BLIND CHANNEL EQUALIZATION AND INSTANTANEOUS BLIND SOURCE SEPARATION

THESIS SUBMITTED
IN ACCORDANCE WITH THE REQUIREMENTS OF
THE UNIVERSITY OF LIVERPOOL
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Shafayat Abrar

Department of Electrical Engineering and Electronics
The University of Liverpool
Liverpool, United Kingdom

JUNE 2010

THE UNIVERSITY OF LIVERPOOL

Date: June 2010

Author:	Shafayat Abrar
Title:	Blind Channel Equalization and Instantaneous Blind Source Separation
Department:	Department of Electrical Engineering and Electronics
Degree:	Ph.D.
Year:	2010

THE AUTHOR DECLARES THAT THIS THESIS EMBODIES HIS OWN RESEARCH WORK, AND THAT IT IS COMPOSED BY HIMSELF. THE AUTHOR ATTESTS THAT PERMISSION HAS BEEN OBTAINED FOR THE USE OF ANY COPYRIGHTED MATERIAL APPEARING IN THIS THESIS (OTHER THAN BRIEF EXCERPTS REQUIRING ONLY PROPER ACKNOWLEDGEMENT IN SCHOLARLY WRITING) AND THAT ALL SUCH USE IS CLEARLY ACKNOWLEDGED.

To my beloved wife, Maria,
for her support, patience, and true love;
and to my adorable kids, Raem and Hufsa.

Abstract

Blind channel equalization and instantaneous blind source separation are two important blind signal processing tasks. Blind equalization refers to the problem of determining the propagation channel impulse response or its inverse when the channel is unknown and the transmitted data is inaccessible. Instantaneous blind source separation is a method of recovering unobserved source signals from observed instantaneous mixtures, exploiting only the assumption of mutual independence between source signals.

Stochastic gradient-based adaptive blind equalization algorithms are relatively simple to implement and are generally capable of delivering satisfactory performance, as is evidenced by their actual use in digital communication systems. Existing adaptive algorithms, however, are either slow in convergence or yield high steady-state misadjustment in equalizer output if forced to converge faster. This thesis proposes several new blind equalization algorithms which are either capable of converging faster or yielding lesser residual inter-symbol interference in steady-state.

Firstly, the blind equalization algorithms have been categorized in two groups: 1) multimodulus algorithms (MMA) and 2) constant modulus algorithms (CMA), depending respectively on whether the carrier-phase offset can jointly be acquired or not. Secondly, the notions of a) dispersion minimization and b) energy maximization have been introduced and under which, new cost-functions have been designed and associated adaptive algorithms have been derived.

Under the notion of dispersion minimization, a new CMA and two new MMA have been obtained. With the aid of computer simulations, it has been shown that newly obtained algorithms are capable of converging faster. Moreover, under the notion of energy maximization, two new algorithms, one for each CMA and MMA, have been obtained as well. New energy maximization based algorithms have been shown to exhibit better steady-state performance. Importantly, this investigation provides first treatment on the design of energy based adaptive blind equalization algorithms.

Finally, iterative joint-diagonalization based instantaneous blind source separation methods have been proposed which make optimal use of third- and fourth-order cumulants. We explore estimation of Given's rotation using both root-finding method and closed-form estimator. Numerous computer experiments verify the theoretical results and make comparison with existing methods.

Acknowledgments

I would like to thank my principal supervisor Prof. Asoke Kumar Nandi for the guidance, independence, trust and help, I received from him. In addition to being a productive supervisor, he is also a splendid human being and I have been thoroughly enriched working with and knowing him. I am also grateful to my second supervisor Dr. Xu (Judy) Zhu for her valuable comments on all of my annual progress reports.

I wish to express my appreciation to Dr. Waleed Al-Nuaimy for serving as one of the examination committee members. Special thanks are due to Dr. Sangarapillai Lambotharan from the Department of Electronics and Electrical Engineering, Loughborough University, for being willing to read my thesis and be my external examiner. I also wish to express my appreciation to Dr. Siavash Amin-Nejad for his supervision and guidance when I demonstrated BS/MS labs on HDL based digital system design.

I acknowledge my UoL colleagues, Hassan Chattha, Waqar Aslam, Sergio Bravo, Obilor Nwamadi, Hussain Jaafar, Moin Ahmed, Abdelrahman Al-Kharabsheh, Nader Abumustafa and Nayel Al-Zubi, for their company and warm friendship. Also, the help with the UNIX computer system, I received from Sonu Punnoose, Jingbo Gao and Sergio Bravo, was certainly invaluable. I would like to acknowledge my neighbors Mr. Mostafa, Dr. Humayun Bakht, Dr. Kashif Kifayat and Mrs. Tahira for their prayers and support.

I want to thank very specially my wife, my kids, my parents and my parents-in-law. I am infinitely grateful for the values they have passed down to me, and for their continuous support and encouragement throughout my studies, which successfully culminate in this thesis. My special gratitude is due to my wife, *Maria*, whose faith and love is my cornerstone. Without her continued encouragement, patience, and support, it would have been impossible for me to achieve this degree.

My work was mainly supported by the Overseas Research Studentship Awards Scheme (ORSAS, UK) and the University of Liverpool (UoL, UK) towards academic expenses. Also, I received financial support from the COMSATS Institute of Information Technology (CIIT, Pakistan) towards living expenses. I gratefully acknowledge all sources.

Shafayat Abrar
Liverpool, Summer 2010.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Blind Channel Equalization

Recent technological advances in the field of telecommunications have made possible the transmission of high data rates even in environments with high interferences. In the context of the actual communications systems, one of the biggest limitations in the transmission rates is the interference originated from the user itself, the so called *intersymbol interference* (ISI). The multiparty propagation (e.g. mobile systems) and limitation in bandwidth (e.g. telephone lines) are some of the most common phenomena that causes ISI. Classically, *equalization* has been the solution to eliminate or reduce this interference. Considering the growing complexity of the communication systems, the design of equalizers has become very important task and subject of many works.

One of the approaches in the theory of equalization is to make use of a training sequence known by both transmitter and receiver. During this period the equalizer has a copy of the transmitted data and it adjusts its parameters to learn the channel impulse response or its inverse. The other approach is the *non supervised equalization*, or simply, *blind equalization*, which is the focus of this work. Blind equalization is mainly characterized by the absence of a learning period. Several of its applications take advantage of this feature to improve spectral efficiency by using the time earlier spent on the training period to transmit information. Note that, in some cases, the transmission of a training sequence is undesired or even impossible, such as in multi-point computer network and radio-digital transmission on microwave band.

The most studied and implemented blind adaptation algorithm of the 1990's is the constant modulus algorithm (CMA) [53, 129]. The CMA seeks to minimize a cost defined by the constant modulus criterion. The constant modulus cost-function penalizes deviations in the modulus (i.e., magnitude) of the equalized signal away from a fixed value. In certain ideal conditions, minimizing the constant modulus cost can be shown to result in perfect (zero-forcing) equalization of the received signal [26]. Remarkably, the constant modulus criterion can successfully equalize signals characterized by source alphabets not possessing a constant modulus [e.g., quadrature amplitude modulation (QAM)], as well as those possessing a constant modulus [e.g., phase shift keying (PSK)]. In 1992, a modified version of CMA appeared [137], which was later termed as multimodulus algorithm (MMA) [139]. The most striking feature of MMA was mentioned to be its capability to jointly recover the carrier-phase offset while the CMA required a separate block for the same purpose. The objective of this thesis is to consider these blind channel equalization methods, to put forward new cost-functions, to derive associated algorithms, to disclose their links, and to carry out a comprehensive study of their properties and performance.

1.2 Blind Source Separation

The problem of *blind source separation* (BSS) arises in many signal processing applications like communications, array processing, speech analysis and speech recognition. In all these instances, the underlying assumption is that several linear mixtures of unknown, random, zero-mean, and statistically independent signals, called *sources*, are observed; the problem consists of recovering the original sources from their mixtures without a priori information of coefficients of the mixtures and knowledge of the sources. The principle involved in the solution to this problem is nowadays called *independent component analysis* (ICA), which can be viewed as an extension of the widely known *principal component analysis* (PCA). The independence between the recovered sources is measured by their *mutual information* (MI). The MI measures the information that one variable contains about another one, i.e., the reduction of uncertainty of a magnitude when another one is known. The MI is zero if and only if the sources are independent.

Comon [32] studied the separability condition for BSS problem, and pointed out that for statistically independent non-Gaussian sources, the separation can be achieved by restoring the independence. He proposed using MI as a tool to measure the independence of the output signals, and to use an Edgeworth expansion to approximate the probability density function in the MI criterion. The Edgeworth expansion of the MI of a standardized (i.e. after whitening) real variable, up to an additive constant I_0 and as a function of standardized *cumulants*. The ICA algorithms based on the maximization of third- and fourth-order cumulants are reported in [32] and [33] by Comon, respectively. In this thesis, we aim to present a weighted form of third- and the fourth-order cumulants, capable of handling the symmetric and asymmetric sources jointly, in an *optimal* manner.

1.3 Thesis Organization and Contributions

Below is the summary of the main original contributions of this work, together with its organization on a chapter-by-chapter basis:

Chapter 3

- A new adaptive constant modulus algorithm, namely cCMA(p), is obtained by minimizing a novel deterministic cost-function. The proposed cost-function constitutes a dispersion measure obtained by combining *a priori* and *a posteriori* equalizer outputs. It is proved, and experimentally endorsed, that for the given equalizer length, the proposed cCMA(p) exhibits better ISI mitigation capability for larger value of parameter p.

- A convergence proof is provided from system theoretic point of view. According to this proof, there exists a range of step-size for the equalizer implementing cCMA(p), for which the modulus of equalizer output can be kept bounded from above infinitely often.

- An easy-to-compute approximate bound is also obtained for the range of step-sizes for which a complex-valued *generic* constant modulus algorithm will remain stable if initialized close to a minimum of the constant modulus cost-function. Simulation results on probability of divergence.

Chapter 4

- Two new families of multimodulus algorithms are presented. The first family, MMA(p,q), is obtained by generalizing the Wesolowski dispersion criterion with two degrees of freedom. It is shown that faster converging algorithms may be obtained by selecting appropriate values of free parameters. The second family, cMMA(p), is obtained by constraining a so-called convex blind equalization criterion to acquire the signal energy on convergence.

- These algorithms are shown to be capable of recovering the carrier-phase jointly with equalization. Analytical reasonings are provided for carrier-phase recovery capability of these families.
- The dispersion constants for the selected members of the families, MMA(p,q) and cMMA(p), are evaluated in the presence and absence of convolutional noise.
- The dynamic convergence analysis is carried out for a generic Bussgang-type blind equalization algorithms leading to the evaluation of dynamic ISI and MSE expressions. Five members of each of MMA(p,q) and cMMA(p) are studied. Simulation results show the correctness of these expressions.

Chapter 5

- A method of l_2-optimization is suggested which leads to the formulation of a type of constant modulus cost-function. The method involved the maximization of equalizer output energy such that the dispersion of equalizer output is minimized with respect to the largest modulus of the transmitted signal. The feasibility of this method is analytically discussed.
- To obtain an adaptive algorithm, the proposed cost-function is modified to contain a differentiable constraint. The derived constant modulus algorithm, which we termed as β-CMA, is shown to be computationally simpler than some existing constant modulus algorithms.
- Experiments are provided to show the superiority of proposed algorithm, β-CMA, over traditional CMA algorithms for the equalization of symbol- and fractionally-spaced channels with APSK signalling.

Chapter 6

- A method of l_2-optimization has been recently suggested which leads to the formulation of a type of multimodulus cost-function. The method involved the maximization of equalizer output energy such that the dispersion in quadrature components of equalizer output are minimized with respect to the corner-points of QAM signal. To obtain an adaptive algorithm, the existing cost-function is modified to contain differentiable constraints. The derived multimodulus algorithm, which we termed as β-MMA, is shown to be computationally simpler than some existing multimodulus algorithms.

- Experiments are provided to show the superiority of derived algorithm, β-MMA, over traditional MMA algorithms for the equalization of symbol- and fractionally-spaced channels with 16/64/256-QAM signalling. Dynamic convergence analysis is also provided and experimentally verified.

Chapter 7

- In the field of blind source separation, joint-diagonalization based approaches constitute an important framework. Recently, some authors have shown how to perform diagonalization by simultaneously using cumulants of third- and fourth-order. In this Chapter, we extend these results to the optimal composition of third- and fourth-order cumulants. We introduce free parameters (or weights) β in combining the cumulants (of pair-wise mixed signals) and evaluate its optimal value such that the mean-square estimation of Given's rotation is minimized. We show that the optimal value of β depends on the *a priori* statistical knowledge of the mixing signals. However, based on several computer experiments, we notice that (even) in the absence of such a priori knowledge, the use of an approximate value of β (obtained directly from the statistics of the observed source) may lead to satisfactory performance and yield better results than some existing algorithms.

- We also obtain a closed-form estimator, which provides a quicker estimate of Given's rotation. Here we also take the statistics of background (additive Gaussian) noise in consideration.

1.4 Publications Derived from this Work

The following publications have arisen from the work detailed in this thesis:

Journal Papers

1. Abrar, S. and Nandi, A.K. "Independent component analysis: Jacobi-like diagonalization of optimized composite-order cumulants", *Proceedings A of The Royal Society*, vol. 465, no. 2105, pp. 1393-1411, May 2009.

2. Abrar, S. and Nandi, A.K. "An adaptive constant modulus blind equalization algorithm and its stochastic stability analysis", *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 55-58, January 2010.

3. Abrar, S. and Nandi, A.K. "Blind equalization of square-QAM signals: A multimodulus approach", to appear in *IEEE Transactions on Communications*, vol. 58, no. 6, pp. 1674-1685, June 2010.

Conference Papers

1. Abrar, S. and Nandi, A.K. "Normalized constant modulus algorithm for blind channel equalization", in *Proceeding of 16th European Signal Processing Conference (EUSIPCO-2008), Lausanne, Switzerland*, Aug. 2008.

2. Abrar, S. and Nandi, A.K. "Composite-order and optimized composite-order ICA algorithms for skewed sources", in *Proceeding of 2nd ICA Research Network International Workshop (ICArn-2008), Liverpool, UK*, pp. 1-4, Sept. 2008.

3. Abrar, S. and Nandi, A.K. "Closed-form optimized composite-order estimator for blind separation of instantaneous linear mixtures", in *Proceeding of 17th European Signal Processing Conference (EUSIPCO-2009), Glasgow, Scotland*, Aug. 2009.

4. Abrar S. and A.K. Nandi, "A blind equalization algorithm for (multimodulus) square-QAM signals", to appear in *Proceeding of 18th European Signal Processing Conference (EUSIPCO-2010), Aalborg, North Denmark*, Aug. 2010.

Under Review

1. Abrar, S. and Nandi, A.K. "Adaptive blind equalization algorithm by l_2 maximization of equalizer output", submitted to *IEEE Communications Letters.*

2. Abrar, S. and Nandi, AK. Àdaptive solution for blind equalization and carrier-phase recovery of square-QAM", submitted to *IEEE Signal Processing Letters.*

Chapter 2

Blind Channel Equalization: Adaptive Methods

2.1 Introduction

In many data communication systems, channel equalization is one of the most important functions for the receivers. Data communication requires that digital signals be transmitted over a specific analog medium between the transmitter and the receiver. Because of practical limitations, analog channels are usually imperfect and can introduce unwanted distortions. Examples of non-ideal analog media include telephone lines, coaxial cables, and wireless channels at various frequencies under different system configurations. For linearly distortive channels, channel equalization is an effective approach of distortion removal and compensation [37].

Adaptive channel equalization was first developed by Lucky [83] at Bell Labs for telephone channels. Adaptive channel equalizers begin adaptation with the assistance of a known training sequence transmitted during the initial stage by the transmitter. During the training phase, a known sequence is transmitted by the transmitter such that the equalizer output can be compared with the desired input to form an error. The equalizer parameters can be adjusted to minimize the mean square symbol error. At the end of the training phase, the equalizer parameters should be near their optimum values such that much of the intersymbol interference has been removed. As the channel input can now be correctly recovered from the equalizer output through a memoryless decision device, real data transmission can begin in the subsequent operation phase [37].

In many communications, signals are transmitted by the sender over time varying channels. As a result, a periodic training signal is necessary to identify or equalize the time-varying channel response. The drawback of this approach is evident in many communication systems where the use of training sequence can represent significant overhead costs and may be unrealistic or impractical. For instance, no training signal is available to receivers attempting to intercept enemy communication. In a multicast or a broadcast system, it is highly undesirable for the transmitter to engage in a training session for a single user by temporarily suspending its normal transmission to a number of other users. As a result, there is a strong and practical need for a special kind of channel equalizer, known as *blind equalizers*, that do not require the transmission of a training sequence. Digital cable TV and cable modems are excellent examples of such systems that directly benefit from blind equalization. In blind equalization, the actual data sequence is unknown to the receiver except for its probabilistic or statistical properties over a known alphabets of transmitted signal [37].

The blind equalization has become an important research problem in digital signal processing primarily because of its desirable features and the challenge it poses to researchers in the field [59, 26]. By eliminating training data and maximizing channel capacity for true information transmission, blind channel equalization presents a bandwidth efficient solution to distortion compensation. Compared with the more traditional approach of training based equalization, blind equalization is a theoretically challenging problem that is gaining appeal. There are a number of different approaches to the problem of blind equalization. In general, blind equalization methods can be classified into direct and indirect approaches [37]. Direct blind equalization approach derives equalizer filters directly from input statistics and the observed output signal of the unknown channel. Indirect blind equalization approach first identifies the underlying channel impulse response before designing an appropriate equalizer filter. **The present thesis focusses on direct and adaptive approach to achieve blind equalization.**

2.1.1 Blind SISO Equalization

Consider a single-input-single-output (SISO) discrete linear systems, where the relationship between the input and the output signal is given as

$$x_n = \sum_{k=0}^{K-1} h_k a_{n-k} + \vartheta_n \tag{2.1.1}$$

where n is discrete time index, $\{a_n\}$ is independently and identically-distributed transmitted sequence, and takes values of quadrature amplitude modulation (QAM) or amplitude phase shift-keying (APSK) symbols with equal probability. The h_k is impulse response of time-invariant moving-average channel, K is the channel length and ϑ_n is additive white Gaussian noise. Now consider a blind equalizer $\boldsymbol{w}_n$ is available at the receiver. The objective of this equalizer is to adjust its coefficients vector $\boldsymbol{w}$ such that the overall channel-equalizer impulse-response $h \circledast w_n = \delta_{n-\varsigma}$ is achieved or well approximated, where $\circledast$ denotes convolution and ς denotes bulk-delay. The key in the development of a blind equalizer is therefore to design the rule of self-adjustment with knowing the transmitted data $\{a_n\}$ and channel coefficients h_k.

There are basically two different approaches to the problem of blind equalization [37]. The *stochastic gradient-based* approach iteratively minimizes a chosen cost-function over all possible choices of the equalizer coefficients, while the *statistical* approach uses sufficient stationary statistics collected over a block of received data for channel identification or equalization. The latter approach often exploits higher order or cyclo-stationary statistical information directly. **The present thesis focusses on the adaptive online equalization methods employing the gradient-search approach.**

For reasons of practicality and ease of adaptation, a linear channel equalizer is typically implemented as a finite impulse-response (FIR) filter. Denote the equalizer parameter vector with N elements as

$$\boldsymbol{w}_n = [w_{n,0}, w_{n,1}, \cdots, w_{n,N-1}]^T,$$

In addition, define the received signal vector as

$$\boldsymbol{x}_n = [x_n, x_{n-1} \cdots, x_{n-N+1}]^T.$$

The output signal of the linear equalizer is thus

$$y_n = \boldsymbol{w}_n^H \boldsymbol{x}_n,$$

where the superscripts T and H represents transpose and conjugate transpose, respectively.

2.1.2 Cost-Functions and Associated Adaptive Functions

The idea behind the cost-function based blind equalization is to optimize, through the choice of equalizer filter coefficients $\boldsymbol{w}$, a certain real-valued cost-function J, such that the equalizer output y_n provides an estimate of the source signal a_n up to some inherent indeterminacies, giving, $y_n = \alpha\, a_{n'}, n' = n - \varsigma$, where $\alpha = |\alpha| e^{\jmath\gamma} \in \mathbb{C}$ and $\varsigma \in \mathbb{Z}$, represent an arbitrary gain and delay, respectively. The phase γ represents an isomorphic rotation of the symbol constellation and hence is dependent on the rotational symmetry of signal alphabets; for example, $\gamma = m\pi/2$ radians, with $m \in \{0, 1, 2, 3\}$ for a square-QAM system. Hence, a blind equalization algorithm tries to solve the following optimization problem:

$$w = \text{arg optmizie}_w J, \text{ with } J = \mathsf{E}[\mathcal{J}(y_n)] \tag{2.1.2}$$

where the cost J is an expression for implicitly embedded higher-order statistics of y_n and $\mathcal{J}(y_n)$ is a real-valued function. Ideally, the cost J makes use of statistics which are significantly modified as the signal propagates through the channel, and the optimization of cost modifies the statistics of the signal at the equalizer output, aligning them with those at channel input. The equalization is accomplished when equalized sequence y_n acquires an identical distribution as that of the channel input a_n [20]. If the implementation method is realized by stochastic gradient-based adaptive approach [60]:

$$w_{n+1} = w_n \pm \mu \left(\frac{\partial \mathcal{J}}{\partial w_n} \right)^*, \tag{2.1.3}$$

where the polarity (+) or (-) depends on whether we want to maximize or minimize the cost, respectively. The updating algorithm becomes

$$w_{n+1} = w_n \mp \mu \Phi(y_n)^* x_n, \quad \text{with } \Phi(y_n) = -\frac{\partial \mathcal{J}}{\partial y_n^*}, \tag{2.1.4}$$

where μ is a step-size governing the speed of convergence and the level of steady-state equalizer performance. The complex-valued *error-function* $\Phi(y_n)$ can be understood as an estimate of the difference between the desired and the actual equalizer outputs. That is, $\Phi(y_n) = \Psi(y_n) - y_n$, where $\Psi(y_n)$[1] is an estimate of the transmitted data a_n. Such nonlinear memory-less estimate, $\Psi(y_n)$, is usually referred to as *Bussgang* nonlinearity and is selected such that, at steady state, when y_n is close to $a_{n-\varsigma}$, the auto-correlation

[1] An admissible estimate of $\Psi(y_n)$ is the conditional expectation $\mathsf{E}[a_{n'}|y_n]$ [101]. Using Bayesian estimation technique, $\mathsf{E}[a_{n'}|y_n]$ was derived for non-Gaussian sources by Bellini [16], Fiori [45] and Haykin [60] assuming uniformly distributed sources. Just recently, in the year 2007, Pinchas and Bobrovsky [107] presented a new approximate expression for $\mathsf{E}[a_{n'}|y_n]$, based on Maximum Entropy principle and Laplace integral method, without assuming uniformly distributed sources; they jointly achieved blind equalization and carrier-phase recovery. In contrast, we rely on cost-function based intrinsically embedded signal/noise statistics for the estimation of $\Psi(y_n)$ in this work.

of y_n becomes equal to the cross-correlation between y_n and $\Psi(y_n)$, i.e.,

$$\mathsf{E}\left[y_n\Phi(y_{n-i})^*\right] = 0, \implies \mathsf{E}\left[y_n\, y_{n-i}^*\right] = \mathsf{E}\left[y_n\Psi(y_{n-i})^*\right] \tag{2.1.5}$$

For two-dimensional signals, the associated algorithms can be divided broadly into two main categories. The algorithms in the first category are those which achieve equalization without carrier-phase recovery while the algorithms in the second category are those which achieve equalization jointly with carrier-phase recovery. Algorithms in the first category (usually) exploit the *modulus* of the equalized signal, $|y_n|$, as a measure to detect and minimize the residual ISI; we refer to these algorithms as *constant modulus algorithms* (CMA) [53, 129]. General design principles for these type of algorithms are recently suggested in [11, 18].

The second category of algorithms exhibits a non-analytic error-function. By non-analytic, we mean that the error-function $\Phi(y_n) = \Psi(y_n) - y_n$ is a decoupled function of the quadrature components of deconvolved sequence y_n. We can write $\Phi(y_n) = \phi(y_{R,n}) + \jmath\phi(y_{I,n})$, so the real ($_R$) and imaginary ($_I$) parts of $\Phi(y_n)$ are respectively obtained from the real and imaginary parts of y_n. Its direct implication is the following optimization problem:

$$\boldsymbol{w} = \text{arg optmizie}_{\boldsymbol{w}} J, \text{ with } J = \mathsf{E}[\mathcal{J}(y_{R,n})] + \mathsf{E}[\mathcal{J}(y_{I,n})] \tag{2.1.6}$$

This *split* cost-function (2.1.6) was first introduced by Benveniste and Goursat [19] under two assumptions: 1) the quadrature components of transmitted data a_n are *independent* of each other and 2) the correlation between the real and imaginary parts of the ISI is *small.*

Later, this *heuristical* strategy was experimented by many researchers and several important observations have been made: i) faster convergence can be obtained for square-QAM in comparison to CMA(2,2) [105, 135, 136, 140], ii) carrier-phase recovery can jointly be achieved [143, 8, 6, 49], iii) moderate-level frequency-offset error can be tolerated [102], iv) lesser fluctuation in equalizer coefficients in steady-state in comparison to CMA(2,2) [5, 81], v) similar form of stationary points as those exhibited by CMA(2,2) [137, 143], vi) (possibly) no undesirable minima [75] and vii) ease in hardware implementation [92]. In contrast to the first category, we refer to algorithms in the second category as *multimodulus algorithms* (MMA). The term multimodulus was first used in [138], probably as a short form of the term *multiple-modulus* which was coined in [120].

If R_R and R_I are respectively assumed to be modulus for in-phase and quadrature components of y_n, then deducing from [11, 25], we establish that the error-function $\phi(y_{L,n})$ of MMA satisfies the following properties:

p1) $\phi(y_{L,n})$ is odd-symmetrical,

p2) $\phi(y_{L,n}) > 0$ for $0 < y_{L,n} < R_L$,

p3) $\phi(y_{L,n}) = 0$ solely at $y_{L,n} = 0, R_L$ and $-R_L$,

p4) $\phi(y_{L,n})$ must be decreasing for $y_{L,n} > R_L$.

where subscript L denotes either R or I. These properties define the requirements for the algorithm to converge to the modulus R_L. Consequently, p2)-p4) guarantee that the error-function $\phi(y_{L,n})$ drives $y_{L,n}$ towards the modulus R_L when $y_{L,n} > 0$ and p1) (together with p2)-p4)) ensures that $\phi(y_{L,n})$ brings $y_{L,n}$ close to minus of modulus (i.e., $-R_L$), when $y_{L,n} < 0$. Thus, an MMA equalizer minimizes the dispersion in $y_{R,n}$ and $y_{I,n}$ away from *four* statistical points $\pm R_R \pm \jmath R_I$; it is the reason that these carrier-phase sensitive algorithms have been termed as *reduced-* or *4-phase-constellation* algorithm [19, 52, 126].

2.2 Notion of Dispersion Minimization

Minimizing (a measure of) dispersion in the equalized sequence y_n is the most studied and practised idea in the context of equalization. The classical *mean-squared-error* (MSE) criterion minimizes the *dispersion* in y_n away from the delayed version of the true channel input a_n, that is $a_{n-\varsigma}$ (where ς is the bulk belay). The criterion is thus given by [41]

$$J_{\mathsf{MSE}} = \mathsf{E}\left[|y_n - a_{n-\varsigma}|^2\right] \tag{2.2.1}$$

The equalizer iteratively minimizes the MSE cost-function (2.2.1) in which the error is defined as $a_{n-\varsigma} - y_n$ and weight update is given by

$$\begin{aligned} w_{n+1} &= w_n + \mu\,(a_{n-\varsigma} - y_n)^*\, x_n, \\ &= w_n + \mu\left[(a_{R,n-\varsigma} - y_{R,n}) - \jmath\,(a_{I,n-\varsigma} - y_{I,n})\right] x_n. \end{aligned} \tag{2.2.2}$$

If the MSE is so small after training that the equalizer output y_n is a close estimate of the true channel input $a_{n-\varsigma}$, then the decision device output $\widehat{a}_{n-\varsigma} = \mathcal{Q}[y_n]$ can replace $a_{n-\varsigma}$

in a decision-directed algorithm (DDA) that continues to track modest time variations in the channel dynamics (this idea was suggested by Niessen in 1967 [100]).

In blind equalization, the channel input $a_{n-\varsigma}$ is unavailable, and thus different minimization criteria are explored. The crudest blind equalization algorithm is the decision-directed scheme that updates the adaptive equalizer coefficients according to

$$\begin{aligned} \boldsymbol{w}_{n+1} &= \boldsymbol{w}_n + \mu \left(\mathcal{Q}[y_n] - y_n \right)^* \boldsymbol{x}_n, \\ &= \boldsymbol{w}_n + \mu \left[\left(\mathcal{Q}[y_{R,n}] - y_{R,n} \right) - \jmath \left(\mathcal{Q}[y_{I,n}] - y_{I,n} \right) \right] \boldsymbol{x}_n. \end{aligned} \tag{2.2.3}$$

where $\mathcal{Q}[y_n] \equiv \mathcal{Q}[y_{R,n}] + \jmath \mathcal{Q}[y_{I,n}]$ is the estimated transmit signal $\widehat{a}_{R,n-\varsigma} + \jmath \widehat{a}_{I,n-\varsigma}$. The performance of the decision-directed algorithm depends on how close $\boldsymbol{w}_n$ is to its optimum setting $\boldsymbol{w}_{[*]}$ under the minimum MSE or the zero-forcing criterion [37]. The closer $\boldsymbol{w}_n$ is to $\boldsymbol{w}_{[*]}$, the smaller the ISI is and the more accurate the estimate $\mathcal{Q}[y_n] = \widehat{a}_{n-\varsigma}$ is to $a_{n-\varsigma}$. Consequently, the algorithm in (2.2.3) is likely to converge to $\boldsymbol{w}_{[*]}$ if $\boldsymbol{w}_n$ is initially close to $\boldsymbol{w}_{[*]}$. The validity of this intuitive argument is shown analytically in [88, 87, 70].

2.2.1 Sato Algorithm and its Variants

The first adaptive blind equalizer for real-valued multilevel PAM signals was introduced by Sato [114] which suggested to minimize dispersion in y_n away from two statistical points $\pm R$. For M-level PAM signals, it is defined by the update

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu \left(R \operatorname{sgn}[y_n] - y_n \right) \boldsymbol{x}_n, \tag{2.2.4}$$

where $R = \mathsf{E}\left[a^2\right] / \mathsf{E}\left[|a|\right]$. In essence, it is identical to the decision-directed algorithm when the PAM input is binary (± 1). Clearly, the Sato algorithm effectively replaces the ideal delayed transmitted data $a_{n-\varsigma}$ with $R \operatorname{sgn}[y_n]$, leading to the cost-function

$$J_{\mathsf{Sato}} = \mathsf{E}\left[\left(y_n - R \operatorname{sgn}[y_n] \right)^2 \right]. \tag{2.2.5}$$

It is clear that the convergence of the Sato algorithm depends on how often the two quantities $a_{n-\varsigma}$ and $R \operatorname{sgn}[y_n]$ have identical signs.

In the year 1984, Benveniste and Goursat [19], while exploiting Sato's principle, generalized the algorithm (2.2.4) to two-dimensional signals (like QAM) and formulated separate error terms for quadrature components, as follows:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu \left[\left(R_R \operatorname{sgn}[y_{R,n}] - y_{R,n} \right) - \jmath \left(R_I \operatorname{sgn}[y_{I,n}] - y_{I,n} \right) \right] \boldsymbol{x}_n, \tag{2.2.6}$$

where $R_L = \mathsf{E}\left[a_L^2\right]/\mathsf{E}\left[|a_L|\right]$. Clearly the algorithm (2.2.6) minimizes the dispersion in y_n away from four statistical points $\pm R_R \pm \jmath R_I$. For this reason, the algorithm (2.2.6) is usually termed as *reduced constellation algorithm* (RCA). For a symmetrical QAM constellation, the cost behind the update (2.2.6) can be expressed as:

$$J_{\mathsf{RCA}} = \mathsf{E}\left[\left(y_{R,n} - R_R \operatorname{sgn}\left[y_{R,n}\right]\right)^2\right] + \mathsf{E}\left[\left(y_{I,n} - R_I \operatorname{sgn}\left[y_{I,n}\right]\right)^2\right]. \tag{2.2.7}$$

However, it should be noted that, in the formulation of (2.2.6), authors have assumed that 1) the real and imaginary parts of the transmitted data a_n are *independent* of each other (as in the case of square-QAM) and 2) the correlation between the real and imaginary parts of the ISI is *small*. The contour plot of the cost (2.2.7) is depicted in Fig. 2.1 for 16-QAM signal. In order to improve the convergence properties of update (2.2.6),

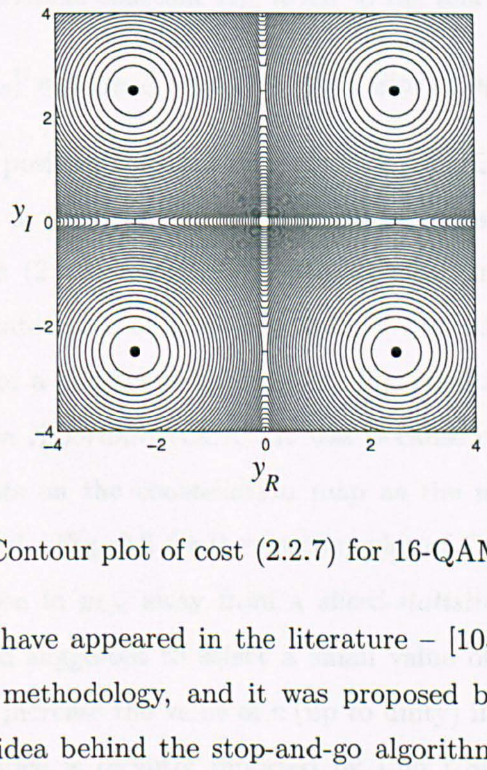


Figure 2.1: Contour plot of cost (2.2.7) for 16-QAM ($R_L = 2.5$).

two important variants have appeared in the literature – [105] and [2]. The first one is called "stop-and-go" methodology, and it was proposed by Picchi and Prati [105] in the year 1987. The idea behind the stop-and-go algorithms is to allow adaptation "to go" only when the error-function is more likely to have the correct sign for the gradient descent direction. Given several criteria for blind equalization, one can expect a more accurate descent direction when more than one of the existing algorithms agree on the sign (direction) of the error-functions. When the error signs differ for a particular output sample, parameter adaptation is "stopped". In [105], Picchi and Prati combined the Sato and the decision-directed algorithms with faster convergence results through the corresponding error-function

$$\begin{aligned}\Phi(y_n) =&\frac{1}{2}\Big[(\mathcal{Q}[y_{R,n}] - y_{R,n}) + |\mathcal{Q}[y_{R,n}] - y_{R,n}| \operatorname{sgn}[\beta \operatorname{sgn}[y_{R,n}] - y_{R,n}] \Big] \\ &+ \frac{\jmath}{2}\Big[(\mathcal{Q}[y_{I,n}] - y_{I,n}) + |\mathcal{Q}[y_{I,n}] - y_{I,n}| \operatorname{sgn}[\beta \operatorname{sgn}[y_{I,n}] - y_{I,n}] \Big]\end{aligned} \tag{2.2.8}$$

where $\beta > 0$. In [7], it is suggested to use $\beta = \max\{|a_L|\}$ to obtain fastest convergence. Note that, given the number of existing algorithms, the stop-and-go methodology can include many different combinations of error-functions, refer to a number of stop-and-go algorithms reported in [58, 3].

The second important variant of algorithm (2.2.6) was proposed by Abrar [2] in the year 2004, who suggested to infuse the error-functions of RCA and DDA. Therefore, the resulting error-function was aware of the dispersion of $y_{L,n}$ away from the decision symbol $\mathcal{Q}[y_{L,n}]$ as well as the statistical constant R_L, it led to the following error-function:

$$\Phi(y_n) = (R_R\, \mathcal{Q}[y_{R,n}]^c \operatorname{sgn}[y_{R,n}] - y_{R,n}) + \jmath\, (R_I\, \mathcal{Q}[y_{I,n}]^c \operatorname{sgn}[y_{I,n}] - y_{I,n}) \tag{2.2.9}$$

where $0 \le c \le 1$ was a positive constant and $R_L = \mathsf{E}\left[a_L^2\right]/\mathsf{E}\left[|a_L|^{1+c}\right]$. Note that by selecting $c = 1$ and $c = 0$, the error-function (2.2.9) becomes equivalent to the error-functions in the updates (2.2.3) and (2.2.6), respectively. In [2], it was shown that by selecting an appropriate value of c, usually $0.25 < c < 0.5$, one can obtain better convergence than RCA for a number of QAM sizes. The resulting algorithm was named as *Compact Constellation Algorithm* (CCA). It was because of the fact that there are as many statistical points on the constellation map as the number of distortion free QAM symbols. Refer to the Fig. 2.2 for the contour plot of CCA for 16-QAM. Clearly, CCA minimizes dispersion in $y_{L,n}$ away from a *sliced-statistical* constant $R_L\, \mathcal{Q}[y_{L,n}]^c$. In [5], Abrar and Axford suggested to select a small value of c (close to zero) at the beginning and smoothly increase the value of c (up to unity) in steady-state. Successful implementation of this idea is recently reported by Lim [78]. Finally, note that the update/error-functions (2.2.6), (2.2.8) and (2.2.9) have been found to be capable of recovering carrier-phase along with blind equalization.

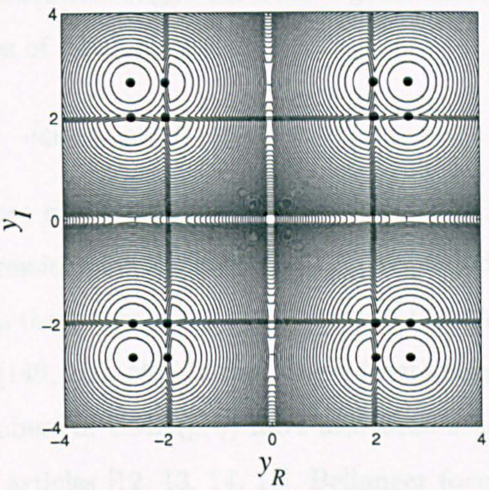


Figure 2.2: Contour plot of the algorithm CCA for 16-QAM ($R_L = 2.0$ and $c = 0.26$).

2.2.2 Godard Algorithms and its Variants

In the year 1980, the Sato cost-function (2.2.5) was generalized by Godard into another class of algorithms that are specified by the cost-function [53]

$$J_{\mathsf{Godard}} = \mathsf{E}\left[\left(|y_n|^p - R^p\right)^2\right], \tag{2.2.10}$$

where $R^p = \mathsf{E}\left[|a|^{2p}\right]/\mathsf{E}\left[|a|^p\right]$. Using the stochastic gradient descent approach, the *Godard (family of) algorithms* is given by

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\left(R^p - |y_n|^p\right)|y_n|^{p-2} y_n^* \boldsymbol{x}_n, \tag{2.2.11}$$

For $p = 2$, the special Godard algorithm was developed as the *constant modulus algorithm* (CMA) independently by Treichler and co-workers [129, 130] using the philosophy of *property-restoral*. In the sequel, we will refer to (2.2.11) as CMA$(p, 2)$. For channel input signal that has a constant modulus $|a_n| = R$, the CMA$(p, 2)$ equalizer penalizes output samples y_n that do not have the desired constant modulus characteristics.

This modulus restoral concept has a particular advantage in that it allows the equalizer to be adapted independent of carrier recovery. Because the CMA$(p, 2)$ cost-function is insensitive to the phase of y_n, the equalizer parameter adaptation can occur independently and simultaneously with the operation of the carrier recovery system. This property also makes CMA$(p, 2)$ applicable to analog modulation signals with constant amplitude such as those using frequency or phase modulation [130].

In [73], Treichler and Larimore suggested a more generalized constant modulus cost-function, with two degrees of freedom, *viz*

$$J_{\mathsf{CMA}(p,q)} = \mathsf{E}\left[\left||y_n|^p - R^p\right|^q\right], \tag{2.2.12}$$

and studied its convergence characteristics for $(p,q) \in \{(1,1),(1,2),(2,1),(2,2)\}$. More rigorous closed-form expressions for the cost $J_{\mathsf{CMA}(p,q)}$ appeared in [123].

It is interesting to note that, only one of the members of CMA(p,q), that is CMA(2,2), has been widely studied [149, 150, 56]. In fact, there exists very few references [72, 142, 64, 80], where other members of CMA(p,q) have also been studied and compared with CMA(2,2). In a series of articles [12, 13, 14, 15], Bellanger focussed on the relative performances of CMA(1,2) and CMA(2,2). In [13], for QPSK signals, he established that if output-signal-to-noise-ratio is less than 8dB, then CMA(2,2) can be employed, otherwise CMA(1,2) criterion is preferable. In [15], for square-QAM signals while acquiring the same level of excess MSE, he quantified that CMA(2,2) may converge in less than half of the number of iterations as required by CMA(1,2).

In the year 2006, Li and Zhang [76] proposed a generalized constant modulus cost-function with three degrees of freedom, as given by

$$\begin{aligned} J_{\mathsf{GCMA}(l,p,q)} &= \mathsf{E}\left[\left||y_n|_l^p - R^p\right|^q\right], \\ \text{where} \quad |y_n|_l &\equiv \left(|y_{R,n}|^l + |y_{I,n}|^l\right)^{1/l}, \quad l \geq 1. \end{aligned} \tag{2.2.13}$$

Although, they have not reported any new (and better) algorithm (with appropriate selection of parameters p, q and l), but they proved affinity among many existing blind equalization algorithms, like CMA(2,2), constant square algorithm [127], sign Godard algorithm (also known as constant diamond algorithm) [133], RCA [19] and multimodulus algorithm [139]. Interestingly, in the year 2007, Goupil and Palicot [55] *independently* proposed the same cost-function (2.2.13); however, they remarkably obtained the *optimal* value of l for several APSK/QAM signals given $p = q = 2$. In Fig. 2.3, we depict the (only) zero-error contours of the cost (2.2.13) for an arbitrary signal and $l = 1, 2, 6$ and $l \to \infty$. Note that, due to minimizing dispersion away from *non-circular* zero-error contours, for $l = 1, 6$ and $l \to \infty$, the resulting algorithms have been shown to be capable of restoring the carrier-phase in the equalized sequence with needing a separate phase-acquisition/tracking block [76, 55].

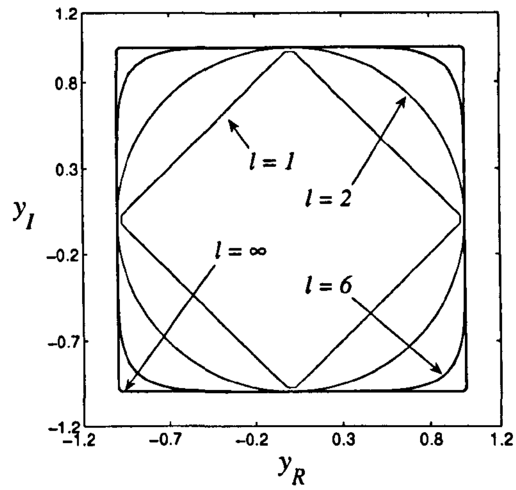


Figure 2.3: Zero-error contours of cost (2.2.7) for $l = 1, 2, 6$ and $l \rightarrow \infty$.

2.2.3 Wesolowski Algorithms and its Variants

In the year 1987, Weoslowski proposed a modified form of Godard dispersion-directed criterion (2.2.10) as follows:

$$J_{\text{Wesolowski}} = \mathsf{E}\left[\left(|y_{R,n}|^p - R^p\right)^2\right] + \mathsf{E}\left[\left(|y_{I,n}|^p - R_I^p\right)^2\right], \tag{2.2.14}$$

The idea behind this cost-function, as compared to the CMA$(p, 2)$ cost-function (2.2.10) is that both the real and imaginary parts of the signal are forced to a constant value and, therefore, the random phase ambiguity of the CMA$(p, 2)$ now becomes only 90°. This is meaningful in pure phase modulation, in which case the CMA may converge to an arbitrarily phase-shifted solution. For QAM though, the 90° symmetry of the constellation makes it possible for both algorithms to converge to a 90° phase-shifted solution. The minimization of (2.2.14) provides the following gradient-based weight update algorithm

$$w_{n+1} = w_n + \mu\left[\left(R_R^p - |y_{R,n}|^p\right)|y_{R,n}|^{p-2}y_{R,n} - \jmath\left(R_I^p - |y_{I,n}|^p\right)|y_{I,n}|^{p-2}y_{I,n}\right]x_n \tag{2.2.15}$$

where $R^p = \mathsf{E}\left[a_L^{2p}\right]/\mathsf{E}\left[|a_L|^p\right]$. Interestingly, a decade later, the cost (2.2.14) was independently proposed by Oh and Chin [102, 103] and Yang *et al.* [139, 138, 141]. In [139, 138], the term *multimodulus algorithm* (MMA) was coined for the update (2.2.15). In the sequel, we will refer to update (2.2.15) as MMA(p,2).

The stationary points of MMA(2,2) and its carrier-phase recovery capability is explored in detail in [137, 143] and [75]. It was shown that MMA(2,2) cost-function has only four local (hence, global) minima. Hence, a gradient search algorithm for minimizing $J_{\mathsf{MMA}(2,2)}$ is expected to converge to the only four global minima that correspond to the desired response. The key lies in the fact that the cost-function of the MMA(2,2) includes a term which contains the phase information of the blind equalizer output [143]. Therefore, the MMA may solve a possible phase-ambiguity problem inherent in the CMA. However, the existence of that phase-sensitive term may appear to result in one potential disadvantage of the MMA(2,2), because four additional saddle points are generated. As a result, the MMA, using the stochastic gradient-descent method, may be first attracted toward the vicinity of one of the saddle points, around which it exhibits slow convergence, before converging to the desired minimum.

Also note that there exist numerous simulation-based studies which have reported the potential of MMA(2,2) and its variants for acquiring faster convergence than existing algorithms. Examples of such studies for 16/64/256-QAM are in abundance and widely known [49, 140, 5, 8, 4, 6]; for the case of 1024-QAM signalling, however, one can refer to [66, 71, 61, 44, 42].

2.3 Notion of Constrained Energy Optimization

Historically, Allen and Mazo [10] were the first who showed that the optimization of output energy of an anchored equalizer is unimodal in achieving an open eye solution for blind equalization. This proposal appeared an year before the appearance of Sato's work (Sato's work appeared in 1975 [114]). According to Allen-Mazo criterion, in contrast to (2.2.1), blind equalization may be achieved by solving

$$\min_{\boldsymbol{w}} \mathsf{E}\left[|y_n|^2\right] \tag{2.3.1}$$

subject to keeping a tap value $w_0 = 1$. Thus no knowledge of the input sequence $\{a_{n-\varsigma}\}$ was required while anchoring was required to avoid all-zero solution. Note that the cost (2.3.1) serves as the first ever proposal for blind equalization; however it simply failed to get any attention. In fact, to the best of our knowledge, no article, discussing the issue of blind equalization, has ever provided any pointer to this important proposal. Probably, the works of Sato and Godard successfully captured all the attentions.

Allen and Mazo [10] have analytically shown that the optimization of (2.3.1) may possibly invert the propagation channel provided the channel belongs to a specific class and the equalizer has a certain structure. Specifically **1)** the channel must be minimum phase, and **2)** the equalizer must be one-sided. A two-sided equalizer (which is the conventional one in data transmission) is certainly capable of giving an even smaller value to the criterion (2.3.1); the resulting taps will simply not be those that invert the channel [9].

Seventeen years later, in 1991, Sethares *et al.* [119] independently proposed the criterion (2.3.1) and argued that, at the point of minimum energy, the equalizer may fail to open the eye even when an open eye solution exist. They concluded in the following words:

> *Minimizing the energy (l_2 minimization) of the equalizer output (under a fixed tap constraint) cannot be guaranteed to open the eye (to reliably unscramble the message) because it tends to converge to an equalizer setting that contains a reflection of the unstable zeros inside the unit circle.*

Quite concurrently, Feyh and Klemt [43] proposed the following similar problem:

$$\min_{\boldsymbol{w}} \sum_{n=1}^{B} |y_n|^2, \quad \text{subject to } \|\boldsymbol{w}\|_2^2 = 1, \tag{2.3.2}$$

where B denoted the number of equalized samples used in the optimization. An eigenvector based block processing algorithm was obtained using subroutines from the NAG library. The simulation studies in [43] reported a failure of (2.3.2) in the following words:

> *If the time span of the adaptive filter $\boldsymbol{w}$ was larger than the symbol duration, the algorithm tended to gather runs of one symbol into one "super-symbol". On the other hand symbol changes were not followed as closely as they should have been, thus single symbols in between two runs of symbols were completely lost.*

Now it is important to know that whether energy optimization is an admissible idea for blind equalization or not. We believe that it was not the failure of the cost (2.3.1), but it was the failure due to the type of constraint used. A properly selected constraint may possibly lead to an admissible criterion!.

Recently, in the year 2009, Meng *et al.* [90] proposed an l_2-maximization based method for blind channel equalization without requiring tap-anchoring. Their cost-function is given as:

$$\max_{\boldsymbol{w}} \mathsf{E}[|y_n|^2], \text{ s.t. } \max(\{|y_{R,n}|\}) = \max(\{|y_{I,n}|\}) \leq \gamma \tag{2.3.3}$$

where $\max(\{|y_R|\})$ and $\max(\{|y_{I,n}|\})$ denote respectively the largest absolute values of the in-phase and quadrature components of equalized sequence $\{y_n\}$, and the parameter γ denotes the maximum quadrature component of the transmitted data a_n. They formulated the cost as a *quadratic programming* problem for blind equalization of square-QAM and reported better results than those obtained from *linear programming* based solutions [38, 86]. Note that the constraints in (2.3.3) have been shown to be *convex* in w [38, 69, 86]. Also note that, due to using separate constraints for in-phase and quadrature components of equalized sequence, the Meng's block-processing equalizer was jointly capable of recovering the carrier-phase.

2.4 Summary

This Chapter has recalled some aspects of cost-function based blind equalization problem, concentrating on **adaptive** methods and has made a review of **constant modulus** and **multimodulus** algorithms. The notion of **dispersion minimization** and **energy optimization** is explained and examples of existing relevant methods are described. The purpose of this thesis is to study such methods and develop new algorithms along similar lines.

Chapter 3

Dispersion Minimization: Adaptive Constant Modulus Algorithms

3.1 Introduction

In this Chapter, we propose a dispersion minimization based family of adaptive constant modulus algorithms for blind equalization of complex-valued communication channels. We first suggest a new composite dispersion measure which is a weighted sum of dispersion measures composed of *a priori* as well as *a posteriori* equalizer outputs (Section 3.2.1). The proposed dispersion measure is then used to define a novel deterministic (minimum-disturbance) constrained optimization criterion (Section 3.2.2). We solve the proposed criterion by satisfying the constraints in a *soft* or *relaxed* manner to obtain a new algorithm, cCMA(p) (Section 3.2.2). The proposed algorithm cCMA(p) is shown to exhibit modulus driven zero-memory continuous nonlinearity, where p is free (positive-valued) parameter.

We discuss the analytical behavior of cCMA(p) in detail. We evaluate the dispersion constants in the presence as well as the absence of convolutional noise, leading to closed-form interesting formulas for specific cases of p (Section 3.3). We discuss the stochastic stability of a generic CMA leading to the derivation of a closed-form simple-to-evaluate bound on step-size (Section 3.4). We discuss the stability of cCMA(p) from system theoretic point of view and present an interesting theorem describing the effect of parameter p on dispersion minimizing performance of cCMA(p) (Section 3.5). We also discuss the influence of parameter p on ISI mitigation capability of cCMA(p). Here, we explicitly prove that the residual ISI level may be lowered further by increasing the value of free parameter (Section 3.6). We finally validate most of our theoretical results through computer simulations (Section 3.7).

3.2 Proposed Cost-Function

3.2.1 A Composite Dispersion Measure

The Godard's cost-function can be written as [53]:

$$\min_{\boldsymbol{w}} \mathsf{E}\left[\left(\xi^{\mathsf{a}}_{n,p}\right)^2\right], \quad \xi^{\mathsf{a}}_{n,p} \triangleq R^p - |y_n|^p \tag{3.2.1}$$

where $\xi^{\mathsf{a}}_{n,p}$ is defined as the pth-order *a priori dispersion error.* The criterion (3.2.1) minimizes the dispersion of the modulus of *a priori* output y_n away from a statistical dispersion constant R. The cost yields the following stochastic gradient algorithm:

$$\text{CMA}(p,2): \ \boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu \xi^{\mathsf{a}}_{n,p}|y_n|^{p-2}y_n^* \boldsymbol{x}_n \tag{3.2.2}$$

where $R^p = \mathsf{E}[|a|^{2p}]/\mathsf{E}[|a|^p]$ [53] in a noise-free scenario. The stochastic gradient algorithm (3.2.2) drops the expectation operator and minimizes (3.2.1) by performing one iteration per symbol period. It is interesting to note that only two members of this family, namely CMA$(1,2)$ and CMA$(2,2)$, have been widely and till recently discussed and studied [12, 13, 15].

For higher-order two-dimensional signals like QAM, the instantaneous error-function $\xi^{\mathsf{a}}_{n,p}|y_n|^{p-2}y_n^*$ is usually *enormous* and leads to a *high* steady-state fluctuation even if the equalizer converges successfully (refer to the steady-state jitter analysis of CMA(p,2) in [110]). However, it is possible to post-process this error-function to estimate some *a posteriori dispersion error* in such a way that useful information can be obtained on the actual error present in the equalization solution [111]. Clearly, if such an error can be estimated, then it should be possible to enhance the equalization so as to reduce the error. Let

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\Phi(y, y^*)^* \boldsymbol{x}_n \tag{3.2.3}$$

be a generic weight-update where $\Phi(y, y^*) = \Psi(y_n, y_n^*) - y_n$ is a memory-less nonlinear blind estimate of prediction error, and the nonlinearity $\Psi(y_n, y_n^*)$ be selected such that, upon convergence, equalizer restores the actual signal energy [17]. Let $s_n = \boldsymbol{w}_{n+1}^H \boldsymbol{x}_n$

be the *a posteriori* output of the equalizer. Taking Hermitian transpose of (3.2.3) and post-multiplying with $\boldsymbol{x}_n$, we obtain

$$s_n = \widetilde{\mu}\,\Psi(y_n, y_n^*) + (1 - \widetilde{\mu})\,y_n, \tag{3.2.4}$$

where $\widetilde{\mu} = \mu\|\boldsymbol{x}_n\|_2^2$. This shows that s_n *is a linear combination of* y_n *and* $\Psi(y_n, y_n^*)$. Hence, the dispersion of s_n away from $\Psi(y_n, y_n^*)$ will be less than the dispersion of y_n away from $\Psi(y_n, y_n^*)$, where $\widetilde{\mu} \in (0, 1]$ controls the extent to which s_n approaches $\Psi(y_n, y_n^*)$. Let

$$\xi_{n,p}^{\mathrm{p}} = R^p - |s_n|^p \tag{3.2.5}$$

be the pth-order *a posteriori dispersion error.* Now consider CMA(2, 2) equalizer where we have $\Psi(y_n, y_n^*) = (\xi_{n,2}^{\mathrm{a}} + 1)y_n$. Exemplary, if we assume $|y_n|^2 = R^2 - \varepsilon, \varepsilon > 0$, then $\xi_{n,2}^{\mathrm{a}} = R^2 - |y_n|^2 = \varepsilon$ and $\xi_{n,2}^{\mathrm{p}} = R^2 - |s_n|^2 = \left[R^2 - (1 + \widetilde{\mu}\varepsilon)^2(R^2 - \varepsilon)\right]$ will be *positive* and *less* than ε, provided

$$0 < \widetilde{\mu} < \frac{R - \sqrt{R^2 - \varepsilon}}{\varepsilon\sqrt{R^2 - \varepsilon}}.$$

Clearly, by considering *a posteriori* quantities it should be possible to enhance the quality of equalization.

