COMPRESSED TRAINING ADAPTIVE EQUALIZATION

Baki B. Yılmaz and Alper T. Erdoğan

Electrical and Electronics Engineering Department,
Koc University,
Istanbul, Turkey

ABSTRACT

We introduce *compressed training adaptive equalization* as a novel approach for reducing number of training symbols in a communication packet. The proposed semi-blind approach is based on the exploitation of the special magnitude boundedness of communication symbols. The algorithms are derived from a special convex optimization setting based on l_∞ norm. The corresponding framework has a direct link with the compressive sensing literature established by invoking the duality between l_1 and l_∞ norms. Through this link, it is possible to adapt various research results in sparse signal processing literature to adaptive equalization problem. In fact, through utilization of such a link, we show that the amount of training data needed is in the order of the logarithm of the channel spread (or equalizer length) in the fractionally spaced equalization scenario. The numerical experiments provided validates the analytical results and the potentials of the proposed approach.

Index Terms— Adaptive Equalization, Sparseness, Semi-Blind Equalization, Compressive Sensing

1. INTRODUCTION

The dispersive effects of communication channel cause time scrambling of digital communication sequences sent from the transmitter. For this reason, the frequency selective behaviour of the communication channel is compensated by the receiver via employing a filter (called equalizer). The coefficients of this filter need to be obtained adaptively as communication channel is unknown to the receiver.

The most typical approach is to send known training symbols within each communication package, as shown in Figure 1. This approach has the drawback of consuming part of available bandwidth for training signals. In the past, a great deal of effort was spent on developing unsupervised or blind algorithms where the goal was the complete elimination of training symbols. Among them, probably the most popular blind approach is Constant Modulus Algorithm (CMA) [1–3] which makes use of the constant modulus property of some digital communication constellations. However, the need for long data lengths for convergence in connection with non-convex property of CMA cost function led to research for other alternatives. In this direction, Vembu et.al. proposed a convex blind equalization algorithm which exploits the use of infinity norm of equalizer outputs [4]. As an algorithmic extension of this convex blind equalization framework, Linear Programming based [5, 6] and sub-gradient optimization based [7] approaches were proposed.

As the blind algorithms typically require relatively long receiver samples for convergence, a compromise between full training and blind approaches can be achieved by the use of semi-blind algorithms which target to reduce the training size by exploiting some side information. In this area, typical approach is to introduce an objective function which is a convex combination of the square error in the training region and the blind objective such as CMA [8, 9] and constant power [10].

In this article, we develop a new semi-blind equalization framework which exploits the least squares approach for training symbols and infinity norm based cost function for the magnitude boundedness property of the digital communication sources. The resulting optimization problems are the duals of the l_1 norm optimization settings encountered in the data processing applications exploiting sparsity, in particular the compressed sensing. Through exploitation of this link, we show that the training requirement can be reduced to the logarithm of the channel spread/equalization. The organization of the article is as follows: Section 2 introduces the communication scenario assumed throughout the article. Compressed training adaptive equalization approach is proposed in Section 3. In Section 4, a performance analysis of our approach is provided. Section 5 offers a numerical example which illustrates the potential of the proposed approach in alignment with the analysis results.

2. COMMUNICATION SETUP

We consider the following communication scenario:

- Blocks of L_D constellation points $\{s_0, s_1, ..., s_{L_D-1}\}$ are sent by the transmitter as illustrated in Figure 1. For simplifying discussion, we assume BPSK constellation

This work is supported in part by TUBITAK 112E057 project.

is used where the symbols take values from the set ∓ 1. The proposed approach can be extended to PAM or complex QAM constellations.

- Training symbols with length L_T are embedded in each block of data, located in indices $\{p, p+1, ..., p+L_T-1\}$.

Fig. 1. A Generic Transmit Block for a Communication System.

