Use of the Newton Method for Blind Adaptive Equalization Based on the Constant Modulus Algorithm

Kenneth Kreutz-Delgado, *Senior Member, IEEE* and Yogananda Isukapalli, *Student Member, IEEE*

***Abstract*—We study the applicability of the second-order Newton gradient descent method for blind equalization of complex signals based on the Constant Modulus algorithm. The Constant Modulus loss function is real with complex valued arguments, and hence nonanalytic. We therefore use the framework of the *Wirtinger* calculus to derive a useful and insightful form of the Hessian for noiseless FIR channels and rederive the known fact that the full Hessian of the Constant Modulus loss function is always singular in a simpler manner. For the implementation of a sub-optimum version of Newton algorithm, we give the conditions under which the leading partial Hessian is non-singular for a noiseless FIR channel model. For this channel model we show that the perfectly equalizing solutions are stationary points of the Constant Modulus loss function and also evaluate the leading partial Hessian and the full Hessian at a perfectly equalizing solution. We also discuss regularization of the full Newton method. Finally, some simulation results are given.**

Keywords: Constant modulus algorithm, blind equalization, complex Newton method, analytic functions, *Wirtinger* calculus, complex Hessian, regularization.

I. Introduction

For finite length FIR channel models it is common to use equalizers with a tapped delay line structure. The taps of the equalizer are updated using adaptive algorithms, which usually depend on the presence of training symbols. Blind equalizers, originally proposed by Sato [1], do not require training symbols, thus potentially improving the information spectral efficiency. A well-known candidate among the blind adaptive algorithms is the Constant Modulus algorithm (CMA) proposed by Godard [2] for two-dimensional digital communication systems (M-ary quadrature amplitude modulation (M-QAM) signals), and by Treichler and Agee [3], for blind equalization of pulse-amplitude modulation (PAM) signals. CMA is a preferred choice for blind equalization because of its robustness and its ease of implementation [4]. Convergence analysis of CMA can be found in [5] and [6]. The speed of convergence is very important for practical systems. It is well-known that the second-order Newton method has fast convergence. However the Newton method requires the computation of the Hessian of the cost function for its implementation. We emphasize that we are working with complex signals as this makes the problem of understanding second-order CMA algorithms significantly more difficult than the real case. In particular, unlike the real case, there are no isolated local minima of the Constant Modulus (CM) loss function in the complex situation and as a consequence the Hessian of the CM loss function is always singular, which means that the fast converging Newton method can not be used in practice without modification. We theoretically show that this intrinsic singularity prevents one from using the Newton algorithm without precautionary modifications. This is accomplished by using the framework and methodology developed in [11] to analyze the Newton method applied to minimizing the CM loss function. Some modified versions of the Newton method as applied to constant modulus algorithm can be found in [7]-[10]. We now briefly summarize the contributions of this paper.

1) The full Hessian of the CM loss function is proven to be intrinsically singular. This fact is actually well known [4], [14], but the proof given in this paper is much simpler and is based on the complex Hessian derived using the framework of Wirtinger calculus.
2) We show that the perfectly equalizing solutions are stationary points of the CM loss function. We compute the form of the Hessian for the noiseless FIR channel model and evaluate it at a perfectly equalizing solution.
3) Conditions are given which ensure that the leading partial Hessian is positive definite, and necessarily full rank.
4) We show that at a perfectly equalizing solution for a noiseless full rank channel, the leading partial Hessian is full rank assuming that the channel input data sequence in sufficiently subgaussian.
5) If the leading partial Hessian is full rank, then the necessarily singular $2N \times 2N$ full Hessian is shown to attain the maximal rank of $2N-1$.
6) We give the form of the vector which spans the one dimensional nullspace of of a maximal rank full Hessian at a perfectly equalizing solution for a noiseless FIR channel.
7) We present a novel phase-enforcing regularization of the full Newton algorithm and discuss its implication for developing practical Newton algorithms as applied to the CMA.

The rest of this paper is organized as follows. In Section II, we introduce our system model. In Section III, we briefly explain the steps in the construction of the Hessian for real valued functions with complex valued arguments and construct

Authors are with the department of electrical and computer engineering at the university of california, san diego. Authors e-mail: kreutz@ece.ucsd.edu, yoga@ucsd.edu. This research was supported in part by the U. S. Army Research Office under the Multi-University Research Initiative (MURI) grant-W911NF-04-1-0224.

the gradients and Hessians for the CM loss function. Stationary points of the CM loss functions are studied in Section IV. The Newton method and a sub-optimum version of it are applied to the CM loss function in Section V. In Section VI, conditions are given for the equalization of a noiseless FIR channel model. Regularization of the maximal-rank Newton method is studied in Section VII. Simulation results are presented in Section VIII, and we conclude the paper in Section IX.

II. System Model

Zero-mean, non-zero bounded fourth-order moment i.i.d signals, $a(n)$, are assumed to be noiselessly transmitted through an unknown linear time invariant channel resulting in a received zero-mean output sequence $x(n)$, which is processed via an N tap FIR equalizer with weights $\mathbf{w}$. The equalizer output sequence is given by

$$y(n) = \sum_{k=0}^{N-1} \bar{w}(k)x(n-k) \triangleq \mathbf{w}^H \mathbf{x}(n). \tag{1}$$

Perfect equalization would yield $y(n) = a(n-\delta)$, for some integer-valued delay $0 \le \delta \le N-1$. Conditions under which perfect equalization of a noiseless fractionally spaced N-tap FIR channel is possible are given in [4]. Similar to [4] we also consider fractionally spaced FIR channels. When there is no possibility of confusion, the notation is simplified by suppressing the sample index n, $y = y(n) \in \mathbb{C}$ and $\mathbf{x} = \mathbf{x}(n) \in \mathbb{C}^N$, so that the length-$N$ FIR equalizer can be written in the simple form

$$y = \mathbf{w}^H \mathbf{x}. \tag{2}$$

We also define

$$\mu_2 \;\; \triangleq \;\; \mathrm{E}\left\{|\, a(n)|^2\right\},\ \mu_4 \triangleq \mathrm{E}\left\{|\, a(n)|^4\right\},\ \kappa \triangleq \frac{\mu_4}{\mu_2^2}.$$

Finally, $|\, y|^2 = \mathbf{w}^H \left(\mathbf{x}\mathbf{x}^H\right) \mathbf{w}$. The CM loss function is given by

$$\ell(\mathbf{w}) = \mathrm{E}\left\{\left(|\, y|^2 - R_2\right)^2\right\},$$

where R_2 is the dispersion constant [2],

$$R_2 \triangleq \frac{\mathrm{E}\left\{|\, a(n)|^4\right\}}{\mathrm{E}\left\{|\, a(n)|^2\right\}} = \frac{\mu_4}{\mu_2} = \mu_2 \kappa = \rho\,. \tag{3}$$

The Constant Modulus loss function can now be written as

$$\ell(\mathbf{w}) = \mathrm{E}\left\{\left(\mathbf{w}^H \mathbf{x}\mathbf{x}^H \mathbf{w} - \rho\right)^2\right\}. \tag{4}$$

To blindly identify the FIR equalizer filter coefficients $\hat{\mathbf{w}}$, the tap weights are chosen to minimize $\ell(\mathbf{w})$,

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathrm{E}\left\{\left(\mathbf{w}^H \mathbf{x}\mathbf{x}^H \mathbf{w} - \rho\right)^2\right\}.$$

The CM loss function (4) is smooth and bounded from below as a function of $\mathbf{w}$, so it is amenable to solution by use of the Newton method.

III. The Hessian of a Real-Valued Function on $\mathbb{C}^N$

In this section we briefly summarize concepts useful for deriving and analyzing the Hessian of a real-valued function on $\mathbb{C}^N$. All real-valued functions, including the CM loss function, can be viewed as functions of both $\mathbf{z}$ and its complex conjugate $\bar{\mathbf{z}}$. The Cauchy-Riemann condition, required for complex differentiation is not satisfied by the mapping $\mathbf{z} \to \bar{\mathbf{z}}$, thus making all real-valued functions on $\mathbb{C}^N$ non-holomorphic. Nonconstant nonholomorphic functions can be written in the form $f(\mathbf{z}, \bar{\mathbf{z}})$, where they are holomorphic in $\mathbf{z} = \mathbf{x} \;+ j\,\mathbf{y}$ for fixed $\bar{\mathbf{z}}$ and holomorphic in $\bar{\mathbf{z}} = \mathbf{x} - j\,\mathbf{y}$ for fixed $\mathbf{z}$ [16]. It can be shown that this fact is true in general for any complex- or real-valued function

$$f(\mathbf{z}) = f(\mathbf{z}, \bar{\mathbf{z}}) = f(\mathbf{x}, \mathbf{y}) = u(\mathbf{x}, \mathbf{y}) + j\, v(\mathbf{x}, \mathbf{y}). \tag{5}$$

This fact underlies the development of the so-called Wirtinger calculus [16]. In [11] a second order complex derivative is developed for the Wirtinger calculus based on the results given in [12] and [13], making use of the fact that there are three vector space coordinate representations for representing complex vectors $\mathbf{z} = \mathbf{x} + j\,\mathbf{y}$. The first is the canonical n-dimensional vector space of complex vectors $\mathbf{z} \in \mathcal{Z} = \mathbb{C}^n$. The second is the canonical $2n$-dimensional real vector space of vectors $\mathbf{r} = \mathrm{col}(\mathbf{x}, \mathbf{y}) \in \mathcal{V} = \mathbb{R}^{2n}$, which arises from the natural correspondence $\mathbb{C}^n \approx \mathbb{R}^{2n}$. The third is a $2n$-dimensional *real* vector space of $\mathbf{c} \in \mathcal{C}$

$$\mathbf{c} = \begin{pmatrix} \mathbf{z} \\ \bar{\mathbf{z}} \end{pmatrix} \in \mathcal{C} \subset \mathbb{C}^{2n} \approx \mathbb{R}^{2n}\,.$$

In the third representation there are actually *two* alternative interpretations of $\mathbf{c} \in \mathcal{C} \subset \mathbb{C}^{2n}$ ($\mathbf{c}$ as belonging to a complex manifold $\mathbb{C}^{2n}$): the $\mathbf{c}$-real case for when we consider the vector $\mathbf{c} \in \mathcal{C} \approx \mathbb{R}^{2n}$ ($\mathbf{c}$ as belonging to a real vector space isomorphic to $\mathbb{R}^{2n}$) as just mentioned, and the $\mathbf{c}$-complex case when we consider a vector $\mathbf{c} \in \mathcal{C} \subset \mathbb{C}^{2n}$. A key insight from [11] is that the nonanalyticity due to the presence of $\bar{\mathbf{z}}$, in a real-valued function is, *not* a nonanalyticity when viewed from the $\mathbf{c}$-real perspective. The Hessian of a real-valued function on $\mathbb{C}^N$ is given by [11]

$$\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}} = \frac{\partial}{\partial \mathbf{c}} \left(\frac{\partial f}{\partial \mathbf{c}}\right)^H = \begin{pmatrix} \mathcal{H}_{\mathbf{zz}} & \mathcal{H}_{\bar{\mathbf{z}}\mathbf{z}} \\ \mathcal{H}_{\mathbf{z}\bar{\mathbf{z}}} & \mathcal{H}_{\bar{\mathbf{z}}\bar{\mathbf{z}}} \end{pmatrix}, \tag{6}$$

where

$$\mathcal{H}_{\mathbf{zz}} \triangleq \frac{\partial}{\partial \mathbf{z}} \left(\frac{\partial f}{\partial \mathbf{z}}\right)^H, \quad \mathcal{H}_{\bar{\mathbf{z}}\mathbf{z}} \triangleq \frac{\partial}{\partial \bar{\mathbf{z}}} \left(\frac{\partial f}{\partial \mathbf{z}}\right)^H,$$

$$\mathcal{H}_{\mathbf{z}\bar{\mathbf{z}}} \triangleq \frac{\partial}{\partial \mathbf{z}} \left(\frac{\partial f}{\partial \bar{\mathbf{z}}}\right)^H, \quad \text{and} \quad \mathcal{H}_{\bar{\mathbf{z}}\bar{\mathbf{z}}} \triangleq \frac{\partial}{\partial \bar{\mathbf{z}}} \left(\frac{\partial f}{\partial \bar{\mathbf{z}}}\right)^H.$$

In what follows $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}$ of (6) is called the *full Hessian* and $\mathcal{H}_{\mathbf{zz}}, \mathcal{H}_{\bar{\mathbf{z}}\mathbf{z}}, \mathcal{H}_{\mathbf{z}\bar{\mathbf{z}}}, \mathcal{H}_{\bar{\mathbf{z}}\bar{\mathbf{z}}}$ are *partial Hessians*. The partial Hessian $\mathcal{H}_{\mathbf{zz}}$ is called the *leading partial Hessian*. The matrices $\mathcal{H}_{\mathbf{zz}}(\mathbf{z})$ and $\mathcal{H}_{\bar{\mathbf{z}}\bar{\mathbf{z}}}(\mathbf{z})$ are together called the block diagonal partial Hessians. In [11] it is pointed out that suboptimal version of the Newton algorithm can be developed using $\mathcal{H}_{\mathbf{zz}}$ rather than $\mathcal{H}_{\mathbf{cc}}$. We call this the pseudo-Newton method. In the next section we discuss the Gradients and Hessians of CM loss function.

