

NONLINEAR ACOUSTIC ECHO CANCELLATION USING ORTHOGONAL POLYNOMIAL

Guan-Yu Jiang and Shih-Fu Hsieh*¹

Department of Communication Engineering

National Chiao Tung University, Hsinchu, Taiwan 300, Republic of China

Tel: 886-3-5731974, E-mail: sjsmi63@yahoo.com.tw, sfhsieh@mail.nctu.edu.tw

ABSTRACT

In order to compensate the nonlinear distortion in the hands-free telephones or teleconferencing system, a memoryless power-series-based polynomial NLMS adaptive filter can be used to cancel nonlinear acoustic echo. Conventional polynomial model employs a power-series expansion. In this paper we propose an orthogonal polynomial adaptive filter to accelerate the convergence rate of the nonlinear adaptive filter. The convergence rates of residual echo power for both power-series and orthogonal polynomials are derived analytically. Computer simulations justify our analysis and show the improved performance of the proposed nonlinear acoustic echo canceller.

1. INTRODUCTION

Hands-free telephone or teleconferencing usually suffers from the annoying acoustic echo problem. A linear adaptive filter is commonly used for acoustic echo cancellation(AEC). However, overdriven the power amplifier of loudspeaker will incur nonlinear distortion. The nonlinear distortion will limit the echo cancellation performance and the conventional linear acoustic echo canceller(AEC) cannot cope with this kind of distortion. Recently, several nonlinear AEC structures have been proposed to compensate this kind of distortion [1]. The cascaded nonlinear AEC structure has fewer coefficients than the Volterra filter [2] and has less computational complexity, if it is updated by the NLMS algorithm. The nonlinear AEC is shown in Fig.1. The power amplifier of loudspeaker is modeled by the nonlinear processor and the echo path is modeled by the linear filter. The power series polynomial cascade FIR is simple to implement but its high correlation among different polynomial orders leads to low convergence rate. In [3] an adaptive orthogonalized power filter is proposed to improve convergence rate. The orthogonal basis is updated online in each iteration and the Gram Schmidt procedure is employed to find out the orthogonalization coefficients, as a result, computational complexity is increased. In this paper we use a fixed orthogonal polynomial to produce the nonlinear components.

With the assumption of a perfect linear filter, convergence analysis of the residual echo power is performed. Its low computational complexity and fast convergence rate makes the orthogonal polynomial filter very promising for nonlinear AEC.

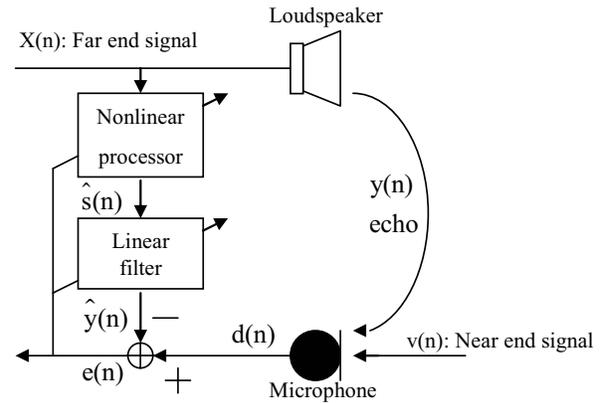


Fig.1 Nonlinear acoustic echo canceller

2. STRUCTURE OF THE NONLINEAR AEC

As shown in Fig 1, the signal from the far end is assumed to be nonlinearly distorted only in the power amplifier of loudspeaker. It is then passing through a room impulse response $\mathbf{h}[n]$. Let $d(n)$ denote the desired signal. The AEC output signal $\hat{y}(n)$ can be written as

$$\hat{y}(n) = \hat{\mathbf{h}}^T(n) \hat{\mathbf{s}}(n)$$

$$\hat{\mathbf{h}}(n) = [\hat{h}_0(n), \hat{h}_1(n), \hat{h}_2(n), \dots, \hat{h}_{L-1}(n)]^T$$

