ADAPTIVE FILTERING IN SUBBANDS USING A WEIGHTED CRITERION

Marc de Courville and Pierre Duhamel

Télécom-Paris, Département Signal
46, rue Barrault 75634 Paris Cedex 13, France
Tel: (+33) 1 45.81.73.65 Fax: (+33) 1 45.88.79.35 E-mail: courvill@sig.enst.fr

ABSTRACT

Classical transform-domain algorithms adapt the filter coefficients (in each "frequency bin") by minimizing a criterion depending on a full-band time-domain error. This paper proposes an algorithm which updates each portion of the frequency response of the adaptive filter according to the error in the same subband. For this purpose, a multirate adaptive filter is used where a subband decomposition of the error is performed using critically sampled lossless perfect reconstruction filter banks. This new algorithm is based on the minimization of a weighted criterion by a stochastic gradient algorithm and leads to improvements in convergence rate compared to both LMS and classical frequency domain algorithms.

1. INTRODUCTION

SQUARE orthonormal transforms have been introduced in adaptive filters for improving the convergence rate of the Least Mean Square (LMS) algorithm. Two such approaches are, on one side, the Frequency domain Block LMS (FBLMS) algorithm that has been derived from the Block LMS (BLMS) using the Discrete Fourier Transform (DFT) and, on the other side, the Transform Domain Adaptive Filter (TDAF) [1]. In all cases, the orthogonal transform is used as a means for decomposing *the input* into approximately decorrelated components.

It is well known that LossLess (LL) Perfect Reconstruction (PR) Filter Banks (FB) also provide approximately decorrelated decompositions of signals (the coefficients of a LL PR FB can be interpreted as a non square matrix of orthonormal vectors [2]). They have the advantage of achieving efficient decorrelation even for a small number of components. Since LL PR FB can be used to implement them, the Discrete Wavelet Transform (DWT) also belongs to this category. First attempts for using filter banks or DWT in transform domain adaptive filtering were reported in [3, 4]. A common characteristic of these schemes is that the "transform" is applied to the inputs of the adaptive filter. Moreover, the variables which are explicitly adapted are the filter coefficients in the transform domain. This imposes some links between the filter length and the transform size, which must be taken into account in the filter bank schemes (and wavelet ones) by some kind of periodization of the input signal.

This paper provides another way of introducing different step sizes in different subbands (thus also improving the convergence rate of the adaptive filter). The main advantages of our algorithm are twofold: first, it keeps the classical filter bank computational structure, and second, it provides more flexibility in the length of the filters in the bank. Finally, our approach shows some connections between *appropriately weighted* least squares minimization and the convergence rate of the adaptive filter.

Note that another set of papers [5, 6, 7] deals with adaptive filtering in subbands mainly in order to allow computational savings. In these algorithms, the adaptive filtering is performed in each subband. Thus, the problem of adapting a single long FIR filter is converted into that of adapting several short filters operating at a lower rate [5]. However, when critical subsampling is used, the output contains undesirable aliasing components which may degrade the adaptation of the algorithm. A possible explanation of the problems encountered with this approach is that one tries to use the subbands in a *fast* convolution algorithm, which can be done only in an approximate way. In order to avoid this problem, our method separates the convergence improvement (which is be obtained by means suited to a subband approach) from the reduction of complexity (which can be obtained by any fast algorithm since our algorithm is basically of a block type).

2. NOTATIONS, CRITERION

This study is undertaken in the context of adaptive identification, all variables being assumed to be complex valued, which corresponds to the most general case. In the following, the operator $(.)^t$ denotes transposition as $(.)^*$ denotes conjugation and $(.)^H = \left((.)^t\right)^*$, while $Diag(X)$ indicates the diagonal matrix whose diagonal elements are the components of vector X. Let W^* be the FIR transversal adaptive filter, of length L. Notations x_n and d_n denote respectively the input and reference signal. N is the block size (number of computations that are grouped together), while KN is the length of the filters in the orthogonal filter bank. Upper case letters denote vectors or matrices of appropriate sizes:

$$\begin{aligned} X_n &= (x_n, x_{n-1}, \cdots, x_{n-KN+1})^t \\ \mathcal{X}_n &= (X_n, X_{n-1}, \cdots, X_{n-L+1})_{KN \times L} \\ W_n &= (w_0(n), w_1(n), \cdots, w_{L-1}(n))^t \end{aligned}$$

This work has been funded by the CCETT Rennes, France. Part of the work described here is filed as a patent. Its number is unknown at time of writing and will be available by the time of the conference.