A. Gradients and Hessians of CM Loss Function

For convenience we define the $N \times N$ complex matrices $\mathsf{A}(\mathbf{w})$, B and $\mathsf{C}(\mathbf{w})$ as

$$\mathsf{A}(\mathbf{w}) \triangleq \mathrm{E}\left\{\left(\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}\right)\mathbf{x}\mathbf{x}^H\right\} = \mathrm{E}\left\{|\,y|^2\,\mathbf{x}\mathbf{x}^H\right\} \in \mathbb{C}^{N\times N}, \tag{7}$$

$$\mathsf{B} \triangleq \mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\} \in \mathbb{C}^{N\times N}, \tag{8}$$

$$\mathsf{C}(\mathbf{w}) \triangleq \mathrm{E}\left\{\left(\mathbf{x}^H\mathbf{w}\right)^2\mathbf{x}\bar{\mathbf{x}}^H\right\} = \mathrm{E}\left\{\bar{y}^2\mathbf{x}\bar{\mathbf{x}}^H\right\} \in \mathbb{C}^{N\times N}. \tag{9}$$

Note that B and $\mathsf{A}(\mathbf{w})$ are necessarily both positive semidefinite, $\mathsf{B} \geq 0$ and $\mathsf{A}(\mathbf{w}) \geq 0$. As shown later, in the noiseless channel case, in order to implement the unmodified pseudo-Newton algorithm, it is necessary to make the stronger assumption that B is positive definite, $\mathsf{B} = \mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\} > 0$, which will generally ensure that $\mathsf{A}(\mathbf{w})$ is positive definite (assuming that $\mathbf{w} \neq \mathbf{0}$). Under the assumption that $\mathbf{x}$, has zero mean, the positive definiteness assumption on B is equivalent to the assumption that $\mathbf{x}$ is a full rank random vector. As $\mathbf{x}$ is a vector of outputs from an LTI channel with a full rank input sequence a, the full rank assumption on $\mathbf{x}$ is equivalent to yet-to-be-specified conditions being placed on the channel model. We call the condition that $\mathsf{B} > 0$, the full rank channel condition. For now, we will assume that $\mathsf{B} > 0$. Unfortunately, as discussed later, this assumption alone will not guarantee *perfect* equalization of a noiseless FIR channel model given in [4]. Following the methodology described in [11], the N-dimensional cogradients and gradients of $\ell(\mathbf{w})$ are readily computed,[1]

$$\begin{aligned}\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}} &= 2\,\mathrm{E}\left\{\left(\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}-\rho\right)\mathbf{w}^H\mathbf{x}\mathbf{x}^H\right\}, \\ &= 2\,\mathrm{E}\left\{\left(|\,y|^2-\rho\right)y\,\mathbf{x}^H\right\},\end{aligned} \tag{10}$$

$$\begin{aligned}\nabla_{\mathbf{w}}\ell(\mathbf{w}) &= \left(\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}}\right)^H = 2\mathrm{E}\left\{\left(\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}-\rho\right)\mathbf{x}\mathbf{x}^H\mathbf{w}\right\}, \\ &= 2\left(\mathsf{A}(\mathbf{w})-\rho\mathsf{B}\right)\mathbf{w},\end{aligned} \tag{11}$$

$$\frac{\partial \ell(\mathbf{w})}{\partial \bar{\mathbf{w}}} = 2\,\mathrm{E}\left\{\left(|\,y|^2-\rho\right)\bar{y}\,\bar{\mathbf{x}}^H\right\} \tag{12}$$

$$\nabla_{\bar{\mathbf{w}}}\ell(\mathbf{w}) = \left(\frac{\partial \ell(\mathbf{w})}{\partial \bar{\mathbf{w}}}\right)^H = 2\left(\overline{\mathsf{A}}(\mathbf{w})-\rho\overline{\mathsf{B}}\right)\bar{\mathbf{w}} \tag{13}$$

$$\frac{\partial \ell(\mathbf{c})}{\partial \mathbf{c}} = 2\left(\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}},\ \frac{\partial \ell(\mathbf{w})}{\partial \bar{\mathbf{w}}}\right) \tag{14}$$

$$\nabla_{\mathbf{c}}\ell(\mathbf{c}) = \left(\frac{\partial \ell(\mathbf{c})}{\partial \mathbf{c}}\right)^H = 2\begin{pmatrix}\left(\mathsf{A}(\mathbf{w})-\rho\mathsf{B}\right)\mathbf{w}\\ \left(\overline{\mathsf{A}}(\mathbf{w})-\rho\overline{\mathsf{B}}\right)\bar{\mathbf{w}}\end{pmatrix}. \tag{15}$$

From the gradients, we can then compute the full Hessian[2]

$$\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{c}) = \mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w}) = \begin{pmatrix}\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) & \mathcal{H}_{\bar{\mathbf{w}}\mathbf{w}}(\mathbf{w})\\ \mathcal{H}_{\mathbf{w}\bar{\mathbf{w}}}(\mathbf{w}) & \mathcal{H}_{\bar{\mathbf{w}}\bar{\mathbf{w}}}(\mathbf{w})\end{pmatrix} \in \mathbb{C}^{2N\times 2N}. \tag{16}$$

It is easy to validate that the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ satisfies the factorization

$$\begin{pmatrix} I & 0\\ -\mathcal{H}_{\mathbf{w}\bar{\mathbf{w}}}\mathcal{H}^{-1}_{\mathbf{ww}} & I\end{pmatrix}\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}\begin{pmatrix} I & -\mathcal{H}_{\bar{\mathbf{w}}\mathbf{w}}\mathcal{H}^{-1}_{\mathbf{ww}}\\ 0 & I\end{pmatrix} = \begin{pmatrix}\mathcal{H}_{\mathbf{ww}} & 0\\ 0 & \widetilde{\mathcal{H}}_{\mathbf{ww}}\end{pmatrix} \tag{17}$$

where

$$\widetilde{\mathcal{H}}_{\mathbf{ww}}(\mathbf{w}) = \mathcal{H}_{\mathbf{ww}}(\mathbf{w}) - \mathcal{H}_{\bar{\mathbf{w}}\mathbf{w}}(\mathbf{w})\mathcal{H}^{-1}_{\bar{\mathbf{w}}\bar{\mathbf{w}}}(\mathbf{w})\mathcal{H}_{\mathbf{w}\bar{\mathbf{w}}}(\mathbf{w}) \in \mathbb{C}^{N\times N} \tag{18}$$

is the Schur complement of $\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) \in \mathbb{C}^{N\times N}$, showing that

$$\mathrm{rank}\left(\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})\right) = \mathrm{rank}\left(\mathcal{H}_{\mathbf{ww}}(\mathbf{w})\right) + \mathrm{rank}\left(\widetilde{\mathcal{H}}_{\mathbf{ww}}(\mathbf{w})\right). \tag{19}$$

The relationship (19) clearly shows that the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ is positive definite if and only if the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ and its Schur complement $\widetilde{\mathcal{H}}_{\mathbf{ww}}(\mathbf{w})$ are both positive definite. Below, we will show that the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ must be singular (at best only positive semidefinite), so that if the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ is positive definite (as required, e.g., for the algorithm proposed in [7]) it must be the case that its Schur complement $\widetilde{\mathcal{H}}_{\mathbf{ww}}(\mathbf{w})$ is singular. The $N \times N$ partial Hessians of $\ell(\mathbf{w})$ are:

$$\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}}\left(\frac{\partial \ell(\mathbf{w})}{\partial \mathbf{w}}\right)^H = 2\left(2\,\mathsf{A}(\mathbf{w})-\rho\,\mathsf{B}\right), \tag{20}$$

$$\mathcal{H}_{\bar{\mathbf{w}}\bar{\mathbf{w}}}(\mathbf{w}) = 2\left(2\,\overline{\mathsf{A}}(\mathbf{w})-\rho\,\overline{\mathsf{B}}\right), \tag{21}$$

$$\mathcal{H}_{\mathbf{w}\bar{\mathbf{w}}}(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}}\left(\frac{\partial \ell(\mathbf{w})}{\partial \bar{\mathbf{w}}}\right)^H = 2\overline{\mathsf{C}}(\mathbf{w}), \tag{22}$$

$$\mathcal{H}_{\bar{\mathbf{w}}\mathbf{w}}(\mathbf{w}) = 2\mathsf{C}(\mathbf{w}). \tag{23}$$

Having computed the partial Hessians, the full Hessian is

$$\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w}) = 2\begin{pmatrix}2\,\mathsf{A}(\mathbf{w})-\rho\,\mathsf{B} & \mathsf{C}(\mathbf{w})\\ \overline{\mathsf{C}}(\mathbf{w}) & 2\,\overline{\mathsf{A}}(\mathbf{w})-\rho\,\overline{\mathsf{B}}\end{pmatrix}. \tag{24}$$

Note that positive definiteness of the leading partial Hessian, $\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) = 2(2\,\mathsf{A}(\mathbf{w})-\rho\,\mathsf{B}) > 0$, is a necessary, but not sufficient, condition for the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ to be positive definite. In reference [14], the Hessian for the general CM loss function of order p for the complex case is computed, but only after first transforming the complex CM loss function of order p into an equivalent real loss function, and the resulting Hessian is much more complicated in form than (24). We now take a closer look at the stationary points of the CM loss function and their relation to the equalization.

IV. Stationary Points of CM Loss Function

From Equation (11) we see that the condition for $\mathbf{w}$ to be a stationary point of the CM loss function (4) is

$$\left(\mathsf{A}(\mathbf{w})-\rho\,\mathsf{B}\right)\mathbf{w} = 0, \tag{25}$$

where ρ is defined in equation (3). Because the matrix $\mathsf{A}(\mathbf{w})-\rho\mathsf{B}$ is Hermitian, the stationarity condition is equivalent (not just sufficient) to

$$\mathbf{w}^H\left(\mathsf{A}(\mathbf{w})-\rho\,\mathsf{B}\right)\mathbf{w} = 0. \tag{26}$$

Assuming the existence of a stationary point, $\mathbf{w}_s \neq 0$ and $\mathsf{B} = \mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\}$ to be full rank, so that $\mathbf{w}_s^H\mathsf{B}\mathbf{w}_s \neq 0$ we can rewrite as (26) as

$$\frac{\mathbf{w}_s^H\mathsf{A}(\mathbf{w})\mathbf{w}_s}{\mathbf{w}_s^H\mathsf{B}\mathbf{w}_s} = \rho. \tag{27}$$

[1] Note that $\frac{\partial \ell}{\partial \bar{\mathbf{w}}} = \overline{\frac{\partial \ell}{\partial \mathbf{w}}}$ and $\left(\frac{\partial \ell}{\partial \bar{\mathbf{w}}}\right)^H = \overline{\left(\frac{\partial \ell}{\partial \mathbf{w}}\right)^H}$.

[2] Note that $\mathcal{H}_{\bar{\mathbf{w}}\bar{\mathbf{w}}} = \overline{\mathcal{H}}_{\mathbf{ww}}$ and $\mathcal{H}_{\mathbf{w}\bar{\mathbf{w}}} = \overline{\mathcal{H}}_{\bar{\mathbf{w}}\mathbf{w}}$.

Then using equations (3), (7) and (8) with $y_s = \mathbf{w}_s^H \mathbf{x}$ we can write the stationary condition (28) as

$$\frac{\mathrm{E}\left\{|\,a(n)|^4\right\}}{\mathrm{E}\left\{|\,a(n)|^2\right\}} = \rho = \frac{\mathrm{E}\left\{|\,y_s(n)|^4\right\}}{\mathrm{E}\left\{|\,y_s(n)|^2\right\}}, \tag{28}$$

showing that a stationary point produces an output that matches the ratio of the 4^{th} and 2^{nd} order moments. Any solution $\mathbf{w}_s$ satisfying (28) is a stationary point. Note that from equations (7) and (8) that any perfect equalizing solution is a stationary point, but other solutions to (26) may exist.

