NONLINEAR PREDICTION OF INFRARED DATA BY THE WIENER SYSTEM

P. Bernabeu, I.Bosch, L.Vergara
Department of Communications, EPSA, University Polytechnic of Valencia,
Plaza Ferrándiz-Carbonell s/n 03801, Alcoi, Alicante, Spain
Tel: +34 96 6528512; Fax: +34 96 6528461
e-mail: pbernabe@dcom.upv.es

ABSTRACT

We consider the use of the Wiener system to perform nonlinear prediction. In this paper we propose a technique to retain the simplicity of the linear prediction by including a memoryless nonlinear function. The design of this later is approached from a Bayesian perspective: we look for the conditional mean of the predicted value, given the output of the linear predictor. Two techniques are proposed: the first one makes use of a closed form solution where some higher-order statistics are to be estimated. The second one is a direct sample estimate of the conditional mean given a data training set. The techniques are applied to improve the signal to noise ratio in the automatic detection of fire by infrared signal processing.

1. INTRODUCTION

The Wiener system has been extensively applied in the area of system identification (see [1]-[4] for illustration of some recent works on this topic). It allows a simple generalization of a linear identification system by including at the output a memoryless nonlinear function. Using only one Wiener system we may obtain significant improvements with respect to the linear case, depending on the statistical distribution of the underlying stochastic processes. On the other hand prediction is a key area of signal processing and time series analysis. In an statistical context, the minimum mean-squared error (MMSE) prediction is the conditional mean of the random variable corresponding to the predicted sample given the past samples from which prediction is to be made. Assuming Gaussianity, the conditional mean is a linear function of the samples, and we have several different standard methods for computing the predictor coefficients. However, there are many applications where Gaussianity is not a realistic hypothesis, then some non-linear schemes are to be devised.

Here we propose the use of the Wiener system (Figure 1) to reduce in a simple manner the linear MMSE (LMMSE) which can be achieved by linear prediction. The linear part of the Wiener system is a LMMSE predictor followed by the memoryless nonlinear function. The design of this later follows a Bayesian approach: we estimate the conditional mean of the predicted value, given the output of the linear predictor.

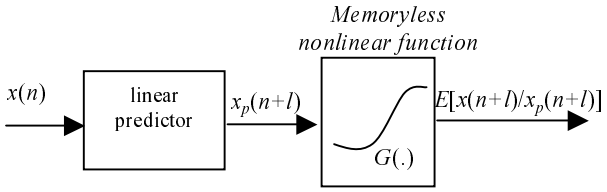

Figure 1: Proposed scheme for nonlinear prediction

In the next section we present two possible ways for designing the memoryless nonlinear estimator and in the last section we apply the proposed structure to a real problem: fire detection by infrared radar, showing the improvement with respect to the classical linear solution.

Part of this work has been presented by the authors in [5].

2. MEMORYLESS FUNCTION DESIGN

Let us assume that we have a training sample record to be used for designing the predictor. The linear part of the structure may be devised following any of the standard methods (for example we may estimate the autocorrelation function and then solve the Wiener-Hopf equations, for a given predictor dimension). Relative to the nonlinear part we propose two alternatives.

The first one is a direct sample estimate of the nonlinearity. Let $x_p(n+l)$ be the output of the linear predictor (i.e., $x_p(n+l)$ is the linear prediction of $x(n+l)$ from the past samples $x(n)$ $x(n\text{-}1)$ ...$x(n\text{-}N+1)$, where n is the linear predictor order and l the prediction time lag), we have to implement the conditional mean.

$$G\left(x_p(n+l)\right) = E\left[x(n+l)/x_p(n+l)\right] \qquad (1)$$

Assuming stationarity, a sample estimate of $G(.)$ may be made in the following manner. We apply the linear predictor to the training record, thus generating a record of sample predictions. The original samples and its corresponding predictions are sorted in ascending order, then a moving average (moving conditional mean estimate) is made on the sorted original data to generate smoothed estimates of $G(.)$ at the corresponding values given by the sorted sample predictions. The width and shape of the moving window controls the degree of smoothing to be obtained in the $G(.)$ estimate. On the other hand, the function $G(.)$ at any desired value may be computed through interpolation of the estimated values.

The above procedure is impractical in the general case of trying to estimate the conditional mean of $x(n+l)$ given de past samples $x(n)$ $x(n\text{-}1)$ ... $x(n\text{-}N+1)$, due to the multidimensionality of the problem. The nonlinearity memoryless condition allows a practical possibility for including it in a calibration procedure by using a training sample record.

