Complex discriminative learning Bayesian neural equalizer

Mirko Solazzi[a], Aurelio Uncini[b, *], Elio D. Di Claudio[b], Raffaele Parisi[b]

[a]*Dipartimento di Elettronica e Automatica, University of Ancona, Via Brecce Bianche, 60131 Ancona, Italy*
[b]*Dipartimento INFOCOM, University of Rome "La Sapienza", Via Eudossiana 18, 00184 Rome, Italy*

Abstract

Traditional approaches to channel equalization are based on the inversion of the global (linear or nonlinear) channel response. However, in digital links the complete channel inversion is neither required nor desirable. Since transmitted symbols belong to a discrete alphabet, symbol demodulation can be effectively recast as a classification problem in the space of received symbols. In this paper a novel neural network for digital equalization is introduced and described. The proposed approach is based on a decision-feedback architecture trained with a *complex-valued* discriminative learning strategy for the minimization of the classification error. Main features of the resulting neural equalizer are the high rate of convergence with respect to classical neural equalizers and the low degree of complexity. Its effectiveness has been demonstrated through computer simulations for several typical digital transmission channels. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Channel equalization; Decision-feedback equalizer; Neural networks; Complex neural networks; Discriminative learning; Bayes decision rule

1. Introduction

Adaptive channel equalization is required to combat intersymbol interference (ISI), nonlinear channel distortion and interference in digital communications [13]. Traditional equalizers attempt to invert the channel response to recover the original signal sequence before the final decision [13]. In alternative, in the last years neural networks have been successfully applied to the equalization task [1,2,11,12,14]. Acting as nonlinear maps between received samples and training symbols [5], in fact, neural nets are able to enhance the received signal before demodulation [6].

The classical approach to equalization is not completely satisfactory in the presence of multipath and/or nonlinearities. As a matter of fact, in these cases the channel inversion is an ill-posed problem, due to a loss of information in the transmission path [1,2,11,12,15]. Bayesian (BA) and maximum likelihood (ML) equalizers are commonly adopted to face this problem [11] and are based on the knowledge of the multidimensional mapping performed by the channel from transmitted symbol sequences onto symbol clusters, defined in a proper output space [2,11]. With ML and BA equalizers, channel inversion is neither required nor desirable. In fact, they can be viewed as classifiers, whose objective is to determine the

---
* Corresponding author. Tel.: +39-06-44585839; fax: +39-06-4873300.

E-mail address: aurel@infocom.uniroma.it (A. Uncini).

Nomenclature

$S[n]$	transmitted symbol at time n	$\tilde{S}[n]$	estimated symbol at time n
h_k	kth sample of channel impulse response	m	length of equalizer feedforward section
		l	length of equalizer feedback section
$H(z)$	channel transfer function	d	decision delay
N	length of channel impulse response	$y[n]$	neural network output at time n
$r[n]$	channel linear output at time n	$\mathbf{w}$	neural net weight vector
$\hat{r}[n]$	channel nonlinear output at time n	$D\{\cdot\}$	decision criterion
$x[n]$	received sample at time n	$D_i(\cdot)$	ith discriminant function
$q[n]$	additive noise sample at time n	$R_i(\cdot)$	ith mis-classification measure

optimal decision boundaries among projected symbol sequences. However, these architectures are very expensive in terms of hardware requirements when compared to traditional equalizers, so the research for better cost/performance trade-offs is fully justified.

Also *radial basis function* (RBF) networks have been exploited in equalizer design, due to their link with Bayesian classifiers in Gaussian environments [2,8]. However, RBF networks are hampered by low speed and limited robustness of the clustering phase [12], which is a critical step of the adaptation process. In fact, if estimated centroids are far from the true positions, local convergence of the algorithm can dramatically slow down or fail.

Alternatively, multilayer feedforward [1,5,9,14] or recurrent [10,12] neural networks can be used as classifiers, using the mean square error (MSE) criterion. However, MSE minimization does not imply a reduced bit error rate (BER) [9] and alternative error functionals may be devised. In particular, the discriminative learning (DL) technique, which optimizes the decision boundaries with a non-parametric approach, has been proved to be highly effective [9].