The trivial solution $\mathbf{w} = \mathbf{0}$ is also a stationary point of the CM loss function $\ell(\mathbf{w})$. However under the assumption that the covariance matrix B is full rank. The trivial solution is a unique local *maximizer* of $\ell(\mathbf{w})$,

$$\mathbf{w} = \mathbf{0} \in \mathcal{W}' \text{ and } \mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{0}) = -2\rho \begin{pmatrix} \mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\} & 0 \\ 0 & \mathrm{E}\left\{\bar{\mathbf{x}}\bar{\mathbf{x}}^H\right\} \end{pmatrix} < 0.$$

This shows that the CM loss function given in Equation (4) is not globally convex (for full rank B) as the condition of global positive semidefiniteness of the Hessian of a smooth globally convex function is violated. Indeed, it is known that the shape of the CM loss function can be quite complex, having multiple local minima which may, or may not, be globally optimal [4]. Because it is a local maximizer of the CM loss function, we can ignore the trivial stationary point $\hat{\mathbf{w}} = \mathbf{0}$ and work instead with the set of nontrivial stationary points,

$$\mathcal{W} \triangleq \mathcal{W}' \setminus \{\mathbf{0}\}\,. \tag{29}$$

Because the CM loss function (4) is smooth and bounded from below, all stable generalized gradient descent methods must converge to a point in the set $\mathcal{W}$ as this set contains all points for which the gradient of $\ell(\mathbf{w})$ is equal to zero. Under the assumption that the output signal covariance matrix B is nonsingular, every nontrivial stationary point satisfies the condition (28). However, this latter fact is not true if B is rank deficient: it will shown below that $\mathbf{B}$ is rank-deficient then there exists $\mathbf{w} \neq 0$ for which condition (28) becomes indeterminant (i.e., "$\rho = 0/0$").

Ideally a solution $\hat{\mathbf{w}} \in \mathcal{W}'$ would have a positive definite full Hessian matrix $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}})$, thereby ensuring that the solution is a unique local minimizer of the CM loss function $\ell(\mathbf{w})$ and amenable to solution via a fast converging Newton algorithm. However, it is well known, and easily verified, that a solution to a CM loss function of the form (4) is not unique, as any minimizer $\hat{\mathbf{w}} \in \mathcal{W}$ of (4) is associated with a connected family of solutions given by

$$[\hat{\mathbf{w}}] = \{\hat{\mathbf{w}}_\alpha \,|\, \hat{\mathbf{w}}_\alpha = \alpha\, \hat{\mathbf{w}},\ \alpha \in \mathbb{C},\ |\alpha| = 1\}\ .$$

Elements of the set $[\hat{\mathbf{w}}]$ differ from each other only by multiplication by a unimodular scalar and every element of this set produces the same (locally minimal) value of the loss function $\ell(\mathbf{w})$. Under the assumption that the output covariance matrix B is nonsingular, the elements of $[\hat{\mathbf{w}}]$ also each satisfy the condition (27). The fact that the CM loss function has connected local minima for the complex case considered here is a significant complication over the real case for which the CM loss function has isolated local minima [4].

In particular, whereas in the real case the full Hessian can be ensured to be positive definite at a local minima, in the complex case the full Hessian can at most be positive semidefinite (as shown below) with nullity (dimension of its nullspace) at least one [14]. Thus

$$\text{rank}(\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}})) \leq 2N - 1 \tag{30}$$

showing that the $2N \times 2N$ matrix rank$(\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}}))$ is *always* singular for complex signals. Below, for a LTI FIR channel, we will verify this fact and also show that the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}})$ has rank *equal* to the maximum possible value of $2N-1$ *only if* the full rank channel condition holds, so that B is positive definite. It is interesting to ask under what conditions the class $[\mathbf{w}]$ is an equivalence class, which would require that $\mathbf{w}' \in [\mathbf{w}]$ if and only if $\ell(\mathbf{w}') = \ell(\mathbf{w})$. Below, we will show that $[\mathbf{w}]$ is an equivalence class for a fractionally spaced noiseless FIR channel model if and only if $\mathsf{B} = \mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\} > 0$.

As a direct consequence of the lack of uniqueness of the solutions to the minimization problem, the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}})$ cannot be positive definite at a minimizing solution $\hat{\mathbf{w}}$ lest we have a contradiction. This explains the phenomenon noted in the simulations of reference [7] that the Schur complement is singular, and hence noninvertible, so that consequently the full Newton algorithm is not implementable.

One can directly show singularity of the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}})$ for *every* nonzero stationary point $\hat{\mathbf{w}} \in \mathcal{W}$. To do so, select any stationary value $\hat{\mathbf{w}} \in \mathcal{W}$ and then form the vector $\mathbf{c} \in \mathcal{C}$ as

$$\mathbf{c} = \begin{pmatrix} \hat{\mathbf{w}} e^{\frac{j\pi}{2}} \\ \overline{\hat{\mathbf{w}} e^{\frac{j\pi}{2}}} \end{pmatrix} = e^{\frac{j\pi}{2}} \begin{pmatrix} \hat{\mathbf{w}} \\ \hat{\bar{\mathbf{w}}} e^{-j\pi} \end{pmatrix} = e^{\frac{j\pi}{2}} \begin{pmatrix} \hat{\mathbf{w}} \\ -\hat{\bar{\mathbf{w}}} \end{pmatrix}. \tag{31}$$

Now from (24) and (27), in a series of steps (32)-(33) (shown at the top of next page), the full Hessian can be shown to be singular with the nonzero vector $\mathbf{c}$ being an element in the nullspace of the Hessian.

V. The Newton and Pseudo-Newton Algorithms

The results of the previous section shows that it is not possible to uniquely compute the Newton algorithm update increment from the Newton algorithm stationarity condition

$$\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{c})\Delta\mathbf{c} = -\nabla_{\mathbf{c}}\ell(\mathbf{c}), \tag{35}$$

as the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ is singular near the desired stationary points of $\ell(\mathbf{c})$. To circumvent the singularity issue associated with the need to invert the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ in the Newton algorithm, reference [7] proposed Newton-like algorithm which is equivalent to the pseudo-Newton algorithm discussed in [11]. The pseudo-Newton algorithm yields an estimation update $\Delta\mathbf{w}$ determined from

$$\Delta\mathbf{w} = -\mathcal{H}^{-1}_{\mathbf{ww}}(\mathbf{w})\nabla_{\mathbf{w}}\ell(\mathbf{w}) = -\mathcal{H}^{-1}_{\mathbf{ww}}(\mathbf{w})\left(\frac{\partial\ell(\mathbf{w})}{\partial\mathbf{w}}\right)^H \tag{36}$$

assuming that the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}})$ is invertible.[3] We can write the pseudo-Newton algorithm (36) as

$$(2\,\mathsf{A}(\mathbf{w}) - \rho\,\mathsf{B})\,\Delta\mathbf{w} = -\,(\mathsf{A}(\mathbf{w}) - \rho\mathsf{B})\,\mathbf{w}\,. \tag{37}$$

[3] The above equation can also be implemented in a computationally less intensive way by writing (36) as $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})\Delta\mathbf{w} = -\nabla_{\mathbf{w}}\ell(\mathbf{w})$, and then solving it for $\Delta\mathbf{w}$.

$$\frac{1}{2}\,\mathbf{c}^H\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}})\mathbf{c} = \begin{pmatrix}\hat{\mathbf{w}}^H & -\hat{\bar{\mathbf{w}}}^H\end{pmatrix}\begin{pmatrix}\mathrm{E}\left\{\left(2\,|y|^2-\rho\right)\mathbf{x}\mathbf{x}^H\right\} & \mathrm{E}\left\{\bar{y}^2\,\mathbf{x}\bar{\mathbf{x}}^H\right\}\\ \mathrm{E}\left\{y^2\,\bar{\mathbf{x}}\mathbf{x}^H\right\} & \mathrm{E}\left\{\left(2\,|y|^2-\rho\right)\bar{\mathbf{x}}\bar{\mathbf{x}}^H\right\}\end{pmatrix}\begin{pmatrix}\hat{\mathbf{w}}\\ -\hat{\bar{\mathbf{w}}}\end{pmatrix} \tag{32}$$

$$= \hat{\mathbf{w}}^H\mathrm{E}\left\{\left(2\,|y|^2-\rho\right)\mathbf{x}\mathbf{x}^H\right\}\hat{\mathbf{w}} - \hat{\mathbf{w}}^H\mathrm{E}\left\{\bar{y}^2\,\mathbf{x}\bar{\mathbf{x}}^H\right\}\hat{\bar{\mathbf{w}}} + \text{complex conjugate} \tag{33}$$

$$= 2\,\mathrm{E}\left\{|\,y|^4\right\} - \rho\,\mathrm{E}\left\{|\,y|^2\right\} - \mathrm{E}\left\{|\,y|^4\right\} + \text{ c.c.} = 0\,. \tag{34}$$

As noted above the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ is at best positive semidefinite, even with the assumption that the leading partial Hessian is full rank, the resulting minimizing solution $\hat{\mathbf{w}}$ found from (36) is not guaranteed to be unique. Thus without further analysis to indicate otherwise, one should assume that the answer produced by the algorithm is realization dependent.

A. General Hessian Positivity Conditions

We now develop sufficient conditions for the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ of the CM loss function to be positive semidefinite at an arbitrary point $\mathbf{w}$. This will be a sufficient condition for both the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ and its Schur complement $\widetilde{\mathcal{H}}_{\mathbf{ww}}(\mathbf{w})$ to be positive semidefinite. Let $\mathbf{c} = \binom{\mathbf{w}}{\bar{\mathbf{w}}}$, $\boldsymbol{\xi} = \binom{\mathbf{v}}{\bar{\mathbf{v}}}$, $y = \mathbf{w}^H\mathbf{x}$, $\eta = \mathbf{v}^H\mathbf{x}$, for arbitrarily chosen vectors $\mathbf{w}$ and $\mathbf{v}$. Note that $(\bar{y}\,\eta + y\,\bar{\eta})^2 = 4\left(\mathrm{Re}\left\{\bar{y}\,\eta + y\,\bar{\eta}\right\}\right)^2 \geq 0$ and $\mathbf{w} = \hat{\mathbf{w}}$ and $\mathbf{v} = \hat{\mathbf{w}}e^{j\frac{\pi}{2}}$ for $\hat{\mathbf{w}} \in \mathcal{W} \Rightarrow (\bar{y}\,\eta + y\,\bar{\eta})^2 = 0$. We have

$$\begin{aligned}\frac{1}{2}\,\boldsymbol{\xi}^H\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})\boldsymbol{\xi} &= \mathbf{v}^H\mathrm{E}\left\{\left(2\,|y|^2-\rho\right)\mathbf{x}\mathbf{x}^H\right\}\mathbf{v} + \\ &\quad \mathbf{v}^H\mathrm{E}\left\{\bar{y}^2\,\mathbf{x}\bar{\mathbf{x}}^H\right\}\bar{\mathbf{v}} + \text{ c.c.},\\ &= \mathrm{E}\left\{\left(2\,|\,y|^2-\rho\right)|\eta|^2\right\} + \mathrm{E}\left\{\bar{y}^2\,\eta^2\right\} + \\ &\quad \mathrm{E}\left\{y^2\,\bar{\eta}^2\right\} + \mathrm{E}\left\{\left(2\,|\,y|^2-\rho\right)|\eta|^2\right\},\\ &= \mathrm{E}\left\{2\,|\,y\eta|^2 + (\bar{y}\,\eta + y\,\bar{\eta})^2 - 2\rho\,|\eta|^2\right\}.\end{aligned}$$

As a check for correctness of the above expression, if we set $\mathbf{w} = \hat{\mathbf{w}}$ and $\mathbf{v} = \hat{\mathbf{w}}e^{j\frac{\pi}{2}}$ for $\hat{\mathbf{w}} \in \mathcal{W}$ we obtain $\frac{1}{2}\,\boldsymbol{\xi}^H\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})\boldsymbol{\xi} = 0$, reproducing our earlier result shown in (31) *et seq.* For a particular choice of $\mathbf{w}$ (equivalently, for a choice of y), the full Hessian is positive definite if and only if the quadratic form $\frac{1}{2}\,\boldsymbol{\xi}^H\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})\boldsymbol{\xi}$ is nonnegative for all $\mathbf{v}$ (equivalently, for all η). Noting that

$$\frac{1}{2}\,\boldsymbol{\xi}^H\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})\boldsymbol{\xi} \geq 2\,\mathrm{E}\left\{|\,y\eta|^2 - \rho\,|\eta|^2\right\},$$

an obvious sufficient condition for this to be true is

$$\mathrm{E}\left\{|\,y\eta|^2 - \rho\,|\eta|^2\right\} = \mathrm{E}\left\{\left(|\,y|^2-\rho\right)|\eta|^2\right\} \geq 0 \tag{38}$$

for all η given a particular choice of y, which is equivalent to the condition

$$\rho \leq \frac{\mathrm{E}\left\{|\,y|^2\,|\eta|^2\right\}}{\mathrm{E}\left\{|\,y|^2\right\}}\,. \tag{39}$$

Expanding (38), we obtain the condition $\mathbf{v}^H\left(\mathsf{A}(\mathbf{w}) - \rho\,\mathsf{B}\right)\mathbf{v} \geq 0$ for all $\mathbf{v}$, which is just the condition that $\mathsf{A}(\mathbf{w}) - \rho\,\mathsf{B}$ be positive semidefinite,

$$\mathsf{A}(\mathbf{w}) - \rho\,\mathsf{B} \geq 0\,. \tag{40}$$

Assuming that B is full rank, the condition (40) is a sufficient condition for the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ to be positive definite. As then

$$\frac{1}{2}\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) = 2\mathsf{A}(\mathbf{w}) - \rho\mathsf{B} = 2\left(\mathsf{A}(\mathbf{w}) - \rho\mathsf{B}\right) + \rho\mathsf{B} \geq \rho\mathsf{B} \geq 0.$$

Note that it cannot be the case that the block diagonal Hessians shown in (24) are nonsingular while at the same time the block off-diagonal matrices vanish as this would contradict the fact that the full Hessian is intrinsically nonsingular.

