ON INTEGRATING SPEECH CODING FUNCTIONS INTO ECHO CANCELLING FILTERS WITH DECORRELATING PROPERTIES

Pascal Scalart1,2, Franck Bouteille1,3

1 FRANCE TELECOM R&D, 2. Av. Pierre Marzin, 22307 Lannion Cedex, France [pascal.scalart@rd.francetelecom.com]
2 Laboratoire d’Analyse des Systèmes de Traitement de l’Information – Ecole Nationale Supérieure des Sciences Appliquées et de Technologie, 22305 Lannion Cedex, France [Pascal.Scalart@enssat.fr]
3 PRESCOM, Rue de Broglie, 22300 Lannion Cedex, France [prescom.lan@wanadoo.fr]

ABSTRACT
The normalized LMS (NLMS) algorithm has been successfully used in many system identification problems. However, the NLMS algorithm is known to exhibit low convergence speed especially when the input data covariance matrix is ill-conditioned. In this paper, we consider a sub-optimal implementation of the affine projection (AP) algorithm based on a prewhitening mechanism, which renders the convergence characteristics less sensitive to the coloring of the input signal spectrum than is the case for the NLMS algorithm. Comparisons with the AP algorithm are given to validate our approach. Implementation details are discussed in the context of hands-free telephony where echo cancelling and speech coding algorithms are integrated on the same DSP board.

1. INTRODUCTION
Adaptive filtering is a key technique in numerous applications such as system identification, prediction, channel equalization, or speech coding. According to the specificities of these applications, a special emphasis must be put on cheap and practical solutions where low computational complexity is necessary. As a result, the least mean square (LMS) algorithm or its normalized version, i.e. the normalized LMS (NLMS), is one of the most referenced algorithms for adjusting the filter coefficients. The reasons for such popularity are its low cost, its simplicity of implementation and its robust performance. However, a common complaint with regard to the NLMS algorithm is its low convergence speed especially when highly colored signals are used to drive the unknown system.

To improve the adaptive behavior of the NLMS, algorithmic modifications that aim to decorrelate the input sequence have been proposed. Recently, several papers [7-8] focused on the decorrelating properties of the affine projection (AP) algorithm, and its extended version for general moving average (MA) and ARMA processes. This paper presents a sub-optimal implementation of the AP algorithm, which also provides an efficient way to solve the previous issues (convergence rate versus complexity). Monte-Carlo simulations are given to demonstrate the usefulness of the proposed solution. Finally, we discuss about some implementation details of the proposed approach in the context of hands-free telephony where echo cancelling and speech coding algorithms are often integrated on the same DSP board. Simulations with the ITU-T G.729 speech codec are given to demonstrate the computational efficiency and practical interest of the proposed algorithm.

2. ANALYSIS OF THE AP ALGORITHM
In the following, vector quantities will be denoted by boldface lowercase letters, matrices by boldface capital letters, and samples by lowercase letters.

2.1. The Affine Projection Algorithm
Starting from the update equation of the original AP algorithm of order $P$, the $L$-dimensional vector $\mathbf{h}(k)$ of the filter coefficient is updated at time $k$ according to

$$
\mathbf{e}(k) = y(k) - \mathbf{X}_p(k) \mathbf{h}(k-1)
$$

$$
\mathbf{h}(k) = \mathbf{h}(k-1) + \mu \mathbf{X}_p(k) [\mathbf{X}_p(k)^T \mathbf{X}_p(k)]^{-1} \mathbf{e}(k)
$$

where $y(k) = [y(k), \ldots, y(k-P+1)]^T$ is a vector of the $P$ past observations, and $\mathbf{X}_p(k) = [x(k), \ldots, x(k-P+1)]$ is a $L \times P$ matrix consisting in the $P$ past input vectors $x(k) = [x(k), \ldots, x(k-L+1)]^T$. The matrix within the square brackets in (1) is symmetric and non-negative definite. Provided that this matrix is non-singular, the system can be solved using the Moore-Penrose pseudoinverse of matrix $\mathbf{X}_p(k)$.

Defining the a posteriori error vector $\mathbf{e}(k)$ at time $k$, as the modeling errors that would have been produced by the filter resulting of the next weight update, we have

$$
\mathbf{e}(k) = y(k) - \mathbf{X}_p(k) \mathbf{h}(k).
$$

When substituting the a priori error vector (1) into the definition of the a posteriori error vector, we obtain

$$
\mathbf{e}(k) = (1 - \mu) \mathbf{e}(k-1).
$$

Recognizing that the lower $P-1$ elements of the a priori error vector $\mathbf{e}(k)$ are the same as the upper $P-1$ elements of $\mathbf{e}(k-1)$, we can re-express $\mathbf{e}(k)$ as

$$
\mathbf{e}(k) = \begin{bmatrix} y(k) - \mathbf{x}(k)^T \mathbf{h}(k-1) \\ (1 - \mu) \mathbf{\hat{e}}(k-1) \end{bmatrix}
$$

where the $P-1$ dimensional vector $\mathbf{\hat{e}}(k-1)$ consists of the first $P-1$ elements of the vector $\mathbf{e}(k-1)$. \hfill (4)
2.2. Decorrelating property of the AP algorithm

