ADAPTIVE NEWTON ALGORITHMS FOR BLIND EQUALIZATION USING THE GENERALIZED CONSTANT MODULUS CRITERION

Wen-Jun Zeng, Xi-Lin Li, and Xian-Da Zhang

Department of Automation, Tsinghua University, Beijing 100084, China

ABSTRACT

Two Newton-type algorithms using the generalized complex modulus (GCM) criterion for blind equalization and carrier phase recovery are proposed. First the partial Hessian and full Hessian of the real GCM loss function with complex valued arguments are calculated by second-order differential. Then an adaptive pseudo Newton learning algorithm and a full Newton learning algorithm are designed. By using the matrix inversion lemma, both Newton algorithms can be implemented with a computational complexity of $\mathcal{O}(L^2)$ efficiently, where L is the length of equalizer. Simulation results demonstrate that the two Newton algorithms can achieve automatic carrier phase recovery and exhibit fast convergence rates.

Index Terms— Blind equalization, generalized constant modulus, Newton algorithm, adaptive signal processing.

1. INTRODUCTION

The constant modulus algorithm (CMA) is widely used for blind equalization of two dimensional modulation schemes [1]. Stochastic gradient descent algorithms are commonly used to minimize the constant modulus (CM) loss function. However, the stochastic gradient method suffers from slow convergence, which constitutes a significant drawback. It is well known that the second-order Newton method has fast convergence. Recently, several Newton-like algorithms minimizing the CM loss function are proposed [2][3]. Unfortunately, the full Hessian of CM loss function at the minima are intrinsically singular [1][3]. This intrinsic singularity prevents one from using the Newton algorithm without precautionary modifications [3].

The other flaw is that the CMA lacks the capability of recovering the carrier phase. To recover the carrier phase, an additional rotator is required at the output of the CMA equalizer, which results in increasing the complexity of implementation of the receiver in steady-state operation. In [4], a family of generalized CMAs (GCMA) was proposed by generalizing the definition of complex modulus. Since the generalized complex modulus (GCM) is sensitive to phase, the GCMA can achieve the phase recovery. However the GCMA is still based on stochastic gradient descent and exhibits poor convergence performance.

In this paper, two Newton-type algorithms are proposed to minimize the GCM loss function, which are referred to as pseudo Newton-GCM and full Newton-GCM. Simulation results validate that the two Newton-type algorithms can recovery the carrier phase and converge fast.

2. SYSTEM MODEL AND GENERALIZED CONSTANT MODULUS CRITERION

2.1. System Model

Consider a baseband communication system described by

$$x(n) = s(n) \otimes h(n) + v(n) \tag{1}$$

where x(n) is the observed sequence, s(n) is the transmitted data symbol, h(n) is the impulse response of the channel, v(n) is zero-mean additive Gaussian white noise (AGWN), and \otimes denotes convolution. The source sequence s(n) is assumed to be zero-mean, independently and identically distributed (i.i.d.). Define an L tap FIR equalizer with weights $\mathbf{w} = [w_0, w_1, \cdots, w_{L-1}]^T$ and a vector $\mathbf{x}(n) = [x(n), x(n-1), \cdots, x(n-L+1)]^T$ with the superscript T denoting transpose, the equalizer output sequence is given

$$y(n) = \mathbf{w}^H \mathbf{x}(n) \tag{2}$$

where the superscript H represents the conjugate transpose. The equalizer w is designed to eliminate the inter-symbol interference (ISI) and the aim of blind equalization is to recover the source sequence without training sequences. In the rest of this paper, when there is no possibility of confusion, the notation is simplified by ignoring the sample index n.

2.2. The Generalized Constant Modulus Criterion

The well known constant modulus loss function for blind equalization is given by [1]

$$J(\mathbf{w}) = E\left[\left(|y|^2 - R_2\right)^2\right] \tag{3}$$

where $|\cdot|$ denotes the modulus of a complex variable and $R_2 = \frac{E[|s(n)|^4]}{E[|s(n)|^2]}$ is the dispersion constant.

This work was supported by the National Natural Science Foundation of China under Grant 60675002.

