Signal Processing 101 (2014) 74–86

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# Newton-like minimum entropy equalization algorithm for APSK systems

Anum Ali [a,*], Shafayat Abrar [b], Azzedine Zerguine [a], Asoke K. Nandi [c]

[a] King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia
[b] COMSATS Institute of Information Technology, Islamabad, Pakistan
[c] Brunel University, Uxbridge, Middlesex UB8 3PH, UK

ARTICLE INFO

*Article history:*
Received 6 August 2013
Received in revised form
29 January 2014
Accepted 5 February 2014
Available online 13 February 2014

*Keywords:*
Constant modulus algorithm
Blind equalizer
Recursive least squares algorithm
Newton's method
Tracking performance
Amplitude phase shift keying

ABSTRACT

In this paper, we design and analyze a Newton-like blind equalization algorithm for the APSK system. Specifically, we exploit the principle of minimum entropy deconvolution and derive a blind equalization cost function for APSK signals and optimize it using Newton's method. We study and evaluate the steady-state excess mean square error performance of the proposed algorithm using the concept of energy conservation. Numerical results depict a significant performance enhancement for the proposed scheme over well established blind equalization algorithms. Further, the analytical excess mean square error of the proposed algorithm is verified with computer simulations and is found to be in good conformation.



## 1. Introduction

In digital communications at high enough data rates, almost all physical channels exhibit inter-symbol interference (ISI). One of the solutions to this problem is blind equalization, which is a method to equalize distortive communication channels and mitigate ISI without supervision. Blind equalization algorithms do not require training at either the startup period or restart after system breakdown. This independence of blind schemes with respect to training sequence results in improved system bandwidth efficiency.

In blind equalization, the desired signal is unknown to the receiver, except for its probabilistic or statistical properties over some known alphabets. As both the channel and its input are unknown, the objective of blind equalization is to recover the unknown input sequence based solely on its probabilistic and statistical properties [1–3]. From the available literature, it can be found that any admissible blind objective (or cost) function has two main attributes: (1) it makes use of the statistics which are significantly modified as the signal propagates through the channel [4] and (2) optimization of the cost function modifies the statistics of the signal at the equalizer output, aligning them with the statistics of the signal at the channel input [5].

One of the earliest methods of blind equalization was suggested by Benveniste et al. [5]. Their proposed method assumed the transmitted signal to be non-Gaussian, independent and identically distributed (i.i.d.) sequence of a known statistical distribution. It sought to match the distribution of its output (deconvolved sequence) with the distribution of the transmitted signal and the adaptation continued until the said objective was achieved. Another approach to this problem was devised by Donoho [4], who defined a partial ordering, measuring the relative *Gaussianity*

* Corresponding author. Tel.: +966 59 779 4592.
*E-mail addresses:* anumali@ieee.org (A. Ali), sabrar@comsats.edu.pk (S. Abrar), azzedine@kfupm.edu.sa (A. Zerguine), asoke.nandi@brunel.ac.uk (A.K. Nandi).

http://dx.doi.org/10.1016/j.sigpro.2014.02.003
0165-1684 

between random variables. He suggested to adjust the equalizer until the distribution of the deconvolved sequence is as *non-Gaussian* as possible. A somewhat informal version of Donoho's method appeared in the work of Wiggins [6,7]. According to Wiggins, assuming the transmitted signal is a non-Gaussian signal with certain distribution $p_a$, the equalizer must be adjusted to make its output signal distribution resemble $p_a$. He termed this approach as a *minimum entropy deconvolution* (MED) criterion.

In this work, the MED criterion is the subject of our concern for designing (and optimizing) cost functions to equalize signals blindly in amplitude phase shift keying (APSK) systems. The APSK signals are very important in modern day communication systems due to their robustness against nonlinear channel distortion and advantageous lower peak-to-average power ratio compared to the conventional quadrature amplitude modulation signals (refer to APSK based systems in [8–14] and references therein).

This paper is organized as follows: Section 2 discusses the baseband communication system model, notion of Gaussianity, and traditional blind equalizers. Section 3 describes the MED criteria for channel equalization, discusses the admissibility of costs tailored for APSK, and stochastic gradient-based optimization. Section 4 formulates the proposed adaptive MED-based blind equalization scheme for APSK systems exploiting Newton-like update. Section 5 provides a steady-state tracking performance of the proposed algorithm in time varying scenario. Simulation results are discussed in Section 6 and conclusions are provided in Section 7.

## 2. System model and traditional blind equalizers

The baseband model for a typical complex-valued data communication system, as shown in Fig. 1, consists of an unknown finite-impulse response filter $\boldsymbol{h}_n$, which represents the physical inter-connection between the transmitter and the receiver. A zero-mean, i.i.d., circularly symmetric, complex-valued data sequence $\{a_n\}$ is transmitted through the channel, whose output $x_n$ is recorded by the receiver. The input/output relationship of the channel can be written as $x_n=\sum_k a_{n-k}h_k+\nu_n$, where the additive noise $\nu_n$ is assumed to be stationary, Gaussian, and independent of the channel input $a_n$. The function of equalizer at the receiver is to estimate the delayed version of original data, $a_{n-\tau}$, from the received signal $x_n$. Let $\boldsymbol{w}_n=[w_{n,0}, w_{n,1}, \ldots, w_{n,N-1}]^T$ be a vector of equalizer coefficients with $N$ elements and $\boldsymbol{x}_n=[x_n, x_{n-1}, \ldots, x_{n-N+1}]^T$ be the vector of channel observations ($T$ denotes transpose operation). The output of the equalizer is then given by $y_n=\boldsymbol{w}_{n-1}^H\boldsymbol{x}_n$ ($H$ denotes the Hermitian conjugate operator). If $\boldsymbol{t}_n=\boldsymbol{h}_n\otimes\boldsymbol{w}_{n-1}^*$ represents the overall channel-equalizer impulse response ($\otimes$ denotes convolution), then

$$y_n=\sum_i w_{n-1,i}^* x_{n-i}=\underbrace{t_{n,\tau}a_{n-\tau}+\sum_{l\neq\tau}t_{n,l}a_{n-l}+\nu_n'}_{\text{signal}+\text{ISI}+\text{noise}} \tag{1}$$

which demonstrates the adverse effect of inter-symbol interference (ISI) and additive noise. The ISI is quantified as [15]:

$$\text{ISI}=\frac{\sum_i|t_{n,i}|^2-\max\{|\boldsymbol{t}_n|^2\}}{\max\{|\boldsymbol{t}_n|^2\}} \tag{2}$$

In subsequent discussions we drop the subscript $n$ from $t$ for notational convenience. The idea behind a *Bussgang* blind equalizer is to minimize (or maximize), through the choice of $\boldsymbol{w}$, a certain cost function $J$ depending on $y_n$, such that $y_n$ provides an estimate of $a_n$ up to some inherent indeterminacies, giving, $y_n=\alpha a_{n-\tau}$, where $\alpha=|\alpha|e^{j\theta}\in\mathbb{C}$ represents an arbitrary complex-valued gain. Hence, a Bussgang blind equalizer tries to solve the following optimization problem:

$$\boldsymbol{w}^\dagger=\arg\underset{\boldsymbol{w}}{\text{optimize}}\,J\quad\text{with } J=\mathsf{E}\mathcal{J}(y_n) \tag{3}$$

The cost $J$ is a function of implicitly embedded statistics of $y_n$ and $\mathcal{J}(\cdot)$ 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 a channel input. The equalization is accomplished when equalized sequence $y_n$ acquires an identical distribution as that of the channel input $a_n$. More formally, we have the following theorem [5]:

**Theorem 1.** *If the transmitted signal is composed of non-Gaussian i.i.d. samples, both the channel and the equalizer are linear time-invariant filters, noise is negligible, and the probability density functions (PDF) of transmitted and equalized signals are equal, then the channel has been perfectly equalized.*

This mathematical result is very important since it establishes the possibility of obtaining an equalizer with the sole aid of signal's statistical properties and without requiring any knowledge of the channel impulse response or training data sequence. Meanwhile, Donoho [4] noted that, as a consequence of the central limit theorem, linear combinations of identically distributed random variables become *more Gaussian* than the individual variables. Therefore, the received signal $x_n$ will have a distribution that is more nearly Gaussian than the distribution of $a_n$. Any suitable objective function capable of measuring *Gaussianity* or *non-Gaussianity* can therefore be used for deconvolution.

One of the measures of Gaussianity is (normalized) kurtosis, $\kappa$, which is a statistic based on second and fourth-order moments. For a circularly-symmetric complex-valued random variable $X$, kurtosis is defined as [15]:

$$\kappa_X=\frac{\mathsf{E}|x|^4}{(\mathsf{E}|x|^2)^2}-2 \tag{4}$$

$\kappa_X$ is greater than zero, equal to zero, and less than zero for super-Gaussian, Gaussian, and sub-Gaussian random

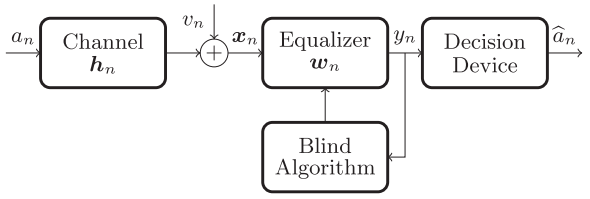


**Fig. 1.** A typical baseband communication system.

variables, respectively. Expectedly most of the existing blind equalizers use second and fourth-order statistics of the equalized sequence. The first ever method which relies on the aforesaid statistics is the constant modulus criterion [16]; it is given by

$$\min_{\boldsymbol{w}}\{\mathsf{E}|y_n|^4-2R^2\mathsf{E}|y_n|^2\}, \tag{5}$$

where $R$ is the *Godard radius* and $R^2=\mathsf{E}|a_n|^4/\mathsf{E}|a_n|^2$ is the *Godard dispersion constant* [17]. For an input signal that has a constant modulus $|a_n|=R$, the constant modulus criterion penalizes output samples $y_n$ that do not have the desired constant modulus characteristics [18].

Shalvi and Weinstein [15] demonstrated that the condition of equality between the PDF of the transmitted and equalized signals, due to Theorem 1, was excessively tight. Under the similar assumptions, as laid in Theorem 1, they discussed the possibility to perform blind equalization by satisfying the condition $\mathsf{E}|y_n|^2=\mathsf{E}|a_n|^2$ and ensuring that a nonzero cumulant (of any order higher than 2) of $a_n$ and $y_n$ are equal. For a two dimensional signal $a_n$ with four-quadrant symmetry (i.e., $\mathsf{E}a_n^2=0$), they suggested to maximize the following exemplary cost containing second and fourth-order statistics:

$$\max_{\boldsymbol{w}}|\mathsf{E}|y_n|^4-2(\mathsf{E}|y_n|^2)^2| \quad \text{s.t. } \mathsf{E}|y_n|^2=\mathsf{E}|a_n|^2 \tag{6}$$

In the next section, we discuss equalization/deconvolution techniques which evolved around the notion of *entropy* as a measure of Gaussianity.

## 3. Minimum entropy deconvolution

In statistics, the most powerful test of the null hypothesis against another is one which maximizes discrimination and minimizes the probability of accepting the null hypothesis as true when actually the alternative is true. In information theory, a statistic that maximizes detectability of a signal while minimizing the probability of a false alarm is most powerful. In seismic community, however, a statistical (rank-discriminating) test against Gaussianity, which can be used to design a deconvolution filter, is loosely termed as a MED criterion [4,6].