A closed form solution for the nonlinear estimator is always desirable. In [6] the author presents a general formula for the (multidimensional) conditional mean that may be used for a polynomial approximation of the nonlinearity. We particularize (see appendix A) the proposed formula for our one-dimensional case, arriving to (for simplicity in the notation we will call in the following $x \equiv x(n+l)$ $x_p \equiv x_p(n+l)$)

$$G(x_p) = \sum_{m=1}^{\infty} \frac{1}{m!} C_m(x, x_p) H_m(x_p) \qquad (2)$$

Where $C_m(x, x_p)$ is a cross-cumulant defined in the form

$$C_m(x, x_p) = cum\left(x, \overbrace{x_p, \ldots, x_p}^{m\ times}\right)$$

and H_m is the Hermite polynomial of order m.

In (2) we have to assume that x_p is a zero-mean, variance normalized, stationary (no n-dependence of the involved cross-cumulants) Gaussian sequence. Note that assuming Gaussianity for the prediction x_p, is a realistic hypothesis, because x_p is the output of the linear prediction filter: a linear combination of N past samples of the original sequence. For a large enough N value (for the application presented in the last section we have observed that N>3 is enough to obtain approximate Gaussian predictions), the central limit theorem guarantees Gaussianity.

To be more specific, a third-order approximation of (2) (i.e., we consider only the first three terms in the sum) leads to the following polynomial approximation (see Appendix A):

$$G_3(x_p) = \frac{1}{6}(a_3 - 3a_1)x_p^3 + \frac{1}{2}a_2 x_p^2 + \left(\frac{5}{2}a_1 - \frac{1}{2}a_3\right)x_p - \frac{1}{2}a_2 \qquad (3)$$

where $a_i = E[x x_p^i]$ are the cross-moments to be estimated from the training data.

3. APPLICATION TO INFRARED RADAR DATA

In [7] the authors present a real data application: fire detection in a wide area by infrared radar. A general scheme for automatic detection is presented where prediction of the infrared level in a given cell, from the values corresponding to the same cell in past scanning times is a key procedure to improve signal to noise ratio and then to improve the probability of detection. The prediction is subtracted from the running value and then detection is made based on the prediction residue. In such a way we take advantage of the highly correlated environment among the different scanning times, thus partly suppressing the infrared background noise. It is shown in [7] that the involved signals are not always Gaussian distributed, so we can try to improve the quality of the prediction by a nonlinear predictor.

On the other hand, prediction lags greater than 1 are necessary to perform automatic detection from a vector whose elements are values measured on the same cell in several consecutive scanning times. In this way, it is possible to incorporate in the detection, information about the expected evolution (usually a slow increasing) of an uncontrolled fire.

We have considered the possibility of applying a Wiener system to this practical problem.

The data were collected by a passive infrared radar, located in a mountainous area in the Southeast of Spain (Alcoy, Alicante). The sensor model was Thermoprofile 6 HT (AGEMA Infrared Systems). Electronic range scanning and an azimuth mechanical system were used. The recorded data depends on three variables: time (or scan number), azimuth and range. We are going to predict in the time domain so we can associate the variable n to the corresponding scan number.

We have selected a data block for training and testing the nonlinear predictor. This implies that stationarity is assumed in the data block as a whole. The selected data block is formed by 400 records of 21 samples each. In a given record, each sample corresponds to a particular cell in a given scan number (the scan period is 1 minute, the actual times were 17:00h to 17:20h, so we have 21 samples in each record). The 400=100x4 records correspond to 100 consecutive ranges with respect to the normal to the sensor and advancing in taken in 4 consecutive azimuths (starting at 7,3 degrees series of 0.23 degrees). The first range corresponds to approximately 10 *Km* in distance and the last one to approximately 900 *m* in distance. We have used the records corresponding to the two first azimuths for training (i.e., for estimating the linear predictor coefficients and the nonlinearity) and the records corresponding to the two last azimuths for testing the predictor performance, i.e., for estimating the prediction-error mean-power (PEMP). We show in figure 2a the normalized PEMP (NPEMP) obtained with the testing data block, for N=8 and different l values, in figure 2b, for N=6 and different l values, figure 2c, for N=4 and different l values, and figure 2d, for N=2 and different l values. Normalization is made by dividing PEMP by the mean-power of the original data x. Finally, the figure 3 shows the memoryless nonlinear function estimated from both the direct sample estimate (the moving window was 100 sample duration) and the third-order polynomial approximation of equation (3) (figure 3a, the linear predictor order is N=8, figure 3b is N=6, figure 3c is N=4 and figure 3d is N=2 , and the prediction lag is l=6 in all cases). The curves are superimposed on the joint distribution $f(x,x_p)$ of the corresponding training data.

