NONLINEAR CHANNEL EQUALIZER USING GAUSSIAN SUM APPROXIMATIONS

Patrick GROHAN *Sylvie MARCOS*

Laboratoire des Signaux et Systèmes, SUPELEC, CNRS
Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France
Tel: + 33 1 69 85 17 35 Fax: + 33 1 69 41 30 60

ABSTRACT

The aim of this paper is to revisit the problem of nonlinear channel equalization. The equalization is here viewed as the estimation, from the observation of the channel output, of the state vector of the channel consisting of the last transmitted symbols. If the probability density function of the state vector given all the available observations, (the *a posteriori* density function) were known, an estimate of the state vector for any performance criterion could be determined. Alspach and Sorenson proposed in [5] an approximation by a weighted sum of Gaussian probability density functions that permits the explicit calculation of the *a posteriori* density from the Bayesian recursion relations. The application of these results to the minimum mean square error solution of the nonlinear channel equalization problem provides a new scheme which consists of the convex combination of the output of several extended Kalman filters operating in parallel.

1. INTRODUCTION

The transmission of a message over a band limited and/or a dispersive channel leads to the distortion of the message. A channel introducing such perturbations is modelled by a finite-memory tapped-delay line whose memory accounts for intersymbol interferences (ISI) and which is followed by an additive, Gaussian and white noise. The aim of equalization is to remove these impairments from the received message in order to recover the transmitted message.

High speed data transmission and satellite communications use amplifier devices which usually work near saturation. These amplifier devices introduce in the transmitted message memoryless nonlinearities which, combined with the effects of transmission and reception filters, become nonlinearities with memory. Such channels are usually modelled by Volterra series [1].

Nonlinear channel equalization has become the subject of considerable research interest during the past few years. Chen *et al.* proposed in [2] a multilayer perceptron and dealt with the problem in terms of classification. Biglieri *et al.* rather proposed in [3] a generalization of the ISI cancellation techniques introduced in the case of linear channel equalization. In a previous work, [4], we investigated nonlinear channel equalization by ISI cancellation techniques using polynomial filters; we pointed out the limits of such structures.

These limits lead us to revisit the problem of nonlinear channel equalization. In this paper, we adopt for the channel model a formulation with state equation and observation equation. The aim of equalization is to estimate the state vector of the channel consisting of the last transmitted symbols, from the received message. For linear channel, the Kalman filter provides a good solution, [6]-[7]-[8]. The direct use of a Kalman filter for nonlinear channel requires the linearization of the input-output relation of the channel and consequently to use the extended Kalman filter. This linearization and the Gaussian assumption for the plant noise impair the performances. Knowledge of the probability density function of the state given all available received data provides the most complete possible description of the state, and from this density any of the common types of estimates (e.g., minimum mean square error or maximum *a posteriori*) can be determined. Except in the linear Gaussian case, it is extremely difficult to determine this density function. In this paper, a weigthed sum of Gaussian probability density functions is used to approximate arbitrarily closely another density function that permits the explicit calculation of the *a posteriori* density from the Bayesian recursion relations, [5]. The linearization of the input-output relation of the channel is also required and consequently, the solution to the minimum mean square error equalization problem gives a new nonlinear channel equalizer which consists of the convex combination of the output of several extended Kalman filters operating in parallel. In this paper, the channel is supposed to be known or to have been identified.

After stating the problem in section 2., the Gaussian sum approximation is defined and applied to the plant noise density function in section 3. In section 4., the new nonlinear channel equalizer is proposed. Results of simulations are commented in section 5. Finally, in section 6., conclusions and outlooks are given.

