

# AFFINE PROJECTION ALGORITHM FOR OVERSAMPLED SUBBAND ADAPTIVE FILTERS

*H. R. Abutalebi<sup>1,3</sup>, H. Sheikhzadeh<sup>2,3</sup>, R. L. Brennan<sup>2</sup>, and G. H. Freeman<sup>3</sup>*

<sup>1</sup>EE Dept., Amirkabir University of Technology, Tehran, Iran

<sup>2</sup>Dspfactory Ltd., Waterloo, Ontario, Canada

<sup>3</sup>E&CE Dept., University of Waterloo, Waterloo, Ontario, Canada

E-mails: abutalebi@aut.ac.ir, {hsheikh, rob.brennan}@dspfactory.com, freeman@pce.uwaterloo.ca

## ABSTRACT

The performance of the Normalized Least Mean Square (NLMS) algorithm for adaptive filtering is dependent on the spectral flatness of the reference input. Thus, the standard NLMS algorithm does not perform well in Over-Sampled Subband Adaptive Filters (OS-SAFs) because colored subband signals are generated even for white input signals. Thus we propose the use of the Affine Projection Algorithm (APA) to adapt the individual subband filters in OS-SAF systems. The OS-SAF using APA for adaptation is implemented on a fast, low-resource over-sampled filterbank. Through both theoretical and experimental analyses, it is demonstrated that a low order APA will significantly improve the convergence behavior, offering a low computational complexity compared to the Recursive Least Squares (RLS) method. We employ a recursive method of calculating the correlation matrix to further decrease the computation cost without affecting the performance.

## 1. INTRODUCTION

The Normalized Least Mean Square (NLMS) algorithm is a popular method used in adaptive filtering. It is a simple, stable adaptation technique of low complexity. However, NLMS convergence is sensitive to the spectral flatness of the reference input and may be very slow when the input signal is colored. In this situation, the use of the Recursive Least Squares (RLS) algorithm speeds up the convergence rate considerably. However, a major drawback of the RLS in practical applications is its high computational requirement [1].

To reduce computations, the Affine Projection Algorithm (APA) has been introduced as a link between NLMS and RLS [2-3]. By employing several input vectors, APA provides faster convergence than NLMS, especially when the reference input of the adaptive filter is highly colored. Furthermore, it requires much fewer computations than the RLS method and is more stable.

On the other hand, in many adaptive applications, Over-Sampled Subband Adaptive Filters (OS-SAFs) have become a common practical solution [4-5] because of the well-known advantages of subband processing, simplified implementation and much reduced distortion (aliasing) as compared to critical sampling implementations [6]. OS-SAF offers a simplified structure that significantly reduces the aliasing level in subbands. In order to reduce group delay while maintaining aliasing at a

low level, it is desirable to use an over-sampling factor (OS) of 2 or more [5].

When adaptive filters are used in these highly over-sampled subband structures, the over-sampled inputs to each subband adaptive filter are colored leading to slow convergence of the NLMS. In these situations, APA is a reasonable alternative adaptation technique. The considerable improvement that is achieved by employing APA in fullband adaptive filters [3], strongly motivates us to propose its use for over-sampled subband structures.

Fig. 1 shows the block diagram of the OS-SAF system in the noise cancellation setup. Due to its desirable properties, a very efficient Weighted OverLap-Add (WOLA) filterbank [5] is employed as the analysis/synthesis filterbank in our research. The WOLA is a highly over-sampled, generalized DFT, uniform filterbank (OS = 2, 4 or more). Here, the number of bands is set to  $K = 32$  (actually,  $K/2$  real bands due to frequency symmetry) and the decimation rate in each subband is  $R = 8$ . The over-sampling rate is  $OS = K/R = 4$ . Shown in Fig. 2, each APA adaptive processor contains an adaptive filter using APA to adapt its weights.

This paper is organized as follows. Section 2 briefly introduces APA. In Section 3, a theoretical formula for time-variations of Mean Squared Error (MSE) is first presented. Then the relationship between convergence improvement and APA order is analyzed including the issue of complexity. The results of system simulation and evaluation in a noise cancellation setup are described in Section 4 and, finally, conclusions of this work are presented in Section 5.