We can clearly observe the advantage of including the nonlinear system when N increases, and to a certain N when l value increases.

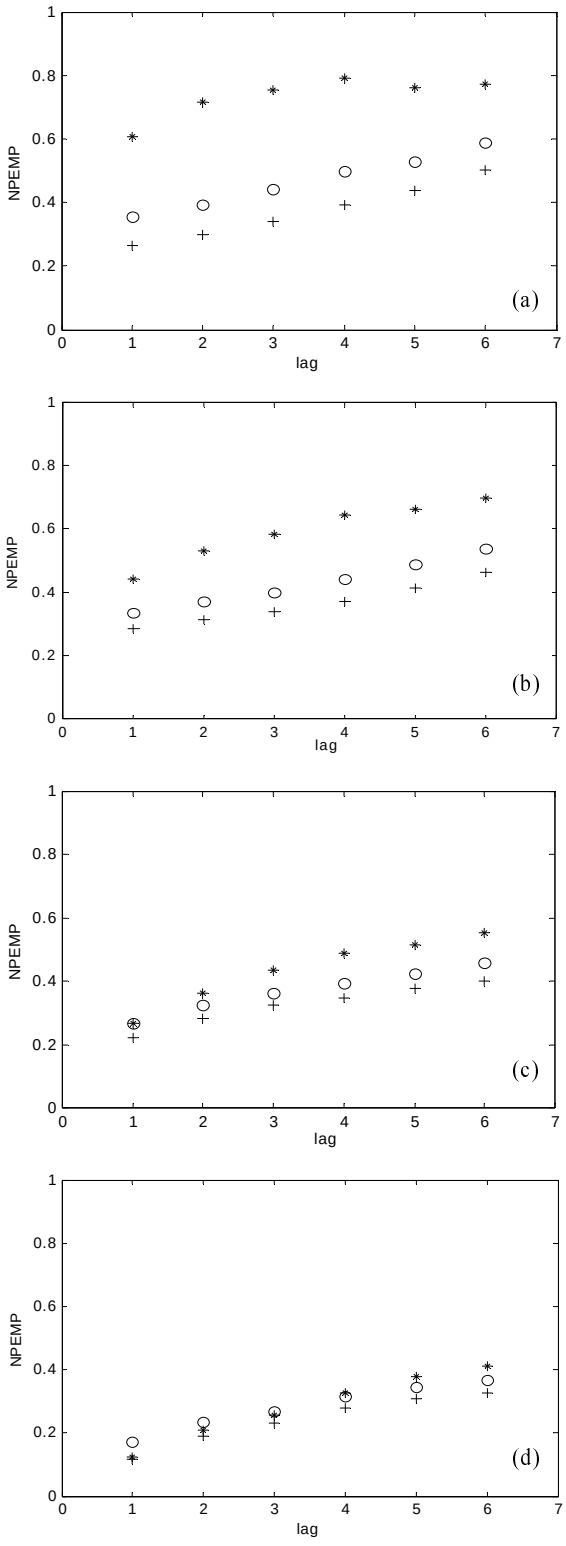

Figure 2. Normalized Prediction Error Mean Power for different values of the prediction time lag l. (a) N=8 in all the cases (b) N=6 in all the cases, (c) N=4 in all the cases, and (d) N=2 in all cases.

* Only the linear predictor

o Linear predictor plus memoryless nonlinearity, third-order polynomial approximation of the conditional mean

+ Linear predictor plus memoryless nonlinearity, sample estimate of the conditional mean.

