CONJUGATE GRADIENT TECHNIQUES FOR MULTICHANNEL ACOUSTIC ECHO CANCELLATION

Lino García

Ingeniería Eléctrica y Electrónica
ESP-UEM, Madrid, Spain
lino.garcia@uem.es

S. Torres-Guijarro, J.A. Beracoechea, F.J. Casajús-Quirós

Señales, Sistemas y Radiocomunicaciones
ETSI Telecomunicación-UPM, Madrid, Spain
{marisol},{beraco},{javier}@gaps.ssr.upm.es

ABSTRACT

Conjugate Gradient (CG) techniques are suitable for resolution of time-variant system identification problems: adaptive equalization, echo cancellation, active noise cancellation, linear prediction, etc. These systems can be seen as optimization problems and CG techniques can be used to solve them. It has been demonstrated that, in the single-channel case, the conjugate gradient techniques provide a similar solution in terms of convergence rate than those provided by the recursive least square (RLS) method, involving higher complexity than the least mean square (LMS) but lower than RLS without stability issues. The advantages of these techniques are especially valuable in the case of high complexity and magnitude problems like multi-channel systems. This work develops CG algorithm for the adaptive MIMO (multiple-input and multiple-output) systems and tests it by solving a multichannel acoustic echo cancellation (MAEC) problem.

1. INTRODUCTION

CG techniques belong to an optimization method family which resolves a purely quadratic problem

$$\mathrm{f}(\mathbf{w}) = \mathbf{w}^T\mathbf{R}\mathbf{w} + \mathbf{r}^T\mathbf{w} + a \tag{1}$$

where $\mathbf{R}$ is a non-negative defined matrix, $\mathbf{r}$ and $\mathbf{w}$ are vectors and a is a scalar[1]. The CG method is a very efficient and well proven optimization technique to find optimum solution for the equation (1). Quadratic problem has a unique solution which is the unique solution of the linear equation [1]

$$\mathbf{R}\mathbf{w}^{\bullet} = \mathbf{r} \tag{2}$$

In section 2 the normal equation for a multichannel adaptive system is given. In section 3 a CG method for a multichannel adaptive system is developed. In section 4 an efficient solution for the problem is presented based on subband decomposition and application of multirate open-loop delay-less techniques.

[1] Matrices are represented by a bold capital letter, vectors are represented by a bold lower case letter and scalars by italic lower case letter. The super indexes ^{T} and ^{H} represent transposition and conjugate transposition (hermitic) of a vector or a matrix respectively and $*$ denote conjugation.

2. MULTICHANNEL ADAPTIVE FILTERING

The case of adaptive multichannel filtering, showed in Figure 1, is in general structurally more difficult than the single-channel case [2], although it might be defined in terms of the so called *normal equation* (2). This equation proves to evaluate the gradient in the solution $\mathbf{w}^{\bullet}$, applying the criteria of mean square error minimization (3).

$$J = E\{ee^*\} = E\{|e|^2\} \tag{3}$$

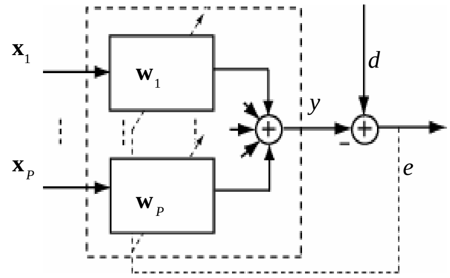

Figure 1: *Multichannel Adaptive Filtering*

The P multichannel filter output error is defined by

$$e = d - y = d - \sum_{p=1}^{P} \mathbf{w}_p^H \mathbf{x}_p = d - \mathbf{w}^H\mathbf{x} \tag{4}$$