In the past, exploiting a posteriori quantities, deterministic and constrained cost-functions have been proposed to obtain variants of normalized CMA for QPSK [104] and APSK/QAM signals [125]. In this work, we instead suggest to formulate a deterministic cost-function constituting both a priori and a posteriori quantities. For this purpose, we propose a $(p_1 + p_2)$th-order *composite dispersion error*

$$\xi_{n,p_1,p_2}^{\mathrm{c}} = R^{p_1+p_2} - |s_n|^{p_1}|y_n|^{p_2}, \tag{3.2.6}$$

where p_1 and p_2 are positive integers. Notice that the $\xi_{n,p_1,p_2}^{\mathrm{c}}$ can be expressed in terms of ξ_{n,p_1}^{p} and ξ_{n,p_2}^{a} as follows:

$$\xi_{n,p_1,p_2}^{\mathrm{c}} = R^{p_1}\xi_{n,p_2}^{\mathrm{a}} + R^{p_2}\,\xi_{n,p_1}^{\mathrm{p}} - \xi_{n,p_1}^{\mathrm{p}}\,\xi_{n,p_2}^{\mathrm{a}} \tag{3.2.7}$$

which indicates that $\xi_{n,p_1,p_2}^{\mathrm{c}}$ is a sort of *weighted composition* of ξ_{n,p_1}^{p} and ξ_{n,p_2}^{a}. Note that expressions (3.2.6) and (3.2.7) are identical.

3.2.2 The Cost-Function and its Relaxed Optimization

After having the equalizer estimate $\boldsymbol{w}_n$, we aim to minimize an instantaneous deterministic cost:

$$\min_{\boldsymbol{w}_{n+1}} \left(\xi^{\mathrm{c}}_{n,p_1,p_2}\right)^2. \tag{3.2.8}$$

It is obvious that the above cost (3.2.8) can perfectly be minimized while leaving $\boldsymbol{w}_{n+1}$ largely undetermined. To fix the degree of freedom in $\boldsymbol{w}_{n+1}$, we can impose that $\boldsymbol{w}_{n+1}$ remains as close as possible to its prior estimate $\boldsymbol{w}_n$, while satisfying the constraints imposed by the new data, i.e., $\xi^{\mathrm{c}}_{n,p_1,p_2} = 0$. Using Lagrange multipliers, we formulate the following constrained optimization problem:

$$\min_{\boldsymbol{w}_{n+1}} \left\{\|\boldsymbol{w}_{n+1} - \boldsymbol{w}_n\|_2^2 + \lambda\, \xi^{\mathrm{c}}_{n,p_1,p_2}\right\} \tag{3.2.9}$$

For a tractable derivation, we suggest to use $p_1 = 2$ and $p_2 = 2p - 2$, it gives

$$\min_{\boldsymbol{w}_{n+1}} \left\{\|\boldsymbol{w}_{n+1} - \boldsymbol{w}_n\|_2^2 + \lambda \left(R^{2p} - |s_n|^2\, |y_n|^{2p-2}\right)\right\} \tag{3.2.10}$$

Now, differentiating (3.2.10) with respect to $\boldsymbol{w}_{n+1}$ and setting the result equal to zero,

$$\boldsymbol{w}^*_{n+1} - \boldsymbol{w}^*_n - \lambda \boldsymbol{x}^*_n \boldsymbol{x}^T_n \boldsymbol{w}^*_{n+1} |\boldsymbol{w}^H_n \boldsymbol{x}_n|^{2p-2} = 0 \tag{3.2.11}$$

Transposing (3.2.11) and post-multiplying it with $\boldsymbol{x}_n$ leads to

$$s_n - y_n - \lambda s_n |y_n|^{2p-2} \|\boldsymbol{x}_n\|_2^2 = 0 \tag{3.2.12}$$

Solving (3.2.12) yields the optimum Lagrange multiplier, $\lambda_{[*]}$,

$$\lambda_{[*]} = \frac{s_n - y_n}{s_n\, |y_n|^{2p-2}\, \|\boldsymbol{x}_n\|_2^2} \tag{3.2.13}$$

and the update reads

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \lambda_{[*]} s^*_n |y_n|^{2p-2} \boldsymbol{x}_n. \tag{3.2.14}$$

At each n, the *hard constraint* in (3.2.9) enforces $s_n = R^p |y_n|^{-p} y_n$. Therefore the optimum Lagrange multiplier in (3.2.13) becomes

$$\lambda_{[*]} = \frac{\xi^{\mathrm{a}}_{n,p}}{R^p\, |y_n|^{2p-2}\, \|\boldsymbol{x}_n\|_2^2} \tag{3.2.15}$$

At this stage, using the approach of [125], we deviate a little and introduce a relaxation factor, η, in (3.2.15) to control the degree of constraint satisfaction. It implies that the constraint on s_n is now retained as a *soft constraint.* By introducing η, we have a *relaxed* Lagrange multiplier, and the corresponding update equation reads

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \eta \frac{\xi^{\mathsf{a}}_{n,p} s^*_n}{R^p} \frac{\boldsymbol{x}_n}{\|\boldsymbol{x}_n\|^2_2} \tag{3.2.16}$$

Taking Hermitian transpose and post-multiplying (3.2.16) with $\boldsymbol{x}_n$ we obtain

$$s_n = \frac{y_n R^p}{R^p - \eta\, \xi^{\mathsf{a}}_{n,p}}, \tag{3.2.17}$$

which leads to

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \eta \frac{\xi^{\mathsf{a}}_{n,p}\, y^*_n}{R^p - \eta\, \xi^{\mathsf{a}}_{n,p}} \frac{\boldsymbol{x}_n}{\|\boldsymbol{x}_n\|^2_2} \tag{3.2.18}$$

The computational complexity of algorithm (3.2.18) is little higher than that of CMA$(p, 2)$. This complexity can be reduced by observing that, for $\eta \ll 1$, the denominator can be approximated as $R^p - \eta\, \xi^{\mathsf{a}}_{n,p} \approx R^p$. Also, by removing the normalization factor and denoting $\mu = \eta/R^p$, the following simplified variant is obtained:

$$\text{cCMA}(p): \quad \boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\, \xi^{\mathsf{a}}_{n,p}\, y^*_n \boldsymbol{x}_n = \boldsymbol{w}_n + \mu\, (R^p - |y_n|^p)\, y^*_n \boldsymbol{x}_n, \tag{3.2.19}$$

We denote (3.2.19) as *pth-order composite constant modulus algorithm*, cCMA(p). The derivation of the dispersion constant in (3.2.19) will be discussed in Section 3.3. The difference between CMA$(p, 2)$ and cCMA(p) is that, the later one lacks the factor $|y_n|^{p-2}$. Obviously, this factor has no effect on the direction of adaptation; removing it, may have an advantageous effect of reducing the magnitude of adaptation. The two algorithms are equivalent only for $p = 2$. Clearly, due to the removal of factor $|y_n|^{p-2}$, at each update, the cCMA(p) requires lesser real-valued multiplications by an amount of $1 + \log_2\left(\frac{p-2}{2}\right)$ than CMA$(p, 2)$ for $p = 2, 4, 6, \cdots$.

3.3 Evaluation of Dispersion Constants

The dispersion constants are considered as the *statistical gain* of equalizer which contain *embedded* information about the true energy of the transmitted signal. According to Bellini [16], *the (dispersion) constant, which controls the equalizer amplification, is chosen to give zero tap-gain increments when perfect equalization is achieved,* i.e., $\mathsf{E}[\Phi(y_n)^*\, x_{n-i}] = 0$. We assume a converged equalizer such that $y_n = a_{n-\varsigma} + u_n$, where u_n is convolutional noise and is assumed to be zero-mean Gaussian [51]. We get

$$\mathsf{E}\left[\Phi(a_{n-\varsigma} + u_n)^* \left(\sum_{k=0}^{K-1} h_k a_{n-k-i} + \vartheta_n\right)\right] = 0.$$

We note that the expected values in the sum are zero whenever $n-\varsigma \neq n-k-i$; on the other hand, considering channel coefficients constant, we get

$$\mathsf{E}\left[\Phi(a_{n-\varsigma}+u_n)^*\left(a_{n-\varsigma}+\vartheta_n\right)\right]=0.$$

If we assume that u_n and ϑ_n are highly correlated then

$$\mathsf{E}\left[\Phi(a_{n-\varsigma}+u_n)^*\left(a_{n-\varsigma}+u_n\right)\right]=0. \tag{3.3.1}$$

Exploiting (3.3.1) and considering $\Phi(z)=(R^p-|z|^p)\,z$, we can obtain an expression for the dispersion constant of cCMA(p). Let the amplitude of $y_n = a_{n'} + u_n$ be $\mathcal{Y}_n$, and it is expressed as $\mathcal{Y}_n = \mathcal{R}_n + \mathcal{U}_n$, where $\mathcal{R}_n = |a_n|$ and $\mathcal{U}_n$ is the random amplitude component due to u_n. Note that, the data and noise sequences are assumed to be independent and identically distributed (i.i.d.) with zero-means.

For a zero-mean complex-valued (narrow-band) Gaussian noise with total variance 2σ, the probability density function of the amplitude, $\mathcal{U}_n$, is known as the *Rayleigh* distribution and is given by [109]

$$\mathrm{p}_{\mathcal{U}}(\tilde{u})=\frac{\tilde{u}}{\sigma^2}\exp\left(-\frac{\tilde{u}^2}{2\sigma^2}\right),\qquad \tilde{u}>0 \tag{3.3.2}$$

Now consider the amplitude $\mathcal{R}_n$ is perturbed by a narrow-band Gaussian noise $\mathcal{U}_n$, the resulting pdf is referred to as *Rice* distribution and is expressed as [109]:

$$\mathrm{p}_{\mathcal{Y}}(\tilde{y})=\frac{\tilde{y}}{\sigma^2}\exp\left(-\frac{\tilde{y}^2+\mathcal{R}^2}{2\sigma^2}\right)I_0\left(\frac{\tilde{y}\mathcal{R}}{\sigma^2}\right),\qquad \tilde{y}>0 \tag{3.3.3}$$

where $I_0(\cdot)$ is the modified Bessel function of first kind with zero order. Consider that the distortion free M-symbol QAM signal a_n comprises L number of unique moduli, such that, $\mathcal{R}_i \in \{R_1,\cdots,R_L\}$ and M_i denotes the number of unique symbols on the ith modulus $\mathcal{R}_i$. As a result, the conditional amplitude pdf of QAM signal is expressed as

$$\mathrm{p}_{\mathcal{Y}}(\tilde{y}|R_i)=\frac{\tilde{y}}{\sigma^2}\exp\left(-\frac{\tilde{y}^2+R_i^2}{2\sigma^2}\right)I_0\left(\frac{\tilde{y}R_i}{\sigma^2}\right) \tag{3.3.4}$$

Using (3.3.4) we can compute statistical moments of equalizer output y_n, like

$$\mathsf{E}\left[|y_n|^p\right]=\mathsf{E}\left[\mathcal{Y}_n^p\right]=\mathsf{E}_{R_i}\!\left[\int_0^\infty \tilde{y}^p\,\mathrm{p}_{\mathcal{Y}}(\tilde{y}|R_i)\mathrm{d}\tilde{y}\right]=\frac{1}{M}\sum_{i=1}^{L}M_i\!\int_0^\infty \tilde{y}^p\,\mathrm{p}_{\mathcal{Y}}(\tilde{y}|R_i)\mathrm{d}\tilde{y}. \tag{3.3.5}$$

For cCMA(1), we obtain

$$R=\frac{\mathsf{E}\left[|y_n|^2\right]}{\mathsf{E}\left[|y_n|\right]}=\frac{\sqrt{\frac{8}{\pi}}\left(2\sigma^3+\sigma\mathsf{E}[|a|^2]\right)}{\mathsf{E}\left[\exp(-\frac{1}{4}|a|^2\sigma^{-2})\left\{(2\sigma^2+|a|^2)I_0(\frac{1}{4}|a|^2\sigma^{-2})+|a|^2I_1(\frac{1}{4}|a|^2\sigma^{-2})\right\}\right]} \tag{3.3.6}$$

Using the following property (for $\alpha \in \{0, 1\}$ and $\beta > 0$):

$$\lim_{x \to 0} \frac{\sqrt{x} \exp\left(\frac{\beta}{x}\right)}{I_\alpha\left(\frac{\beta}{x}\right)} \longrightarrow \sqrt{2\pi\beta} \tag{3.3.7}$$

we can show that

$$\lim_{\sigma \to 0} R \longrightarrow \frac{\mathsf{E}[|a|^2]}{\mathsf{E}[|a|]} \tag{3.3.8}$$

Similarly, for cCMA(2), we obtain

$$R^2 = \frac{\mathsf{E}\left[|y_n|^4\right]}{\mathsf{E}\left[|y_n|^2\right]} = \frac{\mathsf{E}[|a|^4] + 8\sigma^2 \mathsf{E}[|a|^2] + 8\sigma^4}{\mathsf{E}[|a|^2] + 2\sigma^2} \tag{3.3.9}$$

We can show that

$$\lim_{\sigma \to 0} R^2 \longrightarrow \frac{\mathsf{E}\left[|a|^4\right]}{\mathsf{E}\left[|a|^2\right]} \tag{3.3.10}$$

We used MATLAB® for the evaluation of (3.3.6) and (3.3.9). Denoting $\tilde{y}$, σ and R_i with `y`, `s` and `a`, respectively, the following script was used to obtain (3.3.9):

```
syms g s a;
R2 =simple(int(y^5*exp(-(y^2+a^2)/2/s^2)*besseli(0,y*a/s^2),y,0,inf)/...
        int(y^3*exp(-(y^2+a^2)/2/s^2)*besseli(0,y*a/s^2),y,0,inf));
pretty(sort(R2))
>>
           4      2  2      4
          a  + 8 a  s  + 8 s
R2  =     -------------------
               2      2
              a  + 2 s
```

where $I_0(z) =$ `besseli(0,z)`.

In a noiseless environment, we can obtain a closed-form solution for dispersion constants by exploiting *Goupil and Palicot principle* [55], which states that *the dispersion constant R must be selected in such a way that the perfect equalizer is the minimum of the cost-function* $\mathsf{E}[\mathcal{J}]$, mathematically, it says

$$\mathsf{E}\left[\left. \frac{\partial \mathcal{J}(d \cdot a_n)}{\partial d} \right|_{(d=1)} \right] = 0, \tag{3.3.11}$$

where d is the overall system gain.

Note that the cCMA(p) update can directly be obtained by minimizing stochastically the following cost-function:

$$J = \frac{\mathsf{E}\left[|y_n|^{p+2}\right]}{p+2} - \frac{R^p \mathsf{E}\left[|y_n|^2\right]}{2} \tag{3.3.12}$$

Replacing y_n with $d \cdot a_n$ in the cost (3.3.12) and solving (3.3.11), we obtain

$$R^p = \frac{\mathsf{E}\left[|a|^{p+2}\right]}{\mathsf{E}\left[|a|^2\right]}. \tag{3.3.13}$$

which is consistent with (3.3.8) and (3.3.10) for $p = 1$ and $p = 2$, respectively.

3.4 Stochastic Stability of cCMA(p)

Note that, for higher value of p, the cCMA(p) requires a smaller step-size for its stability. In a noise-free scenario, however, the cCMA(p) can be ensured to adapt stochastically stable if $0 < \mu < \mu_{\text{bound}}$, where

$$\mu_{\text{bound}} = \frac{2\mathsf{E}\left[(p+2)|a|^p - 2R^p\right]}{\mathsf{E}\left[\left((p+2)|a|^p - 2R^p\right)^2\right]} \frac{1}{\mathrm{tr}[\boldsymbol{R}]} \tag{3.4.1}$$

and $\boldsymbol{R} = \mathsf{E}\left[\boldsymbol{x}_n \boldsymbol{x}_n^H\right]$ is the autocorrelation matrix. Note that, based on a recent study [36], we deduce that both CMA(p,2) and cCMA(p) are fundamentally always unstable (for all values of p) when the noise is not bounded (as, e.g., Gaussian noise).

3.4.1 Derivation of the Bound

Let $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\Phi(y,y^*)^*\boldsymbol{x}_n$ be a generic weight-update. Subtracting the zero-forcing solution $\boldsymbol{w}_{[*]}$ from both sides, we obtain $\overline{\boldsymbol{w}}_{n+1} = \overline{\boldsymbol{w}}_n - \mu\Phi(y,y^*)^*\boldsymbol{x}_n$, where $\overline{\boldsymbol{w}}_n = \boldsymbol{w}_{[*]} - \boldsymbol{w}_n$. For an algorithm in CM family, a generic error-function can be expressed as $\Phi(y,y^*)^* = y^* f(|y|)$ [11], where $f(|y|)$ is a real function about $|y|$. With $e_a = y_n - a_{n'}e^{\jmath\theta} = (-1)\overline{\boldsymbol{w}}_n^H \boldsymbol{x}_n$ being small enough, a first-order complex-valued Taylor series expansion of Φ can be written as [54]

$$\Phi(y,y^*)^* \approx \Phi(a,a^*)^* + e_a \frac{\partial \Phi(a,a^*)^*}{\partial y} + e_a^* \frac{\partial \Phi(a,a^*)^*}{\partial y^*} \tag{3.4.2}$$

Using *second-order odd-symmetry property* of QAM signal [79, Lemma 1], we can show that

$$\mathsf{E}\left[\frac{\partial \Phi(a,a^*)^*}{\partial y}\right] = \frac{1}{2}\mathsf{E}\left[\frac{f'(|a|)}{|a|}(a^*)^2\right] = 0 \tag{3.4.3}$$

Exploiting (3.4.2)-(3.4.3), we get

$$\overline{\boldsymbol{w}}_{n+1} \approx \overline{\boldsymbol{w}}_n - \mu\left(\Phi(a,a^*)^* + e_a^* \frac{\partial \Phi(a,a^*)^*}{\partial y^*}\right)\boldsymbol{x}_n \tag{3.4.4}$$

Taking the expected value of this expression, we can find a recursion for $\mathsf{E}[\overline{\boldsymbol{w}}_n]$. By virtue of the *Bussgang property,* we have $\mathsf{E}[\Phi(a,a^*)^*\boldsymbol{x}_n] = 0$ (refer to [17] for proof); this leads to

$$\mathsf{E}[\overline{\boldsymbol{w}}_{n+1}] \approx \mathsf{E}[\overline{\boldsymbol{w}}_n] + \mu\mathsf{E}\left[\frac{\partial\Phi(a,a^*)^*}{\partial y^*}\right]\mathsf{E}\left[\boldsymbol{x}_n\boldsymbol{x}_n^H\right]\mathsf{E}[\overline{\boldsymbol{w}}_n] \tag{3.4.5}$$

where we have assumed that the vectors $\boldsymbol{x}_n$ and $\boldsymbol{w}_n$ are independent of each other. The stability of recursion (3.4.5) requires

$$\mathsf{E}\left[\frac{\partial\Phi(a,a^*)^*}{\partial y^*}\right] = \mathsf{E}\left[\frac{1}{2}|a|f'(|a|) + f(|a|)\right] < 0 \tag{3.4.6}$$

as a necessary condition. Denoting $\varpi = (-1)\partial\Phi(a,a^*)^*/\partial y^*$, we express (3.4.5) as follows:

$$\mathsf{E}[\overline{\boldsymbol{w}}_{n+1}] \approx \left(\boldsymbol{I} - \mu\mathsf{E}[\varpi]\,\mathsf{E}\left[\boldsymbol{x}_n\boldsymbol{x}_n^H\right]\right)\mathsf{E}[\overline{\boldsymbol{w}}_n], \tag{3.4.7}$$

where $\boldsymbol{I}$ is the identity matrix. Following the classical reasoning in least-mean-squares adaptive filtering, an estimation of the time constant can be obtained as $\tau \approx (\mu\lambda_{\min}\mathsf{E}[\varpi])^{-1}$ and the algorithm converges in the *mean* if the step-size is selected such that $0 < \mu < (\lambda_{\max}\mathsf{E}[\varpi])^{-1}$, where $\lambda_{\min}$ and $\lambda_{\max}$ are respectively the minimum and maximum eigenvalues of the autocorrelation matrix $\boldsymbol{R} = \mathsf{E}\left[\boldsymbol{x}_n\boldsymbol{x}_n^H\right]$.

Multiplying (3.4.4) by its conjugate transpose and taking the expected value, we obtain a recursion for the autocorrelation $\boldsymbol{\Upsilon}_n \triangleq \mathsf{E}\left[\overline{\boldsymbol{w}}_n\overline{\boldsymbol{w}}_n^H\right]$:

$$\begin{aligned}\boldsymbol{\Upsilon}_{n+1} = \boldsymbol{\Upsilon}_n - \mu\mathsf{E}[\varpi]\,(\boldsymbol{R}\boldsymbol{\Upsilon}_n + \boldsymbol{\Upsilon}_n\boldsymbol{R}) + \mu^2\mathsf{E}[\chi]\,\boldsymbol{R}\\ + \mu^2\mathsf{E}\left[\varpi^2\right](\boldsymbol{R}\boldsymbol{\Upsilon}_n\boldsymbol{R} + \mathtt{tr}[\boldsymbol{R}\boldsymbol{\Upsilon}_n]\,\boldsymbol{R})\end{aligned} \tag{3.4.8}$$

where $\mathtt{tr}[\cdot]$ stands for the trace of the bracketed matrix and $\chi = |\Phi(a,a^*)|^2$. The evaluation of (3.4.8) assumes that the channel is long enough for the fourth-order moments of $\boldsymbol{x}_n$ to be well approximated by those of a Gaussian vector [46]. Now using the *Fisher's diagonalizing theorem* [46], we can transform (3.4.8) to a diagonalized-matrix difference equation

$$\begin{aligned}\boldsymbol{\Omega}_{n+1} = \boldsymbol{\Omega}_n - 2\mu\mathsf{E}[\varpi]\,\boldsymbol{\Lambda}\boldsymbol{\Omega}_n + \mu^2\mathsf{E}[\chi]\,\boldsymbol{\Lambda}\\ + \mu^2\mathsf{E}\left[\varpi^2\right]\left(\boldsymbol{\Lambda}^2\boldsymbol{\Omega}_n + \mathtt{tr}[\boldsymbol{\Lambda}\boldsymbol{\Omega}_n]\,\boldsymbol{\Lambda}\right)\end{aligned} \tag{3.4.9}$$

where, by using orthogonal transformation $\boldsymbol{U}$, we diagonalize $\boldsymbol{R}$ and $\boldsymbol{\Upsilon}$, such that, $\boldsymbol{U}^H\boldsymbol{R}\boldsymbol{U} = \boldsymbol{\Lambda}$ and $\boldsymbol{U}^H\boldsymbol{\Upsilon}\boldsymbol{U} = \boldsymbol{\Omega}$. Defining $\boldsymbol{\lambda} = \mathtt{diag}[\boldsymbol{\Lambda}] = [\lambda_1 \cdots \lambda_N]^T$ and $\boldsymbol{\omega}_n = \mathtt{diag}[\boldsymbol{\Omega}_n] = [\omega_{n,1} \cdots \omega_{n,N}]^T$ and equating the diagonal elements of the matrix on the left side of (3.4.9) with the corresponding diagonal elements of the matrix sum on the right side of this equality yields the vector difference equation

$$\omega_{n+1} = \mu^2 \mathsf{E}[\chi]\,\boldsymbol{\lambda} + \underbrace{\left[\boldsymbol{I} - 2\mu\mathsf{E}[\varpi]\,\boldsymbol{\Lambda} + \mu^2\mathsf{E}\left[\varpi^2\right]\left(\boldsymbol{\Lambda}^2 + \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right)\right]}_{\boldsymbol{F}}\omega_n \tag{3.4.10}$$

The convergence of ω_{n+1} depends on the matrix $\boldsymbol{F}$. It will converge if and only if the eigenvalues of $\boldsymbol{F}$ are all within the unit circle. Following the steps provided in [46] and ensuring eigenvalues lie within the unit circle, the range of step-size that guarantees stability of (3.4.10) is thus obtained as

$$0 < \mu < \frac{\mathsf{E}[\varpi]}{\mathsf{E}[\varpi^2]}\frac{1}{\mathrm{tr}[\boldsymbol{R}]} \tag{3.4.11}$$

The *mean-square* stochastic stability bound (3.4.11) generalizes the work in [99] in two aspects; firstly, we considered complex-valued quantities and (due to which) our result differs from the real-valued case in [99], and secondly, we presented the result for an arbitrary (constant modulus) Bussgang error-function. Also, comparing our result (3.4.11) with the bound evaluated in [40] for real-valued CMA(2,2), it is noticed that our evaluation procedure as well as the result (3.4.11) are noticeably much simple and meaningful. Moreover, the intermediate result (the time constant τ) can be noticed to be in agreement with [15] and further generalization to complex-valued CMA(2,2). In our case, we have $\varpi = \left(\frac{p+2}{2}\right)|a|^p - R^p$, requiring $\forall p \geq 1$,

$$(-1)\mathsf{E}[\varpi] = \frac{\mathsf{E}[|a|^{p+2}]}{\mathsf{E}[|a|^2]} - \left(\frac{p+2}{2}\right)\mathsf{E}[|a|^p] < 0 \tag{3.4.12}$$

which is always true for QAM signals due to their sub-Gaussian nature; substituting the values of $\mathsf{E}[\varpi]$ and $\mathsf{E}\left[\varpi^2\right]$, we can readily obtain (3.4.1).

3.5 Convergence Analysis of cCMA(p)

The convergence behavior of stop-and-go[1] (selective) update CMA$(2,2)$ has been studied by Rupp and Sayed [112]. They showed that for transmitted signals with constant modulus R, the equalizer implementing CMA$(2,2)$ is capable of making its outputs to lie within the circle of radius $R\sqrt{c}$ infinitely often, for some value of c that is slightly

[1]Notice that this stop-and-go principle is devised for the sake of convergence analysis in [112] and it has nothing to do with the conventional Bussgang-type stop-and-go adaptation strategy as appeared in [105] and [58].

larger than one. Due to the similarity between CMA$(p, 2)$ and cCMA(p), we intend to carry out this analysis for cCMA(p) to gain some insight into its convergence behavior. At the end, we would be able to show that a larger value of p in cCMA(p) has an advantageous effect of forcing c to come close to unity.

The corresponding active update steps of cCMA(p) have the form

$$\begin{cases} \text{if} \quad |y_n| \geq R\sqrt{c} \\ \text{then } k = k+1, \\ \qquad e_k = (R^p - |y_k|^p)\, y_k, \\ \qquad \boldsymbol{w}_{k+1} = \boldsymbol{w}_k + \mu_k e_k^* \boldsymbol{x}_k. \end{cases} \tag{3.5.1}$$

Assume we run the above algorithm infinitely often (i.e., $n \to \infty$), and let $\mathcal{K}$ denote the maximum number of active updates that occurred in the process. We now prove that, by properly designing the step-size sequence, $\mathcal{K}$ can be made finite, which in turn means that the condition $|y_n| < R\sqrt{c}$ will hold infinitely often. Let $\boldsymbol{w}_{[*]}$ denote the weight vector of the optimal equalizer and let $z_k = \boldsymbol{w}_{[*]}^H \boldsymbol{x}_k = a_{k'}$ is the optimal output so that $|z_k| = R$. Define further the *a priori* and *a posteriori* estimation errors

$$\begin{aligned} e_k^{\mathrm{a}} &= z_k - y_k = \overline{\boldsymbol{w}}_k^H \boldsymbol{x}_k \\ e_k^{\mathrm{p}} &= z_k - s_k = \overline{\boldsymbol{w}}_{k+1}^H \boldsymbol{x}_k \end{aligned} \tag{3.5.2}$$

where $\overline{\boldsymbol{w}}_k = \boldsymbol{w}_{[*]} - \boldsymbol{w}_k$. We introduce a complex-valued function $f[z_1, z_2]$:

$$f[z_1, z_2] \triangleq \frac{z_1|z_1|^p - z_2|z_2|^p}{z_1 - z_2}, \quad (z_1 \neq z_2) \tag{3.5.3}$$

Using $f[\cdot, \cdot]$ and some algebraic manipulations, we obtain

$$e_k = (f[z_k, y_k] - R^p)\, e_k^{\mathrm{a}} \tag{3.5.4}$$

$$e_k^{\mathrm{p}} = \left(1 - \frac{\mu_k}{\overline{\mu}_k}(f[z_k, y_k] - R^p)\right) e_k^{\mathrm{a}} \tag{3.5.5}$$

where $\overline{\mu}_k = 1/\|\boldsymbol{x}_k\|_2^2$ denotes the reciprocal of the input energy at the iteration k. We have pointed out that the *a posteriori* output s_n (or s_k) is closer to the blind estimate than *a priori* output y_n (or y_k), which requires that $|e_k^{\mathrm{p}}| < |e_k^{\mathrm{a}}|$. To ensure it, we need to select the step-size sequence μ_k so as to guarantee for all k

$$\left|1 - \frac{\mu_k}{\overline{\mu}_k}(f[z_k, y_k] - R^p)\right| < d < 1 \tag{3.5.6}$$

for all possible combinations of z_k and y_k, the value of p and for some positive scalar d. Let $f_R[z_k, y_k]$ and $f_I[z_k, y_k]$ denote the real and imaginary parts of $f[z_k, y_k]$, respectively; from (3.5.6) we obtain

$$\frac{\mu_k^2}{\overline{\mu}_k^2}\left(f_I[z_k, y_k]\right)^2 + \left(1 - \frac{\mu_k}{\overline{\mu}_k}\left(f_R[z_k, y_k] - R^p\right)\right)^2 < 1 \tag{3.5.7}$$

The values of μ_k for which (3.5.7) can be ensured, we have the following theorem:

Theorem 1 [Stop-and-Go cCMA(p)]: Assume y_k stays uniformly bounded from above for all k, say

$$R\sqrt{c} \leq |y_k| \leq PR < \infty \tag{3.5.8}$$

for some $P \geq \sqrt{c} > 1$. Choose a positive number β_o in the interval

$$\frac{36P^{2p} - \epsilon^2 m_o^{4/p}}{36P^{2p} + \epsilon^2 m_o^{4/p}} < \beta_o < 1. \tag{3.5.9}$$

and compute an α_o via

$$\alpha_o = \frac{6(1-\beta_o)P^p}{\epsilon\ m_o^{2/p}}. \tag{3.5.10}$$

Choose further the step-size μ_k for the active update from within the interval

$$\frac{2\,(1-\beta_o)}{\|x\|_2^2\, p\, \epsilon\, m_o^{2/p}\, R^p} < \mu_k < \frac{\alpha_o}{\|x\|_2^2\,(p+1)\, P^p\, R^p} \tag{3.5.11}$$

It then holds that $\mathcal{K} < \infty$. That is, $|y_n| < R\sqrt{c}$ holds infinitely often.

Proof: Let α and β be any two positive numbers satisfying

$$\alpha^2 + \beta^2 < 1 \tag{3.5.12}$$

We need to find a μ_k that satisfies

$$\left|\frac{\mu_k}{\overline{\mu}_k} f_I[z_k, y_k]\right| < \alpha \tag{3.5.13}$$

and

$$\left|1 - \frac{\mu_k}{\overline{\mu}_k}\left(f_R[z_k, y_k] - R^p\right)\right| < \beta \tag{3.5.14}$$

From [112], it is straightforward to prove that[2]

$$f_R[z_k, y_k] \geq R^p\left(1 + \epsilon\, m_o^{2/p}\right)^{p/2},$$
$$|f_I[z_k, y_k]| < R^p(1+p)P^p.$$

[2]Constants ϵ and m_o are related to c as given by $c \triangleq m_o^{-2/p} + \epsilon$; where $m_o = \min_{r \in (0,1)} \left(\frac{1 + r^{p+1}}{1+r}\right)$ and $0 < \epsilon \ll 1$.

We can satisfy (3.5.13) and (3.5.14) by selecting μ_k such that

$$\frac{1-\beta}{\|x\|_2^2}\frac{R^{-p}}{\left(1+\epsilon\, m_o^{2/p}\right)^{p/2}-1} < \mu_k < \frac{\alpha}{\|x\|_2^2}\frac{R^{-p}}{(p+1)P^p-1}$$

Notice that for $0<\epsilon \ll 1$ and $p \geq 1$, we can write

$$\left(1+\epsilon\, m_o^{2/p}\right)^{p/2} \approx 1+\frac{p}{2}\epsilon\, m_o^{2/p}$$

Also notice that $P \geq \sqrt{c} > 1$ and $p \geq 1$ give

$$\frac{1}{(p+1)P^p} < \frac{1}{(p+1)P^p-1}$$

we found a simpler bound on μ_k as given by

$$\frac{2\,(1-\beta)}{\|x\|_2^2\, p\,\epsilon\, m_o^{2/p}\, R^p} < \mu_k < \frac{\alpha}{\|x\|_2^2\,(p+1)\,P^p\,R^p} \tag{3.5.15}$$

which gives

$$\frac{1-\beta}{\alpha} < \frac{p\epsilon\, m_o^{2/p}}{2(p+1)P^p} < \frac{\epsilon\, m_o^{2/p}}{2P^p} \tag{3.5.16}$$

Let for some $\{\alpha_o, \beta_o\}$ we have

$$\frac{1-\beta_o}{\alpha_o} = \frac{1}{3}\cdot\frac{\epsilon\, m_o^{2/p}}{2P^p} = \frac{\epsilon\, m_o^{2/p}}{6P^p} \tag{3.5.17}$$

We obtain the value of α_o as given by

$$\alpha_o = \frac{6(1-\beta_o)P^p}{\epsilon\, m_o^{2/p}} \tag{3.5.18}$$

Then, $\{\alpha_o, \beta_o\}$ satisfy (3.5.16). Substituting into (3.5.12), we see that β_o must be such that

$$\left(\frac{6(1-\beta_o)P^p}{\epsilon\, m_o^{2/p}}\right)^2 + (\beta_o)^2 < 1$$

If we find a β_o that satisfies this inequality, then a pair of $\{\alpha_o, \beta_o\}$ satisfying (3.5.12) and (3.5.16) exists. So consider the following quadratic function

$$g(\beta) = \left(\frac{6(1-\beta)P^p}{\epsilon\, m_o^{2/p}}\right)^2 + (\beta)^2 - 1.$$

It has a negative minimum and it crosses the real axis at the positive roots

$$\beta^{(1)} = \frac{36P^{2p} - \epsilon^2 m_o^{4/p}}{36P^{2p} + \epsilon^2 m_o^{4/p}} < 1, \qquad \beta^{(2)} = 1.$$

Hence, β_o can be chosen as any value in the interval

$$\frac{36P^{2p} - \epsilon^2 m_o^{4/p}}{36P^{2p} + \epsilon^2 m_o^{4/p}} < \beta_o < 1. \tag{3.5.19}$$

For $p = 2$ and $m_o = 3/4$, the above result (3.5.19) can be found consistent with [112, Equation (51)]. The bounds on μ_k are thus justified. □□□

Remarks: From Theorem 1, we can say that, for suitably chosen step-sizes, the stop-and-go cCMA(p) produces a sequence of estimates y_k that lies inside the circle of radius $R\sqrt{c}$ with probability one. We notice that the choice of μ_k is small and it becomes even smaller for large values of p. However at the same time, a larger p is beneficial in making c come close to unity. We have

$$c \triangleq \left[\min_{r\in(0,1)} \left(\frac{1+r^{p+1}}{1+r}\right)\right]^{-\frac{2}{p}} + \epsilon \tag{3.5.20}$$

Notice in Table 3.1, the corresponding values of c are decreasing monotonically and approaching unity with an increase in p.

Table 3.1: Values of p and c

p	1	2	3	4	5	6
$c-\epsilon$	1.4571	1.3333	1.2635	1.2185	1.1868	1.1634
p	7	8	9	10	11	∞
$c-\epsilon$	1.1452	1.1308	1.1190	1.1092	1.1009	1

3.6 ISI Mitigation Capability of cCMA(p)

In this section, we analyze the effect of parameter p on the ISI mitigation capability of cCMA(p). Let $w_{[*]}$ be the zero-forcing solution of the blind equalization problem, such that $w_{[*]} = \arg\min_w \mathsf{E}_y[\mathcal{J}]$ subject to $\|\{w_{[*]}^* \circledast h\}\|_2^2 = 1$, where $h = \{h\}$ is the channel impulse response. Let w_n be a stochastic approximation estimate based on the particular (finite) realization of data y_n at time index n, such that $w_n = \arg\min_w \mathcal{J}$ for which $\|t_n\|_2^2 = \|\{w_n^* \circledast h\}\|_2^2 \neq 1$. The equalization quality in terms of residual ISI can be expressed as [24, Eq. (14)]

$$\text{ISI} \approx \frac{\mathsf{E}\left\|\left\{w_{[*]}^* \circledast h \circledast a - w_n^* \circledast h \circledast a\right\}\right\|_2^2}{\mathsf{E}\left\|\{h \circledast a\}\right\|_2^2} \tag{3.6.1}$$

where $\{h \circledast a\}$ is the received sequence. Defining $\overline{w}_n = w_{[*]} - w_n$ as an error in the estimation of equalizer coefficients, we further obtain [24, Eq. (11)]

$$\mathrm{ISI} \approx \mathsf{E}\|\overline{w}_n\|_2^2 = N \cdot \mathtt{var}(\overline{w}) = \mathtt{tr}\big[\mathsf{E}\left[\overline{w}_n \overline{w}_n^H\right]\big] \tag{3.6.2}$$

Assuming $\|\overline{w}\|_2^2$ is small and using second-order Taylor expansion; the linearization of $\nabla_{w^*} J$ around w, gives

$$\nabla_{w^*}\mathcal{J} \equiv \left.\frac{\partial \mathcal{J}}{\partial w^*}\right|_{w=w_n} \approx \underbrace{\left.\frac{\partial \mathcal{J}}{\partial w^*}\right|_{w=w_{[*]}}}_{\approx 0} + \left(w_n - w_{[*]}\right) \left.\frac{\partial^2 \mathcal{J}}{\partial w^* \, \partial w^T}\right|_{w=w_n} \tag{3.6.3}$$

Equation (3.6.3) can concisely be written as:

$$\nabla \mathcal{J} \approx -\overline{w}_n \nabla^2 \mathcal{J} \tag{3.6.4}$$

Therefore, the error covariance matrix $\mathsf{E}\left[\overline{w}_n \overline{w}_n^H\right]$ can be approximated as

$$\mathsf{E}[\overline{w}_n \overline{w}_n^H] \approx \left(\mathsf{E}\left[\nabla^2 \mathcal{J}\right]\right)^{-1} \cdot \left(\mathsf{E}\left[\nabla \mathcal{J} \nabla \mathcal{J}^H\right]\right) \cdot \left(\mathsf{E}[\nabla^2 \mathcal{J}]\right)^{-H} \tag{3.6.5}$$

Under perfect signal recovery assumption, $y_n = a_n$,

$$\begin{aligned} \mathsf{E}\left[\nabla_{w_k^*}\mathcal{J}(\nabla_{w_l^*}\mathcal{J})^*\right] &= \mathsf{E}[\Phi(a,a^*)^* \, x_{n-l} \, x_{n-k}^* \, \Phi(a,a^*)], \\ &= \begin{cases} \mathsf{E}\left[|\Phi(a,a^*)|^2\right] \cdot \mathsf{E}\left[|x_{n-k}|^2\right], & \text{for } k = l \\ \mathsf{E}\left[|\Phi(a,a^*)|^2\right] \cdot \mathsf{E}\left[x_{n-l} \, x_{n-k}^*\right], & \text{for } k \neq l \end{cases} \end{aligned} \tag{3.6.6}$$

Next we find

$$\begin{aligned} \mathsf{E}\left[\nabla^2_{w_k w_l^*}\mathcal{J}\right] &= \mathsf{E}[\Phi'(a,a^*) \, x_{n-l} \, x_{n-k}^*], \\ &= \begin{cases} \mathsf{E}\left[\Phi'(a,a^*)^*\right] \cdot \mathsf{E}\left[|x_{n-k}|^2\right], & \text{for } k = l \\ \mathsf{E}\left[\Phi'(a,a^*)^*\right] \cdot \mathsf{E}\left[x_{n-l} \, x_{n-k}^*\right], & \text{for } k \neq l \end{cases} \end{aligned} \tag{3.6.7}$$

Combining Equations (3.6.2), (3.6.5)-(3.6.7), we obtain

$$\mathrm{ISI} \approx \frac{\mathsf{E}\left[|\Phi(a,a^*)|^2\right]}{\left|\mathsf{E}\left[\Phi'(a,a^*)^*\right]\right|^2} \, \mathtt{tr}\big[R^{-1}\big] \tag{3.6.8}$$

where $R = \mathsf{E}\left[x_n x_n^H\right]$. Using (3.6.8), we obtain the ISI expression for cCMA(p) as follows:

$$\mathrm{ISI} = \frac{\mathsf{E}\left[(R^p - |a|^p)^2 \, |a|^2\right]}{\left(\left(1 + \frac{p}{2}\right) \mathsf{E}\left[|a|^p\right] - R^p\right)^2} \, \mathtt{tr}\big[R^{-1}\big] \tag{3.6.9}$$

Substituting the value $R^p = \mathsf{E}[|a|^{p+2}]/\mathsf{E}[|a|^2]$ in (3.6.9), we obtain

$$\text{ISI} = \frac{\mathsf{E}\left[|a|^{2p+2}\right]\mathsf{E}\left[|a|^2\right] - \mathsf{E}^2\left[|a|^{p+2}\right]}{\left((1+\frac{p}{2})\,\mathsf{E}\left[|a|^p\right]\mathsf{E}\left[|a|^2\right] - \mathsf{E}\left[|a|^{p+2}\right]\right)^2}\,\mathtt{tr}\left[\boldsymbol{R}^{-1}\right] \;\equiv\; g(p)\,\mathtt{tr}\left[\boldsymbol{R}^{-1}\right] \tag{3.6.10}$$

The above asymptotic performance result is analogous to those in [39, 115, 117, 23]. In Fig. 3.1, we depict the values of metric $g(p)$ against p for some QAM/APSK signals. Note that the metric $g(p)$ has a consistent decreasing trend for larger values of parameter p.

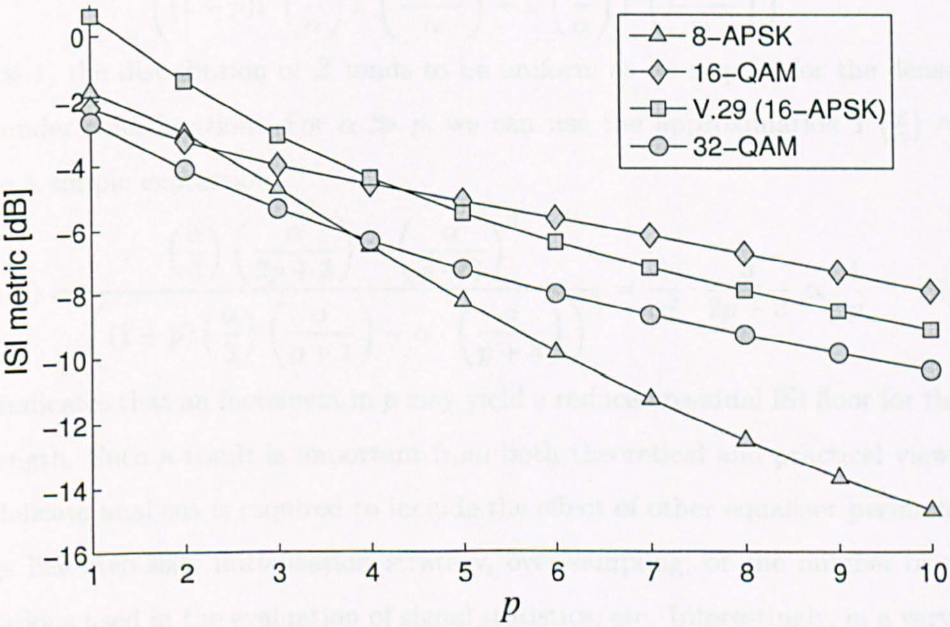


Figure 3.1: ISI metric $g(p)$ vs p for some QAM/APSK signals.

An interesting result may be obtained from (3.6.10), if we consider a real-valued cCMA(p). For a real-valued case, we can show

$$g(p) = \frac{\mathsf{E}\left[a^{2p+2}\right]\mathsf{E}\left[a^2\right] - \mathsf{E}^2\left[a^{p+2}\right]}{\left((1+p)\,\mathsf{E}\left[a^p\right]\mathsf{E}\left[a^2\right] - \mathsf{E}\left[a^{p+2}\right]\right)^2} \tag{3.6.11}$$

Note that the factor $(1+\frac{p}{2})$ is replaced with $(1+p)$ and $|a|$ is replaced with a. Now consider a large (dense) PAM signal, where the PDF of a may be considered to be continuous and uniformly distribution. To illustrate an asymptotic performance for such PAM signal, we consider the particular case of the generalized Gaussian distribution. A random variable Z is said to obey the generalized Gaussian distribution with parameters $\alpha, \beta > 0$ if its PDF is given by

$$p_Z(z) = \frac{\alpha}{2\beta}\exp\left(-\frac{|z|^\alpha}{\beta^\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)^{-1} \tag{3.6.12}$$

where α and β are respectively (shape and size parameters, and $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$ is the Euler Gamma function. It can also be shown that

$$\mathsf{E}[|Z|^p] = \beta^p \Gamma\left(\frac{p+1}{\alpha}\right) \Gamma\left(\frac{1}{\alpha}\right)^{-1} \tag{3.6.13}$$

Exploiting (3.6.13), we obtain the following expression for $g(p)$ in (3.6.11):

$$g(p) = \frac{\Gamma\left(\frac{3}{\alpha}\right)\Gamma\left(\frac{2p+3}{\alpha}\right) - \Gamma^2\left(\frac{p+3}{\alpha}\right)}{\left((1+p)\Gamma\left(\frac{3}{\alpha}\right)\Gamma\left(\frac{p+1}{\alpha}\right) - \Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{p+3}{\alpha}\right)\right)^2} \tag{3.6.14}$$

For $\alpha \gg 1$, the distribution of Z tends to be uniform as we require for the dense PAM signal under consideration. For $\alpha \gg p$, we can use the approximation $\Gamma\left(\frac{p}{\alpha}\right) \approx \frac{\alpha}{p}$; it leads to a simple expression

$$g(p) = \frac{\left(\frac{\alpha}{3}\right)\left(\frac{\alpha}{2p+3}\right) - \left(\frac{\alpha}{p+3}\right)^2}{\left((1+p)\left(\frac{\alpha}{3}\right)\left(\frac{\alpha}{p+1}\right) - \alpha \cdot \left(\frac{\alpha}{p+3}\right)\right)^2} = \frac{1}{\alpha^2} \cdot \frac{3}{2p+3} \propto \frac{1}{p} \tag{3.6.15}$$

which indicates that an increment in p may yield a reduced residual ISI floor for the given filter length. Such a result is important from both theoretical and practical viewpoints. More delicate analysis is required to include the effect of other equalizer perimeters and settings like step-size, initialization strategy, over-sampling, or the number of channel observations used in the evaluation of signal statistics, etc. Interestingly, in a very recent article [106], author has carried out a different but a detailed analysis, incorporated the step-size parameter and finally obtained a closed-form steady-state ISI expression for Bussgang-type *adaptive* blind equalization algorithms.

3.7 Simulation Results

3.7.1 Experiment 1: ISI Performances of CMA(p, 2) and cCMA(p)

We evaluate ISI performances of CMA(p, 2) and cCMA(p) for various values of p. We use a complex-valued seven-tap equalizer and initialize it so that the center tap is set to one and other taps are set to zero. The propagation channel is a (short) voice-band seven-tap telephone channel and is taken from [105]. The signal to noise ratio (SNR) is taken as 30dB at the input of the equalizer. The residual intersymbol interference (ISI) [121] is measured for an 8-APSK signal and compared. The ISI is defined as

$$\mathrm{ISI} = \frac{\sum_l |t_l|^2}{\max\left(\{|t_l|^2\}\right)} - 1, \tag{3.7.1}$$

where $\{t_l\}$ is the overall channel-equalizer impulse-response. Signal alphabets belong to the set $\{\pm 1 \pm \jmath, \frac{1}{\sqrt{2}}(1+\sqrt{3})(\pm 1 \pm \jmath)\}$. Each ISI trace is the ensemble average of 400 independent runs with random initialization of noise and data source. Figure 3.2 depicts the converged constellations for various values of p in cCMA(p); note that the equalized symbols are more aggregated for larger p. Figure 3.3 depicts the residual ISI performances of CMA(p, 2) and cCMA(p) and also shows the values of step-sizes used in the simulations. Note that the performance gets better in terms of steady-state residual ISI when larger p is used in both cases. The CMA(p, 2) yielded stable performance for $p = 1, 2, \cdots, 4$ and it failed to give any stable convergence for $p > 5$. The convergence behavior of cCMA(p) is even more attractive in the sense that it provided a smooth tradeoff between the complexity and performance; we can go up to $p = 9$ in this specific experiment.

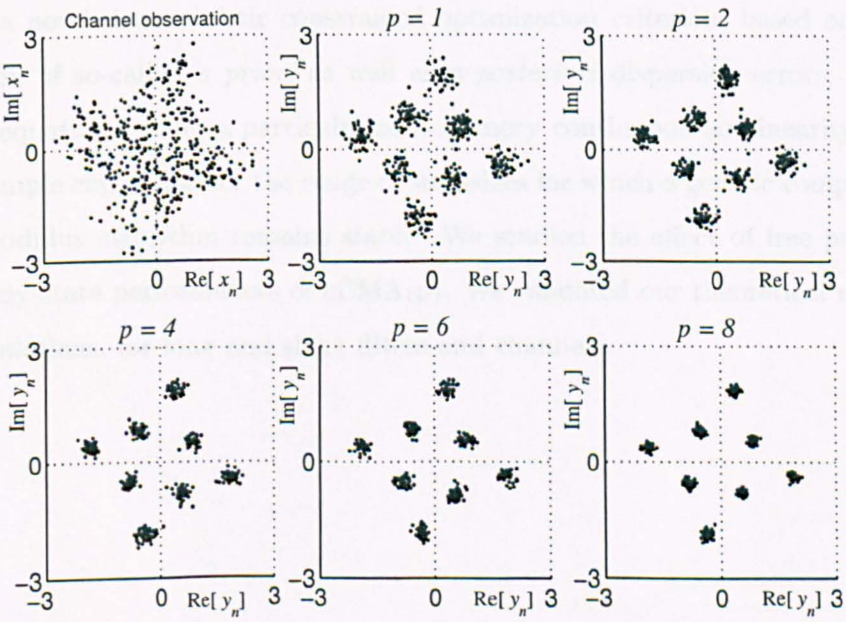


Figure 3.2: Scatter plots for cCMA(p)

3.7.2 Experiment 2: Validating the Stability Bound for cCMA(p)

In this experiment, we validate the upper-bound (3.4.1) for $p = 1$ and 2. In addition to the channel we used in Fig. 3.3, we also consider the first thirty odd-indexed coefficients of a (long) microwave terrestrial channel (`chan2.mat`) taken from SPIB database [1]. In all cases, the simulations were performed with 5000 iterations, $N_{\text{run}} = 400$ runs, and no noise. In Fig. 3.4-3.5, we plot the probabilities of divergence P_{div} for four different equalizer lengths, against the normalized step-size, $\mu_{\text{norm}} = \mu/\mu_{\text{bound}}$. The P_{div} is estimated as $P_{\text{div}} = N_{\text{div}}/N_{\text{run}}$, where N_{div} indicates the number of times equalizer diverged. In our simulations, we label a given run of the algorithm as "diverging" if y_n overflows. Equalizers were initialized as zero-forcing solution and step-sizes were varied in the range $0.5\mu_{\text{bound}} < \mu < 2.0\mu_{\text{bound}}$. *It can be seen that the approximate bound does guarantee a stable performance when $\mu < \mu_{\text{bound}}$, equalizer and channel are long enough, and p is small.*

3.8 Summary

A new constant modulus algorithm, cCMA(p), has been presented for blind equalization of complex-valued communication channels. The proposed algorithm was obtained by solving a novel deterministic constrained optimization criterion, based on the joint minimization of so-called *a priori* as well as *a posteriori* dispersion errors, leading to an update equation having a particular zero-memory continuous nonlinearity. We also derived a simple expression for the range of step-sizes for which a generic complex-valued constant-modulus algorithm remains stable. We studied the effect of free parameter p on the steady-state performances of cCMA(p). We validated our theoretical model with several simulations, for long and short filters and channels.

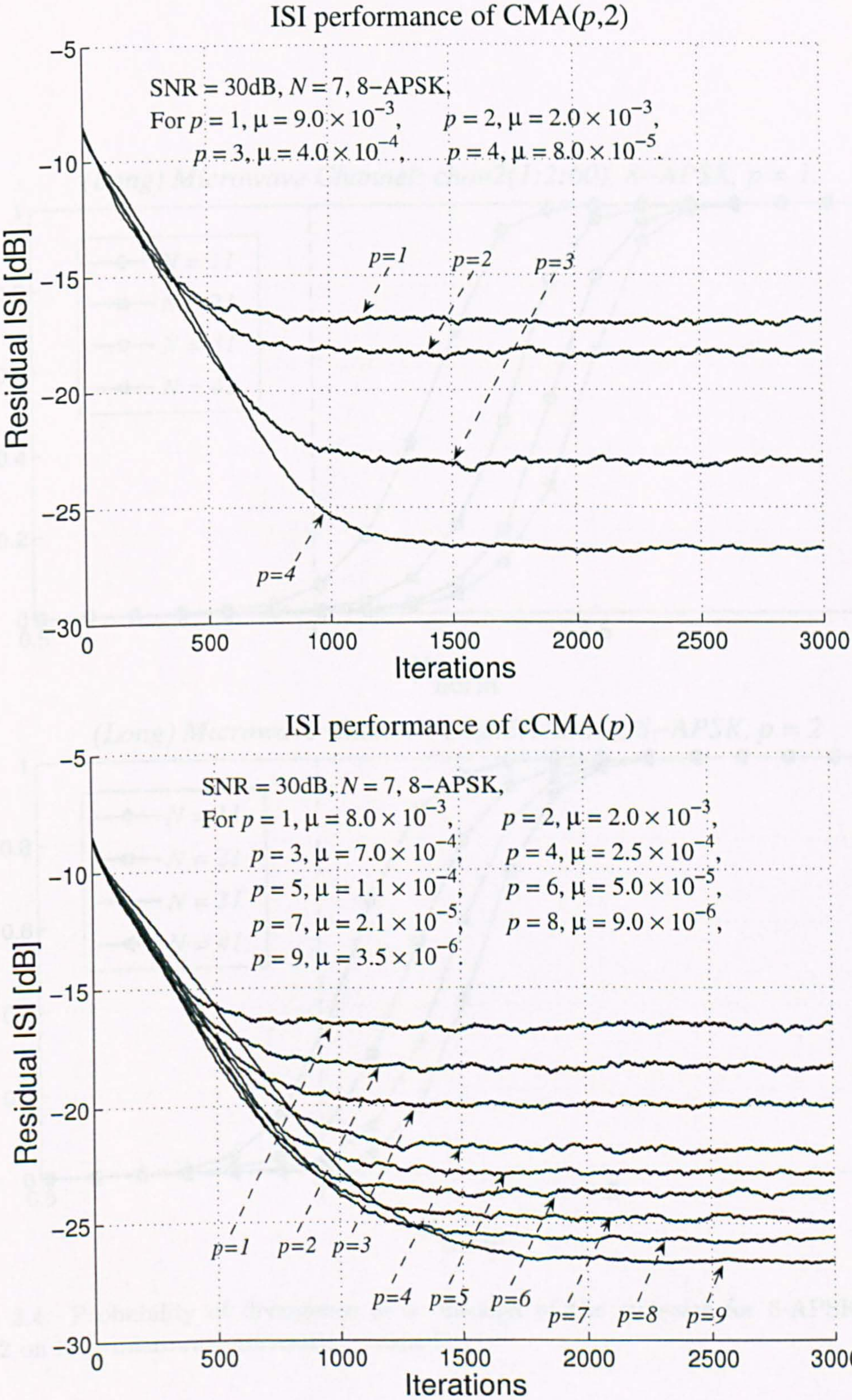


Figure 3.3: Residual ISI traces: (top) CMA$(p,2)$ and (bottom) cCMA(p).