Hogg [19] developed a scale-invariant and the most powerful rank-discriminating test for one member of the generalized Gaussian against another by considering the following PDF:

$$p_Y(y,\alpha)=\frac{\alpha}{2\Gamma\left(\frac{1}{\alpha}\right)}\exp(-|y|^{\alpha}), \quad |y|<\infty \tag{7}$$

where $\alpha>0$ is shape parameter, and $\Gamma(\cdot)$ is the Gamma function. To determine if a random set of samples $\{y_1,y_2,\ldots,y_B\}$ is drawn from the distribution $p_Y(y,\alpha_2)$ as opposed to $p_Y(y,\alpha_1)$, a ratio test was derived (based on the procedure described in [20]) as follows:

$$V_Y(\alpha_1,\alpha_2)=\frac{\left(\frac{1}{B}\sum_{i=1}^{B}|y_i|^{\alpha_1}\right)^{B/\alpha_1}}{\left(\frac{1}{B}\sum_{i=1}^{B}|y_i|^{\alpha_2}\right)^{B/\alpha_2}}\underset{\alpha=\alpha_1}{\overset{\alpha=\alpha_2}{\gtrless}}\chi \tag{8}$$

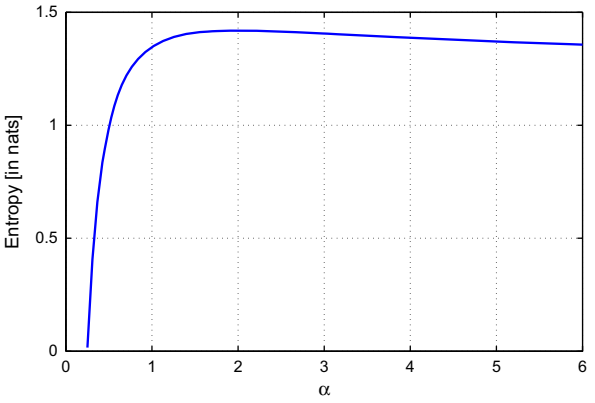


**Fig. 2.** Entropy $H(\alpha)$ versus shape parameter $\alpha$. For $\alpha=2$, entropy is maximum and its value is $\log(\sqrt{2\pi e})$.

where $\chi$ is some threshold. The larger $V$ becomes, the more probable it is that the sample set $Y$ is drawn from the distribution $p_Y(y,\alpha_2)$ and not $p_Y(y,\alpha_1)$.

It would be interesting to look at the entropy, $H(\alpha)$, associated with (7); it is given as follows [21]:

$$H(\alpha)=\log\left(\frac{2}{\alpha}\sqrt{\Gamma\left(\frac{1}{\alpha}\right)^3\Big/\Gamma\left(\frac{3}{\alpha}\right)^2}\right)+\frac{1}{\alpha} \tag{9}$$

This function is illustrated in Fig. 2 for different members of the family. The entropy is maximum for Gaussian (when $\alpha=2$). Note that ($\alpha<2$) and ($\alpha>2$) represent super-Gaussian and sub-Gaussian cases, respectively. As $\alpha\to 0$, $H(\alpha)$ rapidly goes to $-\infty$ which is the entropy of the certain event; however, on the other side, $H(\alpha)$ falls off slowly to another minimum, which is the entropy of the uniform distribution.

From the work of Geary [22], it became known to the statistical community that the test against Gaussianity

$$\frac{\frac{1}{B}\sum_{i=1}^{B}|x_i|^4}{\left(\frac{1}{B}\sum_{i=1}^{B}|x_i|^2\right)^2}$$

is most efficient when no information on the distribution of the random sample is available. Remarkably, Wiggins exploited the same idea and sought to determine the inverse channel $\boldsymbol{w}^{\dagger}$ that maximizes the *kurtosis* of the deconvolved seismic data $y_n$ [6]. Since seismic data are super-Gaussian in nature, given $B$ samples of $y_n$, Wiggins suggested to maximize the following test (or cost):

$$\mathcal{J}_{\mathrm{W}}(y_n)=\frac{\frac{1}{B}\sum_{n=1}^{B}|y_n|^4}{\left(\frac{1}{B}\sum_{n=1}^{B}|y_n|^2\right)^2}\xrightarrow{\text{large } B}\frac{\mathsf{E}|y_n|^4}{(\mathsf{E}|y_n|^2)^2} \tag{10}$$

This scheme seeks the smallest number of large spikes (or impulses) consistent with the data, thus maximizing the order, or equivalently, minimizing the *entropy* or disorder in the data [23]. Wiggins coined the term MED criterion for the test (10).[1] Immediately after Wiggins, Ooe and

[1] The simple structure of the impulsive function led Wiggins to call his technique MED. According to [24], however, the approach which Wiggins actually adopts is not entropy; it is rather a variant of *varimax*

Ulrych [24] realized that better deconvolution results may be obtained for super-Gaussian signals if non-Gaussianity is maximized with some $\alpha$ less than two; they used ($\alpha=1$) and suggested

$$\mathcal{J}_{\mathrm{OU}}(y_n)=\frac{\frac{1}{B}\sum_{n=1}^{B}|y_n|^2}{\left(\frac{1}{B}\sum_{n=1}^{B}|y_n|\right)^2}\xrightarrow{\text{large } B}\frac{\mathsf{E}|y_n|^2}{(\mathsf{E}|y_n|)^2} \tag{11}$$

Later, Gray presented a generic MED criterion with two degrees of freedom [25]:

$$\mathcal{J}_{\mathrm{G}}(y_n)=\frac{\frac{1}{B}\sum_{n=1}^{B}|y_n|^p}{\left(\frac{1}{B}\sum_{n=1}^{B}|y_n|^q\right)^{p/q}}\xrightarrow{\text{large } B}\frac{\mathsf{E}|y_n|^p}{(\mathsf{E}|y_n|^q)^{p/q}} \tag{12}$$

Note that costs (10)–(12) are members of (8) for different values of shape parameters. In the context of digital communication, where the underlying distribution of the transmitted (possibly pulse amplitude modulated) data symbols is closer to a uniform density (sub-Gaussian), we can obtain a blind equalizer by optimizing Gray's cost (12) as follows [26,27]:

$$\boldsymbol{w}^{\dagger}=\max_{\boldsymbol{w}}\mathcal{J}_{\mathrm{G}}(y_n),\quad \text{for } p=2 \text{ and } q>2. \tag{13}$$

Recently, Abrar and Nandi [28] discussed the case $(p,q)=(2,\infty)$ for the blind equalization of the APSK signal:

$$\mathcal{J}_{\mathrm{AN}}(y_n)=\frac{\frac{1}{B}\sum_{n=1}^{B}|y_n|^2}{(\max\{|y_n|\})^2}\xrightarrow{\text{large } B}\frac{\mathsf{E}|y_n|^2}{(\max\{|y_n|\})^2} \tag{14}$$

Maximizing (14) can be interpreted as determining the equalizer coefficients, $\boldsymbol{w}$, which drives the distribution of its output, $y_n$, away from Gaussian distribution toward uniform, thus removing successfully the interference from the received signal. Note that the cost (14) is an optimal, scale-invariant test for the APSK signal against Gaussianity (refer to Appendix A for details).

### 3.1. Admissibility of the cost $\mathcal{J}_{\mathrm{AN}}(y_n)$

In this section, we discuss the admissibility of $\mathcal{J}_{\mathrm{AN}}(y_n)$ (14). By admissibility, we mean that the cost $\mathcal{J}_{\mathrm{AN}}(y_n)$ yields consistent estimates of the exact channel equalizer when the transmitted signal is i.i.d. or in other words the steady-state overall impulse response ($\boldsymbol{t}$) is a delta function with arbitrary delay. Without loss of generality, we can assume that the channel, the equalizer and the transmitted signal are real-valued. Note that $\max\{|y_n|\}=\max\{|a_n|\}\sum_l|t_l|<\infty$ and, owing to i.i.d. property, $\mathsf{E}y_n^2=\mathsf{E}a_n^2\sum_l t_l^2$. Next we can express the cost (14) in $\boldsymbol{t}$ domain as follows (assuming large $B$):

$$\mathcal{J}_{\mathrm{AN}}(\boldsymbol{t})=\frac{\sum_l|t_l|^2}{(\sum_l|t_l|)^2} \tag{15}$$

(*footnote continued*)
rotation which is widely used in obtaining a simple factor structure in factor analysis.

Evaluating the gradient w.r.t. $k$th tap, we obtain

$$\begin{aligned}\frac{\partial}{\partial t_k}\mathcal{J}_{\mathrm{AN}}&=\frac{(\sum_l|t_l|)^2\frac{\partial\sum_j t_j^2}{\partial t_k}-(\sum_l t_l^2)\frac{\partial(\sum_j|t_j|)^2}{\partial t_k}}{(\sum_l|t_l|)^4}\\&=\frac{2(\sum_l|t_l|)^2t_k-2(\sum_l t_l^2)(\sum_j|t_j|)\frac{t_k}{|t_k|}}{(\sum_l|t_l|)^4}\end{aligned} \tag{16}$$

Equating the above to zero, we get $|t_k|\sum_j|t_j|-\sum_l t_l^2=0$. It shows that for a doubly infinite equalizer, the stable global maxima of $\mathcal{J}_{\mathrm{AN}}$ are along the axis, i.e.

$$\{|t_k|\}=\delta_{k-k^{\dagger}},\quad k^{\dagger}=\ldots,-1,0,1,\ldots \tag{17}$$

and unstable equilibria are along the diagonal, located at $\{|t_k|\}=(1/L)\sum_{j\in I_L}\delta_{k-j}$, where $I_L(L\geq 2)$ is any $L$-element subset of the integer set. The surface of cost (15) is depicted in Fig. 3(a) for a two-tap scenario; it can be seen that the cost is maximized only for the solution specified in (17).

Next, incorporating the a priori signal knowledge $\gamma:=\max\{|a_n|\}$, the cost (14) can be written in a *constrained* form as follows:

$$\boldsymbol{w}^{\dagger}=\arg\max_{\boldsymbol{w}}\mathsf{E}|y_n|^2\quad \text{s.t. } \max\{|y_n|\}\leq\gamma. \tag{18}$$

The geometry of the cost (18) is depicted in Fig. 3(b) for a two-tap scenario; it can be seen that the cost is maximized when the two balls, $\sum_i|t_i|^2$ and $\sum_i|t_i|$, coincide. The cost (18) is quadratic, and the feasible region (constraint) is a convex set. The problem, however, is non-convex and may have multiple local maxima. Nevertheless, we have the following theorem (refer to [29] for proof):

**Theorem 2.** *Assume* $\boldsymbol{w}^{\dagger}$ *is a local optimum in*

$$\boldsymbol{w}^{\dagger}=\arg\max_{\boldsymbol{w}}\mathsf{E}|y_n|^2\quad s.t.\ \max\{|y_n|\}\leq\gamma$$

*and* $\boldsymbol{t}^{\dagger}$ *is the corresponding total channel equalizer impulse-response and channel noise is negligible. Then it holds* $\{|t_k|\}=\delta_{k-k^{\dagger}}$, *where* $k^{\dagger}=\ldots,-1,0,1,\ldots$.

Thus an equalizer which is maximizing the output energy and constraining the largest amplitude is able to mitigate the ISI induced by the channel.

### 3.2. Stochastic gradient-based optimization of $\mathcal{J}_{\mathrm{AN}}$

When an equalizer $\boldsymbol{w}$ optimizes a cost $J=\mathsf{E}\mathcal{J}(y_n)$ by the stochastic gradient-based adaptive method, the resulting algorithm is $\boldsymbol{w}_n=\boldsymbol{w}_{n-1}\pm\mu\partial\mathcal{J}/\partial\boldsymbol{w}_{n-1}^{*}$, where $\mu>0$ is the step-size, governing the speed of convergence and the level of steady-state performance [30]. The positive and negative signs are for maximization and minimization, respectively.