0-7803-2431-5/95 $4.00 © 1995 IEEE

$$D_n = (d_n, d_{n-1}, \cdots, d_{n-KN+1})^t$$
$$E_n = (e_n, e_{n-1}, \cdots, e_{n-KN+1})^t = D_n - \mathcal{X}_n W_n^\star$$

The vector of the coefficients of the i^{th} filter of the analysis bank is denoted by H_i, all of them being gathered in an $N \times KN$ matrix $H = (H_0^t, \cdots, H_{N-1}^t)^t$.

Suppose now that the error e_{kN+n}, $0 \leq n \leq N-1$, $k \geq 0$, has been passed through a LL PR FB, thus being decomposed in N subbands $(e_k^0, \cdots, e_k^{N-1})^t$, $k \geq 0$, by H the analysis bank. If the input signal is ergodic and wide sense stationary, the orthogonality property of the LL PR FB ensures that both formulations of a block criterion below are equivalent: ($\mathcal{E}(.)$ denotes expectation):

$$J^{Block} = \mathcal{E}\left(\sum_{n=0}^{N-1} |e_{kN+n}|^2\right) = \mathcal{E}\left(\sum_{i=0}^{N-1} \left|e_k^i\right|^2\right) \quad (1)$$

Of course, minimizing both versions of the criterion would result in the same algorithm, i.e. a BLMS algorithm. However, consider now the minimization of the following weighted least squares criterion, where the quadratic errors in each subband are weighted by some constants:

$$J^{WSAF} = \sum_{i=0}^{N-1} \lambda_i \mathcal{E}\left(\left|e_k^i\right|^2\right) \quad (2)$$

Here, the size of the "transform" (FB) is independent from the filter length and depends only on the block size. Since this approach relies on orthogonality, and since orthogonality of the LL PR FB requires the presence of a subsampling by N, this method is restricted to block algorithms.

3. THE PROPOSED ALGORITHM

Assume that the set of weights λ_i is fixed. Denoting $X_n^{i\,t} = (x_n^i, \cdots, x_{n-L+1}^i) = H_i \mathcal{X}_n$ the non subsampled output of the i^{th} analysis filter, an LMS-like adaptation of the criterion (2) is easily obtained by computing its instantaneous gradient estimate $\hat{J}_k(W)$ relatively to $W^\star$:

$$W_{(k+1)N}^\star = W_{kN}^\star - \mu \frac{\partial \hat{J}_k}{\partial W} = W_{kN}^\star - \mu \sum_{i=0}^{N-1} \lambda_i e_k^i \left(\frac{\partial e_k^i}{\partial W^\star}\right)^\star$$

where μ is the scalar step size controlling convergence rate, leading to:

$$W_{(k+1)N}^\star = W_{kN}^\star + \mu \sum_{i=0}^{N-1} \lambda_i X_{kN}^{i\,\star} e_k^i \quad (3)$$

This can be rewritten in a more compact form as:

$$W_{(k+1)N}^\star = W_{kN}^\star + \mu (H\mathcal{X}_{kN})^H \Lambda^{-2} H E_{kN} \quad (4)$$

where $\Lambda^{-2} = Diag(\lambda_0 \cdots \lambda_{N-1})$. The adaptive scheme is depicted in fig.1 in the special case where $L = N$. It can be checked that when $K = 1$ (H is a square orthonormal transform) and $\forall i\ \lambda_i = 1$ this algorithm is exactly the classical BLMS algorithm. However, at this point, it is not clear how the use of a weighted mean square error instead of a regular one could improve the convergence rate of the algorithm. This is explained in the next section.

4. CHOICE OF THE WEIGHTS

This section is concerned with the influence of the weights λ_i on the convergence rate. In order to understand this mechanism, the case where the filter bank is composed of N ideal Nyquist filters that are adjacent and do not overlap (i.e. $K \to +\infty$) is considered in the following. Let W_o be the L-tap filter to be identified (same length of $W^\star$) and $\delta W_n = W_n - W_o$. We have $D_n = \mathcal{X}_n W_o^\star$.