VI. Noiseless FIR Channel Model

Without further assumptions it is difficult to make additional statements about definiteness conditions for the full and partial Hessians. Therefore, following [4], we will now make the assumption of a noiseless fractionally sampled FIR channel and show that under this assumption we can transform the Hessians into equivalent forms which facilitate further analysis. This will enable us to gain further insight into the structure of the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ as well as the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$. We will determine a necessary and sufficient condition for the full Hessian to be positive-semidefinite at a globally optimal (generally non-unique) solution $\hat{\mathbf{w}}$ of the CMA loss function assuming that the channel can be modeled as a noiseless fractionally spaced FIR filter of length $L = rM$, where $r \geq 1$ is an integer which denotes the number of samples per symbol-interval. The optimal solution allows perfect equalization

$$y = y(n) = \hat{\mathbf{w}}^H\mathbf{x} = \hat{\mathbf{w}}^H\mathbf{x}(n) = a(n-\delta)\,, \quad \hat{\mathbf{w}}, \mathbf{x}(n) \in \mathbb{C}^N, \tag{41}$$

for some integer delay $\delta = 0, 1, \cdots, N-1$. This condition will also ensure that the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ is positive definite, allowing for the implementation of the unmodified pseudo-Newton algorithm. The noiseless L-tap fractionally spaced FIR channel model is given by [4]

$$\begin{aligned}\mathbf{x}(n) &= (x^{(1)}(n), \cdots x^{(r)}(n), x^{(1)}(n-1), \cdots x^{(r)}(n-1), \cdots,\\ &\quad x^{(1)}(n-q+1), \cdots x^{(r)}(n-q+1))^T,\\ &\triangleq (x(n), x(n-1), \cdots, x(n-N+1))^T \in \mathbb{C}^N,\ N = rq,\end{aligned}$$

and

$$x^{(i)}(n) = \sum_{j=0}^{M-1}\bar{c}^{(i)}(j)a(n-j) \quad i = 1, \cdots, r\,. \tag{42}$$

As discussed in [4], this is the r-multichannel model of a noiseless fractionally spaced channel with impulse response of length $L = rM$. Following [4], we recast (42) in vector-matrix form as

$$\mathbf{x}(n) = \mathbf{C}^H\mathbf{a}(n) \in \mathbb{C}^N\,, \tag{43}$$

$$\mathbf{a}(n) \triangleq \begin{pmatrix}a(n) & a(n-1) & \cdots & a(n-P+1)\end{pmatrix}^T \in \mathbb{C}^P$$

and $\mathbf{C} \in \mathbb{C}^{P\times N}$, where $P = M + q - 1$ is the number of transmitted data samples processed by the FIR equalizer. Similar to our earlier notation, we will suppress the sample index n and write the channel model as

$$\mathbf{x} = \mathbf{C}^H \mathbf{a} \tag{44}$$

when there is no possibility of confusion. We call the $P \times N = (M + q - 1) \times rq$ complex matrix $\mathbf{C}$ the channel matrix. Reference [4] describes the channel matrix and give the explicit structural form of $\mathbf{C}$ for symbol rate sampling and for $2\times$ fractionally spaced sampling. Note from Equation (44) that perfect recovery of the transmitted data vector $\mathbf{a}$ from the received data $\mathbf{x}$ is only possible if $\mathbf{C}^H$ is one-to-one, or, equivalently if $\mathbf{C}$ is onto. This condition implies that the $P \times N$ matrix $\mathbf{C}$ has full row rank, $\text{rank}(\mathbf{C}) = P$. Another way to see the importance of this rank condition is to consider the conditions required for perfect channel equalization. Given the channel (44), in order for the condition (41) to be true for any $\mathbf{a}$ we must be able to solve the system

$$\mathbf{e}_\delta = \mathbf{C}\hat{\mathbf{w}} \tag{45}$$

where $\mathbf{e}$ is the canonical vector having zeros in all positions, except for the value 1 in position δ.[4] Johnson et al. [4] argue that (45) should be solvable for every delay $\delta = 0, 1, \cdots, N-1$, again yielding the requirement that $\mathbf{C}$ be onto. We will call the condition of perfect channel equalization for all delays, the all-delay perfect equalization property.

Note that a necessary condition for the $P \times N$ matrix $\mathbf{C}$ to have full row rank is that $N \geq P$, and thus All-delay perfect equalization implies

$$\text{rank}(\mathbf{C}) = P \;\Rightarrow\; N = rq \geq P = M + q - 1. \tag{46}$$

If this condition is violated, $\mathbf{C}$ is not onto and the generic delay property does not hold. In [4] it is pointed out that this condition means that no symbol-rate sampled ($r = 1$) FIR equalizer can yield the all-delay perfect equalization property. Reference [4] also points out that in addition to the condition (46), the r polynomials specified by the channel coefficients in (42) can share no roots in common if $\mathbf{C}$ is to have full row rank. As we shall now see, for the noiseless channel which we have been considering here, the full row rank condition for all-delay perfect reconstruction is in conflict with the condition that the leading partial Hessian be positive definite. The noiseless fractionally spaced FIR channel model (43) yields

$$\begin{aligned}
\mathsf{B} &= \mu_2\, \mathbf{C}^H \mathbf{C} && (47)\\
\mathsf{C}(\mathbf{w}) &= \mathbf{C}^H\, \mathrm{E}\left\{\bar{y}^2 \mathbf{a}\bar{\mathbf{a}}^H\right\} \overline{\mathbf{C}} && (48)\\
\mathsf{A}(\mathbf{w}) &= \mathbf{C}^H\, \mathrm{E}\left\{|y|^2\, \mathbf{a}\mathbf{a}^H\right\} \mathbf{C} && (49)\\
\frac{1}{2}\,\nabla_{\mathbf{w}}\ell(\mathbf{w}) &= \mathbf{C}^H\, \mathrm{E}\left\{\left(|y|^2 - \rho\, I\right) \mathbf{a}\mathbf{a}^H\right\} \mathbf{C}\,\mathbf{w} && (50)\\
\frac{1}{2}\,\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) &= \mathbf{C}^H\, \mathrm{E}\left\{\left(2\,|y|^2 - \rho\, I\right) \mathbf{a}\mathbf{a}^H\right\} \mathbf{C} && (51)
\end{aligned}$$

and the full Hessian can be written as (52) and (53) (shown at the top of next page). Note that the channel output signal covariance matrix $\mathsf{B} = \mathrm{E}\left\{\mathbf{x}\mathbf{x}^H\right\}$ has full rank if and only if the $P \times N$ channel matrix $\mathbf{C}$ has full column rank, $\text{rank}(\mathbf{C}) = N$. This can be viewed as a consequence of the more general fact that

$$\text{rank}(\mathsf{B}) = \text{rank}(\mathbf{C})\,. \tag{54}$$

Furthermore, a necessary condition for the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ to be full rank is that the channel matrix $\mathbf{C}$ again have full column rank. Thus we have shown that for a noiseless fractionally spaced FIR channel, a necessary condition that the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ be full rank is that the output signal covariance matrix $\mathbf{B}$ be full rank (i.e., that we satisfy the *full rank channel condition*) which, in turn, is true if and only if channel matrix has full column rank, $\text{rank}(\mathbf{C}) = N$. Note the important fact that even though we do not know the channel matrix $\mathbf{C}$, we can determine if it fails to have full column rank by determining if the output signal covariance matrix B fails to be full rank from the important condition (54). A necessary condition for $\mathbf{C}$ to have full column rank is that $N \leq P$. Thus we have shown that

$$\begin{aligned}
\text{Full rank } \mathcal{H}_{\mathbf{ww}}(\mathbf{w}) \;&\Rightarrow\; \text{rank}(\mathsf{B}) = N\\
&\Leftrightarrow\; \text{rank}(\mathbf{C}) = N \qquad\qquad (55)\\
&\Rightarrow\; N = rq \leq P = M + q - 1\,.
\end{aligned}$$

Again note that no perfectly equalizing solution exists for $r = 1$ (symbol spaced sampling) as pointed out above and in [4]. The all-delay perfect equalization requirement (46) that $\mathbf{C}$ be onto is in conflict with the requirement (55) that $\mathbf{C}$ have full column rank. For both conditions (46) and (55) to hold simultaneously it must be the case that the channel matrix $\mathbf{C}$ is square, $N = rq = P = M + q - 1$, giving the necessary condition that all-delay perfect equalization and full rank leading partial Hessian imlpies

$$q = \frac{M-1}{r-1}\,. \tag{56}$$

Let us now assume that the full row-rank channel matrix condition for perfect equalization is satisfied and that the channel matrix has full column rank and ask whether the perfect equalizer, $\hat{\mathbf{w}}_*$ is a stationary point of the CM loss function, $\hat{\mathbf{w}}_* \in \mathcal{W}$, and, if so, determine whether or not when evaluated at the perfect solution, the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}}_*)$ has full rank and if the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}}_*)$ is positive semidefinite. For simplicity, and with no loss of generality, we usually assume zero delay, $\delta = 0$, so that the perfect equalization condition is taken as the zero-delay condition

$$\mathbf{e}_0 = \mathbf{C}\hat{\mathbf{w}}_*\,.$$

We then have

$$y = y(n) = \hat{\mathbf{w}}_*^H \mathbf{x} = \hat{\mathbf{w}}_*^H \mathbf{C}^H \mathbf{a} = \mathbf{e_0}^H \mathbf{a} = a(n)\,,$$

as required for perfect equalization. From Equation (50), we

[4] Our vectors are indexed beginning from position 0.

$$\frac{1}{2}\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w}) = \frac{1}{2}\begin{pmatrix}\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) & \mathcal{H}_{\bar{\mathbf{w}}\mathbf{w}}(\mathbf{w})\\ \mathcal{H}_{\mathbf{w}\bar{\mathbf{w}}}(\mathbf{w}) & \mathcal{H}_{\bar{\mathbf{w}}\bar{\mathbf{w}}}(\mathbf{w})\end{pmatrix} = \begin{pmatrix}2\,\mathsf{A}(\mathbf{w})-\rho\,\mathsf{B} & \mathsf{C}(\mathbf{w})\\ \overline{\mathsf{C}}(\mathbf{w}) & 2\,\overline{\mathsf{A}}(\mathbf{w})-\rho\,\overline{\mathsf{B}}\end{pmatrix}, \tag{52}$$

$$= \begin{pmatrix}\mathbf{C} & 0\\ 0 & \overline{\mathbf{C}}\end{pmatrix}^H \begin{pmatrix}\mathrm{E}\left\{\left(2\,|y|^2-\rho\,I\right)\mathbf{a}\mathbf{a}^H\right\} & \mathrm{E}\left\{\bar{y}^2\mathbf{a}\bar{\mathbf{a}}^H\right\}\\ \mathrm{E}\left\{y^2\bar{\mathbf{a}}\mathbf{a}^H\right\} & \mathrm{E}\left\{\left(2\,|y|^2-\rho\,I\right)\bar{\mathbf{a}}\bar{\mathbf{a}}^H\right\}\end{pmatrix}\begin{pmatrix}\mathbf{C} & 0\\ 0 & \overline{\mathbf{C}}\end{pmatrix}. \tag{53}$$

obtain,

$$\begin{aligned}\frac{1}{2}\nabla_{\mathbf{w}}\ell(\hat{\mathbf{w}}_*) &= \mathbf{C}^H\,\mathrm{E}\left\{\left(|a(n)|^2-\rho\,I\right)\mathbf{a}\mathbf{a}^H\right\}\mathbf{e}_0,\\ &= \mathbf{C}^H\begin{pmatrix}\mu_4-\rho\mu_2 & & & \\ & \mu_2^2-\rho\mu_2 & & \\ & & \ddots & \\ & & & \mu_2^2-\rho\mu_2\end{pmatrix}\begin{pmatrix}1\\0\\ \vdots\\ 0\end{pmatrix},\\ &= \mathbf{C}^H\begin{pmatrix}0 & & & \\ & \mu_2^2(1-\kappa) & & \\ & & \ddots & \\ & & & \mu_2^2(1-\kappa)\end{pmatrix}\begin{pmatrix}1\\0\\ \vdots\\ 0\end{pmatrix} = \mathbf{0},\end{aligned}$$

showing that the perfectly equalizing solution $\hat{\mathbf{w}}_*$ is indeed a stationary point of the CM loss function $\ell(\mathbf{w})$ for the noiseless FIR channel model, $\hat{\mathbf{w}}_* \in \mathcal{W}$. We now compute the leading partial Hessian at the perfectly equalizing solution. We have from Equation (51) that

$$\begin{aligned}\frac{1}{2}\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}}_*) &= \mathbf{C}^H\,\mathrm{E}\left\{\left(2\,|a(n)|^2-\rho\,I\right)\mathbf{a}\mathbf{a}^H\right\}\mathbf{C}\\ &= \mathbf{C}^H\left[\mathrm{Diag}\left(2\mu_4-\rho\mu_2,\quad T_D,\quad \cdots,\quad T_D\right)\right]\mathbf{C}\\ &= \mathbf{C}^H\left[\mathrm{Diag}\left(\mu_4,\quad T_D,\quad \cdots,\quad T_D\right)\right]\mathbf{C},\end{aligned} \tag{57}$$

where

$$T_D = \mu_2^2(2-\kappa).$$

Note that the leading partial Hessian is always positive semidefinite if $\kappa < 2$. It is evident that the leading partial Hessian for the noiseless FIR channel is full rank if the $P \times N$ channel matrix $\mathbf{C}$ has full column rank and the normalized kurtosis has value $\kappa < 2$. However, $\mathbf{C}$ is assumed to have full row rank P. Thus

$$\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}}_*) > 0 \Leftarrow \mathrm{rank}\,\mathbf{C} = N = P \quad \text{and} \quad \kappa < 2\,. \tag{58}$$

A complex gaussian process has normalized kurtosis $\kappa_{\text{gauss}} = 2$. We see that a sufficient condition for the perfectly equalizing solution for a full rank channel to be attainable via a pseudo-Newton algorithm is that the signal process $a(n)$ be subgaussian, with kurtosis $\kappa < 2$. This, in fact, is a well-known result [2], [4], but one that is usually derived using approximation arguments. Note that our result is exact for noiseless FIR channel models which satisfy (58). Of course, the models which do satisfy these conditions are rather limited. They must, in particular, satisfy the very limiting necessary condition (56).