## 2. AFFINE PROJECTION ALGORITHM

Since the WOLA filterbank provides near-orthogonal subbands, we shall only consider the first-band adaptive filter for our theoretical analysis. The total (fullband) MSE of the system can be approximated as the sum of subband MSE's. For simplicity subscript  $k$  (representing the  $k^{\text{th}}$  subband signal and filter) is dropped in this analysis.

The APA is a generalization of the NLMS method for adaptation of adaptive filter weights. In NLMS, given the present weight vector  $\mathbf{w}_n$  (of length  $M$ ), the reference input vector

$\mathbf{x}_n = [x_n, x_{n-1}, \dots, x_{n-M+1}]^T$  and the present value of the desired

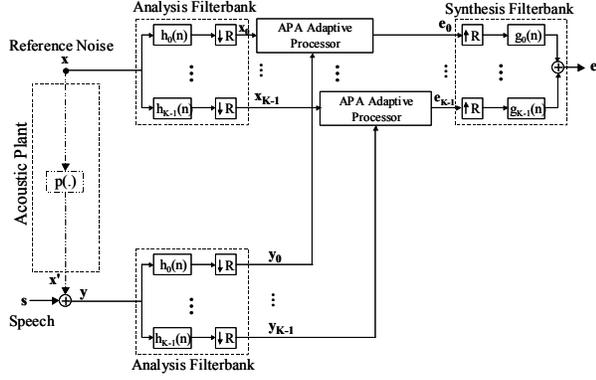


Fig. 1. Block diagram of the SAF system.

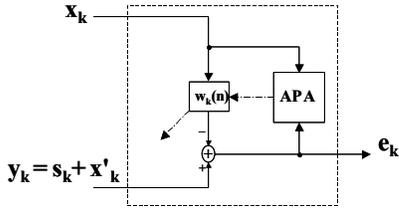


Fig. 2. APA adaptive processor.

signal  $x'_n$  (noise component of the second microphone in a noise cancellation setup), a new coefficient set  $w_{n+1}$  is calculated to minimize  $\|w_{n+1} - w_n\|$  subject to the condition

$$w_{n+1}^H x_n = x'_n. \quad (1)$$

In NLMS, the new adaptive filter weights have to best fit the last input vector to the corresponding desired signal. In APA, this fitting expands to  $P-1$  past input vectors [4]. i.e.,

$$w_n^H x_{n-i} = x'_{n-i}, \quad i = 0, 1, \dots, P-1. \quad (2)$$

Given the second microphone signal vector  $y_n = [y_n, y_{n-1}, \dots, y_{n-P+1}]^T$  and error signal vector  $e_n = [e_n, e_{n-1}, \dots, e_{n-P+1}]^T$ , the adaptation algorithm for  $P^{\text{th}}$  order, APA(P), can be summarized as

$$e_n = y_n - X_n^T w_n^* \quad (3.a)$$

$$C_n = [X_n^H X_n + \delta I] \quad (3.b)$$

$$w_{n+1} = w_n + \mu X_n C_n^{-1} e_n^* \quad (3.c)$$

where  $\mu$  is the adaptation step size limited to  $0 < \mu < 2$ . The input signal matrix  $X_n$  is an  $M$  by  $P$  matrix and has the structure

$$X_n = [x_n, x_{n-1}, \dots, x_{n-(P-1)}]. \quad (4)$$

Also, the scalar  $\delta$  is a regularization parameter used to cope with the ill-conditioning in matrix inversion. It could be chosen experimentally or optimally according to the method proposed in [7].

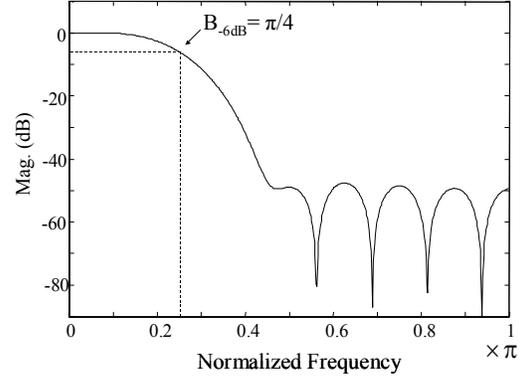


Fig. 3. Typical subband signal spectrum.