4.1. Appropriate choice of the weights

Under the previous assumptions, each term of the sum in (3) deals with adapting a different part of the spectrum of $W_n^\star$. Indeed, the spectra of X_n^i are non overlapping. Thus, this algorithm minimizes the error of each subband independently without any influence from the other ones, and these errors can be adapted with a different step size, without any drawback: our algorithm behaves as several LMS working separately in each subband. Furthermore, it is well known that the best convergence rate of the LMS algorithm is achieved for a convergence step size given by: $\mu\lambda_i = 1/(L\sigma_{x^i}^2)$ where $\sigma_{x^i}^2$ denotes the power of x_k^i.

The parameters μ and λ_i need to be tuned. Each weight λ_i is chosen so that the fastest convergence occurs independently in each subband when $\mu = 1$. Such a choice corresponds to $\lambda_i = 1/(L\sigma_{x^i}^2)$. The actual algorithm using this set of weights and a smaller μ (in order to obtain various trade-offs between convergence rate and residual error) is denoted as the Weighted Subband Adaptive Filter (WSAF). Note that choosing λ_i in order to minimizing each subband error e_k^i with the fastest convergence rate is equivalent to minimizing the weighted criterion (2) with the specific weights given above. This comes in contrast with classical frequency-domain or transform-domain adaptive algorithms which, despite intuition, do not minimize each frequency band independently.

4.2. Convergence rate

This subsection intends to provide a more precise understanding of the underlying mechanism allowing a faster convergence for the WASF.

Let $R_{X^iX^i}$ denote the size L autocorrelation matrix of the non subsampled outputs of the i^{th} filter. Since the filters in the bank are selective, x_n^i has a narrow spectrum. Thus, $R_{X^iX^i}$ seems to be badly conditioned. Apparently this comes in contradiction, with the expected good convergence behavior of the WSAF.

Suppose that no noise is added to d_n. Under the usual assumption that the adaptive filter taps are uncorrelated from the input samples x_n, the WSAF adaptation equation yields:

$$\mathcal{E}(\delta W_{n+N}^\star) = \mathcal{E}\left[I_L - \mu (H\mathcal{X}_n)^H \Lambda^{-2} H\mathcal{X}_n\right] \mathcal{E}(\delta W_n^\star)$$

The convergence rate of the WSAF is thus determined by the eigenvalue spread of matrix $M = \mathcal{E}\left[(H\mathcal{X}_n)^H \Lambda^{-2} H\mathcal{X}_n\right] = \sum_{i=0}^{N-1} \lambda_i R_{X^iX^i}$, (which should be close to one for fast convergence). Asymptotically (i.e. $L \to +\infty$), F_L diagonalizes all matrices $R_{X^iX^i}$. If the analysis bank is composed of non overlapping perfect Nyquist filters, each individual matrix $R_{X^iX^i}$ is singular. However, an eigenvalue will never be zero in all subbands. Therefore, the summation of all matrices $R_{X^iX^i}$ weighted by λ_i is not singular, and for each matrix $R_{X^iX^i}$, the weight λ_i plays the role of a normalization coefficient such that the eigenvalue spread of the summation is reduced.

This development has been checked by simulation by computing the eigenvalue spread of the matrices involved in the convergence. In the case of an AR2 highly correlated input signal (a white gaussian noise filtered by the inverse of $1 - 1.6z^{-1} + 0.81z^{-2}$). The WSAF with weights $\lambda_i = 1/(L\sigma^2_{x^i})$ and increasing filter lengths in the bank is compared to the LMS algorithm. The number of subbands is set to $N = 10$, and the adaptive filter to be modelized has $L = 20$ taps. The length of the filters in the bank used are respectively 20 ($K = 1$) for a DCT_{IV}, 40 ($K = 2$) for an Modulated Lapped Transform (MLT) and 80 ($K = 4$) for an Extended LT (ELT) [2]. The eigenvalue spread reduces respectively from 1400 for R_{XX} (LMS case) to 51.2, 2.7 and 2.2 for the various matrices M. This explains the better convergence behavior of the WSAF compared to the LMS algorithm.

