. There are some features to guide us for choosing the initialization parameters. For instance, we can use the low-pass properties of the blurring function. Additionally, the imaging system does not

absorb or generate energy, so it is assumed that $\sum_{m,n \in W_D} d(m,n) = 1$. We applied three different initializations of the PSF in three cases.

Case 1: No blur ($d(m,n) = \delta(m,n)$). $d_5 = 1.0$, $d_6 = 0$, $d_9 = 0$

Case 2: All PSF coefficients equally: $d_5 = 0.111$, $d_6 = 0.111$, $d_9 = 0.111$

Case 3: Near minimal value of the likelihood function:

$$d_5 = 0.144, d_6 = 0.114, d_9 = 0.1$$

In this part of experiment, we assumed that the AR parameters are known. The estimated blur parameters are given in Table 6.1 for both solutions and for different initial values at 30 dB Signal-to-Noise (SNR) value.

Table 6.1: Estimated blur parameters for different initial values.

Case	Original Values					
	$d_5=0.22$		$d_6=0.12$		$d_9=0.075$	
	Closed Form	Optimization	Closed Form	Optimization	Closed Form	Optimization
1	0.2444	0.2508	0.1310	0.1347	0.0579	0.0526
2	0.2292	0.2360	0.1236	0.1269	0.0691	0.0641
3	0.2236	0.2348	0.1145	0.1175	0.0801	0.0738

From the above experiment it is seen that the PSF needs to be initialized relatively close to the global optimum of the likelihood function. We can measure the restoration performance with a Signal to Noise Ratio (SNR) improvement factor. SNR improvement values for different SNR's are given in Figure 6.1 for the ideal and estimated coefficients.

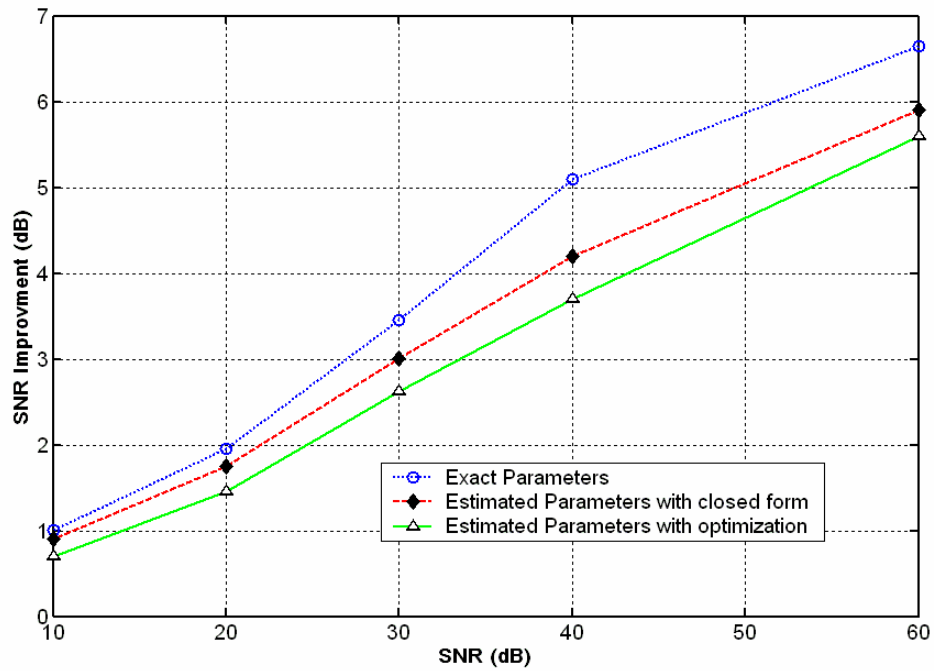
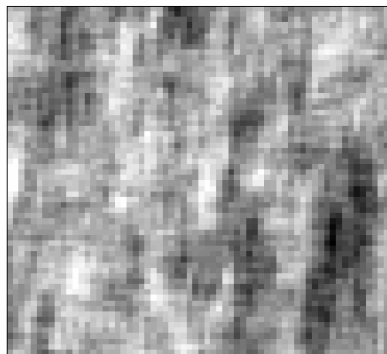
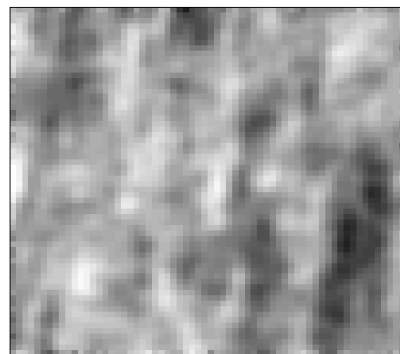


Figure 6-1: SNR improvement values for ideal and estimated values.

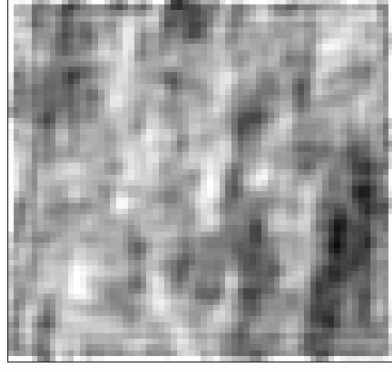
Restoration is performed using the regularized Wiener filter. The original synthetic image, its blurred and noisy version and the restored images are given in Figure 6.2 for an SNR of 30 dB. Restored images have SNR improvement of 3.02 dB for closed form solution and 2.85 dB for optimization solution. In the M step of the optimization technique, 3 or 4 iterations are usually sufficient.



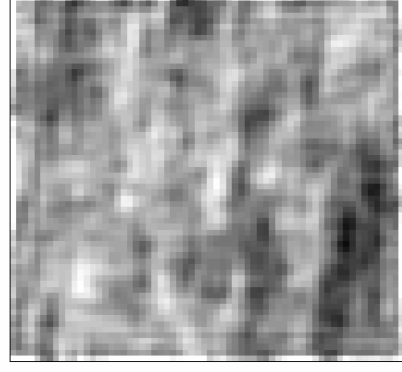
(a)



(b)



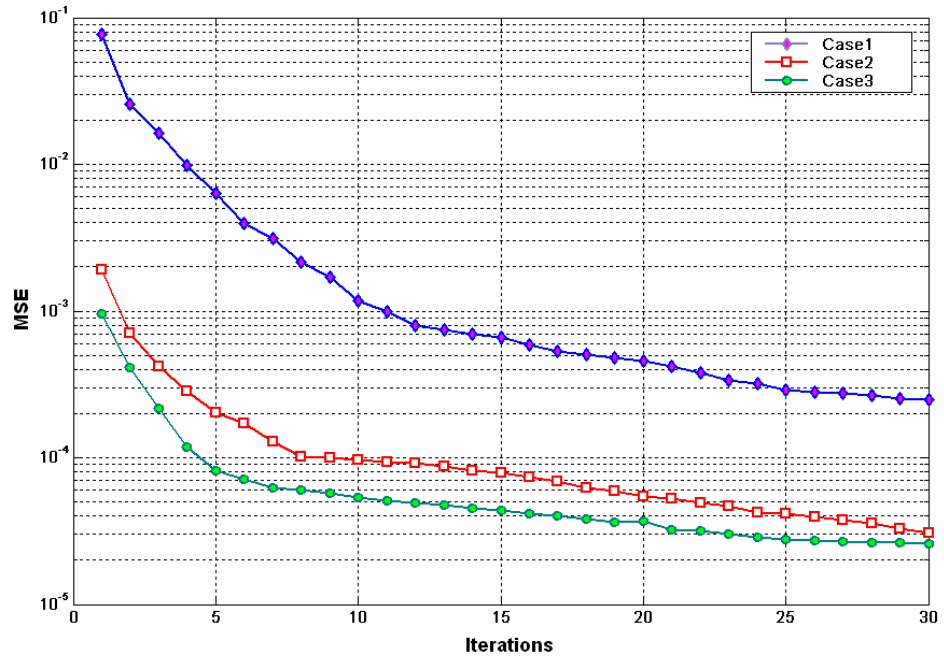
(c)



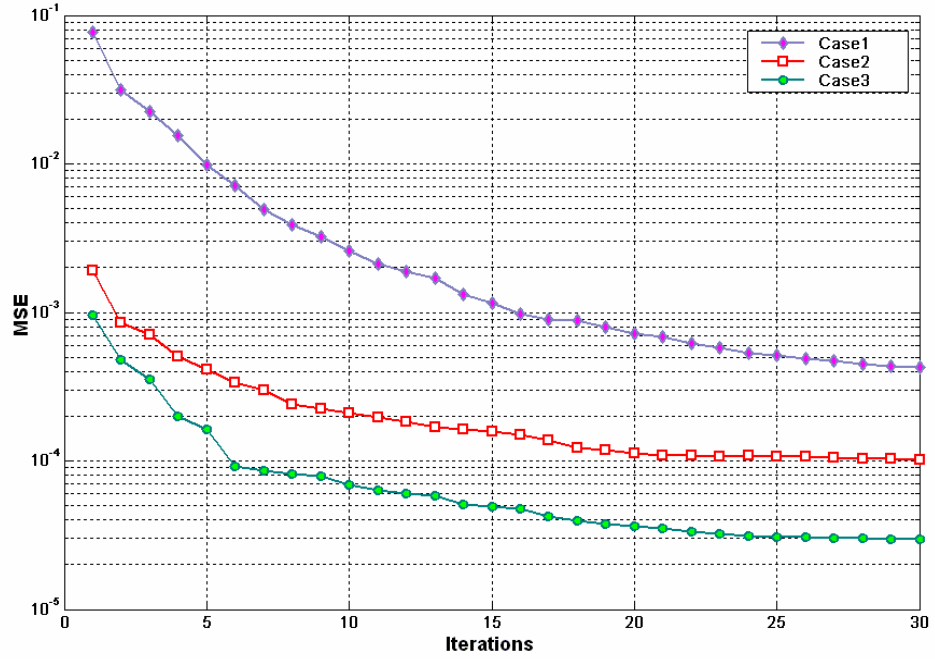
(d)

Figure 6.2: (a) The original synthetic image; (b) Blurred and noisy version with SNR=30 dB.; (c) Restored image with estimated coefficients by closed form solution (SNRimp = 3.02 dB.) ; (d) Restored image with estimated coefficients by optimization solution (SNRimp = 2.85 dB.)

The learning performance of the blur parameters is tested by calculating the MSE after each step. The MSE between $\underline{d}$ and $\hat{\underline{d}}$ was computed for comparison purposes and plotted as a function of iteration for 30 dB SNR value in Figure 6.3 for both closed form and optimization solutions at 3 different cases.



(a)



(b)

Figure 6-3: The MSE between $\underline{d}$ and $\hat{\underline{d}}$ for (a) Closed form solution , (b) Optimization solution.

6.4.2. Real Images

We also applied the closed form and optimization solutions to the real 64×64 blurred and noisy “Lena” images using the same blur parameters defined in (6.37), additive Gaussian noise and first order NSHP AR model. For choosing the initialization parameters, we used the above properties of blurring. The initial image model coefficients were derived from the blurred and noisy version of the image by the Least Squares (LS) method. The identified parameters, which are obtained by the EM-algorithm, are listed in Table 6.2 and Table 6.3. The ratio of the estimated observation noise and model noise (the regularization parameter) is found as

$$\hat{\alpha} = \frac{\hat{\sigma}_w^2}{\hat{\sigma}_v^2} = 0.194 \text{ at final step of the algorithm.}$$

SNR improvement values for different SNR’s were calculated by Wiener filter using the ideal and estimated coefficients and are given in Figure 6.4. Pretty good identification and restoration quality is obtained for SNR values between 60-10 dB.

Figure 6.5 shows the original, blurred and noisy and restored images using the identified values obtained from closed form solution at SNR of 10, 20 30, 40 and 60 dB for Case 2.

Table 6.2: Estimated blur parameters for different initial values.

SNR	Case	Original Values					
		$d_5=0.22$		$d_6=0.12$		$d_9=0.075$	
		Closed Form	Optimization	Closed Form	Optimization	Closed Form	Optimization
60 dB	1	0.167	0.157	0.117	0.115	0.091	0.096
	2	0.199	0.191	0.125	0.128	0.076	0.074
	3	0.218	0.213	0.115	0.117	0.081	0.080
40 dB	1	0.159	0.148	0.107	0.106	0.103	0.107
	2	0.195	0.187	0.13	0.134	0.071	0.069
	3	0.207	0.196	0.121	0.123	0.077	0.078
30 dB	1	0.146	0.144	0.105	0.104	0.109	0.110
	2	0.191	0.185	0.131	0.134	0.071	0.069
	3	0.201	0.193	0.122	0.126	0.078	0.076
20 dB	1	0.143	0.138	0.106	0.103	0.109	0.113
	2	0.181	0.174	0.133	0.139	0.072	0.067
	3	0.193	0.183	0.128	0.131	0.074	0.073
10 dB	1	0.132	0.129	0.104	0.101	0.113	0.117
	2	0.151	0.145	0.134	0.14	0.078	0.074
	3	0.167	0.152	0.132	0.139	0.076	0.073

Table 6. 3: Identification of AR image coefficients.

		$a_{1,1}$	$a_{1,0}$	$a_{-1,1}$	$a_{0,1}$
Estimated from original image by LS		-0.1434	0.1395	0.5215	0.4782
Initialization parameters from observed image by LS		-0.5224	0.2961	0.3475	0.8767
Closed Form	60 dB	-0.1694	0.1602	0.3918	0.5235
	40 dB	-0.1823	0.1599	0.4274	0.5743
	30 dB	-0.2141	0.1654	0.4192	0.6104
	20 dB	-0.2272	0.1764	0.4331	0.6740
	10 dB	-0.2814	0.1986	0.4419	0.6905
Optimization	60 dB	-0.1721	0.1692	0.4012	0.5356
	40 dB	-0.1932	0.1704	0.4363	0.5845
	30 dB	-0.2311	0.1762	0.4476	0.6354
	20 dB	-0.2421	0.1832	0.4510	0.6930
	10 dB	-0.2921	0.2107	0.4643	0.7247

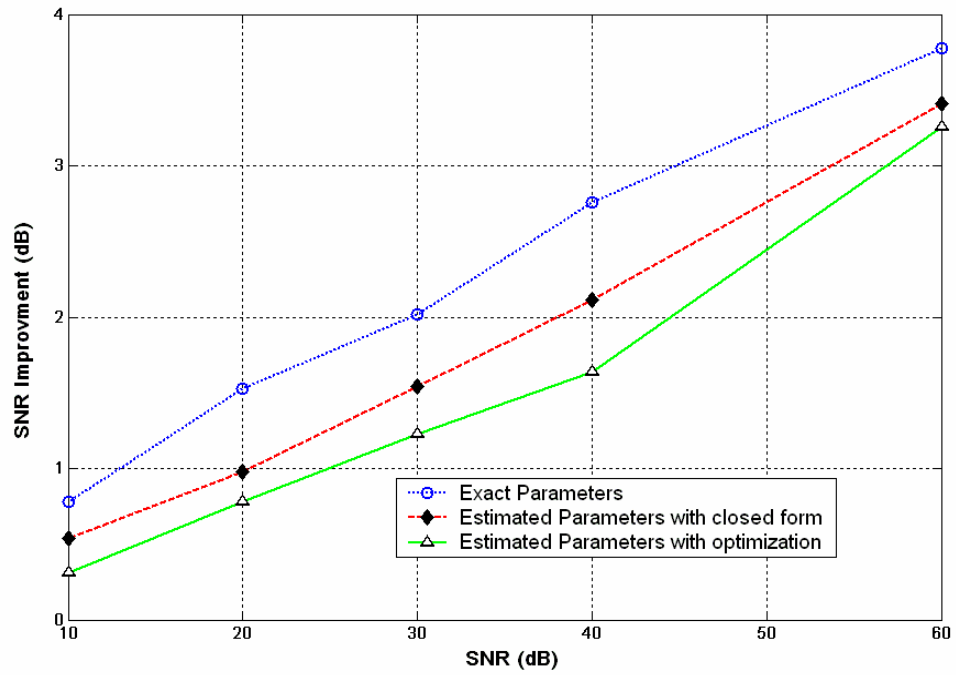


Figure 6.4: SNR improvement values for ideal and estimated values.