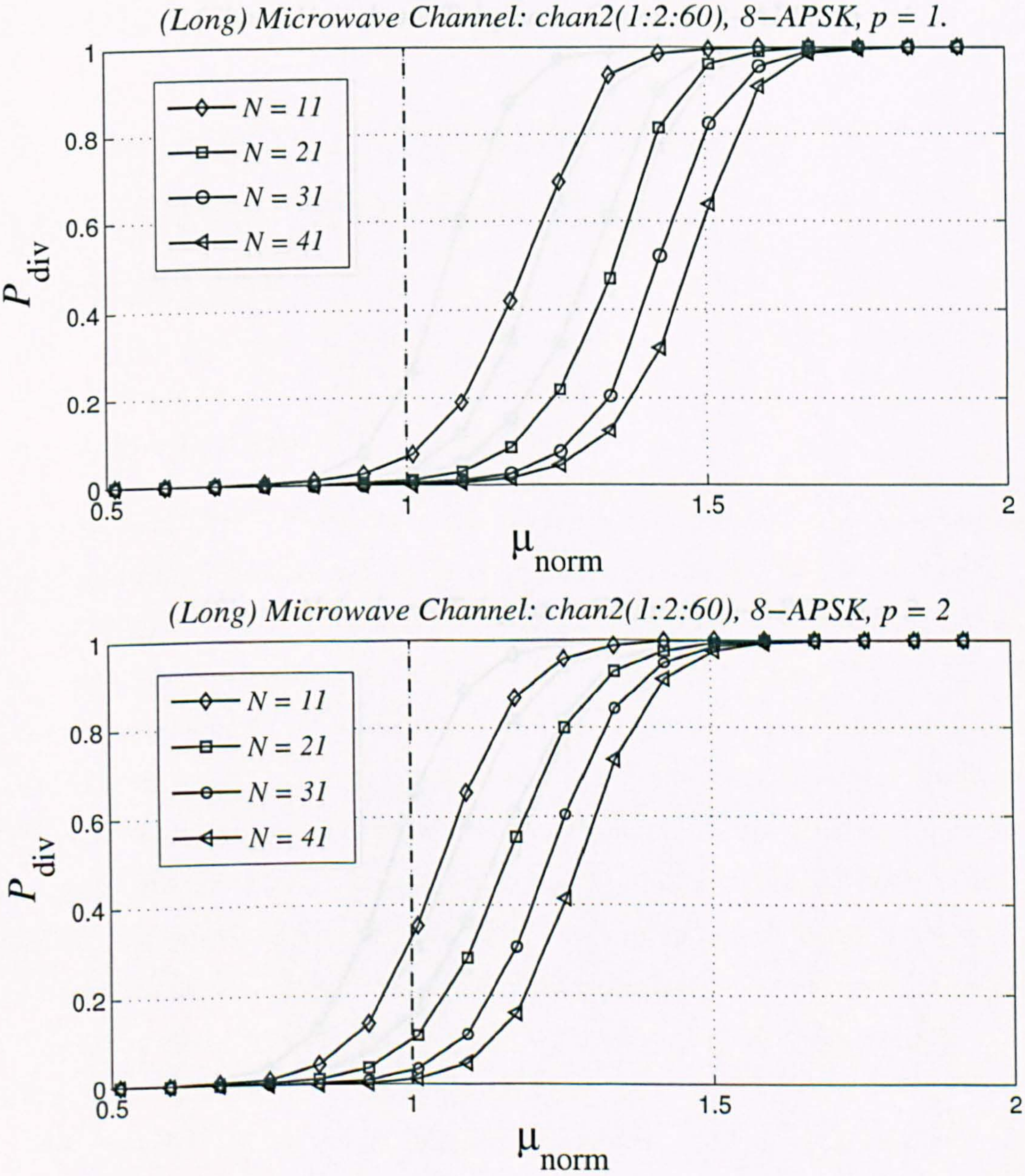


Figure 3.4: Probability of divergence as a function of the step-size for 8-APSK and $p = 1, 2$ on long microwave terrestrial channel.

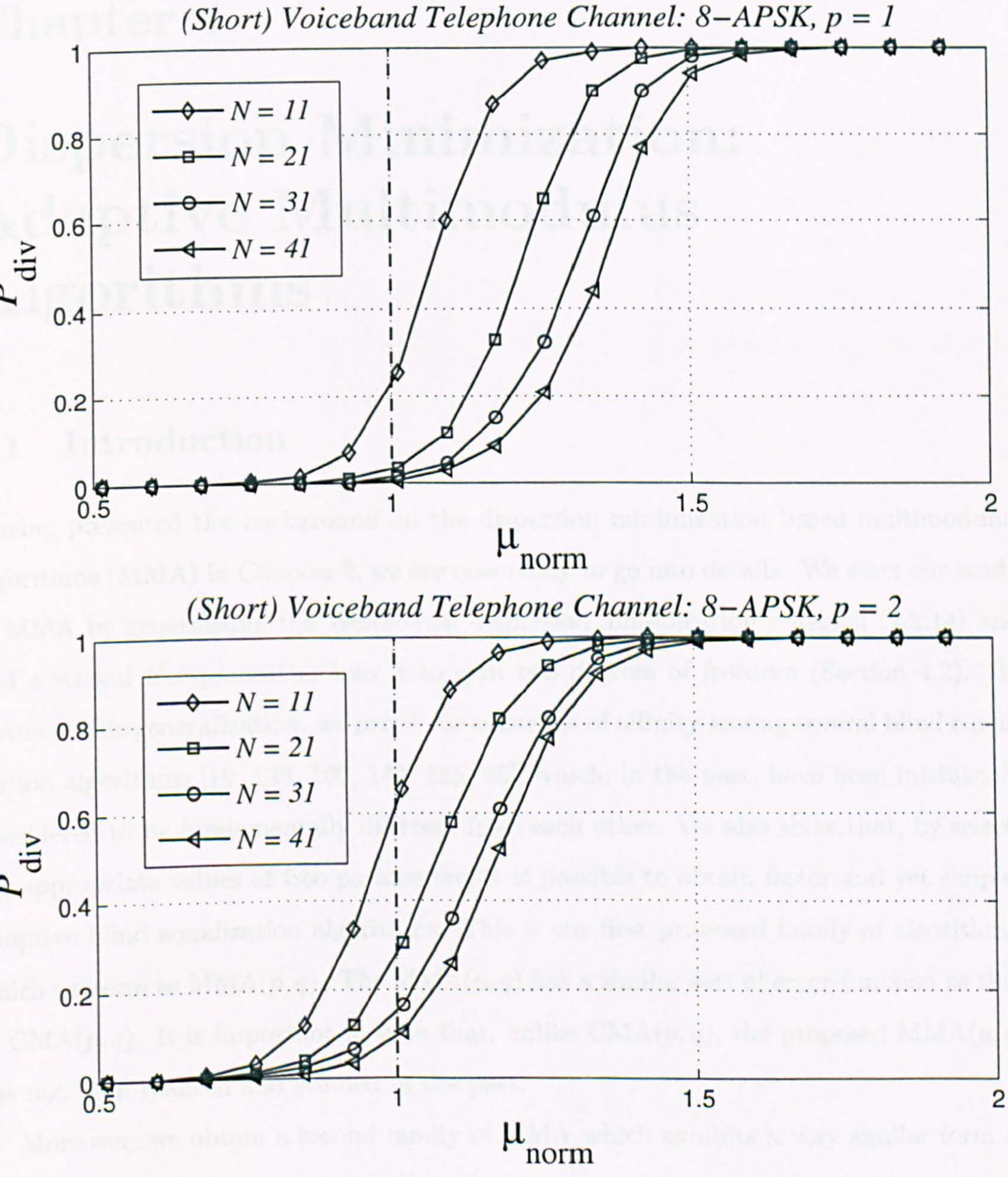


Figure 3.5: Probability of divergence as a function of the step-size for 8-APSK and $p = 1, 2$ on short voice-band telephonic channel.

Chapter 4

Dispersion Minimization: Adaptive Multimodulus Algorithms

4.1 Introduction

Having presented the background on the dispersion minimization based multimodulus algorithms (MMA) in Chapter 2, we are now ready to go into details. We start our study in MMA by generalizing the Wesolowski dispersion minimization criterion (2.2.14) and add a second free parameter into it to gain two degrees of freedom (Section 4.2). By virtue of this generalization, we prove the existence of affinity among several blind equalization algorithms [19, 136, 102, 140, 135, 65], which, in the past, have been mistakenly considered to be fundamentally different from each other. We also show that, by selecting appropriate values of free parameters, it is possible to obtain faster and yet simpler adaptive blind equalization algorithms. This is our first proposed family of algorithms, which we term as MMA(p, q). The MMA(p, q) has a similar sort of error-function as that of CMA(p, q). It is important to note that, unlike CMA(p, q), the proposed MMA(p, q) has not been realized and studied in the past.

Moreover, we obtain a second family of MMA which exhibits a very similar form as that of cCMA(p). To appreciate this similarity, we term the second family of algorithms as cMMA(p). However, unlike cCMA(p), cMMA(p) is obtained by solving a very different optimization problem. In essence, cMMA(p) is obtained by modifying an existing convex cost-function which was suggested by Kennedy and Ding [69] for square-QAM signals. We demonstrate that cMMA(p) is not only capable of achieving successful equalization but is also capable of recovering carrier-phase and restoring true signal energy without needing extra mechanisms or hardware, which are conventionally required by the original Kennedy-Ding algorithm [69]. We explore that cMMA(p) can provide a consistent trade-off between ISI mitigation and computational complexity. We also highlight the existence of a close affinity among cMMA(p) and some existing blind equalization algorithms [95, 8, 6].

In this Chapter, we provide detailed evaluation of dispersion constants for MMA(p, q) and cMMA(p) (Section 4.4). We analytically explore their carrier-phase recovery capabilities (Section 4.2.1 and 4.3.1). And most importantly, we provide dynamic convergence analysis to evaluate the MSE/ISI convergence behavior of the existing as well as proposed MMA equalizers (Section 4.5). A number of Monte-Carlo experiments are provided on several square-QAM signals to validate our analytical findings (Section 4.6).

4.2 The First Proposed Family of Algorithms: MMA(p, q)

Here we present the first of the two proposed families of MMA. For this particular case, we gained motivation from different generalization of Bussgang blind equalization algorithms existing in literature. For example, a) generalization of Sato's [114] cost-function for real-valued signals by Serra and Esteves [118], b) generalization of Godard's constant modulus criterion [53] with two degrees of freedom by Larimore and Treichler [73] and, more recently, work on *non-circular generalized modulus* cost-functions by Li and Zhang [76] and Goupil and Palicot [55] have been a driving force for this present idea.

We propose the following generalized dispersion minimization based (split) cost-function:

$$J_{\mathrm{I}} = \mathsf{E}\left[\left||y_{R,n}|^p - R_R^p\right|^q\right] + \mathsf{E}\left[\left||y_{I,n}|^p - R_I^p\right|^q\right], \tag{4.2.1}$$

where p and q are free (positive) parameters. Cost (4.2.1) can be considered as the generalization of *Wesolowski's* cost-function [136] with two degrees of freedom or the split version of Larimore and Treichler constant modulus cost-function [73]. Note that none of the existing generalizations [73, 76, 55] can be represented in the split form as expressed in (4.2.1). The corresponding stochastic gradient-based adaptive algorithm for the cost-function (4.2.1) is given by

$$\begin{aligned} w_{n+1} = w_n + \mu \Big[\left| \left| y_{R,n}^p \right| - R_R^p \right|^{q-2} \left| y_{R,n}^{p-2} \right| \left(R_R^p - \left| y_{R,n}^p \right| \right) y_{R,n} \\ - \jmath \left| \left| y_{I,n}^p \right| - R_I^p \right|^{q-2} \left| y_{I,n}^{p-2} \right| \left(R_I^p - \left| y_{I,n}^p \right| \right) y_{I,n} \Big] x_n \end{aligned} \tag{4.2.2}$$

where R_R and R_I are chosen in accordance with Bussgang-condition using the statistics of transmitted data (and noise). Clearly, a multitude of algorithms can be obtained for different choices of p and q, providing a possible flexibility in the design of blind equalizers. In the sequel, we will refer to (4.2.2) as MMA(p, q). Now we show that how (4.2.2) generalizes a number of existing adaptive blind equalization algorithms.

1. For $p = 1, q = 2$, the cost (4.2.1) reduces to an equivalent form of Benveniste-Goursat cost-function (1984) [19]. We denote the resulting algorithm as MMA(1,2), *viz*

$$w_{n+1} = w_n + \mu \left[\left(R_R \text{sgn} \left[y_{R,n} \right] - y_{R,n} \right) - \jmath \left(R_I \text{sgn} \left[y_{I,n} \right] - y_{I,n} \right) \right] x_n \tag{4.2.3}$$

2. For $p = q = 1$, the cost (4.2.1) reduces to an equivalent form of the cost-function independently proposed by Weerackody *et al.* (1991) [135] and Im *et al.* (2001) [65]. We denote the resulting algorithm as MMA(1,1), *viz*

$$w_{n+1} = w_n + \mu [\text{sgn} \left[R_R \text{sgn} \left[y_{R,n} \right] - y_{R,n} \right] - \jmath \, \text{sgn} \left[R_I \text{sgn} \left[y_{I,n} \right] - y_{I,n} \right]] \, x_n \tag{4.2.4}$$

3. For $p = q = 2$, the cost simplifies to the cost-function independently proposed by Wesolowski (1987) [136], Chin and Oh (1995) [102], and Yang (1997) [138]. We denote the resulting algorithm as MMA(2,2), *viz*

$$w_{n+1} = w_n + \mu \left[\left(R_R^2 - y_{R,n}^2 \right) y_{R,n} - \jmath \left(R_I^2 - y_{I,n}^2 \right) y_{I,n} \right] x_n \tag{4.2.5}$$

In fact, by considering relevant values of p and q, some new blind equalization algorithms can be obtained. An interesting choice is $(p = 2, q = 1)$, and the resulting MMA$(2, 1)$ update is given by

$$w_{n+1} = w_n + \mu \left[\text{sgn} \left[R_R^2 - y_{R,n}^2 \right] y_{R,n} - \jmath \, \text{sgn} \left[R_I^2 - y_{I,n}^2 \right] y_{I,n} \right] x_n, \tag{4.2.6a}$$

$$= w_n + \mu \left[\text{sgn} \left[R_R - |y_{R,n}| \right] y_{R,n} - \jmath \, \text{sgn} \left[R_I - |y_{I,n}| \right] y_{I,n} \right] x_n. \tag{4.2.6b}$$

Similarly, for $(p=3, q=1)$, we obtain MMA(3,1),

$$w_{n+1} = w_n + \mu \left[\operatorname{sgn}\left[R_R^3 - |y_{R,n}^3|\right] |y_{R,n}|\, y_{R,n} - \jmath \operatorname{sgn}\left[R_I^3 - |y_{I,n}^3|\right] |y_{I,n}|\, y_{I,n} \right] x_n, \tag{4.2.7a}$$

$$= w_n + \mu \left[\operatorname{sgn}\left[R_R - |y_{R,n}|\right] |y_{R,n}|\, y_{R,n} - \jmath \operatorname{sgn}\left[R_I - |y_{I,n}|\right] |y_{I,n}|\, y_{I,n} \right] x_n. \tag{4.2.7b}$$

The expressions (4.2.6b) and (4.2.7b) are respectively simplified versions of (4.2.6a) and (4.2.7a).

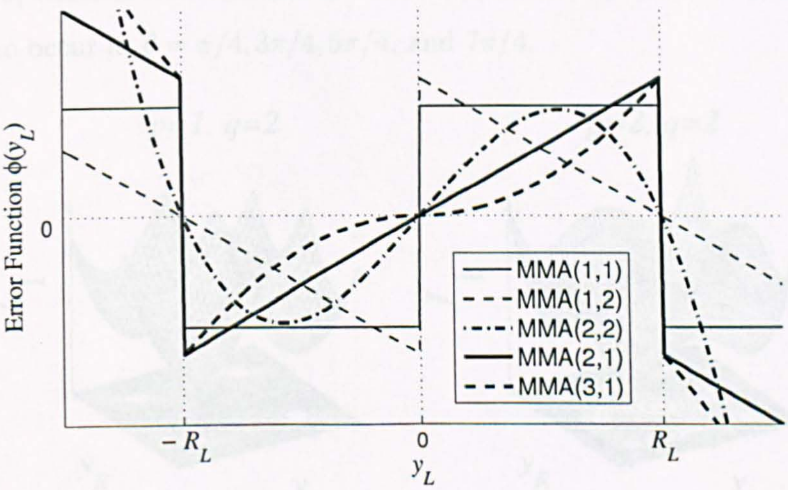


Figure 4.1: Plots of MMA(p, q) error-functions.

Note that the cost (4.2.1) or the updates (4.2.6)-(4.2.7) are not known to have been realized in the past. A possible reason is that, in the past, the existing algorithms (4.2.3) and (4.2.5) have been mistakenly believed to be different from each other. For example, Yang [138] described that the algorithms (4.2.3) and (4.2.5) minimize the dispersion in equalized sequence away from *four points* and *four lines*, respectively. Later, this description propagated to many subsequent important publications [141, 49, 140]. It is just recently that Thaiupathump [126] proved this description was *wrong* and established that the cost-functions of both algorithms exhibit four-point zero-error contour. In Figure 4.1, we depict the error-function $\phi(y_L) = ||y_L^p| - R_L^p|^{q-2}|y_L^{p-2}|(R_L^p - |y_L|^p)y_L$ of MMA(p, q) for an arbitrary square-QAM. Note that all of these error-functions satisfy the properties as discussed in Section 2.1.2.

4.2.1 MMA(p, q): Phase Recovery Capability

In Fig. 4.2, we depict the mesh plots of MMA(p, q) cost-surfaces for four exemplary members. It is clear that MMA(p, q) cost-surface exhibits consecutive four minima (or maxima) $\pi/2$ radians apart. In fact, it is an important feature which enables these costs to fix the phase-offset introduced by the channel with an ambiguity of $\pm 90°$ or its integer multiples. Suppose θ is a residual phase-offset error (in the absence of noise and ISI), it gives $y_n = a_{n'} \exp(\jmath\theta)$. Due to four-quadrant symmetry of square-QAM, it is desirable that the cost J_I exhibits local minima at $\theta = 0, \pi/2, \pi$ and $3\pi/2$; similarly, local maxima are required to occur at $\theta = \pi/4, 3\pi/4, 5\pi/4$, and $7\pi/4$.

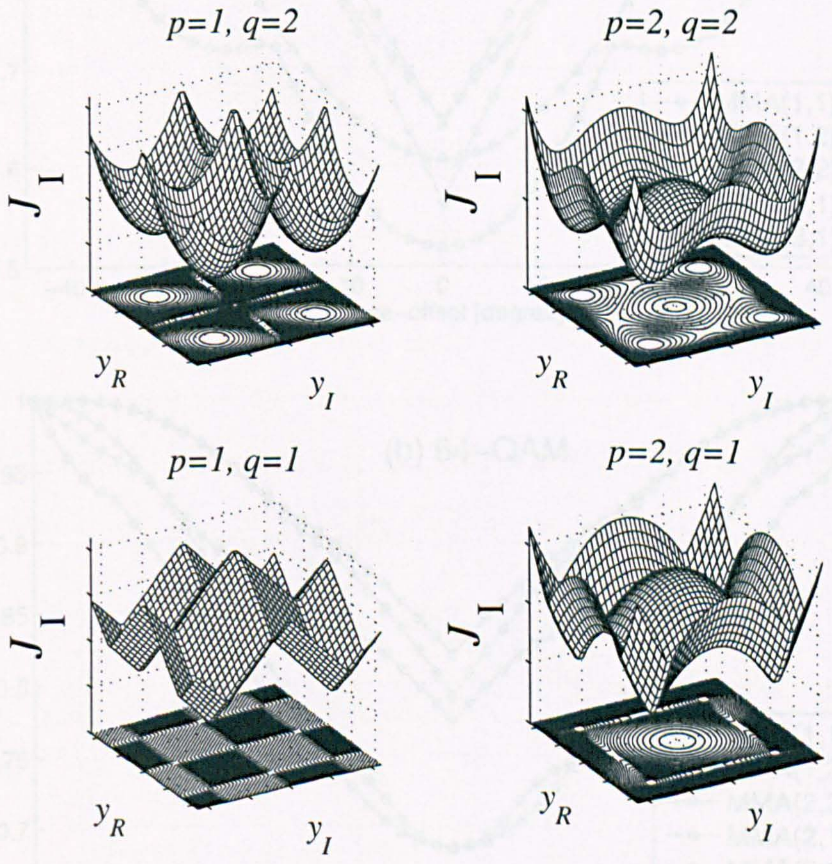


Figure 4.2: Mesh plots of MMA(p, q) cost-function for four exemplary cases.

Consider MMA(2, 2), we can show that [82]

$$J_\mathrm{I}(\theta) = \frac{1}{4}\mathsf{E}\left[a_R^4 + a_I^4 - 6a_R^2 a_I^2\right]\cos(4\theta) + \underbrace{\text{constants}}_{\text{w.r.t. }\theta} \tag{4.2.8}$$

Note that $\mathsf{E}\left[a_R^4 + a_I^4 - 6a_R^2 a_I^2\right]$ is always negative for QAM signals (due to its sub-Gaussian nature); consequently, the cost (4.2.8) exhibits desired stationary points. Since similar results as that in (4.2.8) is difficult to be obtained when $p = 1$ or $q = 1$, we provide the phase-sensitivity of other members of J_I, obtained by simulation, as depicted in Fig. 4.3.

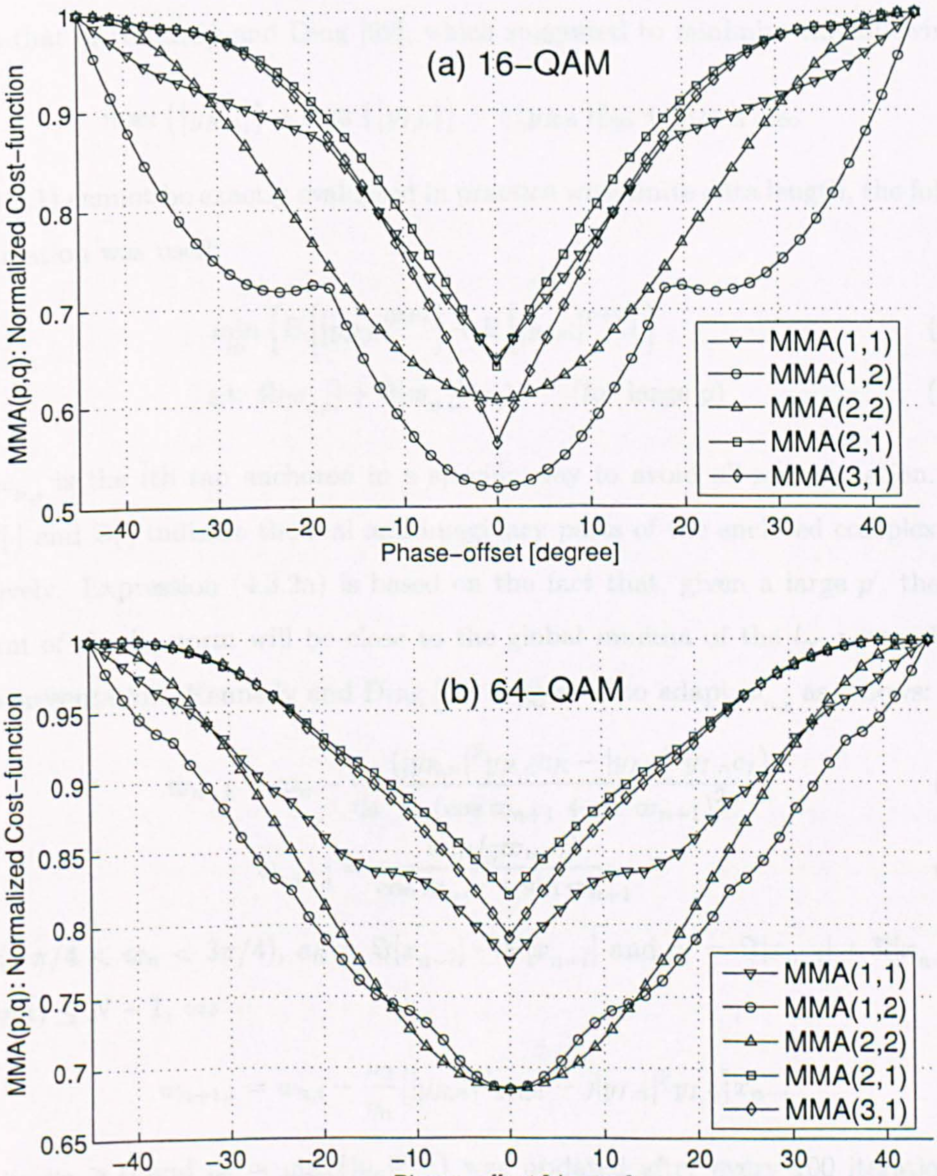


Figure 4.3: Effect of residual phase-offset θ on cost J_I, when $|\theta| \leq (\pi/4)$.

4.3 The Second Proposed Family of Algorithms: cMMA(p)

The cost-function based Bussgang blind equalization algorithms have been studied to be *non-convex* in nature when implemented using *finite-length* adaptive filters and found to be converging to undesirable local minima resulting in insufficient removal of channel distortion [132, 77]. A convex cost-function which has been specifically designed for QAM is that of Kennedy and Ding [69], which suggested to minimize the following:

$$\max\left\{|y_{R,n}|\right\} + \max\left\{|y_{I,n}|\right\} = \|\{y_{R,n}\}\|_\infty + \|\{y_{I,n}\}\|_\infty \tag{4.3.1}$$

Since (4.3.1) cannot be exactly evaluated in practice with finite data length, the following approximation was used:

$$\min_{w}\left\{\mathsf{E}\left[|y_{R,n}|^{p+2}\right] + \mathsf{E}\left[|y_{I,n}|^{p+2}\right]\right\}, \tag{4.3.2a}$$

$$\text{s.t. } \Re[w_{n,\bar{i}}] + \Im[w_{n,\bar{i}}] = 1, \quad \text{(for large } p\text{)} \tag{4.3.2b}$$

where $w_{n,\bar{i}}$ is the $\bar{i}$th tap anchored in a specific way to avoid all-zero situation. Notations $\Re[\cdot]$ and $\Im[\cdot]$ indicate the real and imaginary parts of the enclosed complex entity, respectively. Expression (4.3.2a) is based on the fact that, given a large p', the global minimum of the $l_{p'}$ norm will be close to the global minima of the l_∞ norm. Using a polar representation, Kennedy and Ding [69] suggested to adapt $w_{n,\bar{i}}$ as follows:

$$\varpi_{n+1} = \varpi_n - \frac{\mu_1}{\eta_n}\frac{\left(|y_{R,n}|^p y_{R,n}c_R - |y_{I,n}|^p y_{I,n}c_I\right)}{\left(\cos\varpi_{n+1} + \sin\varpi_{n+1}\right)^2} \tag{4.3.3a}$$

$$w_{n+1,\bar{i}} = \frac{\exp\left(\jmath\varpi_{n+1}\right)}{\cos\varpi_{n+1} + \sin\varpi_{n+1}}, \tag{4.3.3b}$$

where $(-\pi/4 < \varpi_n < 3\pi/4)$, $c_R = \Im[x_{n-\bar{i}}] - \Re[x_{n-\bar{i}}]$ and $c_I = \Im[x_{n-\bar{i}}] + \Re[x_{n-\bar{i}}]$. For $0 \leq (i \neq \bar{i}) \leq N-1$, *viz*

$$w_{n+1,i} = w_{n,i} - \frac{\mu_2}{\eta_n}\left[|y_{R,n}|^p y_{R,n} - \jmath|y_{I,n}|^p y_{I,n}\right]x_{n-i} \tag{4.3.4}$$

where $\mu_1, \mu_2 > 0$ and $\eta_n = \max\{|y_n|^{p+1}\}$ was updated after every 100 iterations. We refer to (4.3.3)-(4.3.4) as Kennedy-Ding algorithm (KDA). Note that the KDA required a separate automatic-gain control (AGC) mechanism to restore the signal's true energy. Based on a number of experiments for 4/16-QAM on various channels, we found that the performance of KDA is largely improved if higher p is selected. However, for higher-order QAM ($\geq$ 64-QAM), the convergence was either not achieved or only achieved by allowing a very small step-size. In this work we aim to modify the constraint (4.3.2b). According to [121], if P_a is the true signal energy, then we can use the constraint $\mathsf{E}\left[|y_n|^2\right] = P_a$ to restore the true energy. This consideration leads to the following optimization problem:

$$J_{\text{II}} = \mathsf{E}\big[\,|y_{R,n}|^{p+2} + |y_{I,n}|^{p+2}\,\big] + \lambda\left(\mathsf{E}\left[|y_n|^2\right] - P_a\right) \tag{4.3.5}$$

where λ is the Lagrangian multiplier. The stochastic gradient descent minimization gives the second family of algorithms, cMMA(p), as follows:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\big[(R_R^p - |y_{R,n}|^p)y_{R,n} - \jmath(R_I^p - |y_{I,n}|^p)y_{I,n}\big]\boldsymbol{x}_n \tag{4.3.6}$$

where $R_L^p = -2\lambda/(p+2)$ is a constant. Note that cMMA(p) would not require separate tap-anchoring and AGC (for true energy conservation), provided the constant R_L^p (or equivalently λ) is properly evaluated. We would discuss this evaluation in detail separately in Section 4.4. Note that cMMA(p) generalizes a number of existing algorithms. Like, for $p = 1$, cMMA(1) is equivalent to the algorithm proposed by Abrar *et al.* [8]. For $p = 2$, cMMA(2) becomes equivalent to MMA(2,2). Moreover, cMMA(p) can be considered as a simplified version of the algorithms proposed by Satorius and Mulligan [116] and Abrar and Shah [6].[1]

4.3.1 cMMA(p): Phase Recovery Capability

Note that the cost $\mathsf{E}[|y_{R,n}|^{p+2}] + \mathsf{E}[|y_{I,n}|^{p+2}]$ is sensitive to phase-offset $\forall p \geq 2$; however, this sensitivity can be expressed analytically *only* for even values of p. For $p = 2$, we

[1] In the year 1993, Satorius and Mulligan [116] made use of uniformly most powerful statistical tests and came up with a phase-sensitive scale-invariant cost-function for square-QAM. The details of the resulting adaptive algorithm, which they termed as Rectangular-Constellation-based Blind Equalization (RECBEQ), appeared six years later in a NASA technical report [95], *viz*

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\beta\left[R y_n^* - (|y_{R,n}|^p y_{R,n} - \jmath|y_{I,n}|^p y_{I,n})\right]\boldsymbol{x}_n \tag{4.3.7a}$$

$$\text{where } R = \frac{\sum_{k=n-B+1}^{n}\{|y_{R,k}|^{p+2} + |y_{I,k}|^{p+2}\}}{\sum_{k=n-B+1}^{n}\{y_{R,k}^2 + y_{I,k}^2\}}, \tag{4.3.7b}$$

$$\beta = \frac{\sqrt{B^{-1}\sum_{k=n-B+1}^{n}\{y_{R,k}^2 + y_{I,k}^2\}}}{\sqrt[q]{B^{-1}\sum_{k=n-B+1}^{n}\{|y_{R,k}^{p+2}| + |y_{I,k}^{p+2}|\}}}. \tag{4.3.7c}$$

where $q = (p+2)/(p+3)$. The algorithm RECBEQ is not known to have appeared in open literature. In the year 2006, Abrar and Shah [6] presented the following family of algorithms by optimizing a constrained cost-function which made use of both *a priori* and *a posteriori* equalizer coefficient vectors:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\left[\frac{y_{R,n}\,(R_R^p - |y_{R,n}|^p)}{R_R^p - \mu\,(R_R^p - |y_{R,n}|^p)} - \jmath\frac{y_{I,n}\,(R_I^p - |y_{I,n}|^p)}{R_I^p - \mu\,(R_I^p - |y_{I,n}|^p)}\right]\frac{\boldsymbol{x}_n}{\|\boldsymbol{x}_n\|_2^2}. \tag{4.3.8}$$

obtain a same expression as given in (4.2.8). For $p = 4$, we get

$$J_{\mathrm{II}}(\theta) = \frac{3}{8}\mathsf{E}\left[a_R^6 + a_I^6 - 5(a_R^2 a_I^4 + a_R^4 a_I^2)\right]\cos(4\theta) + \underbrace{\text{constants}}_{\text{w.r.t. }\theta} \tag{4.3.9}$$

The quantity $\mathsf{E}\left[a_R^6 + a_I^6 - 5(a_R^2 a_I^4 + a_R^4 a_I^2)\right]$ is negative for all square-QAM, like, it yields -3, -495 and -36063 for 4-, 16- and 64-QAM, respectively, ensuring that cMMA(4) exhibits desirable stationary points. Similar evidences for higher even values of p can easily be obtained.

Next we investigate the phase recovery capability for a generic p. Assume a QAM signal a_n contains four alphabets $a(1) = \alpha_R + \jmath\alpha_I$, $a(2) = -\alpha_R + \jmath\alpha_I$, $a(3) = \alpha_R - \jmath\alpha_I$ and $a(4) = -\alpha_R - \jmath\alpha_I$, where $\alpha_R, \alpha_I > 0$. We consider α_R and α_I are not fixed-valued. The cost $|\Re[a_n]|^{p+2} + |\Im[a_n]|^{p+2}$ maps these four points to the same point as (α_R, α_I). The expectation $\mathsf{E}[|\Re[a_n]|^{p+2} + |\Im[a_n]|^{p+2}]$ can be written as $\Sigma_{0,p} \equiv \sum_{i=1}^{4}(|\Re[a_n]|^{p+2} + |\Im[a_n]|^{p+2}])\mathrm{P}[a_n = a(i)]$ to yield

$$\Sigma_{0,p} = |\alpha_R|^{p+2} + |\alpha_I|^{p+2}. \tag{4.3.10}$$

Now assume that the equalizer output is subjected to a phase-offset θ, such that, $y_n = a_{n'}e^{\jmath\theta}$. For the moment, we restrict θ to be in a range such that both $a_{n'}$ and y_n lie in the same quadrant. If $\phi = \arctan(\alpha_I/\alpha_R)$, then this restriction corresponds to

$$-\min[\phi, \pi/2 - \phi] \le \theta \le \min[\phi, \pi/2 - \phi] \tag{4.3.11}$$

We further define

$$\Sigma_{\theta,p} \equiv \sum_{i=1}^{4}(|\Re[y_n]|^{p+2} + |\Im[y_n]|^{p+2}])\mathrm{P}[y_n = a(i)e^{\jmath\theta}] \tag{4.3.12}$$

to describes the effect of phase-offset on the cost (4.3.5), it leads to

$$\begin{aligned}\Sigma_{\theta,p} =& \frac{1}{2}(\alpha_R^2 + \alpha_I^2)^{p/2+1} \\ &\cdot [|\cos(\theta+\phi)|^{p+2} + |\cos(\theta-\phi)|^{p+2} + |\sin(\theta+\phi)|^{p+2} + |\sin(\theta-\phi)|^{p+2}]\end{aligned} \tag{4.3.13}$$

since $\mathrm{P}[y_n = a(i)e^{\jmath\theta}] = \frac{1}{4}$, $\forall i$. For $\theta \ne 0$, $\Sigma_{\theta,p}$ is expected to be greater than $\Sigma_{0,p}$. However, it is observed that $\Sigma_{\theta,p}$ may become smaller than $\Sigma_{0,p}$ for some values of θ, especially around the axes. This behavior is illustrated in Fig. 4.4, where we depict the signal space in its first quadrant. Observe that it is the middle region (specified with

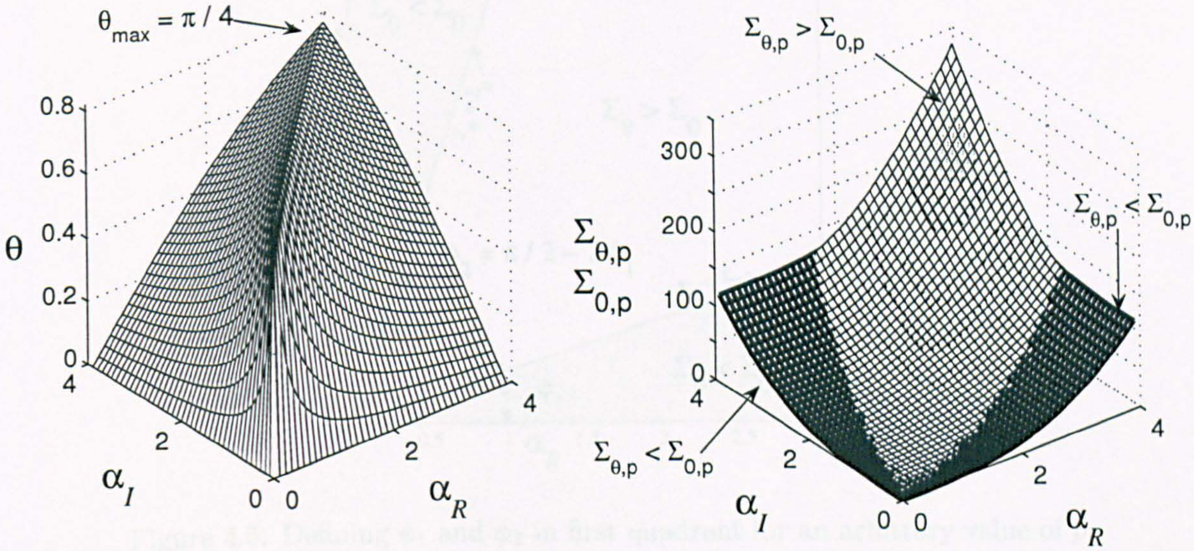


Figure 4.4: The restriction on θ (left) and the behavior of $\Sigma_{0,p}$ and $\Sigma_{\theta,p}$ (right) in the first quadrant.

angle ϕ_2, $\forall p \geq 1$, in the first quadrant), where the value of $\Sigma_{\theta,p}$ is higher than $\Sigma_{0,p}$. We also observe two lines at which $\Sigma_{\theta,p} = \Sigma_{0,p}$.

In Fig. 4.5, at the line closer to x-axis, we have $\phi = \phi_1 = \theta$, whereas at the line closer to y-axis, we have $\phi = \phi_1 + \phi_2 = \pi/2 - \theta$. Equating (4.3.10) and (4.3.13) together and solving for the lower line with substitution $\phi = \phi_1 = \theta$, we get

$$|\cos(2\phi_1)|^{p+2} + |\sin(2\phi_1)|^{p+2} + 1 = 2\frac{|\alpha_R|^{p+2} + |\alpha_I|^{p+2}}{(\alpha_R^2 + \alpha_I^2)^{(p+2)/2}} \tag{4.3.14}$$

Letting $\varrho = p + 2$, we take the ϱth root on both sides of (4.3.14). Under the limit $\varrho = p + 2 \to \infty$, we obtain for (4.3.14),

$$\text{L.H.S.} = \lim_{\varrho\to\infty} (|\cos(2\phi_1)|^{\varrho} + |\sin(2\phi_1)|^{\varrho} + 1)^{1/\varrho} \to 1 \tag{4.3.15}$$

and

$$\text{R.H.S.} = \lim_{\varrho\to\infty} \frac{\sqrt[\varrho]{2(\alpha_R^{\varrho} + \alpha_I^{\varrho})}}{\sqrt{\alpha_R^2 + \alpha_I^2}} \to \frac{\max\{\alpha_R, \alpha_I\}}{\sqrt{\alpha_R^2 + \alpha_I^2}} \tag{4.3.16}$$

Due to considering the lower line, the equation $\max\{\alpha_R, \alpha_I\} \to \sqrt{\alpha_R^2 + \alpha_I^2}$ has a unique solution $\alpha_I \to 0$; which implies $\phi_1 = \arctan(\alpha_I/\alpha_R) \to 0$ and $\phi_2 = \pi/2 - 2\phi_1 \to \pi/2$. It is an important result which describes that, due to the expansion of the middle region, phase recovery capability improves with an increment in p.

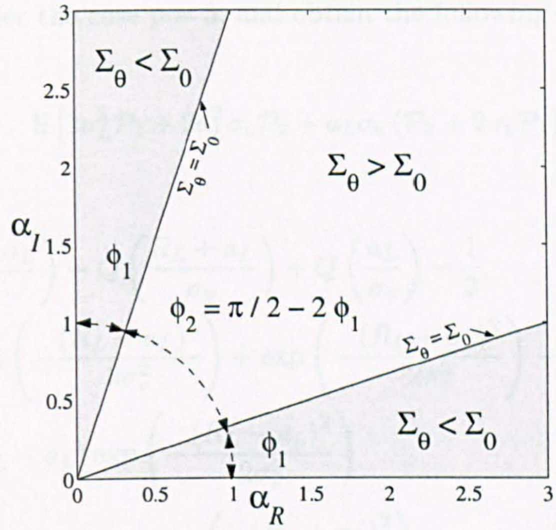


Figure 4.5: Defining ϕ_1 and ϕ_2 in first quadrant for an arbitrary value of p.

4.4 Evaluation of Dispersion Constants

4.4.1 Derivation of R_L for MMA$(p, 1)$

We need to solve the following to obtain the value of R_L for MMA$(p, 1)$ in the presence of convolutional noise,

$$\mathsf{E}\big[a_L(a_L + v)|a_L + v|^{p-2}\mathrm{sgn}\big[|a_L + v| - R_L\big]\big] = 0 \tag{4.4.1}$$

We first consider MMA$(1, 1)$; denoting

$$f_v = \left(\sigma_v\sqrt{2\pi}\right)^{-1}\exp\left(-v^2/(2\sigma_v^2)\right), \tag{4.4.2}$$

we evaluate

$$\begin{aligned}
&\mathsf{E}\left[a_L\,\mathrm{sgn}[a_L + v]\,\mathrm{sgn}\left[|a_L + v| - R_L\right]\right] \\
&= \mathsf{E}\left[-\int_{-\infty}^{-R_L - a_L} + \int_{-R_L - a_L}^{-a_L} - \int_{-a_L}^{R_L - a_L} + \int_{R_L - a_L}^{\infty} a_L f_v \mathrm{d}v\right]
\end{aligned} \tag{4.4.3}$$

and find the following to solve for R_L in MMA$(1, 1)$:

$$\mathsf{E}\left[a_L\left\{Q\left(\frac{R_L - a_L}{\sigma_v}\right) - Q\left(\frac{R_L + a_L}{\sigma_v}\right) + Q\left(\frac{a_L}{\sigma_v}\right) - \frac{1}{2}\right\}\right] = 0 \tag{4.4.4}$$

where $Q(z) = (1/\sqrt{2\pi})\int_z^\infty \exp\left(-0.5t^2\right)\mathrm{d}t$. Next we consider the case $p = 2$ and obtain the following expression to solve for R_L in MMA$(2, 1)$:

$$\begin{aligned}
&2\mathsf{E}\left[a_L^2\left\{Q\left(\frac{R_L - a_L}{\sigma_v}\right) + Q\left(\frac{R_L + a_L}{\sigma_v}\right) - \frac{1}{2}\right\}\right] \\
&\quad + \sqrt{\frac{2}{\pi}}\sigma_v\mathsf{E}\left[a_L\left\{\exp\left(-\frac{(R_L - a_L)^2}{2\sigma_v^2}\right) - \exp\left(-\frac{(R_L + a_L)^2}{2\sigma_v^2}\right)\right\}\right] = 0
\end{aligned} \tag{4.4.5}$$

Finally, we consider the case $p = 3$, and obtain the following expression to solve for R_L in MMA$(3,1)$:

$$\mathsf{E}\left[2a_L^3\mathcal{P}_1 + 2a_L^2\sigma_v\mathcal{P}_2 + a_L\sigma_v\left(\mathcal{P}_3 + 2\sigma_v\mathcal{P}_1\right)\right] = 0, \tag{4.4.6}$$

$$\mathcal{P}_1 = Q\left(\frac{R_L - a_L}{\sigma_v}\right) - Q\left(\frac{R_L + a_L}{\sigma_v}\right) + Q\left(\frac{a_L}{\sigma_v}\right) - \frac{1}{2}, \tag{4.4.7a}$$

$$\mathcal{P}_2 = \sqrt{\frac{2}{\pi}}\left\{\exp\left(-\frac{(R_L - a_L)^2}{2\sigma_v^2}\right) + \exp\left(-\frac{(R_L + a_L)^2}{2\sigma_v^2}\right) - \exp\left(-\frac{a_L^2}{2\sigma_v^2}\right)\right\}, \tag{4.4.7b}$$

$$\begin{aligned}\mathcal{P}_3 = \sqrt{\frac{2}{\pi}}\bigg\{ &(R_L - a_L)\exp\left(-\frac{(R_L - a_L)^2}{2\sigma_v^2}\right) \\ &- (R_L + a_L)\exp\left(-\frac{(R_L + a_L)^2}{2\sigma_v^2}\right) + a_L\exp\left(-\frac{a_L^2}{2\sigma_v^2}\right)\bigg\}.\end{aligned} \tag{4.4.7c}$$

In a noiseless environment, we can obtain a closed-form solution for R_L by exploiting Goupil-Palicot principle [55]. Considering the case of one of the quadrature components, a_L, we evaluate for MMA$(p,1)$:

$$J_{\mathrm{I}} = \frac{2}{\sqrt{M}}\sum_{j=1}^{\sqrt{M}/2}(2j-1)^p\left|d^p - \frac{R_L^p}{(2j-1)^p}\right| \tag{4.4.8}$$

The gradient of (4.4.8) with respect to d can be computed as

$$\frac{\partial J_{\mathrm{I}}}{\partial d} = \frac{2pd^{p-1}}{\sqrt{M}}\left[\sum_{j=z+1}^{\sqrt{M}/2}(2j-1)^p - \sum_{j=1}^{z}(2j-1)^p\right] \tag{4.4.9}$$

where z is a positive integer, which satisfies [65]:

$$\frac{R_L}{2(z+1)-1} < d < \frac{R_L}{2z-1} \tag{4.4.10}$$

According to [55], the coefficient d should converge to 1 in order to recover the *true* energy of the signal. Substituting $(d = 1)$ in (4.4.10) we get $2z - 1 < R_L < 2z + 1$. Thanks to Im *et al.* [65], (4.4.10) can be solved by using the relation $\lceil z\rceil \leq z + 1$; where $\lceil z\rceil$ is the smallest positive integer greater than or equal to z, it gives

$$R_L = 2\lceil z\rceil - 1, \tag{4.4.11}$$

where the value of z is obtained by solving $\partial J_{\mathrm{I}}/\partial d = 0$, *viz*

$$\sum_{j=z+1}^{\sqrt{M}/2}(2j-1)^p - \sum_{j=1}^{z}(2j-1)^p = 0 \tag{4.4.12}$$

For $p = 1$, we can readily get $z = \sqrt{M/8}$. It gives $R_L = 3, 5$ and 11 for 16-, 64- and 256-QAM, respectively. Similarly, for $p = 2$, we get to solve $((M-1)/2) + ((2z - 8z^3)/\sqrt{M}) = 0$; solution of which is obtained as $z = (z_1/12) + (1/z_1)$, where for $z_2 = 0.5\sqrt{M}(M-1)$, z_1 is given as $z_1 = \sqrt[3]{108z_2 + 12\sqrt{81z_2^2 - 12}}$. It gives $R_L = 3, 7$ and 13 for 16-, 64- and 256-QAM, respectively. Also, for $p = 3$, we solve $32z^4 - 16z^2 + 2M - M^2 = 0$ and get $z = 0.25\sqrt{4 + 2\sqrt{2M^2 - 4M + 4}}$. It gives same values for R_L as we obtained for $p = 2$.

4.4.2 Derivation of R_L for MMA$(p, 2)$

Here we discuss the evaluation of R_L for MMA$(p, 2)$, we need to solve

$$\mathsf{E}\left[a_L(a_L + v)|a_L + v|^{p-2}\left(|a_L + v|^p - R_L^p\right)\right] = 0 \tag{4.4.13}$$

to obtain the value of R_L in the presence of convolutional noise. Consider $p = 1$, we readily get $R_L = 0.5\mathsf{E}\left[a_L^2\right]/\mathsf{E}\left[a_L Q\left(-a_L/\sigma_v\right)\right]$, which exploits the fact that $\mathsf{E}\left[a_L v\right] = \mathsf{E}\left[v\right] = \mathsf{E}\left[a_L\right] = 0$. Now consider a high output-SNR case, i.e., $\sigma_v \to 0$, it implies $\mathsf{E}\left[a_L Q\left(-a_L/\sigma_v\right)\right]$ will approach 1 and 0 for $(a_L > 0)$ and $(a_L < 0)$, respectively. Since $\mathsf{E}[|a_L|] = 2\mathsf{E}[a_L]_{(a_L>0)} = 2\mathsf{E}[-a_L]_{(a_L<0)}$, it yields $R_L = \mathsf{E}\left[a_L^2\right]/\mathsf{E}\left[|a_L|\right]$, which is a well-known result [19] and we obtained it in a limiting case (of vanishing convolutional noise). For $p = 2$, we obtain an even simpler result, $R_L = \sqrt{\mathsf{E}\left[a_L^4\right]/\mathsf{E}\left[a_L^2\right] + 3\sigma_v^2}$. Under the limiting case, it simplifies to the value that appeared in [137]. Finally, exploiting Goupil-Palicot principle, we obtain the value of R_L (for a generic p) in a noise free environment, *viz*

$$R_L^p = \frac{\mathsf{E}\left[a_L^{2p}\right]}{\mathsf{E}\left[|a_L|^p\right]}. \tag{4.4.14}$$

From (4.4.14), with $p = 1$, we obtain $R_L = 2.5, 5.25$ and 10.63 for 16-, 64- and 256-QAM, respectively. Similarly, with $p = 2$, we obtain $R_L = 2.86, 6.08$ and 12.34 for 16-, 64- and 256-QAM, respectively.

4.4.3 Derivation of R_L for cMMA(p)

Here we discuss the evaluation of R_L for cMMA(p), we need to solve

$$\mathsf{E}\left[a_L(a_L + v)\left(|a_L + v|^p - R_L^p\right)\right] = 0 \tag{4.4.15}$$

to obtain the value of R_L in the presence of convolutional noise. Consider $p = 1$, we have $\mathsf{E}[a_L(a_L + v)(|a_L + v| - R_L)] = 0$, which readily gives the solution for cMMA(1):

$$R_L = \frac{\mathsf{E}\left[2a_L(a_L^2 + \sigma_v^2)Q\left(-\frac{a_L}{\sigma_v}\right) + \sqrt{\frac{2}{\pi}}a_L^2\sigma_v \exp\left(-\frac{a_L^2}{2\sigma_v^2}\right)\right]}{\mathsf{E}\left[a_L^2\right]} \tag{4.4.16}$$

Under the limit σ_v tends to zero, (4.4.16) simplifies to the value that appeared in [8]. Finally, exploiting Goupil-Palicot principle, we obtain the value of R_L (for a generic p) in a noise free environment, *viz*

$$R_L^p = \frac{\mathsf{E}\left[|a_L|^{p+2}\right]}{\mathsf{E}\left[a_L^2\right]}. \tag{4.4.17}$$

From (4.4.17), for $p = 1$, we obtain $R_L = 2.8, 5.9$ and 11.95 for 16-, 64- and 256-QAM, respectively. For $p = 2$, we obtain similar values of R_L as we obtained in MMA(2,2). Also, for $p = 3$, we obtain $R_L = 2.9, 6.22$ and 12.63 for 16-, 64- and 256-QAM, respectively. Now we determine the value of λ in (4.3.5), in a noise-free scenario, we obtain $\lambda = -0.5(p+2)\mathsf{E}\left[|a_L|^{p+2}\right]/\mathsf{E}\left[a_L^2\right]$, where the negative sign indicates that we need to *maximize* the equalizer output energy while *minimizing* the higher-order moments.

In order to compare the expressions we obtained for R_L, we define the following quantity:

$$\text{Ratio}_R = \frac{R_L(\text{noisy})}{R_L(\text{noise free})} \tag{4.4.18}$$

For noisy scenario, we consider solutions of (4.4.1), (4.4.13) and (4.4.15) for MMA(p, 1), MMA(p, 2) and cMMA(p), respectively. For noise-free scenario, we consider expressions (4.4.11), (4.4.14) and (4.4.17) for MMA(p, 1), MMA(p, 2) and cMMA(p), respectively. In Fig. 4.6, we show the values of Ratio_R versus output-SNR. Note that the Ratio_R deviates from unity significantly at low output-SNR which indicates that a minimum mean-fluctuation in equalizer coefficients can only be ensured by using the values of R_L computed with the consideration of convolutional noise. Finally, we emphasize that, at high output-SNR, the values of R_L obtained for noisy scenario coincide perfectly with those for noise free scenario (i.e., $\text{Ratio}_R \to 1$ when $\sigma_v \to 0$).