4.3. A time varying strategy for the weights and step size

This part provides a good choice for the weights when a white noise is added at the output of the adaptive filter (e.g. in an Acoustical Echo Cancelation context: AEC). Let b_n denote this zero mean white noise and $B_n = (b_n, b_{n-1}, \cdots, b_{n-KN+1})^t$. The modelization error vector is defined as $\varepsilon_n = (\epsilon_n, \epsilon_{n-1}, \cdots, \epsilon_{n-KN+1})^t = B_n - E_n$.

All signals x_n, b_n, d_n and e_n are assumed to be ergodic and wide sense stationary. The filters in the bank are supposed to be non overlapping perfect Nyquist filters. Consequently, the subband outputs are uncorrelated. In the following, d^i_k is used for the i^{th} subband sample associated to block k, resulting of the decomposition of d_n by the analysis bank H. Let us calculate the expectation of the squared norm of δW_{n+N} ($\|\delta W_n\|^2 = \delta W_n^H \delta W_n$). Under the assumption that $x_n \perp b_n$ and $b_n \perp \epsilon_n$ ($\perp$ stands for "is uncorrelated with") and designating by A_n the matrix: $A_n = \Lambda^{-2} H \mathcal{X}_n \mathcal{X}_n^H H^H \Lambda^{-2}$ we have:

$$\begin{aligned}\mathcal{E}\left(\|\delta W_{n+N}\|^2\right) \;=\; & \mathcal{E}\left(\|\delta W_n\|^2\right) + \mu^2 \mathcal{E}\left[(HB_n)^H A_n H B_n\right] \\ & + \mu^2 \mathcal{E}\left[(H\varepsilon_n)^H A_n H \varepsilon_n\right] \\ & - 2\mu \mathcal{E}\left[(H\varepsilon_n)^H \Lambda^{-2} H \varepsilon_n\right] \qquad (5)\end{aligned}$$

At convergence, $\mathcal{E}\left(\|\delta W_{n+N}\|^2\right) = \mathcal{E}\left(\|\delta W_n\|^2\right)$. If $\epsilon^i_k \perp \epsilon^j_k$, $x^i_k \perp x^j_k$, $x^i_k \perp \epsilon^j_k$ for $i \neq j$ and $\mathcal{E}\left(\left|\epsilon^i_k\right|^2 \left|x^i_{n-k}\right|^2\right) = a_i \sigma^2_{\epsilon^i} \sigma^2_{x^i}$ (where a_i is an alternative to the independence hypothesis), it is possible to prove that (5) yields:

$$\sum_{i=0}^{L-1} L\lambda_i \mu \left[\mu \lambda_i \sigma^2_{x^i} \left(\sigma^2_{\epsilon^i} + \sigma^2_{b^i}\right) - \frac{2}{L} \sigma^2_{\epsilon^i}\right] = 0 \qquad (6)$$

As the filter bank is "ideal", the i^{th} subband residual error is due only to the signals x^i_n and b^i_n in that very subband. In this case, each individual term of the sum (6) is zero.

In an AEC system, the adaptive filter aims at attenuating the echo. Therefore, it is reasonable to set the requirement that the average power of the misadjustments due to noise σ_{ϵ^i} for each subband should be asymptotically lower than the power of the actual echo σ_{d^i} in the same subband, say ρ_i times lower. This condition reads: $\sigma_{\epsilon^i} \leq \rho_i \sigma_{d^i}$, $0 \leq i \leq N-1$ and results in the following condition:

$$\mu_i = \mu \lambda_i \leq \frac{2\rho_i}{L\sigma^2_{x^i}\left(a_i \rho_i + \frac{\sigma^2_{b^i}}{\sigma_{d^i}}\right)}, \quad 0 \leq i \leq N-1$$

When a low Signal to Noise Ratio (SNR) occurs, the classical proposed adaptation rule leads to an erratic variation of the adaptive filter taps. This could yield large residual errors, even larger than the actual echo. The step sizes provided in this part enable to struggle against this problem. Indeed, if the SNR is low in a subband, the weight $\mu\lambda_i$ in that specific subband (obtained by exponential window estimations of the various quantities) is also small. Thus, the adaptation process is slowed down in that subband until a good SNR is encountered. On the contrary, with excellent SNR the step size formula reduces to the classical one of subsection 4.1. In a context of AEC for non stationary signal such as speech, this procedure can improve the convergence behavior of the WSAF. The resulting algorithm is called the "improved WSAF". Note that the use of these step sizes can only result in decreasing the adaptation step size, hence cannot lead to instability.