- We assume a two branch receiver (which can correspond to oversampling or multiple antennas) for notational convenience where the results can be easily extended to multiple branches. These branch signals are linearly combined through an adaptive equalizer with L_E taps at each branch.

The block diagram for the communication system described above is given in Figure 2. The signal received at

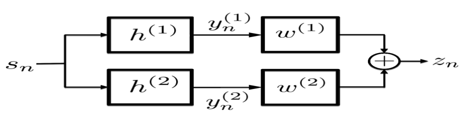

Fig. 2. Equalization Setting with Two Diversity Branches.

each branch is represented by

$$y_n^{(i)} = h_n^{(i)} * s_n \qquad i = 1, 2, \tag{1}$$

where $\{h_n^{(i)} : n \in \{0, \ldots, L_C - 1\}\}$ represents the impulse response of the channel corresponding to the branch i and $\{s_n\}$ is the transmission sequence sent by the transmitter. The equalizer combines these branches through

$$z_n = w_n^{(1)} * y_n^{(1)} + w_n^{(2)} * y_n^{(2)} \tag{2}$$

where $\{w_n^{(i)} : n \in \{0, \ldots, L_E - 1\}\}$ represents the equalizer coefficients for the branch i and $\{z_n\}$ is the equalizer output sequence. It is well known that if the channels of alternative receive branches do not share a common zero, and satisfies the sufficient length requirement $L_E \geq L_C$, the perfect equalization is possible, which will be our standing assumption for this case.

We note that the setup outlined refers to noise free which is the case of interest in the current article, which will be extended in our future article to more general noisy case [11].

3. COMPRESSED TRAINING ADAPTIVE EQUALIZATION

The proposed compressed training adaptive equalization algorithm aims to reduce the training data length by exploiting the bounded magnitude constellation structure of communication sources.

As a recipe to the adaptive equalization problem for obtaining the equalizers coefficients, the following optimization setting is proposed:

$$\textit{Setting I:} \qquad \begin{array}{ll} \text{minimize} & \|\mathbf{z}\|_\infty \\ \text{subject to} & \mathbf{Y}_T \mathbf{w} = \mathbf{s}_T \end{array}$$

where $\mathbf{z} = \begin{bmatrix} z_q & z_{q+1} & \ldots & z_{q+L_D-1} \end{bmatrix}^T$ is the vector containing a collection of L_D equalizer outputs between indexes $\{q\}$ and $\{q + L_D - 1\}$, $\mathbf{Y}_T = \begin{bmatrix} \mathbf{Y}_T^{(1)} & \mathbf{Y}_T^{(2)} \end{bmatrix}$ is the observation matrix, whose components corresponding to different observation branches can be written as the Toeplitz matrix

$$\mathbf{Y}_T^{(i)} = \begin{bmatrix} y_{p+d}^{(i)} & y_{p+d-1}^{(i)} & \cdots & y_{p+d-L_E+1}^{(i)} \\ y_{p+d+1}^{(i)} & y_{p+d}^{(i)} & \cdots & y_{p+d-L_E+2}^{(i)} \\ \ddots & \ddots & \ddots & \ddots \\ y_{p+d+L_T-1}^{(i)} & y_{p+d+L_T-2}^{(i)} & \cdots & y_{p+d+L_T-L_E}^{(i)} \end{bmatrix}$$
$$i = 1, 2,$$

where d is the target equalization delay, and

$$\mathbf{s}_T = \begin{bmatrix} s_p & s_{p+1} & \ldots & s_{p+L_T-1} \end{bmatrix}^T, \tag{3}$$

is the training vector, $\mathbf{w} = \left[\mathbf{w}^{(1)^T} \mathbf{w}^{(2)^T} \right]^T$, is the equalizer coefficient vector whose subcomponents corresponding to different receiver branches can be written as

$$\mathbf{w}^{(i)} = \begin{bmatrix} w_0^{(i)} & w_1^{(i)} & \ldots & w_{L_E-1}^{(i)} \end{bmatrix}^T \qquad i = 1, 2, \tag{4}$$