4.5 Dynamic Convergence Analysis

Due to the nonlinear nature of MMA(p, q) or cMMA(p), especially for odd values of p and q, it is quite difficult to study their general formulation, stationary points, desired

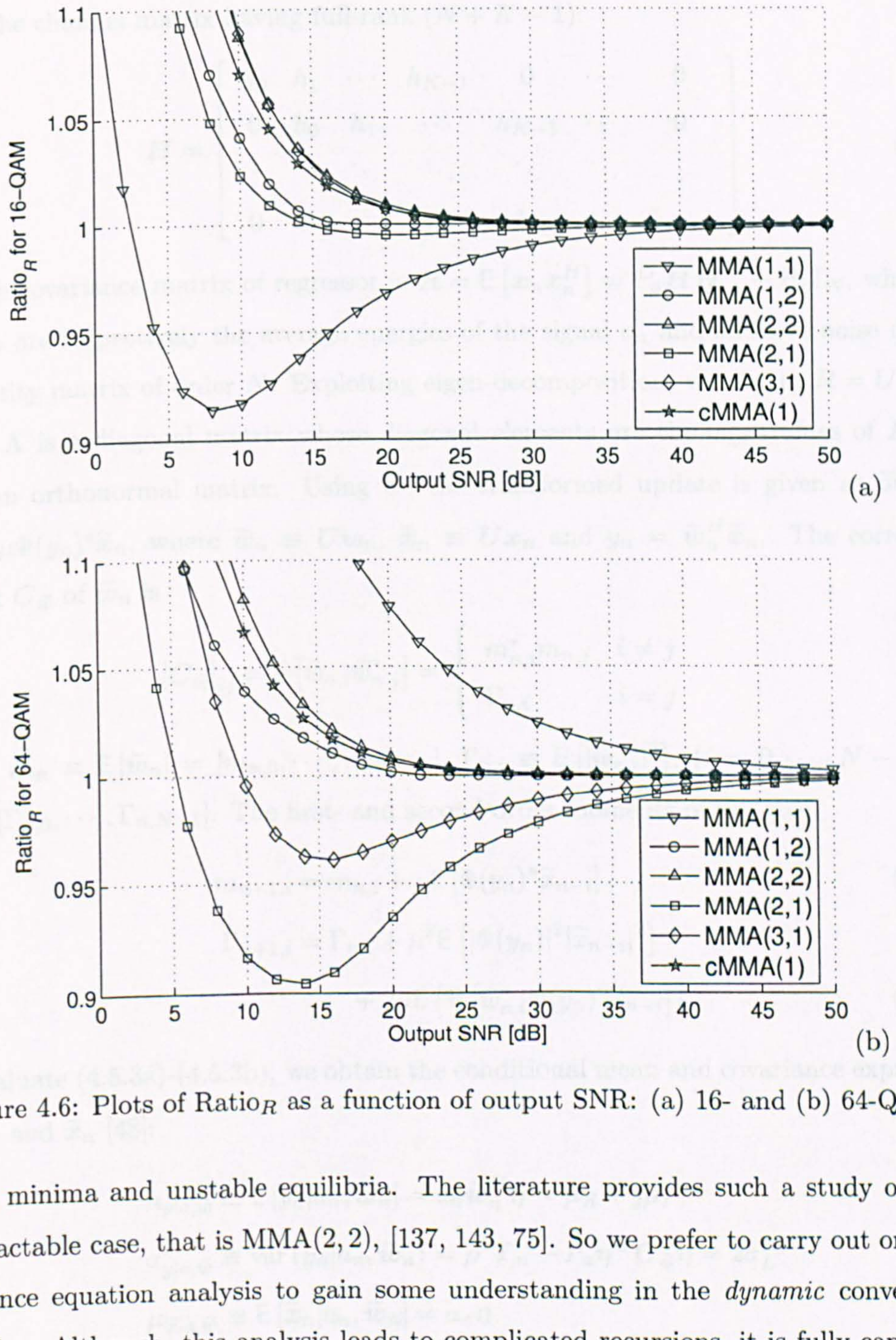


Figure 4.6: Plots of Ratio_R as a function of output SNR: (a) 16- and (b) 64-QAM.

global minima and unstable equilibria. The literature provides such a study only for one tractable case, that is MMA(2, 2), [137, 143, 75]. So we prefer to carry out ordinary difference equation analysis to gain some understanding in the *dynamic* convergence behavior. Although, this analysis leads to complicated recursions, it is fully capable of handling nonlinear and discontinuous error-functions. For the sake of conciseness, we are providing only the final expressions which are necessary for the evaluation of performance measures. In the sequel, we use the notation $[\boldsymbol{B}]_{ij}$ to denote the element of matrix $\boldsymbol{B}$ in its ith row and jth column, and $[\boldsymbol{b}]_i$ to denote the ith element of array $\boldsymbol{b}$. Also we define

$\boldsymbol{H}$ as the channel matrix having full-rank $(N+K-1)$:

$$\boldsymbol{H} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{K-1} & 0 & \cdots & 0 \\ 0 & h_0 & h_1 & \cdots & h_{K-1} & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & h_0 & h_1 & \cdots & h_{K-1} \end{bmatrix} \tag{4.5.1}$$

The covariance matrix of regressor is $\boldsymbol{R} = \mathsf{E}\left[\boldsymbol{x}_n \boldsymbol{x}_n^H\right] = P_a \boldsymbol{H}\boldsymbol{H}^H + P_\vartheta \boldsymbol{I}_N$, where P_a and P_ϑ are respectively the average energies of the signal a_n and additive noise ϑ_n; $\boldsymbol{I}_N$ is identity matrix of order N. Exploiting eigen-decomposition, we obtain $\boldsymbol{R} = \boldsymbol{U}^H \boldsymbol{\Lambda} \boldsymbol{U}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\boldsymbol{R}$, and $\boldsymbol{U}$ is an orthonormal matrix. Using $\boldsymbol{U}$, the transformed update is given as $\widetilde{\boldsymbol{w}}_{n+1} = \widetilde{\boldsymbol{w}}_n + \mu\Phi(y_n)^* \widetilde{\boldsymbol{x}}_n$, where $\widetilde{\boldsymbol{w}}_n \equiv \boldsymbol{U}\boldsymbol{w}_n$, $\widetilde{\boldsymbol{x}}_n \equiv \boldsymbol{U}\boldsymbol{x}_n$ and $y_n = \widetilde{\boldsymbol{w}}_n^H \widetilde{\boldsymbol{x}}_n$. The correlation matrix $\boldsymbol{C}_{\widetilde{w}}$ of $\widetilde{\boldsymbol{w}}_n$ is

$$[\boldsymbol{C}_{\widetilde{w}}]_{ij} = \mathsf{E}\left[\widetilde{w}_{n,i}\widetilde{w}_{n,j}^*\right] = \begin{cases} m_{n,i}^* m_{n,j} & i \neq j \\ \Gamma_{n,i} & i = j \end{cases} \tag{4.5.2}$$

where $\boldsymbol{m}_n \equiv \mathsf{E}[\widetilde{\boldsymbol{w}}_n] = [m_{n,0}, \cdots, m_{n,N-1}]$, $\Gamma_{n,i} \equiv \mathsf{E}\left[|\widetilde{w}_{n,i}|^2\right]$, $(i = 0, \cdots, N-1)$ and $\boldsymbol{\Gamma}_n = [\Gamma_{n,0}, \cdots, \Gamma_{n,N-1}]$. The first- and second-order moments of $\widetilde{\boldsymbol{w}}_n$ give

$$m_{n+1,i} = m_{n,i} + \mu \mathsf{E}\left[\Phi(y_n)^* \widetilde{x}_{n-i}\right], \tag{4.5.3a}$$

$$\begin{aligned} \Gamma_{n+1,i} = \Gamma_{n,i} &+ \mu^2 \mathsf{E}\left[|\Phi(y_n)|^2 |\widetilde{x}_{n-i}|^2\right] \\ &+ 2\mu \mathsf{E}\left[\Re\left[\widetilde{w}_{n,i}^* \Phi(y_n)^* \widetilde{x}_{n-i}\right]\right]. \end{aligned} \tag{4.5.3b}$$

To evaluate (4.5.3a)-(4.5.3b), we obtain the conditional mean and covariance expressions for y_n and $\widetilde{\boldsymbol{x}}_n$ [48]:

$$\mu_{y|a,\widetilde{w}} \equiv \mathsf{E}\left[y_n | a_n, \widetilde{\boldsymbol{w}}_n\right] = a_n \widetilde{\boldsymbol{w}}_n^H \boldsymbol{\eta} = \mu_R + \jmath \mu_I \tag{4.5.4a}$$

$$\sigma^2_{y|a,\widetilde{w}} \equiv \operatorname{var}\left(y_n | a_n, \widetilde{\boldsymbol{w}}_n\right) = \boldsymbol{\rho}^T \boldsymbol{\Gamma}_n - P_a \boldsymbol{\eta}^H \boldsymbol{C}_{\widetilde{w}} \boldsymbol{\eta} = 2\sigma_L^2 \tag{4.5.4b}$$

$$\boldsymbol{\mu}_{\widetilde{x}|a,\widetilde{w}} \equiv \mathsf{E}\left[\widetilde{\boldsymbol{x}}_n | a_n, \widetilde{\boldsymbol{w}}_n\right] = a_n \boldsymbol{\eta} \tag{4.5.4c}$$

$$\boldsymbol{C}_{\widetilde{x}|a,\widetilde{w}} \equiv \operatorname{cov}\left(\widetilde{\boldsymbol{x}}_n, \widetilde{\boldsymbol{x}}_n | a_n, \widetilde{\boldsymbol{w}}_n\right) = \boldsymbol{\Lambda} - P_a \boldsymbol{\eta}\boldsymbol{\eta}^H \tag{4.5.4d}$$

where $\boldsymbol{\rho} = \operatorname{diag}[\boldsymbol{\Lambda}]$ and $\boldsymbol{\eta} = \boldsymbol{U}\boldsymbol{h}$. Using (4.5.3a)-(4.5.3b), the instantaneous mean square error (MSE) is given by [134]:

$$\begin{aligned} \mathrm{MSE}_n &= \mathsf{E}\left[|y_n|^2 | a_n, \widetilde{\boldsymbol{w}}_n\right] + \mathsf{E}[|a_n|^2] - 2\Re\left[\mathsf{E}[y_n a_n^* | a_n, \widetilde{\boldsymbol{w}}_n]\right] \\ &= \boldsymbol{\rho}^T \boldsymbol{\Gamma}_n + P_a - 2P_a \Re\left[\boldsymbol{m}_n^H \boldsymbol{\eta}\right] \end{aligned} \tag{4.5.5}$$

Below we present an approximate expression for the instantaneous residual ISI (deduced from [134, page: 858-9]):

$$\begin{aligned}\mathrm{ISI}_n &\approx \frac{\mathrm{var}\,(y_n|a_n,\widetilde{\boldsymbol{w}}_n)-\mathsf{E}\left[|\sum_i \vartheta_{n-i}\widetilde{w}^*_{n,i}|^2|\widetilde{\boldsymbol{w}}_n\right]}{\mathsf{E}[|a_n|^2]}\\ &\approx \frac{\boldsymbol{\rho}^T\boldsymbol{\Gamma}_n - P_a\boldsymbol{\eta}^H\boldsymbol{C}_{\widetilde{w}}\boldsymbol{\eta}-P_\vartheta\mathbf{1}^T\boldsymbol{\Gamma}_n}{P_a|\boldsymbol{m}_n^H\boldsymbol{\eta}|^2}\end{aligned} \tag{4.5.6}$$

where $\mathbf{1}$ is an N-element column-vector of ones. Making use of (4.5.4) and defining $p_n = \boldsymbol{m}_n^H\boldsymbol{\eta}$, we list the three expectation quantities which appear in (4.5.3a)-(4.5.3b) as follows [74]:

$$\mathsf{E}\left[\widetilde{x}_{n-i}\Phi(y_n)^*\right] = \eta_i\mathsf{E}\left[a_n g_0^*\right] + \left[\boldsymbol{C}_{\widetilde{x}|a}\boldsymbol{m}_n\right]_i(\mathsf{E}\left[g_1\right]-p_n\mathsf{E}\left[a_n g_0^*\right])\sigma_y^{-2} \tag{4.5.7}$$

$$\begin{aligned}\mathsf{E}\left[\widetilde{w}_i^*\widetilde{x}_{n-i}\Phi(y_n)^*\right] &= m^*_{n,i}\eta_i\mathsf{E}\left[a_n g_0^*\right] + (\mathsf{E}\left[g_1\right]-p_n\mathsf{E}\left[a_n g_0^*\right])\\ &\quad\cdot\left[\left[\boldsymbol{C}_{\widetilde{x}|a}\right]_{ii}\Gamma_{n,i} + m^*_{n,i}\sum_{k\neq i}\left[\boldsymbol{C}_{\widetilde{x}|a}\right]_{ik}m_{n,k}\right]\sigma_y^{-2}\end{aligned} \tag{4.5.8}$$

$$\begin{aligned}\mathsf{E}\left[|\widetilde{x}_{n-i}|^2|\Phi(y_n)|^2\right] &= \left[\boldsymbol{C}_{\widetilde{x}|a}\right]_{ii}\mathsf{E}\left[g_2\right] + |\eta_i|^2\mathsf{E}\left[|a_n|^2 g_2\right]\\ &\quad+2\frac{\Re\left[\eta_i^*\left[\boldsymbol{C}_{\widetilde{x}|a}\boldsymbol{m}_n\right]_i(\mathsf{E}\left[a_n^* g_3\right]-p_n\mathsf{E}\left[|a_n|^2 g_2\right])\right]}{\sigma_y^2}\\ &\quad+\left(\mathsf{E}\left[g_4\right]-2\Re\left[p_n^*\mathsf{E}\left[a_n^* g_3\right]\right]+|p_n|^2\mathsf{E}\left[|a_n|^2 g_2\right]-\sigma_y^2\mathsf{E}\left[g_2\right]\right)\\ &\quad\cdot\left[\left[|\boldsymbol{C}_{\widetilde{x}|a}|^2\right]_{ii}\Gamma_{n,i}+\sum_{k\neq j}\left[\boldsymbol{C}_{\widetilde{x}|a}\right]_{ik}\left[\boldsymbol{C}^*_{\widetilde{x}|a}\right]_{ij}m_{n,k}m^*_{n,j}\right]\sigma_y^{-4}\end{aligned} \tag{4.5.9}$$

Next denoting $H_L^{01} = \mathsf{E}\left[\phi(y_{L,n})\right]$, $H_L^{11} = \mathsf{E}\left[y_{L,n}\phi(y_{L,n})\right]$, $H_L^{02} = \mathsf{E}\left[\phi^2(y_{L,n})\right]$, $H_L^{12} = \mathsf{E}\left[y_{L,n}\phi^2(y_{L,n})\right]$, $H_L^{22} = \mathsf{E}\left[y_{L,n}^2\phi^2(y_{L,n})\right]$, g_i's are given by $g_0 = H_R^{01} + \jmath H_I^{01}$, $g_1 = H_R^{11} + H_I^{11} + \jmath(\mu_I H_R^{01} - \mu_R H_I^{01})$, $g_2 = H_R^{02} + H_I^{02}$, $g_3 = H_R^{12} + \mu_R H_I^{02} + \jmath(H_I^{12} + \mu_I H_R^{02})$, and $g_4 = H_R^{22} + H_I^{22} + (\sigma_R^2 + \mu_R^2)H_I^{02} + (\sigma_I^2 + \mu_I^2)H_R^{02}$. Depending on the type of error-function, H_L^{ij} is defined in Table 4.1 for all cases of algorithms, we have used an auxiliary variable $G_L^{\cdot\cdot}$ for this purpose. In Table 4.2, we describe how to compute $G_L^{\cdot\cdot}$ as a function of yet another auxiliary variable $\mathcal{S}_k$, where $\mathcal{S}_k$ is defined as:

$$\mathcal{S}_k(\widehat{\alpha},\widehat{\beta}) \equiv \frac{1}{\sqrt{2\pi}}\int_{\widehat{\alpha}}^{\widehat{\beta}}(\sigma_L x + \mu_L)^k \exp\left[-0.5x^2\right]\mathrm{d}x \tag{4.5.10}$$

We can express $\mathcal{S}_k(\widehat{\alpha},\widehat{\beta}) = \sum_{i=0}^{k}\frac{k!}{i!(k-i)!}\sigma_L^i\mu_L^{k-i}F_i(\widehat{\alpha},\widehat{\beta})$, where $\widehat{\alpha} = (\alpha-\mu_L)/\sigma_L$ and $\widehat{\beta} = (\beta-\mu_L)/\sigma_L$, and $F_k(x,y) \equiv \frac{1}{\sqrt{2\pi}}\int_x^y t^k e^{-0.5t^2}\mathrm{d}t$. So, $F_0(x,y) = Q(x) - Q(y)$ and $F_1(x,y) = (1/\sqrt{2\pi})(e^{-0.5x^2} - e^{-0.5y^2})$. Also $F_k(x,y) = (1/\sqrt{2\pi})(x^{k-1}e^{-0.5x^2} - y^{k-1}e^{-0.5y^2}) + (k-1)F_{k-2}(x,y)$.

Table 4.1: MMA(p,q) and cMMA(p): Values of H_L^{ij} in terms of $G_L^{\cdot\cdot}$.

MMA(1,1): $G_L^i(\alpha,\beta)=\int_\alpha^\beta x^i f_L(x)\mathrm{d}x$ $H_L^{ij}\equiv G_L^i(0,R_L)+G_L^i(-\infty,-R_L)$ $+(-1)^j\left[G_L^i(R_L,\infty)+G_L^i(-R_L,0)\right]$	MMA(1,2): $G_L^{ijk}(\alpha,\beta)=\int_\alpha^\beta x^i\,(\mathcal{R}_k-x)^j\,f_L(x)\mathrm{d}x$ $H_L^{ij}=G_L^{ij1}(-\infty,0)+G_L^{ij2}(0,\infty),\ \mathcal{R}_k=(-1)^kR_L$
MMA(2,1): $G_L^{ij}(\alpha,\beta)=\int_\alpha^\beta x^{i+j} f_L(x)\mathrm{d}x$ $H_L^{ij}=G_L^{ij}(-R_L,R_L)$ $+(-1)^j\left[G_L^{ij}(R_L,\infty)+G_L^{ij}(-\infty,-R_L)\right]$	MMA(2,2): $G_L^{ij}(\alpha,\beta)=\int_\alpha^\beta x^{i+j}\,(R_L^2-x^2)^j\,f_L(x)\mathrm{d}x$ $H_L^{ij}\equiv G_L^{ij}(-\infty,\infty)$
MMA(3,1): $G_L^{ij}(\alpha,\beta)=\int_\alpha^\beta x^{i+2j} f_L(x)\mathrm{d}x$ $H_L^{ij}\equiv G_L^{ij}(0,R_L)+G_L^{ij}(-\infty,-R_L)$ $+(-1)^j\left[G_L^{ij}(R_L,\infty)+G_L^{ij}(-R_L,0)\right]$	cMMA(1): $G_L^{ijk}(\alpha,\beta)=\int_\alpha^\beta x^{i+j}\left(R_L-(-1)^kx\right)^j f_L(x)\mathrm{d}x$ $H_L^{ij}\equiv G_L^{ij1}(-\infty,0)+G_L^{ij2}(0,\infty)$
cMMA(3): $G_L^{ijk}(\alpha,\beta)=\int_\alpha^\beta x^{i+j}\left(R_L^3-(-1)^kx^3\right)^j f_L(x)\mathrm{d}x$ $H_L^{ij}\equiv G_L^{ij1}(-\infty,0)+G_L^{ij2}(0,\infty)$	cMMA(5): $G_L^{ijk}(\alpha,\beta)=\int_\alpha^\beta x^{i+j}\left(R_L^5-(-1)^kx^5\right)^j f_L(x)\mathrm{d}x$ $H_L^{ij}\equiv G_L^{ij1}(-\infty,0)+G_L^{ij2}(0,\infty)$
cMMA(4): $G_L^{ij}(\alpha,\beta)=\int_\alpha^\beta x^{i+j}\,(R_L^4-x^4)^j\,f_L(x)\mathrm{d}x$ $H_L^{ij}\equiv G_L^{ij}(-\infty,\infty)$	where we have denoted $f_L(x)=\frac{1}{\sqrt{2\pi}\sigma_L}\exp\left[-\left(\frac{x-\mu_L}{\sqrt{2}\sigma_L}\right)^2\right]$

4.6 Simulation Results

4.6.1 Experiment: MSE/ISI Analytical/Simulated Performances

To compare performances of existing and new equalizers and to validate the dynamic convergence analysis, we estimate ISI and MSE convergence trajectories for five members of each of the families MMA(p,q) and cMMA(p). We consider 16/64/256-QAM signalling over a complex-valued voice-band telephonic channel [105]. The input-SNR is taken as 30dB for 16-QAM and 34dB for 64/256-QAM. A seven-tap equalizer is used and initialized with central single-spike. We consider estimates of output-SNR, in the computation of R_L, only for the members of MMA$(p,1)$ because the R_L of the addressed members of MMA$(p,2)$ and cMMA(p) have been found to be less sensitive to convolutional noise. We experimentally obtain the steady-state value of output-SNR to be equal to 24.5, 25.0 and 25.5 [dB] for 16-, 64- and 256-QAM, respectively, based on trial-and-error method, such that two conditions are satisfied upon successful convergence: c1) $\mathsf{E}[|y|^2]\approx P_a$ and

Table 4.2: MMA(p,q) and cMMA(p): Values of $G_L^{\cdot\cdot}$ in terms of $\mathcal{S}_k$.

MMA(1,1): $G_L^k(\alpha,\beta)=\mathcal{S}_k(\widehat{\alpha},\widehat{\beta}),\quad (k=0,1,2)$ MMA(2,1): $G_L^{01}(\alpha,\beta)=\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})$ $G_L^{11}(\alpha,\beta)=\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})$ $G_L^{02}(\alpha,\beta)=\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})$ $G_L^{12}(\alpha,\beta)=\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})$ $G_L^{22}(\alpha,\beta)=\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})$	MMA(1,2): $G_L^{01k}(\alpha,\beta)=\mathcal{R}_k\mathcal{S}_0(\widehat{\alpha},\widehat{\beta})-\mathcal{S}_1(\widehat{\alpha},\widehat{\beta}),\ (k=1,2)$ $G_L^{11k}(\alpha,\beta)=\mathcal{R}_k\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})-\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})$ $G_L^{02k}(\alpha,\beta)=\mathcal{R}_k^2\mathcal{S}_0(\widehat{\alpha},\widehat{\beta})-2\mathcal{R}_k\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})$ $G_L^{12k}(\alpha,\beta)=\mathcal{R}_k^2\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})-2\mathcal{R}_k\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})$ $G_L^{22k}(\alpha,\beta)=\mathcal{R}_k^2\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-2\mathcal{R}_k\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})$
MMA(3,1): $G_L^{01}(\alpha,\beta)=\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})$ $G_L^{11}(\alpha,\beta)=\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})$ $G_L^{02}(\alpha,\beta)=\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})$ $G_L^{12}(\alpha,\beta)=\mathcal{S}_5(\widehat{\alpha},\widehat{\beta})$ $G_L^{22}(\alpha,\beta)=\mathcal{S}_6(\widehat{\alpha},\widehat{\beta})$	MMA(2,2): $G_L^{01}(\alpha,\beta)=R_L^2\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})-\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})$ $G_L^{11}(\alpha,\beta)=R_L^2\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})$ $G_L^{02}(\alpha,\beta)=R_L^4\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-2R_L^2\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_6(\widehat{\alpha},\widehat{\beta})$ $G_L^{12}(\alpha,\beta)=R_L^4\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})-2R_L^2\mathcal{S}_5(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_7(\widehat{\alpha},\widehat{\beta})$ $G_L^{22}(\alpha,\beta)=R_L^4\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})-2R_L^2\mathcal{S}_6(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_8(\widehat{\alpha},\widehat{\beta})$
cMMA(1): $c_k\equiv(-1)^k$ $G_L^{01k}(\alpha,\beta)=R_L\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})-c_k\mathcal{S}_2(\widehat{\alpha},\widehat{\beta}),(k=1,2)$ $G_L^{11k}(\alpha,\beta)=R_L\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-c_k\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})$ $G_L^{02k}(\alpha,\beta)=R_L^2\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-2c_kR_L\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})$ $G_L^{12k}(\alpha,\beta)=R_L^2\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})-2c_kR_L\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_5(\widehat{\alpha},\widehat{\beta})$ $G_L^{22k}(\alpha,\beta)=R_L^2\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})-2c_kR_L\mathcal{S}_5(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_6(\widehat{\alpha},\widehat{\beta})$	cMMA(3): $G_L^{01k}(\alpha,\beta)=R_L^3\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})-c_k\mathcal{S}_4(\widehat{\alpha},\widehat{\beta}),(k=1,2)$ $G_L^{11k}(\alpha,\beta)=R_L^3\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-c_k\mathcal{S}_5(\widehat{\alpha},\widehat{\beta})$ $G_L^{02k}(\alpha,\beta)=R_L^6\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-2c_kR_L^3\mathcal{S}_5(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_8(\widehat{\alpha},\widehat{\beta})$ $G_L^{12k}(\alpha,\beta)=R_L^6\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})-2c_kR_L^3\mathcal{S}_6(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_9(\widehat{\alpha},\widehat{\beta})$ $G_L^{22k}(\alpha,\beta)=R_L^6\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})-2c_kR_L^3\mathcal{S}_7(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_{10}(\widehat{\alpha},\widehat{\beta})$
cMMA(4): $G_L^{01}(\alpha,\beta)=R_L^4\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})-\mathcal{S}_5(\widehat{\alpha},\widehat{\beta})$ $G_L^{11}(\alpha,\beta)=R_L^4\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-\mathcal{S}_6(\widehat{\alpha},\widehat{\beta})$ $G_L^{02}(\alpha,\beta)=R_L^8\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-2R_L^4\mathcal{S}_6(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_{10}(\widehat{\alpha},\widehat{\beta})$ $G_L^{12}(\alpha,\beta)=R_L^8\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})-2R_L^4\mathcal{S}_7(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_{11}(\widehat{\alpha},\widehat{\beta})$ $G_L^{22}(\alpha,\beta)=R_L^8\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})-2R_L^4\mathcal{S}_8(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_{12}(\widehat{\alpha},\widehat{\beta})$	cMMA(5): $G_L^{01k}(\alpha,\beta)=R_L^5\mathcal{S}_1(\widehat{\alpha},\widehat{\beta})-c_k\mathcal{S}_6(\widehat{\alpha},\widehat{\beta}),(k=1,2)$ $G_L^{11k}(\alpha,\beta)=R_L^5\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-c_k\mathcal{S}_7(\widehat{\alpha},\widehat{\beta})$ $G_L^{02k}(\alpha,\beta)=R_L^{10}\mathcal{S}_2(\widehat{\alpha},\widehat{\beta})-2c_kR_L^5\mathcal{S}_7(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_{12}(\widehat{\alpha},\widehat{\beta})$ $G_L^{12k}(\alpha,\beta)=R_L^{10}\mathcal{S}_3(\widehat{\alpha},\widehat{\beta})-2c_kR_L^5\mathcal{S}_8(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_{13}(\widehat{\alpha},\widehat{\beta})$ $G_L^{22k}(\alpha,\beta)=R_L^{10}\mathcal{S}_4(\widehat{\alpha},\widehat{\beta})-2c_kR_L^5\mathcal{S}_9(\widehat{\alpha},\widehat{\beta})+\mathcal{S}_{14}(\widehat{\alpha},\widehat{\beta})$

c2) the initially assumed output-SNR (which was required in the computation of R_L) is close to its analytical value (4.6.1).[2] The analytical and simulated ISI/MSE traces for MMA(p,q) and cMMA(p) are depicted in Fig. 4.7-4.12. Each simulated trace is obtained as an ensemble average of over 200 Monte-Carlo realization with independent generation of noise and data symbols. Importantly, note that the analytical ISI/MSE trajectories are in good conformation with those obtained from Monte-Carlo experiments.

In Fig. 4.7, both MMA(2,1) and MMA(3,1) are providing faster convergence for 16-QAM than other three members. For 64-QAM (Fig. 4.8), the MMA(2,2) is converging

[2]Since the convolutional noise has energy $\mathsf{E}[|u|^2]=2\sigma_v^2$, we define

$$\text{output\_SNR}=\frac{P_a}{2\sigma_v^2}\approx\frac{P_a|t_\varsigma|^2}{P_a\sum_{i\neq\varsigma}|t_i|^2+P_\vartheta\|w\|_2^2} \tag{4.6.1}$$

where $\{t\}$ is joint channel-equalizer impulse response. The instantaneous estimation of output-SNR is not simple as it requires information about additive noise and residual ISI. In real scenarios, existing residual ISI estimation methods [22, 91] can be used to estimate output-SNR.

Table 4.3: MMA(p,q): Values of R_L and μ used in simulation.

	R_L			μ		
(p,q)	16	64	256	16	64	256
$(1,1)$	2.91	5.34	11.14	4.8e-4	1.2e-4	4.0e-5
$(1,2)$	2.50	5.25	10.63	2.3e-4	4.8e-5	7.0e-6
$(2,1)$	2.97	6.58	12.81	2.9e-4	3.0e-5	4.8e-6
$(2,2)$	2.86	6.08	12.34	3.4e-5	1.8e-6	7.0e-8
$(3,1)$	2.99	6.76	13.20	1.12e-4	1.0e-5	4.6e-7

Table 4.4: cMMA(p): Values of R_L and μ used in simulation.

	R_L			μ		
p	16	64	256	16	64	256
1	2.80	5.90	11.95	2.0e-4	1.2e-5	1.2e-6
2	2.86	6.08	12.34	4.0e-5	1.44e-6	6.5e-8
3	2.90	6.22	12.63	1.5e-5	2.0e-7	4.0e-9
4	2.92	6.32	12.87	6.0e-6	2.8e-8	3.0e-10
5	-	6.40	13.06	-	4.0e-9	2.0e-11

faster than MMA(2,1) while MMA(3,1) is still the fastest candidate. For 256-QAM (Fig. 4.9), MMA(2,2) is performing better than all other members. Noticeably, the performances of MMA(1,1) and MMA(1,2) are very poor for 64- and 256-QAM. In an another set of experiments, we simulated MMA(p,q) equalizers for a number of fractionally-spaced microwave channels [1] and we have noticed a very similar trend in performance as depicted in Fig. 4.7-4.9.

In Fig. 4.10-4.12, the cMMA(p) is depicted to exhibit a very consistent behavior for all the three addressed QAM sizes. We note that, with an increased value of p, the equalizer is found to be converging faster while requiring more computation in its error-function. It clearly indicates that the cMMA(p) is capable of providing a performance-complexity trade-off. Also note that, unlike KDA which worked satisfactorily only for small QAM sizes (4/16-QAM), the proposed cMMA(p) is working satisfactorily for both small and large QAM sizes.

Due to the nonlinear structure of MMA(p,q), it is difficult to explain why MMA(2,1), MMA(3,1) and MMA(2,2) performed respectively better than others for 16-, 64- and 256-QAM. Or why MMA(1,1) performed poorer for constellations higher than 16-QAM. We intuitively realized that an MMA with a *higher* dispersion constant (R_L) is capable of performing *better* in terms of convergence. For an equalizer implementing MMA(p,q), this intuition is found to be true (only) for 16-QAM, but, on the other hand, we find this intuition very true for cMMA(p), where a higher p always yielded a larger dispersion constant and faster convergence.

4.7 Summary

We have proposed two families of multimodulus Bussgang-type adaptive algorithms, MMA($p, 2$) and cMMA(p), for joint blind equalization and carrier-phase recovery of square-QAM signals over complex-valued transmission channel. The main contribution resided in the generalization of an existing dispersion-directed cost-function as well as the modification in a convex cost-function leading to newer algorithms capable of yielding faster convergence. Evaluation of equalizer gain and dynamic convergence has been described in detail and also shown to be in conformation with simulation results.

Clearly, based on the results reported in our study, it is possible to achieve faster convergence than known adaptive equalizers (especially for 16/64-QAM) and the discussed dynamic convergence analysis can help us select the best equalizer among the members of MMA(p, q) and cMMA(p) for the given channel, equalizer parameters (length, step-size, initialization), QAM-signal, noise condition and computational requirements. Finally, we like to comment that the performances of MMA($p, 2$) and cMMA(p) have been found to be more robust to channel noise than those of MMA($p, 1$).

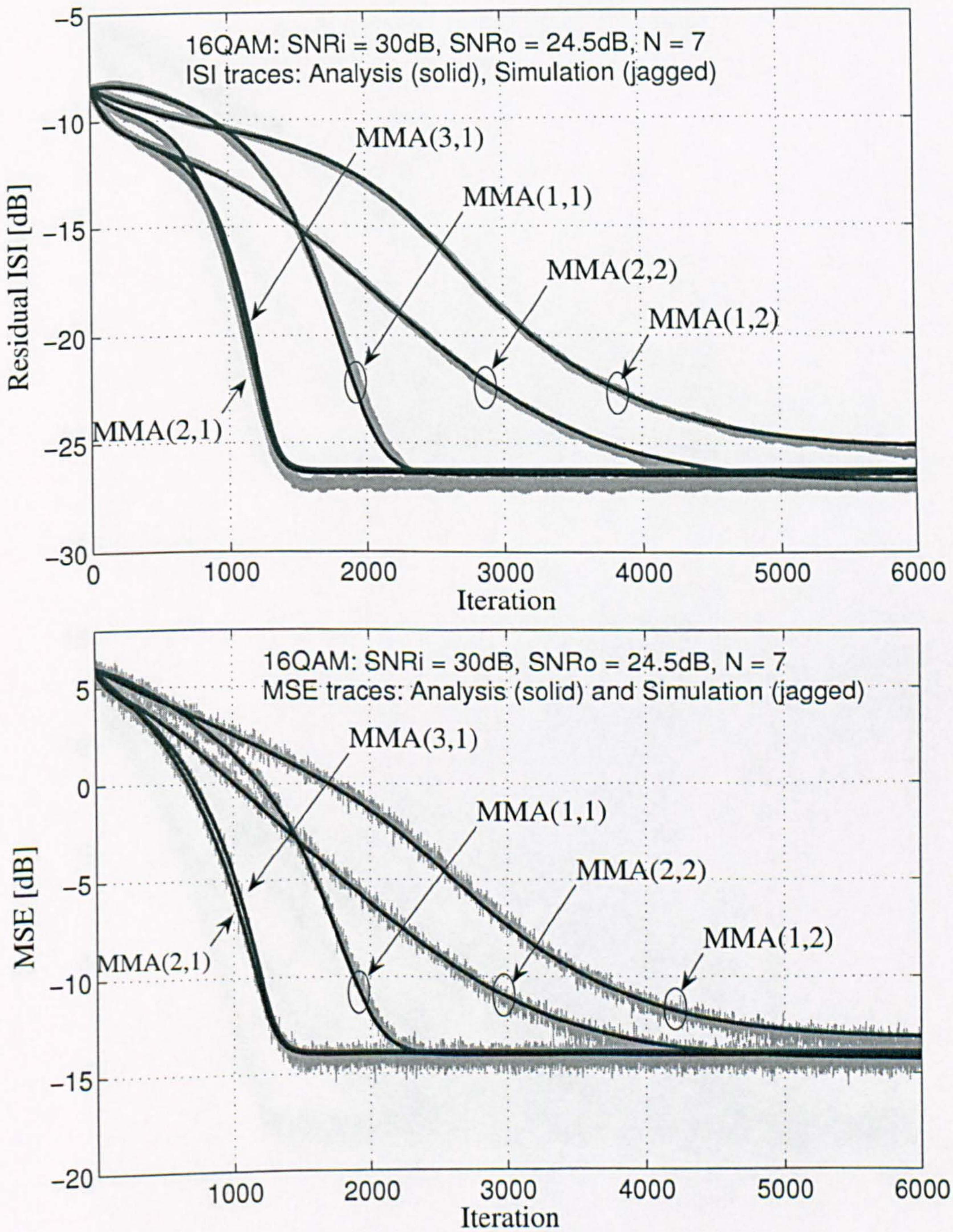


Figure 4.7: Simulated and analytical ISI/MSE traces for MMA(p, q) with 16-QAM.

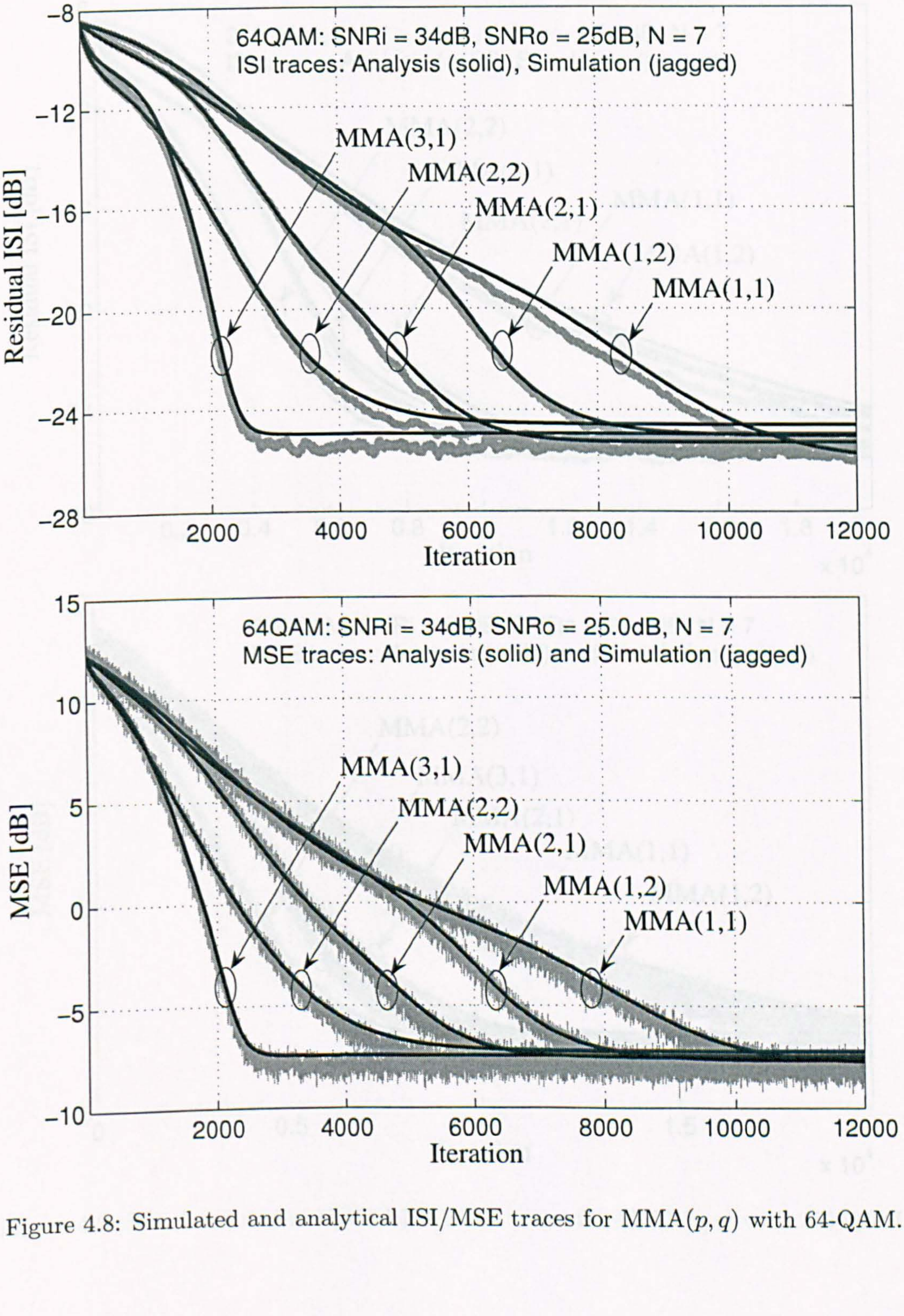


Figure 4.8: Simulated and analytical ISI/MSE traces for MMA(p, q) with 64-QAM.

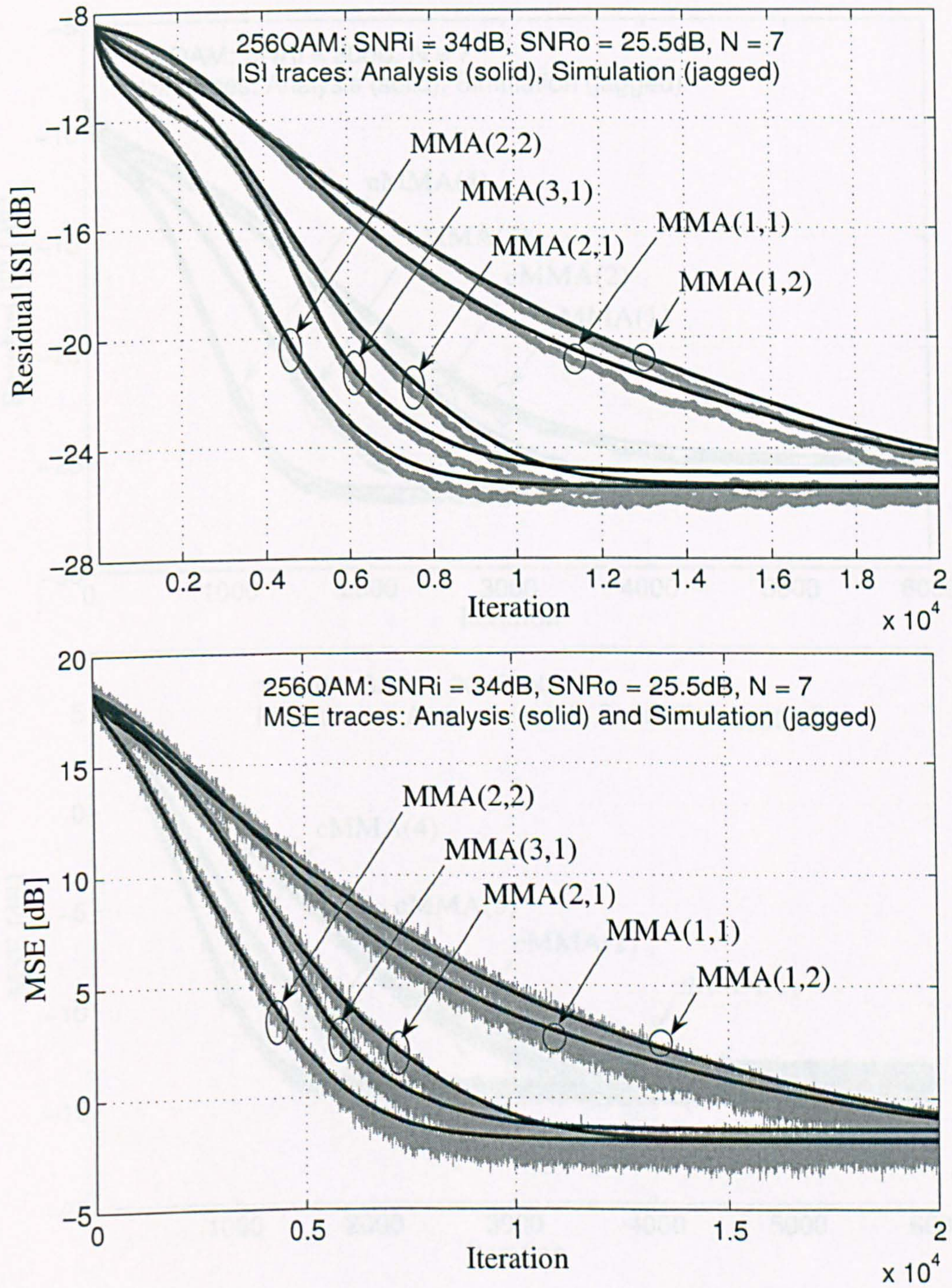


Figure 4.9: Simulated and analytical ISI/MSE traces for MMA(p,q) with 256-QAM.

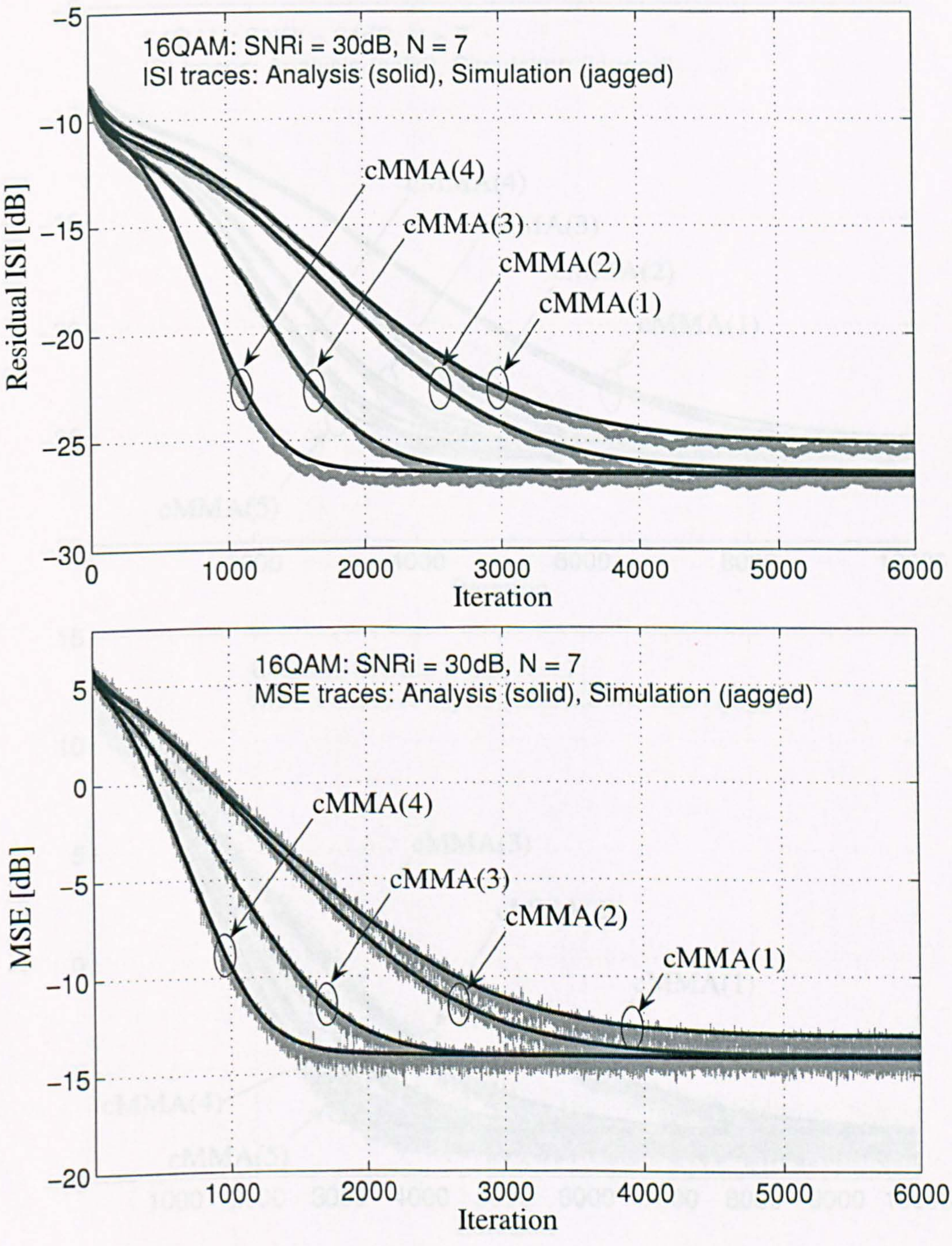


Figure 4.10: Simulated and analytical ISI/MSE traces for cMMA(p) with 16-QAM.

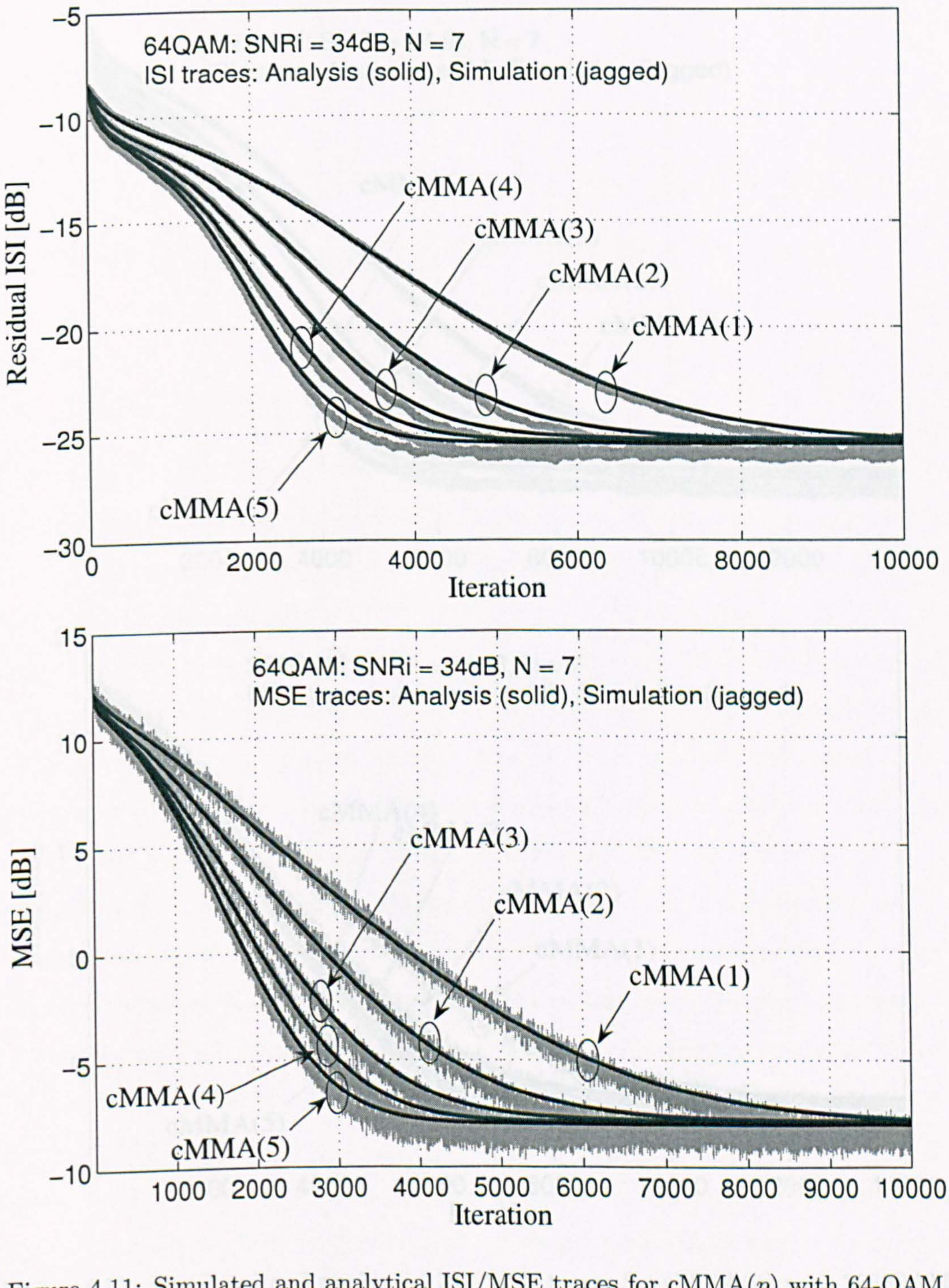


Figure 4.11: Simulated and analytical ISI/MSE traces for cMMA(p) with 64-QAM.

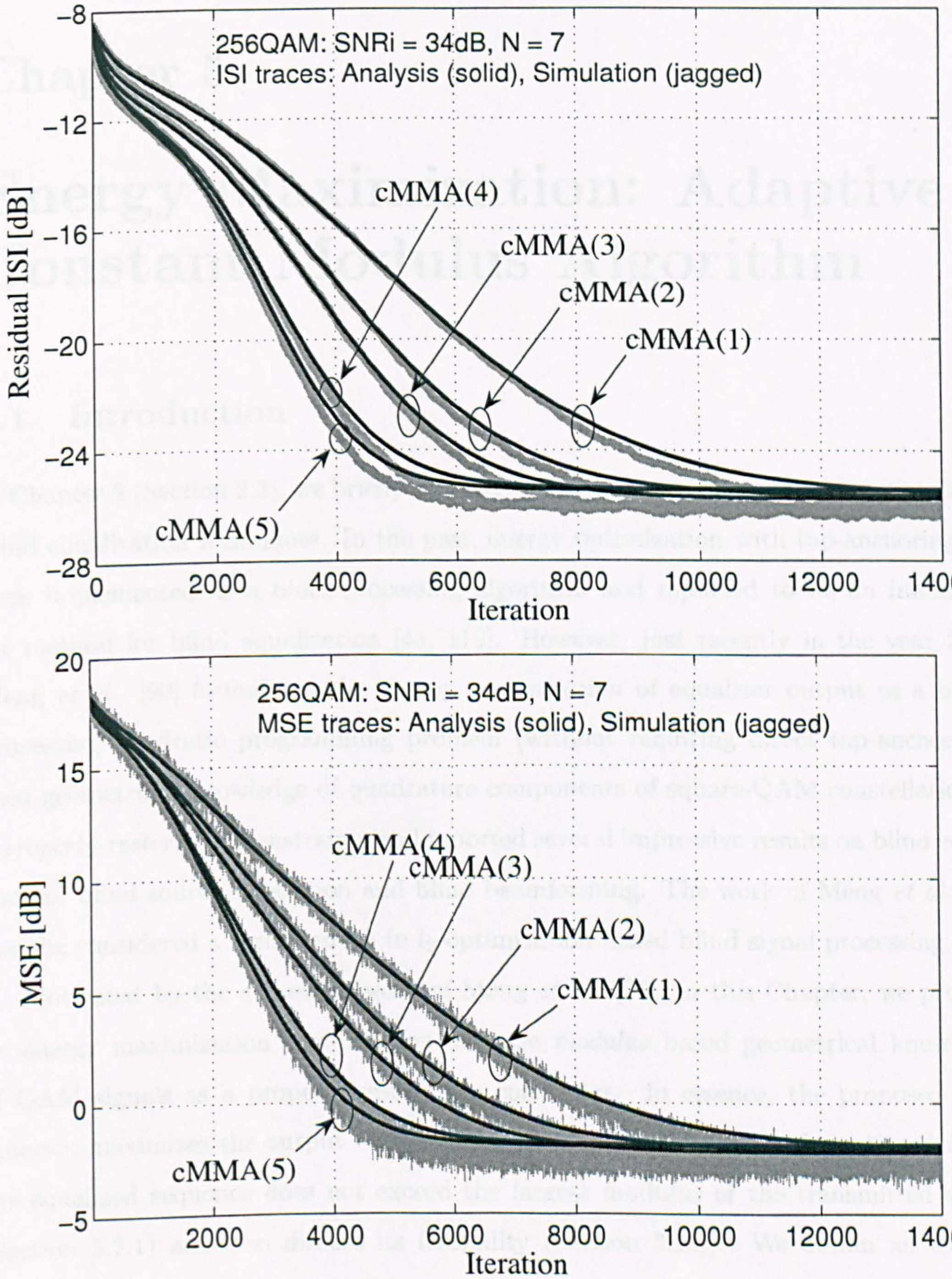


Figure 4.12: Simulated and analytical ISI/MSE traces for cMMA(p) with 256-QAM.

Chapter 5

Energy Maximization: Adaptive Constant Modulus Algorithm

5.1 Introduction

In Chapter 2 (Section 2.3), we briefly summarized constrained energy optimization based blind equalization techniques. In the past, energy optimization with tap-anchoring has been implemented as a block-processing algorithm and reported to be an inadmissible method for blind equalization [43, 119]. However, just recently in the year 2009, Meng *et al.* [90] formulated the *energy maximization* of equalizer output as a block-processing quadratic programming problem (without requiring direct tap-anchoring), used geometrical knowledge of quadrature components of square-QAM constellation as a *property-restoration* constraints and reported several impressive results on blind equalization, blind source separation and blind beamforming. The work of Meng *et al.* [90] may be considered a *breakthrough* in l_2-optimization based blind signal processing.

Motivated by the successful work of Meng *et al.* [90], in this Chapter, we propose an energy maximization cost-function and use *modulus* based geometrical knowledge of QAM signals as a property-restoration constraint. In essence, the proposed cost-function maximizes the output energy under the constraint that the *largest modulus* of the equalized sequence does not exceed the largest modulus of the transmitted signal (Section 5.2.1) and also discuss its feasibility (Section 5.2.2). We obtain an elegant adaptive constant modulus algorithm by optimizing a modified version of the proposed cost-function (Section 5.2.4). We show that the proposed algorithm is fully capable of opening the closed-eye with successful restoration of signal energy. Finally, we provide evidence of good performance in comparison to existing established adaptive methods, like CMA(2,2) and three of its variants through simulations (Section 5.3).