$$\hat{\mathbf{s}}(n) = [\hat{s}_0(n), \hat{s}_1(n), \dots, \hat{s}_{L-1}(n)]^T$$

$$= \begin{bmatrix} p_0(n) & p_1(n) & \cdots & p_N(n) \\ p_0(n-1) & p_1(n-1) & \cdots & p_N(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ p_0(n-L+1) & p_1(n-L+1) & \cdots & p_N(n-L+1) \end{bmatrix} \begin{bmatrix} \hat{a}_0(n) \\ \hat{a}_1(n) \\ \vdots \\ \hat{a}_N(n) \end{bmatrix}$$

$$= \mathbf{P}(n) \hat{\mathbf{a}}(n)$$

¹ This work was supported by the NSC and MediaTek.

where $\hat{\mathbf{h}}(n)$ represents the estimated coefficient vector of the linear filter, $\hat{\mathbf{s}}(n)$ is the output vector of nonlinear filter, p_i is the polynomial basis of order i , N is the order of the polynomial, and $\hat{\mathbf{a}}(n)$ is the estimated coefficient vector of the nonlinear filter. The estimated error is $e(n)=d(n)-\hat{y}(n)$.

The gradient of $e^2(n)$, as derived for linear transversal filter in [3] can be calculated according to:

$$\nabla_h = \frac{\partial e^2(n)}{\partial \hat{\mathbf{h}}(n)} = -2e(n)\hat{\mathbf{s}}(n)$$

$$\nabla_a = \frac{\partial e^2(n)}{\partial \hat{\mathbf{a}}(n)} = -2e(n)\mathbf{P}^T(n)\hat{\mathbf{h}}(n)$$

If the coefficient vectors are updated with step size μ_h and μ_a , a NLMS-type adaptive algorithm is given as follows:

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \frac{\mu_h}{\|\hat{\mathbf{s}}(n)\|_2^2} \hat{\mathbf{s}}(n)e(n) \quad (1)$$

$$\hat{\mathbf{a}}(n+1) = \hat{\mathbf{a}}(n) + \frac{\mu_a}{\|\mathbf{P}^T(n)\hat{\mathbf{h}}(n)\|_2^2 + \delta} \mathbf{P}^T(n)\hat{\mathbf{h}}(n)e(n) \quad (2)$$

At each iteration, the echo signal $e(n)$ is the same for coefficient updating in both (1) and (2) which form the general formula of a joint NLMS-type adaptation of both stages.

3. RESIDUAL ECHO POWER ANALYSIS

For simplicity, we assume the nonlinear loudspeaker and linear room impulse response are time invariant. The near end signal $v(n)$ only contains a white Gaussian noise (WGN), double talk is not present. Here, in order to distinguish convergence rate of orthogonal and power series polynomials we will derive the convergence rate under the assumption of perfect linear coefficient i.e., $\hat{\mathbf{h}}(n) = \mathbf{h}$. Hence the estimation error produced by the nonlinear AEC filter is expressed as

$$\begin{aligned} e(n) &= d(n) - \hat{\mathbf{a}}^T(n)\mathbf{P}^T(n)\mathbf{h} \\ &= v(n) + \mathbf{a}^T\mathbf{P}^T(n)\mathbf{h} - \hat{\mathbf{a}}^T(n)\mathbf{P}^T(n)\mathbf{h} \end{aligned}$$

3.1. Analysis of nonlinear coefficient

We denote the nonlinear coefficient weight error [4] by

$$\boldsymbol{\varepsilon}_a(n+1) = \mathbf{a} - \hat{\mathbf{a}}(n+1)$$

and using (2) we may rewrite $\boldsymbol{\varepsilon}_a(n+1)$ as

$$\boldsymbol{\varepsilon}_a(n+1) = \left[\mathbf{I} - \frac{\mu_a \mathbf{P}^T(n)\mathbf{h}\mathbf{h}^T\mathbf{P}(n)}{\mathbf{T}} \right] \boldsymbol{\varepsilon}_a(n) - \frac{\mu_a \mathbf{P}^T(n)\mathbf{h}}{\mathbf{T}} v(n),$$

where $\mathbf{T} = \|\mathbf{P}^T(n)\mathbf{h}\|_2^2 + \delta$.