Note that a straightforward gradient-based adaptive implementation of $\mathcal{J}_{\mathrm{AN}}$ is not possible. The reason is that the order statistic $\max\{|y_n|\}$ is not a differentiable quantity. In [28], however, authors presented an instantaneous and differentiable version of constrained $\mathcal{J}_{\mathrm{AN}}$ and obtained the following stochastic gradient-based algorithm:

$$\begin{aligned}&\boldsymbol{w}_n=\boldsymbol{w}_{n-1}+\mu f_n y_n^{*}\boldsymbol{x}_n,\\&f_n=\begin{cases}1 & \text{if } |y_n|<\gamma,\\ -\beta & \text{if } |y_n|\geq\gamma.\end{cases}\end{aligned} \tag{19}$$

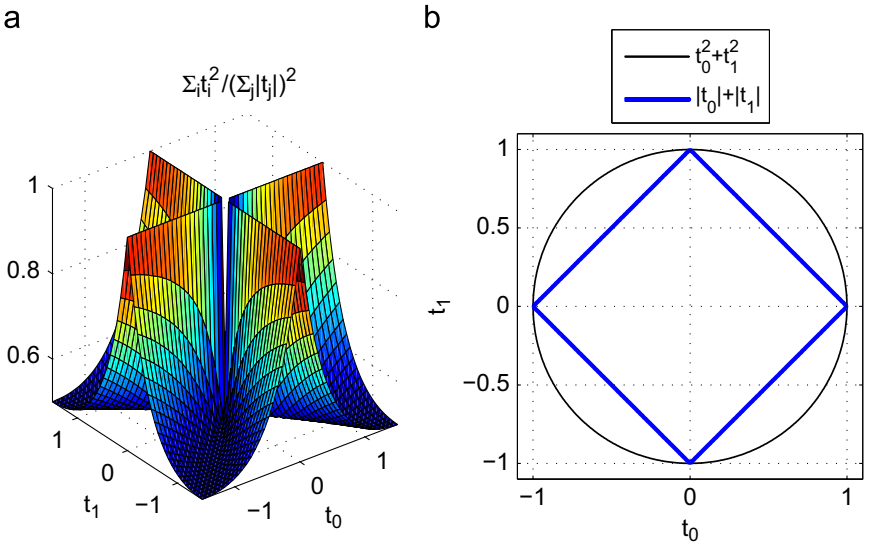


**Fig. 3.** (a) The surface of cost $\mathcal{J}_{\mathrm{AN}}(\boldsymbol{t})$ for a two-tap equalizer and (b) coincidence of unit balls.

where $\beta$ is a constant defined as follows (refer to Appendix B):

$$\beta := \frac{\mathrm{E}|y_n|^2_{\{|y_n|<\gamma\}}}{\mathrm{E}|y_n|^2_{\{|y_n|\geq\gamma\}}} \tag{20}$$

The algorithm (19) was termed as *beta constant modulus algorithm* ($\beta$CMA). The $\beta$CMA may be kept stable in mean-square sense if its step-size satisfies the condition $0<\mu<3/(2\beta\mathrm{tr}(\boldsymbol{R}))$, where $\mathrm{tr}(\cdot)$ is a trace operator and $\boldsymbol{R}=\mathrm{E}\boldsymbol{x}_n\boldsymbol{x}_n^H$ is an autocorrelation matrix [31].

## 4. Adaptive Newton-like optimization of $\mathcal{J}_{\mathrm{AN}}$

In this section, we aim to optimize $\mathcal{J}_{\mathrm{AN}}$ using the Newton-like adaptive method. When Newton-like optimization is used by the equalizer, the update for a minimization scenario is given as follows [32]:

$$\boldsymbol{w}_n = \boldsymbol{w}_{n-1} - \mu\left(\frac{\partial^2\mathcal{J}}{\partial\boldsymbol{w}_{n-1}\partial\boldsymbol{w}_{n-1}^H}\right)^{-1}\left(\frac{\partial\mathcal{J}}{\partial\boldsymbol{w}_{n-1}}\right)^* \tag{21}$$

Firstly, to simplify algebraic manipulation, we suggest the following *instantaneous* cost function:

$$\boldsymbol{w}^\dagger = \arg\min_{\boldsymbol{w}} \mathcal{J}_n \quad \text{where } \mathcal{J}_n := |f_n||y_n|^2-\gamma^2| \tag{22}$$

where $f_n$ is as specified in (19). It is simple to show that solving (22) by the gradient-based method results in $\beta$CMA. For the formulation of a Newton-like scheme, we introduce *exponential weights* (memory) in (22) as follows:

$$\boldsymbol{w}^\dagger = \arg\min_{\boldsymbol{w}} \tilde{\mathcal{J}},$$

$$\text{where } \tilde{\mathcal{J}} := \sum_{k=0}^{n}\lambda^{n-k}|f_k|\cdot\left||\underbrace{\boldsymbol{w}_{n-1}^H\boldsymbol{x}_k}_{=:y_k}|^2-\gamma^2\right| \tag{23}$$

and $\lambda$ is the forgetting factor. When $\lambda$ equals 1, we have a cost which may be considered as a sort of (infinite memory) least-squares problem. For $0<\lambda<1$, however, the cost has effective finite memory $(1/(1-\lambda))$ and it may be optimized adaptively using (21). We readily evaluate gradient $\boldsymbol{g}_n$ and Hessian $\boldsymbol{H}_n$ for (23) as follows:

$$\boldsymbol{g}_n := \frac{\partial\tilde{\mathcal{J}}}{\partial\boldsymbol{w}_{n-1}} = -\sum_{k=0}^{n} f_k y_k \boldsymbol{x}_k^*\lambda^{n-k},$$

$$\boldsymbol{H}_n := \frac{\partial^2\tilde{\mathcal{J}}}{\partial\boldsymbol{w}_{n-1}\partial\boldsymbol{w}_{n-1}^H} = -\sum_{k=0}^{n} f_k\boldsymbol{x}_k\boldsymbol{x}_k^H\lambda^{n-k} \tag{24}$$

Here, we encounter a problem. Note that, for the required steady-state condition $|y_k|\leq\gamma$ implying $f_k=+1, \forall k$, however, it leads to an undesirable situation

$$\boldsymbol{H}_n = -\frac{1}{1-\lambda}\widehat{\boldsymbol{R}}_n \preccurlyeq \boldsymbol{0} \tag{25}$$

where $\widehat{\boldsymbol{R}}_n := \sum_{k=0}^{n}\boldsymbol{x}_k\boldsymbol{x}_k^H \succcurlyeq \boldsymbol{0}$. So, assuming a converging equalizer, the Hessian is found to be negative definite. It means that such an equalizer will try to maximize the cost instead of minimizing it. This is contrary to the problem definition and we conclude that *a straight Newton-like $\beta$CMA equalizer, implementing Hessian as specified in (24), will diverge*. Simulation study is found in conformation with this argument.

One possible way to resolve this matter is to ensure that the recursively computed Hessian remains positive definite. Consider the following solution:

$$\boldsymbol{H}_n = \sum_{k=0}^{n}|f_k|\boldsymbol{x}_k\boldsymbol{x}_k^H\lambda^{n-k} = \lambda\boldsymbol{H}_{n-1}+|f_n|\boldsymbol{x}_n\boldsymbol{x}_n^H \tag{26}$$

Invoking the matrix inversion lemma, we obtain

$$\boldsymbol{P}_n = \boldsymbol{H}_n^{-1} = \frac{1}{\lambda}\left(\boldsymbol{P}_{n-1}-\frac{\boldsymbol{P}_{n-1}\boldsymbol{x}_n\boldsymbol{x}_n^H\boldsymbol{P}_{n-1}^H}{\lambda z_n+\boldsymbol{x}_n^H\boldsymbol{P}_{n-1}\boldsymbol{x}_n}\right) \tag{27}$$

where $z_n=|f_n|^{-1}$. It was shown in [33], that the performance of a Newton-like algorithm may be enhanced if $z_n$ is computed as an iterative estimate, that is $z_n\equiv\widehat{z}_n=\langle|f_n|^{-1}\rangle$ where $\langle\cdot\rangle$ represents some averaged estimate of the enclosed entity. One of the possibilities is $\widehat{z}_n=\widehat{z}_{n-1}+1/(n+1)(|f_n|^{-1}-\widehat{z}_{n-1})$. Further note that $\boldsymbol{g}_n=\lambda\boldsymbol{g}_{n-1}+f_n y_n\boldsymbol{x}_n^*$.

**Table 1**
NL-$\beta$CMA.

$\boldsymbol{w}_0 = [\boldsymbol{0}_{1\times(N-1)/2}, 1, \boldsymbol{0}_{1\times(N-1)/2}]^T$, $\beta > 1$,
$\boldsymbol{P}_0 = \varepsilon \boldsymbol{I}_{N\times N}$, $\boldsymbol{I}$ is identity matrix and $0 < \varepsilon \ll 1$

$$f_n = \begin{cases} 1 & \text{if } |y_n| < \gamma \\ -\beta & \text{if } |y_n| \geq \gamma. \end{cases}$$

$$\widehat{z}_n = \widehat{z}_{n-1} + \frac{1}{n+1}\left(|f_n|^{-1} - \widehat{z}_{n-1}\right)$$

$$\boldsymbol{P}_n = \frac{1}{\lambda}\left(\boldsymbol{P}_{n-1} - \frac{\boldsymbol{P}_{n-1}\boldsymbol{x}_n\boldsymbol{x}_n^H\boldsymbol{P}_{n-1}^H}{\lambda\widehat{z}_n + \boldsymbol{x}_n^H\boldsymbol{P}_{n-1}\boldsymbol{x}_n}\right)$$

$$\boldsymbol{w}_n = \boldsymbol{w}_{n-1} + \mu\boldsymbol{P}_n f_n y_n^* \boldsymbol{x}_n$$

For $\lambda$ close to one, the vector $\boldsymbol{g}_n$ will be dominated by its former estimate $\boldsymbol{g}_{n-1}$, that is $\boldsymbol{g}_n \approx \boldsymbol{g}_{n-1}$, and this leads to a simpler expression $\boldsymbol{g}_n \approx (1-\lambda)^{-1} f_n y_n \boldsymbol{x}_n^*$. With this consideration, the proposed Newton-like $\beta$CMA (NL-$\beta$CMA) update is given as follows:

$$\boldsymbol{w}_n = \boldsymbol{w}_{n-1} + \mu\boldsymbol{P}_n f_n y_n^* \boldsymbol{x}_n, \tag{28}$$

where $\boldsymbol{P}_n$ is as evaluated in (27) and the factor $(1-\lambda)^{-1}$ in $\boldsymbol{g}_n$ is merged with $\mu$. Note that the recursive calculation of $\boldsymbol{P}_n$ keeps the computational complexity of the proposed scheme to $\mathcal{O}(N^2)$ per iteration. The proposed scheme is summarized in Table 1.

## 5. Steady-state performance of NL-$\beta$CMA

In this section, we study the steady-state tracking performance of NL-$\beta$CMA in a non-stationary environment. Although the steady-state performance essentially corresponds to only one point on the learning curve of the adaptive filter, there are many situations where this information is of value by itself. The addressed approach is based on studying the energy flow through each iteration of an adaptive filter [34,35], and it relies on a fundamental error variance relation that avoids the weight-error variance recursion altogether. We may remark that although we focus on the steady-state performance of NL-$\beta$CMA, the same approach can also be used to study the transient (i.e., convergence and stability) behavior of this filter. These details will be provided elsewhere.

