Towards Globally Convergent Blind Equalization of Constant Modulus Signals: A Bilinear Approach

Constantinos B. PAPADIAS † and Dirk T. M. SLOCK ††

†, †† Institut EURECOM, 2229 route des Crêtes
B.P. 193, 06904, Sophia Antipolis Cedex, FRANCE
† Tel/Fax: +33 93002618/ 93002627, E-Mail: papadias@eurecom.fr
†† Tel/Fax: +33 93002606/ 93002627, E-Mail: slock@eurecom.fr

Abstract. We consider the problem of blind equalization of a constant modulus signal. One of the most popular classes of algorithms in this context is the Godard family of blind equalizers [1], which includes among others the Constant Modulus Algorithm (CMA) [2]. A common drawback of these algorithms is that they may converge to undesired equalizer settings if not properly initialized. This is known as the problem of ill convergence and is primarily due to the non-convex form of the cost function of algorithms of this class with respect to the equalizer parameters. We propose a different approach to the problem, namely, a bilinear one, which leads to a different parameterization and to the construction of a convex cost function with respect to the parameters introduced. In a perfectly parameterized case (the equalizer’s order matches exactly the order of the channel inverse), the solution to the problem is unique and permits for a direct calculation of the optimal equalizer. In over-parameterized cases however, there exist multiple solutions to our cost function. However, we propose a method that still allows to determine the channel inverse in this case. Different adaptive schemes are proposed to adaptively compute the solution of our criterion and the influence of additive noise is also discussed.

1. The Constant Modulus Algorithm

The classical setup for adaptive blind-equalization is shown in figure 1. We denote by $a_k$, $x_k$ and $y_k$, the transmitted symbol, received sample and equalizer output, respectively, all at time instant $k$. The equalizer (a FIR filter of order $N-1$) at the same time instant is denoted by a $N \times 1$ vector $W_k = [w_0 \cdots w_{N-1}]^T$ and its output can be written as $y_k = X_k^H W_k$, where $X_k = [x_k x_{k-1} \cdots x_{k-N+1}]^H$, $X_k$ denotes transposed and $H$ complex conjugate transpose. The Godard algorithms are stochastic gradi-

Figure 1: An adaptive blind equalization setup

ent algorithms that attempt to solve the following minimization problem:

$$\min_W J_p(W) = \frac{1}{2p} E[|y|^p - L_p]^2 \quad p \in \{1, 2, \ldots \} \text{ ,}$$

where $E$ denotes statistical expectation and $L_p = \frac{E[|y|^p]}{E[|y|^p]}$. The popular CMA $2 \rightarrow 2$ corresponds to the particular choice $p = 2$ and if one assumes that the constellation modulus equals one ($|y_k| = 1$), is given by:

$$W_{k+1} = W_k - \mu X_k y_k (|y_k|^p - 1) \text{ ,}$$

where $\mu$ denotes the algorithm’s stepsize. Due to the non-convex form of the cost function $J_p(W)$ w.r.t. $W$, the problem [1] admits more than one solution which can be found by setting the gradient of $J_p(W)$ w.r.t. $W$ equal to zero. This gives for $p = 2$:

$$E[(|y_k|^p - 1)y_k X_k] = 0 \text{ .}$$

The stochastic gradient algorithm described by (2) may therefore converge to one of the solutions of (3) which does not correspond to the global but to a local minimum of $J_2(W)$ if it is initialized close to it and if a small stepsize $\mu$ (necessary for the stability of (2)) is used [3]. This is the problem of ill-convergence of CMA’s and in the sequel we will present a different approach that tries to circumvent it.

