AUTOMATIC DELAY SELECTION IN BLIND CHANNEL EQUALIZATION: A PREWHITENING + EIGENVECTOR APPROACH

Roberto López-Valcarce and Fernando Pérez-González

Departamento de Teoría de la Señal y las Comunicaciones
Universidad de Vigo, 36200 Vigo, Spain
valcarce@gts.tsc.uvigo.es, fperez@tsc.uvigo.es

ABSTRACT

We consider the problem of blindly equalizing a single input single output communication channel, assuming that the tap input vector to the equalizer has mutually uncorrelated components. This can be achieved if the received signal is preprocessed by an adaptive all-pole filter (which has been suggested as a means for both MSE improvement and DFE initialization), or by a standard lattice predictor. In the former case, it has been observed recently that an eigenvector of the (prewhitened) input quadricovariance matrix $\mathbf{\Psi}$ may provide a good initial estimate for the equalizer. We provide analytical justification for this fact and show that, if the MSE of a Wiener equalizer for a given delay is small, then this Wiener equalizer is close to an eigenvector of $\mathbf{\Psi}$. The corresponding eigenvalue decreases with the MSE, and thus picking the smallest eigenvalue automatically provides blind delay optimization.

1. INTRODUCTION

To this date, many techniques in blind equalization for communication channels can be found in the literature, among which the Godard [1] or constant modulus (CM) [7] criterion is perhaps the most popular. One drawback of this approach is that, due to its blind nature, preselection of the equalization delay can only be controlled via initialization of the equalizer taps. In practical cases performance may present large fluctuations over the range of delays, and in fact this has been recognized as the reason for the presence of local minima in the associated nonquadratic cost function. Thus, with no a priori knowledge of the channel, convergence of blind algorithms to a setting with delay associated to a high MSE value is possible.

Recently it has been suggested that the components of $\mathbf{x}_n$, the tap input vector to the equalizer, be decorrelated as a previous step [3]. This can be done by using either an all-pole prewhitening filter, or a lattice predictor. In the former case, the resulting CM cost function was analyzed in [5]. There, heuristic evidence showed that certain eigenvector of the quadricovariance matrix of $\mathbf{x}_n$ could provide a good starting point for an iterative CM minimization procedure, since it seemed to lie in a basin of attraction of the global minimum of the (multimodal) CM cost.

Our aim is to provide analytical justification for these observations, by finding the structure of this quadricovariance matrix and exploiting it to obtain its relation to Wiener equalizers. Delay optimization then appears as a by-product.

Supported by the Ramón y Cajal program of the Spanish Ministry of Science and Technology.

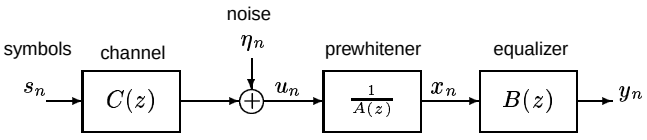

Figure 1: Prewhitened channel model.

2. PREWHITENED CHANNEL MODEL

We consider a single input single output channel-equalizer configuration as shown in Figure 1. The sequence $\{s_n\}$ is assumed to be sub-Gaussian, i.i.d. with zero mean and unit variance, and circular (i.e. $E[s_n^2] = 0$) if complex. The noise $\{\eta_n\}$ is a zero mean white Gaussian process with variance σ_η^2 and independent of $\{s_n\}$. The channel transfer function is FIR with order L: $C(z) = \sum_{k=0}^{L} c_k^* z^{-k}$.

The received signal $u_n = C(z)s_n + \eta_n$ is processed by the all-pole filter $1/A(z)$, whose goal is to whiten its output $\{x_n\}$ [3]. Since $\{u_n\}$ is a moving average (MA) process of order L, perfect whitening of $\{x_n\}$ is achievable as long as the order of $A(z)$ is no less than L. We assume that $1/A(z)$ is properly tuned [4], so that $\{x_n\}$ is a white process.