In this paper, a new neural architecture for digital equalization is introduced and described. The proposed solution is intermediate between the channel inversion and the output symbol classification approaches. A feedforward neural network with decision feedback (DF) [1,11] processes the received sequence. Final decision about received symbols

is made by a Bayesian-like criterion *in the output space*. In particular, *fixed* Gaussian kernels are used to model the output, so that multidimensional clustering is not required. This is in contrast with the classical RBF approach, which partitions the *input space*. The training algorithm is based on a *complex-valued* DL strategy which optimizes decision boundaries among output clusters.

2. Proposed architecture

In digital communication systems the transmitter sends out a sequence of symbols $S[n]$, which are extracted from a given alphabet and assumed independent and equiprobable. Fig. 1 shows the general model of a digital communication system, which introduces both linear and nonlinear distortions. Linear distortions are due to the limited bandwidth of the channel and generate ISI. Nonlinear distortions are produced by nonlinear devices in the transmission path (for example, converters or amplifiers working in saturation).

The linear part of the channel is commonly modeled by a finite impulse response filter [13]:

$$\hat{r}[n] = \sum_{k=0}^{N-1} h_k S[n-k], \tag{1}$$

where N is the length of the channel impulse response. In the presence of nonlinear distortion, assuming that most channel nonlinearities are
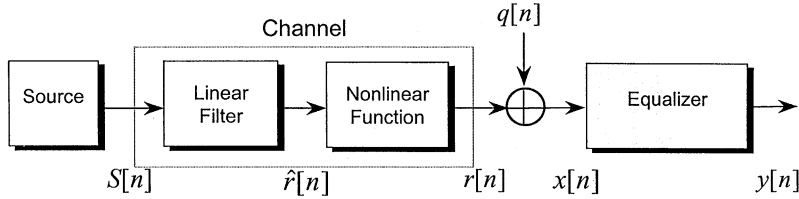
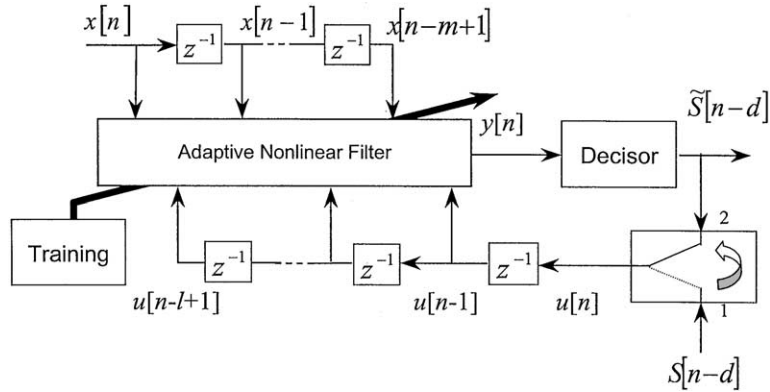Fig. 1. Discrete time model of a communication system.

Fig. 2. Schematic of a decision feedback equalizer.

memoryless, a more general model for the channel is [7,15]:

$$r[n] = f(\hat{r}[n])$$
$$= f(S[n], \ldots, S[n - N + 1]; h_0, \ldots, h_{N-1}),$$
$$(2)$$

where $f(\cdot)$ is a nonlinear function. The signal observed at the receiver end is

$$x[n] = r[n] + q[n], \qquad (3)$$

where $q[n]$ is the additive noise, usually modeled as a Gaussian white process with zero mean and variance σ_0^2.

The proposed equalizer is based on the DF multilayer architecture depicted in Fig. 2. During the learning phase, the feedback links are fed by an internal replica of the transmitted (preamble) sequence. Then the switch commutes from position 1 to position 2 and the equalizer enters into the decision directed mode (DDE) [13] to produce an estimate of $S[n]$. DF links increase the dimension of the input space of the network, thus making the classification task easier [11]. The nonlinear adaptive

filter (formed by a feedforward section of order m and a feedback stage of order l) helps in contrasting nonlinear distortions. The whole filtering process introduces a *decision delay d* [13].