Assuming that $\mathbf{C}$ has full column rank, we can determine the rank of the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ at the perfectly equalizing solution $\hat{\mathbf{w}}_*$ from Equation (52). Equation (52) shows that we can write the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}}_*)$ matrix as

$$\frac{1}{2}\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}}_*) = \begin{pmatrix}\mathbf{C} & 0\\ 0 & \overline{\mathbf{C}}\end{pmatrix}^H \Pi_{\mathbf{cc}}(\hat{\mathbf{w}}_*)\begin{pmatrix}\mathbf{C} & 0\\ 0 & \overline{\mathbf{C}}\end{pmatrix}$$

where

$$\begin{aligned}\Pi_{\mathbf{cc}}(\hat{\mathbf{w}}_*) &\triangleq \begin{pmatrix}\Pi_{\mathbf{aa}}(\hat{\mathbf{w}}_*) & \Pi_{\bar{\mathbf{a}}\mathbf{a}}(\hat{\mathbf{w}}_*)\\ \Pi_{\mathbf{a}\bar{\mathbf{a}}}(\hat{\mathbf{w}}_*) & \Pi_{\bar{\mathbf{a}}\bar{\mathbf{a}}}(\hat{\mathbf{w}}_*)\end{pmatrix}\\ &\triangleq \begin{pmatrix}\mathrm{E}\left\{\left(2\,|y|^2-\rho\,I\right)\mathbf{a}\mathbf{a}^H\right\} & \mathrm{E}\left\{\bar{y}^2\mathbf{a}\bar{\mathbf{a}}^H\right\}\\ \mathrm{E}\left\{y^2\bar{\mathbf{a}}\mathbf{a}^H\right\} & \mathrm{E}\left\{\left(2\,|y|^2-\rho\,I\right)\bar{\mathbf{a}}\bar{\mathbf{a}}^H\right\}\end{pmatrix}.\end{aligned}$$

With the assumption that $\mathrm{rank}(\mathbf{C}) = N$, we have

$$\mathrm{rank}\left(\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}}_*)\right) = \mathrm{rank}\left(\Pi_{\mathbf{cc}}(\hat{\mathbf{w}}_*)\right)\,.$$

From (19) we know that

$$\mathrm{rank}(\Pi_{\mathbf{cc}}) = \mathrm{rank}(\Pi_{\mathbf{aa}}) + \mathrm{rank}(\widetilde{\Pi}_{\mathbf{aa}})\,, \tag{59}$$

where the Schur complement, $\widetilde{\Pi}_{\mathbf{aa}}$, of $\Pi_{\mathbf{aa}}$ is given by

$$\widetilde{\Pi}_{\mathbf{aa}} = \Pi_{\mathbf{aa}} - \Pi_{\bar{\mathbf{a}}\mathbf{a}}\Pi^{-1}_{\bar{\mathbf{a}}\bar{\mathbf{a}}}\Pi_{\mathbf{a}\bar{\mathbf{a}}}\,.$$

Now

$$\frac{1}{2}\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}}_*) = \mathbf{C}^H\Pi_{\mathbf{aa}}(\hat{\mathbf{w}}_*)\mathbf{C}$$

so it is evident that $\Pi_{\mathbf{aa}}$ has already been computed in Equation (57). We have

$$\Pi_{\mathbf{aa}}(\hat{\mathbf{w}}_*) = \Pi_{\bar{\mathbf{a}}\bar{\mathbf{a}}}(\hat{\mathbf{w}}_*) = \mathrm{Diag}\left(\mu_4,\quad T_D,\quad \cdots,\quad T_D\right). \tag{60}$$

It is also straightforward to determine that $\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\hat{\mathbf{w}}_*)$ is the diagonal matrix

$$\begin{aligned}\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\hat{\mathbf{w}}_*) &= \mathrm{Diag}\left(\mu_4,\quad \left|\mathrm{E}\left\{a(n)^2\right\}\right|^2,\quad \cdots,\quad \left|\mathrm{E}\left\{a(n)^2\right\}\right|^2\right),\\ &= \mathrm{Diag}\left(\mu_4,\quad \nu_2^2,\quad \cdots,\quad \nu_2^2\right),\end{aligned} \tag{61}$$

where

$$\nu_2 \triangleq \left|\mathrm{E}\left\{a(n)^2\right\}\right|\,.$$

This gives the diagonal matrix

$$\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\hat{\mathbf{w}}_*)\Pi^{-1}_{\bar{\mathbf{a}}\bar{\mathbf{a}}}(\hat{\mathbf{w}}_*)\Pi_{\mathbf{a}\bar{\mathbf{a}}}(\hat{\mathbf{w}}_*) = \mathrm{Diag}\left(\mu_4,\quad K_D,\quad \cdots,\quad K_D\right),$$

where

$$K_D = \frac{\nu_2^4}{\mu_2^2(2-\kappa)},$$

showing that the Schur complement $\widetilde{\Pi}_{\mathbf{aa}}(\hat{\mathbf{w}}_*)$ is given by the real diagonal matrix,

$$\widetilde{\Pi}_{\mathbf{aa}}(\hat{\mathbf{w}}_*) = \mathrm{Diag}\left(0,\quad C_d,\quad \cdots,\quad C_d\right),$$

where

$$C_d = \mu_2^2(2-\kappa) - \frac{\nu_2^4}{\mu_2^2(2-\kappa)}.$$

Note that if

$$\nu_2^4 < \mu_2^4(2-\kappa)^2$$

then $\mathrm{rank}(\widetilde{\Pi}_{\mathbf{aa}})(\hat{\mathbf{w}}_*) = N-1$. Using the earlier derived condition that we must have $\kappa < 2$ and noting that $\kappa > 0$, we can rewrite our new condition as

$$\kappa < 2 - \frac{\nu_2^2}{\mu_2^2}\,,$$

which is even more stringent than our previously derived condition for the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}}_*)$ to be full rank. Collecting our results, we have shown that for the noiseless FIR channel model, $\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}}_*) > 0$ and $\widetilde{\mathcal{H}}_{\mathbf{ww}}(\hat{\mathbf{w}}_*) \geq 0 \Leftarrow \operatorname{rank} \mathbf{C} = N$ and

$$\kappa < 2 - \frac{\nu_2^2}{\mu_2^2} \Rightarrow \operatorname{rank}(\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\hat{\mathbf{w}}_*)) = 2N - 1\,. \tag{62}$$

Furthermore, from Equation (31) we know that the one-dimensional nullspace of the full Hessian $\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}}_*)$ is spanned by the vector

$$\hat{\mathbf{c}}_{\text{null}} = \begin{pmatrix} \hat{\mathbf{w}}_* e^{\frac{j\pi}{2}} \\ \bar{\hat{\mathbf{w}}}_* e^{\frac{-j\pi}{2}} \end{pmatrix}\,. \tag{63}$$

It is reasonable to assume that the $2N-1$ rank property of the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ at the perfectly equalizing solution $\hat{\mathbf{w}}_*$ shown in (62) implies, giving the smooth nature of the problem, that the full Hessian has rank $2N-1$ and the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ has full rank N throughout an entire open neighborhood of the perfectly equalizing solution $\hat{\mathbf{w}}_*$, thereby implying that $\hat{\mathbf{w}}_*$ is locally optimal over a convex neighborhood. This heuristic argument can be made rigorous via a perturbation analysis about the point $\hat{\mathbf{w}}_*$. This result is important because, as noted in [14], it implies that, at least locally, stabilized generalized gradient descent algorithms, including the pseudo-Newton algorithm, will converge to a solution $\hat{\mathbf{w}}$ which is equivalent to $\hat{\mathbf{w}}_*$, $\hat{\mathbf{w}} \in [\hat{\mathbf{w}}_*]$. However, the solution is not unique, but is one of the equivalent solutions in the set $[\hat{\mathbf{w}}_*]$, which under the full column rank assumption on the channel matrix can be shown to be an equivalence class.

This lack of uniqueness means that without further restrictions on the pseudo-Newton algorithm, we cannot guarantee that we learn $\hat{\mathbf{w}}_*$ itself. As we have seen, the stationary solutions of the CM loss function are equivalent modulo a multiplication by a unimodular scalar. Thus, for $\alpha = e^{j\phi} \in \mathbb{C}$, $|\alpha| = 1$, we have that $\alpha\hat{\mathbf{w}}_*$ is also a solution, $\alpha\hat{\mathbf{w}}_* \in [\hat{\mathbf{w}}_*]$. The pseudo-Newton algorithm does not resolve this ambiguity, but can only at best ensure that we converge to some value in $[\hat{\mathbf{w}}_*]$ as positive-definiteness of the leading principle Hessian $\mathcal{H}_{\mathbf{ww}}(\hat{\mathbf{w}}_*)$ is sufficient to guarantee that the pseudo-Newton algorithm can converge to some element in the set $[\hat{\mathbf{w}}_*]$. The fact that the full Hessian is intrinsically singular means that we can do no better than this without a modification of the pseudo-Newton algorithm (36). This is, nonetheless, a very positive result and explains the positive performance of the pseudo-Newton algorithm described by [7]. Note that if the algorithm converges to the particular solution in $[\hat{\mathbf{w}}_*]$ given by

$$\alpha\,\hat{\mathbf{w}}_* = e^{j\phi}\hat{\mathbf{w}}_*$$

we have

$$y(n) = (\alpha\hat{\mathbf{w}}_*)^H\mathbf{x} = e^{-j\phi}\hat{\mathbf{w}}_*^H\mathbf{C}^H\mathbf{a} = e^{-j\phi}\mathbf{e}_0^H\mathbf{a} = e^{-j\phi}a(n),$$

showing that we can expect to attain perfect equalization modulo a phase-shift of the channel input data.