### 3. CONVERGENCE RATE AND COMPUTATIONAL COST

#### 3.1. Convergence Behavior

Convergence characteristics of APA have been investigated by a few researchers (e.g., see [2-3]). In [3], Sankaran has presented a detailed performance analysis for a generalized version of APA called NLMS with Orthogonal Correction Factors (NLMS-OCF). With the input vector delay  $D=1$ , NLMS-OCF results in APA. The convergence study is based on a simple model of the input signal vector. In addition to an i.i.d. assumption, the angular orientation of the input vectors is assumed to be discrete [3]. The assumptions are not unrealistic (as evident from our simulation results and the results reported in [3]) and make the convergence analysis tractable.

The theoretical convergence behavior of APA can be described by the following formula for the MSE [3]:

$$\xi_n = E[\|y_n - w_n^H x_n\|^2] = E[s_n^2] + \sigma_x^2 \sum_{m=1}^M (1 - \alpha \beta_m)^n \frac{\lambda_m}{\text{tr}(R_{xx})} \quad (5)$$

where  $s_n$  is the speech sample and  $\sigma_x^2$  is the second microphone noise power. Also,  $\alpha = \mu(2 - \mu)$  and  $\beta_m = 1 - q_m^p$ , where  $q_m = 1 - \frac{\lambda_m}{\text{tr}(R_{xx})}$ .

$R_{xx} = E[x_n x_n^H]$  is the correlation matrix of the reference input and  $\lambda_m, m = 1, 2, \dots, M$  are its eigenvalues.

Sankaran has introduced this formula for the convergence behavior of APA in fullband adaptive filters. In this research, we extend the use of (5) for OS-SAFs and investigate convergence properties based on this formula.

Considering a white noise signal as the input to the analysis filterbank, the reference signal of the adaptive filter has a spectrum (shown in Fig. 3) band-limited to  $\pi/4$  since OS=4. The eigenvalues of the input correlation matrix are plotted in Fig. 4 (in ascending order) for adaptive filter length  $M=16$ . For the monotonic part (pass-band through transition-band) of the input spectrum (shown in Fig. 3), the eigenvalue plot can be considered as a mirrored image of the input spectrum. This is

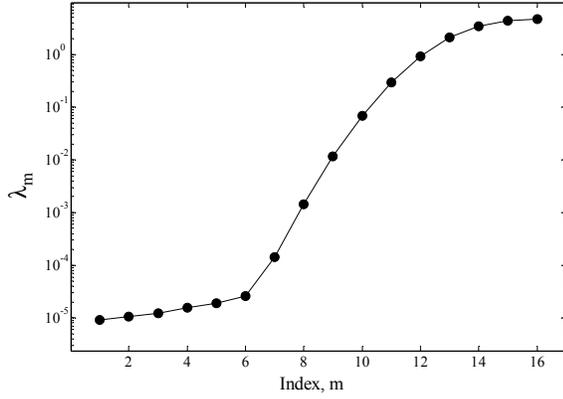


Fig. 4. Eigenvalues of the reference input correlation matrix.

based on the fact that for large  $M$ , a Circulant matrix can approximate the Toeplitz form of correlation matrix. So, the eigenvalues of the correlation matrix will simply approximate spectral samples of the reference input [8].