Consider a non-stationary environment with time-varying optimal weight-vector $\boldsymbol{w}_n^o$, also called the Wiener filter, given by

$$\boldsymbol{w}_n^o = \boldsymbol{w}_{n-1}^o + \boldsymbol{q}_n \tag{29}$$

where $\boldsymbol{q}_n$ is some random perturbation such that $\mathsf{E}\boldsymbol{q}_n\boldsymbol{q}_n^H = \boldsymbol{Q} = \sigma_q^2\boldsymbol{I}$ [36]. Using the Wiener solution, the data $a_n$ can be expressed as $a_n = (\boldsymbol{w}_{n-1}^o)^H\boldsymbol{x}_n + v_n$, where $v_n$ is disturbance and is uncorrelated with $\boldsymbol{x}_n$, ($\mathsf{E}v_n^*\boldsymbol{x}_n = \boldsymbol{0}$) [36]. The purpose of the tracking analysis of an adaptive filter is to study its ability to track such time variations. The weight error vector $\tilde{\boldsymbol{w}}_n$ which quantifies how far away the weight vector is from the Wiener solution is given as

$$\tilde{\boldsymbol{w}}_n = \boldsymbol{w}_n^o - \boldsymbol{w}_n \tag{30}$$

The so-called a posteriori and a priori error are defined as $e_{p,n} = (\tilde{\boldsymbol{w}}_n - \boldsymbol{q}_n)^H\boldsymbol{x}_n$ and $e_{a,n} = \tilde{\boldsymbol{w}}_{n-1}^H\boldsymbol{x}_n$, respectively. First consider a generic update expression $\boldsymbol{w}_n = \boldsymbol{w}_{n-1} + \mu\boldsymbol{P}_n\varphi_n^*\boldsymbol{x}_n$. Subtracting $\boldsymbol{w}_n^o$ from both sides of this update, substituting its value from (29) and exploiting (30), we get

$$\tilde{\boldsymbol{w}}_n = \tilde{\boldsymbol{w}}_{n-1} - \mu\boldsymbol{P}_n\varphi_n^*\boldsymbol{x}_n + \boldsymbol{q}_n \tag{31}$$

Taking Hermitian transpose of (31) and post-multiplying by $\boldsymbol{x}_n$, we obtain:

$$(\tilde{\boldsymbol{w}}_n - \boldsymbol{q}_n)^H\boldsymbol{x}_n = \tilde{\boldsymbol{w}}_{n-1}^H\boldsymbol{x}_n - \mu\boldsymbol{x}_n^H\boldsymbol{P}_n\boldsymbol{x}_n\varphi_n \tag{32a}$$

$$\Rightarrow e_{p,n} = e_{a,n} - \mu\varphi_n\|\boldsymbol{x}_n\|_{\boldsymbol{P}_n}^2 \tag{32b}$$

where $\|\boldsymbol{x}_n\|_{\boldsymbol{P}_n}^2 := \boldsymbol{x}_n^H\boldsymbol{P}_n\boldsymbol{x}_n$ is the weighted Euclidian norm. From (32b), we have $\varphi_n = \mu^{-1}\|\boldsymbol{x}_n\|_{\boldsymbol{P}_n}^{-2}(e_{a,n} - e_{p,n}), \boldsymbol{x}_n \neq \boldsymbol{0}$. Now, substituting in the value of $\varphi_n$ in (31) we get

$$\tilde{\boldsymbol{w}}_n - \boldsymbol{q}_n = \tilde{\boldsymbol{w}}_{n-1} - \tilde{\mu}_n\boldsymbol{P}_n(e_{a,n} - e_{p,n})^*\boldsymbol{x}_n \tag{33}$$

where $\tilde{\mu}_n := (\|\boldsymbol{x}_n\|_{\boldsymbol{P}_n}^2)^+$ is used to define pseudo-inverse of $\|\boldsymbol{x}_n\|_{\boldsymbol{P}_n}^2$. Taking $\boldsymbol{P}_n^{-1}$ weighted norm on both sides and simplifying, we get $\|\tilde{\boldsymbol{w}}_n - \boldsymbol{q}_n\|_{\boldsymbol{P}_n^{-1}}^2$ on the left and following quantity on the right $\|\tilde{\boldsymbol{w}}_{n-1}\|_{\boldsymbol{P}_n^{-1}}^2 - \tilde{\mu}_n|e_{a,n}|^2 + \tilde{\mu}_n|e_{p,n}|^2$. This results in the energy conservation relation and it is summarized as

$$\|\tilde{\boldsymbol{w}}_n - \boldsymbol{q}_n\|_{\boldsymbol{P}_n^{-1}}^2 + \tilde{\mu}_n|e_{a,n}|^2 = \|\tilde{\boldsymbol{w}}_{n-1}\|_{\boldsymbol{P}_n^{-1}}^2 + \tilde{\mu}_n|e_{p,n}|^2 \tag{34}$$

Examining the expectation of the left most term of the energy relation, we have

$$\begin{aligned}
&\mathsf{E}\|\tilde{\boldsymbol{w}}_n - \boldsymbol{q}_n\|_{\boldsymbol{P}_n^{-1}}^2 \\
&= \mathsf{E}\|\tilde{\boldsymbol{w}}_n\|_{\boldsymbol{P}_n^{-1}}^2 + \mathsf{E}\|\boldsymbol{q}_n\|_{\boldsymbol{P}_n^{-1}}^2 - 2\Re\mathsf{E}\boldsymbol{q}_n^H\boldsymbol{P}_n^{-1}\tilde{\boldsymbol{w}}_n \\
&= \mathsf{E}\|\tilde{\boldsymbol{w}}_n\|_{\boldsymbol{P}_n^{-1}}^2 + \mathsf{E}\|\boldsymbol{q}_n\|_{\boldsymbol{P}_n^{-1}}^2 - 2\Re(\mathsf{E}\boldsymbol{q}_n^H\boldsymbol{P}_n^{-1}\tilde{\boldsymbol{w}}_{n-1} \\
&\qquad + \mathsf{E}\|\boldsymbol{q}_n\|_{\boldsymbol{P}_n^{-1}}^2 - \mathsf{E}\boldsymbol{q}_n^H\boldsymbol{x}_n\varphi_n^*) \\
&= \mathsf{E}\|\tilde{\boldsymbol{w}}_n\|_{\boldsymbol{P}_n^{-1}}^2 - \mathsf{E}\|\boldsymbol{q}_n\|_{\boldsymbol{P}_n^{-1}}^2
\end{aligned} \tag{35}$$

where the vanished terms are a result of the following assumptions about the random-walk driving-noise sequence $\{\boldsymbol{q}_n\}$.

(A.1) $\boldsymbol{q}_n$ is i.i.d. and $\mathsf{E}\boldsymbol{q}_n = \boldsymbol{0}$
$\boldsymbol{Q} = \mathsf{E}\boldsymbol{q}_n\boldsymbol{q}_n^H = \sigma_q^2\boldsymbol{I} > \boldsymbol{0}$
$\{\boldsymbol{q}_n\} \perp \{\boldsymbol{x}_n\}$, $\{\boldsymbol{q}_n\} \perp \{\varphi_n\}$

Using (35) with (34), and noting that in steady state $\mathsf{E}\|\tilde{\boldsymbol{w}}_n\|_{\boldsymbol{P}_n^{-1}}^2 = \mathsf{E}\|\tilde{\boldsymbol{w}}_{n-1}\|_{\boldsymbol{P}_n^{-1}}^2$, we obtain

$$\tilde{\mu}_n\mathsf{E}|e_{a,n}|^2 = \mathsf{E}\|\boldsymbol{q}_n\|_{\boldsymbol{P}_n^{-1}}^2 + \tilde{\mu}_n\mathsf{E}|e_{p,n}|^2 \tag{36}$$

This equation can now be solved for the steady-state excess mean-square-error (EMSE), which is defined by

$$\zeta := \lim_{n\to\infty}\mathsf{E}|e_{a,n}|^2 \tag{37}$$

We emphasize that (36) is an exact relation that holds without any approximation or assumption, except for the assumption that the filter is in steady state. The procedure of finding the EMSE through (36) avoids the need for evaluating $\mathsf{E}\|\tilde{\boldsymbol{w}}_n\|^2$ or its steady-state value $\mathsf{E}\|\tilde{\boldsymbol{w}}_\infty\|^2$. To proceed further, we make the following assumptions (refer to [37] for similar treatment):

(A.2) $\mathsf{E}\boldsymbol{P}_n\boldsymbol{x}_n\boldsymbol{x}_n^H = \mathsf{E}\boldsymbol{P}_n\mathsf{E}\boldsymbol{x}_n\boldsymbol{x}_n^H$ (holds in steady state as $\boldsymbol{P}_n$ can be replaced by $\mathsf{E}\boldsymbol{P}_n$, see [38])

(A.3) $\mathrm{E}\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}|e_{a,n}|^2=\mathrm{E}\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}\mathrm{E}|e_{a,n}|^2$ (separation principle [38])
(A.4) $\lim_{n\to\infty}\mathrm{E}\boldsymbol{H}_n^{-1}=(\lim_{n\to\infty}\mathrm{E}\boldsymbol{H}_n)^{-1}$ (justified for *complex Wishart distribution* in [30] and shown to be reasonable via simulations in [38])

From (32), we have $e_{p,n}=e_{a,n}-\mu\varphi_n\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}$, substituting it into (36), we obtain

$$2\mathrm{E}\Re[e^*_{a,n}\varphi_n]=\mu\mathrm{E}\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}|\varphi_n|^2+\mu^{-1}\mathrm{E}\|\boldsymbol{q}_n\|^2_{\boldsymbol{P}_n^{-1}} \tag{38}$$

So solve the right hand side (RHS) of (38), we employ

(A.5) $\mathrm{E}\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}|\varphi_n|^2=\mathrm{E}\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}\mathrm{E}|\varphi_n|^2$ (separation principle [38])

First, let us solve $\mathrm{E}\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}$ as

$$\begin{aligned}\mathrm{E}\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}&=\mathrm{tr}(\mathrm{E}\boldsymbol{P}_n\boldsymbol{x}_n\boldsymbol{x}_n^H)\\&=\mathrm{tr}(\mathrm{E}\boldsymbol{P}_n\mathrm{E}\boldsymbol{x}_n\boldsymbol{x}_n^H)\\&=\mathrm{tr}(\mathrm{E}\boldsymbol{P}_n\boldsymbol{R})\end{aligned} \tag{39}$$

Here $\mathrm{E}\boldsymbol{P}_n$ requires further investigation. It is observed that

$$\begin{aligned}\mathrm{E}\boldsymbol{H}_n&=\sum_{i=1}^{n}\lambda^{n-i}\mathrm{E}|f_n|\boldsymbol{x}_n\boldsymbol{x}_n^H+\varepsilon\lambda^n\boldsymbol{I}\\&=\mathrm{E}|f_n|\mathrm{E}\boldsymbol{x}_n\boldsymbol{x}_n^H\sum_{i=1}^{n}\lambda^{n-i}+\varepsilon\lambda^n\boldsymbol{I}\\&=\rho\boldsymbol{R}\frac{1-\lambda^n}{1-\lambda}+\varepsilon\lambda^n\boldsymbol{I}\quad\text{for }\lambda<1\end{aligned} \tag{40}$$

where $\rho:=\mathrm{E}|f_n|$. Further, exploiting A.4, we get

$$\lim_{n\to\infty}\mathrm{E}\boldsymbol{P}_n\approx\left(\lim_{n\to\infty}\mathrm{E}\boldsymbol{H}_n\right)^{-1}=\frac{1}{\rho}(1-\lambda)\boldsymbol{R}^{-1}\quad\text{for }\lambda<1 \tag{41}$$

by using (41) with (39), the following equation results

$$\mathrm{E}\|\boldsymbol{x}_n\|^2_{\boldsymbol{P}_n}=\mathrm{tr}\left(\frac{1}{\rho}(1-\lambda)\boldsymbol{R}^{-1}\boldsymbol{R}\right)=\frac{1}{\rho}(1-\lambda)\,\mathrm{tr}(\boldsymbol{I})=\frac{1}{\rho}(1-\lambda)N. \tag{42}$$