The equalizer $B(z)$ is an FIR filter of degree N, so that its output y_n can be written as $y_n = \mathbf{b}^H \mathbf{x}_n$, where

$$\mathbf{b} = [\ b_0 \quad b_1 \quad \cdots \quad b_N\ ]^T, \tag{1}$$

$$\mathbf{x}_n = [\ x_n \quad x_{n-1} \quad \cdots \quad x_{n-N}\ ]^T \tag{2}$$

are the equalizer parameter and tap input vectors, respectively. Note that whiteness of $\{x_n\}$ results in

$$E[\mathbf{x}_n \mathbf{x}_n^H] = \sigma_x^2 \mathbf{I}. \tag{3}$$

The Wiener (or MMSE) equalizers are defined as

$$\mathbf{b}_d \triangleq \arg\min_{\mathbf{b}} J(\mathbf{b}, d) = \arg\min_{\mathbf{b}} E[|s_{n-d} - y_n|^2].$$

Their performance strongly depends on the associated delay d. Let us introduce the combined channel-prewhitener transfer function

$$V(z) \triangleq \frac{C(z)}{A(z)} = \sum_{k=0}^{\infty} v_k^* z^{-k}. \tag{4}$$

Then, using (3), one finds that the Wiener equalizers are given by

$$\mathbf{b}_d = \frac{1}{\sigma_x^2}[\ v_d \quad v_{d-1} \quad \cdots \quad v_{d-N}\ ]^H. \tag{5}$$

The resulting MMSE is

$$J_d \triangleq J(\mathbf{b}_d, d) = 1 - \sigma_x^2 \mathbf{b}_d^H \mathbf{b}_d. \tag{6}$$

3. THE EFFECT OF DELAY IN CM RECEIVERS

CM equalizers are obtained by minimization of the CM cost

$$F(\mathbf{b}) = E[(\gamma - |y_n|^2)^2], \quad \text{with} \quad \gamma \triangleq E[|s_n|^4]/E[|s_n|^2]. \tag{7}$$

In contrast with the MMSE J, F is a function of the equalizer output $\{y_n\}$ alone. Therefore the CM criterion is blind in the sense that F can be minimized without the need of training signals. Furthermore, recent results [6] have shown that CM equalizers tend to exist in the neighborhood of Wiener equalizers, provided that the performance of the Wiener solutions is 'sufficiently good'.

The CM cost has long been recognized to be multimodal, and convergence of adaptive algorithms to the global minimum cannot be guaranteed in general. In fact, given the link between CM and Wiener receivers, the local minima of the CM cost are generally seen to be associated to Wiener equalizers corresponding to delays that yield suboptimal performance. Thus, blind delay optimization turns out to be of paramount importance, as the only method for preselection of the equalization delay is through tap initialization.

A study of the CM cost assuming that the received signal has been whitened was presented in [5]. The analysis suggested that an iterative CM minimization procedure could be initialized at the eigenvector of the quadricovariance matrix

$$\mathbf{\Psi} \triangleq E[\mathbf{x}_n \mathbf{x}_n^H \mathbf{x}_n \mathbf{x}_n^H] \tag{8}$$

associated to its smallest eigenvalue. In [5], low-order examples were given in which this eigenvector of $\mathbf{\Psi}$ happened to lie strikingly close to the global minimum of the CM cost function. Note that $\mathbf{\Psi}$ does not depend on the equalizer parameter vector $\mathbf{b}$.