Fig. 5 shows MSE time-variations in the absence of speech for different projection orders. These curves are calculated by (5) using the eigenvalues displayed in Fig. 4 and typical values of other parameters ( $\mu = 0.001$  and  $\sigma_x^2 = 1$ ). As expected, greater values of  $P$  lead to less MSE and faster convergence. However, as  $P$  increases, the amount of convergence improvement diminishes. This can be explained by differentiating (5) with respect to  $P$ , as follows:

$$\frac{\partial}{\partial P} \xi_n = \sigma_x^2 \sum_{m=1}^M \left[ \frac{\partial}{\partial \beta_m} (1 - \alpha \beta_m)^n \frac{\partial}{\partial P} \beta_m \right] \frac{\lambda_m}{\text{tr}(\mathbf{R}_{xx})}, \quad (6)$$

then,

$$\frac{\partial}{\partial P} \xi_n = n \alpha \sigma_x^2 \sum_{m=1}^M [(1 - \alpha \beta_m)^{n-1} q_m^p \text{Ln}(q_m)] \frac{\lambda_m}{\text{tr}(\mathbf{R}_{xx})}. \quad (7)$$

$\text{Ln}(q_m) < 0$ , so  $\frac{\partial}{\partial P} \xi_n < 0$  for every  $n$ ; this means that greater  $P$  results in less MSE. Similarly, it can be shown that

$$\frac{\partial^2}{\partial P^2} \xi_n = n \alpha \sigma_x^2 \sum_{m=1}^M [(1 - \alpha \beta_m)^{n-2} q_m^p (\text{Ln}(q_m))^2 (1 - \alpha + n \alpha q_m^p)] \frac{\lambda_m}{\text{tr}(\mathbf{R}_{xx})}. \quad (8)$$

As shown in Fig. 6 for an arbitrary time instant ( $n = 10^4$ ), as  $P$  increases, both  $\frac{\partial}{\partial P} \xi_n$  and  $\frac{\partial^2}{\partial P^2} \xi_n$  exponentially decay towards zero. This explains why convergence improvement decreases as APA order increases.

### 3.2. Computational Cost

As mentioned before, APA accelerates convergence speed demanding much less complexity compared to the RLS method. While the computational cost of RLS is about  $3M + 3M^2$  ( $M$  being adaptive filter length) [4], fast implementations of APA have a complexity of  $2M + 20P$  [9]. RLS is an  $O(M^2)$  algorithm

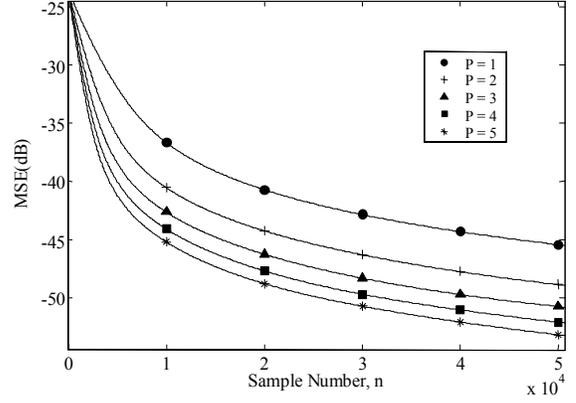


Fig. 5. MSE time-variations of the first subband output calculated by (5) for  $P = 1, 2, \dots, 5$ .

and this makes its computational requirements unacceptable in many applications even for moderate values of  $M$ . Obviously, the APA-order increment also increases the computational complexity moderately.

### 3.3. Approximated Correlation Matrix

The calculation of correlation matrix (Eq. (3.b)) is one of the computationally demanding parts of the algorithm. Here, we propose some modifications in order to make the method more efficient.

Assuming  $P = 2$ ,  $\mathbf{C}_n$  is written as:

$$\mathbf{C}_n = \mathbf{X}_n^H \mathbf{X}_n = \begin{bmatrix} \mathbf{x}_n^H \mathbf{x}_n & \mathbf{x}_n^H \mathbf{x}_{n-1} \\ \mathbf{x}_{n-1}^H \mathbf{x}_n & \mathbf{x}_{n-1}^H \mathbf{x}_{n-1} \end{bmatrix} = \begin{bmatrix} c(n, n) & c(n, n-1) \\ c(n-1, n) & c(n-1, n-1) \end{bmatrix} \quad (9)$$