2. A bilinear approach

Consider the following expansion of the squared modulus of the equalizer’s output:

$$|y_k|^2 = y_k y_k^* = [w_0 w_0^* x_k x_k^* + \cdots + w_0 w_{N-1} x_k x_{N-1}^* + \cdots + (w_{N-1} w_{N-1}^* x_{N-1} x_{N-1}^*)],$$

where $^*$ denotes complex conjugate. If now we introduce a $N^2 \times 1$ bilinear regression vector

$$X_k = [x_k x_k^* \cdots x_k x_{N-1}^* \cdots x_k x_{N-1}^* \cdots x_k x_{N-1}^* [x_k x_{N-1}^* [x_k x_{N-1}^* H],$$

and a $N^2 \times 1$ parameter vector

$$\theta_k = [w_0 w_0^* w_0 w_0^* \cdots w_{N-1} w_{N-1}^* \cdots w_{N-1} w_{N-1}^* w_{N-1} w_{N-1}^*]^T,$$

that contain all the bilinear terms $x_k x_k^*$ and $w_i w_i^*$ of the expansion in (4), respectively, then the squared modulus of the output $y_k$ can be written as:

$$|y_k|^2 = z_k = X_k^H \theta_k \text{ .}$$
Then it follows that the cost function \( J(\theta) \) is quadratic, and therefore convex. The principle behind this parameterization of the problem is the following: in traditional blind equalization, the received discrete-time signal is first passed through a linear channel and then some kind of nonlinearity is applied to its output in order to provide the higher order moments needed for the identification of the channel (or the channel’s inverse) impulse response. Here we have somewhat “interchanged” the order of these two operations in that we first apply a nonlinearity to the received signal (in order to form the bilinear regression vector \( X_k \)) and then pass the resulting process through a linear filter whose output is not submitted to further nonlinearities.

The gradient of \( J(\theta) \) w.r.t. \( \theta \) is given by:
\[
\nabla_{\theta} J(\theta)(\theta) = 2E(X(X^H \theta - 1)) ,
\]
and therefore any solution \( \theta \) to (8) should satisfy the following equation:
\[
E(X X^H) \theta = E(X) .
\]

Now if the matrix \( E(X X^H) \) is invertible, the problem (8) has the following unique solution:
\[
\theta = (E(X X^H)^{-1} E(X) .
\]

In this case the corresponding equalizer can be found from \( \theta \) as follows: a zero-forcing (ZF) FIR equalizer of order (\( N - 1 \)) exists if the channel is all-pole of order \( (N-1) \):
\[
\alpha_k = \sum_{i=0}^{N-1} c_i x_{k-i} = X_k^H C ,
\]
where \( C = [c_0 \cdots c_{N-1}]^T \) contains the coefficients of the impulse response of the channel’s inverse. Then the optimal (ZF) equalizer will equal the inverse impulse response \( W_{\text{opt}} = C \) and the corresponding parameter vector will be given by:
\[
\theta_{\text{opt}} = [c_0^2 \cdots c_{N-2}^2 \quad c_1^2 \cdots c_{N-2}^2 \cdots c_{N-1}^2] .
\]

It is clear that \( \theta_{\text{opt}} \) solves (8) since \( J(\theta_{\text{opt}}) = 0 \). Moreover, as the matrix \( E(X X^H) \) was supposed to be invertible, (8) admits a unique solution which is given by (11). Therefore the following lemma holds:

**Lemma**: When the channel is all-pole of order \( N - 1 \) and the \( N^2 \times N^2 \) matrix \( E(X X^H) \) is invertible, the criterion (8) admits the unique solution \( \theta_{\text{opt}} \) given by (13).

Consider now the \( N \times N \) matrix \( \Theta \) that has as columns the \( N \) consecutive partitions of the vector \( \theta \) of \( N \) elements each:
\[
\Theta = \begin{bmatrix}
\theta & \theta(N) & \cdots & \theta(N^2-N) \\
\theta(1) & \theta(N+1) & \cdots & \theta(N^2-N+1) \\
\vdots & \vdots & \ddots & \vdots \\
\theta(N-1) & \theta(2 N - 1) & \cdots & \theta(N^2-1)
\end{bmatrix} .
\]

Then it follows that \( \Theta_{\text{opt}} = C C^H \). The optimal equalizer setting \( W_{\text{opt}} \) can now be found from \( \Theta_{\text{opt}} \) as follows:
\[
W_{\text{opt}} = \Theta \Theta^H \lambda_{\text{max}} V_{\text{max}} ,
\]
where \( \lambda_{\text{max}} \) and \( V_{\text{max}} \) denote the maximum eigenvalue and the corresponding eigenvector of \( \Theta_{\text{opt}} \) and \( \phi \in (0, 2\pi) \). Note that an ambiguity is inherent in the choice of the optimal equalizer, since the factor \( e^{j \phi} \) cannot be determined. This is a usual phenomenon in blind equalization and can be eliminated by using differential coding at transmission.