4. ANALYTICAL EXPRESSION OF $\mathbf{\Psi}$

Our purpose is to provide analytical justification for the above observations. Relating the eigenvector of $\mathbf{\Psi}$ to the minima of the CM cost (for which no closed-form solution exists) seems a formidable task. Instead, we explore the relation between $\mathbf{\Psi}$ and the Wiener receivers $\mathbf{b}_d$ in the noiseless case, i.e. it is assumed that $\eta_n = 0$ for all n. In that case, $V(z)$ is an allpass transfer function, satisfying

$$\sum_{i=0}^{\infty} v_i^* v_{i-k} = E[x_n x_{n-k}^*] = \sigma_x^2 \delta(k), \tag{9}$$

since $x_n = \sum_{i=0}^{\infty} v_i^* s_{n-i}$. The (i,j) element of $\mathbf{\Psi}$ (counting from zero) is given by

$$\begin{aligned} \mathbf{\Psi}_{ij} &= E[x_{n-i}(\mathbf{x}^H \mathbf{x}_n) x_{n-j}^*] \\ &= \sum_{l=0}^{N} \underbrace{E[|x_{n-l}|^2 x_{n-i} x_{n-j}^*]}_{\triangleq\, r_{lij}}. \end{aligned} \tag{10}$$

The term r_{lij} is given by

$$r_{lij} = \sum_{k,m,p,q} v_k^* v_m v_p^* v_q E[s_{n-k-i} s_{n-m-l}^* s_{n-p-l} s_{n-q-j}^*]. \tag{11}$$

Using the fact that $\{s_n\}$ is a unit variance i.i.d. process, and defining the normalized kurtosis $\kappa_s \triangleq E[|s_n|^4]/E^2[|s_n|^2]$, the expectation in (11) is found to be

$$\begin{aligned} &\kappa_s, && \text{if } k+i = m+l = p+l = q+j && \text{(real \& complex cases)} \\ &1, && \text{if } k+i = m+l \neq p+l = q+j && \text{(real \& complex cases)} \\ &1, && \text{if } k+i = q+j \neq p+l = m+l && \text{(real \& complex cases)} \\ &1, && \text{if } k+i = p+l \neq m+l = q+j && \text{(real case only)} \end{aligned}$$

and zero otherwise. Using this, and introducing the normalized kurtosis of a Gaussian process

$$\kappa_g = \begin{cases} 3, & \text{real case} \\ 2, & \text{complex case} \end{cases}$$

one finds that r_{lij} satisfies

$$\begin{aligned} r_{lij} &= (\kappa_s - \kappa_g) \sum_{k=0}^{\infty} v_k^* |v_{k+i-l}|^2 v_{k+i-j} \\ &\quad + \sum_{k=0}^{\infty} v_k^* v_{k+i-l} \sum_{p=0}^{\infty} v_p^* v_{p+l-j} \\ &\quad + \sum_{k=0}^{\infty} v_k^* v_{k+i-j} \sum_{p=0}^{\infty} |v_p|^2 \\ &\quad + (\kappa_g - 2) \sum_{k=0}^{\infty} v_k^* v_{k+i-l}^* \sum_{m=0}^{\infty} v_m v_{m+l-j} \end{aligned} \tag{12}$$

and, in view of (9),

$$\begin{aligned} r_{lij} &= (\kappa_s - \kappa_g) \sum_{k=0}^{\infty} v_k^* |v_{k+i-l}|^2 v_{k+i-j} \\ &\quad + \sigma_x^4 \delta(i-l)\delta(j-l) + \sigma_x^4 \delta(i-j) \\ &\quad + \sigma_x^4 (\kappa_g - 2)\delta(i-l)\delta(j-l) \end{aligned} \tag{13}$$

Now we can sum over l in (10) to obtain $\mathbf{\Psi}_{ij}$:

$$\begin{aligned} \mathbf{\Psi}_{ij} &= (\kappa_s - \kappa_g) \sum_{k=0}^{\infty} v_k^* v_{k+i-j} \sum_{l=0}^{N} |v_{k+i-l}|^2 \\ &\quad + \sigma_x^4 \sum_{l=0}^{N} [\delta(i-j) + (\kappa_g - 1)\delta(i-l)\delta(j-l)] \\ &= (\kappa_s - \kappa_g) \sum_{k=0}^{\infty} v_{k-i}^* v_{k-j} \sum_{l=0}^{N} |v_{k-l}|^2 \\ &\quad + \sigma_x^4 (N + \kappa_g)\delta(i-j) \end{aligned} \tag{14}$$