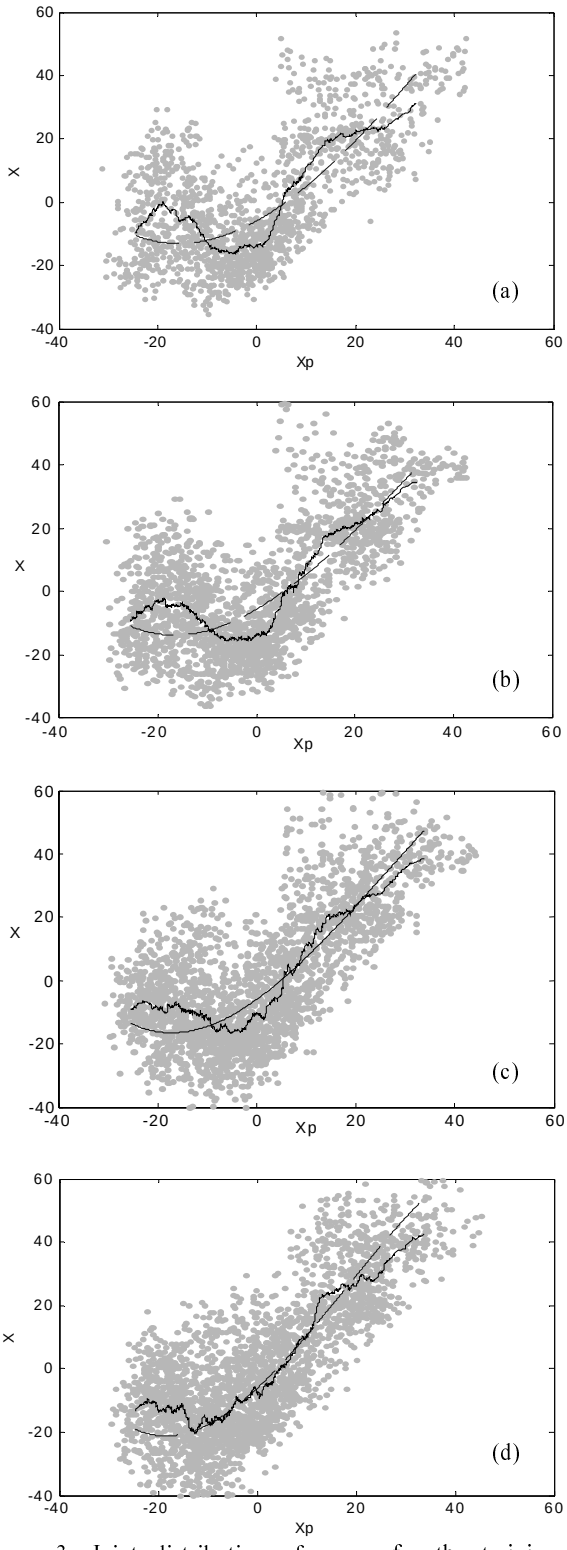

Figure 3. Joint distribution of x, x_p, for the training data. Superimposed are the designed memoryless nonlinear function: Direct sample estimate (solid line) and third-order polynomial fitting equation (3) (dashed line). (a) for N=6 and l=6, (b) for N=4 *and* l=6 and (c) for N=2 and l=6.

Otherwise, despite the N value increases, only if we use the linear predictor the NPEMP value increases, while this last value reminds if we include a nonlinear block.

As expected, the best performance is obtained with the sample estimate of the conditional mean, the polynomial approximation exhibits an intermediate behavior which could be improved by increasing the approximation order. Similar results have been obtained on this application working with other data blocks.

4. CONCLUSIONS

A simple procedure for improving linear prediction has been presented. The memoryless nonlinear function may be directly calibrated from a training block by estimating the conditional mean of the nonlinear prediction given the linear one. A sample estimate is always possible in the above indicated form, but, most interesting, a closed form polynomial approach based on estimates of higher-order statistics is also possible. The practical interest in a real data application has been demonstrated.

5. APPENDIX A

The general multidimensional expression for the conditional mean of a scalar random variable *y,* given a vector of random variables $\underline{x} = [x_1, x_2, \ldots x_N]^T$ is,

$$E[y/\underline{x}] = cum(y) + \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\tau_1=1}^{N} \cdots \sum_{\tau_n=1}^{N} cum(y, x_{\tau_1}, \ldots, x_{\tau_n}) \times \times \frac{(-1)^n}{p(x)} \frac{\delta^n p(x)}{\delta x_{\tau_1} \ldots \delta x_{\tau_n}} \quad \text{(A1)}$$