2. PROBLEM STATEMENT

The transmitted sequence, the channel input $\{d(n)\}$, is composed of indepedent symbols which belong to a finite set. The state vector, $\boldsymbol{D}(n)$ of the channel consists of the last N transmitted symbols, $\{d(n-k)\}_{0\leq k\leq N-1}$, where N represents the channel memory. Consequently, the state vector of the system evolves according to the linear difference equation:

$$\boldsymbol{D}(n) = \boldsymbol{F}\boldsymbol{D}(n-1) + \boldsymbol{G}d(n) \qquad (1)$$

where $\boldsymbol{F}$ is the $N \times N$ shift matrix and $\boldsymbol{G}$, the $N \times 1$ vector

$$\boldsymbol{F} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} \quad \boldsymbol{G} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Thus, the nonlinear channel output can be written as:

$$y(n) = h(\boldsymbol{D}(n), \boldsymbol{p}) + b(n) \qquad (2)$$

where $h(., \boldsymbol{p})$ is a possibly nonlinear function of $\boldsymbol{D}(n)$ with parameters $\boldsymbol{p}$ and $b(n)$ represents a zero-mean white Gaussian noise sequence independent from the input sequence, $d(n)$. For example, if the nonlinear channel is modelled by a transversal Volterra filter of order p_h and memory N, then $h(\boldsymbol{D}(n), \boldsymbol{p})$ is equal to:

$$\sum_{m=1}^{p_h} \sum_{k_1=0}^{N-1} \dots \sum_{k_m \geq k_{m-1}}^{N-1} h_m(k_1, \dots, k_m) d(n-k_1) \dots d(n-k_m) \qquad (3)$$

The vector $\boldsymbol{p}$ is here composed of the Volterra kernels $\{h_m(k_1, \dots, k_m)\}$. According to this formulation, the equalization is equivalent to estimating the state vector $\boldsymbol{D}(n)$ from the observation of the channel output $\boldsymbol{Y}^n = [y(n), y(n-1), \dots, y(0)]$. One can note here that the estimation of $d(n)$ can be obtained at some delayed time $(n+D)$ where $0 \leq D \leq (N-1)$. In the case of a linear channel ($p_h = 1$ in Eq. 3) with additive Gaussian noise, this problem has been resolved by the Kalman observer, [6]-[7]-[8].

As the state vector, $\boldsymbol{D}(n)$, is a random vector, the *a posteriori* density, $p(\boldsymbol{D}(n)|\boldsymbol{Y}^n)$, provides the most complete possible description of the state vector. This density is determined recursively from the Bayesian recursion relations:

$$p(\boldsymbol{D}(n)|\boldsymbol{Y}^n) = c_n p(\boldsymbol{D}(n)|\boldsymbol{Y}^{(n-1)}) p(y(n)|\boldsymbol{D}(n)) \qquad (4)$$

$$p(\boldsymbol{D}(n)|\boldsymbol{Y}^{(n-1)}) = \int p(\boldsymbol{D}(n)|\boldsymbol{D}(n-1)) \, p(\boldsymbol{D}(n-1)|\boldsymbol{Y}^{(n-1)}) d\boldsymbol{D}(n-1) \qquad (5)$$

where the normalizing constant c_n is given by

$$\begin{aligned} 1/c_n &= p(y(n)|\boldsymbol{Y}^{(n-1)}) \\ &= \int p(y(n)|\boldsymbol{D}(n)) p(\boldsymbol{D}(n)|\boldsymbol{Y}^{(n-1)}) d\boldsymbol{D}(n) \end{aligned}$$

The densities $p(y(n)|\boldsymbol{D}(n))$ and $p(\boldsymbol{D}(n)|\boldsymbol{D}(n-1))$ are determined from the equations 1 and 2 and the *a priori* distributions for $d(n)$ and $b(n)$.

It is generally impossible to determine $p(\boldsymbol{D}(n)|\boldsymbol{Y}^n)$ in a closed form using equations 4 and 5, except when the system 1-2, is linear and the *a priori* distributions are Gaussian, the Kalman filter being, then, the solution.

In the following, a procedure is defined that is based on the use of the Gaussian sum representation of the *a posteriori* density function in conjunction with the linearization procedure that has proven so effective in Kalman filter applications.