Given the channel model and the feedforward order m, transmitted symbols that contribute to the equalizer decision at time n are collected in the vector:

$$\mathbf{S}[n] = \{S[n], S[n-1], \ldots, S[n-m-N+2]\}.$$
$$(4)$$

The following vectors

$$\mathbf{x}[n] = \{x[n], x[n-1], \ldots, x[n-m+1]\}, \qquad (5)$$

$$\mathbf{u}[n] = \{u[n], u[n-1], \ldots, u[n-l+1]\} \qquad (6)$$

constitute the inputs of the feedforward and the feedback stages, respectively. Since the oldest feedback symbol is $u[n-l+1]$ and the oldest received symbol in (4) is $S[n-m-N+2]$, the feedback order l must be

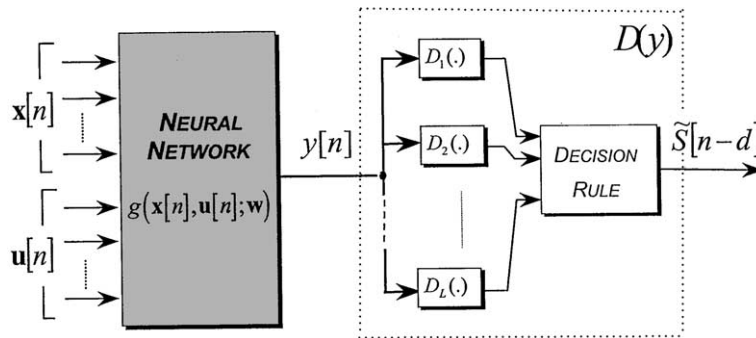$$l \geqslant N + m - d - 2. \qquad (7)$$

Fig. 3. Architecture of the proposed discriminative equalizer.

The proposed architecture is detailed in Fig. 3 and is composed of two parts: a feedforward multilayer neural network and a *discriminative* decision block. The neural network computes an injective mapping $g(\cdot)$ from the complex ($m \times l$)-dimensional input space onto the complex output space Y, in dependence of some weight vector $\mathbf{w}$, to be determined during the supervised training phase. The input–output relationship is

$$y[n] = g(\mathbf{x}[n], \mathbf{u}[n]; \mathbf{w}) \in Y. \tag{8}$$

The decision block produces the estimated symbol sequence:

$$\tilde{S}[n - d] = D\{g(\mathbf{x}[n], \mathbf{u}[n]; \mathbf{w})\}, \tag{9}$$

where $D\{\cdot\}$ represents the decision criterion, to be properly chosen.

3. Discriminative learning

In digital communication systems, transmitted symbols belong to a discrete alphabet $\{s_i; \ i = 1, 2, \ldots, L\}$ and can be recovered by a classification approach [9]. Optimal decision contours separate regions in the output space, mapped by the channel response and the equalizer function.

The proposed equalization approach searches for a trade-off between the classical adaptive channel equalizer and a symbol classifier. The output distribution is modeled by a mixture of Gaussian circular kernels, one for each symbol or *class*. The following *membership function* (or *prototype*) is defined

for the ith class [2,8]:

$$D_i(y) = \exp\left(-\frac{|y - s_i|_2^2}{2\sigma_i^2} \right), \tag{10}$$

where σ_i is the corresponding dispersion parameter. In the proposed approach, prototype dispersions are fixed, while they depend on the noise power in the BA equalizer [2]. Training forces the neural network to produce compact clusters in the output space, within the support region of each Gaussian prototype.

In classification theory, membership functions are usually referred to as *discriminant functions*. Each discriminant function is a generalization of the assumed a posteriori pdf associated with the corresponding symbol. Similar to the Bayes decision rule [11], the decision operator $D(\cdot)$ is defined as

$$D(y) = s_i \quad \text{if } D_i(y) = \max_k D_k(y). \tag{11}$$

This operator involves the calculation of L discriminant functions at each time step and selects the symbol corresponding to the dominant membership function.

The choice of prototypes strongly affects the performance of the learning algorithm. Differently from *least-squares* criterions, where the objective is the *coincidence* between the network outputs and the target symbols s_i, the proposed method attempts to reduce the classification error probability without reaching a minimum MSE.