and $\mathbf{z} = \mathbf{Y}_D \mathbf{w}$, where $\mathbf{Y}_D$ is constructed as $\mathbf{Y}_T$ by considering all receiver outputs rather than the outputs in the training region. Therefore, the proposed optimization setting is about minimizing the peak absolute value for some selected range of equalizer output samples, under the constraint that equalizer outputs are equal to training data in the training region. The main utility of this setting is for the case where the training data is not sufficiently long to determine equalizer coefficients, i.e., $L_T < 2L_E$. In such a case the equation $\mathbf{Y}_T \mathbf{w} = \mathbf{s}_T$ refers to an underdetermined system. Therefore, without additional side information, obtaining a perfect equalizer can not be resolved. In order to clarify why *Setting I* is useful for addressing this *insufficient training data problem*, we defined the combined (communication channel+equalizer) channel from sources to equalizer outputs as

$$\mathbf{g}_n = w_n^{(1)} * h_n^{(1)} + w_n^{(2)} * h_n^{(2)} \tag{5}$$

whose length is $L_G = L_C + L_E - 1$. Based on this definition, we can write $z_n = g_n * s_n$. Therefore, under the assumption that the selected set of output samples are sufficiently long and BPSK constellation at the input, we can write

$$\|\mathbf{z}\|_\infty = \|\mathbf{g}\|_1, \tag{6}$$

where $\mathbf{g} = \begin{bmatrix} g_0 & g_1 & \dots & g_{L_G-1} \end{bmatrix}^T$. Furthermore, in terms of $\mathbf{g}$, the constraint of *Setting I* can be written as $\mathbf{Sg} = \mathbf{s}_T$, where $\mathbf{S}$ is given in (7).

Therefore, in terms of combined impulse response, we can write *Setting I* as follows:

$$\textit{Setting Ig:} \quad \begin{array}{ll} \text{minimize} & \|\mathbf{g}\|_1 \\ \text{subject to} & \mathbf{Sg} = \mathbf{s}_T \end{array}$$

As an important link, we observe *Setting Ig* is equivalent to the optimization setting for Sparse Reconstruction Problem in *Compressed Sensing* literature, where $\mathbf{S}$ would be equivalent to measurement matrix and $\mathbf{s}_T$ vector contains the training symbols. This connection opens up a wealth of possibilities for transferring the results accumulated in sparse information processing literature in the form of algorithms and analysis to adaptive equalization problem.

4. ANALYSIS OF THE PROPOSED APPROACH

The proposed approach offers to reduce the required training length below the number of unknown parameters, $2L_E$. However, the basic question is how much we can reduce the training length, while maintaining perfect equalization solution $\mathbf{g} = \mathbf{e}_{d+1}$ as the unique solution to the *Setting Ig*.

We use the mutual coherence concept to obtain a prescription for the minimum number of training symbols [12]. The mutual coherence of the matrix $\mathbf{\Phi} \in \Re^{L_T \times L_G}$ is defined as

$$\mu(\mathbf{\Phi}) = \max_{\mathbf{1 \le i,j \le M}} \frac{|\mathbf{\Phi}_{:,\mathbf{i}}^{\mathbf{T}} \mathbf{\Phi}_{:,\mathbf{j}}|}{\|\mathbf{\Phi}_{:,\mathbf{i}}\|_{\mathbf{2}} \|\mathbf{\Phi}_{:,\mathbf{j}}\|_{\mathbf{2}}}$$

where the notation $\mathbf{\Phi}_{:,j}$ represents the j^{th} column of $\mathbf{\Phi}$. The following theorem [12] is backbone for the future derivation of lower bound for L_T:

Theorem 4.1. *Let $\mathbf{\Phi} \in \Re^{L_T \times L_G}$ be full rank with $L_T < L_G$. If the system of linear equations $\mathbf{\Phi g} = \mathbf{y}$ has a solution $\mathbf{g}_s$ which obeys*

$$\|\mathbf{g}_s\|_0 < 0.5 \left(1 + \mu(\mathbf{\Phi})^{-1}\right)$$

then it is the unique solution for the optimization problem in Setting Ig.

