

# Adaptively Combined LMS and Logistic Equalizers

Manel Martínez-Ramón, *Member, IEEE*, Antonio Artés-Rodríguez, *Senior Member, IEEE*,  
Ángel Navia-Vázquez, *Senior Member, IEEE*, and Aníbal R. Figueiras-Vidal, *Senior Member, IEEE*

**Abstract**—An adaptive, convex linear combination of the outputs of a standard least mean square (LMS) equalizer and a sigmoidal equalizer is proposed. This procedure results in improving the speed of the LMS equalizer while retaining the low steady-state error of the sigmoidal filter. Appropriate adaption schemes for both of the filters and for the combination parameters are established. Simulations of practical communication applications demonstrate the effectiveness of this adaptive combination.

**Index Terms**—Adaptive, equalization, least mean square, logistic, nonlinear.

## I. INTRODUCTION

EQUALIZATION is a fundamental component of modern telecommunication links. Many excellent textbooks describe the necessity and development of modern equalizers [1], [2]. Notably, Widrow and Hoff's least mean square (LMS) algorithm has been pervasively used in adaptive signal processing applications, including equalization, since its formulation four decades ago [3]. This is due not only to its simple computational requirements, but also to its effective compromise between tracking capabilities and residual error. In particular, LMS equalizers offer a good balance between convergence/tracking capabilities and stationary bit error rate (BER) performance.

Nevertheless, it can be argued that minimizing the power of the difference between transmitted symbol  $d$  or decision  $\hat{d}$  (during supervised and decision directed modes, respectively) and the practical value is not the best mechanism that can be used to keep the steady-state BER as low as possible. This is a direct consequence of minimizing an approximate measure of BER, that is, an instantaneous estimate of the average quadratic error. This is well known in the statistics literature, in which a logistic (sigmoidal) output nonlinearity is frequently used to achieve better decision performance since, for class labels  $\pm 1$  in the binary case, the error to be minimized is  $d - \tanh(\gamma \mathbf{w}^T \mathbf{x})$ , where  $\mathbf{w}$  is the weight vector and  $\mathbf{x}$  is the data vector, and this difference is nearer to the decision error than  $d - \mathbf{w}^T \mathbf{x}$ . However, the presence of the function  $\tanh$  introduces a factor  $1 - \tanh^2(\gamma \mathbf{w}^T \mathbf{x})$  in the expression of the error gradient, and, consequently, training is slower because the adaption speed is reduced when  $|\langle \mathbf{w}^T \mathbf{x} \rangle|$  approaches 1. This creates convergence difficulties in on-line applications, such as equalization schemes.

Manuscript received October 29, 2003; revised February 11, 2004. This work was supported in part by Grant TIC2001-0751-C04-02. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. P. C. Ching.

The authors are with Department of Signal Theory and Communications, Universidad Carlos III de Madrid, Avda. Universidad, 28911 Leganés (Madrid), Spain (e-mail: manel@ieee.org).

Digital Object Identifier 10.1109/LSP.2004.835460

Thus, an interesting approach is to try a combination of both structures in order to improve the tradeoff between speed and steady-state BER. An analogous goal has been previously pursued in similar applications by combining cost functions [4]–[6], but here we proceed in a different manner. We adopt the idea of using model mixtures as proposed in [7] and [8], but using a very fast adaptive convex linear combination of both models: this means that the combined equalizer reduces to one of its components when doing so is adequate for the situation (steady state leads to the sigmoidal scheme, a fast change leads to the linear equalizer), and the combination parameter is adapted to an appropriate value for intermediate cases. Thus, this is an adequate form for the work of the proposed combination. The convex character of the combination avoids the possibility of divergence if the combined schemes are convergent.

## II. THE PROPOSED SCHEME

Here, we adopt a scheme that adaptively combines the outputs of a standard LMS

$$o_1 = \mathbf{w}_1^T \mathbf{x} \quad (1a)$$

and a logistic function

$$o_2 = \tanh(\gamma \mathbf{w}_2^T \mathbf{x}) \quad (1b)$$

filter, where  $o_i$  are the outputs,  $\mathbf{w}_i$  is the plant coefficients, and  $\mathbf{x}$  are the inputs. We apply this when the structure is a finite impulse response (FIR) filter, and also when we desire to use a decision feedback equalizer (DFE) scheme, in which case  $\mathbf{x}$  are inputs and previous decision values.