Now note that, in view of (5)-(6), we can write

$$\sum_{l=0}^{N} |v_{d-l}|^2 = \sigma_x^2 (1 - J_d) \tag{15}$$

Hence, using (14) and (15), the matrix $\mathbf{\Psi}$ can be finally written as

$$\begin{aligned}\mathbf{\Psi} &= (\kappa_s - \kappa_g)\sigma_x^2 \mathbf{V}(\mathbf{I}-\mathbf{J})\mathbf{V}^H + \sigma_x^4(N+\kappa_g)\mathbf{I} \\ &= (\kappa_g - \kappa_s)\sigma_x^2 \mathbf{V}\mathbf{J}\mathbf{V}^H + \sigma_x^4(N+\kappa_s)\mathbf{I} \end{aligned} \quad (16)$$

where $\mathbf{J} \triangleq \mathrm{diag}(J_0, J_1, J_2, \cdots)$, and $\mathbf{V}$ is the $(N+1)\times\infty$ convolution matrix associated to $V(z)$, given by

$$\mathbf{V} \triangleq \begin{bmatrix} v_0^* & v_1^* & v_2^* & \cdots & & & \\ & v_0^* & v_1^* & v_2^* & \cdots & & \\ & & \ddots & \ddots & \ddots & \cdots & \\ & & & v_0^* & v_1^* & v_2^* & \cdots \end{bmatrix}. \quad (17)$$

In the last line of (16) we have made use of the fact that, from (9), one has $\mathbf{V}\mathbf{V}^H = \sigma_x^2\mathbf{I}$.

5. RELATION TO WIENER RECEIVERS

Observe that the k-th component of the vector $\mathbf{V}^H\mathbf{b}_d$ is given by

$$\mathbf{e}_k^H\mathbf{V}^H\mathbf{b}_d = \frac{1}{\sigma_x^2}\sum_{l=0}^{N} v_{d-l}^* v_{k-l}, \quad (18)$$

with $\mathbf{e}_k$ denoting the k-th unit vector. Since the rows of $\mathbf{V}$ are orthogonal, for large N, d, the summation in (18) is approximately zero if $k \neq d$. For $k = d$, we use (15) to obtain

$$\sum_{l=0}^{N} v_{d-l}^* v_{k-l} \approx \sigma_x^2(1-J_d)\delta(k-d). \quad (19)$$

Therefore, in that case $\mathbf{V}^H\mathbf{b}_d \approx (1-J_d)\mathbf{e}_d$, and

$$\begin{aligned}\mathbf{V}\mathbf{J}\mathbf{V}^H\mathbf{b}_d &\approx (1-J_d)\mathbf{V}\mathbf{J}\mathbf{e}_d \\ &= (1-J_d)J_d\mathbf{V}\mathbf{e}_d \\ &= (1-J_d)J_d\sigma_x^2\mathbf{b}_d. \end{aligned} \quad (20)$$

Therefore, from (16) one has

$$\mathbf{\Psi}\mathbf{b}_d \approx \alpha_d\mathbf{b}_d, \quad (21)$$

where

$$\alpha_d = \sigma_x^4[(N+\kappa_s) + (\kappa_g - \kappa_s)(1-J_d)J_d]. \quad (22)$$

This shows that the Wiener receiver $\mathbf{b}_d$ is approximately an eigenvector of $\mathbf{\Psi}$ with eigenvalue α_d, provided that $\mathbf{V}^H\mathbf{b}_d \approx (1-J_d)\mathbf{e}_d$ holds. Note that $\mathbf{V}^H\mathbf{b}_d$ corresponds to the combined channel-prewhitener-equalizer impulse response for the delay d Wiener receiver. Hence, this approximation will be valid if this Wiener receiver is able to significantly remove ISI, or in other words, if J_d is sufficiently small. Also note that if J_d decreases, then α_d decreases as well (since $\kappa_g - \kappa_s > 0$ for sub-Gaussian sources), and therefore it makes sense to choose the eigenvector associated to the smallest eigenvalue.