The ideal combined channel impulse response $\mathbf{e}_{d+1}$ is a solution to $\mathbf{Sg} = \mathbf{s}_T$. According to Theorem 4.1, a sufficient condition for $\mathbf{e}_{d+1}$ whose ℓ_0 norm is one to be the unique solution is given by

$$1 < 0.5 \left(1 + \mu(\mathbf{S})^{-1}\right) \Rightarrow \mu(\mathbf{S}) < 1 \qquad (8)$$

Therefore, the perfect equalization condition is guaranteed if $\mu(\mathbf{S}) < 1$. The following corollary relates the probability of occurrence of this condition as a function of L_T and L_G:

Corollary 4.2. *Let $\mathbf{S} \in \Re^{L_T \times L_G}$ be a Toeplitz matrix with i.i.d. Bernoulli elements. If $L_T > \log_2(L_G(L_G - 1))$ then the mutual coherence condition $\mu(\mathbf{S}) < 1$ is satisfied with probability at least*

$$1 - L_G(L_G - 1) \cdot 2^{-L_T}. \qquad (9)$$

Proof. Assume $\mathbf{G}$ is Gram matrix of $\mathbf{S}$ defined by $\mathbf{G} = \mathbf{S}^T\mathbf{S}$. Every element of $\mathbf{G}$ takes values from $-L_T \le \mathbf{G}_{i,j} \le L_T$.

In order to satisfy the mutual coherence condition $\mu(\mathbf{S}) < 1$, absolute values of all off-diagonal elements of $\mathbf{G}$ must be less than L_T for perfect equalization. Based on this observation, initially, we derive a union bound for $P(\mu(\mathbf{S}) = 1)$.

For deriving the elements of the union bound, we first concentrate on deriving the probability that a super-diagonal element of $\mathbf{G}$ having an absolute value L_T. If we explicitly write the super-diagonal entry $\mathbf{G}_{i,i+1}$ as

$$\mathbf{G}_{(i,i+1)} = s_{p+d-i+1} \times s_{p+d-i} + s_{p+d-i+2} \times s_{p+d-i+1} + s_{p+d-i+L_T} \times s_{p+d-i+L_T-1},$$

the corresponding expression involves $L_T + 1$ different $i.i.d.$ Bernoulli variables, which can create 2^{L_T+1} distinct sequences. Since a super-diagonal entry of $\mathbf{G}$ corresponds to the inner product between two consecutive columns of $\mathbf{S}$, for the simplification of the discussion, let us consider the first two columns of $\mathbf{S}$ as an example:

$$\mathbf{S}_{:,1:2} = \begin{bmatrix} s_{p+d} & s_{p+d-1} \\ s_{p+d+1} & s_{p+d} \\ \vdots & \vdots \\ s_{p+d+L_T-1} & s_{p+d+L_T-2} \end{bmatrix}. \qquad (10)$$

For the inner product between these two columns to be equal to L_T, we need $s_{p+d-1} = s_{p+d} = ... = s_{p+d+L_T-1} = 1$ or $s_{p+d-1} = s_{p+d} = ... = s_{p+d+L_T-1} = -1$, i.e. two sequences where all elements have the same sign. For the inner product of these two columns to be equal to $-L_T$, we need sign alternating sequences, where there are also two alternatives. Therefore, among 2^{L_T+1} different possibilities for $\{s_{p+d-1}, s_{p+d}, \dots s_{p+d+L_T-1}\}$ only 4 of them can cause the condition $|\mathbf{G}_{1,2}| = L_T$. Therefore,

$$P(|\mathbf{G}_{i,i+1}| = L_T) = \frac{4}{2^{L_T+1}} = \frac{1}{2^{L_T-1}} \qquad (11)$$