Traditional MSE criterion minimizes the following error functional:

$$J_{\text{MSE}}(\mathbf{w}) = \frac{1}{2K} \sum_{n=1}^{K} |y(\mathbf{w}) - s_i[n]|_2^2, \tag{12}$$

where $s_i[n]$ is the correct symbol at time n and optimization is performed over $K \geqslant 1$ consecutive time steps. With respect to the MSE criterion, DL reformulates the error functional to take explicitly into account the objective of minimal decision error probability. This requires the introduction of an adequate weighting of the distance between actual outputs and target symbols. Discriminant functions previously defined are incorporated into an objective scalar criterion, suited to optimization procedures. Such a criterion can be derived by defining a proper *mis-classification measure*, which must be a continuous and differentiable function of the network parameter vector $\mathbf{w}$.[1] Adopting Katagiri and Juang solution [9], the mis-classification measure for the generic ith class is defined as

$$R_i(y) = \sqrt[\alpha]{\frac{1}{L-1} \sum_{\substack{j=1 \\ j \neq i}}^{L} D_j(y)^\alpha - D_i(y)}$$

$$= [\Sigma_i(y)]^{1/\alpha} - D_i(y), \tag{13}$$

where

$$\Sigma_i(y) = \frac{1}{L-1} \sum_{\substack{j=1 \\ j \neq i}}^{L} D_j(y)^\alpha \tag{14}$$

with α being a positive number and $D_i(y)$ is given by (10). Formula (13) assumes positive values in correspondence of each wrong decision, and negative values in the opposite case. When $R_i(y)$ is lower than some specified negative threshold η (meaning a correct classification), the global cost function should remain at a constant (small) value.

Finally, the complete DL functional is

$$J_{\text{DL}}(y) = \begin{cases} \exp(\beta R_i(y)) & \text{if } R_i(y) > -|\eta|, \\ \exp(-\beta|\eta|) & \text{otherwise.} \end{cases} \tag{15}$$

_____

[1] The dependence on $\mathbf{w}$ will be omitted in subsequent formulae whenever ambiguities cannot arise.

The threshold η, not present in the original formulation [9], forces network outputs into a compact region between decision boundaries and prototype centers. Values for β are typically chosen in the range [1,4].

4. Learning algorithm

The weight vector can be iteratively adapted by the general descent formula

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \Delta\mathbf{w}[n]. \tag{16}$$

The gradient search procedure adopted in this work gives

$$\Delta\mathbf{w} = -\mu \nabla_w J_{\text{DL}}(y(\mathbf{w}))$$

$$= -\mu \nabla_y J_{\text{DL}}(y) \frac{\partial y(\mathbf{w})}{\partial \mathbf{w}}, \tag{17}$$

where μ is the *learning rate* and the step index n has been omitted for simplicity.

The term $\partial y(\mathbf{w})/\partial \mathbf{w}$ is evaluated by use of the well-known backpropagation algorithm (or least-mean squares, LMS) for complex-valued neural networks [14,15], while the term $\nabla_y J_{\text{DL}}(y)$ is evaluated as

$$\nabla_y J_{\text{DL}}(y) = \begin{cases} \beta J_{\text{DL}}(y) \nabla_y R_i(y) & \text{if } R_i > -|\eta|, \\ 0 & \text{otherwise,} \end{cases}$$
$$\tag{18}$$

$$\nabla_y R_i(y) = [\Sigma_i(y)]^{1/\alpha - 1} \frac{\nabla_y \Sigma_i(y)}{\alpha} - \nabla_y D_i(y), \tag{19}$$

$$\frac{\nabla_y \Sigma_i(y)}{\alpha} = \frac{1}{L-1} \sum_{\substack{j=1 \\ j \neq i}}^{L} D_j(y)^{\alpha-1} \nabla_y D_j(y), \tag{20}$$

$$\nabla_y D_i(y) = -D_i(y) \frac{[y - s_i]}{\sigma_i^2}. \tag{21}$$

Eqs. (18)–(21) must be substituted into (17) to get the final weight update formula.

Optimization of the error functional is carried out in complex arithmetic, by the procedure described in [1,15].