In fact, note from (16) and the facts that $\mathbf{V}\mathbf{V}^H = \sigma_x^2\mathbf{I}$, $0 \leq J_d \leq 1$ for all d, that the eigenvalues of $\mathbf{\Psi}$ satisfy

$$\sigma_x^4(N+\kappa_s) \leq \lambda[\mathbf{\Psi}] \leq \sigma_x^4(N+\kappa_g). \quad (23)$$

But from (22), one has $\alpha_d \to \sigma_x^4(N+\kappa_s)$ as J_d goes to zero. Thus α_d asymptotically approaches the smallest eigenvalue of $\mathbf{\Psi}$.

6. PREORTHOGONALIZED CHANNEL MODEL

Suppose now that, instead of using an all-pole prewhitener, the input autocorrelation matrix is diagonalized by multiplying the received signal vector

$$\mathbf{u}_n = \begin{bmatrix} u_n & u_{n-1} & \cdots & u_{n-N} \end{bmatrix}^T \quad (24)$$

by a matrix $\mathbf{L}$ such that $\mathbf{L}^{-1}\mathbf{L}^{-H} = E[\mathbf{u}_n\mathbf{u}_n^H]$ (i.e. $\mathbf{L}$ is the inverse of the lower triangular Cholesky factor of $E[\mathbf{u}_n\mathbf{u}_n^H]$). Then

$$\mathbf{x}_n = \mathbf{L}\mathbf{u}_n \quad \Rightarrow \quad E[\mathbf{x}_n\mathbf{x}_n^H] = \mathbf{I}.$$

$\mathbf{L}$ can be efficiently implemented as a (normalized) lattice prediction error filter, so that $\mathbf{x}_n$ contains the normalized backward prediction errors of orders 0 through N associated to the process $\{u_n\}$ [2]. We refer to this setting as the preorthogonalized channel model, depicted in Fig. 2. Note that in this case the shift structure present in (2) for the prewhitened case has been lost.

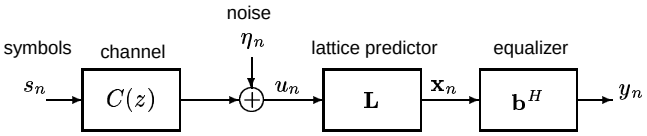

Figure 2: Discrete-time symbol rate channel-equalizer configuration for the preorthogonalized data approach.

Focusing on the noiseless case, we can write

$$\mathbf{x}_n = \mathbf{H}\mathbf{s}_n, \quad (25)$$

where

$$\mathbf{s}_n = \begin{bmatrix} s_n & s_{n-1} & \cdots & s_{n-N-L} \end{bmatrix}^T \quad (26)$$

and $\mathbf{H} = \mathbf{L}\mathbf{C}$, with $\mathbf{C}$ the $(N+1)\times(N+L+1)$ convolution matrix of the channel $C(z)$:

$$\mathbf{C} \triangleq \begin{bmatrix} c_0^* & c_1^* & \cdots & c_L^* & & & \\ & c_0^* & c_1^* & \cdots & c_L^* & & \\ & & \ddots & & & \ddots & \\ & & & c_0^* & c_1^* & \cdots & c_L^* \end{bmatrix}. \quad (27)$$

Note that $\mathbf{H}$ has orthonormal rows:

$$\mathbf{I} = E[\mathbf{x}_n\mathbf{x}_n^H] = \mathbf{H}E[\mathbf{s}_n\mathbf{s}_n^H]\mathbf{H}^H = \mathbf{H}\mathbf{H}^H, \quad (28)$$

In this case, the Wiener receiver for the delay d is just the d-th column of $\mathbf{H}$:

$$\mathbf{b}_d = \mathbf{H}\mathbf{e}_d \quad \Rightarrow \quad J_d = 1 - \mathbf{b}_d^H\mathbf{b}_d, \quad (29)$$

where J_d is the resulting MMSE.