where $\mathbf{x}=\begin{bmatrix}\mathbf{x}_1^T & \mathbf{x}_2^T & \cdots & \mathbf{x}_P^T\end{bmatrix}^T$ is a column vector of $L\times P$ dimension corresponding to the multichannel system inputs; $\mathbf{x}_p=\begin{bmatrix}x(n) & x(n-1) & \cdots & x(n-L+1)\end{bmatrix}^T$ is a column vector of L dimension, equal to the length of the corresponding filter at the input of the p-th channel; $\mathbf{w} = \begin{bmatrix}\mathbf{w}_1^T & \mathbf{w}_2^T & \cdots & \mathbf{w}_P^T\end{bmatrix}^T$ is a column vector of $L\times P$ dimension, corresponding to the multichannel adaptive system coefficients and $\mathbf{w}_p = \begin{bmatrix}w_1 & w_2 & \cdots & w_L\end{bmatrix}^T$ is a L length vector corresponding to the p-th channel adaptive filter; d, y and e are scalars that denoted the desired signal, the adaptive multichannel filter output signal and the error signal, respectively. Although the time variable is only referred explicitly when it helps comprehension, each sample n corresponds to a time instant (sampling period). Combining (4) in (3)

$$J = E\{e^2\} = E\{|d - \mathbf{w}^H\mathbf{x}|^2\} \tag{5}$$

$$= E\{d^2\} - 2\mathbf{w}^H E\{\mathbf{x}d^*\} + \mathbf{w}^H E\{\mathbf{x}\mathbf{x}^H\}\mathbf{w}$$

$$\mathbf{g} = \nabla J = -2[\mathbf{r} - \mathbf{R}\mathbf{w}]$$

where $\mathbf{R}$ is the $LP \times LP$ correlation matrix defined by

$$\mathbf{R} = E\{\mathbf{x}\mathbf{x}^H\} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1P} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{P1} & \mathbf{R}_{P2} & \cdots & \mathbf{R}_{PP} \end{bmatrix} \tag{6}$$

and $\mathbf{r}$ is the $LP \times 1$ cross-correlation vector.

$$\mathbf{r} = E\{\mathbf{x}d^*\} \tag{7}$$

Notice that $\mathbf{R}_{ij} = E\{\mathbf{x}_i\mathbf{x}_j^H\}$. The length of the gradient vector $\mathbf{g}$ defined in (5) is $LP \times 1$.

3. CONJUGATE GRADIENT

The stochastic conjugate gradient techniques for single-channel adaptive filtering have already been presented in [4], where a block stochastic gradient estimation method that averages the last N instantaneous estimated gradients, is described. In this work a new approach is proposed that appropriately rearranges the $\mathbf{x}_p$ vectors in (6) in a $L \times N$ order matrix and expands the d scalar in (7) in a $N \times 1$ vector.

$$\mathbf{X}_p(n) = \begin{bmatrix} x_p(n-N+1) & x_p(n-(N-1)+1) & \cdots & x_p(n) \\ x_p(n-N) & x_p(n-(N-1)) & \cdots & x_p(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ x_p(n-N-L+2) & x_p(n-(N-1)-L+2) & \cdots & x_p(n-L+1) \end{bmatrix}$$

$$\mathbf{d}(n) = \begin{bmatrix} d(n-N+1) & d(n-(N-1)+1) & \cdots & d(n) \end{bmatrix}^T$$

This way the correlation matrix estimation and the cross-correlation vector slightly change to

$$\mathbf{R} = E\{\mathbf{X}\mathbf{X}^H\}/N \tag{8}$$

$$\mathbf{r} = E\{\mathbf{X}\mathbf{d}^*\}/N \tag{9}$$

and $\mathbf{R}_{ij} = E\{\mathbf{X}_i\mathbf{X}_j^H\}/N$. Note that $\mathbf{R}$ matrix and $\mathbf{r}$ vector keep their dimensions $LP \times LP$ and $LP \times 1$, respectively. This method is also known as *sliding data window* and allows the following exponentially decreasing treatment

$$\mathbf{R}_k = \lambda\mathbf{R}_{k-1} + \mathbf{X}_k\mathbf{X}_k^H \tag{10}$$

$$\mathbf{r}_k = \lambda\mathbf{r}_{k-1} + \mathbf{X}\mathbf{d}^* \tag{11}$$

3.1. Conjugate directions

Two vectors $\mathbf{v}_i$ and $\mathbf{v}_j$ are $\mathbf{R}$-*orthogonal*, or *conjugate with respect to* $\mathbf{R}$, if $\mathbf{v}_i^H\mathbf{R}\mathbf{v}_j = 0,\ \forall i \neq j$. $\mathbf{R}$ induces an internal product $\langle\mathbf{v},\mathbf{w}\rangle_{\mathbf{R}} = \mathbf{v}^H\mathbf{R}\mathbf{w}$ and norm $\|\mathbf{v}\|_{\mathbf{R}} = \sqrt{\mathbf{v}^H\mathbf{R}\mathbf{v}}$.