The CMA seeks a equalizer that minimizes the CM loss function. The main drawback of the CMA is that it lacks the capability of carrier phase recovery. Such flaw is attributed to the complex modulus is insensitive to the phase, i.e., the property $|ze^{j\theta}| = |z|$ holds true for any phase angle θ .

Given a complex number $z = z_R + jz_I$, where z_R and z_I represent the real and imaginary parts of z, the generalized modulus of z is defined as [4]

$$|z|_{\ell} = \left(|z_R|^{\ell} + |z_I|^{\ell}\right)^{1/\ell}, \quad \ell \ge 1.$$
 (4)

When $\ell = 2$, the generalized complex modulus reduces to the standard complex modulus. Note that $|ze^{j\theta}|_{\ell} \neq |z|_{\ell}$ for $\ell \neq 2$, which means that the generalized modulus is phase sensitive. Based on this property, the generalized constant modulus criterion which is able to achieve carrier phase recovery for blind equalization is given by [4]

$$J_{\ell}(\mathbf{w}) = E\left[\left(|y|_{\ell}^{2} - R_{\ell}\right)^{2}\right] = E\left[\left(\left|\mathbf{w}^{H}\mathbf{x}\right|_{\ell}^{2} - R_{\ell}\right)^{2}\right]$$
(5)

where R_{ℓ} is the dispersion constant of the GCMA depending on ℓ and the constellation of the source, and can be expressed as

$$R_{\ell} = \frac{E\left[|s(n)|_{\ell}^{4}\right]}{E\left[|s(n)|_{\ell}^{2}\right]}.$$
(6)

3. NEWTON ALGORITHMS FOR THE GENERALIZED CONSTANT MODULUS CRITERION

In [4], a gradient learning algorithm minimizing GCM criterion is given with $\ell = 4$, which is referred to as extended CMA (ECMA). However, the gradient-based learning algorithms suffer from slow convergence. The convergence rate and tracking capacity is very important for practical communication systems. In this paper, we will develop gradient learning algorithms as well as adaptive Newton algorithms minimizing GCM criterion for any value of ℓ .

3.1. Gradient and Hessian of GCM Loss Function

First let us define

$$a(n) = y_R(n) |y_R(n)|^{\ell-2} - jy_I(n) |y_I(n)|^{\ell-2}$$
(7)

and a(n) is abbreviated as a for convenience. The first-order differential of the GCM loss function can be derived as

$$dJ_{\ell}(\mathbf{w}) = 2E\left[|y|_{\ell}^{2-\ell} \left(|y|_{\ell}^{2} - R_{l}\right) \left(a\mathbf{x}^{T} d\mathbf{w}^{*} + a^{*} \mathbf{x}^{H} d\mathbf{w}\right)\right]$$

where the superscript * denotes the conjugate of a complex number or vector. Hence the conjugate gradient and gradient of $J_{\ell}(\mathbf{w})$ are given by

$$\nabla_{\mathbf{w}} J_{\ell}(\mathbf{w}) = \frac{\partial J_{\ell}(\mathbf{w})}{\partial \mathbf{w}^{*}} = 2E \left[|y|_{\ell}^{2-\ell} \left(|y|_{\ell}^{2} - R_{l} \right) a\mathbf{x} \right] \quad (8)$$

$$\nabla_{\mathbf{w}^*} J_{\ell}(\mathbf{w}) = \frac{\partial J_{\ell}(\mathbf{w})}{\partial \mathbf{w}} = 2E \left[\left| y \right|_{\ell}^{2-\ell} \left(\left| y \right|_{\ell}^2 - R_l \right) a^* \mathbf{x}^* \right].$$
(9)

It is clear that $\nabla_{\mathbf{w}} J_{\ell}(\mathbf{w}) = (\nabla_{\mathbf{w}^*} J_{\ell}(\mathbf{w}))^*$. For convenience we define

$$b_R(n) = |y_R(n)|^{\ell-2} + (\ell-2)y_R^2(n)|y_R(n)|^{\ell-4}$$
(10)

$$b_I(n) = |y_I(n)|^{\ell-2} + (\ell-2)y_I^2(n) |y_I(n)|^{\ell-4}.$$
 (11)

