MMSE DECISION FEEDBACK EQUALIZER FROM CHANNEL ESTIMATE

M. Magarini, A. Spalvieri,

Dipartimento di Elettronica e Informazione, Politecnico di Milano,
Piazza Leonardo da Vinci, 32, I-20133 Milano (Italy),
{magarini,spalvier}@elet.polimi.it

Abstract - In digital radio transmission over frequency selective channels, the Minimum Mean-Square Error Decision Feedback Equalizer (MMSE-DFE) is widely recognized as an efficient equalization scheme. In order to compute the coefficients of the feedforward and feedback filters of the MMSE-DFE, both the channel impulse response (CIR) and the variance of the noise have to be estimated. The estimate of the CIR is usually performed by the standard least-square method, where a known training sequence is employed. Then, the estimated CIR is used to evaluate the variance of the noise. In this paper the use of the two above estimates for MMSE-DFE is studied. In particular, an unbiased estimate of the variance of the noise is described.

Keywords - Decision-feedback equalizer, time-varying channel, channel estimation.

I. INTRODUCTION

In digital radio transmission over frequency selective channels equalization of the received signal is required. The Decision Feedback Equalizer (DFE) is known to be an efficient equalization scheme for channels corrupted by severe InterSymbol Interference (ISI) and additive Gaussian noise. In the Minimum Mean-Square Error DFE (MMSE-DFE) the coefficients of the FeedForward Filter (FFF) and of the FeedBack Filter (FBF) are obtained from a spectral factorization where both the Channel Impulse Response (CIR) and the variance of the noise are assumed to be known.

When the MMSE-DFE is employed on a time varying channel, adaptation of the equalizer to the channel has to be performed. In conventional implementations, gradient or least-square methods are used to adapt the equalizer coefficients, using either the detected data or the symbols of the training sequence [1]. In [2] it is shown that in some circumstances, such as moderately rapid fading channels, direct calculation of the equalizer coefficients from the estimated CIR can be more appropriate than adaptive algorithms.

In this paper joint CIR and noise variance estimate is performed at the receiver by using a training sequence. The CIR is obtained by the classical Least Square (LS) method [1]. The problem we are mainly concerned is the estimation of the noise variance from the given CIR estimate. In fact, by assuming perfect channel knowledge, in [3] it is shown that the performance of the MMSE-DFE is sensitive to the quality of the estimate of the
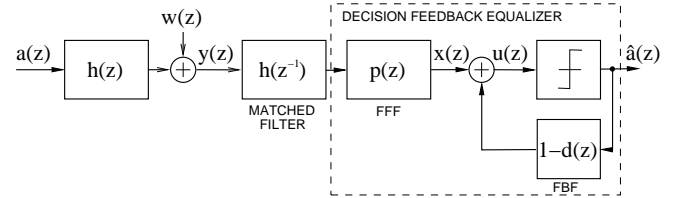

Fig. 1: Discrete-time model of the transmission system with DFE in the receiver.

noise variance.

The estimation of the noise variance could be treated as a problem of single parameter estimation if the CIR were known. In this case, the estimate can be obtained by averaging the square of the difference between the samples of the received signal and the training sequence filtered by the known CIR. The number of elements which are averaged coincides with the length of the observation [1]. The situation is somewhat different when the CIR is not known. We observe that, in order to obtain an unbiased estimate, the scalar that divides the sum of the squared differences has to be chosen equal to the length of the observation diminished by the length of the channel.

The paper is organized as follows. The model of the transmission system is described in section II. Section III presents the equations of the coefficients of the MMSE-DFE. In section IV the estimation method is described. The quality of the estimate is demonstrated in section V by computer simulations. Finally, conclusions are drawn.