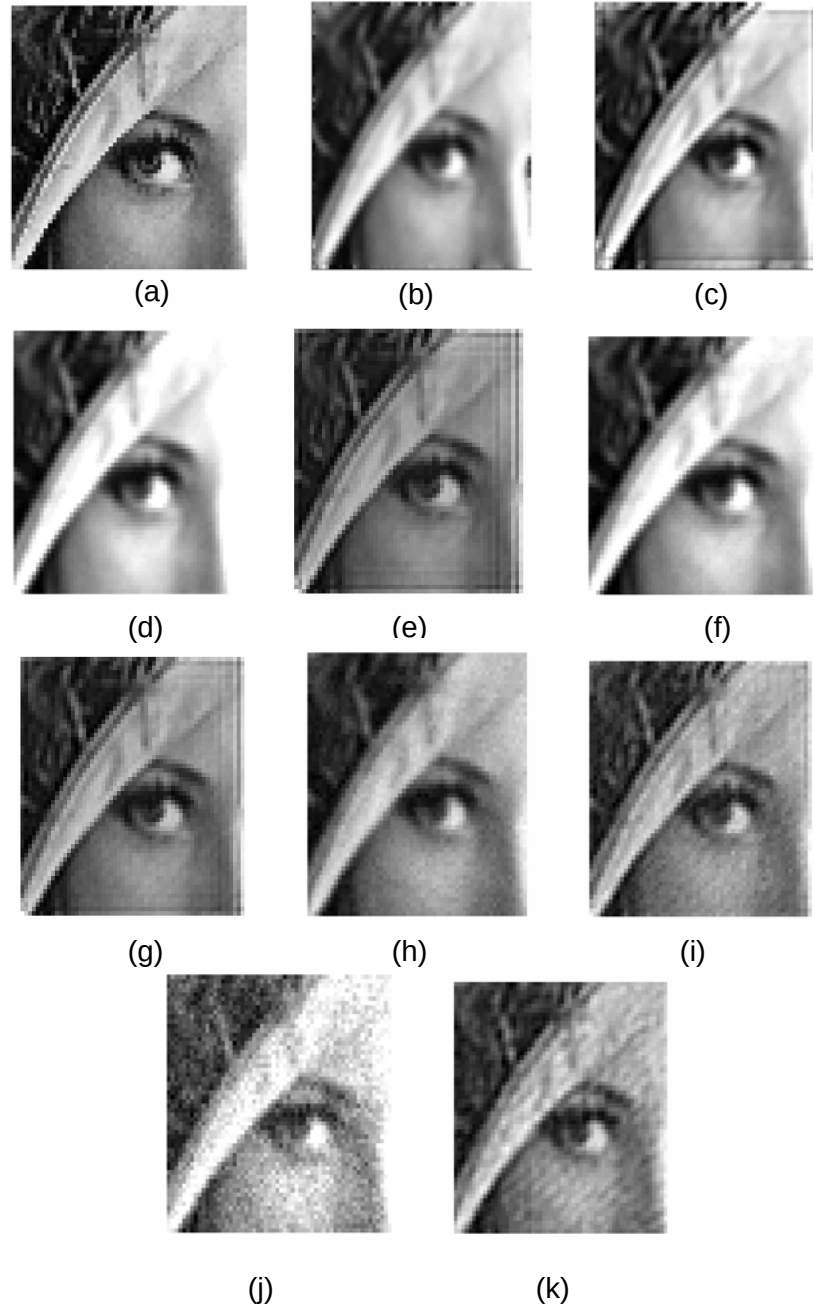
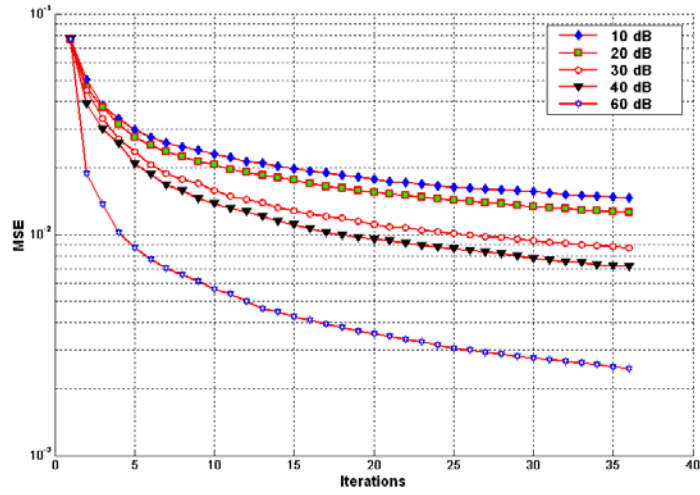
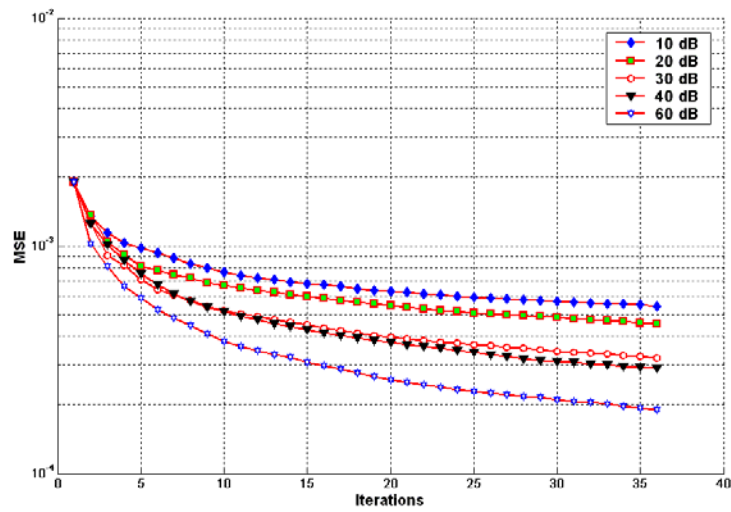


Figure 6-5: (a) The original 64×64 Lena image; (b) Blurred and noisy version with SNR=60 dB; (c) Restored image with learned parameters and SNRimp = 3.36 dB. ; (d) Blurred and noisy version with SNR=40 dB.; (e) Restored image with learned parameters and SNRimp = 2.24 dB. ; (f) Blurred and noisy version with SNR=30 dB.; (g) Restored image with learned parameters and SNRimp = 1.78 dB. ; (h) Blurred and noisy version with SNR=20 dB.; (i) Restored image with learned parameters and SNRimp = 0.99 dB. ; (j) Blurred and noisy version with SNR=10 dB.; (k) Restored image with learned parameters and SNRimp = 0.73 dB.

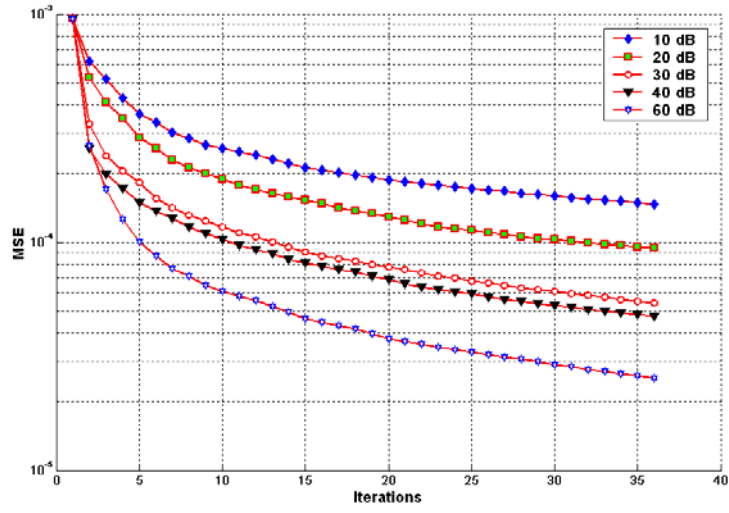
The MSE between $\underline{d}$ and $\hat{\underline{d}}$ was computed as a function of iteration for closed form solution in Figure 6.6 and for optimization solution in Figure 6.7 for different SNR values at each cases.



(a)

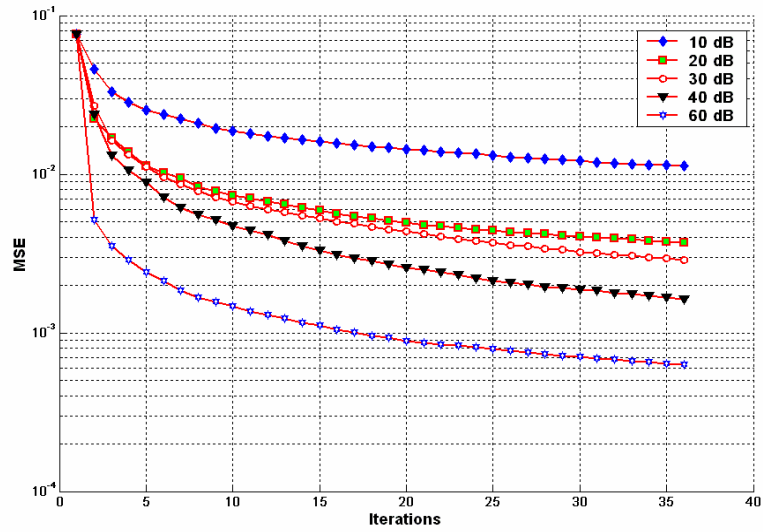


(b)

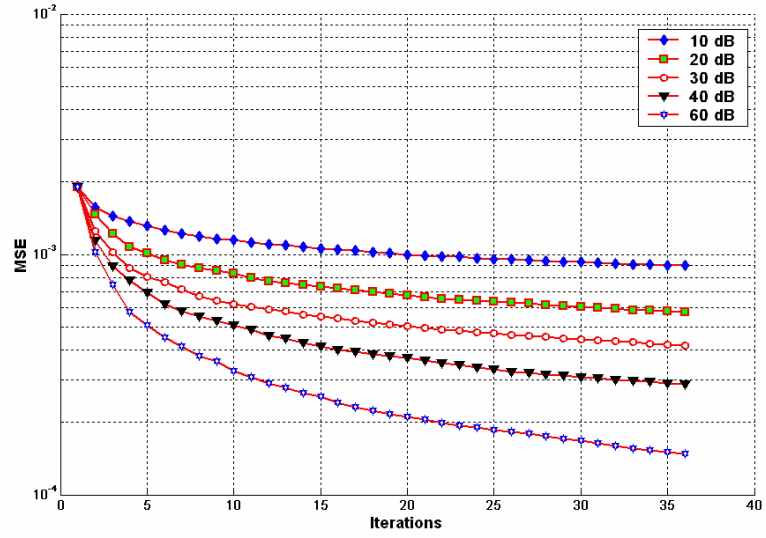


(c)

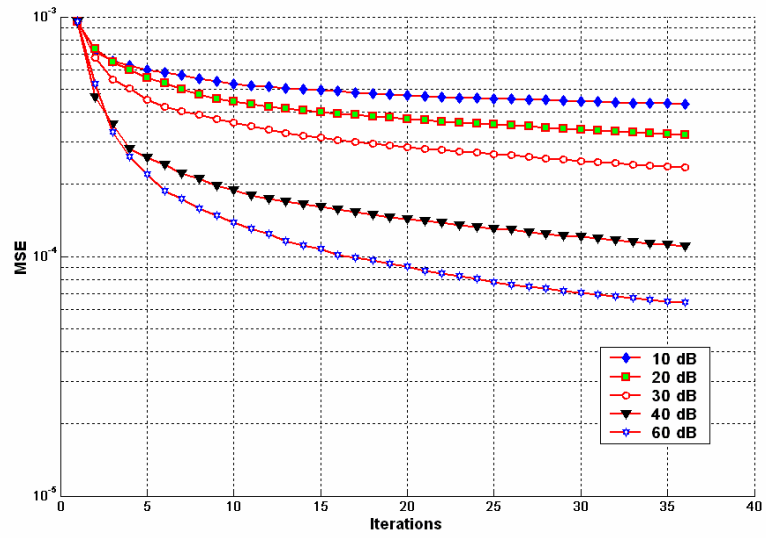
Figure 6-6: The MSE between $\underline{d}$ and $\hat{\underline{d}}$ as a function of iteration for closed form solution.(a) Case1 (b) Case2 (c) Case3.



(a)



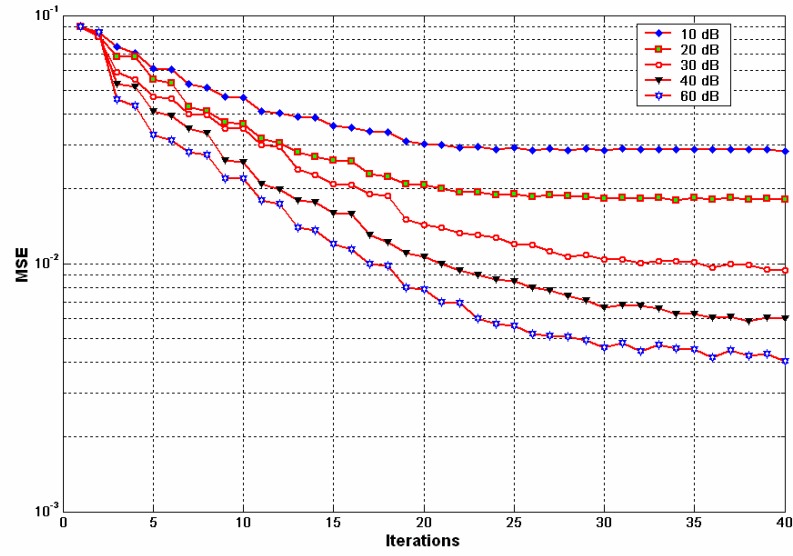
(b)



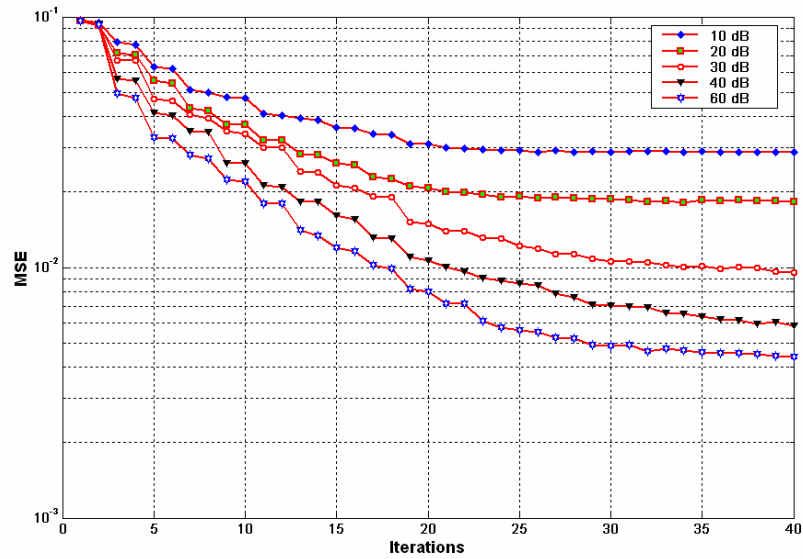
(c)

Figure 6-7: The MSE between $\underline{d}$ and $\hat{\underline{d}}$ as a function of iteration for optimization solution. (a) Case1 (b) Case2 (c) Case3.

The MSE between $\underline{a}$ and $\hat{\underline{a}}$ was computed as a function of iteration for closed form solution and optimization solution in Figure 6.8. for different SNR values.



(a)



(b)

Figure 6-8: The MSE between $\underline{a}$ and $\hat{\underline{a}}$ as a function of iteration for (a) closed form solution. (b) optimization solution.

The steps of the EM algorithm are repeated until convergence. As an empirical rule of thumb, we stopped the iterations when the ratio of the estimated parameter during the previous iteration and the current iteration is less than a threshold ($\epsilon=0.0001$).

6.5. Concluding Remarks

In this chapter, we proposed a new method based on the EM algorithm to learn the unknown image parameters and to restore images, simultaneously. The EM algorithm has slow convergence properties and finding a closed form solution in the M step is generally not easy task. We undertook to solve both problems for blind image deconvolution as a batch processing. We obtained a closed form structure in which the Cholesky factorization was used to find autoregressive image model parameters. Additionally, a new method based on the accelerated EM algorithm was introduced. In the M step of the EM algorithm, we used the FS method to accelerate the EM algorithm. Only a small number FS iteration is required for each M step. Thanks to the FS optimization, it is possible to avoid complicated second derivative of the log-likelihood function by using only the gradient of the log-likelihood function. Also the regularization parameter was estimated and restored image was obtained using the regularized version of the Wiener filter. To cope with the computational load, both recursive processing in time domain, and a frequency domain solution, can be used.

7. RADAR DETECTION APPLICATION: LEARNING OF THE SEA CLUTTER PARAMETERS FOR ADAPTIVE RADAR DETECTION

7.1. Introduction

The understanding and modeling of radar clutter plays an essential role for radar system design and performance evaluation [238] . Sea clutter is modeled as a stochastic process composed of random scattering from the sea surface. The clutter statistics vary widely and generally different type of clutter is modeled using different type of distribution [239] . In the past, Gaussian distribution model was used to model radar clutter at low sea state or in case of using low-resolution radar. But in many applications, clutter cannot be Gaussian and depends on the changing environment. For instance, high-resolution radar systems and low resolution radars at high sea states have target-like spikes that give rise to non-Gaussian heavy tailed observations [240] . Consequently, clutter cannot be considered as homogenous. While traditional radar detector is designed to operate against Gaussian noise, new detection processors are required to reduce the effects of the spikes and to improve detection performance.

Adaptive detection is important for modern radar systems, which are tasked to operate in non-homogeneous and non-stationary environments. Statistical characterization and modeling of radar clutter is of great importance for developing adaptive threshold algorithms. Target detection requires the comparison of the square magnitude with a certain threshold. One approach to setting a detection threshold is based on the Constant False Alarm Rate (CFAR). It is extremely important to maintain a CFAR when the background noise level fluctuates. Several types of CFAR techniques have been proposed in the literature [241] [242] [243] [244] [245] [246] .

Most probably, future advanced radar systems will be able to detect, identify, and estimate the parameters of a target in severe interference backgrounds [247] . Design of adaptive radar detection algorithms requires that the parameters should be learned from the operational environment [248] [249] . Learning of the clutter parameters has been proposed, recently [250] .

In order to detect targets using adaptive CFAR techniques, clutter background statistics should be known a priori. But in practice, background statistics are typically not known a priori. Thus, a joint estimation and detection process is employed whereby the background statistics are estimated before target detection [251] . The ML estimation method is one of the standard tools for parameter estimation. In spite of being computationally expensive, the MLE method application in modern processors might be able to handle the additional processing. In radar detection, various ML based CFAR algorithms have been studied for different types of background. For example, in [252] , a low-loss ML-CFAR algorithm was developed for Weibull background. The two parameters (shape and scale) of the background statistics were estimated using a ML algorithm. An optimal Maximum Likelihood Generalized Gamma (MLGG) CFAR detector was developed in [253] . MLE technique was applied for the estimation of the parameters in the Rice distribution [254] . Also, the Optimal Weibull (OW) CFAR detector was developed using the MLE method [255] .

Mixture distributions seem to be strong candidates for modelling both sea and land clutters. For instance, the contaminated-normal distribution, which is the sum of two Gaussians with different variances, was formulated by Trunk [256] in order to model the sea clutter for non-coherent detection. For modelling sea clutter, Conte and Longo [257] suggested that almost all of the major statistical models might be represented as a Rayleigh mixture model and this suggestion was later verified by Sangston and Gerlach [258] . The optimal and suboptimal target detection problem was re-examined in non-Gaussian clutter represented by the compound Gaussian distribution [259] . Exponential mixture probability density functions (pdfs) were shown to be useful representations of radar sea clutter in [260] . Dogandžić et al. [261] developed ML methods for jointly estimating target and clutter parameters in compound-Gaussian clutter using radar array measurements.

In this chapter, we introduce a new clutter model that is more suitable for the changing environment. In order to model the sea clutter, we propose to use a generalized distribution, which is valid for different background statistics like low and high sea states. A generalized Gaussian Mixture (GM) pdf with zero mean is used in order to form a statistical model of the clutter for each of the inphase (I) and quadrature (Q) channels. Our new background pdf, which is present in the output of the envelope detector, resemble a Rician Mixture. Therefore, we learn (estimate) the required sea clutter parameters from the environment to calculate an adaptive threshold. The derivation of the envelope detector output, which has multi component GM inputs, and the ML estimation of this output are the issues, which have not been addressed before. In addition, the estimation of the mixture parameters to calculate the detection threshold in CFAR was investigated for the first time.