The DL algorithm has only a small computational overhead with respect to a standard neural

Table 1
DL computational overhead with respect to the LMS algorithm

	Additions	Multiplications	Nonlinear functions
LMS	2 (1 complex valued)	0	0
DL overhead			
$D_j(y)$ $j = 1, 2, \ldots, L$	$3L$	$3L$	L (exp)
$\Sigma_i(y)$	$L - 2$	1	$L - 1$ (power of α)
$R_i(y)$	1	0	1 (power of $1/\alpha$)
$J_{\mathrm{DL}}(y)$	0	1	1 (exp)
$\nabla_y J_{\mathrm{DL}}(y)$	$3L - 1$	$3L + 4$	1 (power of $(1/\alpha - 1)$)
Total	$7L - 2$	$6L + 6$	$2L + 2$
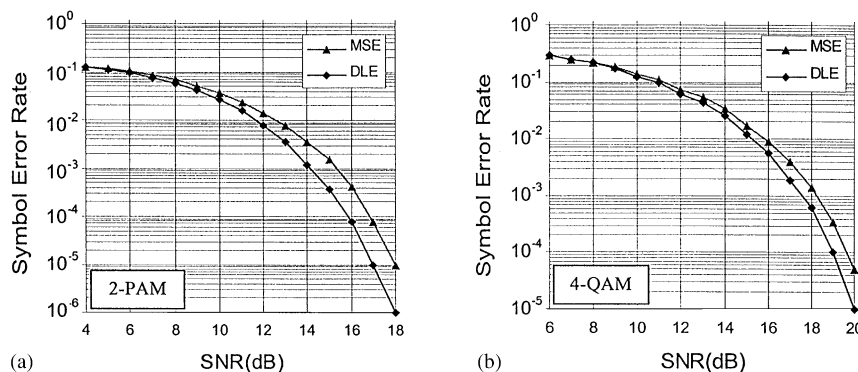
Fig. 4. Channel $H_1(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$: average symbol error rate vs. SNR for (a) 2-PAM and (b) 4-QAM modulations.

equalizer based on the MSE criterion, due to the term $\nabla_y J_{\mathrm{DL}}(y)$ in (17). Table 1 shows in detail this computational overhead. If W is the total number of complex weights in the network and L is the number of symbols, the main cost is given by $4W + 6L$ multiplications for each time step, if a LMS-like algorithm is adopted. In many practical problems $W \gg L$ and this overhead can be neglected. It is important to remark that since α is fixed, nonlinear functions can be conveniently computed through look-up tables.

5. Experimental results

In simulations a two-layer neural network was used. The activation function $F(x)$ was chosen as [3]

$$F(x) = x + a \sin(\pi b x)$$

with $a = 0.03$ and $b = 10$ for the hidden neurons, and $a = 0.05$ and $b = 4$ for the output units. In all experiments, weights were initialized with random complex values extracted from a normal distribution with zero mean and variance $\sigma^2 = 0.1$.

The feedforward filter had length $m = 3$ while the backward filter order l was calculated according to (7). The decision delay was $d = 3$ in all cases.

Training was performed with 2000 output symbols, while performance was evaluated on 10^5 more received symbols and averaged over 10 different network realizations.

Performance of the *DL equalizer* (*DLE*) was tested for different channel models [4,9,10]. Results obtained by an equalizer having the same DF architecture, but trained with the traditional MSE criterion, were also considered for comparison.

The values of α, β and η were empirically optimized. Dispersion parameters σ_i's of discriminant

functions depend on the constellation and were chosen as the half of the (minimum) distance of each symbol from the nearest ones: for instance, $\sigma_i = 1$ for 2-PAM signaling and $\sigma_i = 1/\sqrt{2}$ for 4-QAM. The signal-to-noise ratio (SNR) was defined as

$$\text{SNR} = \frac{E[|r[n]|^2]}{\sigma_0^2}, \tag{22}$$

where $E[\cdot]$ is the expectation operator.