Proceeding to the second term on the RHS of (38), we get

$$\begin{aligned}\mathrm{E}\|\boldsymbol{q}_n\|^2_{\boldsymbol{P}_n^{-1}}&=\mathrm{E}\boldsymbol{q}_n^H\boldsymbol{H}_n\boldsymbol{q}_n=\mathrm{tr}\left(\mathrm{E}\boldsymbol{q}_n\boldsymbol{q}_n^H\boldsymbol{H}_n\right)\\&=\mathrm{tr}\left(\rho\mathrm{E}\boldsymbol{q}_n\boldsymbol{q}_n^H\sum_{i=1}^{n}\lambda^{n-i}\mathrm{E}\boldsymbol{x}_n\boldsymbol{x}_n^H+\varepsilon\lambda^n\mathrm{E}\boldsymbol{q}_n\boldsymbol{q}_n^H\right)\\&=\rho\,\mathrm{tr}(\boldsymbol{QR})\sum_{i=1}^{n}\lambda^{n-i}+\varepsilon\lambda^n\,\mathrm{tr}(\boldsymbol{Q})\\&=\rho\frac{1-\lambda^n}{1-\lambda}\mathrm{tr}(\boldsymbol{QR})+\varepsilon\lambda^n\,\mathrm{tr}(\boldsymbol{Q})\quad\text{for }\lambda<1\end{aligned} \tag{43}$$

so that

$$\lim_{n\to\infty}\mathrm{E}\|\boldsymbol{q}_n\|^2_{\boldsymbol{P}_n^{-1}}=\rho(1-\lambda)^{-1}\,\mathrm{tr}(\boldsymbol{QR})\quad\text{for }\lambda<1 \tag{44}$$

Using (42) and (44) the RHS of (38) becomes

$$\mathrm{RHS}=\frac{\mu}{\rho}N(1-\lambda)\mathrm{E}|\varphi_n|^2+\frac{\rho}{\mu}(1-\lambda)^{-1}\,\mathrm{tr}(\boldsymbol{QR}) \tag{45}$$

The parameter $\rho$ is computed as follows:

$$\begin{aligned}\rho:=\mathrm{E}|f_n|&=\Pr\left[|y_n|<\gamma\right]+\beta\Pr\left[|y_n|>\gamma\right]\\&=1+(\beta-1)\Pr\left[|y_n|>\gamma\right]\\&=1+(\beta-1)\sum_{j=1}^{L}\frac{M_j}{M}\int_{\gamma}^{\infty}p(|y_n|;R_j)\,\mathrm{d}|y_n|\\&=1+(\beta-1)\mathrm{E}Q_{1,0}\left(\frac{|a_n|}{\sigma},\frac{\gamma}{\sigma}\right).\end{aligned} \tag{46}$$

where $Q_{m,v}(a,b)$ is the Nuttall $Q$-function [39] as defined below:

$$Q_{m,v}(a,b):=\int_b^{\infty}x^m e^{-1/2(x^2+a^2)}I_v(ax)\,\mathrm{d}x,\quad a,b>0 \tag{47}$$

where $m>1$, $v\geq 0$ and $I_v(\cdot)$ is the $v$th-order modified Bessel function of first kind. Now to solve $\mathrm{E}|\varphi_n|^2$, where $\varphi_n=f_ny_n$, we see that $f_n$ is a piecewise function and it has got a discontinuity at $|y_n|=\gamma$. We start as

$$\begin{aligned}\mathrm{E}|\varphi_n|^2&=\mathrm{E}f_n^2|y_n|^2\\&=\mathrm{E}|y_n|^2_{\{0<|y_n|<\gamma\}}+\beta^2\mathrm{E}|y_n|^2_{\{\gamma\leq|y_n|<\infty\}}\\&=\mathrm{E}|y_n|^2+(\beta^2-1)\underbrace{\mathrm{E}|y_n|^2_{\{\gamma\leq|y_n|<\infty\}}}_{=:A}\end{aligned} \tag{48}$$

where the factor $A$ is expressed and evaluated as follows:

$$\begin{aligned}A:=\mathrm{E}|y_n|^2_{\{\gamma\leq|y_n|<\infty\}}&=\sum_{j=1}^{L}\frac{M_j}{M}\int_{\gamma}^{\infty}|y_n|^2p(|y_n|;R_j)\,\mathrm{d}|y_n|\\&=\sum_{j=1}^{L}\frac{M_j}{M}\int_{\gamma}^{\infty}\frac{|y_n|^3}{\sigma^2}\exp\left(-\frac{|y_n|^2+R_j^2}{2\sigma^2}\right)I_0\left(\frac{|y_n|R_j}{\sigma^2}\right)\mathrm{d}|y_n|\\&=\sigma^2\sum_{j=1}^{L}\frac{M_j}{M}Q_{3,0}\left(\frac{R_j}{\sigma},\frac{\gamma}{\sigma}\right)=\sigma^2\mathrm{E}Q_{3,0}\left(\frac{|a_n|}{\sigma},\frac{\gamma}{\sigma}\right).\end{aligned} \tag{49}$$

To solve (48), we need to calculate statistical moments of the modulus $|y_n|$. Next we compute the left hand side (LHS) of (38),

$$\begin{aligned}\mathrm{LHS}&=2\mathrm{E}\Re[e^*_{a,n}\varphi]=2\mathrm{E}\Re[(a_n-y_n)^*f_ny_n]\\&=2\mathrm{E}\Re[f_na_n^*y_n]-2\mathrm{E}f_n|y_n|^2\\&=2\left(\underbrace{\mathrm{E}f_n|a_n|^2}_{=:B}-\underbrace{\mathrm{E}\Re[f_na_n^*e_{a,n}]}_{=:C}-\underbrace{\mathrm{E}f_n|y_n|^2}_{=:D}\right)\end{aligned} \tag{50}$$

The term $B$ is computed as follows:

$$\begin{aligned}B&=\mathrm{E}|a_n|^2_{\{0<|y_n|<\gamma\}}-\beta\mathrm{E}|a_n|^2_{\{\gamma\leq|y_n|<\infty\}}\\&=\mathrm{E}|a_n|^2-(\beta+1)\mathrm{E}|a_n|^2Q_{1,0}\left(\frac{|a_n|}{\sigma},\frac{\gamma}{\sigma}\right),\end{aligned} \tag{51}$$

Next, the term $D$ is computed as follows:

$$\begin{aligned}D&=\mathrm{E}|y_n|^2_{\{0<|y_n|<\gamma\}}-\beta\mathrm{E}|y_n|^2_{\{\gamma\leq|y_n|<\infty\}}\\&=\mathrm{E}|y_n|^2-(\beta+1)\underbrace{\mathrm{E}|y_n|^2_{\{\gamma\leq|y_n|<\infty\}}}_{=:A}\end{aligned} \tag{52}$$

where $A$ is as obtained in (49). Next the term $C$ is computed:

$$\begin{aligned}C&=\mathrm{E}\Re[a_n^*e_{a,n}]_{\{0<|y_n|<\gamma\}}-\beta\mathrm{E}\Re[a_n^*e_{a,n}]_{\{\gamma\leq|y_n|<\infty\}}\\&=\underbrace{\mathrm{E}\Re[a_n^*e_{a,n}]}_{=0}-(\beta+1)\mathrm{E}\Re[a_n^*e_{a,n}]_{\{\gamma\leq|y_n|<\infty\}}\\&=-(\beta+1)\left\{\mathrm{E}\left[\frac{|a_n|^2+\zeta}{2}Q_{1,0}\left(\frac{|a_n|}{\sigma},\frac{\gamma}{\sigma}\right)\right]\right.\\&\quad\left.-\frac{\sigma^2}{2}\mathrm{E}Q_{3,0}\left(\frac{|a_n|}{\sigma},\frac{\gamma}{\sigma}\right)\right\}.\end{aligned} \tag{53}$$

Combining (45)–(53), we obtain the following expression to solve for EMSE of the proposed NL-$\beta$CMA, $\zeta^{\mathrm{NL}\text{-}\beta\mathrm{CMA}}$, as follows[2]:

$$\zeta\left(1-\frac{\mu}{\rho}N(1-\lambda)(\beta-1)\right)\mathrm{E}Q_{3,0}\left(\sqrt{\frac{2}{\zeta}}|a_n|,\sqrt{\frac{2}{\zeta}}\gamma\right) - \frac{2}{1+\beta}\left(\frac{\rho\,\mathrm{tr}(\boldsymbol{QR})}{\mu(1-\lambda)}+\frac{\mu}{\rho}N(1-\lambda)(\zeta+\mathrm{E}|a_n|^2)+2\zeta\right) + 2\mathrm{E}(\zeta-|a_n|^2)Q_{1,0}\left(\sqrt{\frac{2}{\zeta}}|a_n|,\sqrt{\frac{2}{\zeta}}\gamma\right)=0. \tag{54}$$

where $\rho$ is as specified in (46).

## 6. Simulation results

In [28], the $\beta$CMA has already been compared and shown to be better than four existing algorithms which include traditional CMA [16] and three of its variants: the (unnormalized) relaxed CMA (RCMA) [40], the Shtrom–Fan algorithm (SFA) [41] and the generalized CMA (GCMA) [42]. In this work, we provide performance comparison of NL-$\beta$CMA (as summarized in Table 1) with Newton-like CMA (NL-CMA) [32] and recursive least square CMA (RLS-CMA) [43]. The performances of CMA and $\beta$CMA are shown as standard benchmark. Moreover, we consider an unorthodox benchmark [44] which is an adaptive method for blind equalization and relies on explicit estimation of PDF using *Parzen* window and is termed as a stochastic gradient algorithm (SQD). For reference CMA, $\beta$CMA, SQD, RLS-CMA and NL-CMA are summarized in Table 2.

Here we would like to highlight the following important implementation details: first, for the SQD scheme the required compensation factor $F(\sigma)$ is computed numerically (see [44] for details). The compensation factor $F(\sigma)$ is a function of the constellation scheme and hence we find the values of $F(\sigma)$ for 16APSK(8,8) using the bisection method (see Fig. 4). The authors of [44] highlight that the use of a smaller $\sigma$ (e.g., 1) results in an increased number of local minima of the cost function. It is for this reason that we use a higher value of $\sigma$ (i.e., $\sigma=15$) in all simulations (the corresponding $F(\sigma)$ is 1.325 as shown in Fig. 4). Secondly, to ensure the stability of NL-CMA, we slightly modify it. According to Miranda et al. [45] one must check the reliability of the error quantity in Hessian. If the statistic of error quantity in Hessian is not reliable, one must adhere to simple autocorrelation matrix. Incorporating this recommendation and owing to sub-Gaussian nature of the transmitted signal, the stability of NL-CMA requires to use

$$z_n=\begin{cases}(2|y_n|^2-R^2)^{-1} & \text{if } (2|y_n|^2-R^2)^{-1}\geq 0\\ 1 & \text{otherwise.}\end{cases}$$

For simulation purposes, equalizer length is set to $N=7$ (unless otherwise noted) and $\boldsymbol{w}_0=[\boldsymbol{0}_{1\times\lfloor(N-1)/2\rfloor},1,\boldsymbol{0}_{1\times\lceil(N-1)/2\rceil}]^T$ i.e., center tap initialization is used. The data is modulated using 8APSK(4,4) and 16APSK(8,8) shown in Fig. 5(b) and (c). For all experiments, the value of $\beta$ is obtained using (B.7) (this value is 1.9319 for 8APSK(4,4) and 3 for 16APSK(8,8)). ISI (as defined in (2)) is used as index to evaluate the performance of different equalization

[2] We have observed that the traditional root-finding methods like bisection or line-search are good enough for solving (54) and secondly, possibly due to the monotonicity of Nuttall $Q$-function, the expression (54) yielded unique solutions in all simulation examples considered in this work.