The ML background parameter learning problem is formulated as maximizing the log-likelihood function or minimizing the minus log-likelihood function. The only way to optimize the log-likelihood function, which is nonlinear in nature, is by means of mathematical programming techniques. First, we propose to use the Gradient descent optimization method to estimate the unknown sea clutter parameters. It is known that the gradient descent algorithm has slow convergence properties, therefore second-order information is essential for fast convergence. Secondly, we apply the Fisher Scoring (FS) optimization method to estimate the unknown sea clutter parameters in a fast way [30] . Performance analysis is presented using real sea clutter data, which was collected by the IPIX radar, at low and high sea states [29] .

7.2. Clutter Modeling

The pdf's of the I and Q channels, which are the inputs of the envelope detector, are denoted by $p(x)$ and $p(y)$, respectively, which may be given as

$$p(x) = \sum_{j=1}^J \alpha_j \frac{1}{(2\pi(\sigma_x^2)_j)^{1/2}} \exp\left[-\frac{x^2}{2(\sigma_x^2)_j}\right] \quad (7.1)$$

$$p(y) = \sum_{k=1}^K \beta_k \frac{1}{(2\pi(\sigma_y^2)_k)^{1/2}} \exp\left[-\frac{y^2}{2(\sigma_y^2)_k}\right] \quad (7.2)$$

Here, J and K denote the component number of x and y . α_j and β_k are the mixture ratios, where $\alpha_j \geq 0$, $\sum_{j=1}^J \alpha_j = 1$, $\beta_k \geq 0$, $\sum_{k=1}^K \beta_k = 1$. $(\sigma_x^2)_j$ and $(\sigma_y^2)_k$ are the variances associated with I and Q channels, respectively. For convenience it is assumed that $(\sigma_x^2)_j$ and $(\sigma_y^2)_k$ are ordered so that $(\sigma_x^2)_1 < (\sigma_x^2)_2 < \dots < (\sigma_x^2)_J$ and $(\sigma_y^2)_1 < (\sigma_y^2)_2 < \dots < (\sigma_y^2)_K$.

Let $x = r \cos \theta$ and $y = r \sin \theta$, so that $r = \sqrt{x^2 + y^2}$ is the output of the envelope detector. If it is assumed that I and Q components of the clutter are independent, we can calculate the output of the envelope detector using the density transformation technique. The joint density of r and θ is related to the joint density of x and y by the Jacobian transformation as given by the following form:

$$p(r, \theta) = |J| p(x, y) \quad (7.3)$$

Here

$$|J| = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{vmatrix} r = r \quad (7.4)$$

and

$$p(r, \theta) = r \left\{ \sum_j \sum_k \frac{\alpha_j \beta_k}{2\pi \sqrt{(\sigma_x^2)_j (\sigma_y^2)_k}} \exp \left[-\frac{r^2 \left[\cos^2 \theta (\sigma_y^2)_k + \sin^2 \theta (\sigma_x^2)_j \right]}{2 (\sigma_x^2)_j (\sigma_y^2)_k} \right] \right\} \quad (7.5)$$

The output pdf of the envelope detector is given by,

$$p(r) = \int_0^{2\pi} p(r, \theta) d\theta \quad (7.6)$$

After the transformation and integration, we calculate the pdf of the output of the envelope detector by the following equation:

$$p(r) = \sum_{j=1}^J \sum_{k=1}^K \frac{\alpha_j \beta_k r}{((\sigma_x^2)_j (\sigma_y^2)_k)^{1/2}} \exp[-A_{jk} r^2] I_0(B_{jk} r^2) \quad (7.7)$$

where $I_n(\bullet)$ is the n^{th} order Modified Bessel function of the first kind of order [262] .

A_{jk} and B_{jk} are respectively defined as:

$$A_{jk} = \frac{1}{4} \left(\frac{1}{(\sigma_x^2)_j} + \frac{1}{(\sigma_y^2)_k} \right) \quad (7.8)$$

$$B_{jk} = \frac{1}{4} abs \left(\frac{1}{(\sigma_y^2)_k} - \frac{1}{(\sigma_x^2)_j} \right) \quad (7.9)$$

Note that despite some similarities, the above pdf $p(r)$ is very different from the Rician Mixtures. It can be seen easily that when $(\sigma_y^2)_k = (\sigma_x^2)_j$, our model is a Rayleigh mixture. Almost all of the major statistical models that have been proposed for the statistics of the sea clutter (the log-normal being an exception) may be represented as a Rayleigh mixture model [257] .

7.3. ML Estimation of the Model Parameters

Adaptive threshold setting problem can be specified as the estimation of the parameter vector ϕ , defined as $\phi = (\phi_1, \phi_2, \dots, \phi_M)^T = (\alpha_i, \beta_k, \sigma_x^2, \sigma_y^2)^T$ from the observed data (output of the envelope detector). Adaptive CFAR threshold is set according to $T = f\left(P_{fa}, \alpha_j, \beta_k, (\sigma_x^2)_j, (\sigma_y^2)_k\right)$ and it is based on estimating the mixture ratios and variances. The ML-CFAR algorithm is implemented as shown in Figure 7.1, by setting the adaptive threshold T [252].

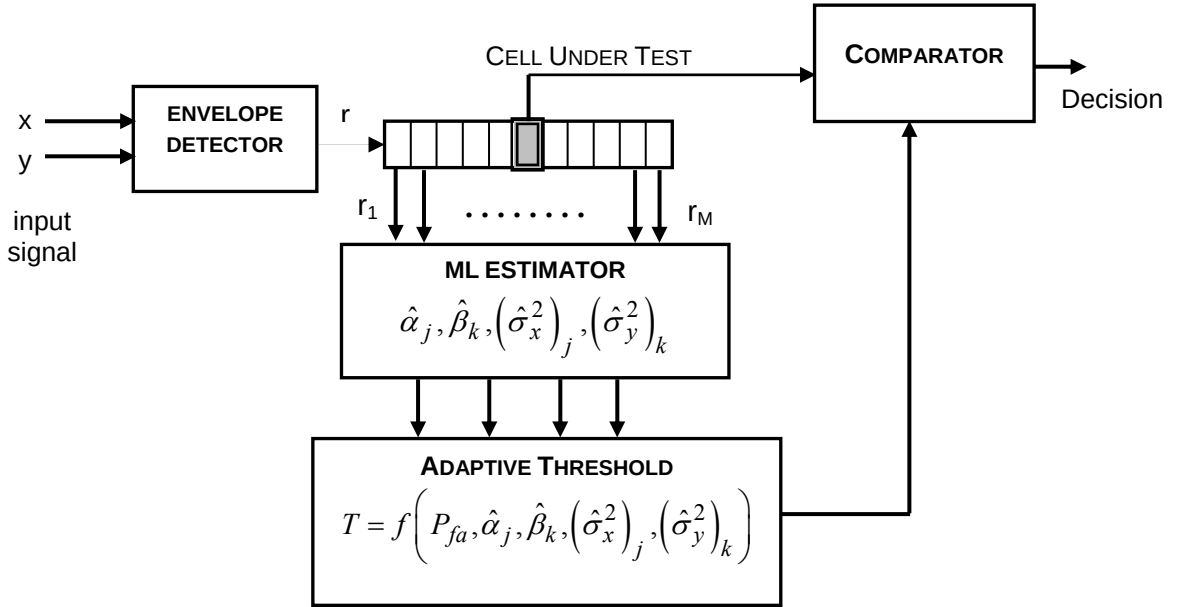


Figure 7-1: ML-CFAR structure.

The ML estimates of unknown parameters are obtained from the M background samples. From M independent observations, the log-likelihood function is given by,

$$-L(\phi) = -\log \prod_{m=1}^M p(r_m; \phi) = -\sum_{m=1}^M \log p(r_m; \phi) \quad (7.10)$$

The classical way to find the parameters involves taking the partial derivatives of the likelihood function with respect to the required parameters, and solving by setting the resulting equations to zero. Since Equation (7.10) does not have a closed form

solution, which is typical for likelihood functions, we resort to numerical optimization techniques.

7.4. Gradient Based Optimization

The gradient of the log-likelihood function should be calculated with respect to each unknown parameter. We define the required gradients in the following equations:

$$\frac{\partial L(\phi)}{\partial \alpha_n} = - \sum_{m=1}^M \frac{\sum_{k=1}^K \frac{\beta_k}{\sqrt{(\sigma_x^2)_n (\sigma_y^2)_k}} r_m \exp[-A_{nk}] I_0(B_{nk})}{p(r_m)} \quad (7.11)$$

$$\frac{\partial L(\phi)}{\partial \beta_n} = - \sum_{m=1}^M \frac{\sum_{j=1}^J \frac{\alpha_j}{\sqrt{(\sigma_x^2)_j (\sigma_y^2)_n}} r_m \exp[-A_{jn}] I_0(B_{jn})}{p(r_m)} \quad (7.12)$$

$$\frac{\partial L(\phi)}{\partial (\sigma_x^2)_n} = - \sum_{m=1}^M \frac{\sum_{k=1}^K \frac{\alpha_n \beta_k}{2 [(\sigma_x^2)_n]^{3/2} [(\sigma_y^2)_k]^{1/2}} r_m \exp[-A_{nk}] \{\varphi_{nmk}\}}{p(r_m)} \quad (7.13)$$

where φ_{nmk} is defined as

$$\varphi_{nmk} = \begin{cases} \frac{R_m^2}{2(\sigma_x^2)_n} [I_0(B_{nk}) + I_1(B_{nk})] - I_0(B_{nk}) & \text{for } \frac{1}{(\sigma_y^2)_k} > \frac{1}{(\sigma_x^2)_n} \\ \frac{R_m^2}{2(\sigma_x^2)_n} [I_0(B_{nk}) - I_1(B_{nk})] - I_0(B_{nk}) & \text{for } \frac{1}{(\sigma_y^2)_k} < \frac{1}{(\sigma_x^2)_n} \end{cases} \quad (7.14)$$

$$\frac{\partial L(\phi)}{\partial (\sigma_y^2)_n} = - \sum_{m=1}^M \frac{\sum_{j=1}^J \frac{\alpha_j \beta_n}{2 [(\sigma_x^2)_j]^{1/2} [(\sigma_y^2)_n]^{3/2}} R_m \exp[-A_{jn}] \{\gamma_{nmj}\}}{p(R_m)} \quad (7.15)$$

In (7.15), γ_{nmj} is defined as

$$\gamma_{nmj} = \begin{cases} \frac{R_m^2}{2(\sigma_y^2)_n} [I_0(B_{jn}) + I_1(B_{jn})] - I_0(B_{jn}) & \text{for } \frac{1}{(\sigma_y^2)_n} > \frac{1}{(\sigma_x^2)_j} \\ \frac{R_m^2}{2(\sigma_y^2)_n} [I_0(B_{jn}) - I_1(B_{jn})] - I_0(B_{jn}) & \text{for } \frac{1}{(\sigma_y^2)_n} < \frac{1}{(\sigma_x^2)_j} \end{cases} \quad (7.16)$$

Now, the gradient-based optimization method can be applied. If the number of components of the mixture distributions is high, it is very difficult to solve the above equations simultaneously. Fortunately, we know that a small value of component number is found to be sufficient to give excellent approximation in several cases of practical interest [263] .

7.5. FS Optimization

The parameters $\alpha_j, \beta_k, (\sigma_x^2)_j, (\sigma_y^2)_k$ are independent from each other, so the FIM will have a block diagonal form as follows:

$$I_f(\phi) = \begin{bmatrix} I_f(\underline{\alpha}) & 0 & 0 & 0 \\ 0 & I_f(\underline{\beta}) & 0 & 0 \\ 0 & 0 & I_f(\sigma_x^2) & 0 \\ 0 & 0 & 0 & I_f(\sigma_y^2) \end{bmatrix} \quad (7.17)$$

where $\underline{\alpha}, \underline{\beta}, \sigma_x^2, \sigma_y^2$ are the vector form of the unknown parameters. The matrices $I_f(\underline{\alpha}), I_f(\underline{\beta}), I_f(\sigma_x^2)$ and $I_f(\sigma_y^2)$ are the corresponding FIM's and are defined as:

$$I_f(\underline{\theta}) = \begin{bmatrix} I_{11}(\theta) & \dots & I_{1n}(\theta) \\ \vdots & \ddots & \vdots \\ I_{m1}(\theta) & \dots & I_{mn}(\theta) \end{bmatrix} \quad (7.18)$$

The FIM entries are defined as follows:

$$I_{ij}(\alpha) = E \left[\frac{\partial L(\phi)}{\partial \alpha_i} \frac{\partial L(\phi)}{\partial \alpha_j} \right] = \int_0^\infty \left(\frac{\partial L(\phi)}{\partial \alpha_i} \frac{\partial L(\phi)}{\partial \alpha_j} \right) p(r) dr \quad (7.19)$$

Notice that the entries of the FIM need to be computed with numerical integration. For many models, computation of the FIM is much simpler than that of the Hessian matrix.

7.6. The Probability of False Alarm

The probability of false alarm (P_{fa}) is a function of the mixture ratios, variances and threshold T . (P_{fa}) is defined as,

$$P_{fa} = \int_T^{\infty} p(r) dr \quad (7.20)$$

where T is the threshold value [264]. Using estimated values, (P_{fa}) is calculated as for $A_{jk} > B_{jk} > 0$,

$$P_{fa} = \sum_{j=1}^J \sum_{k=1}^K \hat{\alpha}_j \hat{\beta}_k - \sum_{j=1}^J \sum_{k=1}^K \frac{\hat{\alpha}_j \hat{\beta}_k}{2 \left((\hat{\sigma}_x^2)_j (\hat{\sigma}_y^2)_k \right)^{1/2}} \left(\int_0^{T^2} e^{-(A_{jk})u} I_0 \left((B_{jk})u \right) du \right) \quad (7.21)$$

Since, the integral in the above equation cannot be calculated analytically, it is evaluated numerically.

7.7. Experimental Results

In order to test our model, we work on the real sea clutter data collected at high and low sea states with McMaster University IPIX radar. These data consist of 131072 samples for I and Q channels at high and low sea states. There are a number of graphs that can be used to check the deviations of the data from the normal distribution. The most useful tool for assessing normality is a quantile quantile (QQ) plot [265]. If the data are from the theoretical distribution, the points on the QQ plot lie approximately on a straight line. It is easily seen from IPIX real radar data that the probability distribution of low sea state data has shorter tails than those of the Normal distribution [27]. By contrast, the probability distribution of high sea state data has longer tails than those of the normal distribution. Figure 7.2 and 7.3 show QQ plot of real data versus standard Normal at low and high sea states, respectively.

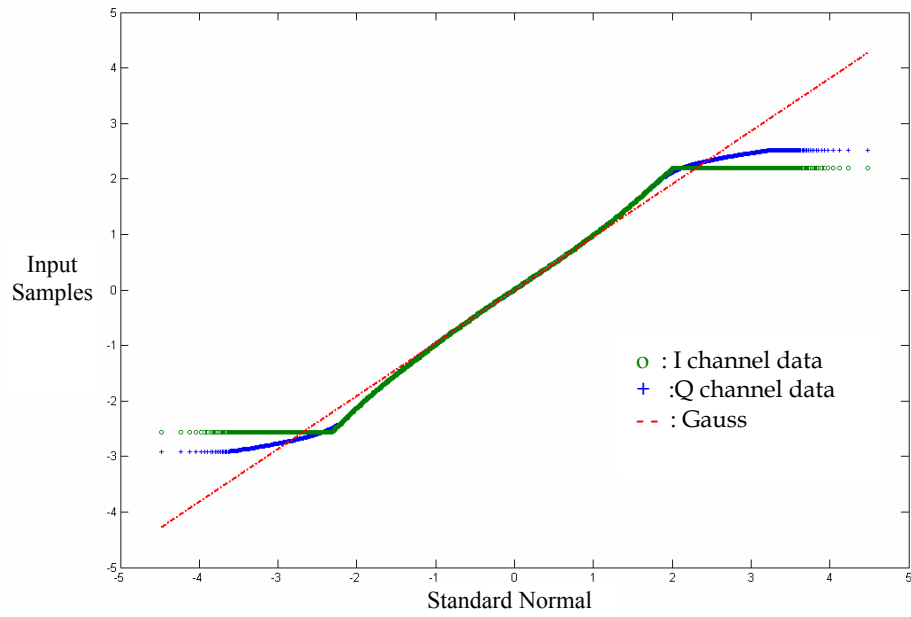


Figure 7-2: QQ plot of sample data versus quantiles of standard normal distribution at low sea states.

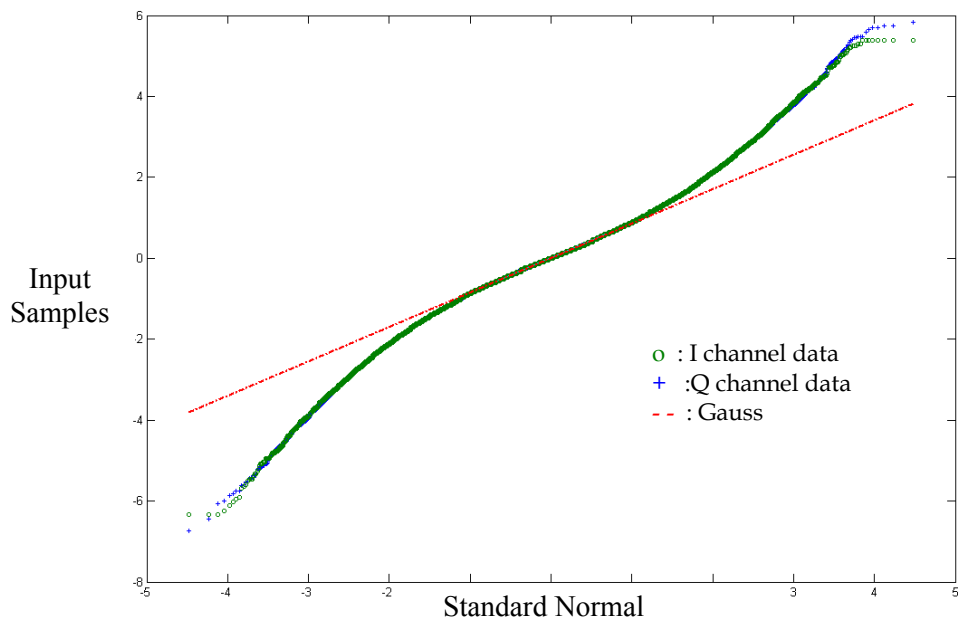


Figure 7-3: QQ plot of sample data versus quantiles of standard normal distribution at high sea states.