In the following, results obtained on some typical channel models are described, both in terms of classification performance and convergence properties.
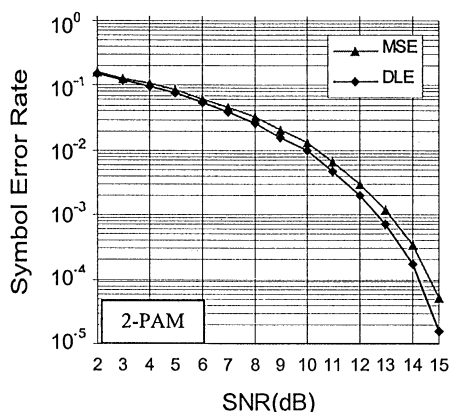
Fig. 5. Channel $H_2(z) = (0.4313 + 0.4311j)(1 - (0.5 + j)z^{-1})$ $(1 - (0.35 + 0.7j)z^{-1})$: average symbol error rate vs. SNR for 2-PAM modulation.
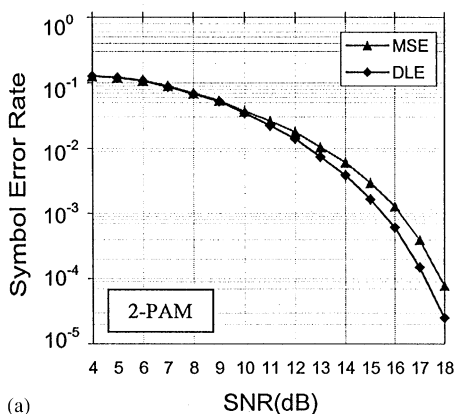
5.1. Classification performance

Test 1. The first example is a typical linear non-minimum-phase channel with transfer function:

$$H_1(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}.$$

Fig. 4 shows the results for 2-PAM and 4-QAM transmission, in terms of symbol error rate vs. the SNR.

Test 2. As a second test a complex-valued linear non-minimum-phase channel was considered, with transfer function:

$$H_2(z) = (0.4313 + 0.4311j)[1 - (0.5 + j)z^{-1}]$$

$$\times [1 - (0.35 + 0.7j)z^{-1}].$$

The curves depicted in Fig. 5 were obtained.

Test 3. This experiment considered a nonlinear channel. According to channel model (3), a nonlinearity was applied to the output $\hat{r}[n]$ of a linear filter, whose transfer function was $H_1(z)$; and the received sequence $x[n]$ was then generated by the following equation:

$$x[n] = \hat{r}[n] + 0.2(\hat{r}[n])^2.$$
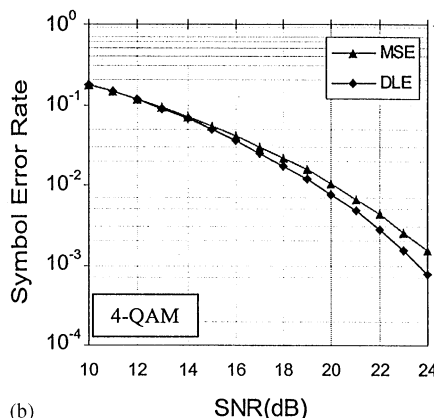
Fig. 6. Nonlinear channel with $H_1(z) = 0.3482 + 0.8704z^{-1} + 0.3482z^{-2}$: average symbol error rate vs. SNR for (a) 2-PAM and (b) 4-QAM modulations.

Both experiments, conducted with 2-PAM (Fig. 6a) and 4-QAM (Fig. 6b) modulations, confirmed the better performance with respect to the MSE criterion, also in the nonlinear case.

Test 4. The last channel was also a nonlinear channel with the linear part given by

$$H_3(z) = -0.2052 - 0.5131z^{-1} - 0.7183z^{-2} + 0.3695z^{-3} + 0.2052z^{-4},$$

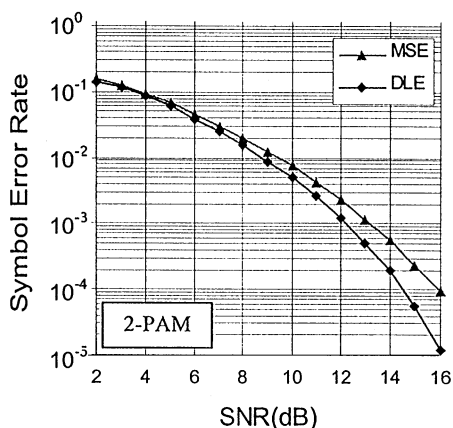while the output was generated as in the previous experiment. Experimental results are shown in Fig. 7.

Fig. 7. Nonlinear channel with $H_3(z) = -0.2052 - 0.5131z^{-1} -0.7183z^{-2} + 0.3695z^{-3} + 0.2052z^{-4}$: average symbol error rate vs. SNR for 2-PAM modulation.