Under the assumption of a noiseless FIR channel, we have evaluated the Full and partial Hessians at the perfectly equalizing solution $\hat{\mathbf{w}}_*$. We now evaluate these Hessians at an arbitrary value of $\mathbf{w}$ under the FIR channel model and determine whether the Hessians evaluated at a point $\mathbf{w}$ are invariant with respect to a multiplication of $\mathbf{w}$ by a unimodular scalar. Note that knowledge of $\mathsf{A}(\mathbf{w})$ and B is sufficient to construct all the relevant Hessians. It is straightforward to compute the output signal covariance matrix B,

$$\mathsf{B} = \mathbf{C}^H\mathrm{E}\left\{\mathbf{a}\mathbf{a}^H\right\}\mathbf{C} = \mu_2\,\mathbf{C}^H\mathbf{C}\,, \tag{64}$$

from which, as noted before, $\operatorname{rank}(\mathsf{B}) = \operatorname{rank}(\mathbf{C})$. To compute the matrix

$$\mathsf{A}(\mathbf{w}) = \mathbf{C}^H\mathrm{E}\left\{|y|^2\,\mathbf{a}\mathbf{a}^H\right\}\mathbf{C}\,, \quad y = \mathbf{w}^H\mathbf{x} = \mathbf{w}^H\mathbf{C}^H\mathbf{a}$$

is rather more work. To do so, we first define the composite channel $\mathbf{f} = \mathbf{C}\mathbf{w}$, so that $y = \mathbf{f}^H\mathbf{a}$. Note that

$$|y|^2 = \bar{y}y = y^Hy = \mathbf{a}^H\mathbf{f}\mathbf{f}^H\mathbf{a} = \sum_{i,j}\bar{a}_i a_j f_i \bar{f}_j\,,$$

so that the k,ℓ-th component of $|y|^2\,\mathbf{a}\mathbf{a}^H$ is given by

$$\left(|y|^2\,\mathbf{a}\mathbf{a}^H\right)_{k,\ell} = \sum_{i,j}\bar{a}_i a_j f_i \bar{f}_j a_k \bar{a}_\ell\,.$$

This enables us to compute the k,ℓ-th component of $\mathrm{E}\left\{|y|^2\,\mathbf{a}\mathbf{a}^H\right\}$ for $k \neq \ell$ as

$$\mathrm{E}\left\{|y|^2\,\mathbf{a}\mathbf{a}^H\right\}_{k,\ell} = \mu_2^2 f_k\bar{f}_\ell + \nu_2^2\bar{f}_k f_\ell$$

and for $k = \ell$ as

$$\begin{aligned}\mathrm{E}\left\{|y|^2\,\mathbf{a}\mathbf{a}^H\right\}_{k,k} &= \mu_4\,|f_k|^2 + \left(\|f\|^2 - |f_k|^2\right)\mu_2^2 \\ &= \left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)|f_k|^2 + \mu_2^2\,|f_k|^2 + \nu_2^2\,|f_k|^2 + \mu_2^2\,\|f\|^2\end{aligned}$$

which we can concisely write as

$$\begin{aligned}\mathrm{E}\left\{|y|^2\,\mathbf{a}\mathbf{a}^H\right\} &= \left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\mathrm{Diag}\left(|f_0|^2, \cdots, |f_{N-1}|^2\right) + \\ &\qquad \mu_2^2\mathbf{f}\mathbf{f}^H + \nu_2^2\bar{\mathbf{f}}\bar{\mathbf{f}}^H + \mu_2^2\,\|\mathbf{f}\|^2\,I \qquad (65)\\ &\triangleq \left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\mathrm{Diag}\,|\mathbf{f}|^2 + \mu_2^2\mathbf{f}\mathbf{f}^H + \nu_2^2\bar{\mathbf{f}}\bar{\mathbf{f}}^H + \mu_2^2\mathbf{f}^H\mathbf{f}\,I \\ &= \left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\mathrm{Diag}\,|\mathbf{C}\mathbf{w}|^2 + \mu_2^2\mathbf{C}\mathbf{w}\mathbf{w}^H\mathbf{C}^H + \\ &\qquad \nu_2^2\bar{\mathbf{C}}\bar{\mathbf{w}}\bar{\mathbf{w}}^H\bar{\mathbf{C}}^H + \mu_2^2\mathbf{w}^H\mathbf{C}^H\mathbf{C}\mathbf{w}\,I. \qquad (66)\end{aligned}$$

Thus

$$\begin{aligned}\mathsf{A}(\mathbf{w}) = \mathbf{C}^H\big(&\left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\mathrm{Diag}\,|\mathbf{C}\mathbf{w}|^2 + \mu_2^2\mathbf{C}\mathbf{w}\mathbf{w}^H\mathbf{C}^H \\ &+ \nu_2^2\bar{\mathbf{C}}\bar{\mathbf{w}}\bar{\mathbf{w}}^H\bar{\mathbf{C}}^H + \mu_2^2\mathbf{w}^H\mathbf{C}^H\mathbf{C}\mathbf{w}\,I\big)\mathbf{C},\end{aligned} \tag{67}$$

which we can write in terms of the composite channel $\mathbf{f} = \mathbf{C}\mathbf{w}$ as

$$\begin{aligned}\mathsf{A}(\mathbf{w}) = \mathbf{C}^H\big(&\left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\mathrm{Diag}\,|\mathbf{f}|^2 + \mu_2^2\mathbf{f}\mathbf{f}^H + \\ &\nu_2^2\bar{\mathbf{f}}\bar{\mathbf{f}}^H + \mu_2^2\mathbf{f}^H\mathbf{f}\,I\big)\mathbf{C}\,.\end{aligned} \tag{68}$$

Note that $\mathsf{A}(\mathbf{w})$ is invariant with respect to multiplication of $\mathbf{w}$ by a unimodular scalar, and therefore so is the leading partial Hessian

$$\frac{1}{2}\,\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) = 2\,\mathsf{A}(\mathbf{w}) - \rho\,\mathsf{B}\,.$$

We can write the leading partial Hessian in terms of the equalizer tap vector $\mathbf{w}$ as

$$\frac{1}{2}\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) = \mathbf{C}^H\Pi_{\mathbf{aa}}(\mathbf{w})\mathbf{C}, \tag{69}$$

$$\Pi_{\mathbf{aa}}(\mathbf{w}) = 2\left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\operatorname{Diag}|\mathbf{Cw}|^2 + 2\mu_2^2\mathbf{Cww}^H\mathbf{C}^H + 2\nu_2^2\bar{\mathbf{C}}\bar{\mathbf{w}}\bar{\mathbf{w}}^H\bar{\mathbf{C}}^H + \mu_2^2(2\mathbf{w}^H\mathbf{C}^H\mathbf{Cw} - \kappa)\, I. \tag{70}$$

Note that, as expected, the resulting expression for the leading partial Hessian evaluated at $\mathbf{w}$ is invariant with respect to multiplication of $\mathbf{w}$ by a unimodular scalar. The matrix $\Pi_{\mathbf{aa}}(\mathbf{w})$ can be written in terms of the composite channel $\mathbf{f} = \mathbf{Cw}$ as

$$\Pi_{\mathbf{aa}}(\mathbf{w}) = 2\left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\operatorname{Diag}|\mathbf{f}|^2 + 2\mu_2^2\mathbf{ff}^H + 2\nu_2^2\bar{\mathbf{f}}\bar{\mathbf{f}}^H + \mu_2^2(2\mathbf{f}^H\mathbf{f} - \kappa)\, I. \tag{71}$$

Note that at the zero-delay perfectly equalizing solution

$$\mathbf{e}_0 = \hat{\mathbf{f}}_* = \mathbf{C}\hat{\mathbf{w}}_* \tag{72}$$

Equations (70) and (71) reduces to (60) as required. Note that a sufficient condition that the leading partial Hessian $\frac{1}{2}\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ be full rank in general, assuming a full-column rank channel matrix $\mathbf{C}$, is that

$$\Pi_{\mathbf{aa}}(\mathbf{w}) = 2\left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\operatorname{Diag}|\mathbf{Cw}|^2 + 2\mu_2^2\mathbf{Cww}^H\mathbf{C}^H + 2\nu_2^2\bar{\mathbf{C}}\bar{\mathbf{w}}\bar{\mathbf{w}}^H\bar{\mathbf{C}}^H + \mu_2^2(2\mathbf{w}^H\mathbf{C}^H\mathbf{Cw} - \kappa)\, I > 0, \tag{73}$$

which we can write in terms of the composite channel $\mathbf{f} = \mathbf{Cw}$ as

$$\Pi_{\mathbf{aa}}(\mathbf{w}) = 2\left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\operatorname{Diag}|\mathbf{f}|^2 + 2\mu_2^2\mathbf{ff}^H + 2\nu_2^2\bar{\mathbf{f}}\bar{\mathbf{f}}^H + \mu_2^2(2\mathbf{f}^H\mathbf{f} - \kappa)\, I > 0\,. \tag{74}$$

Condition (73) is complicated in form and requires knowledge of the channel matrix $\mathbf{C}$ to use as a test to determine if the leading partial Hessian is guaranteed to have full rank when evaluated at a particular value of the equalizer tap-weight vector $\mathbf{w}$. Of course, it is reasonable to expect that positive-definiteness of the leading partial Hessian will depend on the particular form of the channel itself. However, for special situations it may be the case that (other than the full column rank requirement on $\mathbf{C}$) the condition becomes independent of the channel matrix. For example, we have seen that this is the case when the equalizer is the zero-delay perfectly equalizing filter given by (75). Indeed, it is straightforward to show that condition (73) is independent of the channel matrix for any perfectly equalizing value of $\mathbf{w}$,

$$\mathbf{e}_\delta = \hat{\mathbf{f}}_*^{(\delta)} = \mathbf{C}\hat{\mathbf{w}}_*^{(\delta)}\,, \quad \delta = 0, \cdots, N-1. \tag{75}$$

To compute the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ and the Schur complement of the leading partial Hessian $\widetilde{\mathcal{H}}_{\mathbf{ww}}(\mathbf{w})$ requires the additional knowledge of

$$\mathcal{H}_{\bar{\mathbf{w}}\mathbf{w}}(\mathbf{w}) = \mathsf{C}(\mathbf{w}) = \mathbf{C}^H\operatorname{E}\left\{\bar{y}^2\mathbf{a}\bar{\mathbf{a}}^H\right\}\overline{\mathbf{C}} = \mathbf{C}^H\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\mathbf{w})\overline{\mathbf{C}}\,.$$

To compute the k,ℓ-th component of

$$\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\mathbf{w}) = \operatorname{E}\left\{\bar{y}^2\mathbf{a}\bar{\mathbf{a}}^H\right\} = \operatorname{E}\left\{\bar{y}^2\mathbf{a}\mathbf{a}^T\right\}\,,$$

note that

$$\bar{y}^2 = \bar{y}\bar{y} = \bar{y}^T\bar{y} = \bar{\mathbf{a}}^T\mathbf{ff}^T\bar{\mathbf{a}} = \sum_{i,j}\bar{a}_i\bar{a}_j f_i f_j$$

and

$$\left(\bar{y}^2\mathbf{aa}^T\right)_{k,\ell} = \sum_{i,j}\bar{a}_i\bar{a}_j f_i f_j a_k a_\ell\,.$$

The k,ℓ-th component of $\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\mathbf{w}) = \operatorname{E}\left\{\bar{y}^2\mathbf{aa}^T\right\}$ for $k \neq \ell$ can then be computed to be

$$\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\mathbf{w})_{k,\ell} = \operatorname{E}\left\{\bar{y}^2\mathbf{aa}^T\right\}_{k,\ell} = 2\mu_2^2 f_k f_\ell$$

and for $k = \ell$ to be

$$\begin{aligned}\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\mathbf{w})_{k,k} &= \operatorname{E}\left\{\bar{y}^2\mathbf{aa}^T\right\}_{k,k} = \mu_4 f_k^2 + \nu_2^2\left(\mathbf{f}^T\mathbf{f} - f_k^2\right)\\ &= \left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)f_k^2 + 2\mu_2^2 f_k^2 + \nu_2^2\mathbf{f}^T\mathbf{f}\,.\end{aligned}$$

Taken together, this yields

$$\begin{aligned}\Pi_{\bar{\mathbf{a}}\mathbf{a}}(\mathbf{w}) &= \left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\operatorname{Diag}\left(f_0^2, \cdots, f_{N-1}^2\right) + 2\,\mu_2^2\,\mathbf{ff}^T + \nu_2^2\,\mathbf{f}^T\mathbf{f}\, I,\\ &\triangleq \left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\operatorname{Diag}\mathbf{f}^2 + 2\mu_2^2\,\mathbf{ff}^T + \nu_2^2\,\mathbf{f}^T\mathbf{f}\, I,\\ &= \left(\mu_4 - 2\mu_2^2 - \nu_2^2\right)\operatorname{Diag}\left(\mathbf{Cw}\right)^2 + 2\,\mu_2^2\mathbf{Cww}^T\mathbf{C}^T + \nu_2^2\mathbf{w}^T\mathbf{C}^T\mathbf{Cw}\, I\,.\end{aligned} \tag{76}$$

If the channel matrix $\mathbf{C}$ of the noiseless FIR channel model does not have full column rank, then it is evident from Equation (50) that gradient descent-based learning will prematurely turn off (or slow down) if $\mathbf{w}$ is in (or close to) the nullspace of $\mathbf{C}$, $\mathbf{w} \in \mathcal{N}(\mathbf{C})$. Furthermore, it is evident from (51) that the leading partial Hessian is noninvertible, so that the unmodified pseudo-Newton algorithm (36) cannot be implemented. Because B is rank deficient, the set of stationary points $\mathcal{W}'$ changes in undesirable ways; the trivial solution $\hat{\mathbf{w}} = \mathbf{0}$ is not excludable as a isolated local maximum (because the Hessian evaluated at $\mathbf{0}$ is singular) and there can exist nontrivial solutions which do not satisfy the condition (28) but which results in the indeterminate condition $\rho = 0/0$. Finally, a good solution $\hat{\mathbf{w}} \in [\hat{\mathbf{w}}_*]$ can escape the class $[\hat{\mathbf{w}}_*]$ of perfectly equalizing solutions by moving in a direction $\delta\mathbf{w}$ in the nullspace of the channel matrix $\mathbf{C}$ where this change in the equalizer coefficient vector from the good solution $\hat{\mathbf{w}}$ to $\hat{\mathbf{w}} + \delta\mathbf{w}$ does not change the value of the CM loss function (4), yet will result in poorer equalization performance.