It is easily seen from IPIX real radar data that the probability distribution of low sea state data has shorter tails than those of the Normal distribution. By contrast, the probability distribution of high sea state data has a longer tails than those of the

normal distribution. For simplification, we assume that I and Q channels have the same mixture ratios and the same variances.

$$i) \sigma_n^2 = (\sigma_x^2)_n = (\sigma_y^2)_n$$

$$ii) \alpha_n = \beta_n.$$

So it is enough to estimate σ_n^2 and α_n for this example. Quite good fits to the IPIX radar data at high and low sea states have been found using a GM with 3 components and the estimated values are given in Table 7.1. Although choosing the initial values seems to be a problem, there are some restrictions about it. We can summarize these restrictions as:

The sum of the mixture ratios is equal to one.

$$(\sigma_x^2)_1 < (\sigma_x^2)_2 < \dots < (\sigma_x^2)_J \text{ and } (\sigma_y^2)_1 < (\sigma_y^2)_2 < \dots < (\sigma_y^2)_K$$

Table 7.1 : Estimated sea clutter parameters.

		α_1	α_2	α_3	σ_1^2	σ_2^2	σ_3^2	Number Iteration
	Initial Values	0.333	0.333	0.333	1	2	3	
Gradient	High Sea State	0.48	0.4	0.12	0.36	1.27	2.28	134
	Low Sea State	0.19	0.49	0.32	1.87	3.2	5.89	
FS	High Sea State	0.54	0.31	0.15	0.32	1.74	2.19	7
	Low Sea State	0.12	0.47	0.41	1.95	3.95	4.9	

The estimation results obtained using Gradient based and FS methods are similar but Gradient based method needs more iteration than the FS method. The histogram of the real data at low and high sea states and the “fit” of our solution to the real data

$p(r)$ is illustrated in Figure 7.4 and Figure 7.5 using Gradient descent and FS optimization, respectively.

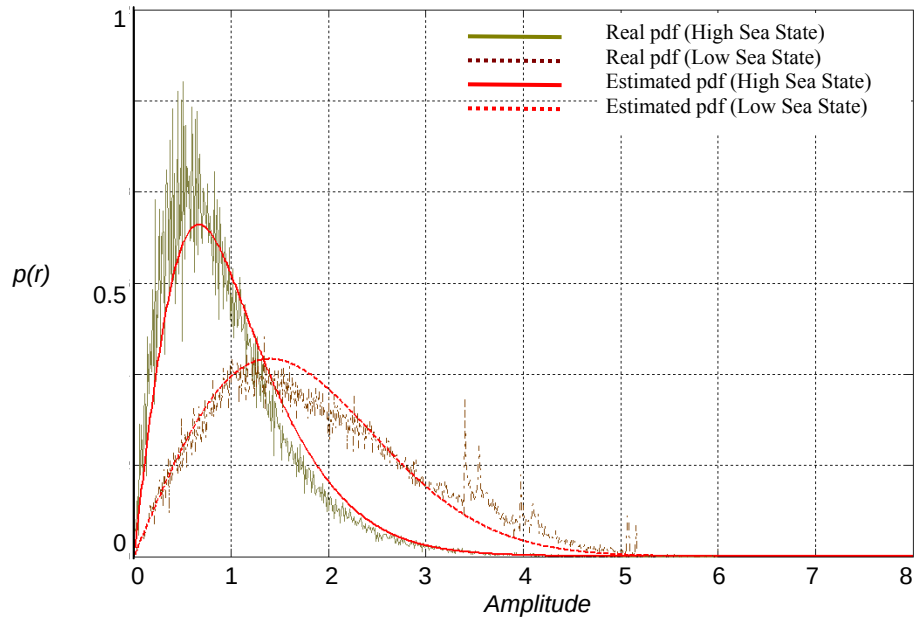


Figure 7-4: Estimated and real pdfs at high and low sea states using Gradient descent method.

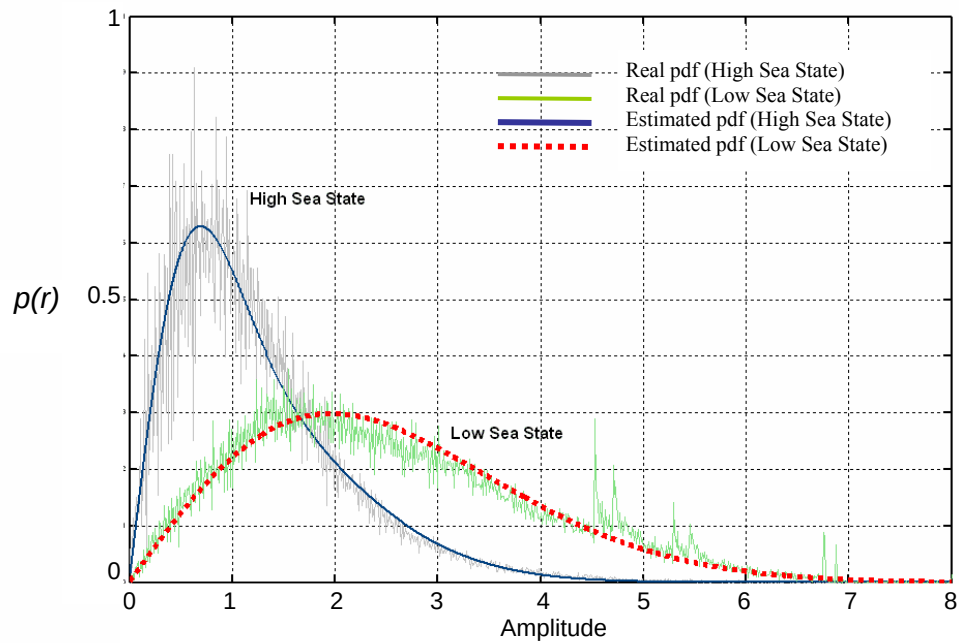


Figure 7-5: Estimated and real pdfs at high and low sea states using FS method.

To assess the fitting of our model to the real data, we plot the Cumulative Distribution Functions (CDFs) of the real data and CDF of our model and illustrate those in Figure 7.6 for Gradient descent. For FS optimization, CDF of the real data and the estimated CDF, are given in Figure 7.7.

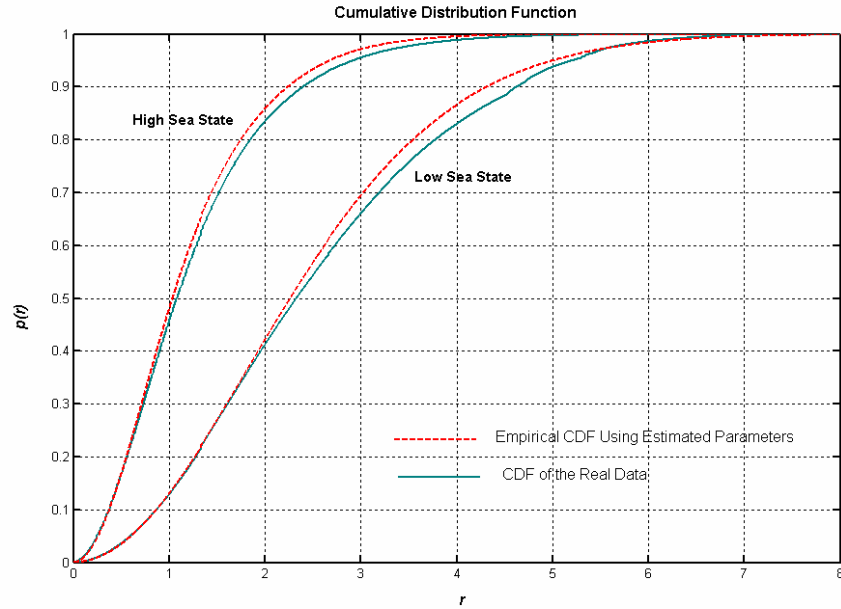


Figure 7-6: Estimated and real CDFs at a high sea state using Gradient descent method

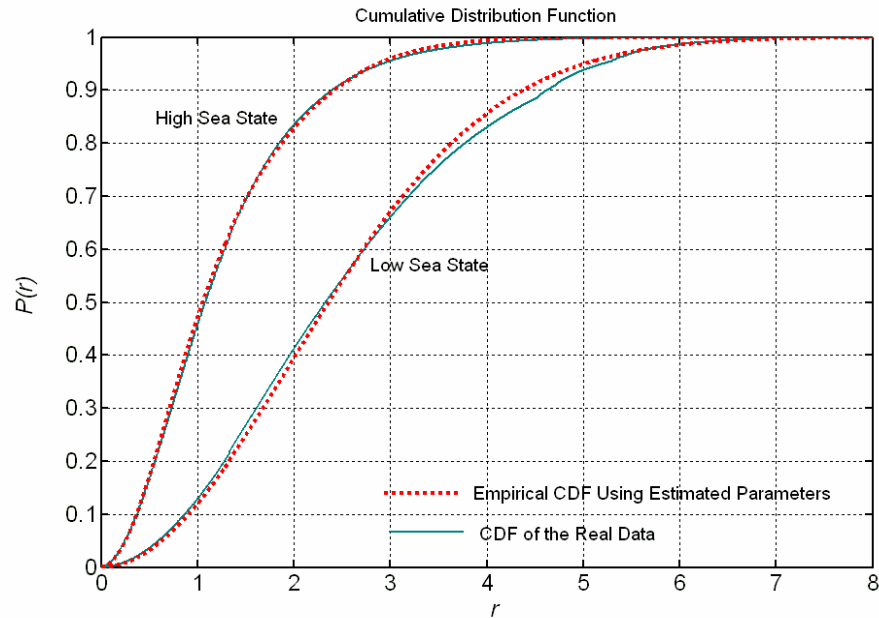


Figure 7-7: Estimated and real CDFs using FS method using FS method.

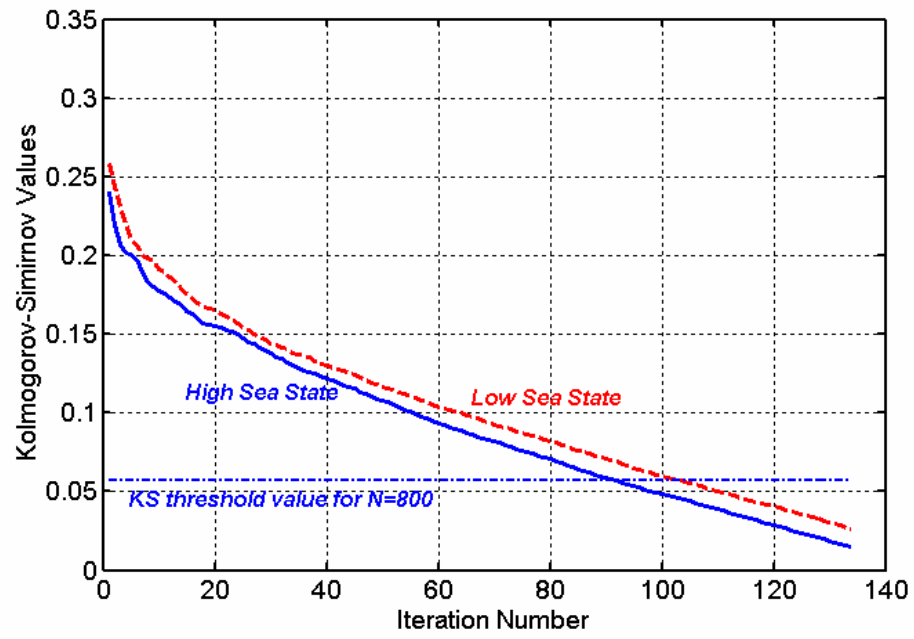
Kolmogorov-Smirnov (KS) goodness of fit test and Kullback Leibler (KL) distance are performed to check if the empirical distribution and the theoretical distribution are close enough. A smaller value of the KS statistic indicates a better fit of the particular distribution to the empirical data. For example, for a significance level 0.01, the KS value should be less than 0.0576 for a sample size of 800 [266] . The KL distance is a natural distance function from a true probability distribution $p = \{p_1, p_2, \dots, p_n\}$, to a target probability distribution $q = \{q_1, q_2, \dots, q_n\}$; and is defined to be

$$KL(p\|q) = \sum_{i=1}^m p_i \ln \left(\frac{p_i}{q_i} \right) \quad (7.22)$$

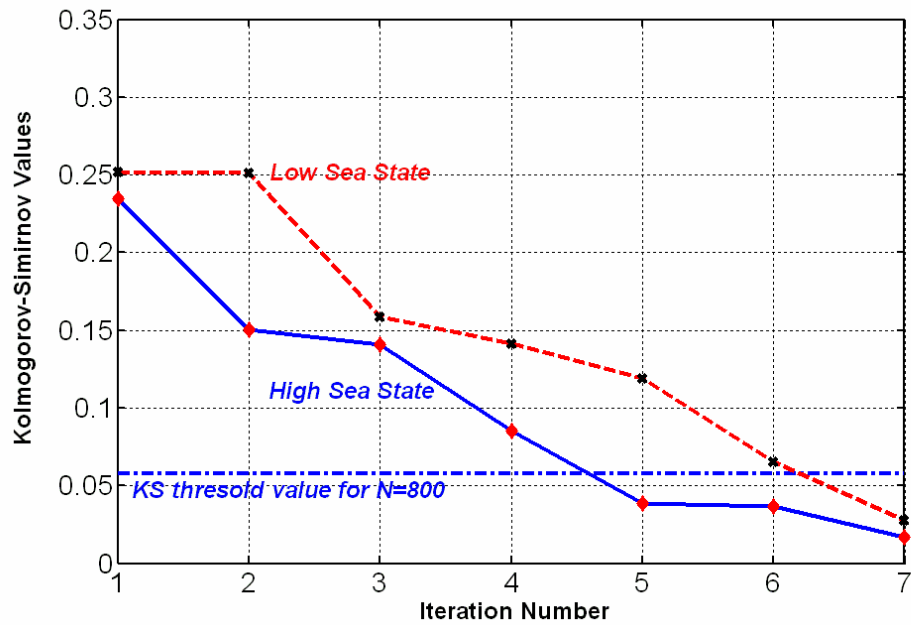
where m is the number of levels of the variables [267] . At the end of the iterations of FS algorithm, KL distance and KS test results are given in Table 7.2. for low and high sea states. KS and KL plots at each iteration are given in Figure 7.8 and 7.9, respectively.

Table 7.2 : KS test results and KL distance

	Sea States	KS	KL
Gradient	Low	0,0213	0,0409
	High	0,0174	0,0453
FS	Low	0,0281	0,0387
	High	0.0167	0,0460

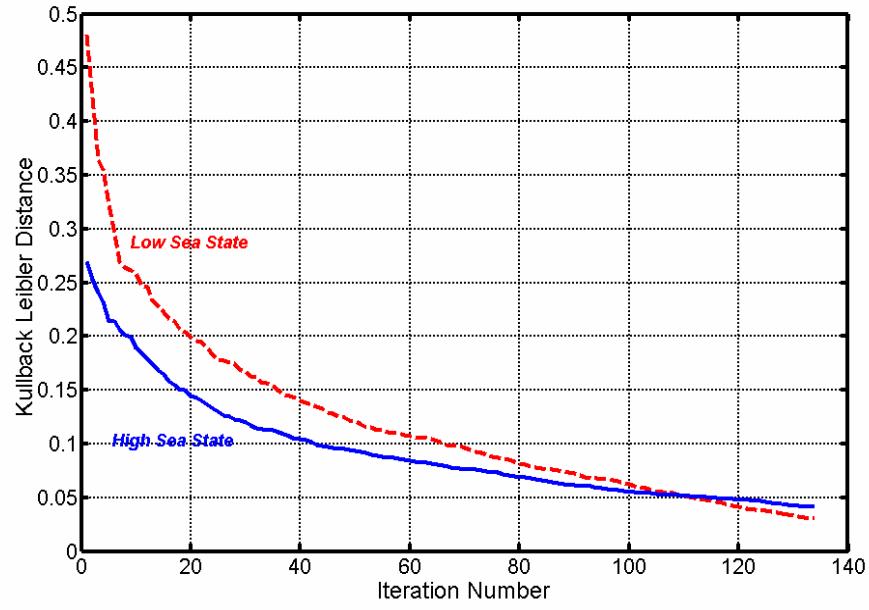


(a)

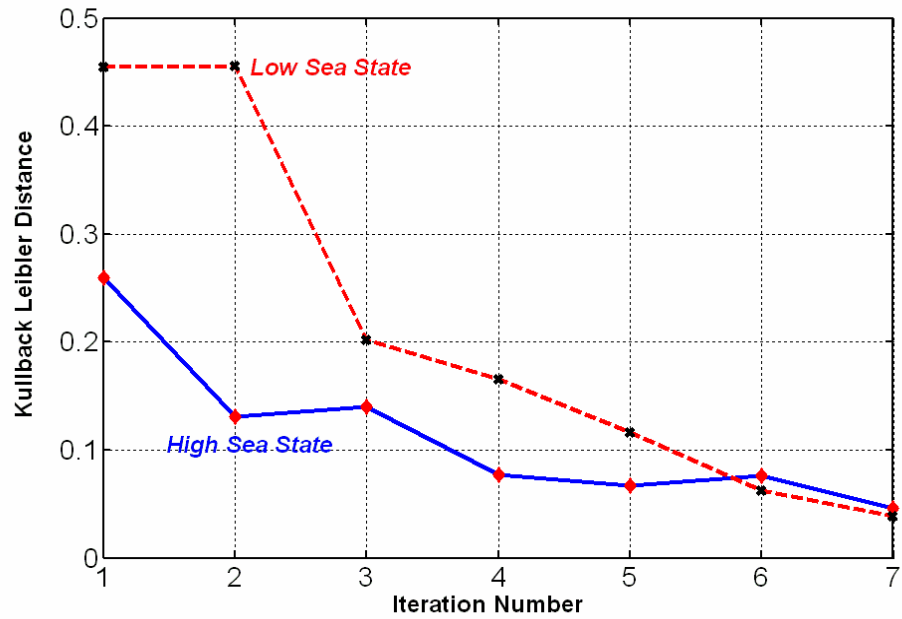


(b)

Figure 7-8 : KS values at each iteration (a) with Gradient based iteration (b) with FS.