Sometimes $b_R(n)$ and $b_I(n)$ are abbreviated as b_R and b_I , respectively. We denote

$$f_{\ell}(n) = |y|_{\ell}^{2-2\ell} \left((4-\ell) |y|_{\ell}^{2} - (2-\ell)R_{l} \right) |a|^{2} + \left(|y|_{\ell}^{2} - R_{l} \right) |y|_{\ell}^{2-\ell} (b_{R} - b_{I})$$
(12)
$$g_{\ell}(n) = |y|_{\ell}^{2-2\ell} \left((4-\ell) |y|_{\ell}^{2} - (2-\ell)R_{l} \right) a^{2} + \left(|y|_{\ell}^{2} - R_{l} \right) |y|_{\ell}^{2-\ell} (b_{R} + b_{I}).$$
(13)

Again, we use $f_{\ell}(n)$ and $g_{\ell}(n)$ to denote f_{ℓ} and g_{ℓ} for convenience, respectively. Note that f_{ℓ} is a real number. By calculating the second-order differential of $J_{\ell}(\mathbf{w})$, we can get the four $L \times L$ partial Hessian matrices:

$$\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) = \frac{\partial^2 J_{\ell}(\mathbf{w})}{\partial \mathbf{w}^* \partial \mathbf{w}^T} = E\left[f_{\ell} \mathbf{x} \mathbf{x}^H\right]$$
(14)

$$\mathcal{H}_{\mathbf{w}\mathbf{w}^*}(\mathbf{w}) = \frac{\partial^2 J_{\ell}(\mathbf{w})}{\partial \mathbf{w}^* \partial \mathbf{w}^H} = E\left[g_{\ell} \mathbf{x} \mathbf{x}^T\right]$$
(15)

$$\mathcal{H}_{\mathbf{w}^*\mathbf{w}}(\mathbf{w}) = \frac{\partial^2 J_{\ell}(\mathbf{w})}{\partial \mathbf{w} \partial \mathbf{w}^T} = E\left[g_{\ell}^* \mathbf{x}^* \mathbf{x}^H\right]$$
(16)

$$\mathcal{H}_{\mathbf{w}^*\mathbf{w}^*}(\mathbf{w}) = \frac{\partial^2 J_{\ell}(\mathbf{w})}{\partial \mathbf{w} \partial \mathbf{w}^H} = E\left[f_{\ell}^* \mathbf{x}^* \mathbf{x}^T\right]$$
(17)

 $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ is referred to as the *leading partial Hessian*. It is clear that $\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) = (\mathcal{H}_{\mathbf{w}^*\mathbf{w}^*}(\mathbf{w}))^*$ and $\mathcal{H}_{\mathbf{ww}^*}(\mathbf{w}) = (\mathcal{H}_{\mathbf{w}^*\mathbf{w}}(\mathbf{w}))^*$. The $2L \times 2L$ full Hessian matrix consisting of the four partial Hessian matrices is given by

$$\mathcal{H}(\mathbf{w}) = \begin{bmatrix} \mathcal{H}_{\mathbf{w}\mathbf{w}}(\mathbf{w}) & \mathcal{H}_{\mathbf{w}\mathbf{w}^*}(\mathbf{w}) \\ \mathcal{H}_{\mathbf{w}^*\mathbf{w}}(\mathbf{w}) & \mathcal{H}_{\mathbf{w}^*\mathbf{w}^*}(\mathbf{w}) \end{bmatrix}.$$
 (18)

3.2. The Gradient Learning Algorithm

According to (8), the stochastic gradient can be written as

$$\hat{\nabla}_{\mathbf{w}} J_{\ell}(\mathbf{w}, n) = |y(n)|_{\ell}^{2-\ell} \left(|y(n)|_{\ell}^2 - R_l \right) a(n) \mathbf{x}(n).$$
(19)

Hence we obtain the stochastic gradient learning algorithm of the GCM criterion

$$\Delta \mathbf{w} = -\mu \hat{\nabla}_{\mathbf{w}} J_{\ell}(\mathbf{w}, n) \tag{20}$$

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \Delta \mathbf{w} \tag{21}$$

where μ is the step size. Its computational complexity is $\mathcal{O}(L)$ per iteration. Although the stochastic gradient learning algorithm is quite popular due to its simplicity, it suffers from slow convergence.