II. SYSTEM MODEL

Figure 1 reports the discrete-time model of a data transmission system over a baseband ISI channel with an DFE at the receive side. The transmission of independent, identically distributed binary data $a_k \in \{-1, +1\}$ is assumed. Let $\nu + 1$ be the duration of the CIR and let $h(z) = \sum_{i=0}^{\nu} h_i z^{-i}$ be its z-transform (z^{-1} represents the unit delay). The autocorrelation of the CIR is $r(z) = h(z)h(z^{-1})$. The discrete-time noise w_k is modelled as zero mean white Gaussian noise with variance σ^2. The z-transform of the observed sequence is

$$y(z) = a(z)h(z) + w(z).$$

In the MMSE-DFE the computation of the taps of the FBF is

based on what is called in [4] the *key equation*:

$$S_0 d(z) d(z^{-1}) = r(z) + \sigma^2, \qquad (1)$$

where $d(z)$ is chosen as the impulse response that is causal, monic, and minimum phase. The scalar S_0 is called the system average energy and is given by

$$S_0 = \sigma^2 \exp\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \log\left(1 + \frac{r(e^{j\omega})}{\sigma^2} \right) d\omega \}.$$

Note that for $\sigma > 0$ the power density spectrum $r(e^{j\omega}) + \sigma^2$ is nonnull everywhere. As a consequence, the existence of $d(z)$ is guaranteed and all its roots are strictly inside the unit circle. The MSE is given by

$$\mathrm{MSE} = E\{u_k^2\}, \qquad (2)$$

where $E\{\cdot\}$ denotes the expected value, and

$$u_k = x_k - \sum_{i=0}^{\nu} d_i a_{k-i}, \qquad (3)$$

is the distortion sequence at the input of the slicer where

$$x(z) = p(z) h(z^{-1}) y(z)$$

is the output of the FFF. In (3) it has been assumed that a_k is the k-th element of the transmitted sequence. Minimization of (2) yields

$$p(z) = \frac{d(z)}{r(z) + \sigma^2} = d^{-1}(z^{-1}). \qquad (4)$$

The MMSE is

$$\mathrm{MSE}_{\mathrm{MMSE\text{-}DFE}} = \frac{\sigma^2}{S_0}.$$

III. ESTIMATE OF THE POWER DENSITY SPECTRUM

A pseudorandom sequence of length N is employed to construct the training sequence. In order to avoid the presence of unknown data symbols that influence the samples of the received sequence, the last $\nu+1$ symbols of the pseudorandom sequence are appended at the beginning of the training sequence, and only N observations are used to work out the estimate.

In the context of estimation problems it is worth adopting vector and matrix notation. By writing the N samples of the observed sequence in the column vector $\mathbf{y} = [y_1, y_2, \ldots, y_N]'$ ($'$ denote transposition), one has

$$\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{w}, \qquad (5)$$

where $\mathbf{A}$ is the $N \times (\nu + 1)$ convolution matrix of the training sequence, $\mathbf{h} = [h_0, h_1, \ldots, h_\nu]'$ is the column vector of the channel coefficients and $\mathbf{w} = [w_1, w_2, \ldots, w_N]'$ is the column vector of the noise samples.

The estimate of the channel coefficients is obtained by applying the LS method [1]

$$\begin{aligned} \hat{\mathbf{h}} &= (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} \\ &= \mathbf{h} + (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{w}, \end{aligned} \qquad (6)$$

where equation (5) has been substituted for $\mathbf{y}$ in the last line of (6). Given $\hat{\mathbf{h}}$ from (6), one can get the estimate of the noise variance as

$$\hat{\sigma}^2 = \frac{1}{\beta} \left\| \mathbf{y} - \mathbf{A}\hat{\mathbf{h}} \right\|^2, \qquad (7)$$

where $\| \cdot \|^2$ represents the squared Euclidean norm in the N dimensional space and β is a scalar factor that is hereafter discussed. If the channel were known the estimate of the noise variance would become a classical problem of a single parameter estimation, hence $\beta = N$. It is proved in [5] that using $\beta = N$ in (15) one gets the ML estimate, which turns out to be biased downward. We prove in the appendix that, in order to obtain an unbiased estimate, one has to set $\beta = N - \nu - 1$. This result can be explained by observing that we are estimating certain parameters of the Gaussian random vector defined in the RHS of (5). Specifically, we estimate the $\nu + 1$ parameters $\{h_0, h_1, \ldots, h_\nu\}$ that determine the mean vector $\mathbf{A}\mathbf{h}$, and the parameter σ^2 that determines the covariance matrix $\sigma^2 \mathbf{I}_N$, where $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. From this perspective we see that the coefficient $(N - \nu - 1)^{-1}$ resembles the coefficient $(N - 1)^{-1}$ which appears in the classical estimate of the variance of a scalar random variable based on the sample mean:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} v_i, \quad \hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (v_i - \hat{\mu})^2.$$