Following the same approach, if we now consider second upper off-diagonal elements of $\mathbf{G}$: the expression for $\mathbf{G}_{i,i+2}$ contains $L_T + 2$ i.i.d. consecutive elements of $\{s_n\}$, which can create 2^{L_T+2} distinct sequences. We can show that there are only 8 vectors out of these 2^{L_T+2} choices for second upper off-diagonal contains an element with absolute value L_T. Therefore, we can write

$$P(|\mathbf{G}_{i,i+2}| = L_T) = \frac{8}{2^{L_T+2}} = \frac{1}{2^{L_T-1}}, \qquad (12)$$

$$\mathbf{S} = \begin{bmatrix} s_{p+d} & s_{p+d-1} & \cdots & s_p & \cdots & s_{p+d-i+1} & s_{p+d-i} & \cdots & s_{p+d-L_C-L_E+2} \\ s_{p+d+1} & s_{p+d} & \cdots & s_{p+1} & \cdots & s_{p+d-i+2} & s_{p+d-i+1} & \cdots & s_{p+d-L_C-L_E+3} \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ s_{p+d+L_T-1} & s_{p+d+L_T-2} & \cdots & s_{p+L_T-1} & \cdots & s_{p+d-i+L_T} & s_{p+d-i+L_T-1} & \cdots & s_{p+d-L_C-L_E+L_T+1} \end{bmatrix} \quad (7)$$

which is same as the probability for the super-diagonal elements. In fact, the result generalizes to any off-diagonal element. Therefore,

$$P(|\mathbf{G}_{i,j}| = L_T) = \frac{1}{2^{L_T-1}}, i \neq j. \quad (13)$$

Since $\mathbf{G} \in \Re^{L_G \times L_G}$ is a symmetric matrix with $\frac{L_G(L_G-1)}{2}$ upper diagonal elements, we can write

$$P\left(\mu(\mathbf{S}) = 1\right) = P\left(\bigcup_{i=1}^{L_G-1} \bigcup_{j=i+1}^{L_G} \{|\mathbf{G}_{i,j}| = L_T\}\right) \quad (14)$$

$$\leq \sum_{i=1}^{L_G-1} \sum_{j=i+1}^{L_G} P(|\mathbf{G}_{i,j}| = L_T) \quad (15)$$

$$= \frac{L_G(L_G-1)}{2} \frac{1}{2^{L_T-1}} = \frac{L_G(L_G-1)}{2^{L_T}}, \quad (16)$$

where the inequality (15) is obtained by union bound. As a result, for the perfect equalization condition:

$$P(\mu(\mathbf{S}) < 1) = 1 - P(\mu(\mathbf{S}) = 1) \quad (17)$$

$$\geq 1 - \frac{L_G(L_G-1)}{2^{L_T}}. \quad (18)$$

Probability lower bound is informative for the perfect equalization if it is grater than 0 which leads to the condition that $L_T > \log_2(L_G(L_G-1))$. □

Therefore, if we set the training length as

$$L_T = 2\log_2(L_G) - \log_2(\upsilon), \quad (19)$$

then the perfect equalization condition holds with probability more than $1 - \upsilon$.

5. NUMERICAL EXAMPLE AND CONCLUSION

We consider a scenario with random channel of length $L_C = 15$ and $L_E = 20$. In Figure 3(a), exact recovery probability (as a function of L_T) is calculated empirically and compared with our lower bound. Moreover, we define a success probability for $\|\mathbf{g}_* - \mathbf{e}_{d+1}\| \leq 10^{-5}$ (similar to [13]) where $\mathbf{g}_*$ is the solution of the optimization setting and plot the success probability of *Setting Ig*. Although mutual coherence condition does not mimic the behaviour of the success probability exactly, it guarantees exact reconstruction with high probability.