The conjugate direction methods [1] provide the solution for a normal equation (2) in terms of a set of L linearly independent direction vectors which are $\mathbf{R}$-orthogonal or conjugate with respect to $\mathbf{R}$.

$$\mathbf{w}^\bullet = \alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k + \cdots + \alpha_L\mathbf{v}_L = \sum_{k=1}^{L}\alpha_k\mathbf{v}_k \tag{12}$$

According to the conjugate direction theorem, the $\{\mathbf{w}_k\}$ succession generated by

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha_k\mathbf{v}_k \tag{13}$$

with

$$\alpha_k = -\frac{\mathbf{g}_k^H\mathbf{v}}{\mathbf{v}_k^H\mathbf{R}\mathbf{v}_k} \tag{14}$$

converges to the unique solution $\mathbf{w}^\bullet$ in $\min\{L,N\}$ steps and, according to span subspace theorem, minimizes the expression (1). The CG method is a conjugate direction method which is obtained by choosing the successive direction vector as a conjugate version of the successive gradient obtained as a method progresses. The following expressions provide a method which can be extended to nonquadratic problems. Starting at any $\mathbf{w}_0$:

$$\mathbf{g}_0 = -[\mathbf{r} - \mathbf{R}\mathbf{w}]$$

$$\mathbf{v}_0 = -\mathbf{g}_0$$

for $k = 1, 2, \ldots, L$

$$\alpha_k = \frac{\|\mathbf{g}_k\|^2}{\|\mathbf{v}_k\|_{\mathbf{R}}^2} = \frac{\|\mathbf{g}_k\|^2}{\mathbf{v}_k^H\mathbf{R}\mathbf{v}_k}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \alpha_k\mathbf{v}_k$$

$$\mathbf{g}_{k+1} = -[\mathbf{r} - \mathbf{R}\mathbf{w}_{k+1}] = \mathbf{g}_k + \alpha_k\mathbf{R}\mathbf{v}_k$$

$$\beta_k = \frac{\|\mathbf{g}_{k+1}\|^2}{\|\mathbf{g}_k\|^2}$$

$$\mathbf{v}_{k+1} = -\mathbf{g}_{k+1} + \beta_k\mathbf{v}_k$$

end

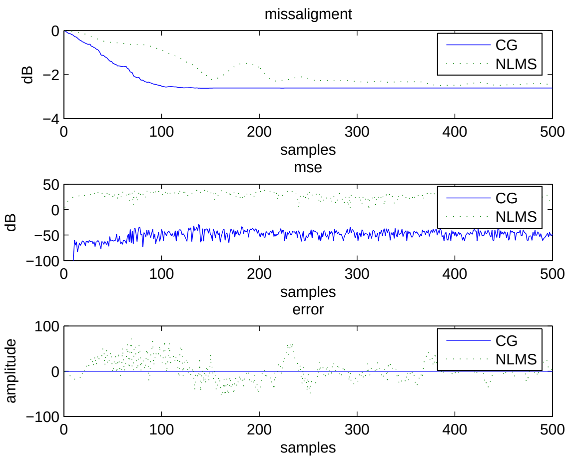

Figure 2: *CG vs NLMS, P = 4*

In [4] an analysis on the window length selection, N, and its impact on the convergence performance is provided. For $N=\sqrt{L}$, results in terms of convergence are similar to RLS.

Figure 2 provides a comparison on the performance of the CG and the NLMS algorithms. The fast convergence rate and low level error of the CG should be noted. The input signals to the system $\mathbf{x}_p = \mathbf{h}_p^H \mathbf{s}$, were generated by convolution of a complex (real part ± 1) zero-mean random sequence $\mathbf{s}$ with $P=4$ filters, which simulate the distortion of the P channels with a raised cosine impulse response $\mathbf{h}_p(n) = 0.5\left[1+\cos\left(2\pi(n-2)/W\right)\right]$, n=1..1000, $W = 2.9$, where W controls the channel's amplitude distortion. A higher W increases distortion and raises $\mathbf{R}$ correlation matrix eigen-values spread in the multichannel equalizer input. Additionally the channel is corrupted by a sequence of Gaussian white noise yielding –30 dB SNR. The length of all filters is L = 64. The adaptation step for the NLMS is $\mu = 0.007$ and the size of the sliding window for the CG is $M = 2L$.