The overall output is

$$o = \lambda o_1 + (1 - \lambda) o_2 \quad (2)$$

where  $\lambda$  is a convex combination parameter. The function of  $\lambda$  in (2) is to make extreme values of  $\lambda$  lead to either a pure LMS or a pure logistic equalization ( $\lambda = 1$  and  $\lambda = 0$ , respectively). If we adapt  $\lambda$  we can go from one extreme to the other to handle, for example, either abrupt changes or steady-state conditions.

We now discuss how to adapt both plants and  $\lambda$  during the decision-directed mode. First, since the LMS filter is the faster component, it is implemented in its usual form, driven by its output decision value  $\hat{d}_1 = \text{sign}(o_1)$  during the decision directed phase, this avoids the effects of the slower convergence of the sigmoidal filter. The sigmoidal component must be guided by the overall  $\hat{d} = \text{sign}(o)$ , because the sole application of its

decision output value  $\hat{d}_2 = \text{sign}(o_2)$  will introduce unnecessary convergence delays. Additionally, we adapt this filter as follows:

$$\mathbf{w}_2[k+1] = \mathbf{w}_2[k] + \mu_2 \left( (1 - \varepsilon)\hat{d}[k] - o_2[k] \right) \mathbf{x}^*[k] \quad (3)$$

where  $1 - \varepsilon$ ,  $\varepsilon \ll 1$  is the traditional target reduction factor used when dealing with sigmoidal outputs to avoid the paralysis effects produced because  $\tanh(\gamma \mathbf{w}_2^T \mathbf{x})$  cannot reach values  $\pm 1$ . Equation (3) is the result of minimizing the square error between  $(1 - \varepsilon)\hat{d}$  and output  $o_2$ , but eliminating factor  $1 - o_2^2$  from the gradient expression

$$\begin{aligned} & \frac{\partial \left[ \left( (1 - \varepsilon)\hat{d} - o_2 \right)^2 \right]}{\partial \mathbf{w}_2} \\ &= -2 \left( (1 - \varepsilon)\hat{d} - o_2 \right) \frac{\partial \tanh(\gamma \mathbf{w}_2^T \mathbf{x})}{\partial \mathbf{w}_2} \\ &= -2 \left( (1 - \varepsilon)\hat{d} - o_2 \right) (1 - \tanh^2(\gamma \mathbf{w}_2^T \mathbf{x})) \gamma \mathbf{x}^*. \end{aligned} \quad (4)$$

This procedure is frequently applied to compensate the speed decrease that obviously appears when  $|o_2|$  is high, and it is clearly equivalent to work with an adaptive step  $\mu_2/(1 - o_2^2)$ . It is easy to show that this is also equivalent to minimizing the Kullback–Leibler divergence, and not the square error, as Hopfield [9] and Hinton [10] have shown in binary cases.

With respect to parameter  $\lambda$ , it is clear that its adaption must be faster than those of  $\mathbf{w}_1$  and  $\mathbf{w}_2$ . Consequently, we use an LMS-type algorithm with a (relatively) high step  $\mu_\lambda$ .

If we minimize

$$C(\lambda) = |\hat{d} - o|^2 \quad (5)$$

the stationary point is given by

$$\lambda = \frac{\hat{d} - o_2}{o_1 - o_2} = \frac{\hat{d} - o_2}{\hat{d} - o_2 - (\hat{d} - o_1)} \quad (6)$$

and this fact creates a practical difficulty for leaving situations in which  $\lambda \approx 0$ , because in these cases  $\hat{d}$  is mainly coming from  $o_2$ , that offers bad tracking capabilities, term  $\hat{d} - o_1$  being not adequate to increase  $\lambda$ . Thus, we prefer to use

$$C'(\lambda) = |\hat{d}_c - o|^2 \quad (7)$$

where

$$\hat{d}_c = \lambda \hat{d}_1 + (1 - \lambda)\hat{d}_2 \quad (8)$$

and the stationary point is

$$\lambda = \frac{\hat{d}_2 - o_2}{\hat{d}_2 - o_2 - (\hat{d}_1 - o_1)} \quad (9)$$

that does not create this difficulty. Then

$$\begin{aligned} \lambda[k+1] &= \lambda[k] + \mu_\lambda \left( \hat{d}_c[k] - o[k] \right) \\ & \left\{ \hat{d}_1[k] - o_1[k] - \hat{d}_2[k] - o_2[k] \right\}^*. \end{aligned} \quad (10)$$

Formulas (3) and (10), plus the standard LMS for  $\mathbf{w}_1$  are the governing equations of our combined equalization scheme.