5. ARITHMETIC COMPLEXITY

The WSAF is dedicated to very long adaptive filters (as in AEC) and small number of subbands (required to keep the overall complexity as low as possible). Thus, this section only treats the case where the adaptive filter length is larger than the filter size in the bank: $L > KN$. In the sequel and for the simulations, an ELT is used as a filter bank. Using overlap techniques based on the Fast Fourier Transform for computing the convolution and the correlation in the update equation of the WSAF leads to the following complexity in terms of real additions ($\alpha_{\mathbb{R}}$) and real multiplications ($\mu_{\mathbb{R}}$) processed per each output sample:

$$\left\{\begin{aligned}\mu_{\mathbb{R}} = \;& 8 - 8K + \tfrac{8L}{N^2} + \tfrac{4L}{KN^2} + \tfrac{12}{N} - \tfrac{9L}{N} + 3\log_2(N) \\ & + 2\log_2(2N) + \tfrac{4L\log_2(2N)}{N} + 4K\log_2(2KN) \\ & + \tfrac{2L\log_2(2KN)}{N} \\ \alpha_{\mathbb{R}} \;& = -2 - 10K + \tfrac{8L}{N^2} + \tfrac{4L}{KN^2} + \tfrac{12}{N} - \tfrac{5L}{N} \\ & + 9\log_2(N) + 6\log_2(2N) + \tfrac{12L\log_2(2N)}{N} \\ & + 12K\log_2(2KN) + \tfrac{6L\log_2(2KN)}{N}\end{aligned}\right.$$

That is to say that for an $L = 1024$ tap adaptive filter, $N = 32$ subbands and an MLT as filter bank (i.e. the $K = 2$ ELT case): $\mu_{\mathbb{R}} = 1013.38$ and $\alpha_{\mathbb{R}} = 3725.38$. Note that these complexities deal with the process of complex data. In an AEC context all variables are real and the complexities are thus approximately halved. In comparison with a small block version of the FBLMS: the SBFBLMS (the adaptive filter is split in equal sized vectors of size the block length, each of them being adapted by an FBLMS), the complexity of the WSAF is 17% smaller.

6. SIMULATIONS AND CONCLUSION

Fig.2 compares the new algorithm -WSAF- to the LMS and SBFBLMS algorithms. The simulation is run in a context of adaptive modelization. The input of the adaptive filter is an USASI noise (stationary noise with the same spectrum as speech in average). The filter to be identified has 128 taps (it is the truncation of the acoustic response of a room), so does the adaptive filter ($L = 128$). White noise is added to the reference signal: the output SNR is 40dB. The parameters of the WSAF algorithm are $N = 32$ and $K = 2$. The filter bank is an MLT in which the filter length is twice the number of subbands ($K = 2$). The step size of each algorithm is chosen in order to enable the fastest convergence rate. The noise is subtracted from the error before computing its mean squared value.

The proposed algorithm is clearly an improvement over the LMS and the SBFBLMS algorithms (for which the parameters were comparably tuned: same number of vectors in the transform and same block sizes) in all the simulations we have run.

Finally, fig.3 shows a classical Normalized LMS (NLMS) algorithm, an improved NLMS algorithm (with the same time-varying strategy as we propose for the WSAF), and an improved WSAF, in a context of AEC. The SNR is 10dB. It clearly appears that, while the improvement provided by the time-varying strategy on the NLMS is important in terms of convergence, the same strategy applied to the WSAF allows a further improvement of about 7dB on the residual error without loss in terms of convergence.

7. REFERENCES

[1] Allen M. Peterson S. Shankar Narayan and Madihally J. Narasimha. Transform domain LMS algorithm. *IEEE Trans. on ASSP*, 3(31):609–615, June 1983.

[2] Henrique S. Malvar. *Signal Processing with Lapped Transforms*. Artech House, Norwood, MA 02062, 1992.