3.3. The Pseudo Newton Learning Algorithm

The Newton algorithm exploits the second-order derivative of the loss function to provide faster convergence. If the leading partial Hessian $\mathcal{H}_{ww}(w)$ is adopted only, we can obtain the pseudo Newton learning rule for GCM criterion (pseudo Newton-GCM). The updated term Δw in the pseudo Newton algorithm is given by

$$\Delta \mathbf{w} = -\mu \mathcal{H}_{\mathbf{w}\mathbf{w}}^{-1}(\mathbf{w}) \hat{\nabla}_{\mathbf{w}} J_{\ell}(\mathbf{w}, n)$$
(22)

where $0 < \mu < 1$ is the step size. Directly computing the inverse of $\mathcal{H}_{ww}(w)$ will lead to a computational complexity of $\mathcal{O}(L^3)$ per iteration. According to (14), we obtain the rank-1 updating structure of the leading partial Hessian $\mathcal{H}_{ww}(w)$:

$$\mathcal{H}_{\mathbf{ww}}(\mathbf{w}, n) = \lambda \mathcal{H}_{\mathbf{ww}}(\mathbf{w}, n-1) + (1-\lambda) f_{\ell}(n) \mathbf{x}(n) \mathbf{x}^{H}(n)$$
(23)

where $0 < \lambda < 1$ is the forgetting factor. We can exploit such rank-1 updating structure to reduce the computational complexity. By using the well known matrix inversion lemma, we obtain the recursive equation of $\mathcal{H}_{\mathbf{w}\mathbf{w}}^{-1}(\mathbf{w}, n)$:

$$\mathcal{H}_{\mathbf{w}\mathbf{w}}^{-1}(\mathbf{w},n) = \lambda^{-1} \mathcal{H}_{\mathbf{w}\mathbf{w}}^{-1}(\mathbf{w},n-1) - \frac{\lambda^{-2} \mathcal{H}_{\mathbf{w}\mathbf{w}}^{-1}(\mathbf{w},n-1) \mathbf{x}(n) \mathbf{x}^{H}(n) \mathcal{H}_{\mathbf{w}\mathbf{w}}^{-1}(\mathbf{w},n-1)}{\left[(1-\lambda)f_{\ell}(n)\right]^{-1} + \lambda^{-1} \mathbf{x}^{H}(n) \mathcal{H}_{\mathbf{w}\mathbf{w}}^{-1}(\mathbf{w},n-1) \mathbf{x}(n)}.$$
(24)

 $\mathcal{H}_{\mathbf{ww}}^{-1}(\mathbf{w})$ is initialized as $\mathcal{H}_{\mathbf{ww}}^{-1}(\mathbf{w}, 0) = \delta \mathbf{I}$, where \mathbf{I} is the identify matrix, and $\delta > 0$ is small enough so that the learning algorithm has reliable initial convergence. It is clear that the computational complexity has been reduced to $\mathcal{O}(L^2)$ per iteration through the recursive computation of $\mathcal{H}_{\mathbf{ww}}^{-1}(\mathbf{w}, n)$.

3.4. The Full Newton Learning Algorithm

Unlike the pseudo Newton algorithm only utilizing the leading partial Hessian, the full Newton algorithm adopts all of the four partial Hessian matrices. In other words, the full Newton algorithm exploits the full Hessian. The full Newton learning rule for GCM criterion (full Newton-GCM) is

$$\begin{bmatrix} \Delta \mathbf{w} \\ \Delta \mathbf{w}^* \end{bmatrix} = -\mu \mathcal{H}^{-1}(\mathbf{w}) \begin{bmatrix} \hat{\nabla}_{\mathbf{w}} J_{\ell}(\mathbf{w}, n) \\ \hat{\nabla}_{\mathbf{w}^*} J_{\ell}(\mathbf{w}, n) \end{bmatrix}$$
(25)

where $\mathcal{H}(\mathbf{w})$ is the full Hessian. According to (14), (15), (16) and (17), we obtain the following recursion for updating the value of the full Hessian matrix:

$$\mathcal{H}(\mathbf{w},n) = \lambda \mathcal{H}(\mathbf{w},n-1) + (1-\lambda) \begin{bmatrix} f_{\ell}(n)\mathbf{x}(n)\mathbf{x}^{H}(n) & g_{\ell}(n)\mathbf{x}(n)\mathbf{x}^{T}(n) \\ g_{\ell}^{*}(n)\mathbf{x}^{*}(n)\mathbf{x}^{H}(n) & f_{\ell}^{*}(n)\mathbf{x}^{*}(n)\mathbf{x}^{T}(n) \end{bmatrix}.$$
(26)

Eq. (27) at the top of the next page means that the full Hessian have the rank-2 update structure. Hence the inverse of the full Hessian matrix can be updated by using the matrix inversion

lemma *twice*. The recursive equation for $\mathcal{H}^{-1}(\mathbf{w}, n)$ is given by (28) and (29). $\mathcal{H}^{-1}(\mathbf{w}, 0)$ is initialized as $\delta \mathbf{I}$ with $\delta > 0$ being small enough. According to (28) and (29), the computational complexity of the full Newton algorithm is $\mathcal{O}(L^2)$ per iteration.

Two important parameters in the Newton algorithms are the step size μ and forgetting factor λ . Although the two parameters can be selected independently, we would like to proposed a rule of thumb for the selection of μ and λ . Both the inverse of μ and the inverse of $1 - \lambda$ are, roughly speaking, measures of the memory of the algorithms. We hope that the two measures of memory equal. Thus we get a rule for the selection of μ and λ :

$$\mu + \lambda = 1. \tag{30}$$

4. SIMULATION RESULTS

The widely used performance index of an equalizer is the ISI (inter-symbol interference) defined as

$$\mathrm{ISI}(\mathrm{dB}) = 10 \log_{10} \left(\frac{\sum_{i} |\eta_{i}|^{2} - \max|\eta_{i}|^{2}}{\max|\eta_{i}|^{2}} \right)$$

where $\eta_n = h_n \otimes w_n$ is the combined channel-equalizer response. However, the conventional ISI cannot reflect the intra-interference between the in-phase component and the quadrature component caused by the carrier phase error. To depict both the inter-symbol interference and intra-symbol interference due to carrier phase error, the ISI performance index is modified as

Modified ISI(dB) =
$$10\log_{10}\left(\frac{\sum_{i}\gamma_{i}^{2} - \max\gamma_{i}^{2}}{\max\gamma_{i}^{2}}\right)$$

where $\boldsymbol{\gamma} = \begin{bmatrix} \operatorname{Re}(\boldsymbol{\eta}) \\ \operatorname{Im}(\boldsymbol{\eta}) \end{bmatrix}$ is a real vector with $\boldsymbol{\eta} = [\eta_{0}, \eta_{1}, \cdots]^{T}$.

A typical FIR voice-band communication channel with impulse response $\{-0.005-j0.004, 0.009+j0.03, -0.024-j0.104, 0.854+j0.52, -0.218+j0.273, 0.04-j0.0749, -0.016+j0.02\}$ is adopted [4]. To achieve phase recovery, we set $\ell = 4$. The transient behavior and steady-state performance of ECMA [4], pseudo Newton-GCM and full Newton-GCM are studied.

The length of the equalizer is set to L = 9. The input data symbol comes from a QAM-16 constellation. The white Gaussian noise is added so that the SNR is 30 dB. To obtain the same steady-state performance, the step sizes are set to $\mu = 1.2 \times 10^{-5}$ for ECMA, and $\mu = 0.002$ for pseudo Newton-GCM and full Newton-GCM. The forgetting factor and the parameter of the initial inverse of Hessian of both Newton algorithms are set to $\lambda = 1 - \mu$ and $\delta = 10^{-2}$, respectively.