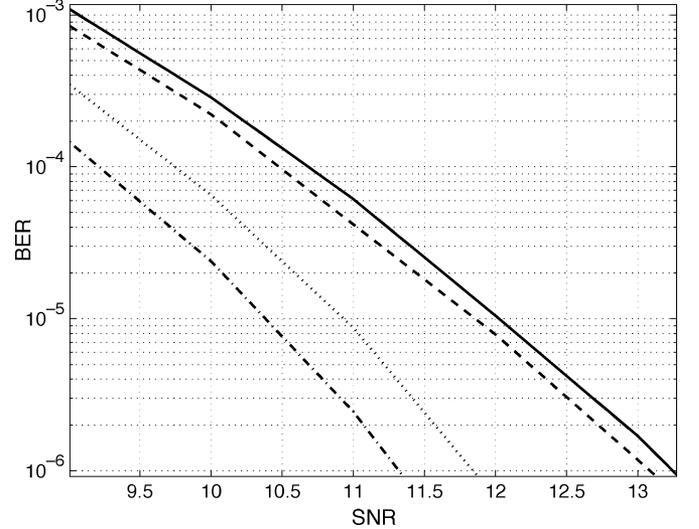


Fig. 1. BER for 3 and 100 km/h. Solid: LMS at 100 km/h. Dash: combination at 100 km/h. Dot: LMS at 3 km/h. Dash-dot: combination and logit at 3 km/h.

### III. SOME EXPERIMENTS

We consider here a typical Universal Mobile Telecommunication Service (UMTS) case, with QPSK transmission. Since we deal with complex symbols, we allow complex values for  $\lambda$  each part creating a convex linear combination for the corresponding (real, imaginary) part of the outputs, and deciding separately for each component (by means of the standard hard decision, applying a sign function). In the simulation examples that follow, we use the usual “saturation” approximation for the real  $\tanh$  function:

$$\text{sat}(u) = \begin{cases} -1, & u < -1 \\ u, & |u| < 1 \\ 1, & u > 1 \end{cases} \quad (11)$$

which is computationally simpler and does not result in significant operational differences.

The results shown here corresponding to a typical UMTS case. The transmitter produces a linear distortion which is represented by means of a FIR filtering having coefficients  $[1, 0.3, -0.2, 0.3, 0.2, -0.1]$ . No phase drift is included. The transmission is QPSK, with a center frequency of 2.15 GHz and a rate of 3.84 Msymbols/s. The transmission channel is channel TU1 (Typical Urban 1) of 3GPP specifications [11]. Doppler effects are simulated according to the specifications and following Shanmugan’s method [12]. Emitter and receiver terminal filters are square-root raised cosine designs with roll-off parameter 0.3. White, zero-mean Gaussian noise is added at the channel output.

The experiment compares the performance of the combined equalizer versus the separate LMS and Logistic components. Both equalizers are  $5 + 5$  DFE schemes (five taps for the input signal plus five taps for the previous decisions). We use  $\gamma = 10$  for the logistic equalizer, which is a reasonably high value (to allow a low residual error), but small enough to avoid local minima effects when considering pedestrian speeds (3 Km/h)

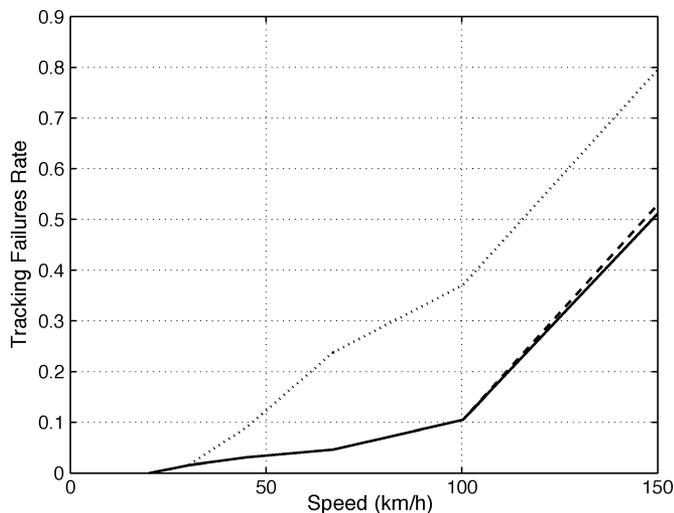


Fig. 2. Tracking failure rate. Solid: LMS. Dash: combination. Dot: logit.