5.2. Convergence properties

Several experiments were performed to evaluate the convergence speed of the algorithm. In fact, a very important feature in equalization is the ability of the equalizer to promptly adapt to channel variations. The convergence rates of the DL and the traditional MSE approaches were compared, using the same networks of previous tests. The proposed algorithm was found to be about 4 times faster, for the same performance in terms of average symbol error rate. Fig. 8 shows four training curves referring to the nonlinear channel model considered in test 2, with SNR = 18 dB and 2-PAM modulation. The DL curves were obtained for different values of the parameter η. For small values of η, the network output is close to the decision boundaries. In contrast, for large η the output is forced around prototype centers; in this case, the DL algorithm behaves similar to the MSE approach. The best performance has been obtained with intermediate settings of η.

Finally, the evidence that MSE minimization does not necessarily imply a minimum symbol error rate was found throughout all simulations. Fig. 9 shows some MSE curves obtained in a typical test. It is clear that DL does not minimize the MSE but scores a better classification performance. Fig. 10 confirms that optimal classification is obtained even if outputs do not converge to the transmitted symbol constellation.
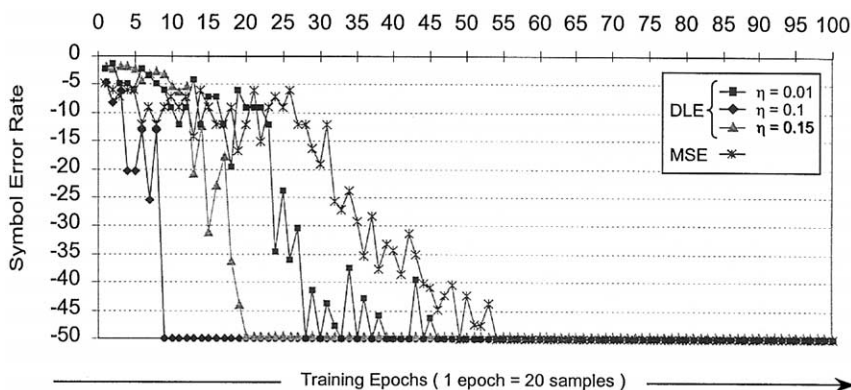

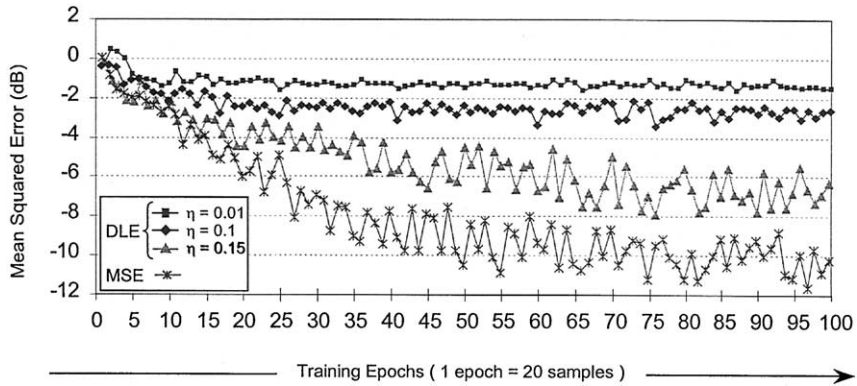Fig. 8. Symbol error rate vs. number of training samples for DL and MSE equalizers.

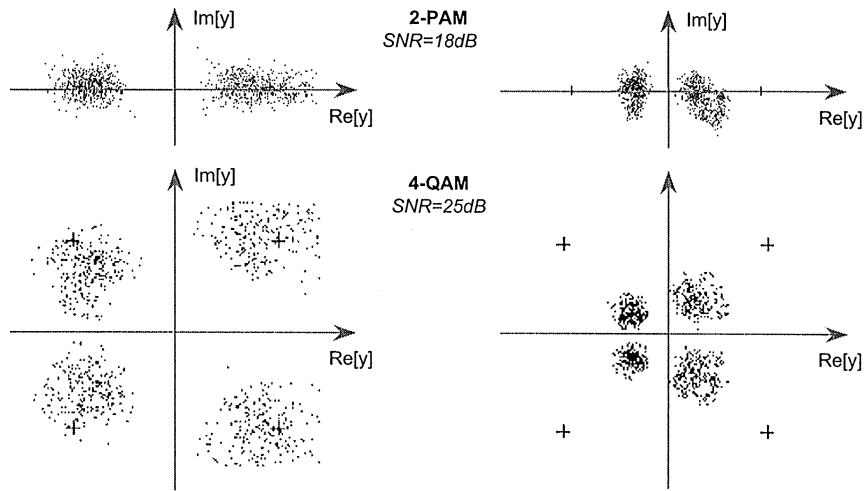Fig. 9. MSE vs. number of training samples for the channel of test 3 (2-PAM signaling and SNR = 18 dB).