**Table 2**
Summary of compared algorithms.

(a) **CMA**

$f_n=R^2-|y_n|^2$
$\boldsymbol{w}_n=\boldsymbol{w}_{n-1}+\mu f_n y_n^*\boldsymbol{x}_n$

(b) $\beta$ **CMA**

$f_n=\begin{cases}1 & \text{if } |y_n|<\gamma\\ -\beta & \text{if } |y_n|\geq\gamma.\end{cases}$
$\boldsymbol{w}_n=\boldsymbol{w}_{n-1}+\mu f_n y_n^*\boldsymbol{x}_n$

(c) **SQD**

$K'_\sigma(x)=-\frac{x}{\sqrt{2\pi}\sigma^3}\exp(-x^2/2\sigma^2)$
$F(\sigma)$: compensation factor
$N_s$: Size of alphabet $\{a\}$
$d_i=(|y_n|^2-F(\sigma)|a_i|^2)$
$\nabla_{\boldsymbol{w}}\mathcal{J}(\boldsymbol{w})=\frac{1}{N_s}\sum_{i=1}^{N_s}K'_\sigma(d_i)y_n^*\boldsymbol{x}_n$
$\boldsymbol{w}_n=\boldsymbol{w}_{n-1}-\mu\nabla_{\boldsymbol{w}}\mathcal{J}(\boldsymbol{w})$

(d) **RLS-CMA**

$\boldsymbol{P}_0=\varepsilon\boldsymbol{I}_{N\times N}$, $0<\varepsilon\ll 1$
$\boldsymbol{z}_n=\boldsymbol{x}_n y_n^*$
$\boldsymbol{g}_n=\boldsymbol{P}_{n-1}\boldsymbol{z}_n/(\lambda+\boldsymbol{z}_n^H\boldsymbol{P}_{n-1}\boldsymbol{z}_n)$
$\boldsymbol{P}_n=(\boldsymbol{P}_{n-1}-\boldsymbol{g}_n\boldsymbol{z}_n^H\boldsymbol{P}_{n-1})/\lambda$
$f_n=R^2-|y_n|^2$
$\boldsymbol{w}_n=\boldsymbol{w}_{n-1}+\mu\boldsymbol{g}_n f_n$

(e) **NL-CMA**

$\boldsymbol{P}_0=\varepsilon\boldsymbol{I}_{N\times N}$, $0<\varepsilon\ll 1$
$f_n=R^2-|y_n|^2$
$z_n=\begin{cases}(2|y_n|^2-R^2)^{-1} & \text{if } (2|y_n|^2-R^2)^{-1}\geq 0\\ 1 & \text{otherwise.}\end{cases}$
$\boldsymbol{P}_n=\frac{1}{\lambda}\left(\boldsymbol{P}_{n-1}-\frac{\boldsymbol{P}_{n-1}\boldsymbol{x}_n\boldsymbol{x}_n^H\boldsymbol{P}_{n-1}^H}{\lambda z_n+\boldsymbol{x}_n^H\boldsymbol{P}_{n-1}\boldsymbol{x}_n}\right)$
$\boldsymbol{w}_n=\boldsymbol{w}_{n-1}+\mu(1-\lambda^n)\boldsymbol{P}_n f_n y_n^*\boldsymbol{x}_n$

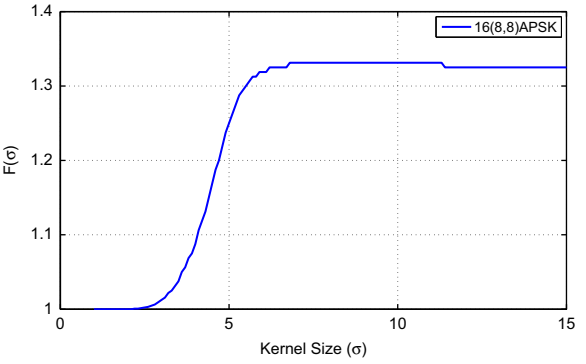


**Fig. 4.** Numerically computed $F(\sigma)$ for 16APSK(8,8).

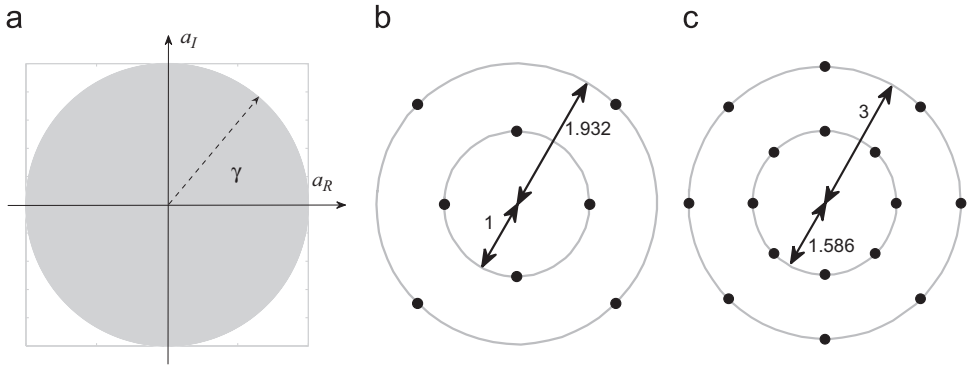


**Fig. 5.** (a) A hypothetical dense APSK, (b) practical 8APSK(4,4), and (c) practical 16APSK(8,8).

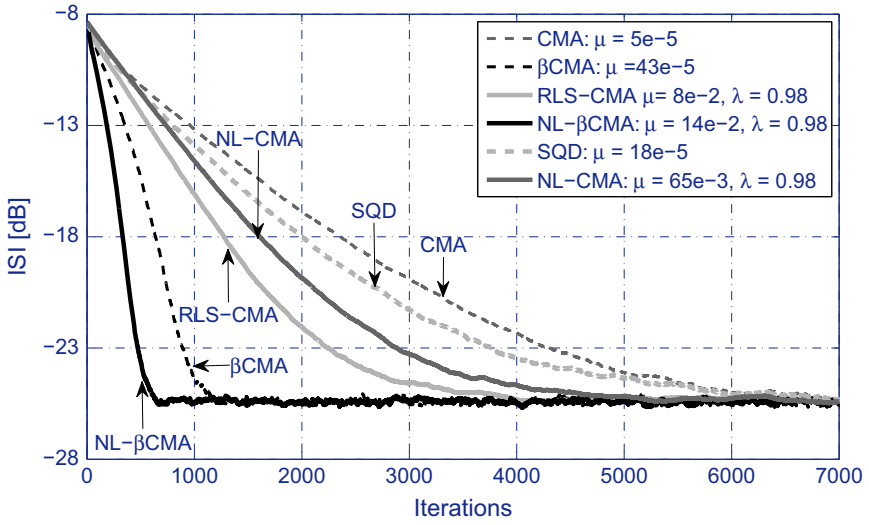


**Fig. 6.** Learning curves for channel with low eigenvalue spread: $N=7$, 16APSK(8,8).

schemes. Experiments comparing the proposed scheme to existing equalization methods are discussed in 6.1 and 6.2. Furthermore, additional experiments are carried to compare the theoretical and practical EMSE of the proposed schemes for both 8APSK(4,4) and 16APSK(8,8). The details of these experiments are discussed in 6.3.

### 6.1. Experiment 1

In this experiment the performance of the proposed scheme is tested for a channel with low eigenvalue spread. The signal is modulated using 16APSK(8,8) and a complex-valued channel with coefficients $\boldsymbol{h}_{\mathrm{R}}=[-0.005, 0.009, -0.024, 0.854, -0.218, 0.049, -0.016]$ and $\boldsymbol{h}_{\mathrm{I}}=[-0.004, 0.03, -0.104, 0.52, 0.273, -0.074, 0.02]$, where $\boldsymbol{h}=\boldsymbol{h}_{\mathrm{R}}+j\boldsymbol{h}_{\mathrm{I}}$ is used [46]. The eigenvalue spread of the channel is 5.83. The signal-to-noise ratio (SNR) is set to 30 dB and stationary environment is assumed (i.e., $\sigma_q=0$, where $\sigma_q$ is the standard deviation of the perturbation $\boldsymbol{q}_n$ in (29)). Learning curves are obtained for 7000 iterations and are ensemble averaged over 200 independent runs. The step sizes and forgetting factors of all schemes are adjusted for comparable performances (i.e., the same steady state ISI) and are mentioned in Fig. 6. It can be observed from Fig. 6 that NL-$\beta$CMA shows the fastest convergence rate followed by $\beta$CMA.

### 6.2. Experiment 2

In this experiment, a channel with higher eigenvalue spread is considered. Generally, an increase in the eigenvalue spread of the channel results in poor performance of equalization schemes. For this experiment, the 16APSK (8,8) modulated signal is passed through a channel with impulse response $\boldsymbol{h}=[0, 0.1, 0.4, 0.8, 0.4, 0.1, 0]$ (the eigenvalue spread of this channel is calculated to be 65.28). Again, SNR is set to 30 dB and a stationary environment is assumed. The step sizes and forgetting factors of all schemes are adjusted for comparable performances and are summarized in Fig. 7. The learning curves are obtained for 40,000 iterations and are ensemble averaged over 200 independent runs. It can be observed from Fig. 7 that the performance of SQD, CMA and $\beta$CMA degrades significantly in comparison with the first experiment. However, it should be noted that NL-CMA, RLS-CMA and NL-$\beta$CMA converge relatively faster. Further, among the three fast

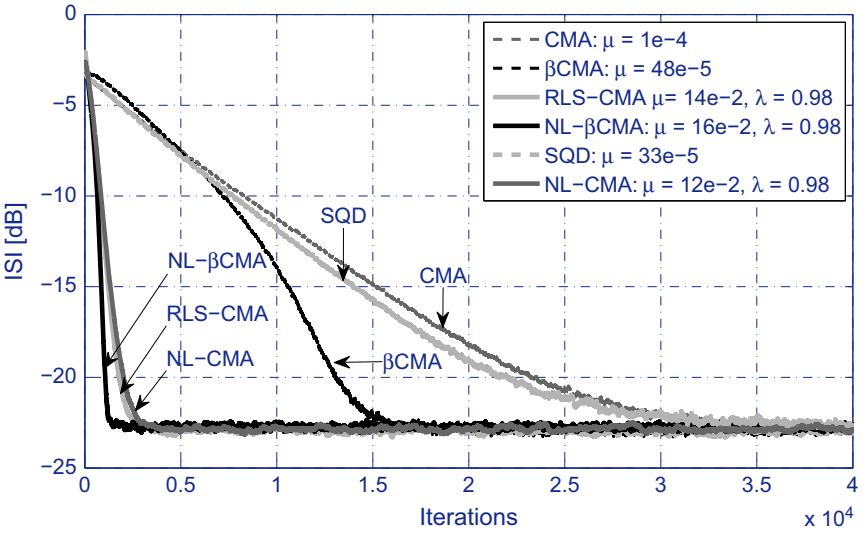


**Fig. 7.** Learning curves for channel with high eigenvalue spread: $N=7$, 16APSK(8,8).

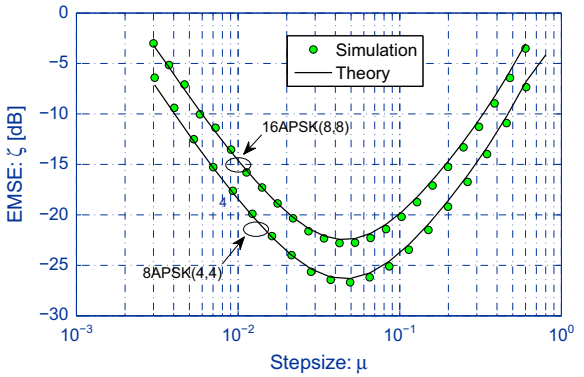


**Fig. 8.** Steady state EMSE using 8APSK(4,4) and 16APSK(8,8).

converging schemes, NL-$\beta$CMA shows the fastest convergence rate.