4. MULTIRATE MULTICHANNEL ADAPTIVE FILTERING

In case of working with very long impulse responses on a multichannel system, e.g., acoustic echo cancellation, the output system delay and the computational effort may make adaptive filtering unfeasible. To overcome this issue a number of methods have been developed based on subband decomposition of the system input signals which allow reducing the complexity of the general system by a factor that approximately equals the order of the number of subbands.

Figure 3 shows the scheme for *delayless* open-loop multichannel adaptive filtering, which adapts in subbands and filters in fullband. The fullband filter coefficients for each channel are obtained through the transformation of the weights in subbands following expression (15). The subband decomposition is performed by an oversampled GDFT (Generalized Discrete Fourier Transform) polyphase filterbank. Only M of the $2M$ subband filters are required to retain signals information.

Note that, the fact of reconstructing the error in fullband rather than in subbands, avoids the use of synthesis subband filters (although it is necessary in (15)), eliminating the associate delay. The fullband filtering is performed partitioning the filters in each channel $\mathbf{w}_p$ in Q segments of equal length. The first segment is processed by direct convolution (which makes the scheme delayless) and the rest of the segments are processed by fast convolutions employing FFTs and inverse FFTs.

In [3] a method for a T transformation is proposed capable of generating the full filter from the M subbands

$$\mathbf{w}_p = real\left\{\sum_{m=1}^{M}\left[\mathbf{h}_m\big|_{\downarrow K} * \mathbf{c}_{mp}\right]_{\uparrow K} * \mathbf{f}_m\right\} \qquad (15)$$

The adaptive filters $\mathbf{c}_{mp}$ in each subband are of order $C = \lceil L+R-1/K \rceil - \lceil R/K \rceil + 1$. L is the length of $\mathbf{w}_p$, R is the length of $\mathbf{h}_m$ and K is the decimate factor. The decomposition developed in [3] reconstructs the fullband filter by inserting a diagonal matrix composed by M $\mathbf{c}_{mp}$ vectors between the analysis and synthesis banks.