According to the direct averaging method, when $\mu_a \ll 1$

$$\boldsymbol{\varepsilon}_a(n+1) \approx \left[\mathbf{I} - \frac{\mu_a \mathbf{R}_{\mathbf{P}^T(n)\mathbf{h}}}{\mathbf{T}} \right] \boldsymbol{\varepsilon}_a(n) + \mathbf{f}(n) \quad (3)$$

where $\mathbf{f}(n) = -\frac{\mu_a \mathbf{P}^T(n)\mathbf{h}}{\mathbf{T}} v(n)$ and $\mathbf{R}_{\mathbf{P}^T(n)\mathbf{h}}$ is the correlation matrix of $\mathbf{P}^T(n)\mathbf{h}$. By applying the unitary similarity transformation, $\mathbf{R}_{\mathbf{P}^T(n)\mathbf{h}}$ is transformed into a simpler form:

$\mathbf{Q}^T \mathbf{R}_{\mathbf{P}^T(n)\mathbf{h}} \mathbf{Q} = \mathbf{D}$, where \mathbf{Q} is an unitary matrix and \mathbf{D} is a diagonal matrix consisting of the eigenvalues λ_i . Let $\mathbf{K}(n) = \mathbf{Q}^T \boldsymbol{\varepsilon}_a(n)$ then we may transform (3) into the form

$$\mathbf{K}(n+1) = \left[\mathbf{I} - \frac{\mu_a}{\mathbf{T}} \mathbf{D} \right] \mathbf{K}(n) + \boldsymbol{\Phi}(n),$$

where $\boldsymbol{\Phi}(n) = \mathbf{Q}^T \mathbf{f}(n)$.

Assuming the initial value $K_i(0)$ of the i -th entry of $\mathbf{K}(n)$ is independent of Φ_i

$$\begin{aligned} E[|K_i(n)|^2] &= (1 - \mu \frac{\lambda_i}{\mathbf{T}})^{2n} |K_i(0)|^2 \\ &+ \sum_{s=0}^{n-1} \sum_{j=0}^{n-1} (1 - \mu \frac{\lambda_i}{\mathbf{T}})^{n-1-s} (1 - \mu \frac{\lambda_i}{\mathbf{T}})^{n-1-j} E[\Phi_i(s)\Phi_i(j)] \\ &= \frac{\mu}{\mathbf{T}} \frac{\sigma_v^2}{2 - \frac{\mu}{\mathbf{T}} \lambda_i} + [|K_i(0)|^2 - \frac{\mu}{\mathbf{T}} \frac{\sigma_v^2}{2 - \frac{\mu}{\mathbf{T}} \lambda_i}] (1 - \mu \frac{\lambda_i}{\mathbf{T}})^{2n} \end{aligned} \quad (4)$$

3.2. Analysis of residual echo power

The mean square error (i.e., residual echo) is given by

$$\begin{aligned} J(n) &= E[|e(n)|^2] \\ &= \sigma_v^2 + E\left[\boldsymbol{\varepsilon}_a^T(n) \mathbf{P}^T(n) \mathbf{h} \mathbf{h}^T \mathbf{P}(n) \boldsymbol{\varepsilon}_a(n) \right] \end{aligned} \quad (5)$$