where  $c(i, j) = \sum_{l=0}^{M-1} x^*(i-l)x(j-l)$ . For moderate or large  $M$  and a slowly varying signal  $x$ ,  $c(i, j)$  is approximated by autocorrelation coefficient  $r_{xx}(i-j)$  scaled by  $M$ . Moreover, we use a smoothing filter (with a close-to-one forgetting factor  $\gamma$ ) to recursively estimate the correlation matrix  $\mathbf{C}_n$  [10]:

$$\hat{\mathbf{C}}_n = (1 - \gamma) \hat{\mathbf{C}}_{n-1} + \gamma M \begin{bmatrix} \mathbf{x}_n^* \mathbf{x}_n & \mathbf{x}_n^* \mathbf{x}_{n-1} \\ \mathbf{x}_{n-1}^* \mathbf{x}_n & \mathbf{x}_{n-1}^* \mathbf{x}_{n-1} \end{bmatrix}. \quad (10)$$

Similarly, in the case of APA(3),  $\hat{\mathbf{C}}_n$  can be estimated as

$$\hat{\mathbf{C}}_n = (1 - \gamma) \hat{\mathbf{C}}_{n-1} + \gamma M \begin{bmatrix} \mathbf{x}_n^* \mathbf{x}_n & \mathbf{x}_n^* \mathbf{x}_{n-1} & \mathbf{x}_n^* \mathbf{x}_{n-2} \\ \mathbf{x}_{n-1}^* \mathbf{x}_n & \mathbf{x}_{n-1}^* \mathbf{x}_{n-1} & \mathbf{x}_{n-1}^* \mathbf{x}_{n-2} \\ \mathbf{x}_{n-2}^* \mathbf{x}_n & \mathbf{x}_{n-2}^* \mathbf{x}_{n-1} & \mathbf{x}_{n-2}^* \mathbf{x}_{n-2} \end{bmatrix}. \quad (11)$$

For higher projection orders, the correlation matrix can be approximated in a similar way. Since all elements along each NW-SE diagonal of  $\hat{\mathbf{C}}_n$  are the same and also  $\hat{\mathbf{C}}_n(i, j) = \hat{\mathbf{C}}_n^*(j, i)$ ,  $\hat{\mathbf{C}}_n$  is a Hermitian-Toeplitz matrix and can be inverted efficiently.

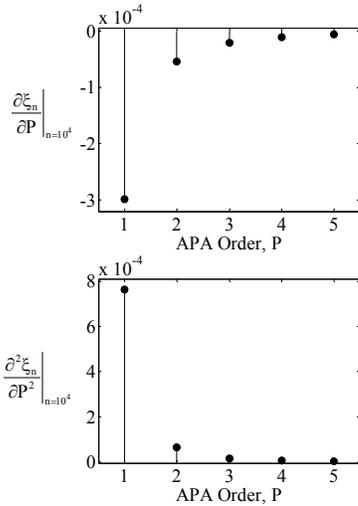


Fig. 6. Values of the first and second partial derivatives of the MSE with respect to P at time instance  $n = 10^4$  and for  $P = 1, 2, \dots, 5$ .

#### 4. SIMULATION

The performance characteristic of APA with different projection orders was evaluated through simulating the OS-SAF in a noise cancellation setup. As mentioned in Section 1, the number of subbands and the decimation rate were chosen as  $K=32$  and  $R=8$ , respectively. A typical low-pass acoustic plant (see Fig. 1) and a white noise signal (sampled at  $F_s = 16$  kHz) for reference input were employed.

MSE time-variations of time-domain (fullband) output signals are plotted in Fig. 7 for  $P=1, 2, \dots, 5$ . Noting that APA(1) is the same as NLMS, APA (with order 2 or more) has increased the convergence rate dramatically. As predicted from theoretical analysis, the amount of improvement saturates as P increases.

It is informative to compare the simulation results to the theoretical ones demonstrated in Fig. 5. Note that while curves in Fig. 5 are for the first-subband MSE, those in Fig. 7 display the time-variations of fullband (time-domain) MSE's. Although the reference input signals are highly colored in OS-SAF, thus violating the independence assumption of input vectors, the theoretical and simulation results are consistent. This can be justified by considering the effect of small  $\mu$  [3]. The use of smaller step sizes reduces the dependency of  $\mathbf{w}_n$  to the present and past input vectors  $\mathbf{x}_i, i \leq n$ . This is just the intended sequel from the independence assumption.