(a)



(b)

Figure 7-9: KL values at each iteration (a) with Gradient based iteration (b) with FS.

We calculate the probability of false alarm values versus threshold (T) using the estimated parameters and we plot those in Figure 7.10. for Gradient based and FS. As expected, P_{fa} is monotonically decreasing with T .

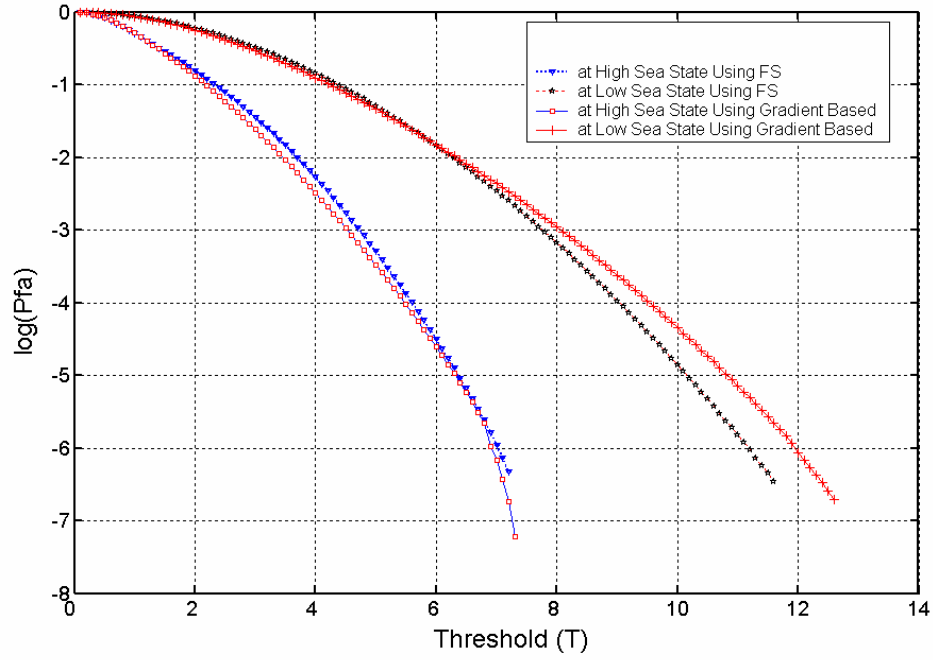


Figure 7-10 : P_{fa} versus T using estimated parameters at high and low sea states.

7.8. Concluding Remarks

We describe a new generalized mixture distribution to model sea clutter for different background statistics. The input statistics of I and Q channels of the envelope detector are assumed to be a GM. Despite some similarities, the output of the envelope detector has a pdf that does not resemble Rician Mixtures. Multi-component GM distributions at envelope detector inputs are very convenient for modeling different types of sea clutter.

The parameters of the background statistics have been learned using a ML algorithm with Gradient based and FS optimization. For CFAR processing, the background sample size M can be of different values (like 20,50,100,500,...). The CFAR loss is related to the variance of the estimated parameters, whose reduction requires a decrease of the CFAR loss. ML method gives better estimates (estimation results) for small sample sizes in comparison to the other methods. A very good fit to the IPIX radar data at high and low sea states, has been found using a GM with 3 components. It can be seen from experimental results that the empirical pdf from estimated values

fits the real data histograms very well. FS optimization reaches the acceptable KS values while Gradient descent method needs much more steps. Additionally, it is possible to avoid complicated Hessian equations by using only the gradient values in FS optimization.

Note that while the proposed method is computationally expensive, modern computers are able to handle the additional processing effort. The computers with more computational power may lead to more widespread use of the GM distribution in radar practice.

8. CONCLUSIONS and SUGGESTIONS for FUTURE WORK

The aim of this research is to investigate learning of unknown parameters from given linear Gaussian observation. We have developed four new approaches for image and blur parameter learning and restoration, simultaneously. The main idea has also been applied successfully, to the radar detection and sea clutter parameter learning problem.

In the first approach, ML blur and image parameter identification problem has been solved using FS method, which is based on Newton type optimization. In the second approach, convergence properties of the FS have been investigated and a new TRFS method, which ensures convergence in every iteration, has been given. In the third approach, the EM learning of unknown parameters has been given for recursive image problem in both closed and optimization form. Finally, in the fourth approach, the same EM learning problem has been examined for batch image processing and an accelerated EM method has been presented.

Sea clutter learning and adaptive radar detection problem can be treated similar to blur identification and image restoration problem. Sea clutter parameters are learned from observation using direct ML estimation method.

For future work, we will also point out some research area in this section. Real-world applications of image restoration techniques do not fit into the linear degradation model. Actually, many degradation mechanisms are of highly nonlinear form. Kalman filtering problem has been used in a linear model for recursive image processing applications to learn unknown parameters. If, however, the model is nonlinear, the new filter structure is referred to as the Extended Kalman Filter (EKF). Freitas *et al.* [268] [269] showed that the EM algorithm can be used for nonlinear system identification. An EM algorithm for nonlinear state space models was developed using Extended Kalman Smoothing (EKS) [270] [271] [272] . The EM

algorithm can be applied to nonlinear (state space) unknown image parameter learning for restoration problem as future work.

One of the major limitations of learning procedure on image processing applications is that it needs a batch algorithm, which is not feasible especially for big size images and for real time applications. On-line learning is a hot topic and should be investigated for blind image restoration.

Another area worthy of further research is parameter learning at GM environment. Generally, the image and observation models are based on Gaussian assumption, which is not the case in real life. Multi-component GM structure is very suitable to express different types of pdf in a single structure. GM models are flexible and powerful modeling tools for multivariate data, providing promising alternatives for the processing of nonlinear signals. For further research, learning of unknown parameters for image restoration can be studied for Gaussian mixture case.

In our image restoration problem, image model parameter matrix A and blur parameter matrix D have sparse property. Sparseness for a matrix is a measure of the degree to which the majority of the elements in such a matrix are zero. Sparse property can be utilized to take some advantage of the large number of zero elements and their locations.

Another important research subject is to find the global minimum. Different initial conditions lead to different solutions corresponding to different local minima. Avoiding dependence on initial conditions is very important. Some algorithms using Simulated Annealing (SA) [273] have also been able to provide global minimization. The SA algorithm is an efficient technique to find the minimum energy configuration of the system [274], but is very expensive. While SA programming avoids poor local minima, but it is still too slow for most problems. In these situations, Deterministic Annealing (DA) is an attractive choice. DA is designed to try to find approximate global minima to the objective, and thus robustify the problem to initial conditions [275] [276]. Ueda and Nakano [277] have proposed Deterministic Annealing EM (DAEM) algorithm for ML estimation problems to overcome a local minima problem associated with the conventional EM algorithm. The DAEM algorithm can

be applied to blind image restoration problem to find unknown image and blur parameters.

9. REFERENCES

- [1] **Castleman, K. R.**, 1979. *Digital Image Processing*, Prentice-Hall, Inc. New Jersey.
- [2] **Cichocki, A., Amari, S.**, 2002, Adaptive Blind Signal and Image Processing, John Wiley&Sons, Ltd, NY.
- [3] **Sezan, M.I., Tekalp, A.M.**, 1990, Image Restoration and Reconstruction, *Optical Engineering*, **29(5)**, 391-404.
- [4] **Molina, R., Núñez, J., Cotijo, F. J., Mateos, J.**, 2001. Image Restoration in Astronomy: A Bayesian Perspective, *IEEE Signal Processing Magazine*, **18(2)**, 11-29.
- [5] **Likas, A.C., Galatsanos, N. P.**, 2004, A Variational Approach for Bayesian Blind Image Deconvolution, *IEEE Transactions on Signal Processing*, **52(8)**, 2222-2233.
- [6] **Segall, C. A, Molina, R., Katsaggelos, A. K.**, 2003. High-Resolution Images from Low-Resolution Compressed Video, *IEEE Signal Processing Magazine*, **20(3)**, 37-48.
- [7] **Ng, M.K. and Bose, N. K.**, 2003. Mathematical Analysis of Super-Resolution Methodology, *IEEE Signal Processing Magazine*, **20(3)**, 62-74.
- [8] **Lamm, P.K.**, 2001. Some Recent Developments and Open Problems in Solution Methods for Mathematical Inverse Problems, *Proceedings XXIV National Congress of Applied and Computational Mathematics*, Belo Horizonte, Brazil, September 10-13.
- [9] **Andrews, H.C. and Hunt, B.R.**, 1977. *Digital Image Restoration*, Englewood Cliffs Prentice-Hall, New Jersey.
- [10] **Legendijk, R.L., Biemond, J.**, 1991. *Iterative Identification and Restoration of Images*, Kluwer Academic Publishers.
- [11] **Banham, M.R., Katsaggelos, A. K.**, 1997. Digital Image Restoration, *IEEE Signal Processing Magazine*, **14(2)**, 24-41.
- [12] **Söderström, T., Stoica, P.**, 1989. *System Identification*, Prentice Hall International Ltd., New Jersey.
- [13] **Murata N., Amari, S.**, 1999. Statistical Analysis of Learning, *Signal Processing*, **74(1)**, 3-28.

- [14] **Russell, S., Norvig, P.**, 1995. *Artificial Intelligence: A Modern Approach*, Prentice-Hall, Inc. New Jersey.
- [15] **Cichocki, A., Unbehauen, R.**, 1993. *Neural Networks for Optimization and Signal Processing*, Wiley, Chichester.
- [16] **Ripley, B.D.**, 1996. *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge.
- [17] **Rumelhart, D., McClelland, J.L.**, 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, PDP Research Group, MIT Press, Cambridge, MA.
- [18] **Saad, D.**, 1999. *On-Line Learning in Neural Networks*, Publications of the Newton Institute.
- [19] **Tenney, R.R., Hebbert, R.S., Sandell, N.S.**, 1977. A Tracking Filter for Maneuvering Sources, *IEEE Trans. On Automatic Control*, **22(2)**, 246-251.
- [20] **Duda, R.O., Hart, P.E., Stork, D.G.**, 2000. *Pattern Classification*. Second Edition. John Wiley and Sons.
- [21] **Mitchell, T.**, 1997. *Machine Learning*. Mc Graw Hill.
- [22] **North, B., Andrew, B., Isard, M., Rittscher, J.**, 2000. Learning and Classification of Complex Dynamics, *IEEE Transactions On Pattern Analysis and Machine Intelligence*, **22 (9)**, 1016-1034.
- [23] **Briegel, T., Tresp, V.**, 2002. A Nonlinear State Space Model for the Blood Glucose Metabolism of a Diabetic, *Automatisierungstechnik*, **50(5)**, 228-236.
- [24] **Li, X. R., Bar-Shalom, Y.**, 1994. A Recursive Multiple Model Approach to Noise Identification, *IEEE Trans on Aerospace and Electronic Systems*, **30(3)**, 671-684.
- [25] **Sarı, N., Sarı, F., Paker, S., Özden, M.T., Kayran, A.H.**, 2000. Radar Imaging Using 2-D Autoregressive Lattice Modelling, *Proceedings of the IEEE Signal Processing and Communications Applications Conference (SIU 2000)*, Antalya, Turkey.
- [26] **Sarı, F., Sarı, N., Mili, L.**, 2004. Modelling of Sea Clutter with Gaussian Mixtures and Estimation of the Clutter Parameter, *Proceedings of the IEEE Signal Processing and Communications Applications Conference (SIU 2004)*, Kuşadası, Turkey, April 28-30, 53-56.
- [27] **Sarı, F., Sarı, N., Mili, L.**, 2004. Learning of the Sea Clutter Parameters for Adaptive Detection, *Proceedings of International Conference on Radar Systems RADAR 2004*, Toulouse, France, October 18-22.

- [28] **Sarı, F.**, 2002. EM Based Identification of Non-Gaussian Sea Clutter and Adaptive Detection, (In Turkish) *Proc. of the IEEE 10th Signal Processing and Communication Applications Conference (SIU)*, 2002, Pamukkale.
- [29] <http://soma.ece.mcmaster.ca/ipix/>
- [30] **Sarı, F., Sar, N., Mili, L.**, 2005. Fast Maximum-Likelihood Sea Clutter Parameter Learning From the Output of the Envelope Detector, *Proc. of the 13th European Signal Processing Conference (EUSIPCO-2005)*, Antalya, Turkey, Sept.4-8.
- [31] **Kundur, D., Hatzinakos, D.**, 1996. Blind Image Deconvolution, *IEEE Signal Processing Magazine*, 13(6), 43-64.
- [32] **Kundur, D., Hatzinakos, D.**, 1998. A Novel Blind Deconvolution Scheme for Image Restoration Using Recursive Filtering, *IEEE Trans. On Signal Processing*, **46(2)**, 375-390.
- [33] **H. Kaufman, A.M. Tekalp**, 1991. Survey of Estimation Techniques in Image Restoration , *IEEE Control Systems Magazine*, **11**, 16-24.
- [34] **Jain, A. K.**, 1981. Advances in Mathematical Models for Image Processing, *Proceedings of the IEEE*, **69**, 502-528.
- [35] **Jain, A. K.**, 1989. *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs.
- [36] **Gelb, A., Kasper, J. F., Nash, R. A., Sutherland, C. A.**, 1974. *Applied Optimal Estimation*, The Analytic Sciences Corporation, The M.I.T. Press.
- [37] **Ljung, L.**, 1999. *System Identification: Theory for the User*, Prentice Hall.
- [38] **Woods, J.W. and Ingle, V.K.**, 1981. Kalman filtering in Two Dimensions: Further Results, *IEEE Trans. on ASSP*, **29**, 188-197.
- [39] **Rabiner, L.R., A** 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, **77(2)**, 257-286.
- [40] **Vageshi, S. V.**, 2000. *Advanced Digital Signal Processing and Noise Reduction*, Second Edition, John Wiley & Sons, Ltd.
- [41] **Ghahramani, Z.**, 1988. Learning Dynamic Bayesian Networks, *Adaptive Processing of Sequences and Data Structures Lecture Notes in Artificial Intelligence*, Springer Verlag, 168-197.
- [42] **Murphy, K.**, 2002. Dynamic Bayesian Networks: Representation, Inference and Learning, Ph.D. Thesis, UC Berkeley, Computer Science Division.

- [43] **Myers, J. W., Laskey, K. B., DeJong, K. A.,** 1999. Learning Bayesian Networks from Incomplete Data using Evolutionary Algorithms, *Processing of Genetic and Evolutionary Computation Conference (GECCO 99)*, Orlando, Florida, USA, July 13-17, 1999.
- [44] **Friedman N.,** 1998. The Bayesian structural EM algorithm. *Processing of the 14th International Conference on Uncertainty in Artificial Intelligence*, San Francisco: Morgan Kaufmann Publishers, 129-138.
- [45] **Penny, W.D., Roberts, S.J.,** 1998. Dynamic Linear Models, Recursive Least Squares and Steepest-Descent Learning, Neural Systems Research Group Technical Report TR-98-13, Department of Electrical Engineering, Imperial College.
- [46] **Bentler, P. M., Liang, J. J.,** 2003. Two-Level Mean and Covariance Structures: Maximum Likelihood via an EM Algorithm, In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications*, 53-70. Mahwah, NJ: Erlbaum.
- [47] **Combettes, P. L., Pesquet, J. C.,** 2004. Image Restoration Subject to a Total Variation Constraint, *IEEE Transactions On Image Processing*, **13(9)**, 1213-1222.
- [48] **Polak E.,** 1997. Optimization: Algorithms and Consistent Approximations, *Applied Mathematical Sciences*, **124**, Springer Verlag, New York.
- [49] **Jorge, N., Wright, S.J.,** 1999. *Numerical Optimization*, Springer Series in Operations Research, New York.
- [50] **Wilkinson, J. H.,** 1965. *The Algebraic Eigenvalue Problem*, Oxford University Press.
- [51] **Fletcher, R.** 1993. *An Overview of Unconstrained Optimization*, NATO ASI on Algorithms for Continuous Optimization, Il Ciocco, Italy, 5-18 September.
- [52] **Luenberger, D.G.,** 1984. *Linear and Nonlinear Programming*, Second Edition, Addison-Wesley, Massachusetts.
- [53] **Nocedal, J., Yuan, Y.,** 1998. Combining Trust Region and Line Search Techniques, In Y. Yuan (Ed.) *Advances in Nonlinear Programming*, 153-175.
- [54] **Gill, P.E., Murry, W. and Wright, M.H.,** 1981. *Practical Optimization*, Academic Press, New York.
- [55] **Nocedal, J.,** 1992. Theory of Algorithms for Unconstrained Optimization, *Acta Numerica*, **1**, 199-242.
- [56] **Nocedal, J.,** 1997. Large Scale Unconstrained Optimization, In *The State of the Art in Numerical Analysis*, A. Watson and I. Duff., (Eds), Oxford University Press, 311-338.