Assume the variation of $\boldsymbol{\varepsilon}_a(n)$ is slow compared with that of $\mathbf{P}^T(n)\mathbf{h}$, hence

$$\begin{aligned} E\left[\boldsymbol{\varepsilon}_a^T(n) \mathbf{P}^T(n) \mathbf{h} \mathbf{h}^T \mathbf{P}(n) \boldsymbol{\varepsilon}_a(n) \right] &\approx E\left[\boldsymbol{\varepsilon}_a^T(n) E\left[\mathbf{P}^T(n) \mathbf{h} \mathbf{h}^T \mathbf{P}(n) \right] \boldsymbol{\varepsilon}_a(n) \right] \\ &= E\left[\boldsymbol{\varepsilon}_a^T(n) \mathbf{R}_{\mathbf{P}^T(n)\mathbf{h}} \boldsymbol{\varepsilon}_a(n) \right] \\ &= E\left[\mathbf{K}^T(n) \mathbf{Q}^T \mathbf{R}_{\mathbf{P}^T(n)\mathbf{h}} \mathbf{Q} \mathbf{K}(n) \right] \\ &= E\left[\mathbf{K}^T(n) \mathbf{D} \mathbf{K}(n) \right] \\ &= \sum_{i=0}^N \lambda_i E\left[|K_i(n)|^2 \right] \end{aligned} \quad (6)$$

From (4) and (6), the mean square error can be written as

$$J(n) = \sigma_v^2 + \frac{\mu_a}{T} \sigma_v^2 \sum_{i=0}^N \frac{\lambda_i}{2 - \frac{\mu_a}{T} \lambda_i} + \sum_{i=0}^N \lambda_i \left[|K_i(0)|^2 - \frac{\frac{\mu_a}{T} \sigma_v^2}{2 - \frac{\mu_a}{T} \lambda_i} \right] \left(1 - \frac{\mu_a}{T} \lambda_i \right)^{2n} \quad (7)$$

3.3. Eigenvalues and the basis of nonlinear component

According to (7) we have known that the convergence rate depends on the eigenvalues λ_i , and the smallest eigenvalue dominates the convergence rate. Next we will show that a smaller eigenvalue spread λ_{spread} (faster convergence) can be achieved by reducing the correlation among the bases. For simplicity, we assume the nonlinear function is an odd function and only contains first and third orders. We

have $\mathbf{R}_{\mathbf{p}^T(n)\mathbf{h}} = \begin{bmatrix} r_{11} & c \\ c & r_{33} \end{bmatrix}$, where c is the correlation between

p_1 and p_3 and the $\lambda_{spread} = \frac{1 + \sqrt{1 - 4(r_{11}r_{33} - c^2)/(r_{11} + r_{33})^2}}{1 - \sqrt{1 - 4(r_{11}r_{33} - c^2)/(r_{11} + r_{33})^2}}$. When

input is uniformly distributed and $\|\mathbf{h}\|_2^2 = 1$, λ_{spread} is 27.78

and $\frac{\lambda_{min}}{\lambda_{max} + \lambda_{min}}$ is 0.03481 for power series polynomial

and λ_{spread} is 14.55 and $\frac{\lambda_{min}}{\lambda_{max} + \lambda_{min}}$ is 0.06417 for

orthogonal polynomial. λ_{spread} of the orthogonal polynomial is about one half of the power series. It is easy to see, the smallest eigenvalue dominates the convergence rate, i.e., the smaller λ_{min} , the slower convergence rate. When c is equal to zero we will have the smallest eigenvalue spread. It means that when p_1 is orthogonal to p_3 , the fastest convergence rate can be attained.

4. SIMULATION RESULTS

To evaluate the performance of the orthogonal polynomial we compare the simulated and theoretical curves. In the following simulation we let the step size $\mu_h = \mu_a = 0.05$, $\delta = 1$, SNR=26 dB, the length of the room impulse response is set to be 128, which is identical to number of taps of the linear filter, the nonlinear filter order is 3. In the first experiment we let the input signal be uniformly distributed and the orthogonal polynomial series can be generated using Gram Schmidt orthonormalization in the interval $(-1,1)$ with

the weighting function 1. The first 3 order orthogonal basis are given as follows:

$$p_0(x) = 1, p_1(x) = x, p_2(x) = x^2 - \frac{1}{3}, p_3(x) = x^3 - \frac{3}{5}x \quad (8)$$

As shown in Fig 2, the theoretical curves are plotted from (7) the simulation results agree well with the theoretical curves. The orthogonal polynomial AEC indeed converges faster due to its smaller eigenvalue spread.