Now, let x and x_p be scalar random variables, then, equation (A1) is simply reduced to,

$$E[x/x_p] = cum(x) - cum(x, x_p) \frac{p'(x_p)}{p(x_p)} + \frac{1}{2!} cum(x, x_p, x_p) \frac{p''(x_p)}{p(x_p)} + \ldots = G(x_p) \quad \text{(A2)}$$

if x_p is Gaussian

$$(-1)^m \frac{p^{(m)}(x_p)}{p(x_p)} = H_m(x_p) \quad \text{(A3)}$$

Then we can write (A2) in the form

$$G(x_p) = \sum_{m=1}^{\infty} \frac{1}{m!} C_m(x, x_p) H_m(x_p) \quad \text{(A4)}$$

where $C_m(x, x_p) = cum\left(x, \overbrace{x_p, \ldots, x_p}^{m \text{ times}}\right)$

In our case, we have to assume that x and x_p are zero-mean, and unit variance , then:

$$\begin{aligned} &\boldsymbol{cum(x, x_p)} = E[xx_p] - E[x]E[x_p] = E[xx_p] = a_1 \\ &\boldsymbol{cum(x, x_p, x_p)} = E[xx_p^2] - 2E[xx_p]E[x_p] - E[x_p^2]E[x] + \\ &+ 2E[x]E[x_p]E[x_p] = E[xx_p^2] = a_2 \\ &\boldsymbol{cum(x, x_p, x_p, x_p)} = E[xx_p^3] - 3E[x_p^2]E[xx_p] - E[x]E[x_p^3] - \\ &- 3E[x_p]E[xx_p^2] + 4E[x]E[x_p]E[x_p^2] + 6E[x_p]^2 E[xx_p] - \\ &- 6E[x]E[x_p]^3 = E[xx_p^3] - 3E[xx_p] = a_3 - 3a_1 \end{aligned} \quad \text{(A5)}$$

with $a_i = E[xx_p^i]$. Therefore the third-order approximation in (A4) is,

$$G_3(x_p) = a_1 H_1(x_p) + \frac{1}{2!} a_2 H_2(x_p) + \frac{1}{3!}(a_3 - 3a_1) H_3(x_p) \quad \text{(A6)}$$

Hermite polynomials can be generated using the recursion equation.

$$H_{m+1}(x_p) = x_p H_m(x_p) - m H_{m-1} \quad \text{with } H_0(x_p) = 1, \text{ and } H_1(x_p) = x \quad \text{(A7)}$$

then,

$$H_1(x_p) = x_p \quad H_2(x_p) = x_p^2 - 1 \quad H_3(x_p) = x_p^3 - 3x_p \quad \text{(A8)}$$

Finally the equation (3) follows

$$\begin{aligned} G_3(x_p) &= a_1 x_p + \frac{1}{2} a_2 (x_p^2 - 1) + \frac{1}{6}(a_3 - 3a_1)(x_p^3 - 3x_p) = \\ &= \frac{1}{6}(a_3 - 3a_1) x_p^3 + \frac{1}{2} a_2 x_p^2 + \left(\frac{5}{2} a_1 - \frac{1}{2} a_3\right) x_p - \frac{1}{2} a_2 \end{aligned} \quad \text{(A9)}$$

6. REFERENCES

[1] Celka, P., Bershad, N.J., Vesin, J-M, "Stochastic gradient identification of polynomial Wiener systems: analysis and application", *IEEE Trans. on Signal Processing*, 2001, 49 (2), pp.301-313.

[2] Nordsjö, A.E., l., Zetterberg, L.H., "Identification of certain time-varying nonlinear Wiener and Hammerstein Systems", *IEEE Trans. on Signal Processing*, 2001, 49 (3), pp.577-592.

[3] Bershad, N.J., Celka, P., McLaughlin, S., "Analysis of stochastic gradient identification of Wiener-Hammerstein Systems for nonlinearities with Hermite polynomial expansions", *IEEE Trans. on Signal Processing*, 2001, 49 (5), pp.1060-1072.

[4] Taleb, A., Sole, J., Jutten, CH., "Quasi-nonparametric blind inversion of Wiener systems", *IEEE Trans. on Signal Processing*, 2001, 49 (5), pp.917-924.

[5] Vergara, L., Bernabeu, P., "Simple approach to nonlinear prediction", *Electron. Lett.*, 2001, 37 (14), pp. 926-928.

[6] Ertugrul Celebi, M., "General formula for conditional mean using higher order statistics", *Electron. Lett.*, 1997, 33 (25), pp. 2097-2099

[7] Vergara, L., Bernabeu, P., "Automatic signal detection applied to fire control by infrared digital signal processing", *Signal Processing*, 2000, 80 (4), pp.659-669.

7. ACKNOWLEDGEMENT

This work has been partially by CICYT through grant No. DPI2000-0619-03-01.