5.2 The Proposed Cost-Function

5.2.1 A New Quadratic Cost with Convex Constraint

Let $t_l = \sum_{k=-K}^{K} h_k w_{n,l-k}^*$ be the lth element of overall impulse response $\{t_l\}$. Now consider the following function:

$$g(\boldsymbol{w}) \triangleq R_a^{-1} \max(\{|y_n|\}) = \sum_l |t_l| = \|\boldsymbol{t}\|_1 \tag{5.2.1}$$

where $R_a = \max(\{|a_n|\})$ is the *largest modulus* of transmitted signal a_n. We have the following theorem:

Theorem 5.1: $g(\boldsymbol{w})$ is *convex* in $\boldsymbol{w}$ [113].
Proof: Suppose, we have $\|\boldsymbol{t}\|_1 = \sum_l |\sum_k h_k w_{l-k}^*|$, noting the $\boldsymbol{t}$ weights are linear function of the $\boldsymbol{w}$ equalizer taps; so write $\boldsymbol{t} \equiv \boldsymbol{t}(\boldsymbol{w})$. Let $\boldsymbol{w}^a \in l_1$, $\boldsymbol{w}^b \in l_1$, and $0 \le \eta \le 1$. Then

$$\begin{aligned} g(\boldsymbol{w}) &= \|\boldsymbol{t}((1-\eta)\boldsymbol{w}^a + \eta\boldsymbol{w}^b)\|_1 \\ &= \sum_l \left| \sum_k h_k((1-\eta)(w_{l-k}^a)^* + \eta(w_{l-k}^b)^*) \right| \\ &\le (1-\eta)\sum_l \left| \sum_k h_k(w_{l-k}^a)^* \right| + \eta \sum_l \left| \sum_k h_k(w_{l-k}^b)^* \right| \\ &= (1-\eta)\|\boldsymbol{t}(\boldsymbol{w}^a)\|_1 + \eta\|\boldsymbol{t}(\boldsymbol{w}^b)\|_1. \ \Box\Box\Box \end{aligned} \tag{5.2.2}$$

We are now ready to formulate l_2-optimization with constraint on the modulus for blind channel equalization as follows:

$$\max_{\boldsymbol{w}} \mathsf{E}\left[|y_n|^2\right] \quad \text{s.t.} \quad g(\boldsymbol{w}) = R_a^{-1} \max(\{|y_n|\}) \le 1. \tag{5.2.3}$$

Note that this problem can be formed using only the channel output $\boldsymbol{x}_n$, which implies its applicability in a blind equalization setting. The cost-function to be maximized in (5.2.3) is quadratic, and the feasible region is a convex set.

5.2.2 Comparison with Allen-Mazo Cost and Admissibility

We define ISI $= \left(\sum_l |t_l|^2 / \max\left(\{|t_l|^2\}\right)\right) - 1$. Introducing the channel autocorrelation matrix $\boldsymbol{\mathcal{H}}$, whose (i,j) element is given by $\mathcal{H}_{ij} = \sum_k h_{k-i} h_{k-j}^*$, $i,j \in \{-K, K\}$, we can show $\sum_l |t_l|^2 = \boldsymbol{w}_n^H \boldsymbol{\mathcal{H}} \boldsymbol{w}_n$. The equalizer has to make one of the coefficients of $\{t_l\}$, say,

$t_0 = \boldsymbol{w}_n^H \boldsymbol{\mathcal{H}}$, to be unity and others to be zero, where $\boldsymbol{h} = [h_{-K}, h_{-K+1}, \cdots, h_0, h_1, \cdots, h_K]^T$ and we assume $N = 2K+1$. We obtain [9]

$$\mathrm{ISI}_n = \boldsymbol{w}_n^H \boldsymbol{\mathcal{H}} \boldsymbol{w}_n \cdot |\boldsymbol{w}_n^H \boldsymbol{h}|^{-2} - 1. \tag{5.2.4}$$

Now consider Allen-Mazo cost-function [10]:

$$\boxed{\min_{\boldsymbol{w}} \mathsf{E}\left[|y_n|^2\right] \quad \text{s.t.} \quad w_0 = 1} \tag{5.2.5}$$

Since $\mathsf{E}[|y_n|^2] = P_a \sum_l |t_l|^2$ and $P_a = \mathsf{E}[|a|^2]$, we can re-express (5.2.5) in an equivalent form as follows:

$$\text{EQ.A}: \quad \min_{\boldsymbol{w}} \sum_l |t_l|^2 \quad \text{s.t.} \quad w_i = 1, \tag{5.2.6}$$

where EQ.A is a label, and i is either zero or $N-1$. Using Lagrangian multiplier λ, we optimize $\sum_l |t_l|^2 + \lambda w_i$ with respect to $\boldsymbol{w}_n^*$ to get $\boldsymbol{w}_n = -\lambda \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{g}$, where $[g]_k = \delta(k-i)$; it leads to

$$\mathrm{ISI}_{\mathrm{EQ.A}} = \boldsymbol{g}^H \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{g} \cdot |\boldsymbol{h}^H \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{g}|^{-2} - 1. \tag{5.2.7}$$

Now consider our proposed cost-function with equality constraint

$$\boxed{\max_{\boldsymbol{w}} \mathsf{E}\left[|y_n|^2\right] \quad \text{s.t.} \quad \max(\{|y_n|\}) \leq R_a} \tag{5.2.8}$$

The cost (5.2.8) can be written in an equivalent form

$$\max_{\boldsymbol{w}} \sum_l |t_l|^2 \quad \text{s.t.} \quad \sum_l |t_l| \leq 1. \tag{5.2.9}$$

Unfortunately, due to the nonlinear constraint in (5.2.9), it is difficult to analyze it in matrix form. Note that, however, by assuming $\sum_{l \neq 0} |t_l| \ll |t_0|$, we can re-express (5.2.9) as follows:

$$\text{EQ.B}: \quad \min_{\boldsymbol{w}} \sum_l |t_l|^2 \quad \text{s.t.} \quad t_0 = 1. \tag{5.2.10}$$

Incorporating Lagrangian multiplier λ in (5.2.10) and differentiating with respect to $\boldsymbol{w}_n^*$, we get $\boldsymbol{w}_n = -\lambda \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{h}$; this solution yields

$$\mathrm{ISI}_{\mathrm{EQ.B}} = \boldsymbol{h}^H \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{h} \cdot |\boldsymbol{h}^H \boldsymbol{\mathcal{H}}^{-1} \boldsymbol{h}|^{-2} - 1. \tag{5.2.11}$$

To appreciate the possible benefit of solution (5.2.11) over (5.2.7), consider a channel $h_{-1} = 1 - \varepsilon$, $h_0 = \varepsilon$ and $h_1 = 0$, where $0 \leq \varepsilon \leq 1$. Without equalizer, we have

$$\mathrm{ISI} = \frac{(1-\varepsilon)^2 + \varepsilon^2}{\max(1-\varepsilon, \varepsilon)^2} - 1. \tag{5.2.12}$$

The ISI approaches zero when ε is either zero or unity. Since $\varepsilon < 0.5$ places the zero inside the unit circle of the z-plane (no precursors) and $\varepsilon > 0.5$ places the zero outside the unit circle (no tail distortion); the initialization of equalizer EQ.A requires

$$[w_0 \quad w_1]^T = \begin{cases} [1 \quad 0]^T & \text{when } 0 < \varepsilon < 0.5 \\ [0 \quad 1]^T & \text{when } 0.5 < \varepsilon < 1 \end{cases} . \tag{5.2.13}$$

From solution (5.2.7), the ISI for equalizer EQ.A is obtained as

$$\mathrm{ISI}_{\mathrm{EQ.A}} = \begin{cases} \varepsilon^2(1-\varepsilon)^{-6}\left(2 - 8\varepsilon + 14\varepsilon^2 - 12\varepsilon^3 + 5\varepsilon^4\right)\Big|_{\varepsilon<0.5} \\ (1-\varepsilon)^2\varepsilon^{-6}\left(1 - 4\varepsilon + 8\varepsilon^2 - 8\varepsilon^3 + 5\varepsilon^4\right)\Big|_{\varepsilon>0.5} \end{cases} \tag{5.2.14}$$

On the other hand, from (5.2.11), the ISI for EQ.B is obtained as

$$\mathrm{ISI}_{\mathrm{EQ.B}} = \frac{\varepsilon^2(1-\varepsilon)^2}{1 - 4\varepsilon + 6\varepsilon^2 - 4\varepsilon^3 + 2\varepsilon^4}. \tag{5.2.15}$$

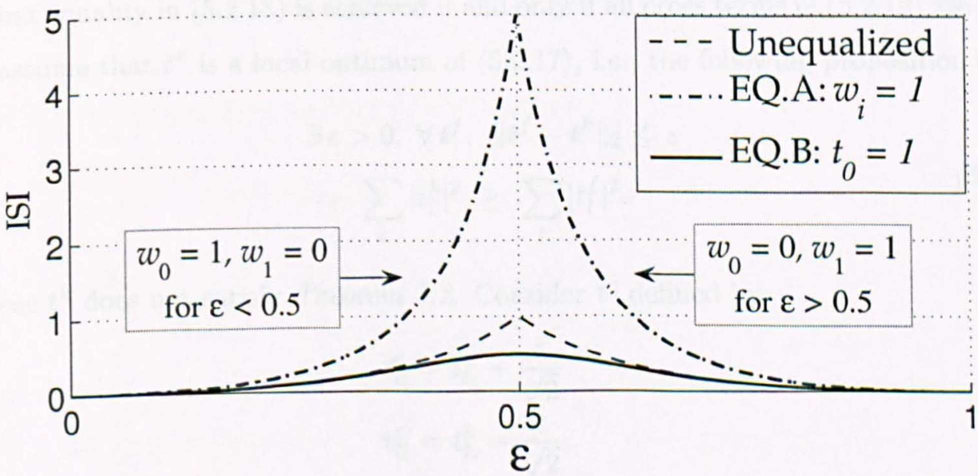


Figure 5.1: ISI of EQ.A and EQ.B compared to the unequalized case.

We depict expressions (5.2.14) and (5.2.15) in Fig. 5.1; note that the equalizer EQ.A has actually made the ISI worse than if it were not used. On the other hand, the proposed equalizer EQ.B has reduced residual ISI and it appears admissible. Note that the problem (5.2.3) is non-convex and may have multiple local maxima. Nevertheless, we have the following theorem.

Theorem 5.2: Assume $\boldsymbol{w}^\dagger$ is a local optimum in (5.2.3), and $\boldsymbol{t}^\dagger$ is the corresponding total channel equalizer impulse-response and channel noise is negligible. Let $l, l^\dagger \in \{-K, \cdots, K\}$, then it holds $|t_l| = \delta_{l-l^\dagger}$.

Proof: Without loss of generality we assume that the channel and equalizer are real-valued. We re-write (5.2.9) with in-equality constraint as follows:

$$\boldsymbol{w}^{\dagger} = \arg\max_{\boldsymbol{w}} \sum_{l} |t_l|^2 \ \text{ s.t. } \ \sum_{l} |t_l| \leq 1. \tag{5.2.16}$$

Now consider the following quadratic problem in $\boldsymbol{t}$

$$\boldsymbol{t}^{\dagger} = \arg\max_{\boldsymbol{t}} \sum_{l} |t_l|^2 \ \text{ s.t. } \ \sum_{l} |t_l| \leq 1. \tag{5.2.17}$$

Assume $\boldsymbol{t}^f$ is a feasible solution to (5.2.17). We have

$$\sum_{l} |t_l|^2 \leq \left(\sum_{l} |t_l|\right)^2 \leq 1 \tag{5.2.18}$$

and

$$\left(\sum_{l} |t_l|\right)^2 = \sum_{l} |t_l|^2 + \sum_{l_1} \sum_{l_2,\, l_2 \neq l_1} |t_{l_1} t_{l_2}| \tag{5.2.19}$$

The first equality in (5.2.18) is achieved if and only if all cross terms in (5.2.19) are zeros. Now assume that $\boldsymbol{t}^k$ is a local optimum of (5.2.17), i.e., the following proposition holds

$$\begin{aligned} &\exists \varepsilon > 0,\ \forall \boldsymbol{t}^f,\ \ \|\boldsymbol{t}^f - \boldsymbol{t}^k\|_2 \leq \varepsilon \\ &\Rightarrow \sum_{l} |t_l^k|^2 \geq \sum_{l} |t_l^f|^2. \end{aligned} \tag{5.2.20}$$

Suppose $\boldsymbol{t}^k$ does not satisfy *Theorem 5.2.* Consider $\boldsymbol{t}^c$ defined by

$$\begin{aligned} t_{l_1}^c &= t_{l_1}^k + \frac{\varepsilon}{\sqrt{2}} \\ t_{l_2}^c &= t_{l_2}^k - \frac{\varepsilon}{\sqrt{2}} \\ t_l^c &= t_l^k,\ \ l \neq l_1, l_2. \end{aligned}$$

We also assume that $t_{l_2}^k < t_{l_1}^k$. Next, we have $\|\boldsymbol{t}^c - \boldsymbol{t}^k\|_2 = \varepsilon$, and $\sum_l |t_l^c| = \sum_l |t_l^k| \leq 1$. However, one can observe that

$$\sum_{l} |t_l^k|^2 - \sum_{l} |t_l^c|^2 = \sqrt{2}\varepsilon \left(t_{l_2}^k - t_{l_1}^k\right) - \varepsilon^2 < 0. \tag{5.2.21}$$

which means $\boldsymbol{t}^k$ is not a local optimum to (5.2.17). Therefore, we have shown by a counterexample that all the local maxima of (5.2.17) should satisfy *Theorem 5.2.* □□□

5.2.3 Energy Maximization With a Differentiable Constraint

The cost (5.2.8) can be expressed as a quadratic problem using B blocks of channel observations

$$\min_{w} -w^H\left(\frac{1}{B}\sum_{n=1}^{B} x_n x_n^H\right) w \quad \text{s.t. } \left|w^H x_n\right| \le R_a,\ \forall n \tag{5.2.22}$$

and can be solved by exploiting established tools [84]. However, due to the widespread popularity of adaptive algorithms in practical receivers [131], we are interested in optimizing (5.2.8) using a stochastic gradient-based method.

For such an implementation, we modify (5.2.8) to involve a *differentiable* constraint and propose the following differentiable cost-function:

$$\boxed{\max_{w} \mathsf{E}\left[|y_n|^2\right] \quad \text{s.t.} \quad \text{fmax}(R_a, |y_n|) = R_a} \tag{5.2.23}$$

where, for some $a, b \in \mathbb{C}$, the function fmax is defined as[1]

$$\text{fmax}(|a|,|b|) \equiv \frac{\big||a|+|b|\big| + \big||a|-|b|\big|}{2} = \begin{cases} |a|, \text{ if } |a| \ge |b| \\ |b|, \text{ otherwise.} \end{cases}$$

Employing the Lagrangian multiplier λ, we obtain

$$\max_{w}\ \mathsf{E}[|y_n|^2] + \lambda(\text{fmax}(R_a, |y_n|) - R_a). \tag{5.2.24}$$

Referring to Fig. 5.2, it is clear that if the equalizer output y_n is inside the circular region, centered at origin with radius R_a, then the constraint $\text{fmax}(R_a, |y_n|) - R_a = 0$, and the cost (5.2.24) simply involves the maximization of $\mathsf{E}[|y_n|^2]$. However, if $|y_n| > R_a$, then the constraint is violated and the computation of new update w_{n+1} (with the aid of Lagrange multiplier λ) will require to bring the *a posteriori* output $s_n = w_{n+1}^H x_n$ inside or onto the perimeter of circle such that $\text{fmax}(R_a, |s_n|) - R_a = 0$. In this manner, the cost (5.2.23) will not only be able to *maximize* $\mathsf{E}[|y_n|^2]$ but also *minimize* the dispersion in y_n away from the *constant modulus* R_a.

[1]The fmax is differentiable (below **sgn** denotes the **signum** function)

$$\frac{\partial\, \text{fmax}(|a|,|b|)}{\partial a^*} = \frac{a\big(1+\mathbf{sgn}(|a|-|b|)\big)}{2|a|} = \begin{cases} a/|a|, & \text{if } |a| > |b| \\ a/(2|a|), & \text{if } |a| = |b| \\ 0, & \text{if } |a| < |b| \end{cases}$$

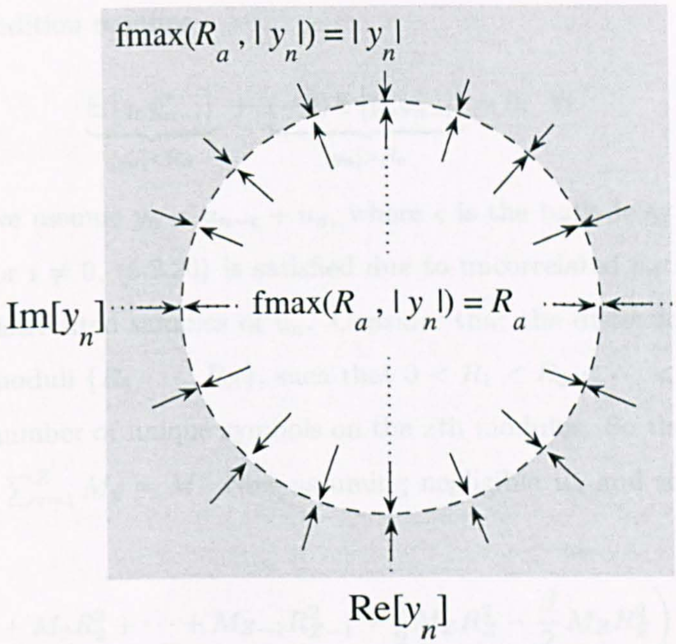


Figure 5.2: Optimization of (5.2.23) results in forcing the instantaneous equalizer output y_n radially towards the modulus R_a (direction of which is shown by a set of arrows with arbitrary magnitudes).

5.2.4 Derivation of an Adaptive Algorithm: β-CMA

Let $\max_w \mathsf{E}[\mathcal{J}]$, $\mathcal{J} \in \mathbb{R}^+$, be a maximization problem, for which the (stochastic approximate) gradient-based solution is obtained as $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\,\partial\mathcal{J}/\partial\boldsymbol{w}_n^*$, where asterisk $(*)$ denotes the complex conjugate of the base entity and $\mu > 0$ is a small step-size. Note that $\partial\mathcal{J}/\partial\boldsymbol{w}_n^* = (\partial\mathcal{J}/\partial y_n^*)^*\boldsymbol{x}_n$, and $\partial\mathcal{J}/\partial y_n^* = 0.5\,(\partial\mathcal{J}/\partial y_{R,n} + \jmath\,\partial\mathcal{J}/\partial y_{I,n})$. The gradient-based maximization of cost (5.2.24) is readily obtained as:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\left(1 + \frac{\lambda}{4}\frac{g_n}{|y_n|}\right) y_n^*\boldsymbol{x}_n \tag{5.2.25}$$

where $g_n \equiv 1 + \mathrm{sgn}\,(|y_n| - R_a)$. If $|y_n| < R_a$, then $g_n = 0$ and we have $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu y_n^*\boldsymbol{x}_n$. If $|y_n| = R_a$, then the *property-restoration condition* $(\max(\{|y_n|\}) = R_a)$ is satisfied and we *stop* the adaptation. Since $g_n = 1$, we require $\lambda = -4|y_n|$ to ensure $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n$. Otherwise when $|y_n| > R_a$, with $g_n = 2$, we get $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\,(1 + \lambda/(2\,|y_n|))\,y_n^*\boldsymbol{x}_n$. As mentioned earlier, in this case, we have to compute λ such that s_n lies inside the circular region without sacrificing the output energy. Such an update can be realized by *minimizing* $|y_n|^2$ and *satisfying* the Bussgang condition. One of the possibilities is $\lambda = -2(1+\beta)|y_n|$, $\beta > 0$, which leads to $\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu(-\beta)y_n^*\,\boldsymbol{x}_n$.

The Bussgang condition requires

$$\underbrace{\mathsf{E}\left[y_n y_{n-i}^*\right]}_{|y_n|<R_a} + \underbrace{(-\beta)\,\mathsf{E}\left[y_n y_{n-i}^*\right]}_{|y_n|>R_a} = 0, \ \ \forall i. \tag{5.2.26}$$

In steady-state, we assume $y_n = a_{n-\varsigma} + u_n$, where ς is the bulk-delay and u_n is convolutional noise. For $i \neq 0$, (5.2.26) is satisfied due to uncorrelated a_n and independent-and-identically distributed samples of u_n. Consider that the distortion free M-symbol a_n comprises Z moduli $\{R_1, \cdots, R_Z\}$, such that $0 < R_1 < R_2 < \cdots < R_{Z-1} < R_Z$, and M_z denotes the number of unique symbols on the zth modulus. So the largest modulus is $R_Z = R_a$ and $\sum_{z=1}^{Z} M_z = M$. Now assuming negligible u_n and solving (5.2.26) for $i = 0$, we get

$$\frac{1}{M}\left(M_1R_1^2 + M_2R_2^2 + \cdots + M_{Z-1}R_{Z-1}^2 + \frac{1}{2}M_ZR_Z^2 - \frac{\beta}{2}\,M_ZR_Z^2\right) = 0 \tag{5.2.27}$$

The last two terms indicate that, when $|y_n|$ is close to R_Z, it would be equally likely to update in either direction. Noting that

$$\frac{1}{M}\sum_{z=1}^{Z} M_zR_z^2 = P_a, \tag{5.2.28}$$

is the energy of the transmitted signal. Next the simplification of (5.2.27) gives a dimensionless value for β

$$\boxed{\beta = 2\frac{M}{M_Z}\frac{P_a}{R_a^2} - 1} \tag{5.2.29}$$

Finally, we summarize our algorithm as follows:

$$\boxed{\begin{aligned} & w_{n+1} = w_n + \mu\,\mathrm{f}\,y_n^*\,x_n, \\ & \mathrm{f} = \begin{cases} 1, & \text{if } |y_n| < R_a \\ 0, & \text{if } |y_n| = R_a \\ -\beta, & \text{otherwise.} \end{cases} \end{aligned}} \tag{5.2.30}$$

Note that the resulting error-function of algorithm (5.2.30) has **1)** finite derivative at the origin, **2)** becomes zero solely at R_a, **3)** is increasing for $|y_n| < R_a$ and **4)** decreasing for $|y_n| > R_a$. In [11], these properties **1)-4)** have been tabulated as essential features of a *generic CMA*. This motivates us to denote (5.2.30) as β-CMA.

5.2.5 Cost-Function Interpretation

If the error-function in (5.2.30) is integrated back with respect to y_n, then the following cost-function is obtained:

$$J_{\beta\text{\_CMA}} = \max_{w} \left\{ \underbrace{\mathsf{E}\left[|y_n|^2\right]}_{|y_n|<R_a} + \underbrace{(-\beta)\mathsf{E}\left[|y_n|^2\right] + c}_{|y_n|>R_a} \right\}. \tag{5.2.31}$$

where c is the constant of integration. So, depending on the value of $|y_n|$ whether it is less or greater than R_a, we need to respectively maximize or minimize the mean equalizer energy. In Fig. 5.3, we depict the mesh and quiver plot of the cost (5.2.31) to demonstrate how the cost maximizes the energy of the equalized sequence while minimizing the dispersion away from the modulus R_a.

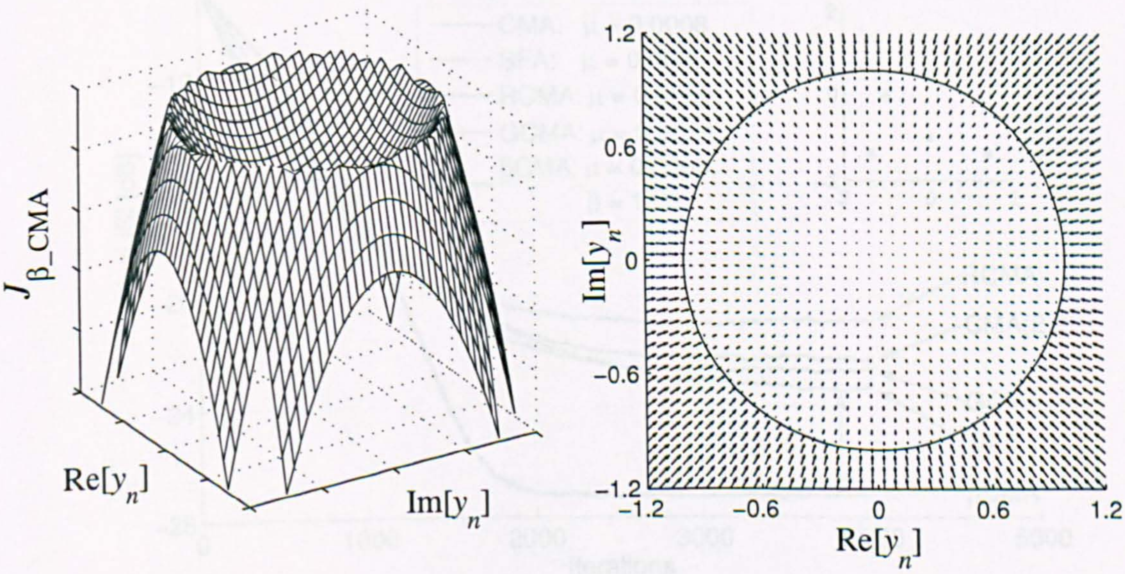


Figure 5.3: Mesh (left) and quiver (right) plots for an arbitrary signal with $R_a = 1$, $\beta = 2$ and $c = 3$. In mesh plot, note that the cost is maximized when $|y_n| = R_a$. In quiver plot, note that the arrows of gradient vectors are directing radially towards the modulus as predicted in Fig. 5.2.

5.3 Simulation Results

5.3.1 Experiment 1: ISI Performance with TSE

We compare β-CMA with the traditional CMA(2,2) [53] and three of its variants: (unnormalized) relaxed CMA (RCMA) [125, Eq.(14)], Shtrom-Fan algorithm (SFA) [122, Eq.(4.7)] and generalized CMA (GCMA) [28, Eq.(11)].[2] We consider transmission of amplitude-phase shift-keying (APSK) signals over a symbol-space complex-valued voice-band telephonic channel (taken from [105]) and evaluate the average ISI traces (at SNR = 30dB). We use a seven-tap equalizer with central spike initialization. Results are summarized in Fig. 5.4-5.6 using three different APSK signals. Note that the proposed β-CMA performed significantly better with much lower ISI floor than other counterparts. Also refer to Fig. 5.7, where we have depicter the scatter plots for converged constellations for all algorithms for 16-APSK. Note that the constellation obtained form proposed algorithm is more aggregated compared to others. In Table 5.1, we provide the average energy recovered by the addressed equalizers in this experiments. Notably, the proposed equalizer β-CMA has recovered the highest amount of energy.

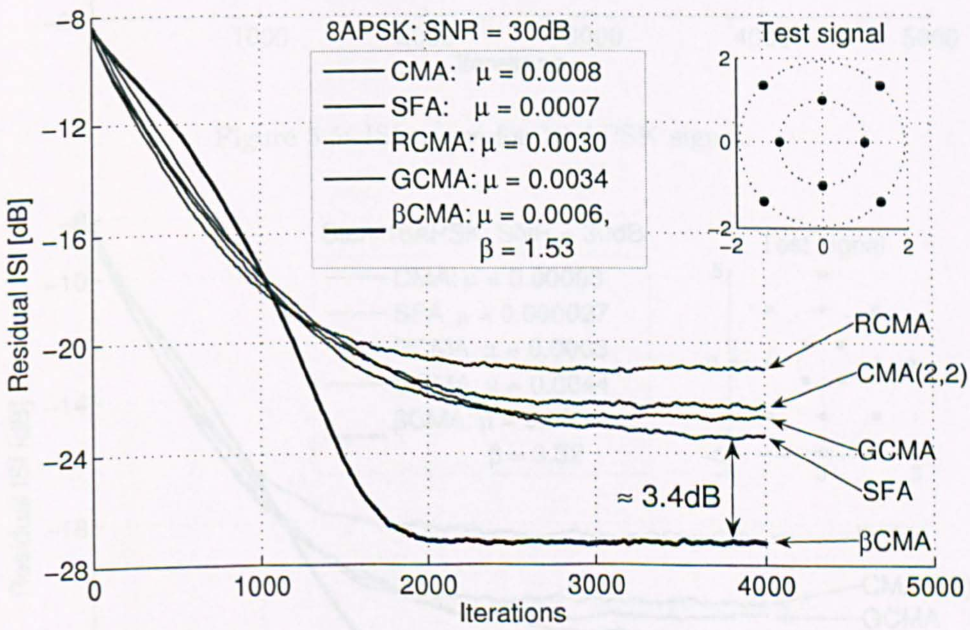


Figure 5.4: ISI traces for 8-APSK signal.

[2]The following four algorithms have been used for comparison with proposed algorithm (5.2.30):

$$\text{RCMA: } \boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\left(R_1 - |y_n|\right) y_n^* \boldsymbol{x}_n, \quad \text{where } R_1 = \mathsf{E}[|a|^3]/\mathsf{E}[|a|^2].$$

$$\text{CMA(2,2): } \boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\left(R_2^2 - |y_n|^2\right) y_n^* \boldsymbol{x}_n, \quad \text{where } R_2^2 = \mathsf{E}[|a|^4]/\mathsf{E}[|a|^2].$$

$$\text{SFA: } \boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\left(R\,\widehat{\zeta}_n - |y_n|^2\right) y_n^* \boldsymbol{x}_n, \quad \text{where } R = \mathsf{E}[|a|^4]/\mathsf{E}^2[|a|^2].$$

$$\text{GCMA: } \boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\left(\widehat{\zeta}_n - |y_n|^2\right)\left(\widehat{\zeta}_n\, y_n^* \boldsymbol{x}_n - |y_n|^2 \widehat{\boldsymbol{\chi}}_n\right)\widehat{\zeta}_n^{-3},$$

$$\text{where } \widehat{\zeta}_n = \widehat{\zeta}_{n-1} + \frac{1}{n}\left(|y_n|^2 - \widehat{\zeta}_{n-1}\right) \quad \text{and} \quad \widehat{\boldsymbol{\chi}}_n = \widehat{\boldsymbol{\chi}}_{n-1} + \frac{1}{n}\left(y_n^* \boldsymbol{x}_n - \widehat{\boldsymbol{\chi}}_{n-1}\right).$$

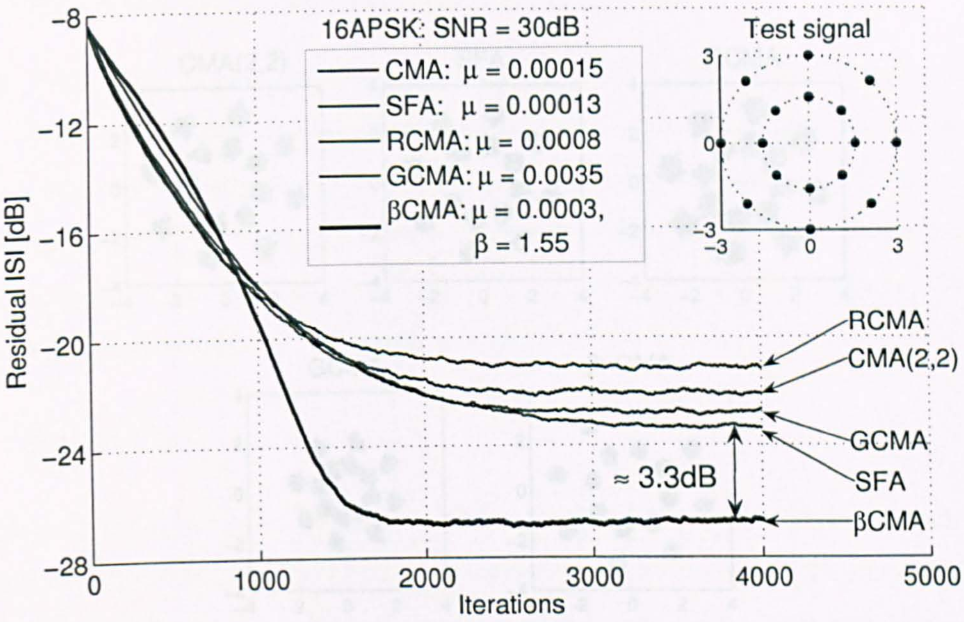


Figure 5.5: ISI traces for 16-APSK signal.

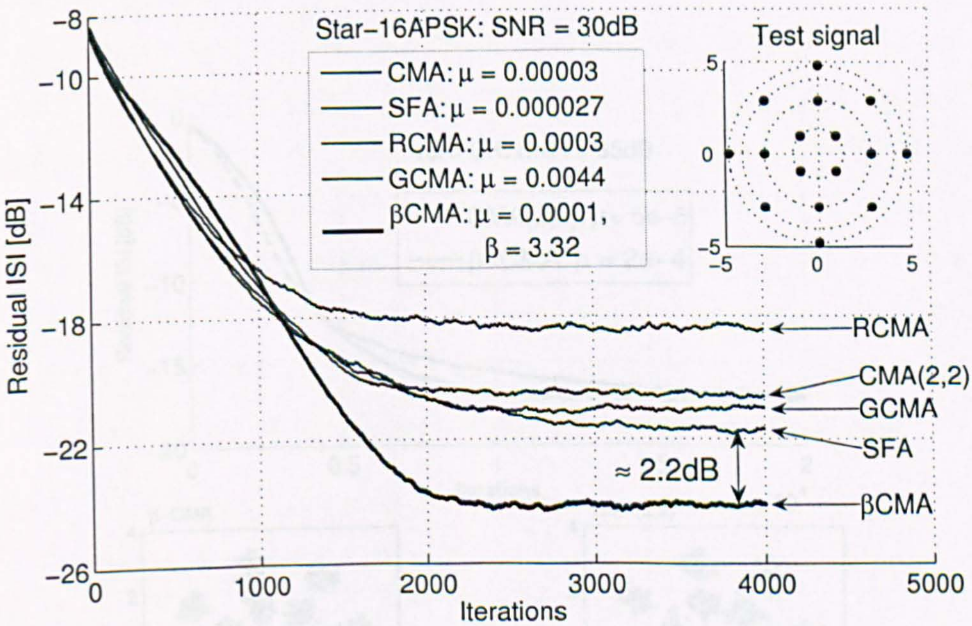


Figure 5.6: ISI traces for star-16-APSK (V.29) signal.

Table 5.1: Average energy recovered (%) @ SNR = 30dB

	RCMA	CMA(2,2)	SFA	GCMA	β-CMA
8-APSK	99.5	99.7	77.0	36.0	99.9
16-APSK	99.6	99.7	70.5	37.0	99.8
star-16-APSK	99.6	99.5	79.0	28.0	99.7

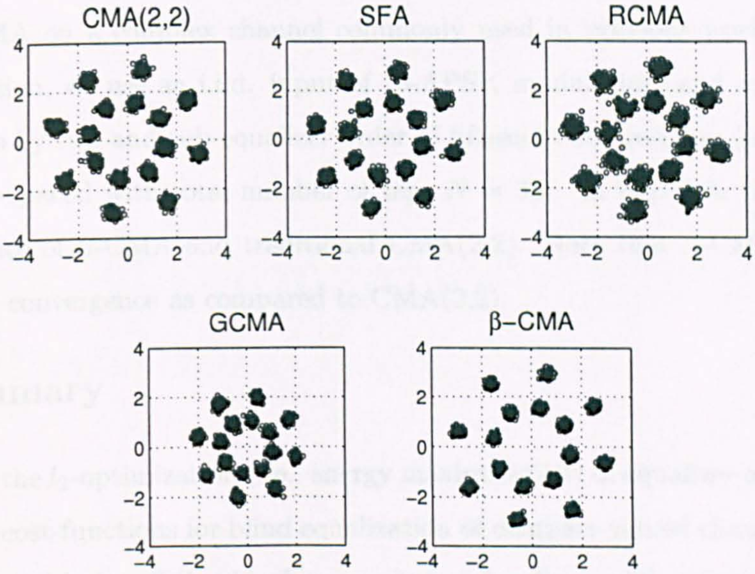


Figure 5.7: Scatter plots for 16-APSK signal each containing last 2000 converged symbols.

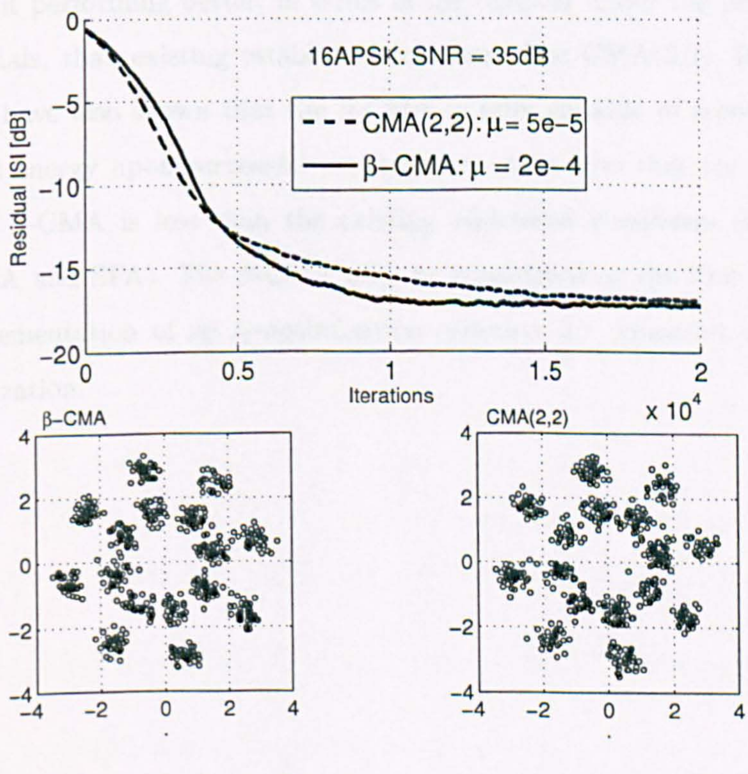


Figure 5.8: ISI traces for 16-APSK signal with fractionally-spaced equalization.

5.3.2 Experiment 2: ISI Performance with FSE

We test β-CMA on a complex channel commonly used in previous works [69, 38, 86]. In the simulation, we use an i.i.d. input of 16-APSK modulation, and assume an over-sampling ratio by two and sub-equalizer order of fifteen at the receiver (so the equalizer is fractionally-spaced with total number of taps $N = 30$). In Fig. 5.8, we compare the ISI performance of β-CMA and traditional CMA(2,2). Note that β-CMA is offering a slightly faster convergence as compared to CMA(2,2).

5.4 Summary

By exploiting the l_2-optimization (i.e., energy maximization) of equalizer output, we have proposed two cost-functions for blind equalization of complex-valued channels, which are respectively suitable for off-line block-processing and online gradient-based implementations. These costs differ in the way whether we constrain a block of equalizer outputs or only its instantaneous value from exceeding the largest modulus of data signal.

We have shown how to obtain an online adaptive algorithm (β-CMA) and have demonstrated it performing better, in terms of ISI removal under the presence of noise for APSK signals, than existing established solutions, like CMA(2,2), RCMA, GCMA and SFA. We have also shown that the β-CMA is fully capable of recovering the true value of signal energy upon successful convergence. Also note that the computational complexity of β-CMA is less than the existing addressed equalizers (like CMA(2,2), RCMA, GCMA and SFA). The β-CMA may be considered as the first ever successful adaptive implementation of an l_2-optimization criterion for (constant modulus) blind channel equalization.

Chapter 6

Energy Maximization: Adaptive Multimodulus Algorithm

6.1 Introduction

This Chapter is the sequel of Chapter 5, where we have proposed and discussed an energy maximization based constant modulus algorithm, β-CMA. Here, we propose and discuss an energy maximization based multimodulus algorithm, β-MMA. Clearly, the proposed algorithm β-MMA is required to achieve blind channel equalization and carrier-phase recovery jointly. The β-MMA is obtained from Meng *et al.* cost-function (Eq. (2.3.3)) with necessary modification to include a differentiable constraint (as we have discussed in Chapter 5, Section 5.2.3).

Unlike Chapter 5, where we have ignored the effect of convolutional noise in the evaluation of gain β, here we consider the effect of convolutional noise to determine β for β-MMA (Section 6.3.2). It is important because, in this Chapter, we evaluate performances of β-MMA and some existing MMA/CMA solutions for higher-order QAM signals, where the effect of convolutional noise cannot be ignored. We provide evidence of good performance exhibited by β-MMA in comparison to existing established methods, like MMA(1,2), MMA(2,2) and CMA(2,2) through computer simulations for higher-order QAM signalling on symbol- and fractionally-spaced channels (Section 6.4).

6.2 Meng-Tuqan-Ding l_2-Optimization Criterion

Recently, in the year 2009, Meng, Tuqan and Ding [90] proposed *an l_2-maximization based method for joint blind channel equalization and carrier-phase recovery without requiring* *tap anchoring*. Their cost-function is given as:

$$\max_{\boldsymbol{w}} \mathsf{E}[|y_n|^2], \text{ s.t. } \max(\{|y_{R,n}|\}) = \max(\{|y_{I,n}|\}) \leq \gamma \tag{6.2.1}$$

where $\max(\{|y_R|\})$ and $\max(\{|y_{I,n}|\})$ denote, respectively, the largest absolute values of the in-phase and quadrature components of equalized sequence $\{y_n\}$, and the parameter γ denotes the maximum quadrature component of the transmitted data a_n. They formulated the cost as an iterative block-processing *quadratic programming* problem for blind equalization of square-QAM and reported better results than those obtained from *linear programming* based solutions [38, 86]. Note that the constraints in (6.2.1) have been shown to be *convex* in $\boldsymbol{w}$ [38, 69, 86]. Also note that, due to using separate constraints for in-phase and quadrature components of equalized sequence, the reported iterative block-processing equalizer was jointly capable of recovering the carrier-phase.

6.3 Differentiable Cost-Function and Adaptive Algorithm

The cost (6.2.1) is not directly suitable for stochastic gradient-based adaptive implementation, since the constraints are applied on a block of equalized sequence. So, in accordance with the ideas discussed in Section 5.2.3, we modify the cost (6.2.1) and present a new *deterministic* cost-function, involving *instantaneous* constraints, for blind equalization and carrier-phase recovery, *viz*

$$\boxed{\begin{aligned} \max_{\boldsymbol{w}} |y_n|^2, \text{ subject to } \mathrm{fmax}(\gamma, |y_{R,n}|) &= \gamma \\ \text{and } \mathrm{fmax}(\gamma, |y_{I,n}|) &= \gamma \end{aligned}}, \tag{6.3.1}$$

where, for some $a, b \geq 0$, fmax is defined as

$$\mathrm{fmax}(a, b) \equiv \frac{|a+b| + |a-b|}{2} = \begin{cases} a, & \text{if } a \geq b \geq 0 \\ b, & \text{if } b \geq a \geq 0 \end{cases}. \tag{6.3.2}$$

The usefulness of fmax in (6.3.1) can be understood by referring to Fig. 6.1, where we depict resulting values of fmax on y_R-y_I plane. Clearly, if y_n falls inside the central square region, that is the region-C, centered at origin with perimeter 8γ, then both constraints in cost (6.3.1) are satisfied; we simply need to maximize $|y_n|^2$. However, when y_n lies outside region-C, then depending on the region where y_n is residing, region-B(D) or A, either one or both of the constraints is/are violated. In such a case, an ideal update $\boldsymbol{w}_{n+1}$ would ensure that the resulting *a posteriori* output $\boldsymbol{w}_{n+1}^H \boldsymbol{x}_n$ lies inside region-C and $\mathsf{E}[|y_n|^2]$ stays close to true signal-energy.

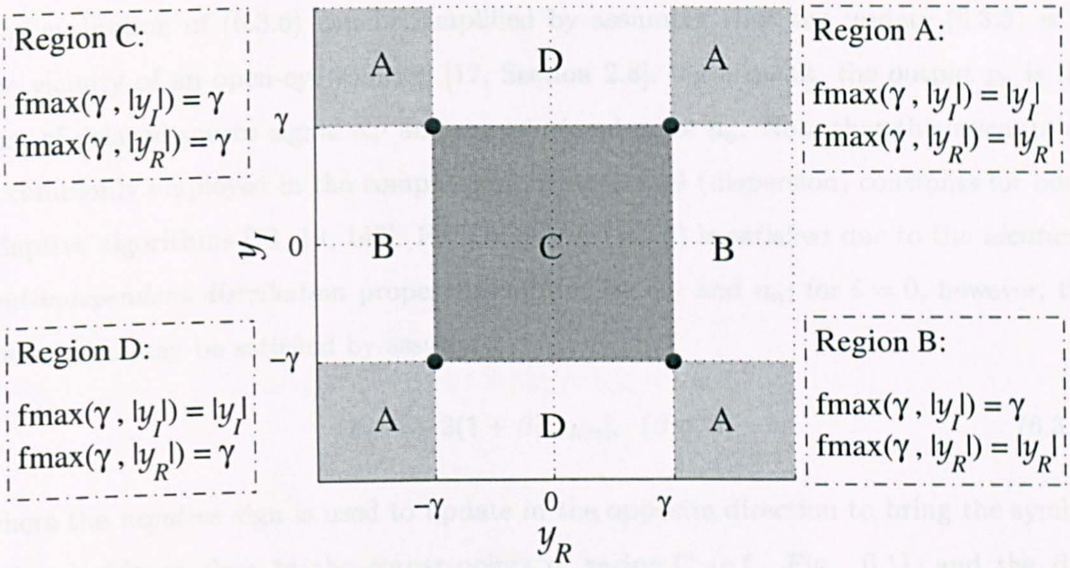


Figure 6.1: Values of $\mathrm{fmax}(\gamma, |y_{L,n}|)$ on y_R-y_I plane for an arbitrary M-ary square-QAM, where $\gamma = \sqrt{M} - 1$, and the four dots designate the (distortion-free) corner symbols of square-QAM.

6.3.1 Derivation of β-MMA

The fmax is differentiable. For some $a, b \in \mathbb{R}$, we can show

$$\frac{\partial}{\partial a}\mathrm{fmax}(|a|, |b|) = \frac{\mathtt{sgn}[a]}{2}(1 + \mathtt{sgn}[|a| - |b|]), \tag{6.3.3}$$

where $\mathtt{sgn}[\cdot]$ is the standard `signum` function. Next employing the Lagrangian multipliers, λ_R and λ_I, we obtain

$$\mathcal{J} = |y_n|^2 + \lambda_R(\mathrm{fmax}(\gamma, |y_{R,n}|) - \gamma) + \lambda_I(\mathrm{fmax}(\gamma, |y_{I,n}|) - \gamma). \tag{6.3.4}$$

The derivative of (6.3.4) with respect to y_n^* gives $\Phi(y_n) = -\frac{1}{4}(\lambda_R g_R + \jmath \lambda_I g_I + 4y_n)$, where $g_L = (1 + \mathtt{sgn}[|y_{L,n}| - \gamma])\mathtt{sgn}[y_{L,n}]$ (subscript L denotes either R or I). From the error-function $\Phi(y_n)$, we devise the following adaptive algorithm:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \frac{\mu}{4}(\lambda_R g_R + \jmath \lambda_I g_I + 4y_n)^* \boldsymbol{x}_n. \tag{6.3.5}$$

If $|y_{L,n}| < \gamma$, then $g_L = 0$; the constraint is effortlessly satisfied. On the other hand, the condition $|y_{L,n}| > \gamma$ yields $g_L = 2\,\mathtt{sgn}[y_{L,n}]$; here, we suggest to compute λ_L such that the Bussgang condition is satisfied. This consideration leads to

$$\underbrace{\mathsf{E}\left[(0.5\lambda_L \mathtt{sgn}[y_{L,n}] + y_{L,n})\, y_{L,n-i}\right]}_{|y_{L,n}|>\gamma} + \underbrace{\mathsf{E}\left[y_{L,n} y_{L,n-i}\right]}_{|y_{L,n}|<\gamma} = 0,\ \forall i. \tag{6.3.6}$$

The evaluation of (6.3.6) can be simplified by assuming that the update (6.3.5) is in the vicinity of an open-eye solution [17, Section 2.8]. As a result, the output y_n is the sum of delayed source signal $a_{n'}$ and convolutional noise u_n. Note that this assumption is commonly employed in the computation of statistical (dispersion) constants for blind adaptive algorithms [53, 19, 140]. For $i \neq 0$, Eq. (6.3.6) is satisfied due to the *identical-and-independent distribution* property exhibited by $a_{n'}$ and u_n; for $i = 0$, however, the constraints may be satisfied by assuming

$$\lambda_L = -2(1+\beta)|y_{L,n}|, \quad (\beta > 0) \tag{6.3.7}$$

where the *negative* sign is used to update in the opposite direction to bring the symbol either inside or close to the corner-points of region-C (c.f. Fig. 6.1), and the β is introduced to limit the growth of $\mathsf{E}[|y_n|^2]$. Due to the four-quadrant symmetry of QAM signal (i.e., $\mathsf{E}[a_R^2] = \mathsf{E}[a_I^2]$ and $\mathsf{E}[a_R a_I] = 0$), note that the expression (6.3.7) directs us to look for a *single* parameter β for both λ_R and λ_I.

6.3.2 Evaluation of Gain β

Denoting v_n as one of the components of u_n, we write $y_{L,n} = a_{L,n'} + v_n$. The v_n is considered to be zero-mean Gaussian with variance σ_v^2 and pdf $f_v = \frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(-\frac{v^2}{2\sigma_v^2}\right)$. Now combining (6.3.6)-(6.3.7), we get

$$\underbrace{(-\beta)\mathsf{E}\left[(a_L+v)^2\right]}_{|a_L+v|>\gamma} = -\underbrace{\mathsf{E}\left[(a_L+v)^2\right]}_{|a_L+v|<\gamma} \Rightarrow \boxed{\beta = \frac{\mathcal{E}_2}{\mathcal{E}_1}} \tag{6.3.8}$$

Further, $\mathcal{E}_1 = \mathsf{E}\left[\int_{-\infty}^{-\gamma-a_L} + \int_{\gamma-a_L}^{\infty} (a_L+v)^2 f_v \mathrm{d}v\right]$ and $\mathcal{E}_2 = \mathsf{E}\left[\int_{-\gamma-a_L}^{\gamma-a_L} (a_L+v)^2 f_v \mathrm{d}v\right]$ are evaluated as

$$\begin{aligned}
\mathcal{E}_2 = \mathsf{E}\Big[&(a_L^2+\sigma_v^2)\left(1 - Q\left(\frac{\gamma+a_L}{\sigma_v}\right) - Q\left(\frac{\gamma-a_L}{\sigma_v}\right)\right) \\
&-\sigma_v(\gamma-a_L)S\left(\frac{\gamma+a_L}{\sigma_v}\right) - \sigma_v(\gamma+a_L)S\left(\frac{\gamma-a_L}{\sigma_v}\right)\Big] \\
\mathcal{E}_1 = \mathsf{E}\Big[&(a_L^2+\sigma_v^2)\left(Q\left(\frac{\gamma+a_L}{\sigma_v}\right) + Q\left(\frac{\gamma-a_L}{\sigma_v}\right)\right) \\
&+\sigma_v(\gamma-a_L)S\left(\frac{\gamma+a_L}{\sigma_v}\right) + \sigma_v(\gamma+a_L)S\left(\frac{\gamma-a_L}{\sigma_v}\right)\Big]
\end{aligned}$$

where $Q(z) \equiv \int_z^{\infty} S(t)\mathrm{d}t$ and $S(t) \equiv (1/\sqrt{2\pi})\exp\left(-0.5t^2\right)$.[1]

Note that the β has an asymptotic value for small noise. Considering a square-QAM, $a_L \in \mathcal{A}_L = \{\pm 1, \pm 3, \cdots, \pm\gamma\}$, and assuming $y_L > 0$ and small v, we evaluate (6.3.8) to get

$$\underbrace{(1+v)^2 + (3+v)^2 + \cdots + (\gamma - 2 + v)^2}_{(\frac{1}{2}|\mathcal{A}_L|-1)\text{ terms}} + \underbrace{\frac{1}{2}(\gamma+v)^2}_{\text{if } v<0} + \underbrace{\frac{1}{2}(-\beta)(\gamma+v)^2}_{\text{if } v>0} = 0, \tag{6.3.9}$$

where $|\mathcal{A}_L| = \gamma + 1$. Assuming a diminishing noise, we get

$$\boxed{\lim_{v\to 0} \beta \to \beta_{\text{lim}} \equiv \frac{\gamma^2+2}{3\gamma} = \frac{M - 2\sqrt{M}+3}{3\sqrt{M}-3}}. \tag{6.3.10}$$

In Fig. 6.2, we demonstrate that $\beta \to \beta_{\text{lim}}$ when $\sigma_v \to 0$.

Finally, we summarize our proposed algorithm as follows:

$$\boxed{\begin{array}{l} \boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\,(\text{f}_R\, y_{R,n} + \jmath\, \text{f}_I\, y_{I,n})^* \boldsymbol{x}_n \\ \text{f}_L = \begin{cases} 1, & \text{if } |y_{L,n}| < \gamma \\ 0, & \text{if } |y_{L,n}| = \gamma \\ -\beta, & \text{if } |y_{L,n}| > \gamma \end{cases} \end{array}}. \tag{6.3.11}$$

Note that the polarity of variable f_L determines the direction of adaptation such that the dispersion in y_n is minimized away from four corner points $\{\pm\gamma \pm \gamma\jmath\}$ and this property has a major role in carrier-phase recovery. In [140], a term *multimodulus* was coined for the algorithm which can jointly solve blind equalization and carrier-phase recovery; we use this terminology to denote (6.3.11) as β-multimodulus algorithm (β-MMA).

Note that another adaptive realization is also possible if constraints are imposed on a posteriori output $s_n \equiv \boldsymbol{w}_{n+1}^H \boldsymbol{x}_n$. Taking conjugate-transpose of (6.3.5) and post-multiplying it with $\boldsymbol{x}_n$, we get

$$s_n = y_n + (\mu/4)\big(\lambda_R\, g_R + \jmath\lambda_I\, g_I + 4y_n\big)\|\boldsymbol{x}_n\|_2^2. \tag{6.3.12}$$

If $(|y_{L,n}| > \gamma)$, then the requirement $(|s_{L,n}| \leq \gamma)$ yields

$$\lambda_L \leq 2\alpha\gamma - 2(1+\alpha)\,|y_{L,n}|, \tag{6.3.13}$$

[1]For the given square-QAM signal and convolutional noise level, the required β can be pre-computed and stored in receiver memory. In equilibrium, $2\sigma_v^2 \approx \sigma_\zeta^2 \|\boldsymbol{w}_\infty\|_2^2$ is closely true, where σ_ζ^2 is the variance of additive noise and $\|\boldsymbol{w}_\infty\|_2^2$ is the steady-state value of $\|\boldsymbol{w}_n\|_2^2$. More accurate estimation of σ_v^2 is possible by carrying out *excess* mean-square analysis [79] to incorporate the contributions of channel eigen-values, equalizer length and step-size.

where $\alpha = 1/\left(\mu\|x_n\|_2^2\right)$. Note that λ_L is favorably negative for $|y_{L,n}| > \gamma$, and is similar to (6.3.7).

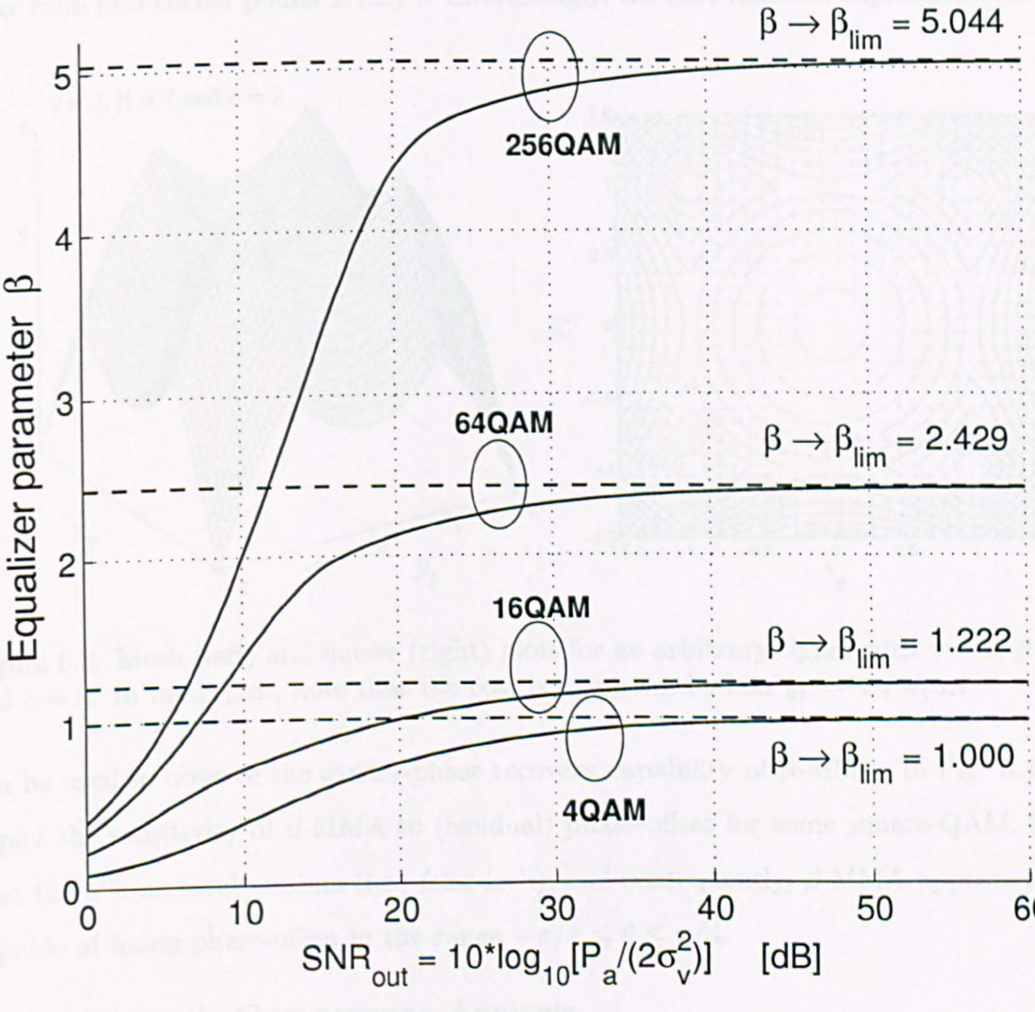


Figure 6.2: Values of β versus $\mathrm{SNR}_{\mathrm{out}}$ for some square-QAM.