3. GAUSSIAN SUM APPROXIMATION

The Gaussian sum representation p_A of a density function associated with a random vector $\boldsymbol{x}$ is defined as:

$$p_A(\boldsymbol{x}) = \sum_{i=1}^{l} \alpha_i \mathcal{N}[\boldsymbol{x} - \boldsymbol{a}_i, \boldsymbol{B}_i] \qquad (6)$$

where

$$\begin{aligned} \mathcal{N}[\boldsymbol{a}, \boldsymbol{B}] &= \exp\{-\frac{1}{2}\boldsymbol{a}^T \boldsymbol{B}^{-1} \boldsymbol{a}\}/(2\pi)^{n/2} |\boldsymbol{B}|^{1/2} \\ \sum_{i=1}^{l} \alpha_i &= 1, \ \alpha_i \geq 0 \ \text{ for all } i \end{aligned}$$

It can be shown, [5], that p_A converges uniformly to any density function of practical concern as the number of terms l increases and the covariance $\boldsymbol{B}_i$ approaches the zero matrix.

Let us consider as an example, the *a priori* density function of the plant noise, $\boldsymbol{w}(n) = \boldsymbol{G}d(n)$ (Eq. 1). If $\{d_l\}_{1\leq l\leq q}$, are the q values, that $d(n)$ can take, associated respectively with the probabilities $\{p_l\}_{1\leq l\leq q}$, then the density function of $\boldsymbol{w}(n)$, $p(\boldsymbol{w}(n))$, is equal to:

$$\begin{aligned} &p_l \quad \text{if } d(n) = d_l, \ 1 \leq l \leq q &(7) \\ &0 \quad \text{otherwhise} &(8) \end{aligned}$$

This density function is approximated by a weighted sum of Gaussian density functions according that this plant noise takes a finite number of discrete values, $\{\boldsymbol{G}d_l\}_{1\leq l\leq q}$. This assumption yields to choose:

$$p(\boldsymbol{w}(n)) = \sum_{l=1}^{q} p_l \mathcal{N}[\boldsymbol{w}(n) - \boldsymbol{w}_l, \boldsymbol{Q}_l] \qquad (9)$$

where $\boldsymbol{w}_l = \boldsymbol{G}d_l$, $\boldsymbol{Q}_l = \epsilon \boldsymbol{I}_N$, $\boldsymbol{I}_N$ the unity matrix, and ϵ choosen small enougth in order that each Gaussian density function is located on a neighborhood of $\boldsymbol{w}_l$ with a probability mass equal to p_l.

4. NONLINEAR CHANNEL EQUALIZER

The main idea of the following procedure is the approximation of the density functions $p(\boldsymbol{D}(n)|\boldsymbol{Y}^n)$ and $p(\boldsymbol{D}(n)|\boldsymbol{Y}^{(n-1)})$ by weighted sums of Gaussian density functions as defined below:.

$$p(\boldsymbol{D}(n)|\boldsymbol{Y}^n) = \sum_{i=1}^{\xi_n} \alpha_{i,n} \mathcal{N}[\boldsymbol{D}(n) - \hat{\boldsymbol{D}}_i(n), \boldsymbol{P}_i(n)] \quad (10)$$

$$p(\boldsymbol{D}(n)|\boldsymbol{Y}^{(n-1)}) = \sum_{i=1}^{\xi'_n} \alpha'_{i,n} \mathcal{N}[\boldsymbol{D}(n) - \hat{\boldsymbol{D}}'_i(n), \boldsymbol{P}'_i(n)] \quad (11)$$