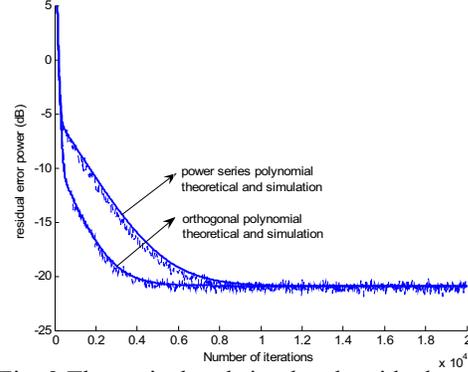


Fig. 2 Theoretical and simulated residual error power curves under the assumption of perfect linear coefficient for uniform input.

Without the assumption of perfect linear coefficients, the simulation results are shown in Fig 3. Because the linear and nonlinear coefficients errors affect each other in the cascade structure, it is difficult to perform the joint error analysis theoretically. In Fig 3, when the nonlinear coefficients have faster convergence, the joint error will also have faster convergence. Hence, the overall performance of nonlinear AEC with orthogonal polynomial is better than that with power series.

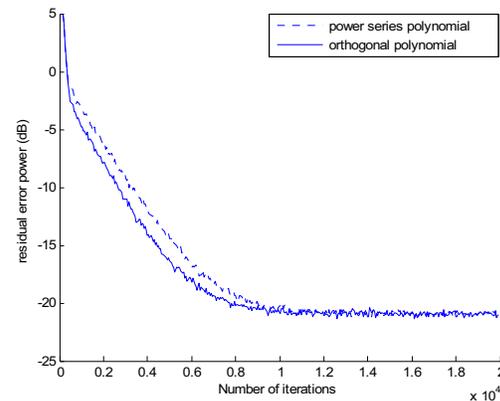


Fig. 3 Simulated residual error power curves without assumption of perfect linear coefficient for uniform input.

Next, we let the input signal be zero mean WGN. The parameters are the same as the first experiment except for the orthogonalized polynomial basis. When the input signal is WGN, the orthogonal basis also can be found by Gram

Schmidt orthonormalization in the interval (-1,1) with the Gaussian weighting function. The orthogonal polynomial basis of order 3 can be shown as follows

$$p_0(x) = 1, p_1(x) = x, p_2(x) = x^2 - \sigma_x^2, p_3(x) = x^3 - 3\sigma_x^2 x$$

We can find the basis is dependent on the input signal. However, in practical implementation, σ_x^2 can be obtained by first order recursion procedure described by

$$E[x^p(n+1)] = \gamma E[x^p(n)] + (1-\gamma)x^p(n),$$

where γ is the forgetting factor. Although the basis is input-dependent it needs to calculate the moment of input signal at each iteration, the computational complexity is slightly increased. The simulation results are shown in Fig 4.

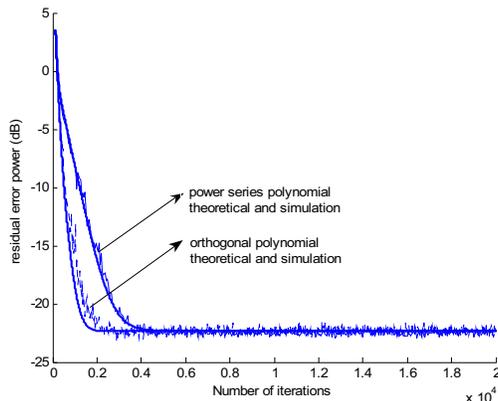


Fig. 4 Theoretical and simulated residual error power curves under the assumption of perfect linear coefficient for WGN input

The joint coefficients errors are also difficult to analyze when the input signal is WGN. In Fig 5 we only present the simulation results. According to the first two experiments, in either case of a uniform or Gaussian input, the nonlinear AEC has better performance when its nonlinear coefficients have faster convergence rate.