- [57] **Cappé, O., Bochoux, V., Moulines E.**, 1998, Quasi Newton Method for ML Estimation of Hidden Markov Model, In *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Vol. IV, Seattle, May, 2265-2268.
- [58] **Battiti, R.**, 1992. First and Second Order Methods for Learning: Between Steepest Descent and Newton's Method, *Neural Computation*, **4**, 141-166.
- [59] **Kay, S. M.**, 1993. Fundamentals of Statistical Signal Processing: Estimation Theory, Printice Hall, New Jersey.
- [60] **Lukšan L., Spedicato E.**, 2000. Variable metric methods for unconstrained optimization and nonlinear least squares, *J. Comput. Appl. Math.* **124**, 61-95.
- [61] **Saad, Y.**, 1996. *Iterative Methods for Sparse Linear Systems*, PWS Publication Co. Boston.
- [62] **Vogel, C.R.**, 2000, A Limited Memory BFGS Method for an Inverse Problem in Atmospheric Imaging, Methods and Applications of Inversion, (Ed.) *P.C. Hansen, B.H. Jacobsan, K. Mosegaard, Lecture Notes in Earth Sciences*, **92**, Springer-Verlag, 292-304.
- [63] **Fessler, J.A.**, 1997, Preconditioning Methods for Shift Invariant Image Reconstruction, *Proceedings International Conference on Image Processing (ICIP '97)*, Washington, DC, USA, October 26-29, 185-188.
- [64] **Kamm, J., Nagy, J.G.**, 1998. Kronecker Product and SVD Approximations in Image Restoration, *Linear Algebra Appl.*, **284**, 177-192.
- [65] **Levenberg, K.**, 1944. A Method For The Solution of Certain Nonlinear Problems in Least Squares, *Quart. Appl. Math.*, **2**, 164-168.
- [66] **Marquardt, D. W.**, 1963. An Algorithm for Least Squares Estimation of Nonlinear Parameters, *SIAM Journal of Applied Math.*, **11**, 431-441.
- [67] **Goldfeld, S.M., Quant, R.E., Trotter, H.F.**, 1966. Maximization by Quadratic Hill-Climbing, *Econometrica*, **34**, 541-551.
- [68] **Fortin, C., Wolkowicz, H.**, 2004. The Trust Region Subproblem and Semidefinite Programming, *Optimization Methods and Software*, **19(1)**, 41-67.
- [69] **Yuan, Ya-xiang**, 2000. A Review of Trust Region Algorithms for Optimization, *Proceedings of the Fourth International Congress on Industrial and Applied Mathematics*, In J. M. Ball, J.C.R. Hunt, (Eds.), Oxford University Press, 271-282.
- [70] **Byrd, R.H., Nocedal, J., Zhu, L.**, 1996. Towards a Discrete Newton Method with Memory for Large-Scale Optimization, In *Nonlinear*

Optimization and Applications. G. Di Pillo and F. Giannessi (Eds.), Plenum: New York.

- [71] **Fletcher, R.**, 1987. *Practical Methods of Optimization*, John Wiley & Sons.
- [72] **Finschi, L.**, 1996. An Implementation of the Levenberg-Marquardt Algorithm, Technical Report, *Institut for Operations Research, Eidgenossische Technische Hochschule*, Zurich, April 16.
- [73] **Mendel, J. M.**, 1990. *Maximum-likelihood Deconvolution A Journey into Model-Based Signal Processing*, Springer-Verlag, New York.
- [74] **Zhang, J-L.**, 2003. On The Convergence Properties of The Levenberg-Marquardt Method, *Optimization, Taylor & Francis*, **52(6)**, 739 – 756.
- [75] **Dan, H., Yamashita, N., Fukushima, M.**, 2001. Convergence Properties of the Inexact Levenberg-Marquardt Method Under Local Error Bound Conditions, Technical Report 2001-003, Department of Applied Mathematics and Physics, Kyoto University.
- [76] **Yamashita N., Fukushima, M.**, 2001. On the Rate of Convergence of the Levenberg-Marquardt Method, *Computing*, **15**, 239-249.
- [77] **Fan, J., Yuan, Y.**, 2000. On the Convergence of a New Levenberg-Marquardt Method, Technical Report, State Key Laboratory of Scientific/Engineering Computing. Institute of Computational Mathematics and Scientific/Engineering Computing, CAS.
- [78] **Baum, L. E., Petrie, T., Soules, G. and Weiss, N.**, 1970. A maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, *The Annals of Mathematical Statistics*, **41**, 164-171.
- [79] **Dempster, A.P., Laird, N. M. and Rubin, D.B.**, 1977. Maximum Likelihood from Incomplete via the EM algorithm, *J. Roy. Stat. Soc.*, **39**, 1-38.
- [80] **Moon, T.K.**, 1996. The Expectation Maximization Algorithm, *IEEE Signal Processing Magazine*, **11**, 47-60.
- [81] **Frenkel L., Feder, M.**, 1999. Recursive Expectation-Maximization (EM) Algorithms for Time-Varying Parameters with Applications to Multiple Target Tracking, *IEEE Trans. On Signal Processing*, **47(2)**, 306-320.
- [82] **Vogel, C.**, 2002. *Computational Methods for Inverse Problems*, SIAM Series: Frontiers in Applied Mathematics.
- [83] **Tikhonov, A. N. and Arsenin, V. A.**, 1977. *Solutions of Ill-posed Problems*, Winston & Sons, Washington.

- [84] **Hanke, M.**, 2001. On Lanczos Based Methods for the Regularization of Discrete Ill-Posed Problems, *BIT*, **41**, 1008-1018.
- [85] **Neumaier, A.**, 1998. Solving Ill-Conditioned and Singular Linear Systems: A Tutorial on Regularization, *SIAM Review*, **40**, 636-666.
- [86] **Engl, H. W., Hanke, M. and Neumaier, A.**, 2000. *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht.
- [87] **Calvetti, D., Lewis, B. And Reichel, L.**, 2001. Krylov Subspace Iterative Methods for Nonsymmetric Discrete Ill-Posed Problems In Image Restoration, *Advanced Signal Processing Algorithms, Architectures, and Implementations XI*, Ed. F.T. Luk, Proceedings of the Society of Photo Optical Instrumentation Engineers (SPIE), 4474, 224-233, The International Society for Optical Engineering, Bellingham, WA.
- [88] **Kärkkäinen, T., Majava, K., Mäkelä, M.**, 2001. Comparison of Formulations and Solution Methods for Image Restoration Problems, *Inverse Problems*, **17**, 1977-1995.
- [89] **Kunisch, K., Zou, J.**, 1998. Iterative Choices of Regularization Parameters in Linear Inverse Problems, *Inverse Problems*, **14**, 1247-1264.
- [90] **Marroquin, J., Mitter, S., Poggio, T.**, 1987. Probabilistic Solution of Ill-Posed Problems in Computational Vision, *Journal of the American Statistical Association*, **82(397)**, 76-89.
- [91] **Calvetti, D., Reichel, L. and Zhang, Q.**, 1995. Iterative solution methods for ill-posed problems, in *Advanced Signal Processing Algorithms*, ed. F. T. Luk, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE), 2563, 338-347, The International Society for Optical Engineering, Bellingham, WA.
- [92] **Gray, R.M.**, 1971. Toeplitz and Circulant Matrices: A Review, Technical Report, No. 6502-1, Stanford University Information Systems Laboratory, Stanford, California.
- [93] **Gray, R.M.**, 1972. On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices, *IEEE Trans. Info. Theory*, **18**, 725-730.
- [94] **Hunt, B.R.**, 1971. A Martix Theory Proof of the Discrete Convolution Theorem, *IEEE Trans. Audio and Electroacoustics*, **AU-19(4)**, 285-288.
- [95] **Gonzalez, R. C., Woods, R.E.**, 2002. *Digital Image Processing*, Second Edition, Prentice Hall, Upper Saddle River, New Jersey.
- [96] **Mendel, J. M.**, 1995. *Lessons in Estimation Theory for Signal Processing, Communications and Control*, Prentice-Hall, New Jersey.
- [97] **De Schutter, J., De Geeter, J., Lefebvre, T., Bruyninx, H.**, 1999. Kalman Filters: A Tutorial, *Journal A*, **40(4)**, 52-59.

- [98] **Galatsanos, N. P. and Katsaggelos, A. K.**, 1992. Methods for Choosing the Regularization Parameter and Estimating the Noise Variance in Image Restoration and Their Relation, *IEEE Transactions on Image Processing*, **1**, 322-336.
- [99] **Hebert, T.J., Lu, K.**, 1995. Expectation Maximization Algorithms, Null Spaces and MAP Image Restoration, *IEEE Transactions on Image Processing*, **4**(8), 1084-1095.
- [100] **Bertero, M., Poggio, T. A., Torre, V.**, 1988. Ill-Posed Problems in Early Vision, *Proceedings of the IEEE*, **76**(8), 869-889.
- [101] **Thompson, A.M., Brown, J.C., Kay, J.W. and Titterton, D.M.**, 1991. A Study of Methods of Choosing the Smoothing Parameter in Image Restoration by Regularization, *IEEE Trans. On Pattern Analy. And Machine Int.*, **13**, 326-339.
- [102] **Sari, F., Çelebi, M.E.**, 1996. 3-D Cell Image Restoration from Optical Microscopy (In Turkish), *Proceedings of the 4th Signal Processing and Applications Conference*, April 1996, Antalya, 83-88.
- [103] **Lim, J. S.**, 1990. *Two Dimensional Signal and Image Processing*, Prentice-Hall, New Jersey.
- [104] **Biemond, J., Lagendijk, R. L., Mersereau, R. M.**, 1990. Iterative Methods for Image Deblurring, *Proceedings of the IEEE*, **78**, 856-883.
- [105] **Adorf, H.M.**, 1995. Hubble Space Telescope Image Restoration in its Fourth Year, *Inverse Problems*, **11**, 639-653.
- [106] **Dias, J. M. B., Silva, T. A. M., Leitão, J. M. N.**, 1998. Adaptive Restoration of Speckled SAR Images using a Compound Random Markov Field, *Proceedings of the International Conference on Image Processing (ICIP)*, **2**, 79-83.
- [107] **Rodriguez, D.**, 2002. Signal Algebra Operators for SAR PSF Hardware Computations, AIRSAR Earth Science and Applications Workshop, NASA Jet Propulsion Laboratory, Pasadena, California, March 2002.
- [108] **Krishnamurthi, V., Liu, Y., Bhattacharyya, S., Turner, J., Holmes, T.**, 1995. Blind Deconvolution of Fluorescence Micrographs by Maximum Likelihood Estimation, *Applied Optics*, **34**(29), 6633-6647.
- [109] **Haykin, S.**, 1994. *Blind Deconvolution*, Printice Hall.
- [110] **Likas, A., Galatsanos, N.P.**, 2003. A Variational Method for Bayesian Blind Image Deconvolution, *Proceedings of the IEEE Int. Conf. On Image Processing*, Barcelona, Spain,
- [111] **Nakagaki, R., Katsaggelos, A. K.**, 2003. A VQ-Based Blind Image Restoration Algorithm, *IEEE Transactions On Image Processing*, **12**(9), 1044-1053.

- [112] **Bronstein M. M., Bronstein A.M., Zibulevsky, M., Zeevi Y.Y.,** 2005, Blind Deconvolution of Images Using Optimal Sparse Representations, *IEEE Transactions on Image Processing*, **14(6)**, 726-736.
- [113] **Bronstein A. M., Bronstein M.M., Zibulevsky, M.,** 2005, Relative Optimization for Blind Deconvolution, *IEEE Transactions on Signal Processing*, **53(6)**, 2018-2026.
- [114] **Kaufman, H. et al.,** 1983. Estimation and Identification of Two Dimensional Images, *IEEE Trans. Aut. Cont.*, **28**, 715-750.
- [115] **Cannon, M.,** 1976. Blind Deconvolution of Spatially Invariant Image Blurs with Phase, *IEEE Trans. ASSP*, **24**, 58-63.
- [116] **Stockham, T.G., Cannon, T.M., Ingebretsen, R.B.,** 1975. Blind Deconvolution through Digital Signal Processing, *Proc. IEEE*, **64(4)**, 678-692.
- [117] **Tekalp, A.M. , Kaufman, H. and Woods, J.W.,** 1986. Identification of Image and Blur Parameters for the Restoration of Noncausal Blurs, *IEEE Transactions on Acoustics, Speech and Signal Processing*, **ASSP-34(4)**, 963-972.
- [118] **Sarı, N. and Kayran, A.H.,** 2005. 2D ARMA Parameter Identification Using A Hybrid Lattice Design, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Philadelphia, PA USA, 18-23 March.
- [119] **Sarı, N, Sarı, F., Kayran, A.H.,** 2005. Simultaneous Identification of Image and Blur Parameters with the 2-D Hybrid Lattice Structure, *Proc. Of the 13th IEEE Signal Processing and Communication Applications Conference (SIU'2005)*.
- [120] **Reeves, S. J. and Mersereau, R. M.,** 1992. Blur identification by the method of generalized cross-validation, *IEEE Transactions on Image Processing*, **1**, 301-311.
- [121] **Lertrattanapanich, S., Bose, N. K.,** 2002. High Resolution Image Formation From Low Resolution Frames Using Delaunay Triangulation, *IEEE Trans. Image Proc.*, **11(12)**, 1427-1441.
- [122] **Lagendijk, R.L, Katsaggelos, A.K. and Biemond, J.,** 1988. Iterative Identification and Restoration of Images, *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, 992-995.
- [123] **Lay, K.T. and Katsaggelos, A.K.,** 1990. Image Identification and Restoration Based on Expectation-Maximization Algorithm, *Optical Engineering*, **29(5)**, 436-445.

- [124] **Legendijk, R.L., Tekalp, A.M. and Biemond, J.,** 1990. Maximum Likelihood Image and Blur Identification : A Unifying Approach, *Optical Engineering*, **29(5)**, 422-435.
- [125] **Katsaggelos, A.K., Legendijk, R.L., Biemond, J.,** 1988. Constrained Iterative Identification and Restoration of Images, *Proc EUSIPCO* 1988, 1585-1588.
- [126] **Porat, B.,** 1994. *Digital Processing of Random Signals: Theory and Methods*, Prentice Hall, Inc.
- [127] **Nahi, N.E.,** 1969. *Estimation Theory and Applications*, John Wiley & Sons, Inc.
- [128] **Scharf L. L.,** 1991. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley Publishing Company, Inc.
- [129] **Bard, Y.,** 1970. Comparison of gradient methods for the solution of nonlinear parameter estimation problems, *SIAM Journal on Numerical Analysis*, **7**, 157-186.
- [130] **Legendijk, R.L., Biemond, J. and Boekee, D.E.,** 1990. Identification and Restoration of Noisy Blurred Images Using the Expectation-Maximization Algorithm, *IEEE Tans. Acoust. Speech Sig. Proc.*, **38(7)**, 1180-1191.
- [131] **Katsaggelos, A.K.,** 1991 (editor), *Digital Image Restoration*, Springer Series in Information Sciences, 23, Springer-Verlag.
- [132] **Titterington, D., Smith, A. And Markov, U.,** 1985. *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.
- [133] **Wu, C.,** 1983. On the Convergence Properties of the EM Algorithm, *Ann. Statist.*, **11(1)**, 95-103.
- [134] **Xu, L.,** 1997. Comparative Analysis on Convergence Rates of the EM Algorithm and Its Two Modifications for Gaussian Mixtures, *Neural Processing Letters*, **6**, 69-76.
- [135] **Kontaxakis, G., Tzanakos, G.,** 1993. Study of the Convergence Properties of the EM Algorithm- A New Stopping Rule, *IEEE Nuclear Science Symposium and Medical Imaging Conference*, Orlando, FL, USA, 2:1163-1165.
- [136] **Demoment, G.** 1989. Image Reconstruction and Restoration: Overview Of Common Estimation Structures and Problems, *IEEE Trans. ASSP*, **37**, 2024-2036.
- [137] **Molina, R., Katsagelos, A. K., Mateos, J.,** 1999. Bayesian and Regularization Methods for Hyperparameter Estimation in Image Restoration, *IEEE Trans on Image Proc.*, **8(2)**, 231-246.

- [138] **Djafari, A.M.**, 1993, On The Estimation of Hyperparameters in Bayesian Approach of Solving Inverse Problems, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 5, 27-30 April, 495 – 498.
- [139] **Youmaran, R., Adler, A.**, 2004. Combining Regularization Frameworks for Image Deblurring: Optimization of Combined Hyper-parameters, *Canadian Conference on Electrical and Computer Engineering (CCECE)*, Niagara Falls, Canada, 723-726.
- [140] **Monalakis, D.G., Ingle, V.K., Kogon, S. M.**, 2000. *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*, McGraw-Hill Companies, Inc.
- [141] **Betsekas, D.P.**, 1982. *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Academic Press.
- [142] **Gould N.I. M., Lucidi S., Roma M., Toint P. L.**, 2000. Exploiting Negative Curvature Directions in Linesearch Methods for Unconstrained Optimization, *Optimization Methods and Software*, **14**, 75-98.
- [143] **Radner, R.A. and Walker, H.F.**, 1984. Mixtures Densities, Maximum Likelihood, and the EM Algorithm, *SIAM Review*, **26**, 195-239.
- [144] **Yamamoto, T.**, 2000. Historical Developments in Convergence Analysis for Newton's and Newton-like Methods, *Journal of Computational and Applied Mathematics*, **124**, 1-23.
- [145] **Galántai, A.**, 2000. The Theory of Newton's method, *Journal of Computational and Applied Mathematics*, **124**, 25-44.
- [146] **Devidas M., George E. O.**, 1999. Monotonic Algorithms for Maximum Likelihood Estimation in Generalized Linear Models, *The Indian Journal of Statistics*, **61**, Series B(3), 382–396.
- [147] **Rojas, M.**, 1998. A Large-Scale Trust-Region Approach to the Regularization of Discrete Ill-Posed Problems, Ph.D. Thesis, Rice University, Houston, Texas.
- [148] **Osborne, M. R.**, 1987. Estimating Nonlinear Models by Maximum Likelihood for the Exponential Family, *SIAM J. Sci. Statist. Comput.* **8**, 446–456.
- [149] **Sagara, N., Fukushima, M.**, 2005. A Trust Region Method for Nonsmooth Convex Optimization, *Journal of Industrial and Management Optimization*, **1(2)**, 171-180.
- [150] **Liu, X., Yuan, Y.**, 2001. A Robust Trust Region Algorithm for Solving General Nonlinear Programming, *Journal of Comp. Math.*, **19**, 309-322.