Fig.1 shows the equalized data symbols of CMA, Newtonlike CMA [2], ECMA, pseudo Newton-GCM and full Newton-

$$\begin{bmatrix} f_{\ell}(n)\mathbf{x}(n)\mathbf{x}^{H}(n) & g_{\ell}(n)\mathbf{x}(n)\mathbf{x}^{T}(n) \\ g_{\ell}^{*}(n)\mathbf{x}^{*}(n)\mathbf{x}^{H}(n) & f_{\ell}^{*}(n)\mathbf{x}^{*}(n)\mathbf{x}^{T}(n) \end{bmatrix} = \alpha(n) \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{x}^{*}(n) \end{bmatrix} \begin{bmatrix} \mathbf{x}^{H}(n), \mathbf{x}^{T}(n) \end{bmatrix} + \beta(n) \begin{bmatrix} \mathbf{x}(n) \\ -\mathbf{x}^{*}(n) \end{bmatrix} \begin{bmatrix} \mathbf{x}^{H}(n), -\mathbf{x}^{T}(n) \end{bmatrix}$$

with $\alpha(n) = \frac{1}{2} [f_{\ell}(n) + g_{\ell}(n)], \quad \beta(n) = \frac{1}{2} [f_{\ell}(n) - g_{\ell}(n)].$ (27)

$$\mathcal{G}^{-1}(\mathbf{w},n) = \lambda^{-1} \mathcal{H}^{-1}(\mathbf{w},n-1) - \frac{\lambda^{-2} \mathcal{H}^{-1}(\mathbf{w},n-1) \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{x}^*(n) \end{bmatrix} [\mathbf{x}^H(n), \mathbf{x}^T(n)] \mathcal{H}^{-1}(\mathbf{w},n-1)}{[(1-\lambda)\alpha(n)]^{-1} + \lambda^{-1} [\mathbf{x}^H(n), \mathbf{x}^T(n)] \mathcal{H}^{-1}(\mathbf{w},n-1) \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{x}^*(n) \end{bmatrix}}$$
(28)
$$\mathcal{H}^{-1}(\mathbf{w},n) = \mathcal{G}^{-1}(\mathbf{w},n) - \frac{\mathcal{G}^{-1}(\mathbf{w},n) \begin{bmatrix} \mathbf{x}(n) \\ -\mathbf{x}^*(n) \end{bmatrix} [\mathbf{x}^H(n), -\mathbf{x}^T(n)] \mathcal{G}^{-1}(\mathbf{w},n)}{[(1-\lambda)\beta(n)]^{-1} + [\mathbf{x}^H(n), -\mathbf{x}^T(n)] \mathcal{G}^{-1}(\mathbf{w},n) \begin{bmatrix} \mathbf{x}(n) \\ -\mathbf{x}^*(n) \end{bmatrix}}$$
(29)



Fig. 1. Constellation of signals at steady-state.

GCM at steady state. One can find that both CMA and Newtonlike CMA converge to rotated solutions, while the three methods using GCM criterion can achieve phase recovery. Fig.2 illustrates the learning curve of the ensemble-averaged modified ISI over 200 independent trials at SNR=30 dB. Fig.2 validates that the full Newton-GCM has the fastest convergence rate, while the gradient-based ECMA has the slowest convergence rate. It is worthy to note that the full Newton-GCM performs much better than the pseudo Newton-GCM.

5. CONCLUSION

To overcome the two drawbacks of CMA, two Newton algorithms which are referred to as the pseudo Newton-GCM the full Newton-GCM are presented. The pseudo Newton-GCM uses the leading partial Hessian, while the full Newton-GCM uses the full Hessian of the GCM loss function. The two proposed Newton algorithms are able to recover the carrier phase and exhibit much faster convergence rates than the ECMA.



Fig. 2. Learning curve of the modified ISI

Moreover, both Newton algorithms have efficient online recursive forms with a computational complexity of $O(L^2)$, where L is the length of the equalizer.

6. REFERENCES

- C. R. Johnson, etal, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927-1950, 1998.
- [2] G. Yan and H. Fan, "A newton-like algorithm for complex variables with applications in blind equalization," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 553-556, 2000.
- [3] K. K-Delgado and Y. Isukapalli, "Use of the Newton method for blind adaptive equalization based on the constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3983-3995, 2008.
- [4] X.-L. Li and X.-D. Zhang, "A family of generalized constant modulus algorithm for blind equalization," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 1913-1917, 2006.