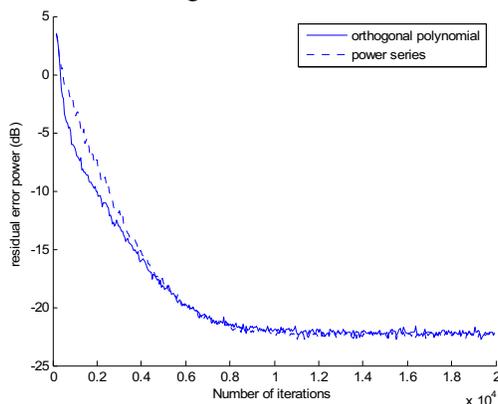


Fig. 5 Simulated residual error power curves without assumption of perfect linear coefficient for WGN input.

In the last simulation we compare the performance between the orthogonal polynomial in (8) and the power series

polynomial when the input is a real speech signal. The probability density function of speech signal is neither uniform nor Gaussian, therefore (8) may not have perfect orthogonality but it still has better performance than the power series polynomial as shown in Fig 6. The performance is measured by ERLE, defined as

$$ERLE(\text{dB}) = 10 \log_{10} \frac{E[d^2(n)]}{E[e^2(n)]}.$$

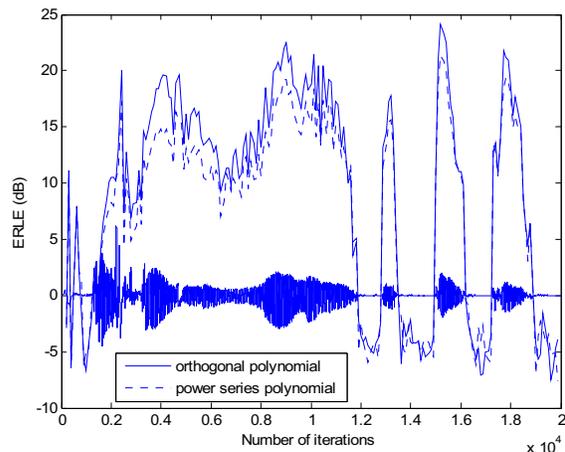


Fig. 6 ERLE comparison between orthogonal and power series polynomial under the assumption of perfect linear coefficients for speech input.

5. SUMMARY

In this paper we have presented the orthogonal basis polynomial for nonlinear AEC. For both uniform and Gaussian inputs the simulation results agree well with theoretical curves. Convergence rate analysis indicates that a smaller eigenvalue spread is closely related to the correlation between polynomial bases. The proposed orthogonal basis has better performance than conventional power basis without increasing computational complexity.

6. REFERENCES

- [1] J. P. Costa, A. Lagrange and A. Arliaud, "Acoustic Echo Cancellation Using Nonlinear Cascade Filters," *ICASSP*, vol. 5, pp.389-392, April 2003.
- [2] A. Stenger, L. Trautmann and R. Rabenstein, "Nonlinear Acoustic Echo Cancellation with 2nd Order Adaptive Volterra Filters," *ICASSP Proceeding*, vol. 2, pp. 877-880, Nov. 1999.
- [3] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear Acoustic Echo Cancellation Using Adaptive Orthogonalized Power Filters," *ICASSP*, vol.3, pp. 105-108, March 2005.
- [4] S. Haykin: *Adaptive Filter Theory*, 4th ed., Prentice-Hall, 2002.
- [5] A. Stenger and W. Kellermann, "Nonlinear Acoustic Echo Cancellation with Fast Converging Memoryless Preprocessor," *ICASSP*, vol.2, pp.805-808, June 2000.