Identification of the (generally non-minimum phase) transmission channel is therefore achievable by minimizing the cost function \( J(\theta) \) w.r.t. \( \theta \). This should not be an astonishing result, since the matrix \( E(X X^H) \) contains \( N^2 \) order moments of the received signal (it is known that identification of a non-minimum phase channel is not possible at the baud rate by use of only second order statistics).

The solution of (8) can be calculated either in a batch or in an adaptive way. In the first case one just has to estimate the quantities \( E(X X^H) \) and \( E(X) \) and use (11) to calculate the corresponding estimate for \( \theta \). In the second case, an adaptive filtering algorithm can be employed for the parameter vector \( \theta \) that uses at each iteration \( \theta_k \) as the regression vector and the scalar \( 1 \) as the “desired” sample. An LMS-like algorithm (stochastic gradient minimization) for \( \theta_k \) will be as follows:
\[
\begin{align*}
\psi_k &= 1 - X^H_k \theta_k \\
\theta_{k+1} &= \theta_k + \mu \psi_k x_k ,
\end{align*}
\]
where \( \psi_k \) is the a priori error at time instant \( k \) and \( \mu \) the stepsize parameter that controls both the convergence speed and steady-state error of the algorithm. As the parameter vector \( \theta \) is of length \( N^2 \), the complexity of the above algorithm will be \( N^2 \) multiplications/iteration. Similarly, an RLS-like algorithm can be used that will allow for faster convergence (in approximately \( N^2 \) iterations) but at the expense of a higher computational complexity \((O(N^3)\) multiplications/iteration):
\[
\begin{align*}
\psi_k &= 1 - \lambda_k \theta_k \\
R_k &= \lambda_k^{-1} R_{k-1} - \lambda_k^{-1} R_{k-1} X_k (1 + \lambda_k^{-1} R_{k-1} X_k)^{-1} \times \lambda_k^{-1} R_{k-1} X_k \\
\theta_{k+1} &= \theta_k + R_k^{-1} \psi_k x_k ,
\end{align*}
\]
where \( R_k \) is \( N^2 \times N^2 \) and \( \lambda \) is the forgetting factor. A multi-channel Fast Transversal Filter algorithm [4] can be used in order to reduce the complexity to \((O(N)\) multiplications/iteration (see figure 2).

When the received samples \( x_k \) are real (which corresponds to a QAM modulation and a real channel), a reduction in the required number of parameters for the bilinear method is possible. Namely, in this case the expansion in (4) will have only \( N(3N-1)/2 \) terms, and therefore the regression and parameter vectors will be defined as :
\[
X_k = [\begin{array}{c}
x_k^2 x_k x_k - 1 \quad x_k x_k - 2 \quad x_k x_k - 2 \quad x_k - 2 \quad x_k - (N+1)\end{array}]^T .
\]

and
\[
\theta = [g(0)^T \ g(1)^T \ g(2 N - 1)^T]^T,
\]
respectively. All the above mentioned equations are still valid in this case, the only difference being the dimensions of \( \theta \), \( X \) and the matrix \( E(X X^H) \). All Hermitian transposes can also be replaced by simple transposes in this case. Figure 2 shows the setup for bilinear equalization in the real case, where the \((N(3N-1)/2)\) entries of \( \theta \) are organized in \( N \) equalizers of respective lengths \( N, N-1, \ldots, 1 \). Note the multichannel structure that allows also for a multichannel FFT algorithm as mentioned above.