where $\hat{\boldsymbol{D}}_i(n)$ and $\hat{\boldsymbol{D}}'_i(n)$ are $N \times 1$ vectors and $\boldsymbol{P}_i(n)$ and $\boldsymbol{P}'_i(n)$ $N \times N$ matrices, defined below.
The *a priori* density function of the plant noise, $\boldsymbol{w}(n)$, is also supposed to be approximated by a weighted sum of Gaussian density functions as defined in section 3. and the *a priori* density function of the additive noise, $b(n)$, is supposed to be Gaussian with variance σ_b^2. With these assumptions and the linearization of the nonlinearity $h(.,\boldsymbol{p})$ relative to $\hat{\boldsymbol{D}}'_i(n)$, the Bayesian recursion relations 4-5 can be derived easily, [5], and yield for parameters ξ'_n, $\alpha'_{i,n}$, $\hat{\boldsymbol{D}}'_i(n)$, $\boldsymbol{P}'_i(n)$, ξ_n, $\hat{\boldsymbol{D}}_i(n)$, $\boldsymbol{P}_i(n)$,$\alpha_{i,n}$, in the case of the Gaussian sum approximations 10-11, the following relations:
Prediction resulting from the derivation of Eq. 5

$$\begin{aligned} \xi'_n &= q\xi_{n-1} && (12)\\ \alpha'_{i,n} &= p_l \alpha_{j,n-1} && (13)\\ \hat{\boldsymbol{D}}'_i(n) &= \boldsymbol{F}\hat{\boldsymbol{D}}_j(n-1) + \boldsymbol{w}_l && (14)\\ \boldsymbol{P}'_i(n) &= \boldsymbol{F}\boldsymbol{P}_j(n-1)\boldsymbol{F}^T + \boldsymbol{Q}_l && (15) \end{aligned}$$

Estimation resulting from the derivation of Eq. 4

$$\begin{aligned} \xi_n &= \xi'_n && (16)\\ \hat{\boldsymbol{D}}_i(n) &= \hat{\boldsymbol{D}}'_i(n) + \boldsymbol{K}_i(n)(y(n) - h(\hat{\boldsymbol{D}}'_i(n), \boldsymbol{p})) && (17)\\ \boldsymbol{P}_i(n) &= \boldsymbol{P}'_i(n) - \boldsymbol{K}_i(n)\boldsymbol{H}_i(n)\boldsymbol{P}'_i(n) && (18)\\ \boldsymbol{K}_i(n) &= \boldsymbol{P}'_i(n)\boldsymbol{H}_i(n)^T[\boldsymbol{H}_i(n)\boldsymbol{P}'_i(n)\boldsymbol{H}_i(n)^T + \sigma_b^2]^{-1} && (19)\\ \boldsymbol{H}_i(n) &= \frac{\partial h(\boldsymbol{D}(n), \boldsymbol{p})}{\partial \boldsymbol{D}(n)}\Big|_{\boldsymbol{D}(n) = \hat{\boldsymbol{D}}'_i(n)} && (20)\\ \alpha_{i,n} &= \frac{\alpha'_{i,n}\beta_{i,n}}{\sum_{i=1}^{\xi'_n} \alpha'_{i,n}\beta_{i,n}} && (21)\\ \beta_{i,n} &= \mathcal{N}[y(n) - h(\hat{\boldsymbol{D}}'_i(n), \boldsymbol{p}), \boldsymbol{H}_i(n)\boldsymbol{P}'_i(n)\boldsymbol{H}_i(n)^T + \sigma_b^2] \end{aligned}$$

The estimated state vector, $\hat{\boldsymbol{D}}(n)$, of the state vector $\boldsymbol{D}(n)$, solution to the minimum mean square error estimation problem is given by the conditional expectation, $\mathrm{E}[\hat{\boldsymbol{D}}(n)|\boldsymbol{Y}^n]$ with associated error covariance matrix, $\bar{\boldsymbol{P}}(n)$ defined as $\mathrm{E}[(\boldsymbol{D}(n) - \hat{\boldsymbol{D}}(n))(\boldsymbol{D}(n) - \hat{\boldsymbol{D}}(n))^T|\boldsymbol{Y}^n]$. It yields the following equations:

$$\hat{\boldsymbol{D}}(n) = \sum_{i=1}^{\xi_n} \alpha_{i,n}\hat{\boldsymbol{D}}_i(n) \quad (22)$$

$$\bar{\boldsymbol{P}}(n) = \sum_{i=1}^{\xi_n} \alpha_{i,n}(\boldsymbol{P}_i(n) + (\hat{\boldsymbol{D}}_i(n) - \hat{\boldsymbol{D}}(n))(\hat{\boldsymbol{D}}_i(n) - \hat{\boldsymbol{D}}(n))^T) \quad (23)$$