Following a similar approach to that of section 4, one finds that the quadricovariance matrix $\mathbf{\Psi}$ satisfies in this case

$$\mathbf{\Psi} = (\kappa_g - \kappa_s)\mathbf{H}\mathbf{J}\mathbf{H}^H + (N+\kappa_s)\mathbf{I}, \quad (30)$$

where now $\mathbf{J} = \mathrm{diag}(J_0, J_1, \cdots, J_{N+L})$. Again, the vector $\mathbf{H}^H\mathbf{b}_d$ contains the impulse response of the channel-predictor-equalizer overall system. If the MMSE J_d is small, then this response approximates a pure delay, so that $\mathbf{H}^H\mathbf{b}_d \approx (1-J_d)\mathbf{e}_d$, and therefore $\Psi\mathbf{b}_d \approx \alpha_d\mathbf{b}_d$ with

$$\alpha_d = (N+\kappa_s) + (\kappa_g - \kappa_s)(1-J_d)J_d. \quad (31)$$

Observe from (30) that the eigenvalues of $\mathbf{\Psi}$ satisfy now

$$N + \kappa_s \leq \lambda[\mathbf{\Psi}] \leq N + \kappa_g, \tag{32}$$

so that α_d in (31) approaches the smallest eigenvalue of $\mathbf{\Psi}$ as $J_d \to 0$.

7. A NUMERICAL EXAMPLE

To illustrate the results of the previous sections, let us consider a sixth-order real-valued channel $C(z)$ with impulse response

$$[\ 1 \quad -0.22 \quad 0.42 \quad -1.06 \quad -0.08 \quad 0.78 \quad 1.12\ ]$$

whose zero pattern and frequency response are shown in Figs. 3 and 4. Note that this channel is mixed phase and rather severe, since its roots are quite close to the unit circle. Fig. 5 shows the MMSE obtained with the two preprocessors (an all-pole prewhitener and a lattice predictor) assuming a noise-free channel and equalizer of order $N = 19$.

The optimum delays for the prewhitened and preorthogonalized approaches are respectively $d = 19$ and $d = 15$; the respective delay optimized MMSE values are -19 dB and -16 dB. This difference in performance is due to the fact that the prewhitened configuration makes more efficient use of the available degrees of freedom: the receiver comprises an IIR transfer function with L poles and N zeros. On the other hand, the receiver in the preorthogonalized setting reduces to an FIR filter of order N: the effect of the transfer matrix $\mathbf{L}$ of the lattice predictor is merely to provide a change of basis for the representation of the equalizer parameters. Thus, for the same value of N, the prewhitened configuration will always yield lower MMSE values. Also note that if, as it is usually required, the equalizer order is much larger than the channel memory (i.e. $N \gg L$), the all-pole prewhitener of Fig. 1 will be considerably less costly to implement than the lattice predictor in Fig. 2.

Fig. 6 shows the taps of the Wiener receivers ($\mathbf{b}_{19}$ for the prewhitened case and $\mathbf{b}_{15}$ for the preorthogonalized case) normalized to unit norm, and the taps of the eigenvectors of $\mathbf{\Psi}$ corresponding to its smallest eigenvalue in each case. The closeness of these eigenvectors to the delay optimized Wiener receivers indicates that they could be used as starting points for further processing by an adaptive blind equalization algorithm.