### 3. Over-parameterized case

In this section we consider the case where the true channel is all-pole of order \( N - 1 \) (AP (\( N - 1 \))) and the equalizer FIR of order \( M - 1 \) (\( \text{FIR} (M - 1) \)), \( M > N \). This case does not seem to
be of practical interest, since in practice the transmission channel is a FIR filter and therefore an equalizer of any length will never exceed the order of the channel’s inverse impulse response. However, it merits special attention because it gives insight in the behaviour of the bilinear method. Suppose that the equalizer is FIR(N). Then both equalizer settings \( W^1 = [C^T \theta]^T \) and \( W^2 = [0\; C^T \theta]^T \) are zero forcing. If the corresponding bilinear parameter vectors are denoted by \( \theta^1 \) and \( \theta^2 \), respectively, then they will both satisfy (10). Moreover, any vector of the form \( a\theta^1 + b\theta^2 \), where \( a \) and \( b \) are scalars such that \( a + b = 1 \) will also satisfy (10):

\[
E(XX^H(a\theta^1 + b\theta^2)) = EX, \quad a + b = 1.
\]

In the general case \( M = N + L \), the eq. (10) will be satisfied by any vector of the form:

\[
\theta = \sum_{i=1}^{L} a_i \theta^i, \quad \text{with} \quad \sum_{i=1}^{L} a_i = 1,
\]

where \( \theta^i \) is the bilinear parameter vector corresponding to \( W^i = [0_{1 \times (i-1)}\; C^T \; 0_{1 \times (L+i)}]^T \). This means that when the equalizer is overparameterized w.r.t. the inverse of the channel impulse response, the matrix \( E(XX^H) \) is singular, and as a consequence, the problem (8) has an infinite number of solutions. Running an adaptive algorithm like the one in [16] or in [17] will converge to one of these solutions. The same will happen if one uses a batch technique to estimate \( E(XX^H) \) and \( E(X) \) and a pseudo-inverse for the matrix inversion needed in [11]. In all cases, the solution obtained will no longer correspond to a rank-1 matrix \( \Theta \); (14) and (15) will no longer yield the optimal ZF equalizer. Therefore the problem of ill-convergence appears (under a different form) also in the bilinear method. In the case of a FIR channel (which is a realistic one), one might think that the problem should not arise since the impulse response of the channel is theoretically of infinite length and thus an equalizer long enough should be able to approximate fairly well the ZF equalizer. However, as in practice the channel’s inverse impulse response will be very close to zero out of a specific interval, if the equalizer’s length is bigger than the number of samples in this interval, the same over-parameterization problem will exist and the matrix \( E(XX^H) \) will be very ill-conditioned. This makes the choice of the equalizer length \( N \) in this case a problem of critical importance, as, unlike in conventional equalization, the longest possible length will not necessarily yield the best possible equalizer! In the next section we present a method to calculate the ZF equalizer from any solution of the form (21) when \( L \) is given.

4. A subspace fitting approach for the calculation of a ZF equalizer

Our task is to try to extract a ZF equalizer from the matrix \( \Theta \) that corresponds to a vector \( \theta \) of the form in (21). As will be shown, this will be possible due to the known specific structure of \( \Theta \). The channel inverse is assumed to be of length \( N \) and the equalizer of length \( M = N + L \). Consider the eigenvalue decomposition of the matrix \( \Theta \):

\[
\Theta = \sum_{i=1}^{M} \lambda_i V_i V_i^H, \tag{22}
\]

where the (real) eigenvalues \( \lambda_i \) are in descending order of magnitude and \( V_i \) is the eigenvector corresponding to \( \lambda_i \). When \( M = N(L + 1) \), we saw that the ZF equalizer can be found based on a rank-1 decomposition of \( \Theta \). When \( L > 0 \) we will try to determine the ZF equalizer based on a rank-\( (L+1) \) decomposition of \( \Theta \) by the following subspace fitting approach: we first construct an extended equalizer vector of \( M + L \) entries:

\[
W^e = [w_{-L} \cdots w_{-1} \; w_0 \cdots w_{M-1}]^T, \tag{23}
\]

where \( w_{-L} \cdots w_{-1} \) and \( w_0 \cdots w_{M-1} \) stand for the additional coefficients. If the length of the channel inverse is indeed \( N \), then ideally \( W^e = [0 \cdots 0 \; w_0 \cdots w_{N-1} \; 0 \cdots 0]^T \).