- [151] **Dennis, J. E. and Vicente, L. N.**, 1997. On the Convergence Theory of Trust-Region-Based Algorithms for Equality-Constrained Optimization, *SIAM Journal on Optimization*, **7**, 527-550.
- [152] **Wang, Y. and Yuan, Y.**, 2005. Convergence and Regularity of Trust Region Methods for Nonlinear Ill-Posed Inverse Problems, *Inverse Problems*, **21**, pp. 821-838.
- [153] **Lin, C.J. and Moré, J.J.**, 1999. Newton's Method for Large Bound-Constrained Optimization Problems, *SIAM Journal on Optimization*, **9**, 1100-1127.
- [154] **Moré, J.J.**, 1988. Trust Regions and Projected gradients, Systems Modelling and Optimization, M.Iti and K. Yajima, eds., *Lecture Notes in Control and Inform. Sci.*, **113**, 1-13, Springer, New-York.
- [155] **Yuan, Y.**, 1999. Problems on convergence of unconstrained optimization algorithms, in: Y. Yuan, ed., Numerical Linear Algebra and Optimization (Science Press, Beijing, New York, 1999), pp. 95--107.
- [156] **El-Alem, M.**, 1995. A Robust Trust-Tegion Algorithm with a Non-monotonic Penalty Parameter Scheme for Constrained Optimization, *SIAM Journal on Optimization*, **5**, 348-378.
- [157] **Aitkin M. , Foxall R.**, 2003. Statistical Modelling of Artificial Neural Networks Using the Multi-layer Perceptron, *Statistics and Computing*, **13**, 227-239.
- [158] **Wolfe, P.**, 1972. On the Convergence of Gradient Methods Under Constraint, *IBM J. Res. Develop.*, 407-411.
- [159] **Gertz, E. M.**, 2004. A Quasi-Newton Trust-Region Method, *Mathematical Programming*, **100(3)**, 447-470.
- [160] **Vicente, L.N.**, 1996. A Comparison Between Line Searches and Trust Regions for Nonlinear Optimization, *Investigação Operacional*, **16**, 173-179.
- [161] **Wang, Y.F., Yuan, Y-X., Zhang, H-C.**, 2002. A Trust Region-CG Algorithm for Deblurring Problem in Atmospheric Image Reconstruction, *Science in China*, **45**, 731-740.
- [162] **Wang, Y.F., Yuan, Y.**, 2003. A Trust Region Method for Solving Distributed Parameter Identification Problems, *Journal of Computational Mathematics*, **21(6)**, 759-772.
- [163] **Rojas, M. and Steihaug, T.**, 2002. Large-Scale Optimization Tehniques for Nonnegative Image Restorations, *Proceedings of SPIE*, 4791, 237-242.

- [164] **Sarı, F.**, 2002. Newton Type Optimization for ML Blur Identification and Restoration”, *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, 4, 3716-3719.
- [165] **Sarı F., Çelebi M.E.**, 2004. A New Trust Region Fisher Scoring Optimization For Image and Blur Identification, 12th European Signal Processing Conference (EUSIPCO), September 6-10, Vienna, Austria, 505-508.
- [166] **Smyth, G. K.**, 2002. An Efficient Algorithm for REML in Heteroscedastic Regression, *Journal of Graphical and Computational Statistics*, **11(4)**, 836-847.
- [167] **Xu, L. and Jordan, M.I.**, 1996. On Convergence of the EM Algorithm for Gaussian Mixtures, *Neural Computation*, **8**, 129-151.
- [168] **Moon, T.K., Stirling, W. C.**, 2000. *Mathematical Methods and Algorithms for Signal Processing*, Printice Hall, New Jersey.
- [169] **Cramér, H.**, 1946. *Mathematical Methods of Statistics*, Princeton University Press.
- [170] **O’Sullivan, J. A., Blahut, R.E., Snyder, D.L.**, 1998. Information-Theoretic Image Formation, *IEEE Transactions on Information Theory*, **44(6)**, 2094-2123.
- [171] **Kenney, C.S., Manjunath, B.S., Zuliani, M., Hwer, G.A., Nevel, A.V.**, 2003. A Condition Number for Point Matching with Application to Registration and Postregistration Error Estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **25 (11)**, 1437-1454.
- [172] **Li, M. , Chen, X., Li, X., Ma, B., Vitányi, P.M.B.**, 2004. The Similarity Metric, *IEEE Trans. on Information Theory*, **50(12)**, 3250-3264.
- [173] **Smith, S.T.**, 2005. Covariance, Subspace, and Intrinsic Cramér–Rao Bounds, *IEEE Transactions on Signal Processing*, **53(5)**, 1610 – 1630.
- [174] **Hanke, M.**, 1997. A Regularizing Levenberg-Marquardt Scheme, with Application to Inverse Groundwater Filtration Problems, *Inverse Problems*, **13**, 79-95.
- [175] **Chan, L.W.**, 1996. Levenberg-Marquardt Learning and Regularization, *Progress in Neural Information Processing*, Springer-Verlag, 139-144.
- [176] **Dennis J.E., Schnabel R.B.**, 1983. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ.
- [177] **Wang, Y.F., Yuan, Y.**, 2002. On the Regularity of a Trust Region-CG Algorithm for Nonlinear Ill-Posed Problems, Proceedings of the Third

Asian Mathematical Conference, In T. Sunada, P.W. Sy and Y. Lo (Eds.), World Scientific, Singapore, 562—580.

- [178] **Bertsekas, D. P.**, 1999. *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific.
- [179] **Bortoletti, A., Di Fiore, C., Fanelli, S., Zellini, P.**, 2003. A New Class of Quasi-Newtonian Methods for Optimal Learning in MLP-Networks, *IEEE Transactions on Neural Networks*, **14(2)**, 263-273.
- [180] **Salakhutdinov, R., Roweis, S., Ghahramani, Z.**, 2003. Optimization with EM and Expectation-Conjugate-Gradient, *Proceedings of the International Conference on Machine Learning*, Washington DC, USA, 672-679.
- [181] **Titterton, D.M.**, 1984. Recursive Parameter Estimation Using Incomplete Data, *Journal of the Royal Statistical Society Series B*, **46(2)**, 257-267.
- [182] **Hanke, M., Nagy, J.G. and Vogel, C.**, 2000. Quasi-Newton Approach to Nonnegative Image Restorations, *Linear Algebra and its Applications*, **316**, 223-226.
- [183] **Arian, E. and Ta'san, S.**, 1996. Analysis of the Hessian for Aerodynamic Optimization: Inviscid Flow, NASA CR-198328 ICASE, TR-96-28.
- [184] **Mignolet, M.P., Spanos, P.D.**, 1992. Simulation of Homogeneous Two-Dimensional Random Fields: Part-I AR and ARMA Models, *Transactions of the ASME*, **59**, 260-269.
- [185] **Sarı, F. and Celebi, M.E.**, 2004. Recursive Learning of Image Parameters for Totally Blind Image Restoration, *Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference*, 28-30 April, 339 – 342.
- [186] **Sarı F. and Çelebi, M. E.**, 2005. EM Based Learning for Image Restoration: A Closed Form Solution, *Proceedings of the IEEE 13th Signal Processing and Communication Applications Conference (SIU)*.
- [187] **Sarı F. and Çelebi, M. E.**, 2003. Recursive Learning of Image Parameters and Restoration of Images Using EM Based Learning Algorithm”, *IJCI Proceedings of International Conference on Signal Processing*, ISSN 1304-2386, 2003.
- [188] **Woods, J.W. and Radevan, C.H.**, 1977. Kalman filtering in Two Dimensions, *IEEE Transactions on Information Theory*, **23**, 473-482.
- [189] **Citrin, S.; Azimi-Sadjadi, M.R.**, 1992. A Full-Plane Block Kalman Filter for Image Restoration, *IEEE Transactions on Image Processing*, **1(4)**, 488–495.

- [190] **Suresh, B.; Shenoi, B.**, 1981. New Results in Two-Dimensional Kalman Filtering with Applications to Image Restoration, *IEEE Transactions on Circuits and Systems*, **28(4)**, 307 – 319.
- [191] **Azimi-Sadjadi, M.; Wong, P.**, 1987. Two-Dimensional Block Kalman Filtering for Image Restoration, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **35(12)**, 1736 – 1749.
- [192] **Galatsanos, N.P.; Chin, R.T.**, 1991. Restoration of Color Images by Multichannel Kalman Filtering, *IEEE Transactions on Signal Processing*, **39(10)**, pp. 2237 – 2252.
- [193] **Tekalp, A.M.; Pavlovic, G.**, 1990. Multichannel Image Modeling and Kalman Filtering for Multispectral Image Restoration, *Signal Processing*, 19(3), pp. 221-232.
- [194] **Çelebi, M. E.**, 1990. *Robust Recursive Estimation with Applications to Image Processing*, Ph.D. Thesis, Polytechnic University.
- [195] **Çelebi, M. E., Kurz, L.** 1996. Robust Locally Optimal Filters: Kalman, and Bayesian Estimation Theory, *Information Sciences*, **92**, 1-32.
- [196] **Meinel, E. S.**, 1986. Origins of linear and Nonlinear Recursive Restoration Algorithms, *Optical Society of America A*, **3**, 787-799.
- [197] **Gupta N. K., Mehra, R.K.**, 1974. Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations, *IEEE Trans. on Automatic Control*, **AC-19(6)**.
- [198] **Bergboer, N.H., Verdult, V. And Verhagaen, M. H. G.**, 2002. An efficient Implementation of LTI State-Space Models by Local Gradient Search, *Proceedings of the 41st IEEE Conference on Decision and Control*, Las Vegas.
- [199] **Segal M. and Weinstein, E.**, 1989. A New Method for Evaluating the Log-Likelihood Gradient, the Hessian, and the Fisher Information Matrix for Linear Dynamic Systems, *IEEE Trans. Info. Theory*, **35**, 682-687.
- [200] **Charalambous, C.D. and Hibey, J.L.**, 2001. Exact filters for Newton-Raphson Parameter Estimation Algorithms for Continous-Time Partially Observed Stochastic Systems, *Systems and Control Letters*, 42, 101-115.
- [201] **Dethlefsen, C.**, 2003. Markov Random Field Extensions Using State Space Models, *Bayesian Statistics*, **7**, Oxford University Press.
- [202] **Chen, C. F.**, 1981. The EM Algorithm to the Multiple Indicators and Multiple Causes Model via the Estimation of the Latent Variable, *Journal of the American Statistical Association*, **76 (375)**, 704-408.

- [203] **Watson, M. W. , Engle, R. F.**, 1983. Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models, *Journal of Econometrics*, **23**, 385-400.
- [204] **Jafarian, H. Z., Pasupathy, S.**, 1999. EM-Based Recursive Estimation of Channel Parameters”, *IEEE Trans. On Communications*, **47(9)**, xx.
- [205] **Shumway, R.H., Stoffer, D.S.**, 1982. An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm, *Journal of Time Series Analysis*, **3(4)**, 253-264.
- [206] **Ghahramani, Z., Hinton, G.E.**, 1996. Parameter Estimation for Linear Dynamical Systems, Technical Report CRG-TR-96-2, Department of Computer Science, University of Toronto.
- [207] **Wu, L. S.Y., Pai, J. S., Hosking, J. R. M.**, 1996. An Algorithm for Estimating Parameters of State-Space Models, *Statistics&Probability Letters*, **28**, 99-106.
- [208] **Rabiner, L. R., Juang, B. H.**, 1986. An Introduction to Hidden Markov Models, *IEEE Acoustic Speech & Signal Processing Magazine*, **3(1)**, 4-16.
- [209] **Ghahramani, Z.**, 2001. An Introduction to Hidden Markov Models and Bayesian Networks, *International Journal of Pattern Recognition and Artificial Intelligence*, **15(1)**, 9-42.
- [210] **Smyth, P., Heckerman, D., and Jordan, M. I.**, 1997. Probabilistic Independence Networks for Hidden Markov Probability Models, *Neural Computation*, **9**, 227-269.
- [211] **Kailath, T.**, 1975. A Survey of Data Smoothing, *Automatica*, **11**, 109-111.
- [212] **Ljung, L., Kailath, T.**, 1976. A Unified Approach to Smoothing Formulas, *Automatica*, **12**, 147-157.
- [213] **Moheimani, S. O. R., Savkin, A. V., Petersen, I. R.**, 1998. Robust Filtering, Prediction, Smoothing, and Observability of Uncertain Systems, *IEEE Transactions On Circuits And Systems—I: Fundamental Theory And Applications*, **45(4)**.
- [214] **Rauch, H. E.**, 1963. Solutions to the linear smoothing problem, *IEEE Trans. On Automatic Control*, **8**, 371-372.
- [215] **Anderson, B.D.O. and Moore, J.B.**, 1979. *Optimal Filtering*, Printice Hall, New Jersey.
- [216] **Gannot, S., Burshtein, D. and Weinstein, E.**, 1998. Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms, *IEEE Trans. on Speech and Audio Processing*, **6**, 373-385.

- [217] **Boik, R. J.**, 2004. Montana State University, Department of Mathematical Sciences, Lecture Notes: Statistics 550 Spring 2004. http://www.math.montana.edu/~rjboik/classes/550/notes_550.pdf
- [218] **H. Neudecker**, 2003. On Two Matrix Derivatives by Kollo and von Rosen, *SORT* **27(2)**, 153-164, July-December 2003.
- [219] **D.A. Wilson, A. Kumar**, 1982. Derivative Computations for the Log Likelihood Fuction, *IEEE Trans. on Automatic Control*, **AC-27(1)**.
- [220] **Abromowitz, M., Stegun, I.A.**, 1965. *Handbook of Mathematical Functions*, Dover Pub., New York.
- [221] **Angwin, D. L. and Kaufman, H.**, 1989. Image Restoration Using Reduced Order Models, *Signal Processing*, **16**, 21-28.
- [222] **Lange, K.**, 1995. A Gradient Algorithm Locally Equivalent to the EM Algorithm, *Journal of Royal Statistics Soc. B*, **57(2)**, 425-437.
- [223] **Jamshidian, M. , Jennrich, R. I.**, 1997. Acceleration of the EM Algorithm by Using Quasi-Newton methods, *Journal of the Royal Society, Series B*, **49**.
- [224] **Meng X. L. and van Dyk D.** 1998. Fast EM-type implementations for mixed effects models. *J. of the Royal Statistical Society series B*, **60**,559–578.
- [225] **Louis, T. A.**, 1982. Finding the observed information matrix when using the em algorithm, *Journal of Royal Statistical Society, Series B*, vol. 44, pp. 226.233.
- [226] **Salakhutdinov, R. and Roweis, S.**, 2003. Adaptive Overrelaxed Bound Optimization Methods,. *International Conference on Machine Learning (ICML)*.
- [227] **Petersen, K. B. and Winther, O.** 2005. The EM Algorithm in Independent Component Analysis, *International Conference on Acoustics, Speech, and Signal Processing*.
- [228] **McLachlan, G.J., Krishnan, T.**, 1997. *The EM Algorithm and Extensions*, John Wiley & Sons, Inc. New York.
- [229] **Jamshidian, R.M, and Jennrich, R.I**, 1993. Conjugate Gradient Acceleration of the EM Algorithm, *Journal of the American Statistical Association*, **88**, 221-228.
- [230] **Ikeda, S.**, 2000. Acceleration of the EM algorithm, *Systems and Computers in Japan*, **31**, 10-18.
- [231] **J. Browne, A.R. De Pierro**, 1996. A Row-Action Alternative to the EM Algorithm for Maximizing Likelihood in Emission Topography, *IEEE Trans. Med. Imaging* , **15**, 687-699.

- [232] **Liu, C., Sun, D.X.**, 1997. Acceleration of EM algorithm for mixtures models using ECME, *ASA Proceedings of The Stat. Comp. Session*, 109-114.
- [233] **Sarı, F., Çelebi, M. E.**, 2004. A New Accelerated EM Based Learning of the Image Parameters and Restoration, *IEEE International Joint Conference on Neural Networks (IJCNN)*, Vol.4, 25-29 July 2004, 2513 – 2518.
- [234] **J. L. Goldberg**, 1992. *Matrix Theory with Applications*, McGraw-Hill.
- [235] **Park, C.**, 2005. Parameter Estimation of Incomplete Data in Competing Risks Using the EM Algorithm, *IEEE Trans. On Reliability*, **52(4)**, 282-290.
- [236] **Gould N.I. M., Lucidi S., Roma M., Toint P. L.**, 2000. Exploiting Negative Curvature Directions in Linesearch Methods for Unconstrained Optimization, *Optimization Methods and Software*, **14**, 75-98.
- [237] **Graham, A.**, 1981. *Kronecker Products and Matrix Calculus: with Applications*, Ellis Horwood Series in Mathematics and Its Applications.
- [238] **Shnidman D.A.**, 1999. Generalized Radar Clutter Model, *IEEE Trans. On Aerospace and Electronic Systems*, 35(3), 857-865.
- [239] **Noga, J. L.**, 1998. Bayesian State-Space Modelling of Spatio-Temporal Non-Gaussian Radar Returns, Ph.D. Thesis, Department of Engineering University of Cambridge.
- [240] **Gini F., Greco M.**, 2002. Covariance Matrix Estimation for CFAR Detection in Correlated Heavy Tailed Clutter, *Signal Processing*, **82**, 1847-1859.
- [241] **Rohling, H.**, 1983. Radar CFAR Thresholding in Clutter and Multiple Target Situations, *IEEE Trans. On Aerospace and Electronic Systems*, **19(4)**, 608-621.
- [242] **Weiss, M.**, 1982. Analysis of Some Modified Cell-Averaging CFAR Processors in Multiple-Target-Situations, *IEEE Trans. On Aerospace and Electronic Systems*, **18**, 102-114.
- [243] **Soltani, F., Barkat, M.**, 1997. CFAR Binary Integration Detection in Nonhomogeneous Partially Correlated Clutter, *IEE Proc -Radar. Sonar, Navig.*, **144(5)**, 293-300.
- [244] **Farrouki, A., Barkat, M.**, 2005. Automatic Censoring CFAR Detector Based on Ordered Data Variability for Nonmhomogenous Environments, *IEE Proc -Radar. Sonar, Navig.*, **152(1)**, 43-51.
- [245] **Pulsone, N. B., Raghavan, R S.**, 1999. Analysis of an Adaptive CFAR Detector in Non-Gaussian Interference, *IEEE Trans. On Aerospace and Electronic Systems*, **35(3)**, 903-916.