Once the estimate of the channel is available, we compute the estimate of the autocorrelation of the CIR as

$$\hat{r}(z) = \hat{h}(z)\hat{h}(z^{-1}). \qquad (8)$$

Note that $\hat{r}(z)$ is a biased estimate of $r(z)$. We observe that

$$\mathrm{cov}(\hat{\mathbf{h}}) = E\{(\hat{\mathbf{h}} - \mathbf{h})'(\hat{\mathbf{h}} - \mathbf{h})\} = E\{\boldsymbol{\eta}'\boldsymbol{\eta}\} = \sigma^2 (\mathbf{A}'\mathbf{A})^{-1}, \qquad (9)$$

which, for a pseudorandom training sequence, gives the result [6]

$$E\{\boldsymbol{\eta}'\boldsymbol{\eta}\} = \frac{\sigma^2}{N+1} \left(\mathbf{I}_{\nu+1} + \frac{\mathbf{U}_{\nu+1}}{N - \nu + 1} \right), \qquad (10)$$

where $\mathbf{U}_{\nu+1}$ is a $(\nu + 1) \times (\nu + 1)$ all ones matrix. The k-th coefficient of $\hat{r}(z)$ is obtained as

$$\hat{r}_k = \sum_{i=0}^{\nu} (h_i + \eta_i)(h_{i+k} + \eta_{i+k}),$$

hence, using (10) and $E\{\boldsymbol{\eta}\} = \mathbf{0}$, one concludes that
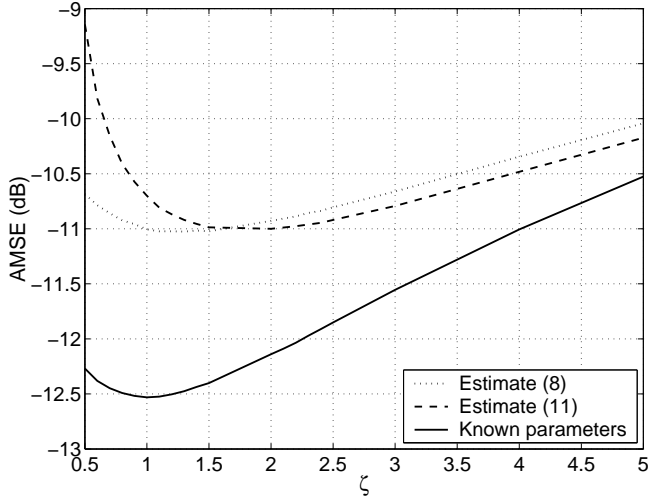
$$E\{\hat{r}_k\} = r_k + b_k,$$

Fig. 2: AMSE versus ζ for $N = 15$ at SNR= 25 dB.

Fig. 3: AMSE versus ζ for $N = 31$ at SNR= 25 dB.

where

$$b_k = \sigma^2(\nu + 1 - k)\left(\frac{1}{(N - \nu + 1)(N + 1)} + \frac{\delta_k}{N + 1}\right),$$

where δ_k is the Kronecker delta function. Therefore, an unbiased estimate of $r(z)$ is

$$\hat{r}(z) = \hat{h}(z)\hat{h}(z^{-1}) - b(z). \tag{11}$$

IV. EXPERIMENTAL RESULTS

In the experiments we compute the FBF by the spectral factorization $S_0 d(z) d(z^{-1}) = \hat{r}(z) + \zeta \hat{\sigma}^2$ and compare the results obtained by using $\hat{r}(z)$ given in (8) and (11). The factor ζ is adopted in view of the results obtained in [3, 7]. The estimate $\hat{\sigma}^2$ is computed from (7) with $\beta = N - \nu - 1$. As far as the computation of the coefficients of the FFF is concerned, note that formula (4) holds when the FFF has an infinite number of coefficients. When the FFF has a finite number of coefficients, equation (4) is no longer valid and minimization of (2) is performed as reported in [8]. In the results that follows, an FIR FFF with 91 taps has been considered. The estimated CIR given in (6) is used as the discrete-time matched filter.