6.3.3 Cost-Function Interpretation and Phase Recovery Capability

If the error-function in (6.3.11) is integrated back with respect to y_n, then the following cost-function is obtained:

$$J_{\beta\text{-MMA}} = \max_{w} \left\{ \underbrace{\mathsf{E}\left[y_{R,n}^2\right]}_{|y_{R,n}|<\gamma} + \underbrace{(-\beta)\mathsf{E}\left[y_{R,n}^2\right] + c_R}_{|y_{R,n}|>\gamma} + \underbrace{\mathsf{E}\left[y_{I,n}^2\right]}_{|y_{I,n}|<\gamma} + \underbrace{(-\beta)\mathsf{E}\left[y_{I,n}^2\right] + c_I}_{|y_{I,n}|>\gamma} \right\} \tag{6.3.14}$$

where c_R and c_I are constants of integration. So, depending on the value of $|y_{L,n}|$ whether it is less or greater than γ, we need to respectively maximize or minimize $\mathsf{E}\left[y_{L,n}^2\right]$. In

Fig. 6.3, we depict the mesh and quiver plot of the cost (6.3.14) to demonstrate how the cost maximizes the energy of the equalized sequence while minimizing the dispersion away from four corner points $\pm\gamma \pm \jmath\gamma$. Interestingly, the cost-function expression (6.3.14)

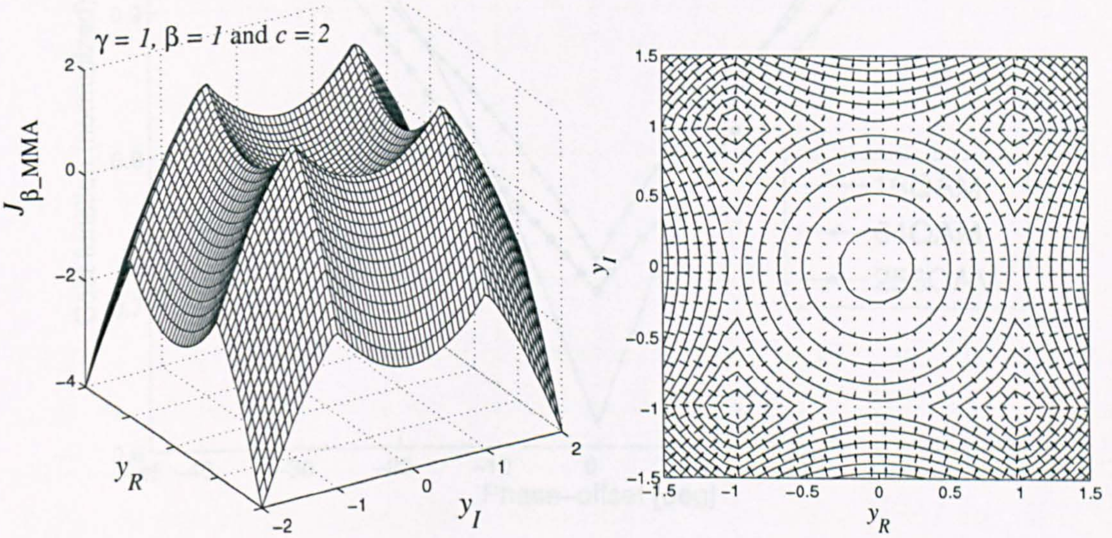


Figure 6.3: Mesh (left) and quiver (right) plots for an arbitrary signal with $\gamma = 1$, $\beta = 1$ and $c = 2$. In mesh plot, note that the cost is maximized when $y_L = \pm\gamma = \pm 1$.

can be used to observe the carrier-phase recovery capability of β-MMA. In Fig. 6.4, we depict the sensitivity of β-MMA to (residual) phase-offset for some square-QAM. Note that there is no local minima (i.e., false-lock), and consequently, β-MMA appears to be capable of fixing phase-offset in the range $-\pi/4 \le \theta \le \pi/4$.

6.3.4 Dynamic Convergence Analysis

First of all, note that the proposed β-MMA has a remarkable similarity with the algorithm MMA(2, 1) we discussed in Chapter 4. Notably by substituting $\beta = 1$ in weight update expression of β-MMA, we obtain exactly MMA(2, 1). Based on this observation, we find that ordinary difference equations (ODE) of β-MMA can easily be obtained from that of MMA(2, 1) by simply introducing the parameter β in the definition of the auxiliary variable H_L^{ij}, it gives

$$
\begin{aligned}
H_L^{ij} &= G_L^{ij}(-\gamma,\gamma) \;+\; (-\beta)^j \left[G_L^{ij}(\gamma,\infty) + G_L^{ij}(-\infty,-\gamma)\right] \\
\text{where } \; G_L^{ij}(x_1,x_2) &= \int_{x_1}^{x_2} x^{i+j} f_L(x)\mathrm{d}x.
\end{aligned}
\tag{6.3.15}
$$

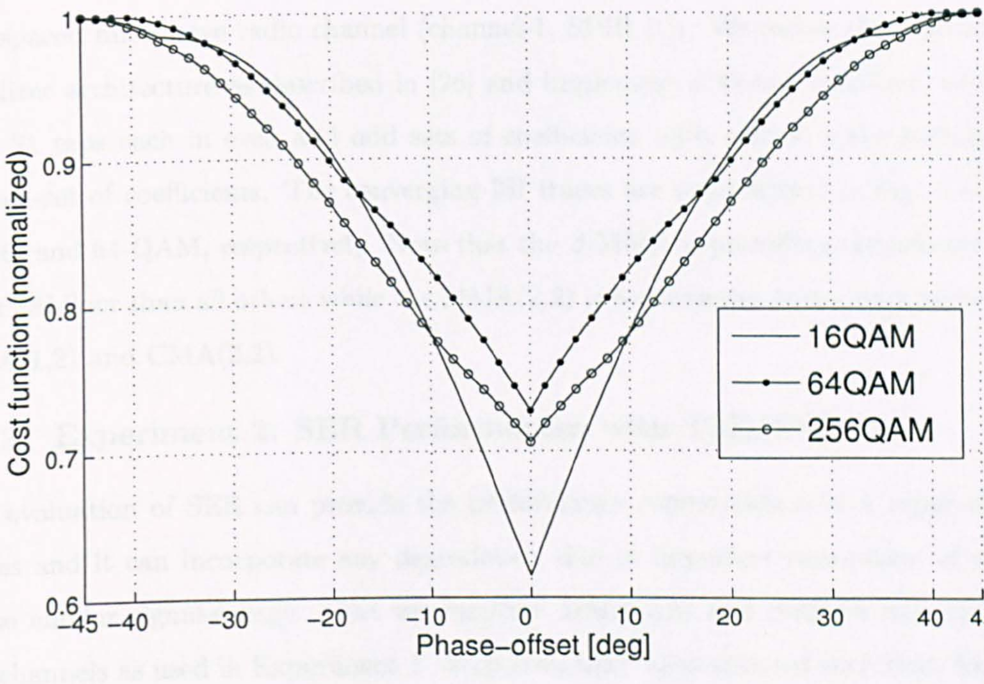


Figure 6.4: Sensitivity of β-MMA to (residual) phase-offset for some square-QAM.

6.4 Computer Experiments

We simulate adaptive equalizers implementing MMA(1,2), CMA(2,2), MMA(2,2), and the proposed one β-MMA, while considering square-QAM transmission over complex-valued symbol- and fractionally spaced (normalized) channels, evaluating (transient) ISI [121, Eq.(50)] and (steady-state) symbol-error rate (SER) performances.

6.4.1 Experiment 1: ISI Performances with TSE/FSE

Firstly, we consider symbol-spaced equalization (TSE) of a telephonic channel [105]. A seven-tap equalizer is used with central spike initialization. Each of the traces has been obtained by taking average of 300 Monte-Carlo realizations with independent generation of noise and data samples. Also, the step-sizes have been selected such that all algorithms reached steady-state requiring almost equal number of iterations. We have marked the point of convergence by a dashed vertical line.

The converging ISI traces are summarized in Fig. 6.5(a)-(b) for 16- and 64-QAM, respectively. Note that the β-MMA is providing much lower ISI floor than all others while the traditional MMA(2,2) is performing better than MMA(1,2) and CMA(2,2).

Secondly, we consider fractionally-spaced equalization (FSE) of a long (300-coefficients) $T/2$-spaced microwave radio channel (channel-1, SPIB [1]). We follow the multichannel equalizer architecture as described in [26] and implement a 42-tap equalizer, where we have 21 taps each in even and odd sets of coefficients with central spike initialization in even-set of coefficients. The converging ISI traces are summarized in Fig. 6.6(a)-(b) for 16- and 64-QAM, respectively. Note that the β-MMA is providing remarkably much lower ISI floor than all others while the MMA(2,2) is consistently performing better than MMA(1,2) and CMA(2,2).

6.4.2 Experiment 2: SER Performances with TSE/FSE

The evaluation of SER can provide the performance comparison over a range of SNR values and it can incorporate any degradation due to imperfect restoration of carrier-phase and/or signal-energy. Here we simulate MMA(2,2) and β-MMA over the same two channels as used in Experiment 1. Step-sizes have been selected such that, for TSE, both MMA(2,2) and β-MMA acquire stable convergence around 3000th, 6000th and 20,000th iteration for 16-, 64- and 256-QAM, respectively, and for FSE, both MMA(2,2) and β-MMA acquired stable convergence around 1500th, 3000th and 10,000th iteration for 16-, 64- and 256-QAM, respectively. At lower SNR_{in}, the β exhibits smaller value which slows down the convergence and may lead to an unfairly better solution. To avoid it, we increased μ such that the product of β and μ is kept constant for all SNR_{in}. In Fig. 6.7(a), we depict SER performances for 16/64/256-QAM over the telephonic channel. Observe that at lower SNR values, both MMA(2,2) and β-MMA performed almost identical; but, for higher SNR values, β-MMA outperformed MMA(2,2) for all QAM sizes. In Fig. 6.7(b), we depict SER results over the microwave radio channel. Again, we observe that β-MMA is yielding much lower SER than MMA(2,2).

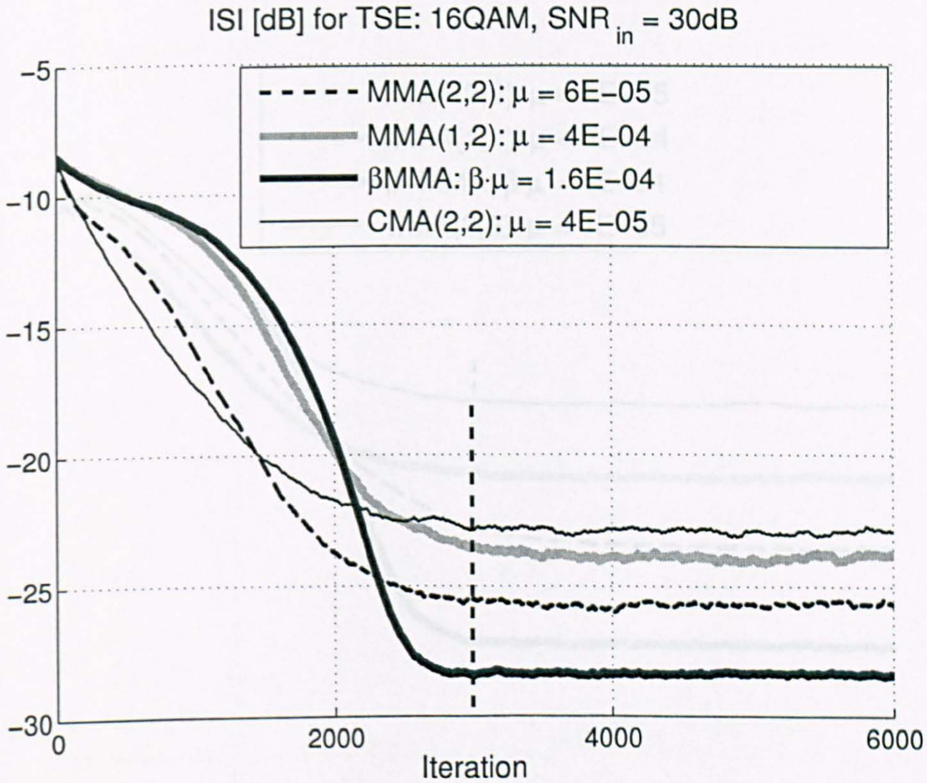
6.4.3 Experiment 3: Validating MSE/ISI Convergence Analysis

In this experiment, we compare the analytical ISI/MSE performance of β-MMA with those obtained from Monte-Carlo simulations for 16/64-QAM. We use the same experimental setup we used in subsection 6.4.1. Results are summarized in Fig. 6.8, note that the simulated and analytical traces are in full conformation with each other for both MSE and ISI performance measures.

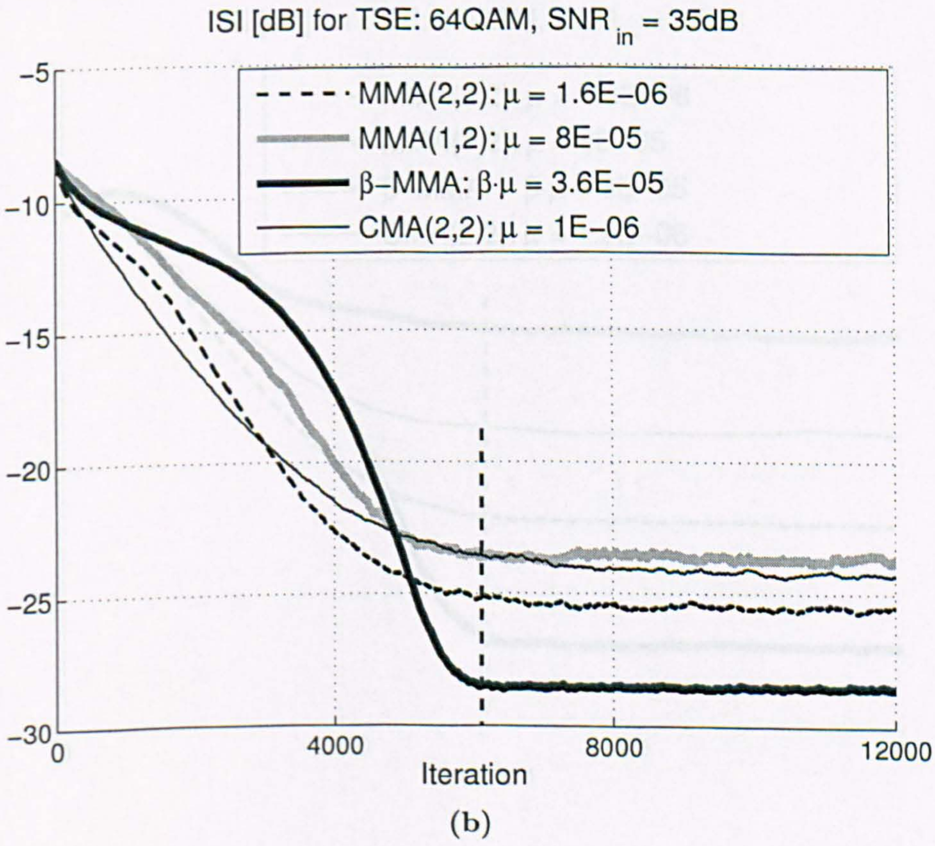
6.5 Summary

We have proposed a new MMA algorithm by exploiting the l_2-optimization (i.e., energy maximization) cost-function for joint blind equalization, carrier-phase recovery and energy restoration of square-QAM signals. The quadrature components of the equalized sequence are constrained not to exceed the largest real part of the transmitted signal. We optimized the cost to yield an adaptive multimodulus algorithm, which we termed as β-MMA. The parameter β is evaluated by considering the presence of convolutional noise at equalizer output. We have experimentally showed that the resulting new algorithm (β-MMA) can give better solution in terms of removing ISI and lower SER values under the presence of noise than existing established adaptive algorithms like MMA(1,2), CMA(2,2) and MMA(2,2).

We have also shown that the β-MMA is fully capable of recovering the true value of signal energy upon successful convergence. Also note that the computational complexity of β-MMA is less than the existing addressed equalizers (like MMA(1,2), CMA(2,2) and MMA(2,2)). The β-MMA may be considered as the first ever successful adaptive implementation of an l_2-optimization criterion for joint blind channel equalization and carrier-phase recovery.



(a)



(b)

Figure 6.5: Plots of ISI convergence on symbol-spaced channel with 16/64-QAM.

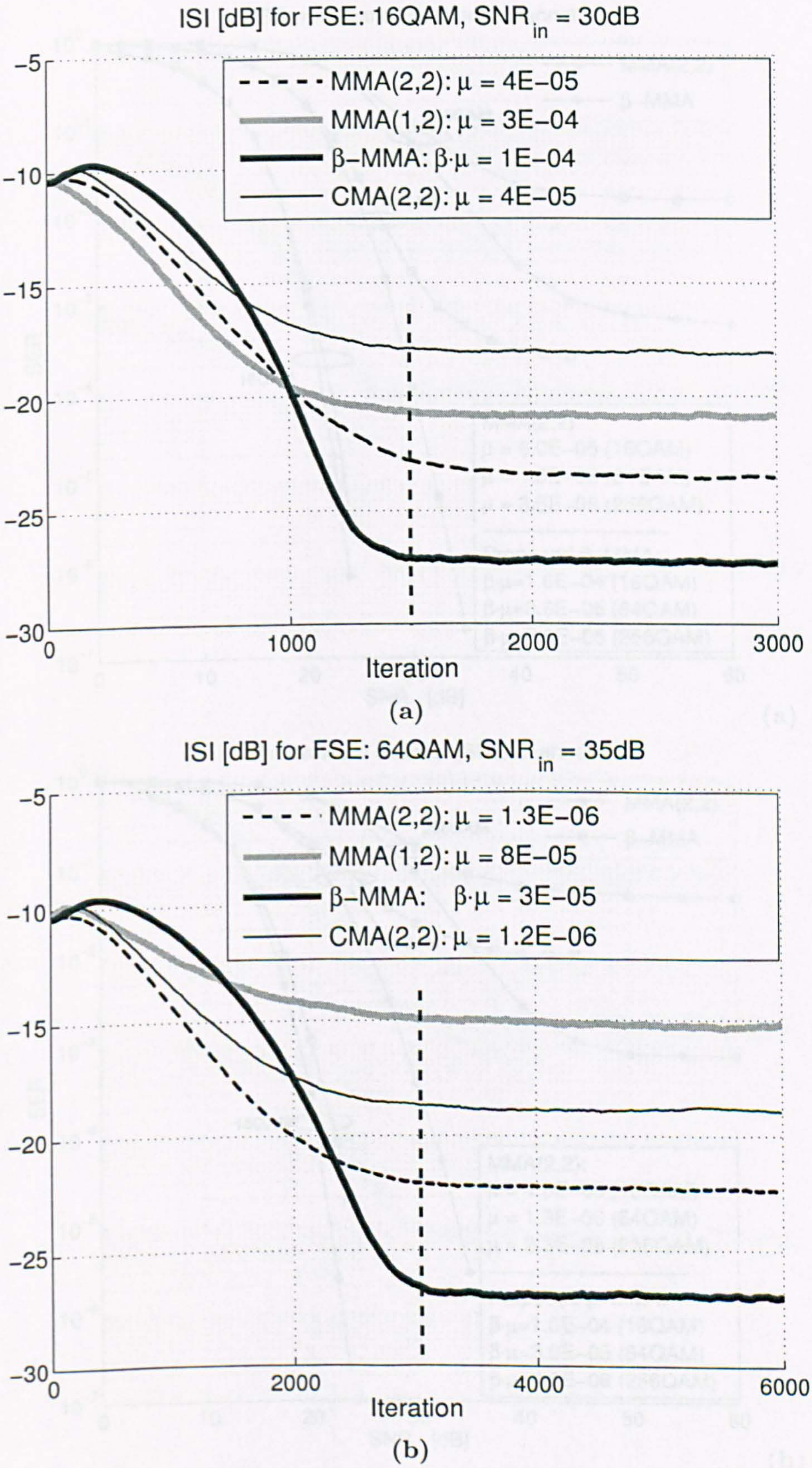
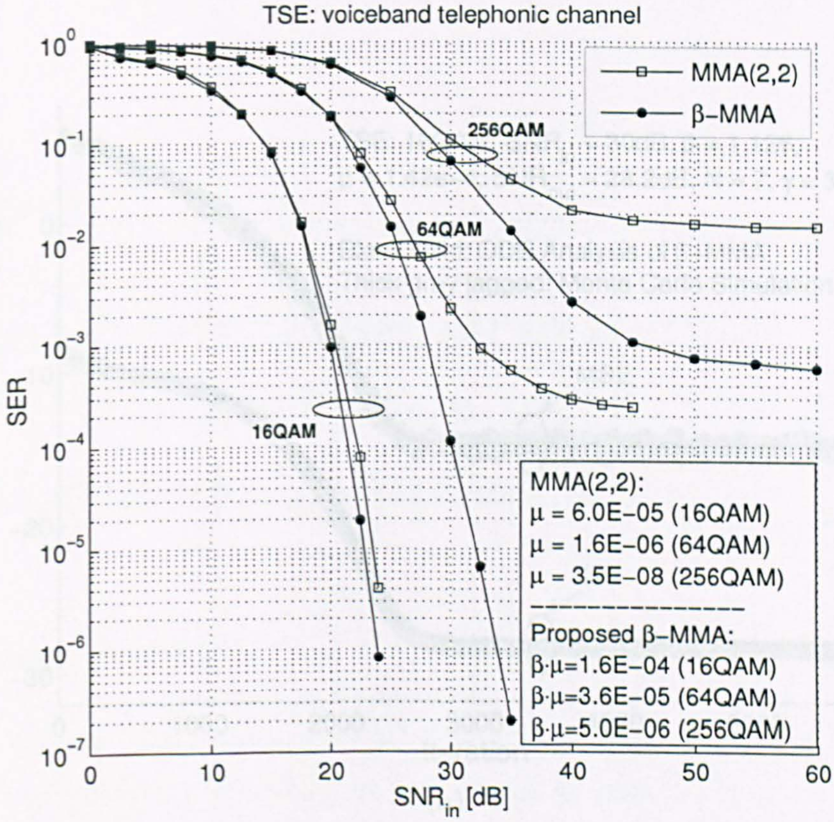
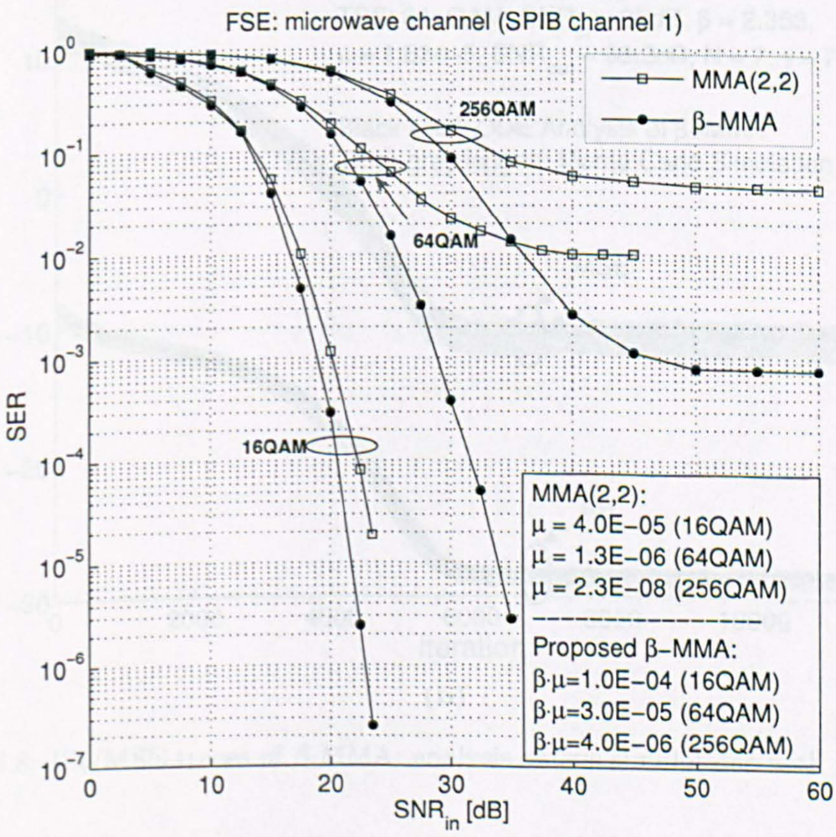


Figure 6.6: Plots of ISI convergence on fractionally-spaced channel with 16/64-QAM.



(a)



(b)

Figure 6.7: Plots of SER versus $\mathrm{SNR_{in}}$ for MMA(2,2) and β-MMA with 16/64/256-QAM.

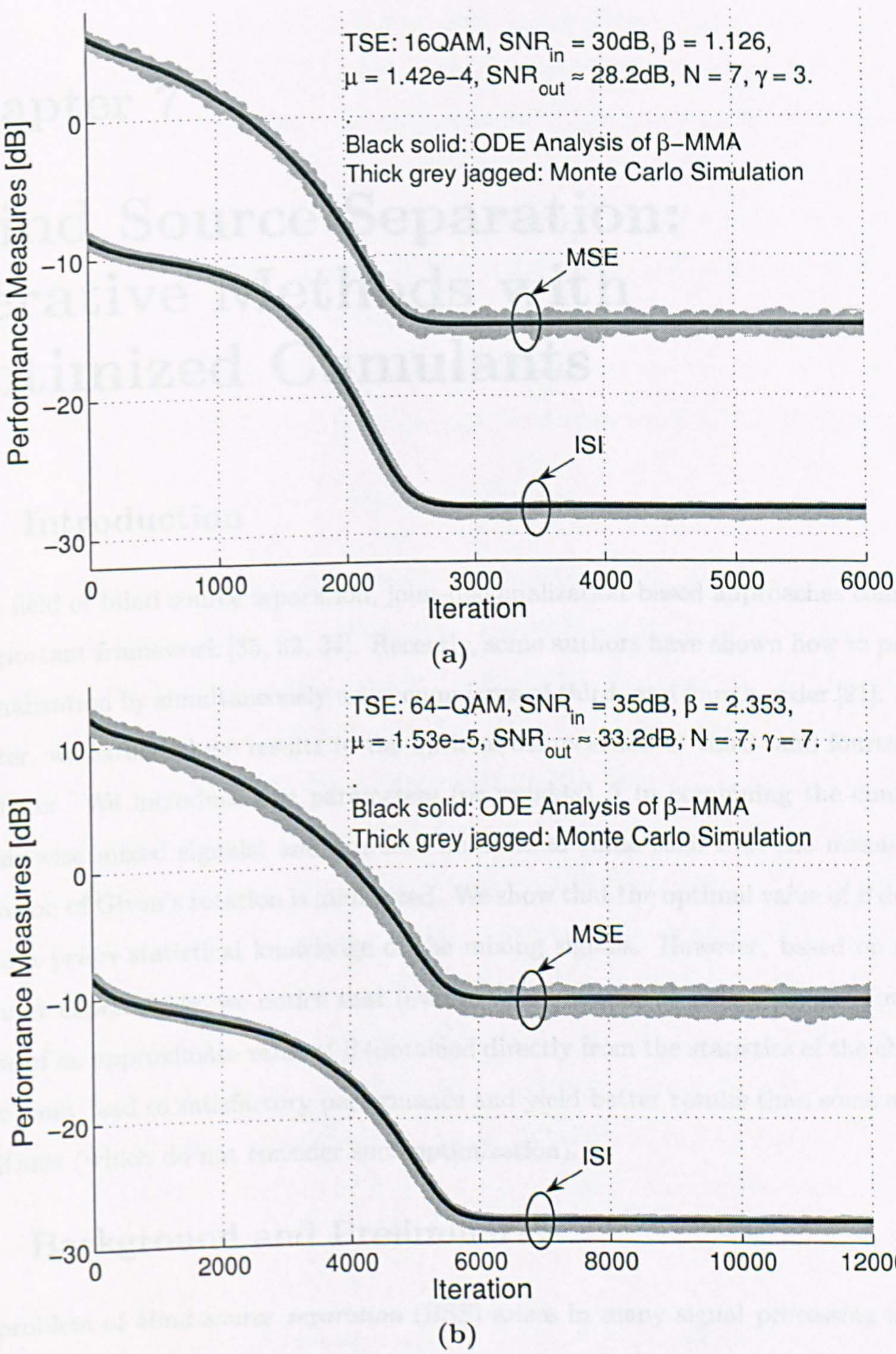


Figure 6.8: ISI/MSE traces of β-MMA: analysis versus simulations with 16/64-QAM.

Chapter 7

Blind Source Separation: Iterative Methods with Optimized Cumulants

7.1 Introduction

In the field of blind source separation, joint-diagonalization based approaches constitute an important framework [35, 32, 34]. Recently, some authors have shown how to perform diagonalization by simultaneously using cumulants of third- and fourth-order [21]. In this Chapter, we extend these results to the optimal composition of third- and fourth-order cumulants. We introduce free parameters (or weights) β in combining the cumulants (of pair-wise mixed signals) and evaluate its optimal value such that the mean-square estimation of Given's rotation is minimized. We show that the optimal value of β depends on the *a priori* statistical knowledge of the mixing signals. However, based on several computer experiments, we notice that (even) in the absence of such a priori knowledge, the use of an approximate value of β (obtained directly from the statistics of the observed source) may lead to satisfactory performance and yield better results than some existing algorithms (which do not consider such optimization).

7.2 Background and Preliminaries

The problem of *blind source separation* (BSS) arises in many signal processing applications like communications, array processing, speech analysis and speech recognition. In all these instances, the underlying assumption is that several linear mixtures of unknown, random, zero-mean, and statistically independent signals, called *sources*, are observed;

the problem consists of recovering the original sources from their mixtures without a priori information of coefficients of the mixtures and knowledge of the sources. The principle involved in the solution to this problem is nowadays called *independent component analysis* (ICA), which can be viewed as an extension of the widely known *principal component analysis* (PCA). The independence between the recovered sources is measured by their *mutual information* (MI). The MI measures the information that one variable contains about another one, i.e., the reduction of uncertainty of a magnitude when another one is known. The MI is zero if and only if the sources are independent. A natural criterion to measure the mutual independence between M variables (say $\mathbf{y} = \{y_i\}_{i=1}^{M}$) is the divergence between the joint probability density and the product of the marginal ones. If we follow the Kullback-Leibler divergence, it ends up with the MI [32]:

$$I[\mathbf{y}] = \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{i=1}^{M} p_i(y_i)} d\mathbf{y} \tag{7.2.1}$$

where $p(\mathbf{y})$ and $p_i(y_i)$ are the multivariate and marginal PDF of $\mathbf{y}$ and y_i, respectively. Consider an M-input M-output memoryless channel described by $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n)$, where $n \in \mathbb{Z}$ is the discrete time, $\mathbf{x}(n)$ is an $M \times 1$ vector of the observed signals, $\mathbf{s}(n)$ is an $M \times 1$ vector of the (original) sources, and $\mathbf{A} \in \mathbb{R}^{M \times M}$ is an unknown (invertible) mixing matrix. Our goal is to determine a separation matrix $\mathbf{B} \in \mathbb{R}^{M \times M}$ such that $\mathbf{y}(n) = \mathbf{B}\mathbf{x}(n) = \mathbf{B}\mathbf{A}\mathbf{s}(n) = \mathbf{C}\mathbf{s}(n)$ recovers the source signal up to a permutation and scaling, where $\mathbf{C}$ ia a global matrix representing a mixing-nonmixing structure. Source separation is typically carried out in two-step. First, whitening or standardization projects the observed vector $\mathbf{x}(n)$ on the signal subspace and yields a set of second-order decorrelated, normalized signals $\mathbf{z}(n) = \mathbf{W}\mathbf{x}(n)$ such that $\mathsf{E}[\mathbf{z}\mathbf{z}^T] = \mathbf{I}_M$. As a result, the source and whitened vectors must be related through a unitary transformation $\mathbf{z}(n) = \mathbf{Q}\mathbf{s}(n)$. The separation problem thus reduces to the computation of unitary matrix $\mathbf{Q}$, which is accomplished in a second step. The ICA approach to BSS consists of computing $\mathbf{Q}$ such that the entries of the separator output $\mathbf{y}(n) = \mathbf{C}\mathbf{s}(n) = \mathbf{Q}^T\mathbf{W}\mathbf{x}(n) = \mathbf{Q}^T\mathbf{z}(n) = \tilde{\mathbf{Q}}\mathbf{z}(n)$ are as independent as possible.[1]

[1]The philosophy behind this two-step strategy is described by Cardoso [27]; he emphasized that components that are as independent as possible according to some measure of independence are not necessarily uncorrelated because exact independence cannot be achieved in most practical applications. Thus, if de-correlation is desired, it must be enforced explicitly. Interestingly, this approach leads to a simple implementation; the whitening matrix $\mathbf{W}$ can be obtained straightforwardly by computing the matrix square root of the inverse covariance matrix of $\mathbf{x}$, and the orthonormal matrix $\mathbf{Q}^T$ can be obtained

Comon [32] studied the separability condition for this problem, and pointed out that for statistically independent non-Gaussian sources, the separation can be achieved by restoring the independence. He proposed using MI as a tool to measure the independence of the output signals, and to use an Edgeworth expansion to approximate the probability density function in the MI criterion. The Edgeworth expansion of the MI of a standardized (i.e. after whitening) real variable, up to an additive constant I_0 and as a function of standardized *cumulants*[2], is given as follows [32]:

$$-I[\mathbf{y}] \approx I_0 + \sum_i \left(4\kappa_{iii}^2(\mathbf{y}) + \kappa_{iiii}^2(\mathbf{y}) + 7\kappa_{iii}^4(\mathbf{y}) - 6\kappa_{iii}^2(\mathbf{y})\kappa_{iiii}(\mathbf{y})\right) \tag{7.2.2}$$

where $\kappa_{iii}(\mathbf{y})$ and $\kappa_{iiii}(\mathbf{y})$ are the third-order and fourth-order marginal cumulants of each entry of $\mathbf{y}$, i.e., $\kappa_{iii}(\mathbf{y}) = \mathsf{E}[y_i^3]$ and $\kappa_{iiii}(\mathbf{y}) = \mathsf{E}[y_i^4] - 3\mathsf{E}^2[y_i^2]$. Comon [32] have shown that the cumulants are *contrast*. By definition, a contrast $\mathcal{J}(\mathbf{y})$ is a mapping from the set of densities $\{p_i(y_i), \mathbf{y} \in \mathbb{E}^M\}$ to $\mathbb{R}$, where M is the number of sources, such that if $\mathbf{y}$ has independent components, then $\mathcal{J}(\mathbf{y}) \geq \mathcal{J}(\mathbf{A}\mathbf{y})$, $\forall \mathbf{A}$ nonsingular, with equality if and only if $\mathbf{A}$ is nonmixing; also, $\mathcal{J}(\mathbf{y})$ is invariant to permutation or scaling of the components of $\mathbf{y}$. Thus the maximization of specific cumulants would result into a successful blind separation for particular type of sources[3]. For example, if the sources are asymmetrical (skewed) then the maximization of third-order cumulant $\kappa_{iii}^2(\mathbf{y})$ would be enough to ensure successful separation; similarly, for symmetric sources, the maximization of fourth-order $\kappa_{iiii}^2(\mathbf{y})$ would be sufficient. This is

$$\mathcal{J}(\mathbf{y}) = \begin{cases} \sum_i \kappa_{iii}^2(\mathbf{y}), & \text{for asymmetrical sources} \\ \sum_i \kappa_{iiii}^2(\mathbf{y}), & \text{for symmetrical sources} \end{cases} \tag{7.2.3}$$

There exist number of ways to find the unmixing matrix $\widetilde{\mathbf{Q}}$ such that the contrast (7.2.3) is maximized. The ICA algorithms based on the maximization of third- and fourth-order cumulants are reported in [32] and [33] by Comon, respectively. The contrast (7.2.3) is

by the Jacobi technique subject to the maximization of some suitable criterion for independence. In short, whitening (second-order independence) solves the BSS problem up to an orthogonal transformation.

[2] For zero-mean random variables X_i, X_j, X_k, X_l, third- and the fourth-order cumulants are defined respectively as: $\kappa_{ijk}(X) \stackrel{\text{def}}{=} \mathsf{Cum}(X_i, X_j, X_k) = \mathsf{E}[X_iX_jX_k]$ and $\kappa_{ijkl}(X) \stackrel{\text{def}}{=} \mathsf{Cum}(X_i, X_j, X_k, X_l) = \mathsf{E}[X_iX_jX_kX_l] - \mathsf{E}[X_iX_j]\mathsf{E}[X_kX_l] - \mathsf{E}[X_iX_k]\mathsf{E}[X_jX_l] - \mathsf{E}[X_iX_l]\mathsf{E}[X_jX_k]$.

[3] If a certain criterion for blind source separation qualifies to be a contrast then it is not necessary that it approximates the MI too. Similarly, if a certain statistical quantity approximates MI then it does not necessarily qualifies for a contrast. Notice that the expression (7.2.2) which is a pretty good approximation to MI is not a proven contrast (as a whole).

discriminating over the set of random vectors $\mathbf{y}$ if they have at most one null third-order (resp. fourth-order) marginal cumulant for skewed (resp. symmetrical) sources [32, 33]. Just recently, Blaschke & Wiskott showed that the *joint* use of third- and fourth-order cumulants is an *admissible* choice for a contrast [21]:

$$\max_{\tilde{\mathbf{Q}}} \mathcal{J}(\mathbf{y}): \quad \mathcal{J}(\mathbf{y}) = \sum_i \left(4\kappa_{iii}^2(\mathbf{y}) + \kappa_{iiii}^2(\mathbf{y})\right) \tag{7.2.4}$$

The Blaschke-Wiskott's ICA algorithm is known as CuBICA. The contrast (7.2.4) provided a good mean to handling the symmetric and asymmetric sources simultaneously. In the same year when CuBICA appeared, Comon obtained a more generalized result; he showed that CuBICA is a special case of the following contrast [34, Theorem 13]:

$$\mathcal{J}(\mathbf{y}) = \sum_i w_i \mathcal{J}_i(\mathbf{y}) \tag{7.2.5}$$

where, $\forall i$, w_i is a strictly positive number and $\mathcal{J}_i(\mathbf{y})$ is a contrast; i.e., *the weighted sum of contrasts is also a contrast.* The literature witnesses several efforts where researchers proposed a number of contrasts, obtained by the weighted sum of contrasts. For example, in [94], Moreau and Thirion-Moreau suggested a weighted contrast using the fourth-order statistics, as given by:

$$\mathcal{J}(\mathbf{y}) = \begin{cases} \varepsilon \sum_{i=1}^{M} w_i \kappa_{iiii}(\mathbf{y}) \\ \varepsilon \sum_{i=1}^{M} w_i \mathsf{E}[y_i^4] \end{cases} \tag{7.2.6}$$

where ε indicates the sign of kurtosis and w_i are free parameters. In (7.2.6), it was assumed that all sources have same sign of kurtosis. For the case $M = 2$, they showed how to obtain the optimal values of weight parameters. Moreover, some higher-order (including the third-order) generalization and complex-valued extension of (7.2.6) were also discussed in [94, 93]. In [124], Stoll and Moreau proposed a yet another generalized form of weighted fourth-order contrast function, as given by

$$\mathcal{J}(\mathbf{y}) = \sum_{i=1}^{M} \kappa_{iiii}^2(\mathbf{y}) + 2\Bigg(w_1 \sum_{\substack{i,j=1 \\ j\neq i}} \kappa_{iiij}^2(\mathbf{y}) + w_2 \sum_{\substack{i,j=1 \\ j>i}} \kappa_{iijj}^2(\mathbf{y}) + w_3 \sum_{\substack{i,j,k=1 \\ k\neq j\neq i \\ k>j}} \kappa_{iijk}^2(\mathbf{y})\Bigg) \tag{7.2.7}$$

The contrast (7.2.7) is an attempt to combine the autocumulants and cross-cumulants. In general, autocumulants and cross-cumulants are assumed to be maximized and minimized, respectively. In [124], it was experimentally shown that it is possible to get better performance (in some cases) by selecting appropriate *binary* values for weight parameters. However, it was not suggested analytically that how to obtain the optimal values of these parameters for given source statistics. Moreover, *closed-form* near-optimal/optimal fourth-order estimators using *weighted-centroid* appeared recently in [96] and [146].

It is interesting to notice that all of the aforementioned weighted contrasts were based on fourth-order statistics; fourth-order statistics are more suitable for symmetric sources than asymmetric ones. Notice that asymmetric sources arise in many practical scenarios, such as in sonar signal processing [101] or source separation of urban images [151] (see also [67] and [68]). In some cases, digitized speech signals have non-zero skewness; and separation of such signals get benefit from third-order statistics [30]. Also, in biomedical applications, skewness is sometime more important to just non-Gaussianity for certain categories of signals, say, certain artifacts (like eye-blinking) and, some known components in electrocardiograms and electroencephalograms are not symmetric.

Due to the importance of asymmetry, in this Chapter, we present a weighted form of third- and the fourth-order contrast (7.2.4) which is capable of handling the symmetric and asymmetric sources jointly in an *optimal* manner. The proposed weighted contrast and the derivation of optimal weight parameter is described in Section 7.3, the resulting ICA algorithm is given in Section 7.4, and the performance comparisons are provided in Section 7.6. We conclude briefly in Section 7.7. All simulations were done with MATLAB; analytical calculations in Section 7.4 were supported by Symbolic toolbox of MATLAB.

7.3 Proposed Weighted Contrast And Optimal Weight Parameter

At the end of last section, we emphasized over the importance of weighted contrasts in source separation and the presence of asymmetrical sources in various engineering problems. In this section, for the blind separation of mixture of symmetric and asymmetric sources, we propose a weighted contrast which jointly diagonalizes the third- and the fourth-order cumulants as a generalization of the contrast (7.2.4) as given by

$$\mathcal{J}(\mathbf{y}) = \sum_{i=1}^{M} \left(w_{S,i}\kappa_{iii}^2(\mathbf{y}) + w_{K,i}\kappa_{iiii}^2(\mathbf{y}) \right) \tag{7.3.1}$$

The contrast (7.3.1) does not arise directly from the MI criterion, but it is weighted combination of two solutions (refer to equation (7.2.3)) which are not only contrast but also approximate MI under specific assumptions. However, in the absence of those assumptions, it is possible to obtain better results using (7.3.1) with appropriately selecting the values of free parameters. This improved efficacy is possible because the other contrasts (7.2.3) are only *approximate* MI solutions. The algebraic nature of cumulants is tensorial (with symmetry) [89]; thanks to the multilinearity of cumulants $\kappa_{...}(\mathbf{y})$ in $\kappa_{...}(\mathbf{z})$, the criterion (7.3.5) becomes an implicit function of the elements of the unitary matrix $\widetilde{\mathbf{Q}}$, we obtain:

$$\kappa_{iii}(\mathbf{y}) = \sum_{jkl} \widetilde{Q}_{ij}\widetilde{Q}_{ik}\widetilde{Q}_{il}\kappa_{jkl}(\mathbf{z}) \tag{7.3.2a}$$

$$\kappa_{iiii}(\mathbf{y}) = \sum_{jklm} \widetilde{Q}_{ij}\widetilde{Q}_{ik}\widetilde{Q}_{il}\widetilde{Q}_{im}\kappa_{jklm}(\mathbf{z}) \tag{7.3.2b}$$

where the unitary transformation matrix $\widetilde{\mathbf{Q}} = \mathbf{Q}^T$ is modeled as Givens rotation which is a rotation around the origin within the plane of two selected components μ and ν, and has the matrix form,

$$\widetilde{\mathbf{Q}} = \prod_{\mu=1}^{M-1} \prod_{\nu=\mu+1}^{M} \widetilde{\mathbf{Q}}^{\mu\nu} \tag{7.3.3a}$$

$$\widetilde{Q}_{ab}^{\mu\nu} = \begin{cases} \cos\phi_{\mu\nu}, & \text{for } (a,b) \in \{(\mu,\mu),(\nu,\nu)\} \\ -\sin\phi_{\mu\nu}, & \text{for } (a,b) \in \{(\mu,\nu)\} \\ \sin\phi_{\mu\nu}, & \text{for } (a,b) \in \{(\nu,\mu)\} \\ \delta_{ab}, & \text{otherwise} \end{cases} \tag{7.3.3b}$$

with Kronecker symbol δ_{ab} and rotation angle $\phi_{\mu\nu}$. $\widetilde{\mathbf{Q}}$ is a product of $(M(M-1)/2)$ Givens (or plane) rotation matrices $\widetilde{\mathbf{Q}}^{\mu\nu}$. The estimation of the plane rotation $\phi_{\mu\nu}$ is obtained by an iterative Jacobi method over the set of orthonormal matrices. The orthonormal transforms are thus obtained as a sequence of plane rotations. Each plane rotation is applied to a pair of coordinates, such that, $y_\mu \leftarrow y_\mu \cos\phi_{\mu\nu} + y_\nu \sin\phi_{\mu\nu}$ and $y_\nu \leftarrow -y_\mu \sin\phi_{\mu\nu} + y_\nu \cos\phi_{\mu\nu}$, while leaving the other coordinates unchanged[4]. Thus,

[4]This data-based Jacobi algorithm for ICA works through a sequence of *sweeps* on the whitened data until a given orthogonal contrast is optimized; sweep is defined to be a one complete pass through all the $M(M-1)/2$ possible pairs of distinct indices. In simple words, the Jacobi-iteration spans the whole set of rotation matrices in a sequential manner. It is mentioned in [27] that the updating step on a pair μ, ν partially undoes the effect of previous optimizations on pairs containing either μ or ν. For this reason, it is necessary to go through several sweeps before optimization is completed.

the Jacobi approach considers a sequence of two-dimensional ICA problems. Considering the subspace of only two selected components, the Givens rotation matrix becomes:

$$\tilde{\mathbf{Q}} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \tag{7.3.4}$$

That is, for $M = 2$, we get

$$\mathcal{J}(\phi) = w_{S,1}\kappa_{111}^2(\mathbf{y}) + w_{S,2}\kappa_{222}^2(\mathbf{y}) + w_{K,1}\kappa_{1111}^2(\mathbf{y}) + w_{K,2}\kappa_{2222}^2(\mathbf{y}) \tag{7.3.5}$$

Notice that the number of free parameters have been reduced to four. It can further be reduced to three, if the contrast is normalized by any one of the four parameters; however, the expression (7.3.5) has the advantage that any of the free parameters can be set to zero. In the blind scenario, where we usually have no a priori knowledge of mixing signals, the tuning of these weight parameters is not trivial. Suppose it is known that the mixing sources are highly skewed then we can limit the search space of these parameters by setting $w_{S,1} = 1$, $w_{K,1} = w_{K,2} = 0$, and look for the optimal value of $w_{S,2}$. Similarly, if it is known that the sources are symmetrical then we can set $w_{S,1} = w_{S,2} = 0$, $w_{K,1} = 1$, and look for the optimal value of $w_{K,2}$.

In an earlier attempt, Nandi studied the contrast (7.3.5) for a two-source scenario and put forward the following suggestions [98]:

1. $w_{S,1} = w_{S,2}$ and $w_{K,1} = w_{K,2}$,
2. $w_{S,j} \propto \sum_i \kappa_{iii}^2(\mathbf{s})$ and $w_{K,j} \propto \sum_i \kappa_{iiii}^2(\mathbf{s})$, for $i, j = 1, 2$,
3. $\alpha w_{S,i} + w_{K,i} = 1$ (normalized), $i = 1, 2$ and $\alpha \in \mathbb{R}$.

Exemplary, he suggested the following simple and suboptimum expressions ($i = 1, 2$):

$$w_{S,i} = \frac{\kappa_{111}^2(\mathbf{s}) + \kappa_{222}^2(\mathbf{s})}{\alpha(\kappa_{111}^2(\mathbf{s}) + \kappa_{222}^2(\mathbf{s})) + \kappa_{1111}^2(\mathbf{s}) + \kappa_{2222}^2(\mathbf{s})} \tag{7.3.6a}$$

$$w_{K,i} = \frac{\kappa_{1111}^2(\mathbf{s}) + \kappa_{2222}^2(\mathbf{s})}{\alpha(\kappa_{111}^2(\mathbf{s}) + \kappa_{222}^2(\mathbf{s})) + \kappa_{1111}^2(\mathbf{s}) + \kappa_{2222}^2(\mathbf{s})} \tag{7.3.6b}$$

The expressions (7.3.6) did provide better results in number of experiments; however, due to their heuristic nature, it is difficult to consider them a suitable choice for a general ICA problem. Interestingly, the study of the *single* weight parameter for the optimized use of a fourth-order contrast function has been studied in [94, 96, 146]. Motivated by the convincing results reported in these works, we also limit our search to a single weight parameter. We select $w_{S,1} = w_{S,2} = 1$ and $w_{K,1} = w_{K,2} = \beta$, which lead to the following contrast:

$$\mathcal{J}(\phi) = \kappa_{111}^2(\mathbf{y}) + \kappa_{222}^2(\mathbf{y}) + \beta\Big(\kappa_{1111}^2(\mathbf{y}) + \kappa_{2222}^2(\mathbf{y})\Big) \tag{7.3.7}$$

The optimal value of the single weight parameter β can be obtained by performing *small error analysis*; i.e., the value of β is optimum if it minimizes the asymptotic (large-sample) mean-square error. Thanks to the work reported in [94], this analysis can easily be carried out. First, we consider that the mixing matrix is orthonormal so that the prewhitening stage is not necessary. Further, the asymptotic analysis is carried out for the case of two real sources, subject to the planar (Givens) rotation. In such a case, it is assumed that a first stage has already realized the normalization of the observation vector, i.e., $\mathbf{z}$ is supposed to be a white vector.

Thus, we have to estimate an angle ϕ according to the maximization of $\mathcal{J}(\cdot)$, i.e., $\widehat{\phi} = \arg\max_{\phi} \mathcal{J}(\phi)$, where $\widehat{\phi}$ is an estimate of the true (separation) value $\widetilde{\phi}$. In practice, the maximization of contrast function does not provide the exact value of the parameter $\widetilde{\phi}$, since the true cumulants are actually approximated by the sample estimates. Replacing the expectations by sample averages leads to the empirical version of $\mathcal{J}(\mathbf{y})$, which is denoted $\widehat{\mathcal{J}}(\mathbf{y})$ and is given by

$$\widehat{\mathcal{J}}(\phi) = \widehat{\kappa}_{111}^2(\mathbf{y}) + \widehat{\kappa}_{222}^2(\mathbf{y}) + \beta\Big(\widehat{\kappa}_{1111}^2(\mathbf{y}) + \widehat{\kappa}_{2222}^2(\mathbf{y})\Big) \tag{7.3.8}$$

where $\widehat{\kappa}_{iii}(\mathbf{y}) = \frac{1}{N}\sum_{k=1}^{N} y_i^3(k)$, and $\widehat{\kappa}_{iiii}(\mathbf{y}) = -3 + \frac{1}{N}\sum_{k=1}^{N} y_i^4(k)$, $i = 1, 2$. As a result, an estimation error is involved in the estimation of the true value $\widetilde{\phi}$. The estimated angle $\widehat{\phi}$ is actually the solution of the estimating equation $\widehat{\mathcal{J}}'(\widehat{\phi}) = \partial\widehat{\mathcal{J}}(\widehat{\phi})/\partial\phi|_{\phi=\widehat{\phi}} = 0$. Approximating this derivative around the true value $\widetilde{\phi}$ by means of its Taylor series expansion yields: $\widehat{\mathcal{J}}'(\widehat{\phi}) \approx \widehat{\mathcal{J}}'(\widetilde{\phi}) + \widehat{\mathcal{J}}''(\widetilde{\phi})(\widehat{\phi} - \widetilde{\phi})$, where $\widehat{\mathcal{J}}''(\widetilde{\phi}) = \partial\widehat{\mathcal{J}}'(\phi)/\partial\phi|_{\phi=\widetilde{\phi}}$ and $\widehat{\mathcal{J}}'(\widetilde{\phi}) = \partial\widehat{\mathcal{J}}(\phi)/\partial\phi|_{\phi=\widetilde{\phi}}$. Assuming $\widehat{\phi}$ to be in the neighborhood of $\widetilde{\phi}$, we obtain $\widehat{\mathcal{J}}'(\widetilde{\phi}) \approx -\widehat{\mathcal{J}}''(\widetilde{\phi})(\widehat{\phi} - \widetilde{\phi})$. The *mean square error* (MSE) is given by

$$\mathrm{MSE} = \frac{\mathsf{E}\left[\left(\widehat{\mathcal{J}}'(\widetilde{\phi})\right)^2\right]}{\left(\mathsf{E}\left[\widehat{\mathcal{J}}''(\widetilde{\phi})\right]\right)^2} \tag{7.3.9}$$

When $\widehat{\phi} = \widetilde{\phi}$, $\mathbf{y} = \mathbf{s}$. The MSE expression (7.3.9) is generalized and is thus valid for any two-dimensional contrast for ICA problem. Further, the *strong law of large number* ensures that $\widehat{\mathcal{J}}''(\widetilde{\phi})$ converges with probability one to its expected value. As $N \to \infty$, we have

$$\mathsf{E}[\widehat{\mathcal{J}}''(\widetilde{\phi})] \to -(B_0 + B_1\beta) \tag{7.3.10}$$

where

$$B_0 = 3\left(\kappa_{111}^2(\mathbf{s}) + \kappa_{222}^2(\mathbf{s})\right) \tag{7.3.11a}$$

$$B_1 = 4\left(\kappa_{1111}^2(\mathbf{s}) + \kappa_{2222}^2(\mathbf{s})\right). \tag{7.3.11b}$$

Next we obtain $\mathsf{E}[(\widehat{\mathcal{J}}'(\widetilde{\phi}))^2]$ as follows:

$$\mathsf{E}\left[\left(\widehat{\mathcal{J}}'(\widetilde{\phi})\right)^2\right] \to \frac{A_0 - 2A_1\beta + A_2\beta^2}{N} \tag{7.3.12}$$

where

$$A_0 = 9\left(\kappa_{111}^2(\mathbf{s})\mathsf{E}\left[s_1^4\right] + \kappa_{222}^2(\mathbf{s})\mathsf{E}\left[s_2^4\right] - 2\kappa_{111}^2(\mathbf{s})\kappa_{222}^2(\mathbf{s})\right) \tag{7.3.13a}$$

$$\begin{aligned} A_1 = 12\big(&\kappa_{111}^2(\mathbf{s})\kappa_{2222}(\mathbf{s})\mathsf{E}\left[s_2^4\right] + \kappa_{222}^2(\mathbf{s})\kappa_{1111}(\mathbf{s})\mathsf{E}\left[s_1^4\right] \\ &-\kappa_{111}(\mathbf{s})\kappa_{1111}(\mathbf{s})\mathsf{E}\left[s_1^5\right] - \kappa_{222}(\mathbf{s})\kappa_{2222}(\mathbf{s})\mathsf{E}\left[s_2^5\right]\big) \end{aligned} \tag{7.3.13b}$$

$$A_2 = 16\left(\kappa_{1111}^2(\mathbf{s})\mathsf{E}\left[s_1^6\right] + \kappa_{2222}^2(\mathbf{s})\mathsf{E}\left[s_2^6\right] - 2\kappa_{1111}(\mathbf{s})\kappa_{2222}(\mathbf{s})\mathsf{E}\left[s_1^4\right]\mathsf{E}\left[s_2^4\right]\right) \tag{7.3.13c}$$

The MSE depends on the statistics of the sources and on the parameter β. We now easily derive the optimum value of β, denoted β^*, such that the MSE is minimum by solving the equation $\frac{\partial}{\partial\beta}\mathrm{MSE} = 0$, i.e.,

$$\beta^* = \frac{A_0B_1 + A_1B_0}{A_2B_0 + A_1B_1} = \frac{4A_0\left(\kappa_{1111}^2(\mathbf{s}) + \kappa_{2222}^2(\mathbf{s})\right) + 3A_1\left(\kappa_{111}^2(\mathbf{s}) + \kappa_{222}^2(\mathbf{s})\right)}{3A_2\left(\kappa_{111}^2(\mathbf{s}) + \kappa_{222}^2(\mathbf{s})\right) + 4A_1\left(\kappa_{1111}^2(\mathbf{s}) + \kappa_{2222}^2(\mathbf{s})\right)} \tag{7.3.14}$$

which indicates that the β^* depends on the statistics of mixing source and is independent of the coefficients of unknown mixing matrix. Hence, given the source statistics, we can obtain a contrast with minimum asymptotic MSE. In the scenario that nothing is known a priori about the source statistics, a possible simple strategy is to use statistical properties of the observed sources for the evaluation of β^*, and the separation can be repeated until β^* converges.