- [246] **Watts S., 2000.** The Performance of Cell-Averaging CFAR Systems in Sea Clutter, *Proc.of the IEEE International Radar Conference*, 7-12 May 2000, 398 – 403.
- [247] **Tsakalides, P., Raspanti, R., Nikias, C. L., 1999.** Angle/Doppler Estimation in Heavy-Tailed Clutter Backgrounds, *IEEE Trans. On Aerospace and Electronic Systems*, **35(2)**, 419-436.
- [248] **Haykin, S., Bakker, R., Currie, B., 2002.** Uncovering Nonlinear Dynamics: The Case Study of Sea Clutter, *Proc. IEEE*, **90(5)**,860-861,
- [249] **Haykin, S., 2001.** Signal Processing: Where Physics and Mathematics Meet, *IEEE Signal Processing Magazine*, July, 6-7.
- [250] **Haykin, S., 2002.** Adaptive Radar, *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, Orlando, Florida, USA, May 13 – 17.
- [251] **Blackman, S., Popoli, R.,** *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.
- [252] **Ravid, R., Levanon, N., 1992.** Maximum-Likelihood CFAR for Weibull Background, *IEE Proceedings-F*, **139(3)**, 256-264.
- [253] **Gigli, G., Lampropoulos G. A., 2002.** A New Maximum Likelihood Generalized Gamma CFAR Detector, *Proc. of the International IEEE Geoscience and Remote Sensing Symposium,. IGARSS '02.* Vol 6, 24-28 June 2002, pp. 3399–3401.
- [254] **Talukdar, K.K., Lawing, W.D., 1991.** Estimation of the Parameters of the Rice Distribution, *Journal of Acoust. Soc. Am.*, **89(3)**, 1193-1197.
- [255] **Anastassopoulos V., Lampropoulos G.A., 1995.** Optimal CFAR Detection in Weibull Clutter, *IEEE Trans. On Aerospace and Electronic Systems*, **31(1)**, 52-64.
- [256] **Trunk, G.V., 1976.** Non-Rayleigh Sea Clutter: Properties and Detection of Targets, *Naval Research Laboratory Report 7986*, Washington D.C.
- [257] **Conte, E., Longo, M., 1987.** Characterization of radar clutter as a spherically invariant random process, *IEE Proceedings, Pt. F*, **134(2)**, 191-197.
- [258] **Sangston, K. J., and Gerlach, K. R., 1994.** Coherent Detection of Radar Targets in a Non-Gaussian Background, *IEEE Trans. On Aerospace and Electronic Systems*, **30(2)**, 330-340.
- [259] **Sangston K. J., Gini F., Greco M. V., Farina A., 1999.** Structures for Radar Detection in Compound Gaussian Clutter, *IEEE Trans. On Aerospace and Electronic Systems*, **35(2)**, 445-458.
- [260] **Stein D. W. J., 1999.** A Robust Exponential Mixture Detector Applied to Radar, *IEEE Trans. On Aerospace and Electronic Systems*, **35(2)**, 519-532.

- [261] **Dogandžić A., Nehorai, A., Wang J.,** 2004. Maximum Likelihood Estimation of Compound-Gaussian Clutter And Target Parameters, Proc. 12th Annu. Workshop Adaptive Sensor Array Processing (ASAP 2004), 16-18 March 2004, Lincoln Lab., Lexington, MA.
- [262] **Erdélyi, A.,** 1954. *Tables of Integral Transforms*, Volume I-II, McGraw-Hill Book.
- [263] **Blum, R.S., Zhang, Y., Sadler, B.M., Kozick, R.J.,** 1999. On the Approximation of Correlated No-Gaussian Noise pdfs Using Gaussian Mixtures, *The 1st Conference on the Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, Washington D.C.
- [264] **Barton, D.K.,** 1988. *Modern Radar System Analysis*, Artech House.
- [265] **Evans, M., Hastings, N., Peacock, B.,** 2000. *Statistical Distributions*, 3rd ed., Wiley, New York.
- [266] **Conover, W.J.,** 1980. *Practical Nonparametric Statistics*, New York: Wiley,.
- [267] **Cover, T.M., Thomas, J.A.,** *Elements of Information Theory*, Wiley Interscience, New York, NY, 1991.
- [268] **Ghahramani Z., Roweis, S. T.,** 1999. Learning Nonlinear Dynamical Systems Using an EM Algorithm, (Eds) M.S. Kearns, S.A. Solla, D.A. Cohn, *Advances in Neural Information Processing Systems*, **11**, Cammbridge, MA: MIT Press.
- [269] **Freitas, J.F.G. , Niranjana, M., Gee, A. H.,** 2000. Dynamic Learning with the EM Algorithm for Neural Networks, *Journal of VLSI Signal Processing Systems*, Volume 26 , Issue 1-2, August. Pages: 119 – 131.
- [270] **Haykin, S.,** 2001. *Kalman Filtering and Neural Networks*, John Wiley & Sons.
- [271] **J.F.G. de Freitas, M. Niranjana and A. H. Gee,** 1998. Nonlinear State Space Learning with EM and Neural Networks, *IEEE Workshop on Neural Networks for Signal Processing (NNSP 98)*, (Eds.) T. Constantinides, S.Y. Kung, M. Niranjana, E. Wilson, 254-263.
- [272] **J.F.G. de Freitas, M. Niranjana and A. H. Gee,** 1998. The EM Algorithm and Neural Networks for Nonlinear State Space Estimation, Cambridge University Engineering Department Technical Report, CUED/F-INGENG/TR 313.
- [273] **Geman, S. and Geman, D.** 1984, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. PAMI*, **6**, 721-741.
- [274] **S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi,** 1983. *Science*, **220**, 671-680,

- [275] **M. Takada and R. Nakano**, 2002. Multi-Thread Search with Deterministic Annealing EM Algorithm, *Proc. Int. Joint Conf. on Neural Networks (IJCNN '02)*, LNCS 2189, 1039-1044.
- [276] **T. Kawai, R. Nakano** , 2004. Threshold Based Multi-Thread EM Algorithm”, *IEEE International Joint Conference on Neural Networks (IJCNN)*, 25-29 July 2004.
- [277] **N. Ueda and R. Nakano**. 1998. Deterministic Annealing EM Algorithm. *Neural Networks*, **11**, 271–282.

APPENDIX A : DERIVATION OF (4.5-4.6)

We will make use of the following identities,

$$\frac{\partial}{\partial \theta_i} [\log \det |P|] = \text{tr} \left(P^{-1} \frac{\partial P}{\partial \theta_i} \right) \quad (\text{A.1})$$

$$\frac{\partial P^{-1}}{\partial \theta_i} = -P^{-1} \frac{\partial P}{\partial \theta_i} P^{-1} \quad (\text{A.2})$$

where $E(y^T z) = \text{tr} [E(z y^T)]$ for y, z $N \times 1$ vectors and (A.2) have been used and $E[g g^T] = P$.

$$\begin{aligned} [I(\theta)]_{i,j} = & \text{tr} \left(P^{-1} \frac{\partial P}{\partial \theta_i} \right) \text{tr} \left(P^{-1} \frac{\partial P}{\partial \theta_j} \right) + \text{tr} \left(P^{-1} \frac{\partial P}{\partial \theta_i} \right) E \left\{ g^T \frac{\partial P^{-1}}{\partial \theta_j} g \right\} \\ & + E \left\{ g^T \frac{\partial P^{-1}}{\partial \theta_i} g \right\} \text{tr} \left(P^{-1} \frac{\partial P}{\partial \theta_j} \right) + E \left\{ g^T \frac{\partial P^{-1}}{\partial \theta_i} g g^T \frac{\partial P^{-1}}{\partial \theta_j} g \right\} \end{aligned} \quad (\text{A.3})$$

$$E \left\{ g^T \frac{\partial P^{-1}}{\partial \theta_i} g \right\} = -\text{tr} \left(P^{-1} \frac{\partial P}{\partial \theta_i} \right) \quad (\text{A.4})$$

$$E \left\{ g^T \frac{\partial P^{-1}}{\partial \theta_j} g \right\} = -\text{tr} \left(P^{-1} \frac{\partial P}{\partial \theta_j} \right) \quad (\text{A.5})$$

To evaluate the last term in (A.3), the following relation can be used.

$$E\left[y^T A y y^T B y\right]=tr\left(AC\right)tr\left(BC\right)+2tr\left(ACBC\right) \quad (\text{A.6})$$

where $C = E\left(y y^T\right)$ and A and B are symmetric matrices. Therefore the last term in (A.3) becomes,

$$tr\left(P^{-1} \frac{\partial P}{\partial \theta_i}\right) tr\left(P^{-1} \frac{\partial P}{\partial \theta_j}\right)+2tr\left(P^{-1} \frac{\partial P}{\partial \theta_i} P^{-1} \frac{\partial P}{\partial \theta_j}\right) \quad (\text{A.7})$$

and finally substituting (A7) in (A3) we obtain the following

$$\left[I(\theta)\right]_{i,j}=2tr\left(P^{-1} \frac{\partial P}{\partial \theta_i} P^{-1} \frac{\partial P}{\partial \theta_j}\right) \quad (\text{A.8})$$

In DFT domain, (A.8) takes the form given below

$$\left[I(\theta)\right]_{(i,j)}=2\sum_{u,v}\left\{\frac{1}{P^2(u,v)} \frac{\partial P(u,v)}{\partial \theta_i} \frac{\partial P(u,v)}{\partial \theta_j}\right\} \quad (\text{A.9})$$

APPENDIX B : RTS RECURSION

For $k=1,2,\dots,N$, forward filtering equations are the same as Kalman filtering equations defined in (3.14)-(3.18). For $k=N,N-1,\dots,1$, the backward smoothing equations make a backward pass through the data to calculate $\mu_{k-1|N}$ and $M_{k-1|N}$ as given in the following formulations.

$$\mu_{k-1|N} = \hat{f}_{k-1} + S_{k-1} \left[\mu_{k-1|N} - A\hat{f}_{k-1} \right] \quad (\text{B.1})$$

$$M_{k-1|N} = M_{k-1|k-1} - S_{k-1} \left[M_{k|N} - M_{k|k-1} \right] S_{k-1}^T \quad (\text{B.2})$$

where S_{k-1} is defined as

$$S_{k-1} = M_{k-1|k-1} A^T M_{k|k-1}^{-1} \quad . \quad (\text{B.3})$$

APPENDIX C: DERIVATION OF (5.14)

$$M_{k-1,k|N} = E_{\theta} \left\{ \left(f_{k-1} - \mu_{k-1|N} \right) \left(f_k - \mu_{k|N} \right)^T \right\} \quad (\text{C.1})$$

After the modification which is mentioned at section 5.2.1.1., we can write,

$$M_{k-1,k|N} = E_{\theta} \left\{ \left(f_{k-1} - \hat{f}_{k-1} \right) \left(f_k - \left(\hat{f}_{k|k-1} + K_k \left[g_k - D \hat{f}_{k|k-1} \right] \right) \right)^T \right\} \quad (\text{C.2})$$

If we substitute (2.10) and (2.11) into the (C-2), we obtain the following equation:

$$M_{k-1,k|N} = E_{\theta} \left\{ \tilde{f}_{k-1} \left(A \tilde{f}_{k-1} + v_{k-1} - K_k D A \tilde{f}_{k-1} - K_k D v_{k-1} - K_k w_k \right)^T \right\} \quad (\text{C.3})$$

Here, $\tilde{f}_{k-1} = f_{k-1} - \hat{f}_{k-1}$.

$$M_{k-1,k|N} = E_{\theta} \left\{ \tilde{f}_{k-1} \left(I - K_k D \right)^T \tilde{f}_{k-1}^T A^T + \tilde{f}_{k-1} v_{k-1}^T - \tilde{f}_{k-1} v_{k-1}^T D^T K_k^T - \tilde{f}_{k-1} w_k^T K_k^T \right\} \quad (\text{C.4})$$

$$M_{k-1,k|N} = E_{\theta} \left\{ \tilde{f}_{k-1} \tilde{f}_{k-1}^T \right\} \left(I - K_k D \right)^T A^T \quad (\text{C.5})$$

$$M_{k-1,k|N} = M_{k-1|k-1} \left(I - K_k D \right)^T A^T \quad (\text{C.6})$$

APPENDIX D: REQUIRED DERIVATIONS FOR (5.17-5.28)

$$\frac{\partial(R^{-1}D)}{\partial d(m,n)} = R^{-T} E_{m,n} \quad (\text{D.1})$$

$$\frac{\partial(Q^{-1}A)}{\partial a(k,l)} = Q^{-T} E_{k,l} \quad (\text{D.2})$$

$$\frac{\partial}{\partial d(m,n)}(D^T R^{-1} D) = E_{m,n}^T R^{-1} D + D^T R^{-1} E_{m,n} \quad (\text{D.3})$$

$$\frac{\partial}{\partial a(k,l)}(A^T Q^{-1} A) = E_{k,l}^T Q^{-1} A + A^T Q^{-1} E_{k,l} \quad (\text{D.4})$$

$$\frac{\partial Q^{-1}}{\partial \sigma_v^2} = -\frac{1}{(\sigma_v^2)^2} I \quad , \quad \frac{\partial R^{-1}}{\partial \sigma_w^2} = -\frac{1}{(\sigma_w^2)^2} I \quad (\text{D.5})$$

$$\frac{\partial Q}{\partial \sigma_v^2} = I \quad , \quad \frac{\partial R}{\partial \sigma_w^2} = I \quad (\text{D.6})$$

$$\frac{\partial(Q^{-1}A)}{\partial \sigma_v^2} = -\frac{1}{(\sigma_v^2)^2} A \quad , \quad \frac{\partial(R^{-1}D)}{\partial \sigma_w^2} = -\frac{1}{(\sigma_w^2)^2} D \quad (\text{D.7})$$

$$\frac{\partial}{\partial \sigma_v^2}(A^T Q^{-1} A) = -\frac{1}{(\sigma_v^2)^2} (A^T A) \quad (\text{D.8})$$

$$\frac{\partial}{\partial \sigma_w^2} (D^T R^{-1} D) = -\frac{1}{(\sigma_w^2)^2} (D^T D) \quad (\text{D.9})$$

APPENDIX E: THE DERIVATIVES OF $\Gamma(\theta)$ WITH RESPECT TO EACH UNKNOWN PARAMETERS

The derivatives of $\Gamma(\theta)$ with respect to each unknown parameters are given as:

$$\frac{\partial \Gamma(\theta)}{\partial d(m, n)} = \frac{1}{2} \text{tr} \left[\left(C_{f|g} + \mu_{f|g} \mu_{f|g}^T \right) \left(E_{mn}^T R^{-1} D + D^T R^{-1} E_{mn} \right) \right] - g^T R^{-1} E_{mn} \mu_{f|g} \quad (\text{E.1})$$

$$\begin{aligned} \frac{\partial \Gamma(\theta)}{\partial a(k, l)} = \frac{1}{2} \text{tr} \left\{ \left(C_{f|g} + \mu_{f|g} \mu_{f|g}^T \right) \frac{\partial}{\partial a(k, l)} \left[(I - A)^T Q^{-1} (I - A) \right] \right. \\ \left. - \left[(I - A)(I - A)^T \right]^{-1} \frac{\partial}{\partial a(k, l)} \left[(I - A)(I - A)^T \right] \right\} \end{aligned} \quad (\text{E.2})$$

$$\frac{\partial \Gamma(\theta)}{\partial \sigma_v^2} = \frac{\partial \Gamma(\theta)}{\partial Q(i, i)} = \frac{1}{2} \text{tr} \left\{ Q^{-1} \frac{\partial Q}{\partial Q(i, i)} + \left(C_{f|g} + \mu_{f|g} \mu_{f|g}^T \right) \frac{\partial}{\partial Q(i, i)} \left[(I - A)^T Q^{-1} (I - A) \right] \right\} \quad (\text{E.3})$$

$$\begin{aligned} \frac{\partial \Gamma(\theta)}{\partial \sigma_w^2} = \frac{\partial \Gamma(\theta)}{\partial R(i, i)} = \frac{1}{2} \text{tr} \left\{ R^{-1} \frac{\partial R}{\partial R(i, i)} + \left(C_{f|g} + \mu_{f|g} \mu_{f|g}^T \right) \frac{\partial}{\partial R(i, i)} \left(D^T R^{-1} D \right) \right\} \\ + \frac{1}{2} g^T \left[\frac{\partial R}{\partial R(i, i)} g - 2 \frac{\partial}{\partial R(i, i)} \left(R^{-1} D \right) \mu_{f|g} \right] \end{aligned} \quad (\text{E.4})$$

APPENDIX F: REQUIRED DERIVATIONS FOR CHAPTER 6

$$\frac{\partial(R^{-1}D)}{\partial d(m,n)} = R^{-1}E_{m,n} \quad (\text{F.1})$$

$$\frac{\partial}{\partial d(m,n)}(D^T R^{-1}D) = E_{m,n}^T R^{-1}D + D^T R^{-1}E_{m,n} \quad (\text{F.2})$$

$$\frac{\partial}{\partial a(k,l)}[(I-A)^T Q^{-1}(I-A)] = -Q^{-1}\{E_{k,l}^T(I-A) + E_{k,l}(I-A)^T\} \quad (\text{F.3})$$

$$\frac{\partial}{\partial a(k,l)}[(I-A)(I-A)^T] = -[E_{k,l}(I-A)^T + E_{k,l}^T(I-A)] \quad (\text{F.4})$$

$$\frac{\partial Q}{\partial Q(i,i)} = E_{i,i} \quad (\text{F.5})$$

$$\frac{\partial}{\partial Q(i,i)}[(I-A)^T Q^{-1}(I-A)] = -(I-A)^T Q^{-1}E_{i,i}Q^{-1}(I-A) \quad (\text{F.6})$$

$$\frac{\partial R}{\partial R(i,i)} = E_{i,i} \quad (\text{F.7})$$

$$\frac{\partial}{\partial Q(i,i)}(D^T R^{-1}D) = -D^T R^{-1}E_{i,i}R^{-1}D \quad (\text{F.8})$$

$$\frac{\partial R^{-1}}{\partial R(i,i)} = -R^{-1}E_{i,i}R^{-1} \quad (\text{F.9})$$

$$\frac{\partial}{\partial R(i,i)}(R^{-1}D) = -R^{-1}E_{i,i}R^{-1}D \quad (\text{F.10})$$

$E_{i,j}$ is an elementary matrix, which is a block-Toeplitz matrix, defined that is all zero except for a 1 at location (i,j) in each block.

CURRICULUM VITAE

Faruk SARI was born in İzmir in 1968. He got the Bachelor of Science degree from Yıldız University Kocaeli Engineering Faculty, Electronics and Communication Engineering Department in 1989. At the same year he was accepted to the Master of Science programme in the Electronics and Communication Department of Yıldız University and got M.Sc. degree in 1991. In 1994 he started his Ph.D. education in Electronics and Communication Department of Science and Technology Institute of Istanbul Technical University. He joined the research team of U.S. Navy, Naval Research Laboratory (NRL) and Virginia Tech, Alexandria Research Center in 2002-2003. He is currently working in a defence company as a R&D Director.