We adopt as a benchmark the time discrete additive white Gaussian noise channel with $\nu = 6$ studied in [9]. The impulse response of the channel is $h(z) = 0.176 + 0.316z^{-1} + 0.476z^{-2} + 0.532z^{-3} + 0.476z^{-4} + 0.316z^{-5} + 0.176z^{-6}$. Note that the channel is characterized by three zeros on the unit circle. We consider transmission of data packets whose length is 200 bits. By performing Monte Carlo simulations we measured the MSE at SNR=25 dB (SNR=r_0/σ^2). The Average MSE (AMSE) is obtained by averaging the measured MSE over 10^5 data packets.

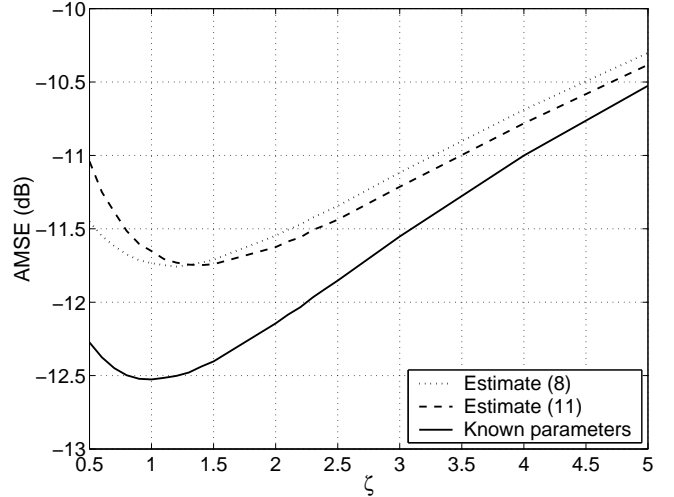Figures 2 and 3 report the AMSE versus ζ at SNR= 25 dB respectively for $N = 15$ and $N = 31$. From the figures one observes that for both the estimates of $r(z)$ the minimum of the AMSE is not achieved in correspondence of $\zeta = 1$, but for $\zeta > 1$. For this reason, once the unbiased estimate of the noise variance is available, it could be appropriate to slightly fictitiously increase its value in order to improve the performance. We observe that the AMSE of the two estimates is mainly influenced by the bias introduced in the estimated signal power (the zero lag coefficient $\hat{r}_0 + \zeta \hat{\sigma}^2$), which can be controlled by the parameter ζ. From these results it also appears that cancelling the bias from the tails of the autocorrelation does not improve the performance. Therefore, in the following results the estimate (8), which does not require any corrective term, is adopted. Figure 4 shows the first error event rate (FEER), which measures the probability of error given correct detection of the previous bits. Figure 4 gives the bit error rate (BER), where the effect of error propagation is included. From the figures we observe that the minimum of the error probability occurs for a value of ζ which is greater than the value of ζ for which the minimum AMSE occurs. From the presented results, it appears that it is worth biasing upward the estimate of the signal power.