One can note here that the equations 14-15 and 17-20 are those of the extended Kalman filter and consequently that $\hat{\boldsymbol{D}}(n)$ (see Eq. 22) is the convex combination of ξ_n extended Kalman filters operating in parallel. The major drawback of this scheme is the growth of the number of terms used in the Gaussian sum approximation (see Eq. 12). A way to reduce this number of terms is to consider that the Gaussian sum approximation for $p(\boldsymbol{D}(n)|\boldsymbol{Y}^n)$ is reduced to only one Gaussian density function with mean $\hat{\boldsymbol{D}}(n)$ and covariance matrix $\bar{\boldsymbol{P}}(n)$ after the estimation procedure. This means that after the estimation procedure, ξ_n is forced to be equal to 1, $\hat{\boldsymbol{D}}_i(n)$ equal to $\hat{\boldsymbol{D}}(n)$, $\boldsymbol{P}_i(n)$ equal to $\bar{\boldsymbol{P}}(n)$ and $\alpha_{i,n}$ equal to 1. This is a proper assumption because the density function $p(\boldsymbol{D}(n)|\boldsymbol{Y}^n)$ must be only located on one possible state vector. This assumption reduced to q the number of extended Kalman filters. The figure 1 shows the scheme resulting with equally probable (± 1) binary channel input symbols, $d(n)$. This scheme is evidently followed by a thresholding device in order to recover estimated symbols belonging to the channel input alphabet.
To start the procedure, it is necessary that the probability density function prescribed for the initial state, $p(\boldsymbol{D}(0))$, be correctly represented as a Gaussian sum. We propose to use:

$$p(\boldsymbol{D}(0)) = \sum_{i=1}^{q^N} (\prod_{l=1}^{N} p_l) \mathcal{N}[\boldsymbol{D}(0) - \hat{\boldsymbol{D}}_i(0), \boldsymbol{P}_i(0)] \quad (24)$$

where $\hat{\boldsymbol{D}}_i(0)$ is equal to one of the q^N possible combinations of a $N \times 1$ vector with components choosen among $\{d_l\}_{1 \le l \le q}$ and $\boldsymbol{P}_i(0) = \epsilon_0 \boldsymbol{I}_N$ with ϵ_0 small.

5. SIMULATIONS

In this section, the novel equalizer is compared with the transversal Volterra equalizer, solution to the minimum mean square error equalization problem. Two nonlinear channels modeled by Volterra filters of order 2 and memory 2 are considered. For both channels, the linear kernels are [1; 0.5] whereas the quadratic kernels are $[0.1; 0.1; 0.025]$ for the first one and $[0.9; 0.9; 0.225]$ for the second one. The nonlinear ISI introducted by the second channel are consequently more important than the ISI introduced by the first one. A binary sequence is chosen to drive the channel and the equalizer scheme represented in figure 1 is therefore used. The tables 1 and 2 collect the bit error rates at the output of a transversal linear equalizer with memory 9, of a transversal Volterra equalizer of order 2 and with memory 9 and of the novel equalizer.

For the first channel, when no noise is present at the output, all the considered schemes can equalize the channel. Whereas, when a noise with signal to noise (SNR) equal to 10dB is present at the output, one can note the improvement of the performances in terms of bit error rate by the novel scheme. Although, when the order of the transversal Volterra equalizer grows (here from 1 to 2), the bit error rate decreases, the new equalizer divides the bit error rate about by 3. One can think that the increase in the order of the Volterra equalizer may give results equivalent to the ones of the new equalizer but, then, the complexity of the Volterra equalizer would be useless increased. The introduction of a delay, D, in the estimation of the transmitted symbol also permits to reduce the bit error rate.