8. CONCLUSION

We have explored the structure of the quadricovariance matrix of the input vector to the equalizer, assuming that the components of the input vector have been decorrelated. This can be accomplished by an all-pole prefilter or a lattice predictor. In both cases it is seen that Wiener receivers (corresponding to different delays) with small associated MMSE are approximately eigenvectors of this matrix. The corresponding eigenvalue decreases with the MMSE, so that picking the eigenvector associated to the smallest eigenvalue automatically provides blind delay optimization.

In practice, the quadricovariance matrix must be estimated from finite length data records. Future work should address the robustness of the approach to this fact as well as to noise. Numerical stability issues should also be investigated.

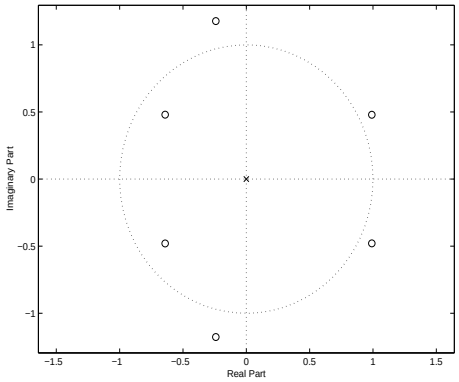

Figure 3: Zero plot of $C(z)$.

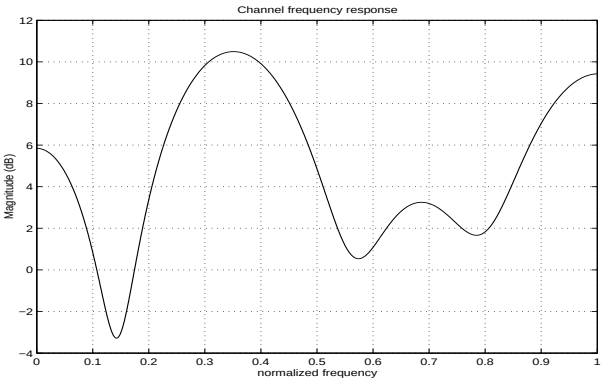

Figure 4: Frequency response of $C(z)$.

9. REFERENCES

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

[2] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 1996.

[3] J. Labat, O. Macchi and C. Laot, 'Adaptive decision feedback equalization: can you skip the training period?', *IEEE Trans. Communications*, vol. 46, pp. 921-930, Jul. 1998.

[4] R. López-Valcarce and F. Pérez-González, 'On blind adaptive algorithms for IIR equalizers', *Proc. 9th DSP Workshop*, Oct. 2000. Online proceedings in http://spib.ece.rice.edu/SPTM/DSP2000/program.html .

[5] R. López-Valcarce and F. Pérez-González, 'Analysis of the prewhitened constant modulus cost function', *Proc. IEEE ICASSP 2002*, vol. 3, pp 2641-2644.

[6] P. Schniter and C. R. Johnson, Jr., 'Bounds for the MSE performance of constant modulus estimators', *IEEE Trans. Information Theory*, vol. 46, pp. 2544-2560, Nov. 2000.

[7] J. R. Treichler and B. G. Agee, 'A new approach to multipath correction of constant modulus signals', *IEEE Trans. ASSP*, vol. 31, pp. 459-472, Apr. 1983.