Fig. 10. Output clusters for the MSE (left) and DL (right) equalizers.

6. Conclusion

Traditional approaches to adaptive digital channel equalization employ the MSE criterion. Anyway, MSE minimization does not lead to minimum symbol error rate, which is of interest in digital equalization problems. Moreover, MSE-based equalizers generally exhibit a low speed of convergence. In this paper, a new equalizer based on the DL criterion has been introduced. The proposed architecture guarantees a better classification performance without imposing convergence of outputs toward constellation symbols. Experimental tests, performed with different channel models and modulation schemes, confirmed the validity of the new equalizer, in terms of both symbol error rate and speed of convergence.

References

[1] N. Benvenuto, M. Marchesi, F. Piazza, A. Uncini, Non linear radio links equalized using blind neural networks, IEEE International Conference on Acoustic Speech and Signal Processing, Toronto, Canada, May 1991, pp. 1521–1524.

[2] S. Chen, S. McLaughlin, B. Mulgrew, Complex valued RBF network, Part II: application to digital communications channel equalization, Signal Process. 36 (1994) 175–188.

[3] Cheolwoo You, Daesik Hong, Nonlinear blind equalization schemes using complex-valued multilayer feedforward neural networks, IEEE Trans. Neural Networks 9 (6) (November 1998) 1442–1455.

[4] F.F. Cocchi, E.D. Di Claudio, R. Parisi, G. Orlandi, Improved decision feedback equalizer using discriminative neural learning, Proceedings of the International Conference on Neural Networks & Brain, Beijing, China, October 27–30, 1998, pp. 623–625.

[5] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.

[6] G.J. Gibson, S. Chen, C.F.N. Cowan, P.M. Grant, Adaptive equalization of finite nonlinear channels using multilayer perceptrons, Signal Process. 20 (1990) 107–109.

[7] G.J. Gibson, S. Siu, C.F.N. Cowan, Multilayer perceptron structures applied to adaptive equalisers for data communications, Proceedings of the IEEE International Conference on Acoustic Speech and Signal Processing, Glasgow, Scotland, 1989, pp. 1183–1186.

[8] S. Haykin, Neural Networks—A Comprehensive Foundation, IEEE Press, New York, 1994.

[9] S. Katagiri, B.H. Juang, Discriminative learning for minimum error classification, IEEE Trans. Signal Process. 40 (12) (1992) 3043–3054.

[10] G. Kechriotis, E. Zervas, E.S. Manolakos, Using recurrent neural networks for adaptive communications channel equalization, IEEE Trans. Neural Networks 5 (2) (1994) 267–278.

[11] B. Mulgrew, Nonlinear signal processing for adaptive equalization and multi-user detection, Proceedings of the EUSIPCO-98, Vol. 1, Rhodes, Greece, September 7–12, 1998, pp. 537–544.

[12] R. Parisi, E.D. Di Claudio, G. Orlandi, B.D. Rao, Fast adaptive digital equalization by recurrent neural networks, IEEE Trans. Signal Process. 45 (11) (1997) 2731–2739.

[13] J.G. Proakis, Digital Communications, 2nd Edition, McGraw-Hill International, New York, 1989.

[14] D.E. Rumelhart, G.E. Hinton, R.G. Williams, Learning internal representations by error propagation, in: D.E. Rumelhart, J.L. McLelland (Eds.), Parallel Distributed Proceedings: Exploration in the Microstructure of Cognition, Vol. 1, M.I.T. Press/Bradford books, Cambridge, MA, 1986, pp. 318–364.

[15] A. Uncini, L. Vecci, P. Campolucci, F. Piazza, Complex-valued neural networks with adaptive spline activation function for digital radio links nonlinear equalization, IEEE Trans. Signal Process. 47 (2) (February 1999) 505–514.