### 6.3. Experiment 3

In this experiment, analytical and empirical performance of the proposed scheme is compared in a non-stationary environment. In steady state, we assume that the proposed scheme converges in the mean to a zero forcing solution, i.e., the mean of combined channel-equalizer response converges to a delta function with arbitrary delay and phase rotation. We consider a non-stationary channel with $\sigma_q = 10^{-3}$, where $\sigma_q$ is the standard deviation of the perturbation $\boldsymbol{q}_n$ in (29). The equalizer length is set to $N=11$ and the steady state EMSE is evaluated for 20,000 symbols in a noise-free environment. The forgetting factor is set to 0.98 and step size is varied to obtain the desired curves for both 8APSK(4,4) and 16APSK (8,8). A close match between the analytic and measured EMSE for both constellation types is observed as depicted in Fig. 8.

## 7. Conclusion

Exploiting the notion of minimum entropy deconvolution, a cost function was specifically derived for the blind channel equalization of amplitude phase shift keying signal in digital communication systems. The cost was optimized adaptively using Newton's method and it yielded a Newton-like constant modulus algorithm NL-$\beta$CMA. Tracking analysis of the proposed algorithm was performed based on Sayed-Rupp's feedback approach. Simulation results demonstrated significant improvement in performance compared with conventional algorithms. Further, observations of excess mean square error demonstrated good agreement between theoretical and practical findings.

## Appendix A. Derivation of $\mathcal{J}_{\mathrm{AN}}(\boldsymbol{y}_n)$

Consider a continuous APSK signal, where signal alphabets $\mathcal{A} = a_R + ja_I$ are uniformly distributed (theoretically) over a circular region of radius $\gamma$, with the center at the origin. The joint PDF of $a_R$ and $a_I$ is given by (refer to Fig. A1(a))

$$p_{\mathcal{A}}(y) = \begin{cases} \dfrac{1}{\pi\gamma^2}, & \sqrt{a_R^2 + a_I^2} = |y_n| \le \gamma, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1}$$

Now consider the transformation $\tilde{\mathcal{Y}} = |\mathcal{A}| = \sqrt{a_R^2 + a_I^2}$ and $\Theta = \angle(a_R, a_I)$, where $\tilde{\mathcal{Y}}$ is the modulus and $\angle(i,j)$ denotes the angle in the range $(0, 2\pi)$ that is defined by the point $(i,j)$. The joint distribution of the modulus $\mathcal{Y}$ and $\Theta$ can be obtained as $p_{\tilde{\mathcal{Y}},\Theta}(\tilde{y},\tilde{\theta}) = \tilde{y}/(\pi\gamma^2), \tilde{y} \ge 0, 0 \le \tilde{\theta} < 2\pi$. Since $\tilde{\mathcal{Y}}$ and $\Theta$ are independent, we obtain $p_{\tilde{\mathcal{Y}}}(\tilde{y} : H_0) = 2\tilde{y}/\gamma^2, \tilde{y} \ge 0$, where $H_0$ denotes the hypothesis that signal is distortion-free. Let $\tilde{\mathcal{Y}}_1, \tilde{\mathcal{Y}}_2, \ldots, \tilde{\mathcal{Y}}_B$ be a sequence, of size $B$, obtained by taking modulus of randomly generated distortion-free signal alphabets $\mathcal{A}$, where subscript $n$ indicates discrete time index. Let $\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_B$ be the order statistic of sequence $\{\tilde{\mathcal{Y}}\}$. Let $p_{\tilde{\mathcal{Y}}}(\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_B : H_0)$ be a $B$-variate density of the continuous

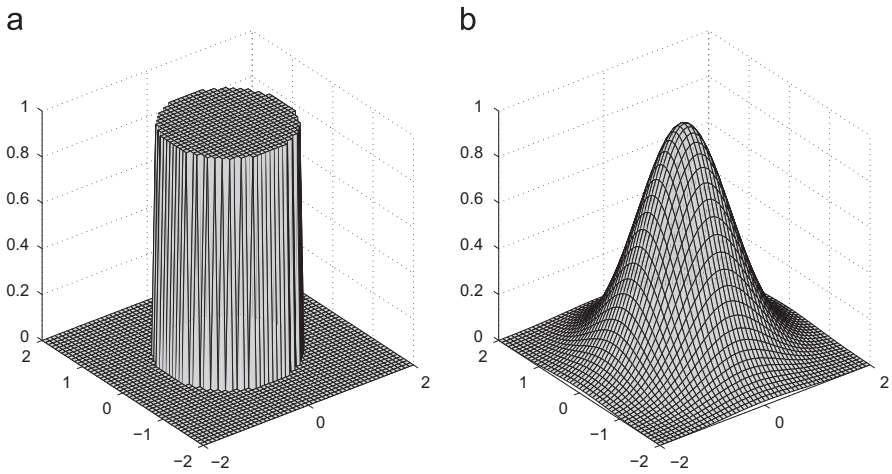


**Fig. A1.** PDFs (not to scale) of (a) hypothetical (dense) APSK and (b) Gaussian distributed received signal.

type, then, under the hypothesis $H_0$, we obtain

$$p_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_0) = \frac{2^B}{\gamma^{2B}} \prod_{k=1}^{B} \tilde{y}_k. \tag{A.2}$$

Next we find scale-invariant PDF $p^{\mathrm{si}}_{\tilde{y}}(\tilde{y}_1,\tilde{y}_2,\ldots,\tilde{y}_B : H_0)$ for given $B$ realizations of $\tilde{y}$ as follows (below $\alpha$ is some positive scale):

$$\begin{aligned} p^{\mathrm{si}}_{\tilde{y}}&(\tilde{y}_1,\ldots,\tilde{y}_B : H_0) \\ &= \int_0^{\infty} p_{\tilde{y}}(\alpha\tilde{y}_1,\ldots,\alpha\tilde{y}_B : H_0)\alpha^{B-1}\,\mathrm{d}\alpha \\ &= \frac{2^B}{\gamma^{2B}} \prod_{k=1}^{B} \tilde{y}_k \int_0^{Ra/(z_B-z_1)} \alpha^{2B-1}\,\mathrm{d}\alpha \\ &= \frac{2^{B-1}}{B(z_B-z_1)^{2B}} \prod_{k=1}^{B} \tilde{y}_k, \end{aligned} \tag{A.3}$$

where $z_1, z_2, \ldots, z_B$ are the order statistic of elements $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_B$, so that $z_1 = \min\{\tilde{y}\}$ and $z_B = \max\{\tilde{y}\}$. Now consider the next hypothesis ($H_1$) that the signal suffers from multi-path interference as well as with additive Gaussian noise (refer to Fig. A1(b)).

Thus, the in-phase and quadrature components of the received signal are modeled as normal-distributed; owing to central limit theorem, it is theoretically justified. It means that the modulus of the received signal follows the Rayleigh distribution,

$$p_{\tilde{y}}(\tilde{y} : H_1) = \frac{\tilde{y}}{\sigma^2_{\tilde{y}}} \exp\left(-\frac{\tilde{y}^2}{2\sigma^2_{\tilde{y}}}\right), \quad \tilde{y} \geq 0, \;\; \sigma_{\tilde{y}} > 0. \tag{A.4}$$

The $B$-variate densities $p_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_1)$ and $p^{\mathrm{si}}_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_1)$ are obtained as

$$p_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_1) = \frac{1}{\sigma^{2B}_{\tilde{y}}} \prod_{k=1}^{B} \tilde{y}_k \exp\left(-\frac{\tilde{y}_k^2}{2\sigma^2_{\tilde{y}}}\right), \tag{A.5a}$$

$$p^{\mathrm{si}}_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_1) = \frac{1}{\sigma^{2B}_{\tilde{y}}} \prod_{k=1}^{B} \tilde{y}_k \int_0^{\infty} \exp\left(-\frac{\alpha^2 \sum_{k'=1}^{B} \tilde{y}_{k'}^2}{2\sigma^2_{\tilde{y}}}\right) \alpha^{2B-1}\,\mathrm{d}\alpha. \tag{A.5b}$$

Substituting $u = \frac{1}{2}\alpha^2 \sigma_{\tilde{y}}^{-2} \sum_{k'=1}^{B} \tilde{y}_{k'}^2$, we obtain

$$p^{\mathrm{si}}_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_1) = \frac{2^B \Gamma(B) \prod_{k=1}^{B} \tilde{y}_k}{2(\sum_{k=1}^{B} \tilde{y}_k^2)^B} \tag{A.6}$$

The scale-invariant uniformly most powerful test of $p^{\mathrm{si}}_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_0)$ against $p^{\mathrm{si}}_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_1)$ provides us, see [20]:

$$\begin{aligned} O(\tilde{y}_1) &= \frac{p^{\mathrm{si}}_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_0)}{p^{\mathrm{si}}_{\tilde{y}}(\tilde{y}_1,\ldots,\tilde{y}_B : H_1)} = \frac{1}{B!}\left[\frac{\sum_{k=1}^{B} \tilde{y}_k^2}{(z_B - z_1)^2}\right]^B \\ &= \frac{1}{B!}\left[\frac{\sum_{n=1}^{B} |y_n|^2}{(\max\{|y_n|\} - \min\{|y_n|\})^2}\right]^B \end{aligned} \tag{A.7}$$

In the present context, where $y_n$ is the deconvolved sequence, we have $\min\{|y_n|\} = 0$. Further taking $B$th root of (A.7), ignoring constants and some manipulations, we get (14).