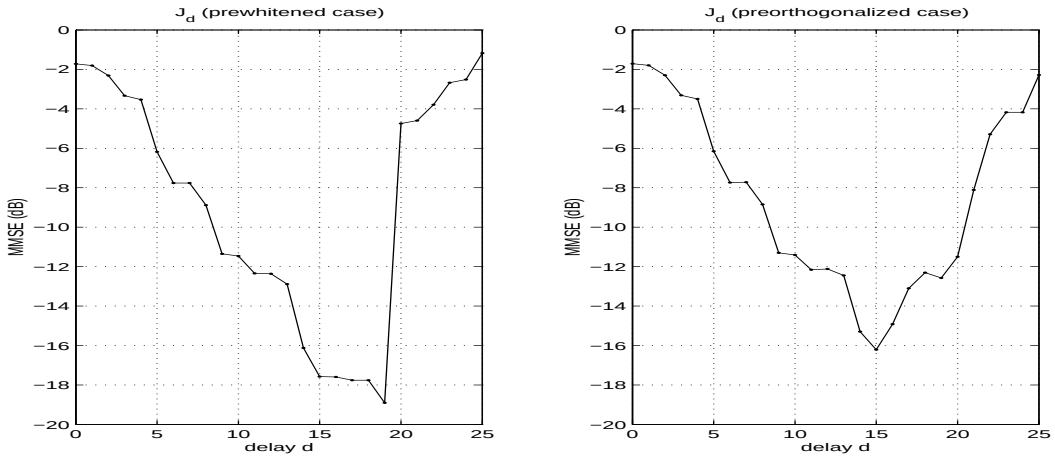

Figure 5: MMSE obtained with the two approaches and $N = 19$.

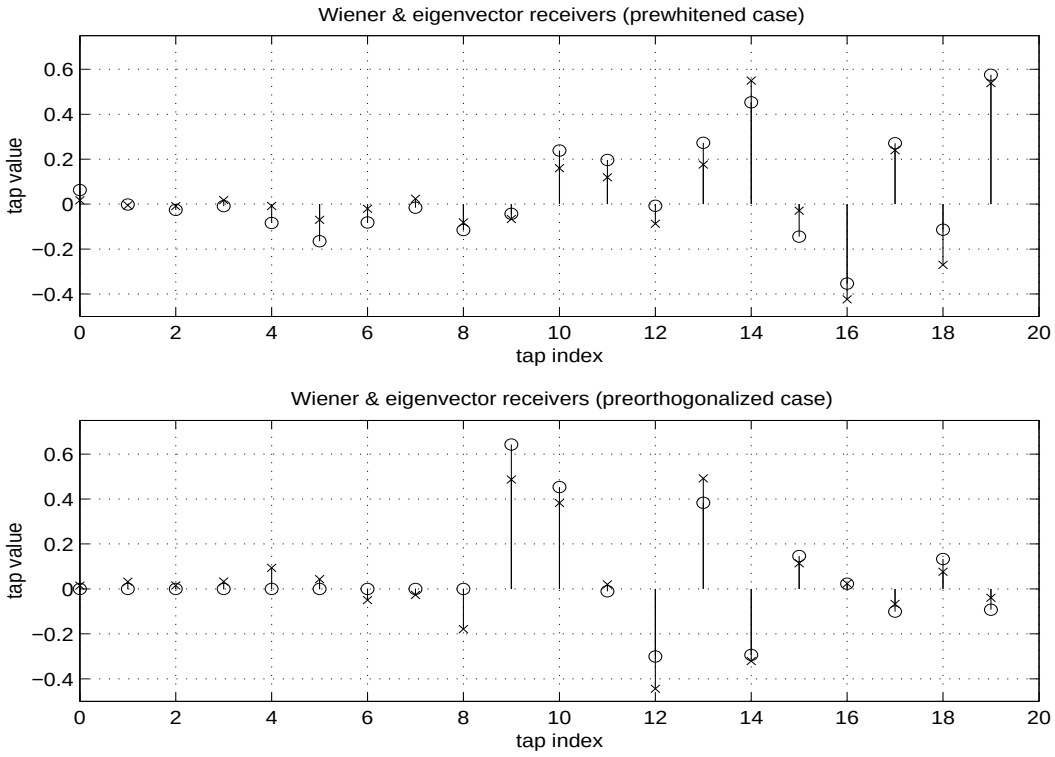

Figure 6: Normalized Wiener receivers ($\circ$) and eigenvector-based equalizers ($\times$). ($N = 19$)