Notice a resemblance among the expressions (7.3.6) and the optimal β (7.3.14), the denominators in these expressions can be seen to be equal, if α is selected to be

$3A_2/(4A_1)$. Also notice that if the two sources s_1 and s_2 have the same statistics, then $\beta^* = 1$, and this is quite natural because in that case, nothing enables to make any statistical distinction between s_1 and s_2. The experiments in Section 7.6 will illustrate the validity of the asymptotic expression (7.3.9) and the performance improvements that can be derived from the use of the optimal weight parameter (7.3.14).

7.4 Derivation of the Proposed ICA Algorithm

We consider the case of real mixtures and assume that the angle of rotation is required to lie in the interval $(-\pi/2, \pi/2]$. Considering the pairwise estimation of angle of rotation, the optimization problem reduces to a one-dimensional search. Looking carefully at the optimization criterion (7.3.5) reveals that stationary points can be obtained by mere polynomial rooting technique (as initially suggested in [32]). First, we reformulate expressions (7.3.2) in a matrix form for the ease of derivation as given by

$$\kappa_{ijk}(\mathbf{y}) = \begin{bmatrix} \widetilde{Q}_{i1} \\ \widetilde{Q}_{i2} \end{bmatrix}^T \begin{bmatrix} \kappa_{111}(\mathbf{z}) & \kappa_{121}(\mathbf{z}) & \kappa_{112}(\mathbf{z}) & \kappa_{122}(\mathbf{z}) \\ \kappa_{211}(\mathbf{z}) & \kappa_{221}(\mathbf{z}) & \kappa_{212}(\mathbf{z}) & \kappa_{222}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \widetilde{Q}_{j1}\widetilde{Q}_{k1} \\ \widetilde{Q}_{j2}\widetilde{Q}_{k1} \\ \widetilde{Q}_{j1}\widetilde{Q}_{k2} \\ \widetilde{Q}_{j2}\widetilde{Q}_{k2} \end{bmatrix} \tag{7.4.1a}$$

$$\kappa_{ijkl}(\mathbf{y}) = \begin{bmatrix} \widetilde{Q}_{i1}\widetilde{Q}_{k1} \\ \widetilde{Q}_{i1}\widetilde{Q}_{k2} \\ \widetilde{Q}_{i2}\widetilde{Q}_{k1} \\ \widetilde{Q}_{i2}\widetilde{Q}_{k2} \end{bmatrix}^T \begin{bmatrix} \kappa_{1111}(\mathbf{z}) & \kappa_{1112}(\mathbf{z}) & \kappa_{1211}(\mathbf{z}) & \kappa_{1212}(\mathbf{z}) \\ \kappa_{1121}(\mathbf{z}) & \kappa_{1122}(\mathbf{z}) & \kappa_{1221}(\mathbf{z}) & \kappa_{1222}(\mathbf{z}) \\ \kappa_{2111}(\mathbf{z}) & \kappa_{2112}(\mathbf{z}) & \kappa_{2211}(\mathbf{z}) & \kappa_{2212}(\mathbf{z}) \\ \kappa_{2121}(\mathbf{z}) & \kappa_{2122}(\mathbf{z}) & \kappa_{2221}(\mathbf{z}) & \kappa_{2222}(\mathbf{z}) \end{bmatrix} \begin{bmatrix} \widetilde{Q}_{l1}\widetilde{Q}_{j1} \\ \widetilde{Q}_{l2}\widetilde{Q}_{j1} \\ \widetilde{Q}_{l1}\widetilde{Q}_{j2} \\ \widetilde{Q}_{l2}\widetilde{Q}_{j2} \end{bmatrix} \tag{7.4.1b}$$

To avoid trigonometric functions in $\widetilde{\mathbf{Q}}$ (7.3.4), we adopt the following form from [32]:

$$\widetilde{\mathbf{Q}} = \frac{1}{\sqrt{1+\theta^2}} \begin{bmatrix} 1 & \theta \\ -\theta & 1 \end{bmatrix} \tag{7.4.2}$$

where θ is an auxiliary variable defined as $\theta \overset{\text{def}}{=} \tan\phi$. Now, we expand the squares of third-order cumulants as a function of θ:

$$w_{S,1}\kappa_{111}^2(\mathbf{y}) + w_{S,2}\kappa_{222}^2(\mathbf{y}) = \frac{1}{(1+\theta^2)^3}\sum_{i=0}^{6} c_i\theta^i \tag{7.4.3}$$

where $c_6 = w_{S,2}b_0^2 + w_{S,1}b_3^2$, $c_5 = 2(w_{S,1}b_2b_3 - w_{S,2}b_0b_1)$, $c_4 = w_{S,1}(2b_1b_3 + b_2^2) + w_{S,2}(2b_0b_2 + b_1^2)$, $c_0 = w_{S,1}b_0^2 + w_{S,2}b_3^2$, $c_3 = 2(w_{S,1} - w_{S,2})(b_0b_3 + b_1b_2)$, $c_2 = w_{S,2}(2b_1b_3 + b_2^2) +$

$w_{S,1}(2b_0b_2 + b_1^2)$, and $c_1 = 2(w_{S,1}b_0b_1 - w_{S,2}b_2b_3)$. We define, for the sake of simplicity, $b_0 = \kappa_{111}(\mathbf{z})$, $b_1 = 3\kappa_{112}(\mathbf{z})$, $b_2 = 3\kappa_{122}(\mathbf{z})$ and $b_3 = \kappa_{222}(\mathbf{z})$. Similarly, we expand the squares of fourth-order cumulants:

$$w_{K,1}\kappa_{1111}^2(\mathbf{y}) + w_{K,2}\kappa_{2222}^2(\mathbf{y}) = \frac{1}{(1+\theta^2)^4}\sum_{i=0}^{8} d_i\theta^i \tag{7.4.4}$$

where $d_8 = w_{K,1}a_4^2 + w_{K,2}a_0^2$, $d_7 = 2(w_{K,1}a_3a_4 - w_{K,2}a_1a_0)$, $d_6 = 2(w_{K,1}a_4a_2 + w_{K,2}a_0a_2) + w_{K,2}a_1^2 + w_{K,1}a_3^2$, $d_5 = 2w_{K,1}(a_1a_4 + a_2a_3) - 2w_{K,2}(a_3a_0 + a_1a_2)$, $d_4 = (w_{K,1} + w_{K,2})(2a_3a_1 + 2a_4a_0 + a_2^2)$, $d_3 = 2w_{K,1}(a_3a_0 + a_1a_2) - 2w_{K,2}(a_1a_4 + a_2a_3)$, $d_2 = w_{K,2}(2a_4a_2 + a_3^2) + w_{K,1}(2a_0a_2 + a_1^2)$, $d_1 = 2(w_{K,1}a_1a_0 - w_{K,2}a_3a_4)$, and $d_0 = w_{K,1}a_0^2 + w_{K,2}a_4^2$. We define: $a_0 = \kappa_{1111}(\mathbf{z})$, $a_1 = 4\kappa_{1112}(\mathbf{z})$, $a_2 = 6\kappa_{1122}(\mathbf{z})$, $a_3 = 4\kappa_{1222}(\mathbf{z})$ and $a_4 = \kappa_{2222}(\mathbf{z})$. Combining (7.4.3) and (7.4.4), we obtain

$$\mathcal{J}(\theta) = \frac{1}{(1+\theta^2)^4}\sum_{i=0}^{8} e_i\theta^i \tag{7.4.5}$$

where $e_8 = c_6 + d_8$, $e_7 = c_5 + d_7$, $e_6 = c_4 + d_6 + c_6$, $e_5 = c_3 + d_5 + c_5$, $e_4 = c_2 + d_4 + c_4$, $e_3 = c_1 + d_3 + c_3$, $e_2 = c_0 + d_2 + c_2$, $e_1 = d_1 + c_1$, and $e_0 = d_0 + c_0$. Taking the derivative of (7.4.5) with respect to θ and setting that to zero, we obtain the following:

$$\Omega(\theta) = \sum_{k=0}^{8} f_k\theta^k = 0 \tag{7.4.6}$$

where $f_8 = e_7$, $f_7 = 2e_6 - 8e_8$, $f_6 = 3e_5 - 7e_7$, $f_5 = 4e_4 - 6e_6$, $f_4 = 5e_3 - 5e_5$, $f_3 = 6e_2 - 4e_4$, $f_2 = 7e_1 - 3e_3$, $f_1 = 8e_0 - 2e_2$ and $f_0 = -e_1$. The expression (7.4.6) is eighth-order; however, the proposal $w_{S,1} = w_{S,2}$ and $w_{K,1} = w_{K,2}$ will help it to get reduced to fourth-order. By selecting $w_{S,1} = w_{S,2}$, we notice that $c_6 = c_0$, $c_5 = -c_1$, $c_3 = 0$ and $c_4 = c_2$; similarly, with $w_{K,1} = w_{K,2}$, we obtain $d_8 = d_0, d_7 = -d_1, d_6 = d_2, d_5 = -d_3$, it makes us write (7.4.5) as follows:

$$\mathcal{J}(\theta) = \frac{\begin{pmatrix} (c_0 + d_0)(\theta^8 + 1) - (c_1 + d_1)(\theta^7 - \theta) \\ +(c_0 + c_2 + d_2)(\theta^6 + \theta^2) - (d_3 + c_1)(\theta^5 - \theta^3) \\ +(2c_2 + d_4)\theta^4 \end{pmatrix}}{(1+\theta^2)^4} \tag{7.4.7}$$

Thanks to Comon [32] for the substitution, $\xi = \theta - 1/\theta$, which simplifies (7.4.7) into a reduced-order form as a function of auxiliary variable ξ, given by:

$$\mathcal{J}(\xi) = \frac{1}{(4+\xi^2)^2}\sum_{i=0}^{4} h_i\xi^i \tag{7.4.8}$$

where $h_4 = c_0 + d_0, h_3 = -(c_1 + d_1), h_2 = 5c_0 + c_2 + 4d_0 + d_2, h_1 = -(4c_1 + 3d_1 + d_3)$ and $h_0 = d_4 + 4c_2 + 4c_0 + 2d_0 + 2d_2$. Taking the derivative of (7.4.8) with respect to ξ and setting that to zero, we obtain the following polynomial:

$$\Omega(\xi) = \sum_{k=0}^{4} g_k \xi^k = 0 \tag{7.4.9}$$

where $g_4 = -h_3/8$, $g_3 = 2h_4 - h_2/4$, $g_2 = 3h_3/2 - 3h_1/8$, $g_1 = h_2 - h_0/2$ and $g_0 = h_1/2$. The explicit analytical solutions to the roots of polynomial (7.4.9) can be found with standard algebraic procedure such as Ferrari's formula [108]. Preliminary experiments point out that, although complex-valued roots may appear as favorite, the best real-valued candidate root should typically be preferred. At most two of the roots correspond to the maxima of $\mathcal{J}(\xi)$. Similar to the findings in [35], there are in general only two real roots to polynomial $\Omega(\xi)$ and the contrast $\mathcal{J}(\xi)$ admits in general a single maximum. After finding the roots of (7.4.9) and the corresponding value of the contrast, we retain ξ that maximizes the contrast, and compute the θ (which corresponds to the tangent of the rotation angle) by solving

$$\theta^2 - \xi\theta - 1 = 0, \tag{7.4.10}$$

we retain θ which satisfies $\theta \in (-1, 1]$ and finally compute the rotation matrix (7.4.2). Next, we consider two cases of the proposed ICA algorithm as follows:

1. First, we assume constant values for free parameters: $w_{S,1} = w_{S,2} = 4$, $w_{K,1} = w_{K,2} = 1$, (or equivalently, $w_{S,1} = w_{S,2} = 1$, $w_{K,1} = w_{K,2} = \frac{1}{4}$). It results into the same contrast as in (7.2.4); however, unlike CuBICA, we have used root-finding method for the maximization of contrast. The resulting ICA algorithm is named *composite-order ICA algorithm (based on third- and fourth-order cumulants), COICA.*

2. Secondly, we assume $w_{S,1} = w_{S,2} = 1$ and $w_{K,1} = w_{K,2} = \beta$, where β is an optimized weight parameter, as we derived in Section 7.4. The resulting ICA algorithm is named *optimized composite-order ICA algorithm (based on third- and fourth-order cumulants), OCOICA.*

7.5 A Closed Form Estimator and Background Noise

In the fundamental real-valued two-source scenario, the ICA problem reduces to the identification of a single parameter, the unknown angle characterizing the Given's-rotation mixing matrix. A variety of closed-form methods for the estimation of this angle have been proposed in the literature [31, 144, 57, 147, 50, 148, 62, 97, 145]. These estimators consist of simple formulas involving straightforward operations on certain statistics of the whitened sensor output. Most of these share the common feature of being based on the fourth-order statistics of the whitened sensor output. It is interesting to note that all of the aforesaid closed-form estimators were based on fourth-order statistics.

We restate the weighted contrast (7.3.7) as follows:

$$\mathcal{J}(y) = \kappa_{111}^2(\mathrm{y}) + \kappa_{222}^2(\mathrm{y}) + \beta\left(\kappa_{1111}^2(\mathrm{y}) + \kappa_{2222}^2(\mathrm{y})\right) \tag{7.5.1}$$

Owing to [21], it is possible to express the contrast (7.5.1) as the function of ϕ as follows:

$$\mathcal{J}(\phi) = \mathcal{A}_0 + \mathcal{A}_4 \cos(4\phi + \phi_4) + \mathcal{A}_8 \cos(8\phi + \phi_8) \tag{7.5.2}$$

where $\mathcal{A}_0$, $\mathcal{A}_4$ and $\mathcal{A}_8$ are constants and depend upon the statistics of mixing signals. On the other hand, ϕ_4 and ϕ_8 are such constants which also contain the information of Given's rotation. Note that the first term $\mathcal{A}_0$ plays no role in the estimation of ϕ. Similarly constants $\mathcal{A}_8$ and ϕ_8 do not comprise of third-order statistical information and contribute no significant role if the mixing sources are asymmetrical in nature. One can refer to [21] for the detailed expressions for these constants which are obtained for the specific case $\beta = 1$). Importantly, the constant ϕ_4 in the middle term not only comprise of third- and fourth-order statistics but can be used to obtain a closed form estimator for the separation of mixed symmetrical/asymmetrical sources [93]. The angle ϕ that maximizes $\mathcal{A}_4 \cos(4\phi + \phi_4)$ is

$$\hat{\phi} = -\frac{\phi_4}{4} = -\frac{1}{4}\arctan(\mathcal{S}_0 + \beta\mathcal{S}_1, \mathcal{C}_0 + \beta\mathcal{C}_1) \tag{7.5.3}$$

which exploits the relations $\mathcal{S}_0 + \beta\mathcal{S}_1 = \sin\phi_4$ and $\mathcal{C}_0 + \beta\mathcal{C}_1 = \cos\phi_4$. Also $\arctan(y, x)$ is the unique angle $\alpha \in (-\pi, \pi]$ for which $\cos(\alpha) = (x/\sqrt{x^2+y^2})$ and $\sin(\alpha) = (y/\sqrt{x^2+y^2})$.

Constant $\mathcal{S}_0, \mathcal{S}_1, \mathcal{C}_0$ and $\mathcal{C}_1$ are computed as:

$$\begin{aligned}
\mathcal{S}_0 &= 24\big(\kappa_{111}(\mathbf{z})\kappa_{112}(\mathbf{z}) - \kappa_{122}(\mathbf{z})\kappa_{222}(\mathbf{z})\big)\\
\mathcal{S}_1 &= 4\Big(7\big(\kappa_{1111}(\mathbf{z})\kappa_{1112}(\mathbf{z}) - \kappa_{1222}(\mathbf{z})\kappa_{2222}(\mathbf{z})\big)\\
&\quad + 6\kappa_{1122}(\mathbf{z})\big(\kappa_{1112}(\mathbf{z}) - \kappa_{1222}(\mathbf{z})\big) + \kappa_{1111}(\mathbf{z})\kappa_{1222}(\mathbf{z}) - \kappa_{1112}(\mathbf{z})\kappa_{2222}(\mathbf{z})\Big)\\
\mathcal{C}_0 &= 6\Big(\kappa_{111}^2(\mathbf{z}) + \kappa_{222}^2(\mathbf{z}) - 3\big(\kappa_{112}^2(\mathbf{z}) + \kappa_{122}^2(\mathbf{z})\big) - 2\big(\kappa_{111}(\mathbf{z})\kappa_{122}(\mathbf{z}) + \kappa_{112}(\mathbf{z})\kappa_{222}(\mathbf{z})\big)\Big)\\
\mathcal{C}_1 &= 7\big(\kappa_{1111}^2(\mathbf{z}) + \kappa_{2222}^2(\mathbf{z})\big) - 36\kappa_{1122}^2(\mathbf{z}) - 2\kappa_{1111}(\mathbf{z})\kappa_{2222}(\mathbf{z}) - 32\kappa_{1112}(\mathbf{z})\kappa_{1222}(\mathbf{z})\\
&\quad - 12\big(\kappa_{1111}(\mathbf{z})\kappa_{1122}(\mathbf{z}) + \kappa_{1122}(\mathbf{z})\kappa_{2222}(\mathbf{z})\big) - 16\big(\kappa_{1112}^2(\mathbf{z}) + \kappa_{1222}^2(\mathbf{z})\big)
\end{aligned}$$

If we consider $\beta = 1$, then the closed-form estimator (7.5.3) provides an equivalent formulation of ICA algorithm as CuBICA34a [21]; however, note that unlike CuBICA34a, which is a search-based algorithm, the proposed expression (7.5.3) is a closed-form estimator.

In practice, the mixing model should also take into account a possible additive noise. This is considered hereafter because we want to take into account both the measurement noises and errors resulting from the first stage of whitening. Hence, now, the mixing model we consider reads $\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{g}$, where $\mathbf{g}$ is the vector of additive noise. In a two-source scenario, each noise $g_i, i \in \{1,2\}$ is a zero-mean, independent and identically distributed Gaussian random signal with equal power, i.e., $\mathsf{E}[g_1^2] = \mathsf{E}[g_2^2] = \sigma^2$. Moreover, $g_i, i \in \{1,2\}$ are assumed statistically mutually independent and independent of the sources $s_i, i \in \{1,2\}$.

The formula of optimal weight β^* remains same as we derived in Section 7.3:

$$\beta^* = \frac{A_0B_1 + A_1B_0}{A_2B_0 + A_1B_1}, \tag{7.5.4}$$

However, the auxiliary variables now include the statistics of noise, *viz*

$$A_0 = 9\left(B_0\sigma^6 + 3B_0\sigma^4 + (3B_0 + d_0)\,\sigma^2 + B_0 + d_0 - 2c_0^2\right) \tag{7.5.5a}$$

$$A_1 = 12\left(d_1\sigma^6 + (3d_1 - 10d_0)\,\sigma^4 - (d_3 - d_2 - 3d_1 + 20d_0)\,\sigma^2 - d_3 + d_2 + d_1 - 10d_0\right) \tag{7.5.5b}$$

$$\begin{aligned}
A_2 = 16\,\Big(&\left(\tfrac{15}{4}B_1 - 18c_1\right)\sigma^8 + (15B_1 - 72c_1)\,\sigma^6 + \left(d_5 - d_4 + \tfrac{90}{4}B_1 - 108c_1\right)\sigma^4\\
&+ (d_7 + d_6 + 2d_5 - 2d_4 - 72c_1 + 15B_1)\,\sigma^2 + d_7 + d_6 + d_5 - d_4 - 2c_1^2 - 18c_1 + \tfrac{15}{4}B_1\Big)
\end{aligned} \tag{7.5.5c}$$

where auxiliary variables are defined as

$$B_0 = 3\left(\kappa_{111}^2(\mathrm{s}) + \kappa_{222}^2(\mathrm{s})\right) \tag{7.5.6a}$$

$$B_1 = 4\left(\kappa_{1111}^2(\mathrm{s}) + \kappa_{2222}^2(\mathrm{s})\right) \tag{7.5.6b}$$

$$c_0 = \kappa_{111}(\mathrm{s})\kappa_{222}(\mathrm{s}) \tag{7.5.6c}$$

$$c_1 = \kappa_{1111}(\mathrm{s})\kappa_{2222}(\mathrm{s}) \tag{7.5.6d}$$

$$d_0 = \kappa_{111}^2(\mathrm{s})\kappa_{1111}(\mathrm{s}) + \kappa_{222}^2(\mathrm{s})\kappa_{2222}(\mathrm{s}) \tag{7.5.6e}$$

$$d_1 = 3\left(\kappa_{111}^2(\mathrm{s})\kappa_{2222}(\mathrm{s}) + \kappa_{222}^2(\mathrm{s})\kappa_{1111}(\mathrm{s})\right) \tag{7.5.6f}$$

$$d_2 = \kappa_{111}^2(\mathrm{s})\kappa_{2222}^2(\mathrm{s}) + \kappa_{222}^2(\mathrm{s})\kappa_{1111}^2(\mathrm{s}) \tag{7.5.7a}$$

$$d_3 = \kappa_{111}(\mathrm{s})\kappa_{1111}(\mathrm{s})\kappa_{11111}(\mathrm{s}) + \kappa_{222}(\mathrm{s})\kappa_{2222}(\mathrm{s})\kappa_{22222}(\mathrm{s}) \tag{7.5.7b}$$

$$d_4 = 6\left(\kappa_{1111}^2(\mathrm{s})\kappa_{2222}(\mathrm{s}) + \kappa_{2222}^2(\mathrm{s})\kappa_{1111}(\mathrm{s})\right) \tag{7.5.7c}$$

$$d_5 = 15\left(\kappa_{1111}^3(\mathrm{s}) + \kappa_{2222}^3(\mathrm{s})\right) \tag{7.5.7d}$$

$$d_6 = 10\left(\kappa_{111}^2(\mathrm{s})\kappa_{1111}^2(\mathrm{s}) + \kappa_{222}^2(\mathrm{s})\kappa_{2222}^2(\mathrm{s})\right) \tag{7.5.7e}$$

$$d_7 = \kappa_{1111}^2(\mathrm{s})\kappa_{111111}(\mathrm{s}) + \kappa_{2222}^2(\mathrm{s})\kappa_{222222}(\mathrm{s}) \tag{7.5.7f}$$

which indicates that the β^* depends on the statistics of mixing source and additive noise, and is independent of the coefficients of unknown mixing matrix. Hence, given the source and noise statistics, we can obtain a contrast with minimum asymptotic m.s.e. Finally, with the help of β^* (7.5.4), the optimum value of Givens rotation is estimated as:

$$\hat{\phi}^* = -\frac{1}{4}\arctan(\mathcal{S}_0 + \beta^*\mathcal{S}_1, \mathcal{C}_0 + \beta^*\mathcal{C}_1) \tag{7.5.8}$$

The estimator (7.5.8) is named *Closed-Form OCOICA (CF-OCOICA).*

7.6 Simulation Results

In order to illustrate the potential benefits of the proposed algorithms, some computer simulations are now presented. We intend to compare the performance of COICA and OCOICA with the joint diagonalization of third-order cumulant matrices (Com3) [33], the joint diagonalization of fourth-order cumulant matrices (Com4) [32], and the joint diagonalization of third- and fourth-order cumulant matrices (CuBICA) [21]. The performance measure, interference-to-signal ratio (ISR), introduced in [29], has been used in our simulation to characterize the restitution quality quantitatively. The performance index reads

$$\mathrm{ISR} = \sum_{i=1}^{M} \left(\frac{\sum_{j=1}^{n} |c_{ij}|^2}{\max_j |c_{ij}|^2} - 1 \right) \tag{7.6.1}$$

where c_{ij} represents the element (i, j) of the global mixing-unmixing matrix $\mathbf{C}$. In the two-signal case, the ISR approximates the MSE of the angle estimates around any valid separation solution [146].

We consider two cases of sources:

1. *A parameterized source:* We borrow a parameterized source $s(\alpha)$ from [93]. This is a discrete i.i.d. signal that takes its values in the set $\{-1, 0, \alpha\}$ with the respective probability $\{1/(1+\alpha), (\alpha-1)/\alpha, 1/(\alpha(1+\alpha))\}$. The real parameter α is called cumulant parameter, $\alpha \geq 1$. It can easily be shown that $\mathsf{E}[s] = 0$, $\mathsf{E}[s^2] = 1$, $\kappa_3(s) = \alpha - 1$ and $\kappa_4(s) = \alpha^2 - \alpha - 2$. Note that various (discrete) distributions can be obtained with appropriate values of α as given by:

   - $\alpha > 2$ gives Leptokurtic ($\kappa_4 > 0$)
   - $\alpha < 2$ gives Platykurtic ($\kappa_4 < 0$)
   - $\alpha = 2$ gives Mesokurtic ($\kappa_4 = 0$)
   - $\alpha > 1$ gives Asymmetrical ($\kappa_3 \neq 0$)
   - $\alpha = 1$ gives Symmetrical ($\kappa_3 = 0$)

2. *Synthetic sources:* Random sources with desired skewness and kurtosis are generated by *Fleishman's method* [47]. The sources used in simulation are labeled 1 to 30 and are listed in Table I. Except for signals, labeled $5, 6, 7$ and 11, all signals are skewed; similarly, except for signals, labeled $7, 12$ and 13, all signals have non-zero kurtosis. All signals are drawn with zero-mean and unit-variance.

   Fleishman proposed a form of the transformation on normal deviate $\omega \sim \mathcal{N}(0, 1)$ which is $s = a + b\omega + c\omega^2 + d\omega^3$. If $\kappa_3(s)$ and $\kappa_4(s)$ are the desired skewness and kurtosis, respectively, then the four coefficients $\{a, b, c, d\}$ are computed by solving

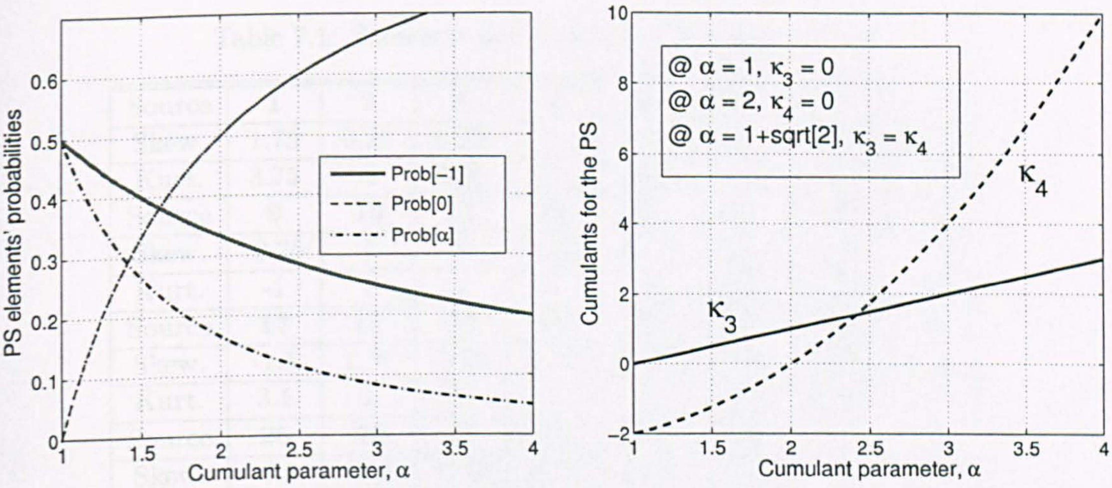


Figure 7.1: Parameterized source: (a) elements probabilities, and (b) skewness and kurtosis versus α.

simultaneously the following four equations:[5]

$$\begin{aligned}
\mu_s &= 0 = a + c, \\
\sigma_s^2 &= 1 = b^2 + 6bd + 2c^2 + 15d^2, \\
\kappa_3(s) &= 2c(b^2 + 24bd + 105d^2 + 2), \\
\kappa_4(s) &= 24[bd + c^2(1 + b^2 + 28bd) + d^2(12 + 48bd + 141c^2 + 225d^2)]
\end{aligned}$$

7.6.1 Experiment 1: ISR versus Cumulant Parameter

For case-1 of sources, the sample size (number of observations) is held constant, and we plot the estimated ISR performance as a function of cumulant parameter α. The mixing matrix is two-by-two, and both signals are drawn (independently) from the case-1 of sources. The mixing matrix is taken fixed for all Monte-Carlo experiments, and is given by:

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0.9 & 1 \end{pmatrix} \tag{7.6.2}$$

The condition number of the matrix (7.6.2) is 38; which is pretty high for these types of simulation. The ISR is computed for 50 different values of α ranging from 1 to 2.5 as

[5]The solution of these equations can be found within a space, which can be described by a parabola defined by [47]

$$\kappa_3^2(s) < 0.0629576\kappa_4(s) + 0.0717247$$

This is the limitation of the Fleishman's method.

Table 7.1: Skewness and kurtosis of the source data

Source	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Skew.	1.75	0.25	0.25	1	0	0	0	-1.75
Kurt.	3.75	-1	2.25	2	3.75	-1	0	3.75
Source	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Skew.	-0.25	1	0	0.75	0.25	1.5	-1.75	1.5
Kurt.	-1	1	2	0	0	3.75	3.75	3.5
Source	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Skew.	-1.5	1.25	-1.25	0.75	-0.75	0.5	-0.5	0.5
Kurt.	3.5	2	2	2	2	-0.25	-0.25	-0.5
Source	**25**	**26**	**27**	**28**	**29**	**30**		
Skew.	-0.5	0.25	-0.25	-0.25	0.25	-0.25		
Kurt.	-0.5	-0.25	-0.25	-1	-0.75	-0.75		

depicted in Fig. 7.2. The sample size N is taken to be 5000 for all algorithms and each trace of ISR is averaged over 700 Monte-Carlo realizations.

Notice that the ISR floor of Com3 is decreasing with an increase in α, which is quite natural, as the magnitude of third-order cumulant increases with α. For α close to 1, the poor performance of Com3 is due to the fact that the third-order cumulants do not bring sufficient statistical information since their values are near zero. For $\alpha > 2$, however, the performance of Com3 can be seen to be much better than Com4 and slightly better than CuBICA and COICA. The better performance of Com3 over Com4 (for $\alpha > 2$) is quite justified based on the findings of [63], where it was shown that by considering the asymmetric nature of sources, one can gain better performance over solely fourth-order schemes. Moreover, the better performance of Com3 in comparison to CuBICA or COICA makes it clear that merely a joint (un-optimized) use of third- and fourth-order cumulants can not guarantee a better performance.

Notice that the performance of Com4 becomes very poor in the neighborhood of $\alpha = 2$. This is not surprising, because, the fourth-order cumulants do not bring sufficient statistical information since their values are near zero. Moreover, in spite of the difference in their algorithmic formulation, the performance of CuBICA and COICA can be seen to be exactly the same for all values of α.

Notice that the performance of the OCOICA, in comparison to Com3, Com4, CuBICA and COICA, is almost (or totally) insensitive to the variation in the statistics of the sources. For $1 \leq \alpha < 1.5$, the ISR floor of OCOICA is almost equal to those of

Com4, CuBICA and COICA. Moreover, for $\alpha \geq 1.5$, the ISR floor of OCOICA is lower than those of others by at least 5 dB. The performance gain, achieved by OCOICA, is significant. It is important to notice that the calculation of β in OCOICA algorithm was carried out by directly using the observed (mixed) data and no *a priori* information of source statistics was assumed to be known to the algorithm.

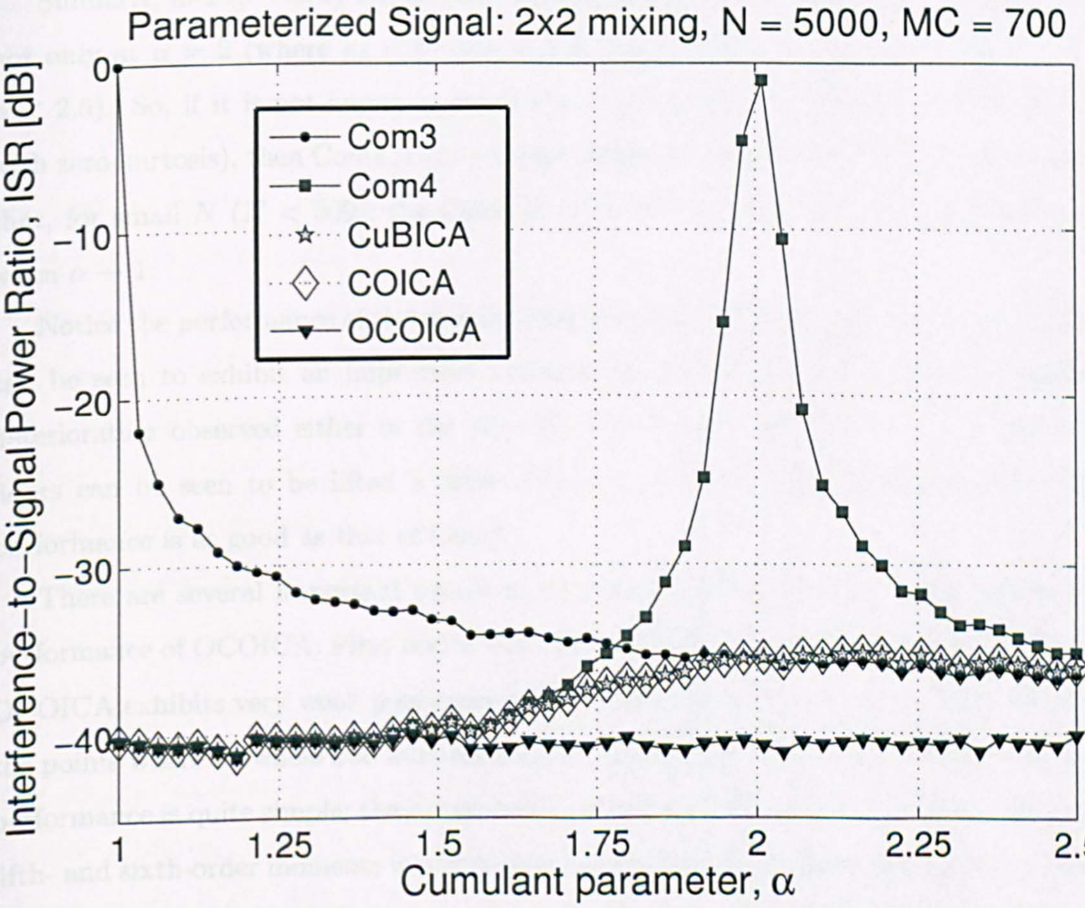


Figure 7.2: ISR performance for parameterized signal in 2-signal mixing scenario.

7.6.2 Experiment 2: ISR versus Cumulant Parameter for various Sample Sizes

This experiment provides a detailed account on the ISR performances, partially investigated in Experiment 1. Here, we obtained the ISR performances of Com3, Com4, COICA and OCOICA for various sample sizes ($N = 125 \times 2^i, i = 0, 1, \cdots, 7$) versus cumulant parameter $\alpha \in [1, 3]$. Results obtained for Com3, Com4, COICA and OCOICA are depicted in Figures 7.3(a), 7.3(b), 7.4(a) and 7.4(b), respectively. Notice in Fig.

7.3(a) that Com3 exhibits a satisfactory performance for all values of N (even as small as $N = 125$). However, when $\alpha \to 1$, the distribution comes close to symmetry, and Com3 stops working and results with an ISR as high as 5dB. So, if it is not known *a priori* that the sources are *symmetrical*, which is the case here when $\alpha \to 1$, then Com3 is not an appropriate ICA algorithm for BSS.

Similarly, in Fig. 7.3(b) notice that Com4's performance deteriorates (significantly) not only at $\alpha = 2$ (where $\kappa_4 = 0$) but in the large vicinity around it (which is $1.5 < \alpha < 2.5$). So, if it is not known *a priori* that the sources are *mesokurtic* (distribution with zero kurtosis), then Com4 is not an appropriate ICA algorithm for BSS. Also notice that, for small N ($N < 500$), the Com4 is found to be unable to perform satisfactory when $\alpha \to 1$.

Notice the performance of the proposed algorithm COICA in Fig. 7.4(a); the COICA can be seen to exhibit an impressive behavior for all values of N with no significant deterioration observed either in the vicinity of $\alpha = 1$ or that of $\alpha = 2$. Though ISR floors can be seen to be lifted a little around $\alpha = 2$, but still (in this vicinity) the performance is as good as that of Com3.

There are several important points to be noticed in Fig. 7.4(b), which depicts the performance of OCOICA. First notice that, for the number of samples less than 500, the OCOICA exhibits very weak performance in the vicinity of $\alpha = 1$ and $\alpha = 2$. These are the points where skewness and kurtosis vanish, respectively. The reason behind the weak performance is quite simple; the computation of optimal weight, β^*, requires to compute fifth- and sixth-order moments which cannot be obtained with reasonable accuracy using a small set of samples. Moreover, instead of mixing signal statistics, the algorithm is using observed signal properties to compute β^*, which is already a suboptimum way to go with. Secondly, notice that, *after being held constant irrespective of the value of* α, the ISR floor starts increasing. Let the point (i.e., the value of α) after which ISR floor departs from a constant level be termed as *take-off* value of α, denoted as α_N. Before we answer why this departure behavior occurs, we would like to highlight that if N_1 and N_2 are two different sample sizes, with $N_2 > N_1$, then $\alpha_{N_2} > \alpha_{N_1}$. This is a very important property of OCOICA, which tells that, *if a large enough sample-set is ensured, then irrespective of the value of* α, *OCOICA is capable of giving a successful source separation, with separation quality much better than those of Com3, Com4 and*

COICA (including CuBICA).

It is evident from this experiment that, with the parameterized signal in consideration, the performance of OCOICA will ultimately go degraded after a certain $\alpha > \alpha_N$ for some sample-size N. The reason is that with an increase in the value of α, both the skewness and kurtosis increase. The variance in the estimation error of β^* increases with the magnitude of kurtosis, since the variance in the estimation error of fifth- and sixth-order moments depends on the magnitude of kurtosis.

7.6.3 Experiment 3: ISR versus Sample Size

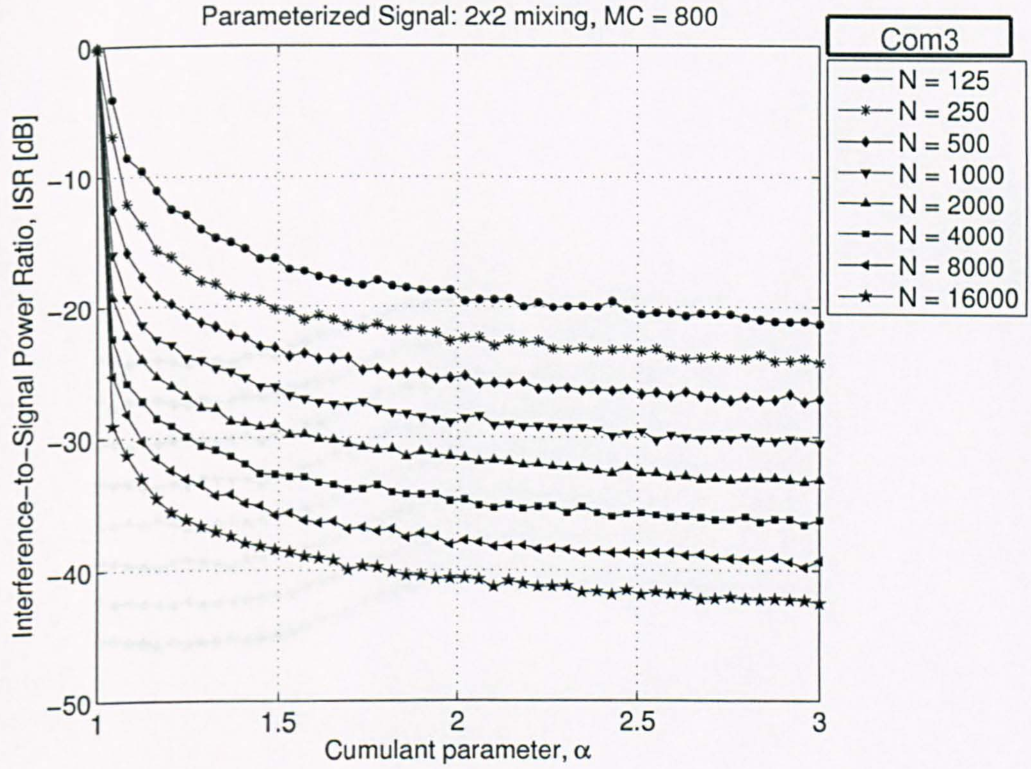
For case-1 of sources, now the cumulant parameter is held constant, and we plot the estimated ISR performance as a function of sample size. In Fig. 7.5(a), we use $\alpha = 1.7$ (this case corresponds to a negative fourth-order cumulant), whereas in Fig. 7.5(b), we use $\alpha = 2.5$ (this case corresponds to a positive fourth-order cumulant). The mixing matrix is again two-by-two, and both signals are drawn (independently) from the case-1 of sources. The mixing matrix, as specified in (7.6.2), was used in the simulation. This experiment shows that the proposed algorithm is capable of giving better results for both skewed sub-Gaussian and skewed super-Gaussian signals when the number of observations is moderate, i.e., all around 500 (or more).

7.6.4 Experiment 4: Contrast Parameter versus Cumulant Parameter

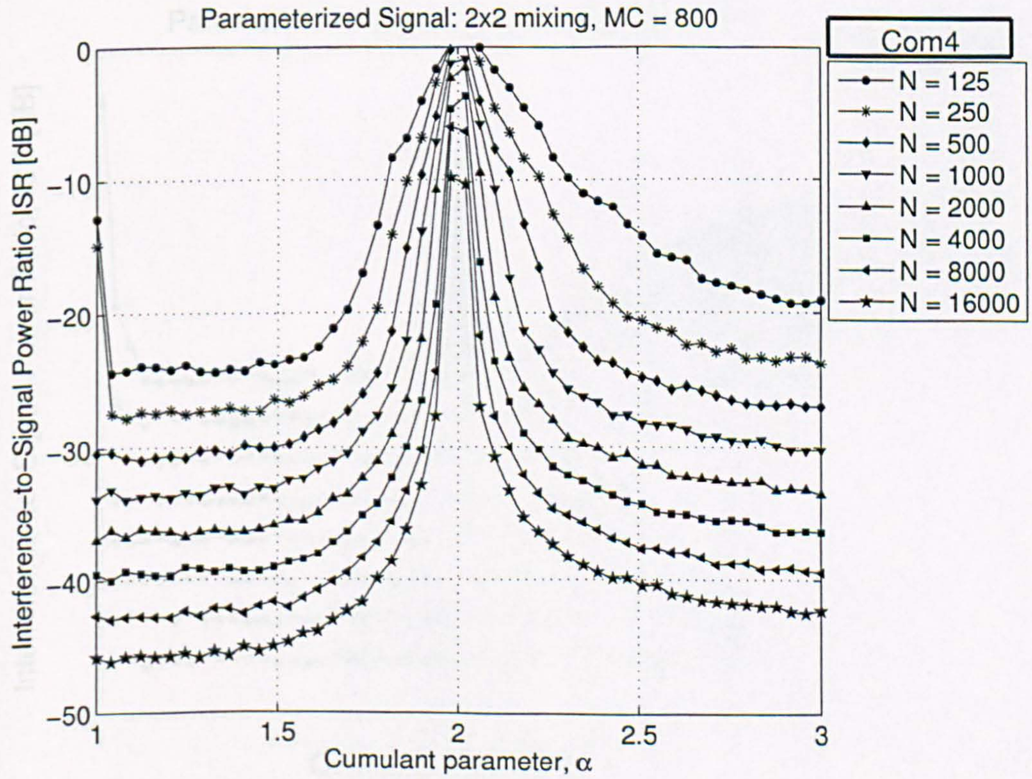
The optimal value of the contrast parameter β, β^*, can be computed in a closed form for case-1 source. It can easily be shown that $\mathsf{E}[s^5] = \alpha^3 - \alpha^2 + \alpha - 1$ and $\mathsf{E}[s^6] = \alpha^4 - \alpha^3 + \alpha^2 - \alpha + 1$, which lead us to obtain the following expression:

$$\beta^* = \frac{3}{4}\frac{1}{(1+\alpha)(2-\alpha)} \tag{7.6.3}$$

The expression (7.6.3) is plotted in Fig. 7.6. To conform the analytical value of β^* in (7.6.3), we obtained its estimated values under the same mixing scenario as specified in Experiment 1 except that $N = 500,000$. The analytical and simulated values can be seen to be conforming with each other for all values of α.

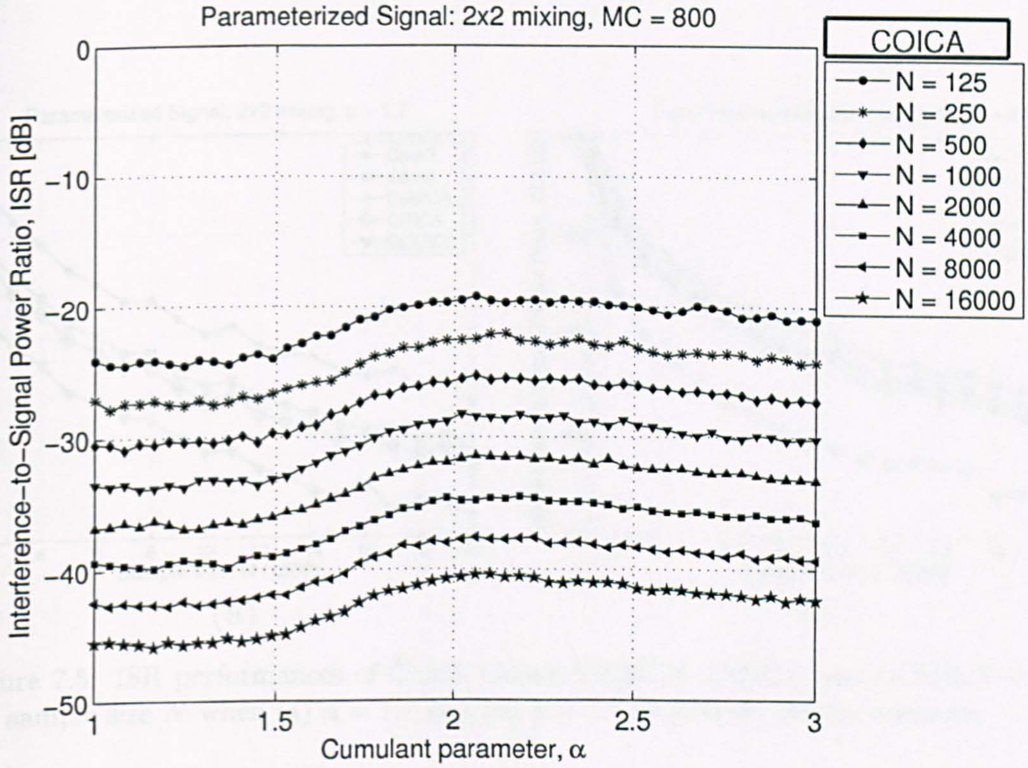


(a)

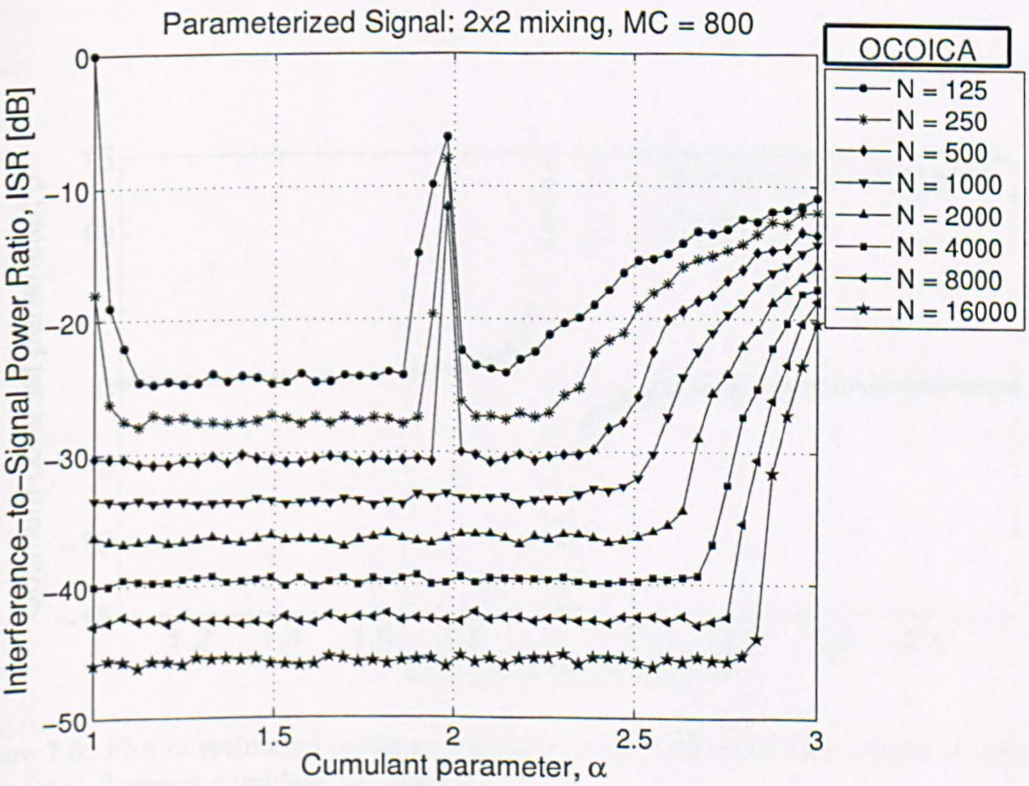


(b)

Figure 7.3: ISR performance for various N versus the cumulant parameter α for Com3 and Com4.



(a)



(b)

Figure 7.4: ISR performance for various N versus the cumulant parameter α for COICA and OCOICA.

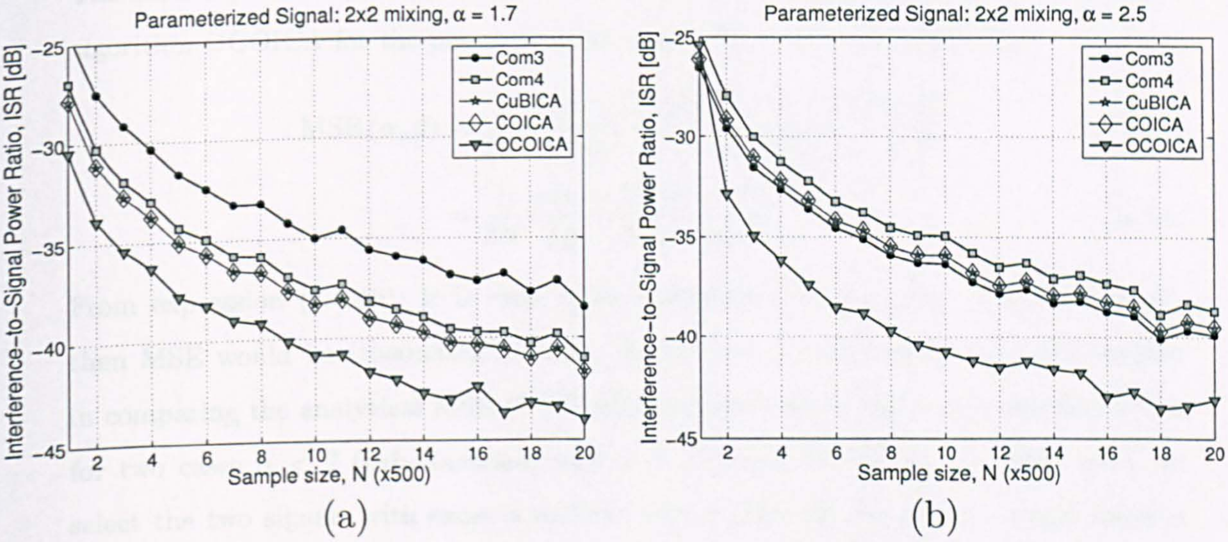


Figure 7.5: ISR performances of Com3, Com4, CuBICA, COICA and OCOICA w.r.t. the sample size N when (a) $\alpha = 1.7$ and (b) $\alpha = 2.5$ in 2-signal mixing scenario.

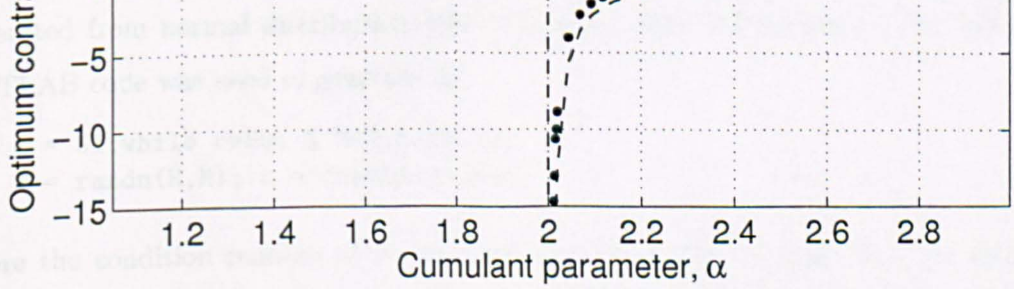


Figure 7.6: Plot of estimated (open circles) and theoretical (solid line) values of contrast parameter β versus cumulant parameter α.