We then create a Toeplitz matrix \( W \) as follows:

\[
W = \begin{bmatrix}
w_0 & w_1 & \cdots & w_{-L} \\
w_1 & w_0 & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
w_{M-1} & w_{M-2} & \cdots & w_{M-1-L}
\end{bmatrix} \tag{24}
\]

We will try to fit the matrix \( W \) to a subspace of the space \( \mathbb{R}^M \) created by the first \( L \) eigenvectors of \( \Theta \). This fitting may be accomplished by minimizing the following criterion:

\[
\min_{Q,W^e} ||W - VQ||_F^2, \tag{25}
\]

where \( ||.||_F \) denotes the Frobenius norm of a matrix \( ||A||_F^2 = tr(A^H A) \), \( Q \in \mathbb{R}^{(L+1) \times (L+1)} \) and \( V \) is a matrix containing the \( L + 1 \) first eigenvectors of \( \Theta \):

\[
V = [V_1 \cdots V_{L+1}]. \tag{26}
\]

Minimization w.r.t. \( Q \) only yields:

\[
Q = (V^H V)^{-1} V^H W. \tag{27}
\]

The problem (25) now becomes:

\[
\min_{W^e} \frac{||P^e W^e||_F^2}{tr(W^e V)} = \min_{W^e} \frac{tr(W^e P^e W)}{tr(W^e V)} \tag{28}
\]

Noting that \( tr(W^H V P^e W) = tr(W^H W) - tr(W^e P^e V) \) the problem (28) can be approximated by the problem:

\[
\max_{W^e} \frac{tr(W^H V P^e V^H W)}{tr(W^e V)} = \max_{W^e} F(W, V) = tr(W^H V P^e V^H W). \tag{29}
\]

The quantity \( F(W, V) \) can be written as:

\[
F(W, V) = W^H \begin{bmatrix}
0 & 0_{L \times (L+1)} & 0_{L \times (L+1)} \\
0_{(L+1) \times L} & 0_{(L+1) \times L} & 0_{(L+1) \times L} \\
0 & 0_{L \times (L+1)} & 0_{L \times (L+1)}
\end{bmatrix} W^e + W^H \begin{bmatrix}
0 & 0_{L \times (L+1)} & 0_{L \times (L+1)} \\
0_{(L+1) \times L} & 0_{(L+1) \times L} & 0_{(L+1) \times L} \\
0 & 0_{L \times (L+1)} & 0_{L \times (L+1)}
\end{bmatrix} W^e + \cdots
\]

\[
= \max_{W^e} W^H \begin{bmatrix}
0 & 0_{L \times (L+1)} & 0_{L \times (L+1)} \\
0_{(L+1) \times L} & 0_{(L+1) \times L} & 0_{(L+1) \times L} \\
0 & 0_{L \times (L+1)} & 0_{L \times (L+1)}
\end{bmatrix} W^e, \tag{30}
\]

which gives the following expression for (29):

\[
\max_{W^e} W^H \begin{bmatrix}
0 & 0_{L \times (L+1)} & 0_{L \times (L+1)} \\
0_{(L+1) \times L} & 0_{(L+1) \times L} & 0_{(L+1) \times L} \\
0 & 0_{L \times (L+1)} & 0_{L \times (L+1)}
\end{bmatrix} \begin{bmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_{L+1}
\end{bmatrix} W^e, \tag{31}
\]
where \( \Omega_i = [\Omega_i]^T \). The solution to (31) is:
\[
W^* = \max \text{eigen} \text{vector} \text{ of } \Lambda = \sum_{i=1}^{L-1} \Omega_i \Omega_i^T . \tag{32}
\]

When \( L \) is known, (32) will give the optimal ZF equalizer (in the absence of additive noise).

5. The influence of additive noise

When additive noise is present in the received signal, it will corrupt both the quantities \( E[X,Y] \) and \( E[X] \) and therefore the solution (11) will be biased and no longer correspond to a ZF equalizer. However, for a given length, one can still use (13) for is sensible to the choice of the equalizer length, as already expressed in Section 4.1. The solution (11) will be biased and no longer correspond to a ZF equalizer. Then a unique solution is found, but is suboptimal. Further investigation of this aspect is the object of ongoing research.

Acknowledgement

After the publication [5] we realized that the principle of the bilinear cost function had already appeared in [7] in order to analyze the behaviour of CMA. We are still however not aware of any paper that discusses the acquisition of the ZF equalizer [15],[32] and the critical role of over-parameterization (21) in the behaviour of the bilinear method.

References