VII. Regularizing the Maximal-Rank Full Newton Algorithm

In this section we consider the special case of regularizing the full Hessian with a rank of $2N-1$ (maximum possible rank). This corresponds to the case of a full rank channel, $\mathsf{B} = \operatorname{E}\left\{\mathbf{xx}^H\right\}$ (i.e., the channel matrix $\mathbf{C}$ is one-one). We consider minimizing the following regularized loss function

$$\ell'(\mathbf{w},\gamma) = \ell(\mathbf{w}) + 2\,\gamma\left(\operatorname{Re}\boldsymbol{\beta}^H\mathbf{w}\right)^2 = \ell(\mathbf{w}) + 2\gamma\left(\frac{\boldsymbol{\beta}^H\mathbf{w} + \mathbf{w}^H\boldsymbol{\beta}}{2}\right)^2, \tag{77}$$

which we obtain from the CM loss function by the addition of a phase enforcing penalty term. Driving the penalty term in (77) to zero sets an overall phase requirement on the scalar $\boldsymbol{\beta}^H\mathbf{w}$ via the requirement

$$\operatorname{Re}\boldsymbol{\beta}^H\mathbf{w} = \left(\frac{\beta^H\mathbf{w} + \mathbf{w}^H\beta}{2}\right) = 0 \tag{78}$$

for a prespecified vector $\boldsymbol{\beta}$ which can be attained by a multiplication of $\mathbf{w}$ by an appropriate unimodular scalar. For example, if one chooses $\boldsymbol{\beta} = \mathbf{e}_0$, then the requirement becomes equivalent to demanding that the first component of

$\mathbf{w}$, $w(0)$, is purely imaginary. Without any *a priori* reason to impose a pure imaginary or pure real requirement on any particular component of $\mathbf{w}$, a more balanced choice of $\boldsymbol{\beta}$ would be $\boldsymbol{\beta} = \mathbf{1} = (1, \cdots, 1)^T$, yielding the phase requirement that

$$\mathrm{Re}\,\mathbf{1}^H\mathbf{w} = \sum_{k=1}^{N} \mathrm{Re}\,w(k) = 0\,.$$

Via a simple unimodular (and hence harmless) rescaling, this requirement can be imposed on each member of the sequence of estimates, $\hat{\mathbf{w}}_k$, provided by the pseudo-Newton algorithm, on occasional members of the sequence of estimates, or only on the final converged, estimation provided by the pseudo-Newton algorithm. Note that the penalty term removes the ambiguity associated with the phase and subsequently has the benefit of removing Hessian singularity.

For positive, nonzero values of γ in (77), this modification will ensure a nonsingular full Hessian, enabling the full Newton algorithm to be implemented, typically with less cost than computing the pseudoinverse. Note that the penalty term is nonnegative $\left(\mathrm{Re}\,\boldsymbol{\beta}^H\mathbf{w}\right)^2 \geq 0$ and has a minimum value of zero. Let

$$\boldsymbol{\beta}^H\mathbf{w} = \left|\boldsymbol{\beta}^H\mathbf{w}\right| e^{j\theta} = \left|\boldsymbol{\beta}^H\mathbf{w}\right| (\cos(\theta) + j\,\sin(\theta))\,.$$

Then for $\theta \neq n\frac{\pi}{2}$, $n = \pm 1, \pm 2, \cdots$, we have $\left(\mathrm{Re}\,\boldsymbol{\beta}^H\mathbf{w}\right)^2 \neq 0$. Note that multiplication of $\mathbf{w}$ by the unimodular scalar $\alpha = e^{j(\frac{\pi}{2}-\theta)}$ results in a vector $\mathbf{w}' = \alpha\,\mathbf{w}$, for which the penalty function attains its minimum value of zero,

$$\begin{aligned}\left(\mathrm{Re}\,\boldsymbol{\beta}^H\mathbf{w}'\right)^2 &= \left(\mathrm{Re}\,\alpha\boldsymbol{\beta}^H\mathbf{w}\right)^2 = \left(\mathrm{Re}\,e^{j(\frac{\pi}{2}-\theta)}\left|\boldsymbol{\beta}^H\mathbf{w}\right| e^{j\theta}\right)^2 \\ &= \left(\mathrm{Re}\,\alpha\boldsymbol{\beta}^H\mathbf{w}\right)^2 \left(\mathrm{Re}\,j\left|\boldsymbol{\beta}^H\mathbf{w}\right|\right)^2 = 0.\end{aligned}$$

This shows that if $\hat{\mathbf{w}}$ is a local minimizer of the un-regularized CM loss function

$$\ell(\mathbf{w}) = \ell'(\mathbf{w}, 0)$$

then there exists a unimodular scalar α, such that $\hat{\mathbf{w}}' = \alpha\,\hat{\mathbf{w}} \in [\hat{\mathbf{w}}]$ is also a local minimizer of the CM loss function, but one for which the regularizing term vanishes, $\left(\mathrm{Re}\,\boldsymbol{\beta}^H\hat{\mathbf{w}}'\right)^2 = 0$. Furthermore, for all $\mathbf{w}$ in a sufficiently small neighborhood of $\hat{\mathbf{w}}'$

$$\begin{aligned}\ell'(\mathbf{w},\gamma) = \ell(\mathbf{w}) + 2\gamma\left(\mathrm{Re}\,\boldsymbol{\beta}^H\mathbf{w}\right)^2 &\geq \ell(\hat{\mathbf{w}}') + 2\gamma\left(\mathrm{Re}\,\boldsymbol{\beta}^H\mathbf{w}\right)^2 \\ &\geq \quad \ell(\hat{\mathbf{w}}') = \ell'(\hat{\mathbf{w}}',\gamma)\,,\end{aligned}$$

showing that $\hat{\mathbf{w}}'$ is a local minimum of the regularized loss function $\ell'(\mathbf{w},\gamma)$. Within the set of equivalent local minimizers $[\hat{\mathbf{w}}]$ of the un-regularized CM loss function, there are two isolated points which have the property that they are also minimizers of the regularized loss function $\ell'(\mathbf{w},\gamma)$.

It is straightforward to compute the relevant quantities needed to implement the pseudo-Newton and Newton algorithms for purposes of learning a minimum of the regularized loss function. The gradient of the regularized CM loss function is given by

$$\nabla_{\mathbf{c}}\ell'(\mathbf{c},\gamma) = \begin{pmatrix}\nabla_{\mathbf{w}}\ell'(\mathbf{w},\gamma) \\ \overline{\nabla_{\mathbf{w}}\ell'(\mathbf{w},\gamma)}\end{pmatrix}, \tag{79}$$

where

$$\begin{aligned}\frac{1}{2}\nabla_{\mathbf{w}}\ell'(\mathbf{w},\gamma) &= \frac{1}{2}\nabla_{\mathbf{w}}\ell(\mathbf{w}) + \gamma\,\boldsymbol{\beta}\,\mathrm{Re}\left(\boldsymbol{\beta}^H\mathbf{w}\right) \\ &= (\mathsf{A}(\mathbf{w}) - \rho\mathsf{B})\,\mathbf{w} + \gamma\,\boldsymbol{\beta}\,\mathrm{Re}\left(\boldsymbol{\beta}^H\mathbf{w}\right)\,.\end{aligned}$$

Furthermore, The leading partial Hessian $\mathcal{H}'_{\mathbf{ww}}(\mathbf{w},\gamma)$ is given by

$$\begin{aligned}\frac{1}{2}\mathcal{H}'_{\mathbf{ww}}(\mathbf{w},\gamma) &= \frac{1}{2}\mathcal{H}_{\mathbf{ww}}(\mathbf{w}) + \frac{\gamma}{2}\boldsymbol{\beta\beta}^H \\ &= 2\,\mathsf{A}(\mathbf{w}) - \rho\,\mathsf{B} + \frac{\gamma}{2}\boldsymbol{\beta\beta}^H\,,\end{aligned}$$

and the full Hessian of the regularized CM loss function is given by

$$\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w},\gamma) = \mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w}) + \gamma\begin{pmatrix}\boldsymbol{\beta}\\ \bar{\boldsymbol{\beta}}\end{pmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \bar{\boldsymbol{\beta}}\end{pmatrix}^H = \mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w}) + \gamma\,\boldsymbol{\xi\xi}^H.$$

Note that if the unregularized full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ is positive semidefinite, then the regularized full Hessian will be positive definite for any nonzero value of γ provided that ξ has a nontrivial component in the one-dimensional null space of $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w})$. If the vector $\boldsymbol{\xi}$ is chosen at random (equivalently, if $\boldsymbol{\beta}$ is chosen at random), then generically the rank of the regularized Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w},\gamma)$ will be increased by one over the rank of the unregularized Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$. In particular, if the unregularized Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ has maximal rank, $\mathrm{rank}\,\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w}) = 2N-1$, then generically the regularized Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w},\gamma)$ will have full rank, $\mathrm{rank}\,\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w},\gamma) = 2N$. In actual implementation of the full Newton algorithm one often additionally stabilizes the algorithm via the addition of a small regularizing term $\epsilon\,\boldsymbol{I}$, in which case $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w},\gamma)$ takes the following form

$$\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w},\gamma,\epsilon) = \mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w}) + \epsilon\,\boldsymbol{I} + \gamma\,\boldsymbol{\xi\xi}^H. \tag{80}$$

The principal steps involved in the implementation are:

Algorithm A1:

$$\begin{aligned}\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w}_k,\gamma\,\epsilon) &= \text{Full Hessian given in (80)},\\ \nabla_{\mathbf{c}}\ell'(\mathbf{c}_k,\gamma) &= \text{Full gradient given in (79)},\\ \text{Solve:}\quad \mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}{}'(\mathbf{w}_k,\gamma,\epsilon)\Delta\mathbf{c} &= -\nabla_{\mathbf{c}}\ell'(\mathbf{c}_k,\gamma)\,,\\ \mathbf{c}_{k+1} &= \mathbf{c}_k + \mu\Delta\mathbf{c}\,.\end{aligned}$$

The above algorithm implements the full Newton method. By replacing the full Hessian with the leading partial Hessian and the full gradient with the partial gradient, the partial Newton method can be implemented. Typically the Newton method uses a step size $\mu = 1$, however if the algorithm proves to be difficult to stabilize, as is commonly the case when stochastic approximations are made to the gradient and Hessians needed to implement the algorithm, then one takes $0 < \mu < 1$.

A solution $\hat{\mathbf{w}}_*$ is a stationary point of the unregularized CM loss function and minimizes the γ penalty term, it satisfies the conditions

$$\nabla_{\mathbf{w}}\ell(\hat{\mathbf{w}}_*) = \mathbf{0} \quad \text{and} \quad \mathrm{Re}\left(\boldsymbol{\beta}^H\hat{\mathbf{w}}_*\right) = 0\,.$$

Thus we have

$$\nabla_{\mathbf{w}}\ell'(\hat{\mathbf{w}}_*,\gamma,0) = \nabla_{\mathbf{w}}\ell(\hat{\mathbf{w}}_*) + 2\gamma\boldsymbol{\beta}\,\mathrm{Re}\left(\boldsymbol{\beta}^H\hat{\mathbf{w}}_*\right) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

showing that $\hat{\mathbf{w}}_*$ is indeed a stationary point of the regularized CM loss function $\ell'(\mathbf{w},\gamma)$, as claimed.

VIII. Simulation Results

In this section we simulate the blind equalization of two full rank channels using the stochastic gradient, partial Hessian and full Hessian approaches. The performance metric is taken to be the intersymbol interference (ISI),

$$ISI = \frac{\sum_i |s_i|^2 - |s|^2_{max}}{|s|^2_{max}}, \tag{81}$$

where s_i is i^{th} element of the convolution between the channel and the equalizer and $s_{max} = \max_i s_i$.