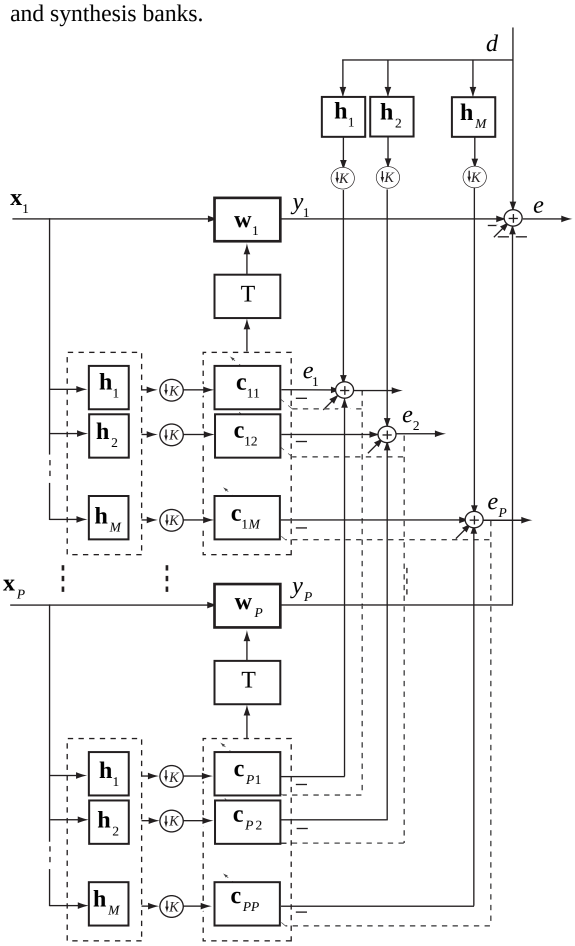

Figure 3: *Subband Adaptive Filtering with Fullband*

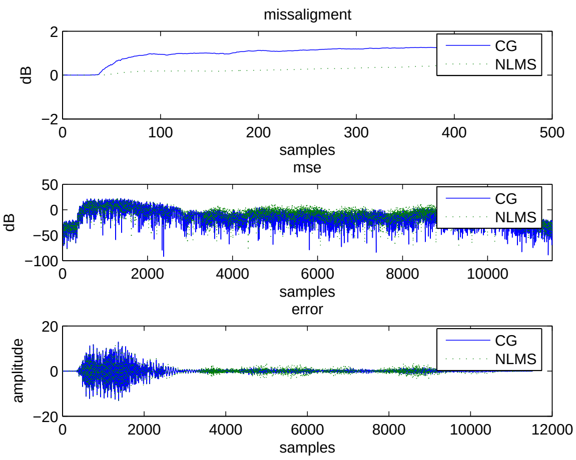

Figure 4: *CG vs NLMS Subbands, P = 2*

This configuration has the advantage of being able to adapt with different parameters, and moreover, different filtering algorithms in each subband. shows the results for the stereophonic acoustic echo cancellation application illustrated in Figure 5.

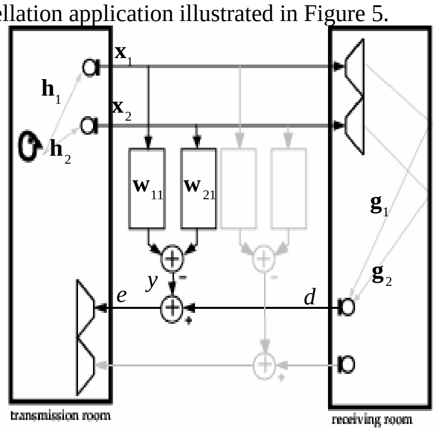

Figure 5: *Multichannel Echo Cancellation, P = 2*

Both rooms in the figure have the same dimensions [3 3 3][2]. The source **s** is located in [.75 1 1] position and the microphones in [1.25 1.45 1] and [1.25 1.55 1]. Notice that proximity between them produces strong correlation between $\mathbf{x}_1$ and $\mathbf{x}_2$ signals. The speakers, in the receiving room, are located in [2.5 1 1] and [2.5 2 1] and the microphone, of the simulated channel, in [1.25 1.45 1]. The wall reflection coefficients for both rooms are 0.9, 0.9, 0.9, 0.9, 0.7 y 0.7. The elevation, azimuth and beam semi-aperture of the speakers radiation pattern are 45º, 45º and 60º respectively. The corresponding speaker's parameters are 45º, 135º and 80º. The length of each of the four filters is 1056 samples and the sampling frequency is 8 kHz. 32 subbands have been employed from which only half are necessary, with a decimate factor of 24 samples and a 768 length prototype filter. The GDFT has been implemented with polyphase filtering for an efficient implementation. Results for the other channel can be obtained in the same way. The bottom of illustrates the low error level that returns in the form of echo when applying GC versus that obtained with NLMS.

[2] [x y z] coordinates in meters

5. CONCLUSIONS

A novel scheme for multichannel adaptive filtering based on complex subband adaptation has been proposed and proved in a MAEC case. In this scheme the polyphase filtering techniques and partitioning of the fullband response have been applied to reduce the computational requirements while keeping the system delayless. This architecture provides independent adjustment of the algorithm parameters on each subband. Moreover, each subband may work with different algorithms. A multichannel stochastic CG algorithm has been proposed and its performance has been shown to be much better than LMS family algorithms. This approach requires a high computational effort, although it does not require dense matrix operations, but offers a high parallel operation capacity and good performance which make the technique suitable for employment on real time applications.

6. REFERENCES

[1] D. G. Luenberger, *Introduction to Linear and Nonlinear Programing*, MA: Addison-Wesley, Reading, Mass., 1984.

[2] H. Buchner, J. Benesty, and W. Kellermann, Multichannel Frequency-Domain Adaptive with Application to Multichannel Acoustic Echo Cancellation, Adaptive Signal Processing. Application to Real World Problems, Chapter 4, Springer, 2003.

[3] J. P. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand, The Complex Subband Decomposition and its Application to the Decimation of Large Adaptive Filtering Problems, IEEE Trans. on Signal Processing, Vol. 50, No. 11, November 2002.

[4] G. K. Boray and M. D. Srinath, Conjugate Gradient Techniques for Adaptive Filtering, IEEE Trans. on Circuits and Systems –I: Fundamental Theory and Application, Vol. 39, No. 1, January 1992.