We have repeated this simulation for several acoustic plants (lowpass, highpass, with different values of delay) with three types of input noise (white, babble, and pink) achieving consistent results. This confirms the independence of the results from the acoustical plant and input noise spectrum.

#### 5. CONCLUSION

OS-SAF is a common practical choice for many adaptive systems. However, over-sampling leads to coloring of the signals at the adaptive filter input. This decelerates the convergence rate

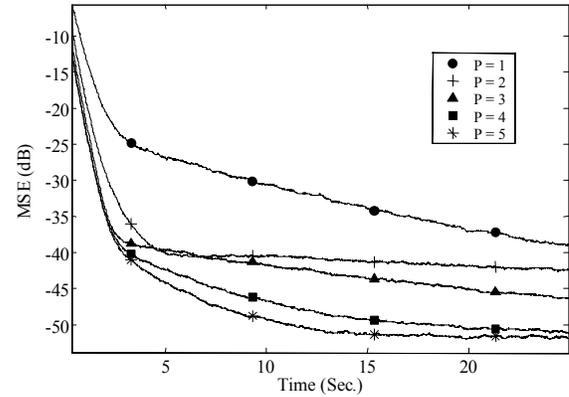


Fig. 7. Simulation results: MSE time-variations of the fullband noise-cancelled output for  $P = 1, 2, \dots, 5$ .

of the NLMS technique which is sensitive to the whiteness of the input signal.

In this research, we employed the APA as an effective adaptation method for OS-SAF. APA has less dependency on the input spectrum and results in faster convergence rates. Also, its computational cost is much less than RLS. Our theoretical analysis and simulation results show that the achievable amount of convergence improvement diminishes as projection order increases. Considering the computational load-convergence rate tradeoff in APA, and the saturating behavior of convergence improvement, reasonable values of P for real-time implementation of the algorithm will be 2 or 3.

We also further improved the efficiency of APA by approximating the correlation matrix through a computationally simple algorithm.

#### 6. REFERENCES

- [1] S. Haykin, *Adaptive Filter Theory*, 3<sup>rd</sup> ed., Prentice Hall Inc., 1996.
- [2] M. Montazeri, and P. Duhamel, "A set of algorithms linking NLMS and block LMS algorithms," *IEEE Tran. Signal Processing*, vol. 43, no. 2, Feb. 1995.
- [3] S. G. Sankaran, "On ways to improve adaptive filter performance," Ph.D. thesis, Virginia polytechnic Institute and State University, Dec 1999.
- [4] S. Weiss, "On adaptive filtering in oversampled subbands," Ph.D. Thesis, Signal Processing Division, University of Strathclyde, Glasgow, May 1998.
- [5] R. Brennan, and T. Schneider, "A flexible filterbank structure for extensive signal manipulation in digital hearing aids," in *Proc. IEEE Int. Symp. Circuits and Systems*, pp. 569-572, 1998.
- [6] A. Gilloire, and M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments and application to acoustic echo cancellation," *IEEE Tran. Signal Processing*, vol. SP-40, no. 8, pp. 1862-1875, Aug. 1992.
- [7] V. Myllyla, and G. Schmidt, "Pseudo-optimal regularization for affine projection algorithms," in *Proc. of the ICASSP*, pp. II.1917-II.1920, 2002.
- [8] R. M. Gray, "On the asymptotic eigenvalue distribution for Toeplitz matrices," *IEEE Tran. Inform. Theory*, vol. IT-18, pp.725-730, Nov. 1972.
- [9] S. L. Gay, and S. Tavathia, "Fast affine projection algorithm," in *Proc. of the ICASSP*, vol. V, pp. 3023-3026, 1995.
- [10] S. D. Stearns, "Fundamentals of adaptive signal processing," in *Advanced Topics in Signal Processing* (J. S. Lim, and A. V. Oppenheim Ed.), Prentice Hall Inc., 1988.