For the second channel for which the nonlinear ISI are important, even when there is no noise at the output, only the new scheme is able to correctly equalize this channel.

6. CONCLUSION

In this paper, a new nonlinear channel equalizer is proposed which gives good performances compared with the ones of the transversal Volterra equalizer. However, although this new scheme is full of promise, it is not fully adaptive in the sense that the parameters, $\boldsymbol{p}$, of the nonlinear function $h(., \boldsymbol{p})$ and the additive noise variance must be known. Identification techniques, as used in [7], may be applied directly in parallel with this scheme. Nevertheless, some drawbacks in the uniqueness of the parameters, $\boldsymbol{p}$, may appear. Our research is moving towards an adaptive scheme.

REFERENCES

[1] S. Benedetto, E. Biglieri, R. Daffara, Modeling and performance evaluation of nonlinear satellite links – A Volterra series approach, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 15, pp. 494-506, July 1979.

[2] S. Chen, G. J. Gibson, C.F.N. Cowan, P. M. Grant, Adaptive Equalization of Finite Nonlinear Channels Using Multilayer Perceptrons, *Signal Processing,*, vol 20, pp. 107-119, 1990.

[3] E. Biglieri, A. Gersho, R.D. Gitlin, T.L. Lim, Adaptive Cancellation of Nonlinear Intersymbol Interference for Voiceband Data Transmission, *IEEE Journal on Selected Areas in Communications*, vol. SAC-2, no. 5, Sept. 1984.

[4] P. Grohan, S. Marcos, Limits and Cures of Nonlinear Channel Equalization by Polynomial Filters, *IEEE Workshop on Nonlinear Signal and Image Processing*, vol. 1, Halkidiki, Greece, pp. 372-375, June 1995.

[5] D. L. Alspach, H. W. Sorenson, Nonlinear Bayesian Estimation Using Gaussian Sum Approximations, *IEEE Trans. on Automatic Control*, vol. 17, no. 4, pp. 439-447, Sept. 1972.

[6] R. E. Lawrence, H. Kaufman, The Kalman Filter for the Equalization of a Digital Communications Channel, *IEEE Trans. Commun. Technol.*, pp. 1137-1141, Dec. 1971.

[7] B. Mulgrew, C. F. N. Cowan, An Adaptive Kalman Equalizer: Structure and Performance, *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 35, no. 12, pp. 1727-1735, Dec. 1987.

[8] P. Grohan, S. Marcos, Structures and Performances of Several Adaptive Kalman Equalizers, *IEEE Workshop on Digital Signal Processing*, Loen, Norway, pp. 454-457, Sept. 1996.

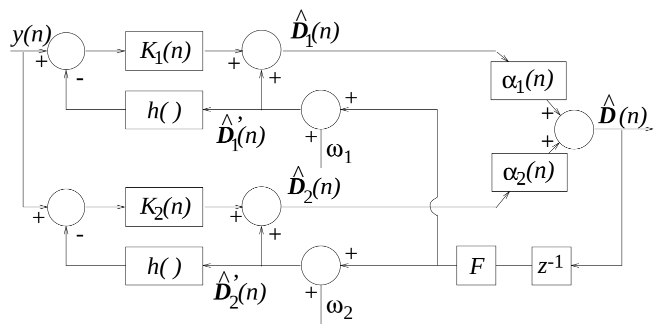

Figure 1. New nonlinear channel equalizer for a binary channel input

Channel 1	$D=0$	$D=0$	$D=1$
SNR	$+\infty$	10dB	10dB
No equalizer	0	0.0359	0.4613
Linear equalizer	0	0.0080	0.0066
Volterra equalizer	0	0.0077	0.0056
Novel equalizer	0	0.0024	0.0015

Table 1. Bit error rates

Channel 2	$D=0$
SNR	$+\infty$
No equalizer	0.2494
Linear equalizer	0.2326
Volterra equalizer	0.2127
Novel equalizer	0.0001

Table 2. Bit error rates