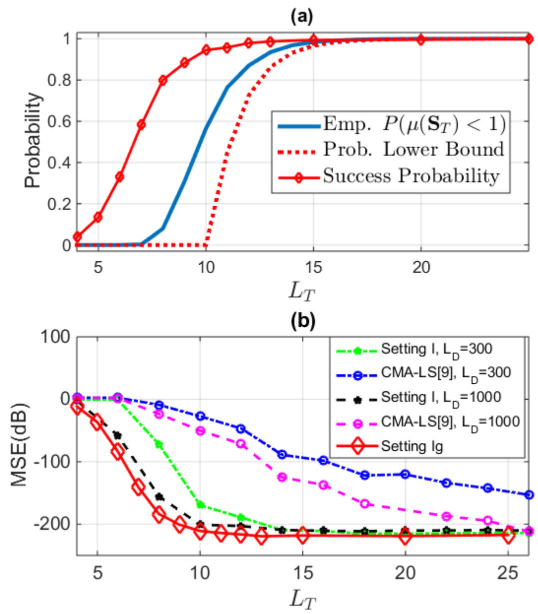

Fig. 3. (a)The mutual coherence probability, (b)Mean Square Error performance of the proposed approach and CMA-LS [9]

In Figure 3(b), we compare our approach with semi-blind algorithm in [9] for changing L_T and L_D in terms of mean square error performance. We observe our algorithm outperforms the other algorithm for every combination of packet and training symbol length. We also observe enlarging the communication packet length results in equivalence of *Setting Ig* and *Setting I* as assumed.

As a conclusion, the proposed approach provides a convex adaptive equalization approach with a clearly prescribed training length, which is in the order of the logarithm of the channel spread (or equalizer length) for noiseless/high-SNR scenarios, as opposed to training length proportional to the equalizer length. The link established with the compressed sensing literature is also very fruitful, which help us inherit the algorithm and analysis related results from the sparse signal processing literature. The same link can also be utilized for the algorithm development and analysis for the noisy scenario as shown in [11].

6. REFERENCES

[1] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. on Communications*, vol. 28, pp. 1867–1875, Nov. 1980.

[2] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[3] C. Johnson et.al., "Blind equalization using the constant modulus criterion: A review," *Proc. of IEEE*, vol. 47, pp. 1927–1950, June 1998.

[4] S. Vembu, S. Verdu, R. Kennedy, and W. Sethares, "Convex cost functions in blind equalization," *IEEE Trans. on Signal Processing*, vol. 42, pp. 1952–1960, August 1994.

[5] Z. Ding and Z. Luo, "A fast linear programming algorithm for blind equalization," *IEEE TCOM.*, vol. 48, pp. 1432–1436, September 2000.

[6] Zhi-Quan Luo, Mei Meng, Kon Max Wong, and Jian-Kang Zhang, "A fractionally spaced blind equalizer based on linear programming," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1650–1660, July 2002.

[7] Alper T. Erdogan and Can Kizilkale, "Fast and low complexity blind equalization via subgradient projections," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2513–2524, July 2005.

[8] A Lee Swindlehurst, "A semi-blind algebraic constant modulus algorithm," *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol. 4, pp. iv–445–iv–448 vol.4, May 2004.

[9] V. Zarzoso and P. Comon, "Semi-blind constant modulus equalization with optimal step size," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, March 2005, vol. 3, pp. iii/577–iii/580 Vol. 3.

[10] V. Zarzoso and P. Comon, "Blind and semi-blind equalization based on the constant power criterion," *Signal Processing, IEEE Transactions on*, vol. 53, no. 11, pp. 4363–4375, Nov 2005.

[11] Baki B. Yılmaz and Alper T. Erdoğan, "Compressed training adaptive equalization: Algorithms and analysis," In Preparation.

[12] Michael Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[13] Dennis Amelunxen, Martin Lotz, Michael B McCoy, and Joel A Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference*, p. iau005, 2014.