[3] Srinath Hosur and Ahmed H. Tewfik. Wavelet transform domain LMS algorithm. In *Proceedings of the Int. Conf. on Audio Speech and Sig. Proc.*, pages III.508–III.510, Minneapolis, USA, April 1993.

[4] Miloš Doroslovački and Huong Fan. Wavelet-based adaptive filtering. In *Proceedings of the Int. Conf. on Audio Speech and Sig. Proc.*, pages III.488–III.491, Minneapolis, USA, April 1993.

[5] André Gilloire and Martin Vetterli. Adaptive filtering in subbands with critical sampling: Analysis, experiments, and application to acoustic echo cancellation. *IEEE Trans. on Circ. and Syst.*, 40(8):1862–1875, August 1992.

[6] W. Kellermann. Kompensation akustischer echos in frequenzteilbandern. *Frequenz*, 39:209–215, July 1985.

[7] Mariane R. Petraglia and Sanjit K. Mitra. Performance analysis of adaptive filter structures based on subband decomposition. In *Proceedings of the Int. Symp. on Circ. And Systems*, pages I.60–I.63, Chicago USA, June 1993.

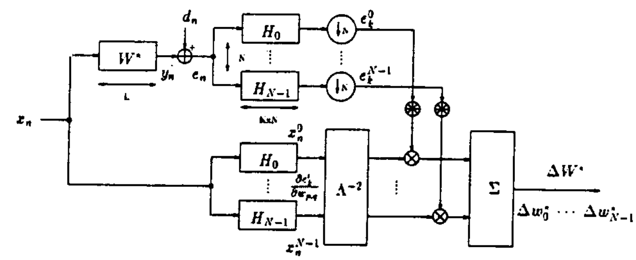

Figure 1: Multirate adaptive filter scheme for $K = 1$

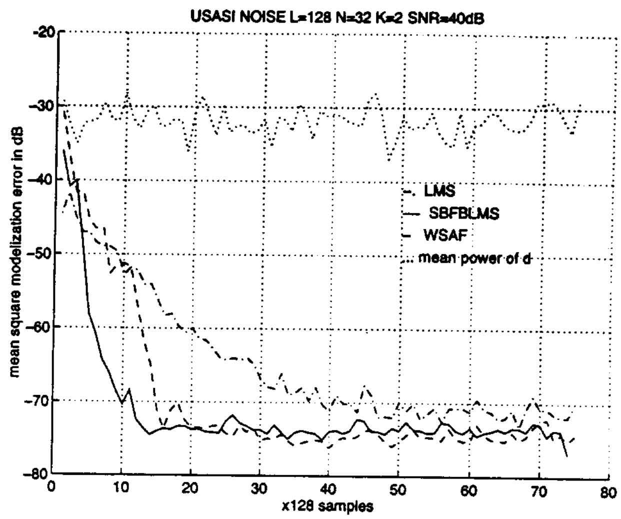

Figure 2: Comparison of the convergence curves of the WSAF, SBFBLMS and LMS algorithms

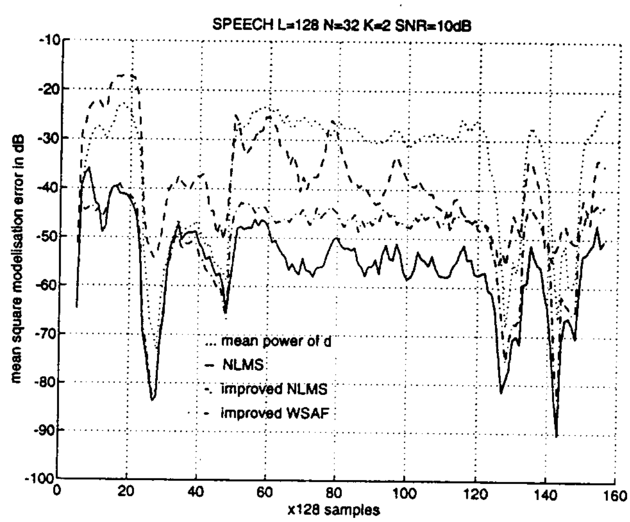

Figure 3: Comparison of the convergence curves of the LMS, improved LMS and improved WSAF algorithms