## Appendix B. Evaluation of $\beta$ for APSK

Expressing $y_n \approx a_n - e_{a,n}$, note that the amplitude $|a_n|$ is perturbed by $e_{a,n}$; since $e_{a,n}$ is assumed to be zero-mean (complex-valued) narrowband Gaussian, the modulus $|y_n|$ becomes Rician distributed and its PDF (conditioned on $|a_n|$) can be expressed as

$$p(|y_n|; |a_n|) = \frac{|y_n|}{\sigma^2} \exp\left(-\frac{|y_n|^2 + |a_n|^2}{2\sigma^2}\right) I_0\left(\frac{|y_n||a_n|}{\sigma^2}\right) \tag{B.1}$$

where $\sigma^2 := \mathsf{E}(\Re[e_a])^2 = \mathsf{E}(\Im[e_a])^2$, and $I_0$ is zeroth order modified Bessel function of first kind (where $\Re[\cdot]$ and $\Im[\cdot]$ refer to real and imaginary components of the enclosed complex-valued entity). Using (B.1), a $k$th-order moment of modulus $|y_n|$ can be computed as follows:

$$\mathsf{E}|y_n|^k = \sum_{j=1}^{L} \frac{M_j}{M} \int_0^{\infty} |y_n|^k p(|y_n|; R_j)\,\mathrm{d}|y_n| \tag{B.2}$$

Consider a distortion-free $M$-symbol complex-valued constellation $\{a\}$ which comprises $L$ number of unique moduli,

that is $|a_n| \in \{R_1, R_2, \ldots, R_L\}$, satisfying $0 < R_1 < R_2 < \cdots < R_L = \gamma$. Now let $M_j$ be the number of unique symbols on the $j$ th modulus $R_j$, this implies $\mathrm{E}|a_n|^2 = (1/M)\sum_{j=1}^{L} M_j R_j^2$ is the per-symbol average energy of constellation $\{a\}$.

$$\beta := \frac{\mathrm{E}|y_n|^2_{\{|y_n|<\gamma\}}}{\mathrm{E}|y_n|^2_{\{|y_n|\geq\gamma\}}} = \frac{\mathrm{E}|y_n|^2 - \mathrm{E}|y_n|^2_{\{|y_n|\geq\gamma\}}}{\mathrm{E}|y_n|^2_{\{|y_n|\geq\gamma\}}} \tag{B.3}$$

Using (B.1), we obtain $\mathrm{E}|y_n|^2 = 2\sigma^2 + \mathrm{E}|a_n|^2$ and $\mathrm{E}|y_n|^2_{\{|y_n|\geq\gamma\}} = \sigma^2 \mathrm{E} Q_{3,0}(|a_n|/\sigma, \gamma/\sigma)$, where $Q_{m,v}(a,b)$ is the Nuttall $Q$-function as defined below [39]:

$$Q_{m,v}(a,b) := \int_b^\infty x^m e^{-1/2(x^2+a^2)} I_v(ax)\,\mathrm{d}x, \quad a,b>0 \tag{B.4}$$

for $v \geq 0, m \geq 1$ and $I_v(\cdot)$ is the $v$th-order modified Bessel function of first kind. An approximate expression for $Q_{3,0}(a,b)$ is given as follows [47]:

$$Q_{3,0}(a,b) \approx \frac{2+(a+b)^2+b^2}{\sqrt{8\pi ab}} \exp\left(-\frac{(b-a)^2}{2}\right) + \frac{a(3+a^2)+b(1+a^2)}{4\sqrt{ab}} \mathrm{erfc}\left(\frac{b-a}{\sqrt{2}}\right). \tag{B.5}$$

Exploiting (B.5), we can evaluate $\beta$ under the limit $\sigma \to 0$. Denoting $z = (\gamma - |a_n|)/(\sqrt{2}\sigma)$, note that for $\gamma > |a_n|$, we have $\lim_{\sigma\to 0} \exp(-z^2) = \lim_{\sigma\to 0} \mathrm{erfc}(z) \to 0$. And for $\gamma = |a_n|$, we have $\lim_{\sigma\to 0} \exp(-z^2) = \lim_{\sigma\to 0} \mathrm{erfc}(z) \to 1$. So it is simple to show that:

$$\lim_{\sigma\to 0} \mathrm{E}|y_n|^2_{\{|y_n|\geq\gamma\}} = \lim_{\sigma\to 0} \sigma^2 \mathrm{E} Q_{3,0}\left(\frac{|a_n|}{\sigma}, \frac{\gamma}{\sigma}\right) \to \frac{M_L \gamma^2}{2M}. \tag{B.6}$$

Finally, the asymptotic value of $\beta$ for the APSK signal under the condition of vanishing convolutional noise of variance $2\sigma^2$ is given by

$$\lim_{\sigma\to 0} \beta \to \frac{2M\mathrm{E}|a_n|^2 - M_L\gamma^2}{M_L\gamma^2}. \tag{B.7}$$

## References


[1] S. Haykin, Blind Deconvolution, Snewblock Prentice Hall, New Jersey, 1994.
[2] C.R. Johnson Jr., P. Schniter, T.J. Endres, J.D. Behm, D.R. Brown, R.A. Casas, Blind equalization using the constant modulus criterion: a review, Proc. IEEE 86 (10) (1998) 1927–1950.
[3] Z. Ding, Y. Li, Blind Equalization and Identification, Marcel Dekker Inc., New York, 2001.
[4] D. Donoho, On minimum entropy deconvolution, in: Proceedings of 2nd Applied Time Series Symposium, March 1980, pp. 565–608.
[5] A. Benveniste, M. Goursat, G. Ruget, Robust identification of a nonminimum phase system: blind adjustment of a linear equalizer in data communication, IEEE Trans. Autom. Control 25 (3) (1980) 385–399.
[6] R.A. Wiggins, Minimum entropy deconvolution, in: Proceedings of International Symposium Computer Aided Seismic Analysis and Discrimination, 1977.
[7] R.A. Wiggins, Minimum entropy deconvolution, Geoexploration 16 (1978) 21–35.
[8] A. Morello, V. Mignone, DVB-S2: the second generation standard for satellite broad-band services, IEEE Proc. 94 (1) (2006) 210–227.
[9] R. De Gaudenzi, A. Guillén i Fàbregas, A. Martinez, Turbo-coded APSK modulations design for satellite broadband communications, Int. J. Satell. Commun. Netw. 24 (4) (2006) 261–281.
[10] C.-Y. Kao, M.-C. Tseng, C.-Y. Chen, The performance analysis of backward compatible modulation with higher spectrum efficiency for DAB Eureka 147, IEEE Trans. Broadcast. 54 (1) (2008) 62–69.
[11] C. Shaw, M. Rice, Turbo-coded APSK for aeronautical telemetry, IEEE Aerosp. Electron. Syst. Mag. 25 (4) (2010) 37–43.
[12] Z. Liu, Q. Xie, K. Peng, Z. Yang, APSK constellation with Gray mapping, IEEE Commun. Lett. 15 (12) (2011) 1271–1273.
[13] H. Wang, Y. Li, X. Yi, D. Kong, J. Wu, J. Lin, APSK modulated CO-OFDM system with increased tolerance toward fiber nonlinearity, IEEE Photon. Tech. Lett. 24 (13) (2012) 1085–1087.
[14] S. Fan, H. Wang, Y. Li, W. Du, X. Zhang, J. Wu, J. Lin, Optimal 16-Ary APSK encoded coherent optical OFDM for long-haul transmission, IEEE Photon. Tech. Lett. 25 (13) (2013) 1199–1202.
[15] O. Shalvi, E. Weinstein, New criteria for blind equalization of non-minimum phase systems, IEEE Trans. Inf. Theory 36 (2) (1990) 312–321.
[16] D.N. Godard, Self-recovering equalization and carrier tracking in two-dimensional data communications systems, IEEE Trans. Commun. 28 (11) (1980) 1867–1875.
[17] L.R. Litwin, Blind channel equalization, IEEE Potentials 18 (4) (1999) 9–12.
[18] J.R. Treichler, B.G. Agee, A new approach to multipath correction of constant modulus signals, IEEE Trans. Acoust. Speech Signal Process. 31 (2) (1983) 459–471.
[19] R.V. Hogg, More light on the kurtosis and related statistics, J. Am. Stat. Assoc. 67 (338) (1972) 422–424.
[20] Z. Sidak, P.K. Sen, J. Hajek, Theory of Rank Tests, 2 ed. Academic Press, London, UK, 1999.
[21] J.P. Marques de Sa, L.M.A. Silva, J.M.F. Santos, L.A. Alexandre, Minimum Error Entropy Classification, Springer-Verlag, Berlin, Heidelberg, 2013.
[22] R.C. Geary, Testing for normality, Biometrika 34 (3/4) (1947) S209–S242.
[23] A.T. Walden, Non-Gaussian reflectivity, entropy, and deconvolution, Geophysics 50 (12) (1985) 2862–2888.
[24] M. Ooe, T.J. Ulrych, Minimum entropy deconvolution with an exponential transformation, Geophysical Prospecting 27 (1979) 458–473.
[25] W. Gray, Variable norm deconvolution (Ph.D. thesis), Stan. Univ., 1979.
[26] E.H. Satorius, J.J. Mulligan, Minimum entropy deconvolution and blind equalisation, IEEE Electron. Lett. 28 (16) (1992) 1534–1535.
[27] E.H. Satorius, J.J. Mulligan, An alternative methodology for blind equalization, Digital Signal Process.: Rev. J. 3 (3) (1993) 199–209.
[28] S. Abrar, A.K. Nandi, Adaptive minimum entropy equalization algorithm, IEEE Commun. Lett. 14 (10) (2010) 966–968.
[29] S. Abrar, A. Zerguine, A.K. Nandi, Adaptive blind channel equalization, in: C. Palanisamy (Ed.), Digital Communication, InTech Publishers, Rijeka, Croatia, 2012, pp. 93–118.
[30] S. Haykin, Adaptive Filtering Theory, Prentice-Hall, Englewood Cliffs, NJ, 1996.
[31] A.K. Nandi, S. Abrar, Adaptive blind equalization based on the minimum entropy principle, in: 5th International Conference on Computers and Devices for Communication (CODEC), December 2012.
[32] G. Yan, H.H. Fan, A Newton-like algorithm for complex variables with applications in blind equalization, IEEE Trans. Signal Process. 48 (2) (2000) 553–556.
[33] S. Abrar, A. Zerguine, Enhancing the convergence speed of a multi-modulus blind equalization algorithm, in: IEEE SCONEST, 2004, pp. 41–44.
[34] M. Rupp, A.H. Sayed, A time-domain feedback analysis of filtered-error adaptive gradient algorithms, IEEE Trans. Signal Process. 44 (6) (1996) 1428–1439.
[35] N.R. Yousef, A.H. Sayed, A feedback analysis of the tracking performance of blind adaptive equalization algorithms, in: IEEE Conference on Decision and Control, vol. 1, 1999, pp. 174–179.
[36] A.H. Sayed, T.Y. Al-Naffouri, Mean-square analysis of normalized leaky adaptive filters, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2001.
[37] X. Wang, G. Feng, Performance analysis of RLS linearly constrained constant modulus algorithm for multiuser detection, Signal Process. 89 (2009) 181–186.
[38] A.H. Sayed, Fundamentals of Adaptive Filtering, Wiley-Interscience and IEEE Press, New York, 2003.
[39] A.H. Nuttall, Some integrals involving the $Q$-function, Naval Underwater Systems Center Technical Report 4297, April 1972.

[40] O. Tanrikulu, A.G. Constantinides, J.A. Chambers, New normalized constant modulus algorithms with relaxation, IEEE Signal Process. Lett. 4 (9) (1997) 256–258.
[41] V. Shtrom, H.H. Fan, New class of zero-forcing cost functions in blind equalization, IEEE Trans. Signal Process. 46 (10) (1998) 2674.
[42] F.R.P. Cavalcanti, A.L. Brandao, J.M.T. Romano, The generalized constant modulus algorithm applied to multiuser space-time equalization, in: Proceedings of IEEE-SPAWC, 1999, pp. 94–97.
[43] C. Yuxin, T. Le-Ngoc, B. Champagne, X. Changjiang, Recursive least squares constant modulus algorithm for blind adaptive array, IEEE Trans. Signal Process. 52 (May (5)) (2004) 1452–1456.
[44] M. Lazaro, I. Santamaria, D. Erdogmus, K.E. Hild, C. Pantaleon, J.C. Principe, Stochastic blind equalization based on PDF fitting using parzen estimator, IEEE Trans. Signal Process. 53 (2) (2005) 696–704.
[45] M.D. Miranda, M.T.M. Silva, V.H. Nascimento, Avoiding divergence in the Shalvi–Weinstein algorithm, IEEE Trans. Signal Process. 56 (11) (2008) 5403–5413.
[46] G. Picchi, G. Prati, Blind equalization and carrier recovery using a 'stop-and-go' decision-directed algorithm, IEEE Trans. Commun. 35 (9) (1987) 877–887.
[47] S. Abrar, A. Ali, A. Zerguine, A.K. Nandi, Tracking performance of two constant modulus equalizers, IEEE Commun. Lett. 17 (5) (2013) 830–833.