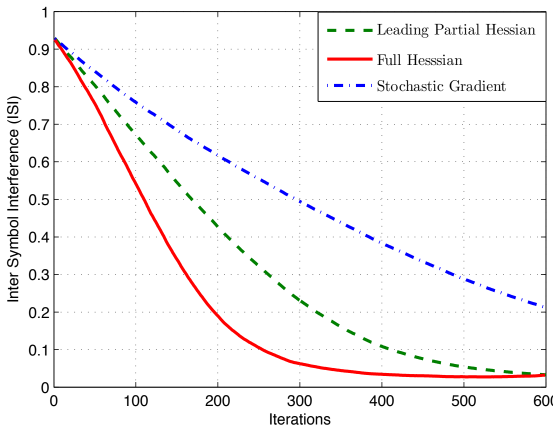

Fig. 1. The ISI comparison of the gradient, the leading partial Hessian and the full Hessian algorithms for channel.1

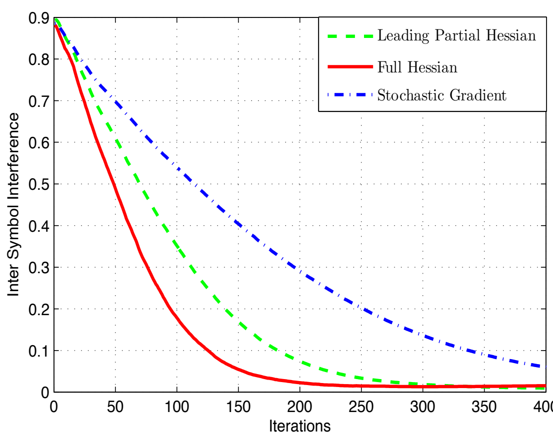

Fig. 2. The ISI comparison of the gradient, the leading partial Hessian and the full Hessian algorithms for channel.2

The channel used in simulations shown in Figure.1 ("channel 1") is the same as the channel simulated in reference [7] and has an impulse response of $[0.04, -0.05, 0.07, -0.21, -0.5, 0.72, 0.36, 0, 0.21, 0.03, 0.07]$ $\exp(j\pi/5)$. The channel length is $M = 11$. The input to the channel comes from a $4 - QAM$ constellation with $\rho = 1$ and we consider a symbol-rate sampled, $r = 1$ model. Because the channel is sampled at symbol rate, perfect noiseless equalization using an FIR filter is not possible [4]. We compare our results to [7] which implemented the pseudo-Newton algorithm (36). We use an equalizer of length $q = 21$ (hence $N = rq = 21$) with a corresponding leading partial Hessian of dimension $N \times N = 21 \times 21$ and full Hessian of dimension $2N \times 2N = 42 \times 42$. Note that $P = M + q - 1 = 31$, so that the $P \times N$ channel matrix $\mathbf{C}$ is a 31×21 matrix. For the channel 1 and for any randomly generated channel the channel matrix $\mathbf{C}$ has full column rank and hence $\mathbf{B}$ is full rank thus satisfying the ratio of moments condition given by (28).

The full Hessian, $\mathcal{H}^{\mathbb{C}\,\prime}_{\mathbf{cc}}(\mathbf{w}, \gamma\,\epsilon)$, is statistically approximated using the available received data. In our simulations we calculated the stochastic approximations each step as

$$\begin{aligned}
\mathcal{H}_{\mathbf{ww}}(\mathbf{w}_k) &= 2\,\mathrm{E}\left\{\left(2\,\mathbf{w}_k^H \mathbf{x}\mathbf{x}^H \mathbf{w}_k - \rho\right)\mathbf{x}\mathbf{x}^H\right\}\\
&= \frac{2(1-\lambda)}{1-\lambda^n}\sum_{i=1}^{n}\lambda^{n-i}\left[\left(2\,\mathbf{w}_k^H \mathbf{x}(i)\mathbf{x}^H(i)\mathbf{w}_k - \rho\right)\mathbf{x}(i)\mathbf{x}^H(i)\right],\\
\mathcal{H}_{\bar{\mathbf{w}}\bar{\mathbf{w}}}(\mathbf{w}_k) &= \overline{\mathcal{H}}_{\mathbf{ww}}(\mathbf{w}_k),\\
\mathcal{H}_{\mathbf{w}\bar{\mathbf{w}}}(\mathbf{w}_k) &= 2\,\mathrm{E}\left\{y^2\,\bar{\mathbf{x}}\mathbf{x}^H\right\}\\
&= \frac{2(1-\lambda)}{1-\lambda^n}\sum_{i=1}^{n}\lambda^{n-i}\left[y(i)^2\,\bar{\mathbf{x}}(i)\mathbf{x}^H(i)\right],\\
\mathcal{H}_{\bar{\mathbf{w}}\mathbf{w}}(\mathbf{w}_k) &= \overline{\mathcal{H}}_{\mathbf{w}\bar{\mathbf{w}}}(\mathbf{w}_k),\\
\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w}_k) &= \begin{pmatrix}\mathcal{H}_{\mathbf{ww}}(\mathbf{w}_k) & \mathcal{H}_{\bar{\mathbf{w}}\mathbf{w}}(\mathbf{w}_k)\\ \mathcal{H}_{\mathbf{w}\bar{\mathbf{w}}}(\mathbf{w}_k) & \mathcal{H}_{\bar{\mathbf{w}}\bar{\mathbf{w}}}(\mathbf{w}_k)\end{pmatrix}.
\end{aligned}$$

Alternatively one can implement an efficient on-line version of the algorithm along the lines of [7]. After obtaining the approximate Hessian, instead of directly inverting it, we use the following equation to solve for $\Delta\mathbf{c}$ (step 3 in Algorithm A1):

$$\mathcal{H}^{\mathbb{C}\,\prime}_{\mathbf{cc}}(\mathbf{w}_k, \gamma, \epsilon)\Delta\mathbf{c} = -\nabla_{\mathbf{c}}\ell'(\mathbf{c}_k, \gamma)\,\mathbf{c}_k.$$

Generally, the primary computation cost in implementing the algorithm comes from computing $\Delta\mathbf{c}$. When implementing the full Newton method we are dealing with a matrix of size $2N \times 2N$ as opposed to a $N \times N$ matrix when dealing with partial Newton method, hence the computational complexity of the full Hessian algorithm increases by no more than 4 times of the partial Hessian algorithm.

In Figure. 1, it can be seen that the regularized version of the full Hessian (which was shown to be rank-deficient) outperforms the leading partial Hessian as well as the stochastic gradient based algorithm. The parameters used in the implementation are $\epsilon = 10(0.9979)^{it}+0.01$, $\gamma = 0.99^{it}$, where 'it' implies the iteration number and the forgetting factor in the stochastic approximation is $\lambda = 0.95$. The curves shown are each averaged over 1000 trials. Initial conditions are set to be the same for all three algorithms: the 11^{th} co-efficient of the equalizer is chosen to be 1 and the rest of the co-efficients to be 0 (note that there are 21 taps in the equalizer).

The impulse response of the second simulated channel shown in Figure. 2 is given by an $M = 8$ FIR channel model $[0.03, -0.15, 0.09, -0.21, -0.5, -0.72, .36, -0.03]\exp(j\pi/3)$ and we use an equalizer of length $N = 12$. Again the input to the channel comes from a $4 - QAM$ constellation with $\rho = 1$ and we focus on a symbol-rate sampled, $r = 1$ model. The plots shown are each averaged over a 1000 trials.

In Figure 3 we plot the a phase of the function $\mathrm{Re}\mathbf{1}^H\mathbf{w}$, for both the phase constrained and unconstrained implementation of Newton algorithm (all the parameters are same as that of

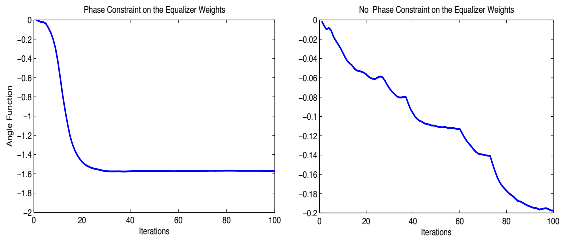

Fig. 3. Phase constraint imposed on the function of equalizer weights

channel 1 of Figure 1). The constrained system quickly converges to a constant phase, while the un-constrained system's phase function varies in a random manner.

The simulations shown are typical for a wide range of channels that we evaluated. The regularized full Newton algorithm yields improved performance while converging to a global phase constraint placed on the equalizer taps, with an increase in computational cost of no more than 4 times than the pseudo-Newton algorithm.

IX. Conclusions

In this paper, we have examined the use of full Newton and pseudo-Newton algorithms for blind equalization of a noiseless complex linear time invariant (LTI) channel model to minimize the constant modulus (CM) loss function [4]. Noting the fact that the CM loss function is real with complex arguments, we use the framework of the second order Wirtinger calculus to derive a novel and insightful form of the full Hessian expressed in terms of its partial Hessians. We show that the perfectly equalizing solutions are stationary points of the CM loss function and we compute the forms of the partial Hessians (and hence the full Hessian) for noiseless LTI FIR channel model and evaluate them at a perfectly equalizing solution.

A necessary condition for the leading partial Hessian $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ to be positive definite is that the channel matrix $\mathbf{C}$ have full column-rank and if the channel matrix also has full row-rank, so that perfectly equalizing solutions exist, then a sufficient condition for $\mathcal{H}_{\mathbf{ww}}(\mathbf{w})$ to be positive definite at a perfectly equalizing solution is that the input data sequence be subgaussian $\kappa < 2$ [4]. Under this condition, the pseudo-Newton algorithm, based on the use of leading partial Hessian, can converge to the set of values $[\hat{\mathbf{w}}_*]$ which differ from the perfectly equalizing solution $\hat{\mathbf{w}}_*$ by a unimodular scalar factor. We also computed the form of the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$ at a perfectly equalizing solution of a full rank channel, showing that it attains the largest possible rank of $2N-1$, assuming a full rank leading partial Hessian, provided that the channel input data sequence in sufficiently subgaussian. In this case, the perfectly equalizing solutions are shown to be the local minima of the CM loss function and we also give the form of the vector which spans the nullspace of of the full Hessian $\mathcal{H}^{\mathbb{C}}_{\mathbf{cc}}(\mathbf{w})$.

We performed simulations on full rank channels and found that the full Newton method, based on a regularized full Hessian performs better than the pseudo-Newton method. The regularization of the full Hessian method is designed to enforce a phase constraint on the equalizer weights while not perturbing a CM optimal solution. The cost of the full Newton method is at most 4 times that of the pseudo-Newton method.

Finally, we note an outstanding problem, the solution of which would make the above developments relatively complete: For an LTI FIR channel model it remains to be shown which of the stationary solutions in $\mathcal{W}$ have a positive definite leading partial Hessian and/or a full Hessian of maximal rank $2N-1$ under the condition that the channel matrix $\mathbf{C}$ has full column rank. This is equivalent to identifying all of the local minima of the CM loss function under the channel matrix full-column rank condition. In this paper, we have determined these properties only for the perfectly equalizing solutions under the assumption that the channel matrix is square and full rank.

References

[1] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *IEEE Trans Communications.*, vol. 23, no. 6, pp. 679-682, Jun. 1975.

[2] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communications systems," *IEEE Trans Communications.*, vol. 28, no. 11, pp. 1867-1875, Nov. 1980.

[3] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans Acoustics, Speech, and Signal Processing.*, vol. 31, no. 2, pp. 459-472, Apr. 1983.

[4] C. R. Johnson Jr., P. Schniter, T .J. Edres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proceedings of the IEEE.*, vol. 86, no. 10, pp. 1927-1950, Oct. 1998.

[5] O. Dabeer and E. Masry, "Convergence Analysis of the Constant Modulus Algorithm Onkar Dabeer, Member,"*IEEE Trans Info Thoery.*, vol. 49, no. 6, pp. 1447-1464, June. 2003.

[6] R. Cusani and A. Laurenti, "Convergence Analysis of the CMA Blind Equalizer," *IEEE Trans Communnications.*, vol. 43, no. 2/3/4, pp. 1304-1307, Apr. 1995.

[7] G. Yan and H. Fan, "A newton-like algorithm for complex variables with applications in blind equalization," *IEEE Trans Signal Processing.*, vol. 48, no. 2, pp. 553-556, Feb. 2000.

[8] T. A. Schirtzinger and W.K. Jenkins, *Designing adaptive equalizers based on the constant modulus errorcriterion,* Proceedings of *IEEE ISCAS* 1995.

[9] T. A. Schirtzinger and W.K. Jenkins, *A comparison of three algorithms for blind equalization based on the constant modulus error criterion,* Proceedings of *IEEE ICASSP* 1995.

[10] M. T. M. Silva and M. D. Miranda, "Tracking issues of some blind equalization algorithms," *IEEE Signal Proc. Letters.*, vol. 11, no. 9, pp. 760-763, Sep. 2004.

[11] K. K-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," *Course Lecture Supplement No. ECE275CG-F05v1.2d.*, Sep-Dec 2005, Dept. of Electrical and Computer Engineering, UC San Diego. Available at: http://dsp.ucsd.edu/ ∼kreutz/PEI05.html

[12] D.H. Brandwood, "A Complex Gradient Operator and its Application in Adaptive Array Theory," *IEE Proceedings Microwaves, Optics, and Antennas.*, vol. 130, no. 1, pp. 11-16, Feb. 1983.

[13] A. van den Bos, "Complex Gradient and Hessian," *IEE Proc.-Vis. Image Signal Processing.*, vol. 141, no. 6, pp. 380-382, Dec. 1994.

[14] Z. Ding, R. A. Kennedy, B. D. O. Anderson, and C. R. Johnson Jr., "Ill-convergence of godard blind equalizers in data communication systems," *IEEE Trans Communications.*, vol. 39, no. 9, pp. 1313-1327, Sep. 1991.

[15] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd Edition, Johns Hopkins University Press, 1996.

[16] R. Remmert, *Theory of Complex Functions,*, Springer-Verlag, 1991.