V. CONCLUSIONS

Direct computation of the coefficients of the MMSE-DFE requires the knowledge of the CIR and of the noise variance. When these parameters are not known a priori, they have to be estimated. While the problem of the CIR estimation is hugely treated in the literature [1, 8], it does not seem to be the same for the noise variance. The main contribution of this paper is the derivation of the sample estimate of the noise variance, from the received sequence and the estimated CIR. The results show that the AMSE is slightly improved when the estimated noise variance is slightly biased upward. As far as the error probability is concerned, our results confirm what was already observed in [3, 7] in the context of known parameters. Specifically, the er-
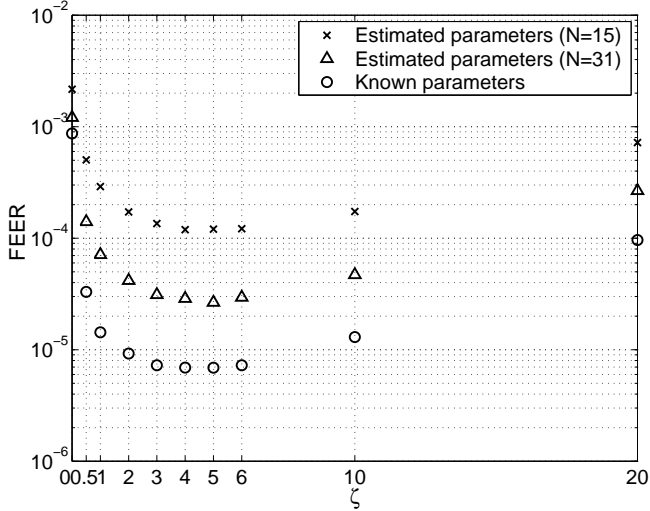
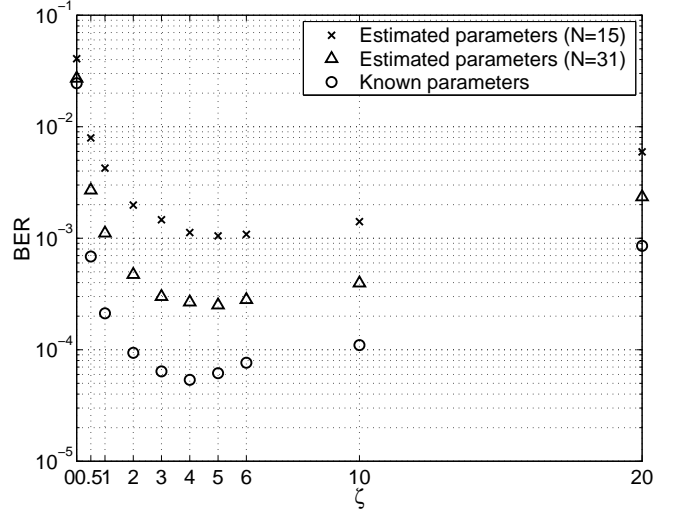Fig. 4: Simulated FEER versus ζ at SNR= 25dB.

Fig. 5: Simulated BER versus ζ at SNR= 25dB.

ror probability is substantially improved by adding a substantial upward bias to the estimated signal power.

APPENDIX

By substituting equation (6) in (7) one has

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{\beta} \left\| \mathbf{w} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{w} \right\|^2 \\
&= \frac{1}{\beta} \left\| \left(\mathbf{I}_N - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \right) \mathbf{w} \right\|^2 .
\end{aligned} \tag{12}
$$

The key of the proof is to observe that

$$
\mathbf{P} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'
$$

is a projection matrix [10]. $\mathbf{P}$ satisfies the two properties $\mathbf{P}^2 = \mathbf{P}$, $\mathbf{P}' = \mathbf{P}$, and its rank coincides with that of $\mathbf{A}$, that is $\nu + 1$. In fact, the space spanned by $\mathbf{P}$ is defined by the $\nu + 1$ columns of the matrix $\mathbf{A}$. It is easy to show that $\mathbf{P}_\perp = \mathbf{I}_N - \mathbf{P}$ is the projection matrix in the space orthogonal to that spanned by $\mathbf{P}$. The rank of $\mathbf{P}_\perp$ is $N - \nu - 1$. Hence, equation (12) can be rewritten as

$$
\hat{\sigma}^2 = \frac{1}{\beta} \mathbf{w}' \mathbf{P}_\perp \mathbf{w}. \tag{13}
$$