7.6.5 Experiment 5: MSE versus Contrast Parameter

The MSE expression (7.3.9) can be used to obtain the MSE performance of the proposed algorithm OCOICA for the parameterized signal. We obtain the following:

$$\mathrm{MSE}(\alpha,\beta) = \frac{1}{2N}\frac{\alpha(\alpha-1)^2\left(4\left(\alpha^2-\alpha-2\right)\beta+3\right)^2}{\left(4\left(\alpha^2-\alpha-2\right)\beta+3\left(\alpha-1\right)\right)^2} \tag{7.6.4a}$$

$$= \frac{1}{2N}\frac{\alpha(\alpha-1)^2(\beta-\beta^*)^2}{\left(\beta-\beta^*\left(\alpha-1\right)\right)^2} \tag{7.6.4b}$$

From expression (7.6.4b), it is clear that, irrespective of the value of α, if $\beta = \beta^*$, then MSE would be (theoretically) zero. In this set of experiment, we are interested in comparing the analytical MSE (7.3.9) with those obtained from computer simulation for two cases $\alpha < 2$ (sub-Gaussian) and $\alpha > 2$ (super-Gaussian). In both cases, we select the two signals with same α and are mixed through the unitary transformation with $\phi = 15°$. The results of MSE performance are depicted in Figure 7.7(a) and (b) for $\alpha = 1.8$ and 2.2, respectively. The values of sample-size and Monte-Carlo runs are mentioned in the figures. It is clear from these results that our analytical findings are in complete conformation with simulation results and the use of optimal weight parameter β can provide significant improvement for both types of distributions (sub- and super-Gaussian).

7.6.6 Experiment 6: ISR versus Sample Size

For case-2 of sources (as listed in Table I), we experimented with several combination of signals and estimated the ISR performance as a function of sample size. The mixing matrix was taken to be of order $M \times M$, where $M = 2$, 5 and 10. These matrices were generated from normal distribution with zero-mean and unit-variance. The following MATLAB code was used to generate $\mathbf{A}$:

```
c = 51;while c>50; % M=2,5 or 10;
A = randn(M,M); c = cond(A); end;
```

where the condition number of $\mathbf{A}$ was constrained to be less than 50. The mixture is first whitened via PCA based on the singular value decomposition of the observed data matrix. The ISR performance is obtained for several sample sizes N. The curves have been averaged ν independent Monte Carlo runs. The value of the product νN is selected to be $1 \cdot 10^8$, $1 \cdot 10^6$ and $5 \cdot 10^5$ for $M = 2$, 5 and 10, respectively. The

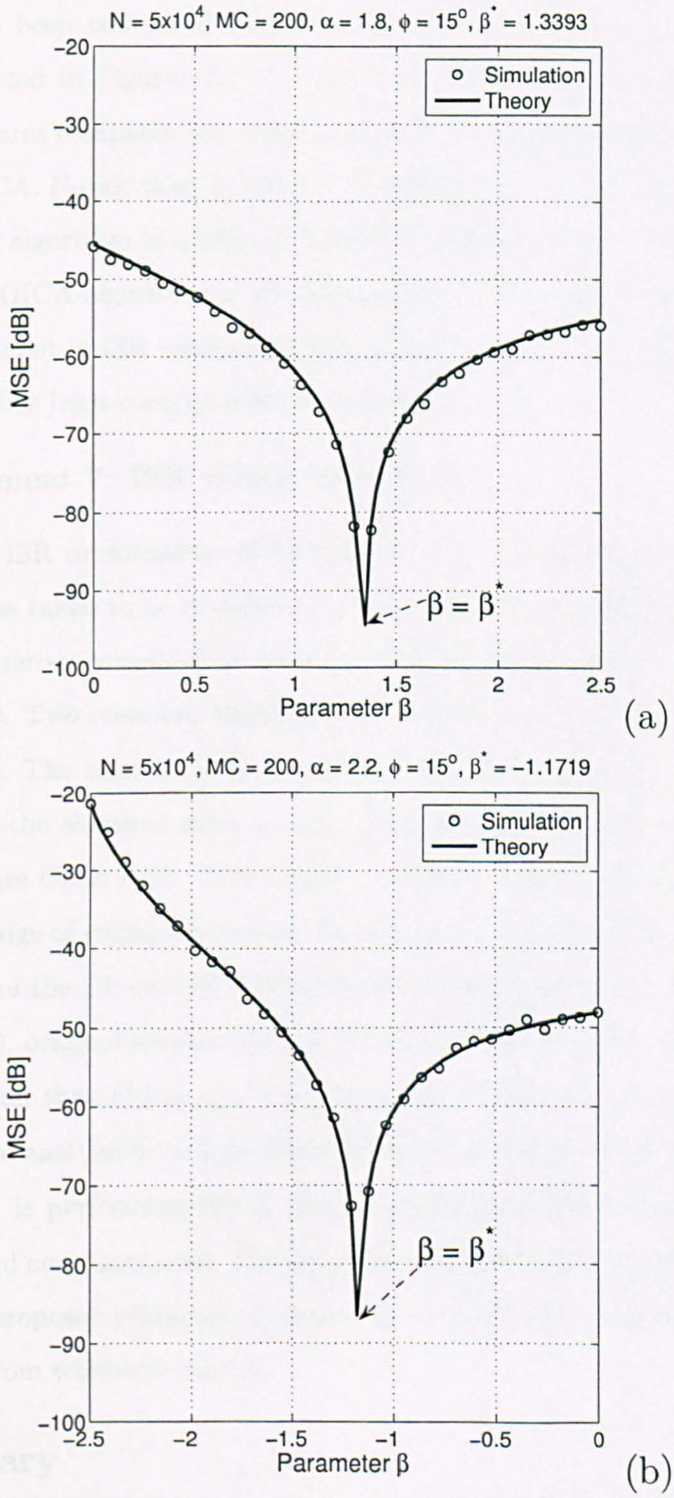


Figure 7.7: Analytical and estimated MSE of OCOICA versus contrast parameter β when (a) $\alpha = 1.8$ and (b) $\alpha = 2.2$ in 2-signal mixing scenario.

values of β have been computed from the statistical knowledge of observed-sources. Results are depicted in Figures 7.8, 7.9 and 7.10 for $M = 2$, 5 and 10, respectively. Each of these figures compares the performances of COICA and OCOICA with Com3, Com4 and CuBICA. Notice that, in spite of the differences in optimization method, the proposed COICA algorithm is similar to CuBICA in performance. Secondly, notice that the proposed OCOICA algorithm is outperforming all other algorithms . Finally notice that the improvement in ISR reduction achieved by OCOICA is consistent even though the parameter β has been computed from observed-sources.

7.6.7 Experiment 7: ISR versus Sample Size

We estimate the ISR performance of CF-OCOICA as a function of sample size. The mixing matrix was taken to be of order 2×2. Matrices (**A**) are generated from normal distribution with zero-mean and unit-variance. The condition number of **A** is constrained to be less than 50. Two cases are considered – no noise $\sigma = 0$ and with noise $\sigma = 0.0316$ [i.e., SNR=30dB]. The mixture is first whitened via PCA based on the singular value decomposition of the observed data matrix. The curves have been averaged over 2000 independent Monte Carlo runs. The weight parameter β has been computed from the statistical knowledge of whitened-sources. Results are depicted in Figure 7.11 comparing the performance of the CF-OCOICA with those of Com3, Com4 and CuBICA.

In Figure 1(a), original sources are highly asymmetric in nature, that is why Com3 is performing better than Com4 and CuBICA, while in Figure 1(b), original sources are moderately skewed and CuBICA is performing better than Com4 and Com3. Notice that, the CF-OCOICA is performing better than Com3, Com4 and CuBICA algorithms in both noise-free and noisy scenarios. Finally notice that the improvement in ISR reduction achieved by the proposed estimator is consistent even though the free parameter β has been computed from whitened-sources.

7.7 Summary

This Chapter explored the combination of third- and fourth-order cumulant based tensor diagonalization in an optimal sense. A free parameter β is introduced in combining the third- and fourth-order cumulants and its optimal value is calculated such that the mean square estimation of Given's rotation is minimized. Computer simulation for

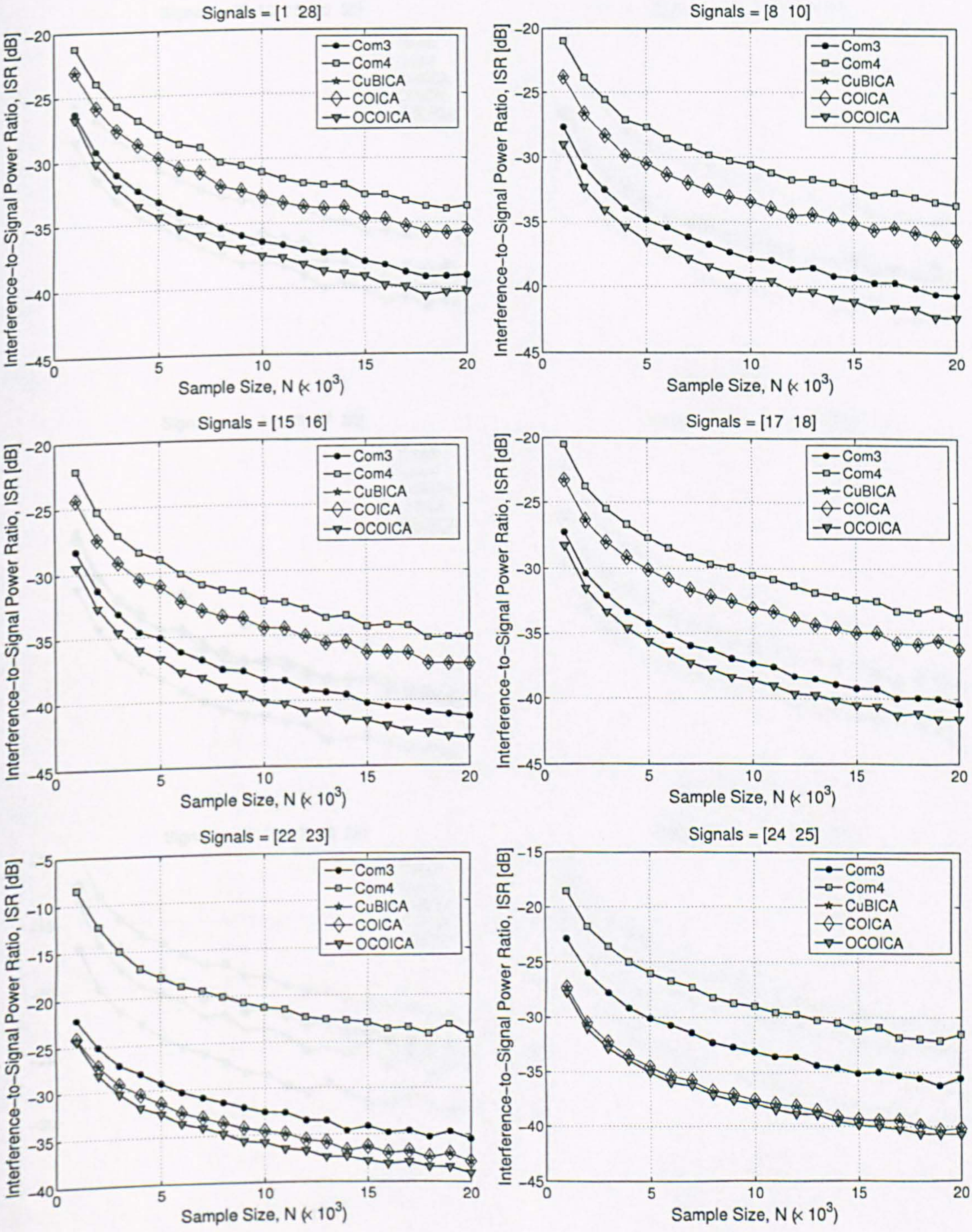


Figure 7.8: ISR performance for 2-source mixing scenario.

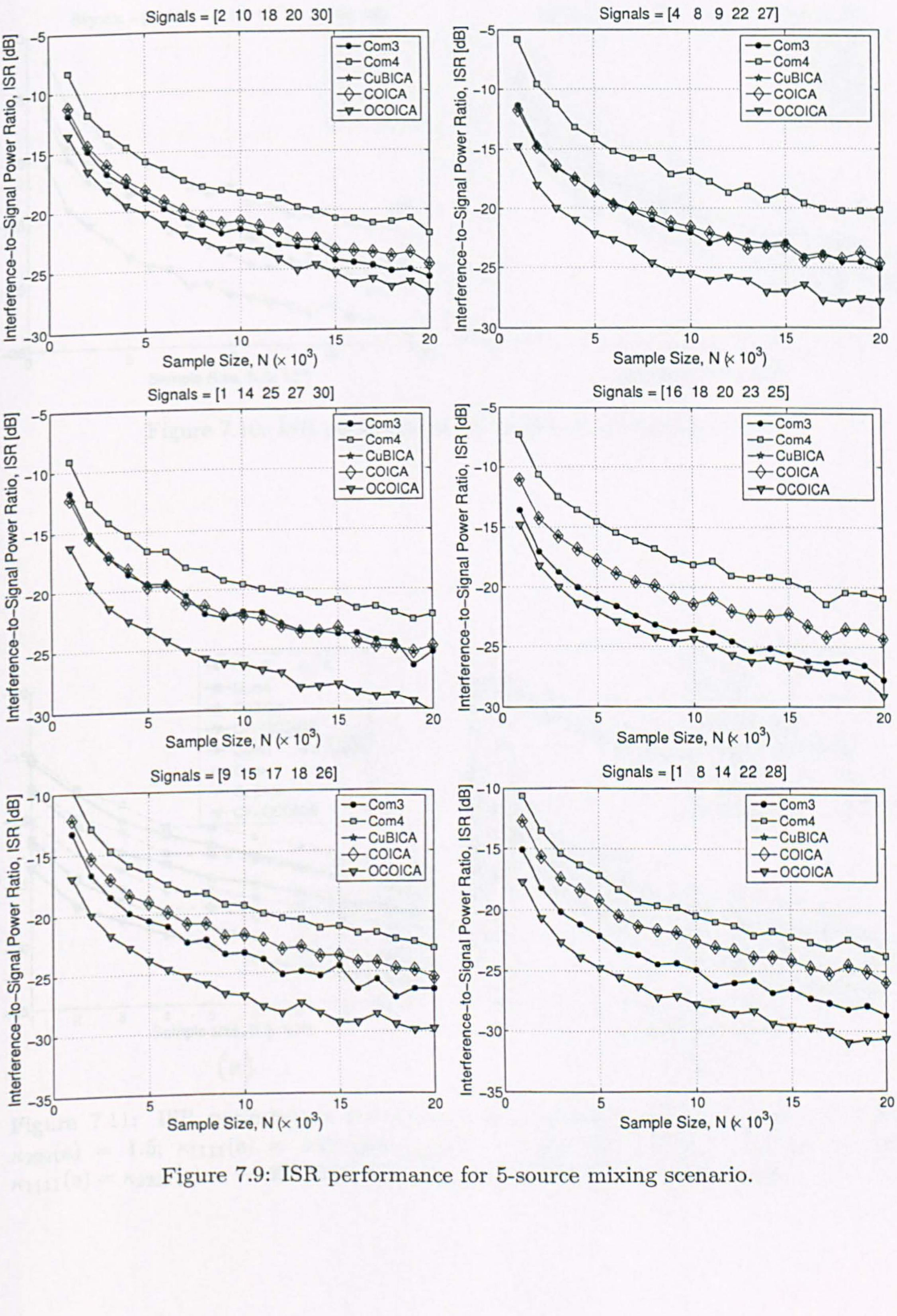


Figure 7.9: ISR performance for 5-source mixing scenario.

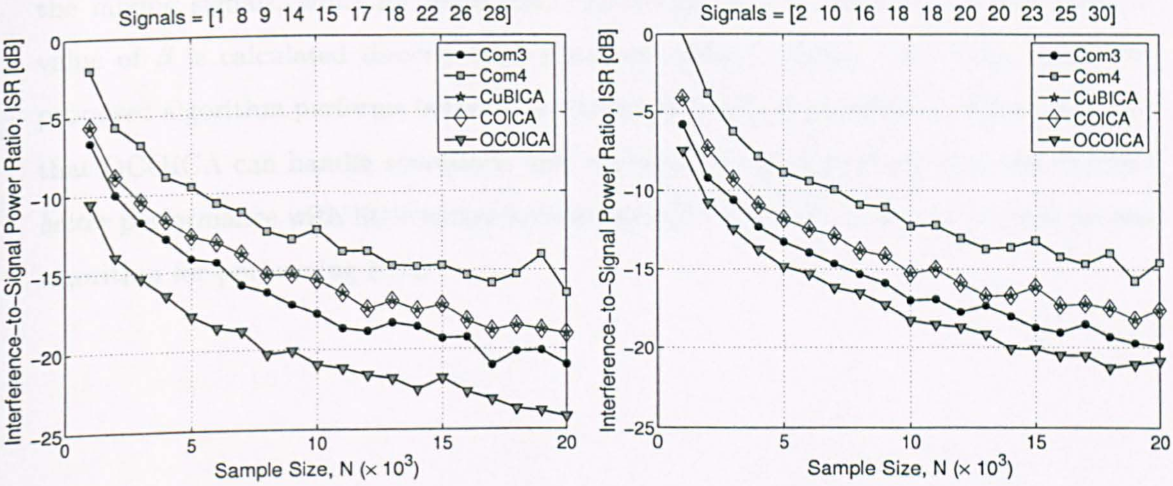
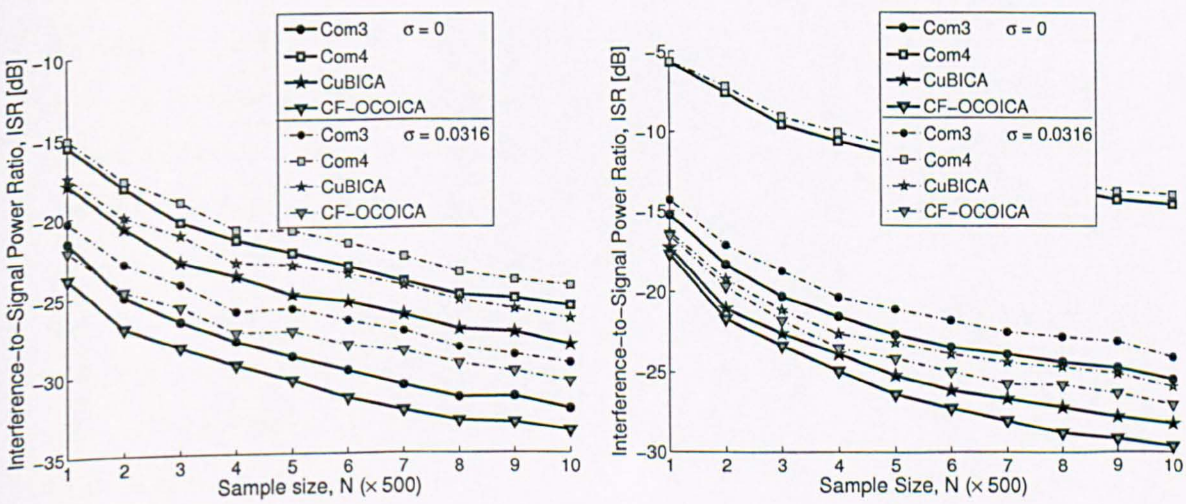


Figure 7.10: ISR performance for 10-source mixing scenario.



(a) (b)

Figure 7.11: ISR performance for 2-source mixing scenario. (a) $\kappa_{111}(\mathbf{s}) = -1.75$, $\kappa_{222}(\mathbf{s}) = 1.5$, $\kappa_{1111}(\mathbf{s}) = 3.75$ and $\kappa_{2222}(\mathbf{s}) = 3.5$; (b) $\kappa_{111}(\mathbf{s}) = -\kappa_{222}(\mathbf{s}) = 0.5$, $\kappa_{1111}(\mathbf{s}) = \kappa_{2222}(\mathbf{s}) = -0.25$. Solid lines, no noise; dashed lines, SNR=30dB.

the separation of two and more real and skewed sources is provided. For the case of two real-source, the optimal value of β is calculated from the *a priori* knowledge of the mixing signals, while for more than two-source mixing scenario, the approximate value of β is calculated directly from observed (mixed) signals. For both cases, the proposed algorithm performs better than three existing ICA algorithms. We have shown that OCOICA can handle *symmetric and asymmetric* distributed sources, and exhibits *better* performance with little excess computational overhead. It may be a good *general* algorithm for performing ICA.

Chapter 8

Conclusions and Outlook

8.1 Summary and Conclusions

This thesis has explored the problem of designing cost-functions and deriving associated adaptive algorithms for blind channel equalization (Chapter 2-6). It also explored the problem of obtaining iterative diagonalization of cumulant matrices for instantaneous blind source separation (Chapter 7).

For blind channel equalization, the thesis focused on cost-function based stochastic gradient-based adaptive algorithms for blind channel equalization as well as carrier-phase recovery in APSK/QAM communication system. Cost-function based algorithms are at the center of adaptive blind equalization and they implicitly incorporate higher-order statistics of signal and noise components. The choice of these cost-functions has a direct impact on the complexity and performance of the associated algorithms. In Chapter 2, the cost-function based blind equalization algorithms have been broadly classified into two groups: **1)** those which can only equalize and are insensitive to carrier-phase offset, and are termed as constant modulus algorithms (CMA), and **2)** those which can equalize and jointly remove the carrier-phase offset, and are termed as multimodulus algorithms (MMA). Next, the notions of a) **dispersion minimization** and b) **energy maximization** have been provided to design new cost-functions for blind channel equalization with/without carrier-phase recovery.

In Chapter 3, a new constant modulus algorithm, cCMA(p), has been presented for blind equalization of complex-valued communication channels. The proposed algorithm was obtained by solving a novel deterministic constrained optimization criterion, based on joint minimization of so-called *a priori* and *a posteriori* dispersion errors, leading to an update equation having a particular zero-memory continuous nonlinearity. The dispersion constant of cCMA(p) has been evaluated in closed-form with the consideration of convolutional noise (Section 3.3). This is a new result, because in the past, dispersion constants in CMA type algorithms have usually been evaluated and computed by assuming a noise-free scenario. Further, the stability of cCMA(p) has been studied and an easy-to-compute (generic) bound is derived for the range of step-sizes for which the proposed algorithm may be kept stable if initialized in the vicinity of zero-forcing solution (Section 3.4). An interesting theorem is provided in Section 3.5. According to that, by properly designing a range of step-size, the cCMA(p) may be kept converged such that $0 < |y_n| < R\sqrt{c}$ holds infinitely often, where R is dispersion constant and c is a constant which approaches unity for larger p. Further, the effect of free parameter p on the steady-state ISI performances has also been studied and it is shown that a metric in residual ISI is inversely proportional to p (Section 3.6). The implication of this result is that, for the given filter length, a lower residual ISI floor may be obtained by selecting a larger p and an appropriate step-size without sacrificing the convergence speed. Most of the theoretical results have been validated by computer simulations, for long and short equalizers and channels with APSK signaling (Section 3.7).

In Chapter 4, two new families of MMA algorithms, MMA(p, q) and cMMA(p), are presented for joint blind equalization and carrier-phase recovery of square-QAM signals over complex-valued transmission channel. The main contribution resided in the generalization of an existing dispersion-directed cost-function as well as the modification in a convex cost-function leading to newer algorithms capable of yielding faster convergence. Evaluation of dispersion constants and dynamic convergence has been described in detail and also shown to be in conformation with simulation results. Clearly, based on the results reported in this study, it is possible to obtain fast converging MMA equalizers (especially for 16/64-QAM). Also the discussed dynamic convergence analysis can help us select the best equalizer among the members of MMA(p, q) and cMMA(p) for the given channel, equalizer parameters (length, step-size, initialization), QAM-signal, noise condition and computational requirements. Finally, it is observed that MMA($p, 2$) and cMMA(p) are more robust to channel noise than MMA($p, 1$). Section 4.6 provided detailed simulation results using symbol-spaced as well as fractionally-space channels in the presence of additive noise.

Under the notion of energy maximization (Chapter 5 and 6), this thesis presented both CMA as well as MMA class of adaptive algorithms, where the energy of the equalized sequence is maximized subject to the restoration of some signal property (or properties). In Chapter 5, by exploiting the energy maximization principle, two new cost-functions have been proposed for blind channel equalization, which are respectively suitable for off-line block-processing and online gradient-based (constant modulus) implementations. These costs differ in the way, whether a block of equalizer outputs or only its instantaneous value is constrained from exceeding the largest modulus of data signal. An online adaptive algorithm (β-CMA) is obtained and has been shown to be performing better, in terms of ISI removal under the presence of noise for APSK signals, than existing established solutions like CMA(2,2) and three of its variants. It is also shown that the β-CMA is fully capable of recovering the true value of signal energy upon successful convergence. The computational complexity of β-CMA is noticeably much less than those of others addressed equalizers. The β-CMA may be considered as the first ever successful adaptive implementation of an l_2 optimization criterion for (constant modulus) blind channel equalization.

Similarly, in Chapter 6, a new adaptive multimodulus algorithm, β-MMA, is presented by exploiting energy maximization principle for joint blind equalization, carrier-phase recovery and energy restoration of square-QAM signals. The quadrature components of the equalized sequence are constrained not to exceed the largest real part of the transmitted signal. The parameter β is evaluated by considering convolutional noise at equalizer output. It is experimentally shown that the proposed β-MMA can yield better solution in terms of removing ISI and lower SER values under the presence of noise than existing established adaptive algorithms like MMA(1,2), CMA(2,2) and MMA(2,2). It is also shown that the β-MMA is fully capable of recovering the true value of signal energy upon successful convergence. Also note that the computational complexity of β-MMA is less than the existing addressed equalizers (like MMA(1,2), CMA(2,2) and MMA(2,2)). The β-MMA may be considered as the first ever successful adaptive implementation of an l_2 optimization criterion for joint blind channel equalization and carrier-phase recovery.

Finally, in Chapter 7, the optimal combination of third- and fourth-order cumulant based tensor diagonalization is studied for blind source separation. A free parameter β was introduced in combining the third- and fourth-order cumulants and its optimal value was calculated such that the mean square estimation of Given's rotation is minimized. Computer simulation for the separation of two and more real and skewed sources was provided. For the case of two real-source, the optimal value of β is calculated from the *a priori* knowledge of the mixing signals, while for more than two-source mixing scenario, the approximate value of β is calculated directly from observed (mixed) signals. For both cases, the proposed algorithm performs better than three existing ICA algorithms. A closed-form estimator was also obtained. The proposed algorithm can handle *symmetric and asymmetric* distributed sources, and exhibits *better* performance with little excess computational overhead; it may be a good *general* algorithm for performing ICA.

8.2 Future Suggestions and Outlook

The following is a list of possible points which could lead the continuation of the present investigations:

- The investigation of the equilibria of the cost-functions has been disregarded in this work. Such study is classified as static convergence analysis. The static convergence analysis of CMA(p,2) has been rigorously provided in [37]. Similar study can be carried out for our proposed algorithm cCMA(p).

- According to the analysis provided in [37], for a given channel and step-size, there is an optimum length for an equalizer to minimize the intersymbol interference. The results imply that a longer-length blind equalizer does not necessarily outperform a shorter one, as contrary to what is conventionally conjectured. Future work may investigate the determination of this length blindly.

- Timing error has been assumed perfect in this thesis. Future work may focus to updating the equalizer taps and timing offset jointly to minimize the mean dispersion and inter symbol interference.

- Some communication systems embed pilot tones into data spectrum to aid the receiver in synchronization, which may result in a DC offset of the basedband data [128]. Future work may focus on the estimation of the DC offset in CMA and MMA types of algorithms.

- In Chapter 5, the proposed cost (5.2.3) was not optimized. It may be optimized using quadratic programming by linearizing the constraint. It would be an important future topic to explore. Similar to the work of Meng *et al.* [90], the block-processing optimization of proposed cost (5.2.3) may be extended to other applications like blind source separation and blind beamforming.

- The cross-correlation and constant modulus algorithm has been proven to be an effective approach in the problem of joint blind equalization and source separation in a multi-input and multi-output system [85]. Future work may focus on studying the feasibility of this method for the proposed energy maximization based algorithms, β-CMA and β-MMA.

Bibliography

[1] *Signal processing information base (SPIB), Rice University*, [Online]. Available: http://spib.rice.edu/spib/microwave.html.

[2] S. Abrar, *Compact constellation algorithm for blind equalization of QAM signals*, Proc. IEEE Intl. Netw. Commun. Conf. (INCC) (2004), 170–174.

[3] ______, *Stop-and-go algorithms for blind channel equalization in QAM data communication systems*, Proc. Nat. Conf. on Emerg. Techn. (NCET) (2004), 170–174.

[4] S. Abrar and I. M. Qureshi, *Blind equalization of cross-QAM signals*, IEEE Sig. Process. Lett. **13** (2006), no. 12, 745–748.

[5] S. Abrar and R. A. Axford, Jr., *Sliced multi-modulus blind equalization algorithm*, ETRI Jnl. **27** (2005), no. 3, 257-266.

[6] S. Abrar and S. I. Shah, *New multimodulus blind equalization algorithm with relaxation*, IEEE Sig. Process. Lett. **13** (2006), no. 7, 425–428.

[7] S. Abrar, A. Zerguine, and M. Bettayeb, *Recursive least-squares backpropagation algorithm for stop-and-go decision-directed blind equalization*, IEEE Trans. Neural Networks **13** (2002), no. 6, 1472-1481.

[8] S. Abrar, A. Zerguine, and M. Deriche, *Soft constraint satisfaction multimodulus blind equalization algorithms*, IEEE Sig. Process. Lett. **12** (2005), no. 9, 637–640.

[9] J. B. Allen and J. E. Mazo, *Comparison of some cost functions for automatic equalization*, IEEE Trans. Commun. **21** (1973), no. 3, 233–237.

[10] ______, *A decision-free equalization scheme for minimum-phase channels*, IEEE Trans. Commun. **22** (1974), no. 10, 1732–1733.

[11] B. Baykal, O. Tanrikulu, A. G. Constantinides, and J. A. Chambers, *A new family of blind adaptive equalization algorithms*, IEEE Sig. Process. Lett. **3** (1999), no. 4, 109–110.

[12] M. Bellanger, *A simple comparison of constant modulus and wiener criteria for equalization with complex signals*, Dig. Sig. Process. **14** (2004), no. 5, 429–437.

[13] ______, *Which constant modulus criterion is better for blind adaptive filtering : CM(1,2) or CM(2,2)?*, Proc. IEEE Intl. Symp. Circuits and Systems (2005), IV–29:I–V32.

[14] ———, *Coefficient bias in constant modulus adaptive filters*, Proc. IEEE Intl. Symp. Circuits and Systems (2006), 1079–1082.

[15] ———, *On the performance of two constant modulus algorithms in equalization with non-CM signals*, Proc. IEEE Intl. Symp. Circuits and Systems (2007), 3475–3478.

[16] S. Bellini, *Bussgang techniques for blind equalization*, Proc. IEEE Global Commun. Conf. (GLOBECOM) (1986), 1634–1640.

[17] ———, *Bussgang techniques for blind deconvolution and equalization*, pp. 8–59, Blind deconvolution, S. Haykin (Ed.), Prentice Hall, 1994.

[18] F. Benedetto, G. Giunta, and L. Vandendorpe, *A blind equalization algorithm based on minimization of normalized variance for DS/CDMA communications*, IEEE Trans. Veh. Tech. **57** (2008), no. 6, 3453–3461.

[19] A. Benveniste and M. Goursat, *Blind equalizers*, IEEE Trans. Commun. **COM-32** (1984), 871–883.

[20] A. Benveniste, M. Goursat, and G. Ruget, *Robust identification of a noninimum phase system: blind adjustment of a linear equalizer in data communications*, IEEE Trans. Automatic Control **AC-25** (1980), no. 3, 385–389.

[21] T. Blaschke and L. Wiskott, *CuBICA: Independent component analysis by simultaneous third- and fourth-order cumulant diagonalization*, IEEE Trans. Sig. Process. **52** (2004), no. 5, 1250–1256.

[22] G. Bremer and R. Wachtel, *System for evaluating transmission line impairments*, US Patent 4503545 (1985).

[23] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky, *Relative optimization for blind deconvolution*, IEEE Trans. Sig. Process. **53** (2005), no. 6, 2018–2026.

[24] A. M. Bronstein, M. M. Bronstein, M. Zibulevsky, and Y. Y. Zeevi, *QML blind deconvolution: asymptotic analysis*, Proc. Intl. Conf. Ind. Comp. Anal. and Blind Sig. Sep., Lecture Notes in Comp. Science, Springer (2004), no. 3195, 677–684.

[25] C. R. Johnson, Jr., *Admissibility in blind adaptive channel equalization*, IEEE Cont. Syst. Mag. **1** (1991), 3–15.

[26] C. R. Johnson, Jr., P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, *Blind equalization using the constant modulus criterion: A review*, Proc. IEEE **86** (1998), 1927-1950.

[27] J. F. Cardoso, *High-order contrasts for independent compoent analysis*, Neural Computation **11** (1999), 157–192.

[28] F. R. P. Cavalcanti, A. L. Brandao, and J. M. T. Romano, *The generalized constant modulus algorithm applied to multiuser space-time equalization*, Proc. IEEE Work. Sign. Process. Adv. Wireless Commun. (SPAWC), Annapolis, MD (1999), 94–97.

[29] P. Chevalier, *On the performance of higher order blind source separation methods*, Proc. IEEE-ATHOS Workshop Higher-Order Statistics (1995), 30–34.

[30] S. Choi, R. Liu, and A. Cichocki, *A spurious equilibria-free learning algorithm for the blind separation of non-zero skewness signals*, Neural Process. Lett. **7** (1998), 61–68.

[31] P. Comon, *Separation of stochastic processes*, Proc. Workshop Higher-Order Spectral Analysis (1989), 174179.

[32] P. Comon, *Independent component analysis, a new concept?*, Sig. Process. **36** (1994), no. 3, 287314.

[33] P. Comon, *Tensor diagonalization, a useful tool in signal processing*, Proc. 10th IFAC Symp. Syst. Ident. (IFAC-SYSID) (1994), 77–82.

[34] P. Comon, *Contrasts, independent component analysis, and blind deconvolution*, Int. Jnl. Adapt. Cont. Sig. Process. **18** (2004), 225–243.

[35] P. Comon and J. F. Cardoso, *Eigenvalue decompostion of a cumulant tensor with applications*, SPIE: Adv. Sig. Process. Alg., Arch. and Impl. **1348** (1990), 361372.

[36] O. Dabeer and E. Masry, *Convergence analysis of the constant modulus algorithm*, IEEE Trans. Inf. Theory **49** (2003), no. 6, 14471464.

[37] Z. Ding and Y. Li, *Blind equalization and identification*, Marcel Dekker Inc., New York, 2001.

[38] Z. Ding and Z. Q. Luo, *A fast linear programming algorithm for blind equalization*, IEEE Trans. Commun. **48** (2000), no. 9, 1432–1436.

[39] D. Donoho, *On minimum entropy decopnvolution*, Proc. 2nd Appl. Time Series Symp. (1980), 565–608.

[40] G. Dziwoki, *An upper bound of the step size for the gradient constant modulus algorithm*, Proc. SPIE Intl. Soc. Opt. Eng.: Photonics Applications in Astronomy, Communications, Industry and High-Energy Physics Experiments IV **6159** (2006), (41–1 to 41–6).

[41] D. D. Falconer, *Jointly adaptive equalization and carrier-recovery in two-dimensional digital communication systems*, Bell Sys. Tech. Jnl. **55** (1976), no. 3, 317–334.

[42] C. P. Fan, W. H. Liang, and W. Lee, *Efficient fast blind equalization with two-stage single/multilevel modulus and DD algorithm for 64/256/1024-QAM wired cable communications*, Jnl. Chinese Institute of Engineers **32** (2009), no. 1, 1–15.

[43] G. Feyh and R. Klemt, *Blind equalizer based on autocorrelation lags*, Proc. IEEE Asilomar Conf. Sig. Sys. Comp. (ACSSC) (1990), 268–272.

[44] J. M. Filho, M. T. M. Silva, M. D. Miranda, and V. H. Nascimento, *A region-based algorithm for blind equalization of QAM signals*, Proc. IEEE Work. Stat. Sig. Process., Cardiff, Gales (2009), 685–688.

[45] S. Fiori, *A contribution to (neuromorphic) blind deconvolution by flexible approximated Bayesian estimation*, Sig. Process. **81** (2001), 2131–2153.

[46] B. Fisher and N. J. Bershad, *The complex LMS adaptive algorithm-transient weight mean and covariance with applications to the ALE*, IEEE Trans. Commun. **ASSP-31** (1983), no. 1, 34–44.

[47] A.I. Fleishman, *A method for simulating non-normal distributions*, Psychometrika **43** (1978), no. 4, 521–532.

[48] L. M. Garth, *A dynamic convergence analysis of blind equalization algorithms*, IEEE Trans. Commun. **49** (2001), no. 4, 624–634.

[49] L. M. Garth, J. Yang, and J. J. Werner, *Blind equalization algorithms for dual-mode CAP-QAM reception*, IEEE Trans. Commun. **49** (2001), no. 3, 455–466.

[50] M. Ghogho, A. Swami, and T. Durrani, *Approximate maximum likelihood blind source separation with arbitrary source PDFs*, Proc. IEEE Stat. Sig. and Array Process. (2000), 368372.

[51] M. Ghosh, *A signed-error algorithm for blind equalization of real signals*, Proc. IEEE Intl. Conf. Acoust., Speech, Sig. Process. (ICASSP) **6** (1998), 3365–3368.

[52] D. N. Godard and P. E. Thirion, *Method and device for training an adaptive equalizer by means of an unknown data signal in a QAM transmission system*, US Patent 4227152 (1980).

[53] D.N. Godard, *Self-recovering equalization and carrier tracking in two-dimensional data communications systems*, IEEE Trans. Commun. **COM-28** (1980), 1867–1875.

[54] A. Goupil and J. Palicot, *A geometrical derivation of the excess mean square error for Bussgang algorithms in a noiseless environment*, Sig. Process. **84** (2004), 311315.

[55] ———, *New algorithms for blind equalization: The constant norm algorithm family*, IEEE Trans. Sig. Process. **55** (2007), no. 4, 1436–1444.

[56] M. Gu and L. Tong, *Geometrical characterizations of constant modulus receivers*, IEEE Trans. Sig. Process. **47** (1999), no. 10, 2745–2756.

[57] F. Harroy and J. L. Lacoume, *Maximum likelihood estimators and Cramer-Rao bounds in source separation*, Sig. Process. **55** (1996), no. 2, 167–177.

[58] D. Hatzinakos, *Blind equalization using stop-and-go adaptation rules*, Optical Engineering **31** (1992), no. 6, 1181–1188.

[59] S. Haykin, *Blind deconvolution*, Prentice Hall, 1994.

[60] ———, *Adaptive filtering theory*, Prentice-Hall, 1996.

[61] G. Haza and R. Makowski, *Joint blind adaptive equalization based on shell partitioned multimodulus with soft switching and orthogonal basis for 256- and 1024-QAM*, Proc. Eur. Sig. Process. Conf. (EUSIPCO), Poznan, Poland (2007).

[62] F. Herrmann and A. Nandi, *Blind separation of linear instantaneous mixture using closed-forms estimators*, Sig. Process. **81** (2001), no. 7, 15371556.

[63] F. Herrmann and A. K. Nandi, *Maximisation of squared cumulants for blind source separation*, IEE Elec. Lett. **36** (2000), no. 19, 1664–1665.

[64] C. I Hwang and D. W. Lin, *A family of low-complexity blind equalizers*, IEEE Trans. Commun. **52** (2004), no. 3, 395–405.

[65] G. H. Im, C. J. Park, and H. C. Won, *A blind equalization with the sign algorithm for broadband access*, IEEE Commun. Lett. **5** (2001), no. 2, 70–72.

[66] D.W. Lin K. C. Hung and C. N. Ke, *Variable-step-size multimodulus blind decision-feedback equalization for high order QAM based on boundary MSE estimation*, Proc. IEEE Intl. Conf. Acoust., Speech, Sig. Process. (ICASSP) (2004).

[67] J. Karvanen and V. Koivunen, *Blind separation using absolute moments based adaptive estimating function*, Proc. 3rd Intl. Conf. Ind. Comp. Anal. and Sig. Sep. (ICA'2001) (2001), 218223.

[68] ______, *Independent component analysis via optimum combining of kurtosis and skewness-based criteria*, Journal of the Franklin Institute **341** (2004), 401–418.

[69] R. A. Kennedy and Z. Ding, *Blind adaptive equalizers for quadrature amplitude modulated communication systems based on convex cost functions*, Opt. Eng. **31** (1992), no. 6, 1189–1199.

[70] R. Kumar, *Convergence of a decision-directed adaptive equalizer*, Proc. 22nd IEEE Conf. Decision and Control **22** (1983), no. 1, 1319–1324.

[71] T. Kurakake, N. Nakamura, and K. Oyamada, *A 1024-QAM demodulator utilizing blind equalization*, IEIC Tech. Report (in Japanese) **103** (2004), no. 125, 55–60.

[72] O. W. Kwon, C. K. Un, and J. C. Lee, *Performance of constant modulus adaptive digital filters for interference cancellation*, Sig. Process. **26** (1992), 185–196.

[73] M. G. Larimore and J. R. Treichler, *Convergence behavior of the constant modulus algorithm*, Proc. IEEE Intl. Conf. Acoust., Speech, Sig. Process. (ICASSP) **1.4** (1983), 13–16.

[74] W. Lee and K. Cheun, *Convergence analysis of the stop-and-go blind equalization algorithm*, IEEE Trans. Commun. **47** (1999), no. 2, 177–180.

[75] X. L. Li and W. J. Zeng, *Performance analysis and adaptive Newton algorithms of multimodulus blind equalization criterion*, Sig. Process. **89** (2009), 2263–2273.

[76] X. L. Li and X. D. Zhang, *A family of generalized constant modulus algorithms for blind equalization*, IEEE Trans. Commun. **54** (2006), no. 11, 1913–1917.

[77] Y. Li and Z. Ding, *Convergence analysis of finite length blind adaptive equalizers*, IEEE Trans. Sig. Process. **43** (1995), no. 9, 2120–2129.

[78] W. G. Lim, *New soft transition dual mode type algorithms for blind equalization*, Proc. IEEE Wireless Commun. Networking Conf. (WCNC) (2007), 504508.

[79] B. Lin, R. He, X. Wang, and B. Wang, *The excess mean-square error analysis for Bussgang algorithm*, IEEE Sig. Process. Lett. **15** (2008), 793–796.

[80] L. Lin and B. Farhang-Boroujeny, *Convergence analysis of blind equalizer in a filter-bank-based multicarrier communication system*, IEEE Trans. Sig. Process. **54** (2006), no. 10, 4061–4067.

[81] F. Liu, L. D. Ge, and Y. J. Wu, *Dual-mode multi-modulus algorithms for blind equalisation of QAM signals*, European Trans. Telecommun. **19** (2008), no. 8, 917–921.

[82] R. López-Valcarce, *Cost minimisation interpretation of fourth power phase estimator and links to multimodulus algorithm*, Elect. Lett. **40** (2004), no. 4, 278.

[83] R. W. Lucky, *Automatic equalization for digital communications*, Bell Sys. Tech. Jnl. **44** (1965), 547–588.

[84] D. G. Luenberger, *Linear and nonlinear programming*, 2 ed., Addison-Wesley Publishing Company, London, 1984.

[85] Y. Luo and J. A. Chambers, *Steady-state mean-square error analysis of the cross-correlation and constant modulus algorithm in a MIMO convolutive system*, IEE Proc. Vision, Image and Sig. Process. **149** (2002), no. 4, 196–203.

[86] Z. Q. Luo, M. Meng, K. M. Wong, and J. K. Zhang, *A fractionally spaced blind equalizer based on linear programming*, IEEE Trans. Sig. Process. **50** (2002), no. 7, 1650–1660.

[87] O. Macchi and E. Eweda, *Convergence analysis of self-adaptive equalizers*, IEEE Trans. Info. Theory **IT-30** (1983), 162–176.

[88] J. E. Mazo, *Analysis of decision-directed equalizer convergence*, Bell Syst. Tech. Jnl. **59** (1980), no. 10, 1857–1876.

[89] P. McCullagh, *Tensor methods in statistics*, Chapman and Hall, London, 1987.

[90] C. Meng, J. Tuqan, and Z. Ding, *A quadratic programming approach to blind equalization and signal separation*, IEEE Trans. Sig. Process. **57** (2009), no. 6, 2232–2244.

[91] M. Miller, *Method and apparatus for determining inter-symbol interference for estimating data dependent jitter*, US Patent 7310392 (2007).

[92] M. Mizuno, J. Okello, and H. Ochi, *A high throughput pipelined architecture for blind adaptive equalizer with minimum latency*, IEICE Trans. Fund. Elect. Commun. Comp. Sc. **E86-A** (2003), no. 8, 2011–2019.

[93] E. Moreau, *A generalization of joint-diagonalization criteria for source separation*, IEEE Trans. Sig. Process. **49** (2001), no. 3, 530–541.

[94] E. Moreau and N. Thirion-Moreau, *Nonsymmetrical contrasts for sources separation*, IEEE Trans. Sig. Process. **47** (1999), no. 8, 2241–2252.

[95] J. J. Mulligan and E. H. Satorius, *Algorithm for equalizing rectangular signal constellations*, Tech. Rep. NPO-20324, NASA, Jet Propulsion Laboratory, California Institute of Technology, California **23** (1999), no. 3, 422–424.

[96] J. J. Murillo-Fuentes and F. J. Gonzalez-Serrano, *A sinusoidal contrast function for the blind separation of statistically independent sources*, IEEE Trans. Sig. Process. **52** (2004), no. 12, 3459–3463.

[97] J. J. Murillo-Fuentes and F. J. Gonzlez-Serrano, *A sinusoidal contrast function for the blind separation of statistically independent sources*, IEEE Trans. Sig. Process. **52** (2004), no. 12, 34593463.

[98] A. K. Nandi, *Personal notes*, 2000.

[99] V. H. Nascimento and M. T. M. Silva, *Stochastic stability analysis for the constant-modulus algorithm*, IEEE Trans. Sig. Process. **56** (2008), no. 10, 4984–4989.

[100] C. W. Niessen, *Automatic channel equalization algorithm*, Proc. IEEE **55** (1967), no. 5, 698–698.

[101] C. L. Nikias and A. P. Petropulu, *Higher-order spectra analysis a nonlinear signal processing framework*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[102] K. N. Oh and Y. O. Chin, *Modified constant modulus algorithm: Blind equalization and carrier phase recovery algorithm*, Proc. IEEE Int. Conf. Commun. (ICC) **1** (1995), 498–502.

[103] ———, *New blind equalization techniques based on constant modulus algorithm*, Proc. IEEE Global Telecommun. **2** (1995), 865–869.

[104] C. B. Papadias and D. T. M. Slock, *Normalized sliding window constant modulus and decision-directed algorithms: A link between blind equalization and classical adaptive filtering*, IEEE Trans. Sig. Process. **45** (1997), no. 1, 231–235.

[105] G. Picchi and G. Prati, *Blind equalization and carrier recovery using a 'stop-and-go' decision-directed algorithm*, IEEE Trans. Commun. **COM-35** (1987), 877–887.

[106] M. Pinchas, *A closed approximated formed expression for the achievable residual inter symbol interference obtained by blind equalizers*, Sig. Process. **90** (2010), 1940–1962.

[107] M. Pinchas and B. Z. Bobrovsky, *A novel hos approach for blind channel equalization*, IEEE Trans. Wireless Commun. **6** (2007), no. 3, 875–886.

[108] W. H. Press, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes in c. the art of scientific computing*, Cambridge Univ. Press, 1992.

[109] J. G. Proakis, *Digital Communications*, 3 ed., New York: McGraw-Hill, 1995.

[110] R. A. Axford, Jr., *Refined techniques for blind equalization of phase shift keyed (PSK) and quadrature amplitude modulated (QAM) digital communications signals*, Ph.D. thesis, Univ. of California, San Diego, 1995.

[111] M. Rupp and S. C. Douglas, *A posteriori analysis of adaptive blind equalizers*, Proc. IEEE Asilomar Conf. Sig. Sys. Comp. (ACSSC) **1** (1998), 369–373.

[112] M. Rupp and A. H. Sayed, *On the convergence analysis of blind adaptive equalizers for constant modulus signals*, IEEE Trans. Commun. **48** (2000), no. 5, 795–803.

[113] W. Rupprecht, *Orthogonalfilter und adaptive daten-signalentzerrung*, R. Oldenbourg Verlag München Wien, 1987.

[114] Y. Sato, *A method of self-recovering equalization for multilevel amplitude modulation systems*, IEEE Trans. Commun. **COM-23** (1975), 679–682.

[115] E. H. Satorius and J. J. Mulligan, *Minimum entropy deconvolution and blind equalisation*, Elect. Lett. **28** (1992), no. 16, 1534–1535.

[116] E. H. Satorius and J. J. Mulligan, *An alternative methodology for blind equalization*, Dig. Sig. Process.: A Rev. Jnl. **3** (1993), no. 3, 199–209.

[117] E. H. Satorius and J. J. Mulligan, *Asymptotic analysis of scale-invariant cost functions for blind adaptive processing*, Proc. IEEE Asilomar Conf. Sig. Sys. Comp. (ACSSC) **2** (1994), 1468–1472.

[118] J. Serra and N. Esteves, *A blind equalization algorithm without decision*, Proc. IEEE Intl. Conf. Acoust., Speech, and Sig. Process. (ICASSP) **9** (1984), no. 1, 475–478.

[119] W. A. Sethares, R. A. Kennedy, and Z. Gu, *An approach to blind equalization of non-minimum phase systems*, Proc. IEEE Intl. Conf. Acoust., Speech, and Sig. Process. (ICASSP), Toronto, Ont., Canada (1991), 1529–1532.

[120] W. A. Sethares, G. A. Rey, and C. R. Johnson, Jr., *Approach to blind equalization of signal with multiple modulus*, Proc. IEEE Intl. Conf. Acoust., Speech, Sig. Process. (ICASSP) (1989), 972–975.

[121] O. Shalvi and E. Weinstein, *New criteria for blind deconvolution of nonminimum phase systems (channels)*, IEEE Trans. Inf. Theory **36** (1990), no. 2, 312–321.

[122] V. Shtrom and H. H. Fan, *New class of zero-forcing cost functions in blind equalization*, IEEE Trans. Sig. Process. **46** (1998), no. 10, 2674.

[123] J. J. Shynk and C. K. Chan, *Performance surfaces of the constant modulus algorithm based on a conditional gaussian model*, IEEE Trans. Sig. Process. **41** (1993), no. 5, 1965–1969.

[124] B. Stoll and E. Moreau, *A generalized ICA algorithm*, IEEE Sig. Process. Lett. **7** (2000), no. 4, 90–92.

[125] O. Tanrikulu, A. G. Constantinides, and J. A. Chambers, *New normalized constant modulus algorithms with relaxation*, IEEE Sig. Process. Lett. **4** (1997), no. 9, 256–258.

[126] T. Thaiupathump, *New algorithms for blind equalization and blind source separation/phase recovery*, Ph.D. thesis, Univ. of Pennsylvania, 2002.

[127] T. Thaiupathump, L. He, and S. A. Kassam, *Square contour algorithm for blind equalization of QAM signals*, Sig. Process. **86** (2006), no. 11, 3357–3370.

[128] A. Touzni and T. Endres, *Behavior and corrections of constant modulus equalization with a DC offset*, Proc. Eur. Sig. Process. Conf. (EUSIPCO) **III** (2002), 327–330.

[129] J. R. Treichler and B. G. Agee, *A new approach to multipath correction of constant modulus signals*, IEEE Trans. Acoust., Speech, Sig. Process. **ASSP-31** (1983), 459–471.

[130] J. R. Treichler and M. G. Larimore, *New processing techniques based on the constant modulus adaptive algorithm*, IEEE Trans. Acoust., Speech, Sig. Process. **ASSP-33** (1985), no. 2, 420–431.

[131] J. R. Treichler, M. G. Larimore, and J. C. Harp, *Practical blind demodulators for high-order QAM signals*, Proceeding of the IEEE **86** (1998), no. 10, 1907–1926.

[132] S. Vembu, S. Verdu, R. Kennedy, and W. Sethares, *Convex cost functions in blind equalization*, IEEE Trans. Sig. Process. **42** (1994), no. 8, 1952–1959.

[133] V. Weerackody and S. A. Kassam, *A simple hard-limited adaptive algorithm for blind equalization*, IEEE Trans. Circt. Syst.-II **39** (1992), no. 7, 482–487.

[134] V. Weerackody, S. A. Kassam, and K. R. Laker, *Convergence analysis of an algorithm for blind equalization*, IEEE Trans. Commun. **39** (1991), no. 6, 856–865.

[135] ———, *Sign algorithms for blind equalization and their convergence analysis*, Jnl. Circt., Syst., Sig. Process. **10** (1991), no. 4, 393–431.

[136] K. Wesolowski, *Self-recovering adaptive equalization algorithms for digital radio and voiceband data modems*, Proc. Eur. Conf. Circuit Theory and Design (1987), 19–24.

[137] ———, *Analysis and properties of the modified constant modulus algorithm for blind equalization*, European Trans. Telecommun. **3** (1992), no. 3, 225–230.

[138] J. Yang, *Multimodulus algorithms for blind equalization*, Ph.D. thesis, University of British Columbia, Aug. 1997.

[139] J. Yang, J. J. Werner, and G. A. Dumont, *The multimodulus blind equalization algorithm*, Proc. IEEE Intl. Conf. Dig. Sig. Process. (DSP) **1** (1997), 127–130.

[140] ———, *The multimodulus blind equalization and its generalized algorithms*, IEEE Jr. Sel. Areas Commun. **20** (2002), no. 5, 997–1015.

[141] J. Yang, J. J. Werner, D. D. Harman, and G. A. Dumont, *Blind equalization for broadband access*, IEEE Commun. Magazine **37** (1999), no. 4, 87–93.

[142] N. R. Yousef and A. H. Sayed, *A unified approach to the steady-state and tracking analyses of adaptive filters*, IEEE Trans. Sig. Process. **49** (2006), no. 2, 314–324.

[143] J. T. Yuan and K. D. Tsai, *Analysis of the multimodulus blind equalization algorithm in QAM communication systems*, IEEE Trans. Commun. **53** (2005), no. 9, 1427–1431.

[144] V. Zarzoso, *Closed-form higher-order estimators for blind separation of independent source signals in instantaneous linear mixtures*, Ph.D. thesis, Univ. of Liverpool, Liverpool, UK, Oct. 1999.

[145] V. Zarzoso, J. J. Murillo-Fuentes, R. Boloix-Tortosa, and A. K. Nandi, *Optimal pairwise fourth-order independent component analysis*, IEEE Trans. Sig. Process. **54** (2006), no. 8, 3049–3063.

[146] V. Zarzoso, J.J. Murillo-Fuentes, R. Boloix-Tortosa, and A.K. Nandi, *Optimal pairwise fourth-order independent component analysis*, IEEE Trans. Sig. Process. **54** (2006), no. 8, 3049–3063.

[147] V. Zarzoso and A. K. Nandi, *Blind separation of independent sources for virtually any source probability density function*, IEEE Trans. Sig. Process. **47** (1999), no. 9, 2419–2432.

[148] V. Zarzoso, A. K. Nandi, F. Herrmann, and J. Millet-Roig, *Combined estimation scheme for blind source separation with arbitrary source PDFs*, Electron. Lett. **37** (2001), no. 2, 132–133.

[149] H. H. Zeng, L. Tong, and C. R. Johnson, Jr., *Relationships between the constant modulus and Wiener receivers*, IEEE Trans. Inf. Theory **44** (1998), no. 4, 1523–1538.

[150] H. H. Zeng, L. Tong, and C. R. Johnson, Jr., *An analysis of constant modulus receivers*, IEEE Trans. Sig. Process. **47** (1999), no. 11, 2990–2999.

[151] C. Ziegaus and E. W. Lang, *Statistics of natural and urban images*, Lecture Notes in Computer Science **1327** (1997), 219224.