in the adopted environment. Selecting  $\mu_2 = 10^{-3}$  provides a reasonable compromise between BER (very near the attainable minimum) and the speed of convergence for this pedestrian speed.  $\mu_1 = 10^{-2}$  provides an equivalent convergence rate for the LMS equalizer, since numerical values in the corresponding gradient expressions become identical. Finally, a relatively high value is needed for  $\mu_\lambda$  in order to allow an effective adaptive combination of both equalizers. This value is decomposed into two (high) values: 0.1 when the magnitude of  $\lambda$  increases, and 0.01 when it decreases. To select a lower value for the transition from LMS to logistic components (when  $\lambda$  decreases) is to apply a “cautious” policy; the system is more tuned to changes than to reducing the steady-state error. The algorithm works in the supervised mode during the first 1000 samples, and then it is switched to the decision directed mode.

Fig. 1 shows BER versus signal-to-noise ration (SNR) for all the three equalizers in the above context for speeds of 3 km/h and 100 km/h. Remember that the convergence speeds are the same. The advantage of the combined scheme is obvious: at 3 km/h, it works as a logistic equalizer, with a performance improvement of around 1 dB. Predictably, this advantage decreases when the speed increases, because the LMS part tends to be more dominant. However, even at 100 km/h the combined equalizer demonstrate an advantage of approximately 0.5 dB. In comparison, the pure logistic equalizer does not track the changes and yields nearly random decisions (its BER is not shown in the figure).

To verify the efficacy of the combined equalizer, Fig. 2 shows the tracking failure rates for the three equalizers for

SNR = 9 dB. These rates are measured over 1000 runs of 10 000 symbols under decision-directed conditions. It can be seen that LMS and the combined schemes have almost nearly equivalent tracking capabilities, much better than those of the logistic equalizer.

We also remark that we have verified that the presence of nonlinearities in the transmitter increases the advantages of the combined equalizer over pure LMS.

#### IV. CONCLUSIONS

We have shown that by applying adequate algorithmic controls it is possible to linearly combine a standard LMS equalizer and a logistic equalizer in order to obtain the tracking capabilities of the former and the excellent steady-state performance of the latter. The key is to control the LMS component by itself, the logistic element with the overall output, and the combination parameter,  $\lambda$ , quickly enough to avoid being trapped in local minima. Simulation examples that mimic practical communication situations demonstrate these advantages.

Furthermore, we note that the idea of using adaptive combinations of adaptive schemes is completely general; we are actively pursuing the use of such a method in other applications.

#### REFERENCES

- [1] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.
- [2] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2001.
- [3] B. Widrow and M. E. Hoff, Adaptive switching circuits, in *Wescon Conv. Record, Pt 4*, pp. 96–140, 1960.
- [4] J. Chambers and A. Avlonitis, “A robust mixed-norm adaptive filter algorithm,” *IEEE Signal Processing Lett.*, vol. 4, pp. 46–48, Jan. 1997.
- [5] C. Rusu, M. Helsingius, and P. Kuosmanen, “Joint LMS and LMF threshold technique for data echo cancellation,” in *Proc. COST 254 Int. Workshop on Intelligent Comms. and Multimedia Terminals*, J. Tasic, Ed., Ljubljana, Slovenia, 1998, pp. 127–130.
- [6] A. Zerguine, M. Bettayeb, and C. F. N. Cowan, “Hybrid LMS-LMF algorithm for adaptive echo cancellation,” *Proc. Inst. Elect. Eng. Vision, Image and Signal Proc.*, vol. 146, no. 4, pp. 173–180, Oct. 1999.
- [7] A. C. Singer and M. Feder, “Universal linear prediction by model order weighting,” *IEEE Trans. Signal Processing*, vol. 47, pp. 2685–2699, Oct. 1999.
- [8] S. S. Kozat and A. C. Singer, “Multi-stage adaptive signal processing algorithms,” in *Proc. 1st IEEE Sensor Array and Multichannel Signal Processing Workshop*, Cambridge, MA, 2000, pp. 380–384.
- [9] J. J. Hopfield, “Learning algorithm and probability distributions in feed-forward and feedback networks,” *Proc. Nat. Academy of Science*, vol. 84, pp. 8429–8433, 1987.
- [10] G. E. Hinton, “Connectivist learning procedures,” *Artific. Intell.*, vol. 40, pp. 185–234, 1989.
- [11] T. S. G. Working Group 4, Radio Access Network, “UTRA (UE) TTD Radio Transmission and Reception. TS 25.102 v3.0.0,” 3rd Generation Partnership Project 3GPP, Tech. Rep., 1999.
- [12] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Channels*. New York: Plenum, 1994.