From this equation one observes that $\hat{\sigma}^2$ is completely estimated in the $(N - \nu - 1)$-th dimensional space which is orthogonal to that defined by the columns of the matrix $\mathbf{A}$. A property of a projection matrix is that its eigenvalues are 1 or 0. Since the number of eigenvalues different from 0 defines the rank of a matrix, it follows that $\mathbf{P}_\perp$ has $N - \nu - 1$ eigenvalues equal to 1. Hence, $\mathbf{P}_\perp$ can be expressed as

$$
\mathbf{P}_\perp = \mathbf{Q}'\mathbf{\Lambda}\mathbf{Q}, \tag{14}
$$

where $\mathbf{Q}$ is the $N \times N$ eigenvectors matrix, and

$$
\mathbf{\Lambda} = \left[\begin{array}{cc} \mathbf{I}_{N-\nu-1} & \mathbf{0}_{N-\nu-1,\nu+1} \\ \mathbf{0}_{\nu+1,N-\nu-1} & \mathbf{0}_{\nu+1,\nu+1} \end{array} \right],
$$

where $\mathbf{0}_{l,m}$ denotes a $l \times m$ matrix with zero entries. By substituting equation (14) in equation (13) we obtain

$$
\hat{\sigma}^2 = \frac{1}{\beta} \mathbf{n}' \mathbf{\Lambda} \mathbf{n}, \tag{15}
$$

where $\mathbf{n} = \mathbf{Q}\mathbf{w}$. Since the columns of $\mathbf{Q}$ define an orthonormal basis, $\mathbf{n}$ is statistically equivalent to $\mathbf{w}$. The mean value of the estimate of the noise variance is obtained by taking the expectation of (15) as

$$
\begin{aligned}
E\{\hat{\sigma}^2\} &= \frac{1}{\beta} E\{\mathbf{n}'\mathbf{\Lambda}\mathbf{n}\} \\
&= \frac{1}{\beta} \sum_{i=1}^{N-\nu-1} E\{n_i^2\} = \frac{N - \nu - 1}{\beta} \sigma^2.
\end{aligned}
$$

From this equation it follows that one has to set $\beta = N - \nu - 1$ in order to obtain an unbiased estimate. Note that with $\beta = N$ one has the ML estimate of the noise variance [5], and that the condition $\mathbf{A}'\mathbf{A} = N\mathbf{I}_{\nu+1}$ required in [5] is not necessary both for our unbiased estimate and for the ML estimate.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, New Jersey, 1996.

[2] S. Fechtel and H. Meyer, "An investigation of channel estimation and equalization techniques for moderately rapid HF channels," in *Proceedings of International Conference on Communications*, Denver, CO, vol. 2, pp. 768-772, June 1991.

[3] M. Magarini, A. Spalvieri and G. Tartara, "Sensitivity of the mean-square DDFSD to a noisy estimate of the noise

variance," in *Proceedings of Vehicular Technology Conference (VTC '01-Fall)*, Atlantic City, NJ, vol. 2, pp. 892-896, October 2001.

[4] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney Jr., "MMSE decision-feedback equalizers and coding-part I: equalization results," *IEEE Trans. Commun.*, vol. 43, pp. 2582-2594, October 1995.

[5] C. H. Aldana and J. M. Cioffi, "Channel tracking for multiple input, single output systems using EM algorithm," in *Proceedings of International Conference on Communications*, Helsinki, Finland, vol. 2, pp. 586-590, June 2001.

[6] S. N. Crozier, D. D. Falconer, S. A. Mahmoud, "Least sum of squared errors (LSSE) channel estimation," *IEE Proc.*, vol. 38, pp. 371-378, August 1991.

[7] P. S. Bednarz and J. M. Cioffi, "Decision feedback equalization for channels with error correcting capabilities," in *Proc. IEEE International Conference on Communications*, Montreal, Canada, 1997, pp. 1607 -1612.

[8] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 1995.

[9] R. R. Anderson and G. J. Foschini, "The minimum distance for MLSE digital data systems of limited complexity," *IEEE Trans. Inform. Theory*, vol. 21, pp. 544-551, September 1975.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, John Hopkins University Press, Baltimore, 1996.