In the following we will consider the case $\mu = 1$ for which fastest convergence rate is obtained for time-invariant systems. Referring to (4), we see that for $\mu = 1$, the last $P-1$ components of $e(k-1)$ are equal to zero. Introducing the Least Square optimal forward predictor of the input vector $x(k)$ given (covariance method) by

$$a(k) = \left[ X_r^2(k-1) X_r(k-1) \right]^T X_r^2(k-1) x(k),$$

(5)

the filter solution (1) can now be rewritten as

$$h(k) = h(k-1) + \frac{u(k)}{u^T(k)u(k)} [y(k) - x^T(k)h(k-1)]$$

(6)

with the decorrelated input vector given by

$$u(k) = x(k) - X_r(k-1)a(k)$$

(7)

Assuming that the optimal filter $h_{opt}$ of the unknown system is a FIR filter of order lower than $L$, i.e. $y(k) = h_{opt}^T x(k) + n(k)$, we can re-write equation (6) as

$$\Delta h(k) = \left( I + \frac{u u^T}{u^T u} \right) \Delta h(k-1) - \frac{u}{u^T u} \left[ n(k) - \sum_{i=0}^{\infty} a(i)n(k-i) \right]$$

(8)

where $\Delta h (k) = h(k) - h_{opt}$ represents the error in the estimated filter coefficients at time $k$. Thus, the noise sequence $n(k)$ is filtered through a filter with coefficients given by relation (5). If the noise is white, the filtered noise power is increased by a factor $\left[ I + a(k)^T a(k) \right]$ (see [7-8]).

2.3. Derivation of the « Pseudo AP » algorithm

In (6), the gradient has the direction of $u(k)$. To further optimize the computational load of the proposed algorithm, we approximate this vector by a tapped-delay line which stores every sample $u(k)$ of the residual error available at the output of a forward linear prediction filter [3,4], i.e. $\tilde{u}(k) = [u(k) \ldots u(k-P+1)]^T$. Hence, we get the following stochastic gradient version of the AP algorithm, denoted in the following « Pseudo AP » algorithm, given by

$$h(k) = h(k-1) + \frac{\tilde{u}(k)}{\tilde{u}^T(k)\tilde{u}(k)} [y(k) - x^T(k)h(k-1)].$$

(9)

The block diagram of the proposed system is depicted in Figure 1. In relation with previous work on decorrelation filters for the NLMS algorithm, it turns out that the stochastic gradient technique defined by (9) does not correspond to classical decorrelation structures that have been already proposed in [3-6]. In [5], a symmetric coupled adaptive predictor-canceler system is proposed where the decorrelation filtering has to be carried twice (on the input sequence and on the identification error). In [6], several approaches were proposed for the application of the decorrelation filters either requiring the inverse of the decorrelation filter or an auxiliary loop for the adaptation of the filter coefficients. Again, these algorithms do not have the convergence and tracking characteristics of the algorithm defined by (9) which is more related to the interpretation of the AP algorithm presented by Rupp [7] (equivalent to the USWC RLS algorithm of Slock [2] with $\mu = 0$), and its statistical behavior analysis described in [8]. However, the gradient in [5-6] has the direction of the decorrelated vector given by (7). Our approximation (9) of this vector leads to a simple implementation with few computational efforts.

![Fig. 1. Block diagram of the « Pseudo AP » algorithm.](image)

A considerable reduction of the computational efforts is reached if the predictor coefficients are not newly computed at each sampling instant, but are adapted periodically to the short-term characteristics of the excitation signal. Additional robustness can be achieved by implementing the decorrelation filter in a lattice structure that enables the coefficients at a given stage to be computed independently of those following that stage. Thus, using the Gram-Schmidt type of orthogonality provided by the lattice/minimum MSE combination, the optimal number of decorrelation filter coefficients (which is related to the AP projection order) is periodically estimated from the final prediction-error (FPE) criteria as

$$\hat{P} = \text{Argmin}_{P=1,2, \ldots, P_{\text{max}}} \left[ \frac{N+P+1}{N-P-1} ||\tilde{u}(k)||^2 \right]$$

(10)

where $N$ denotes the block size, and $P_{\text{max}}$ the maximum prediction order set to a value of 10 according to the implementation constraints (for a sampling frequency of 8 kHz). The block diagram of the identification system is depicted in Figure 1.

2.4. Comparative performance

In the experiments reported in this section, the input signal is a Gaussian $AR(1)$ process, and the unknown system to be identified is a 128-coefficients FIR filter. Perfect modeling of the unknown system is assumed, i.e. $L = 128$. 

The block diagram of the proposed system is depicted in Figure 1. In relation with previous work on decorrelation filters for the NLMS algorithm, it turns out that the stochastic gradient technique defined by (9) does not correspond to classical decorrelation structures that have been already proposed in [3-6]. In [5], a symmetric coupled adaptive predictor-canceler system is proposed where the decorrelation filtering has to be carried twice (on the input sequence and on the identification error). In [6], several approaches were proposed for the application of the decorrelation filters either requiring the inverse of the decorrelation filter or an auxiliary loop for the adaptation of the filter coefficients. Again, these algorithms do not have the convergence and tracking characteristics of the algorithm defined by (9) which is more related to the interpretation of the AP algorithm presented by Rupp [7] (equivalent to the USWC RLS algorithm of Slock [2] with $\mu = 0$), and its statistical behavior analysis described in [8]. However, the gradient in [5-6] has the direction of the decorrelated vector given by (7). Our approximation (9) of this vector leads to a simple implementation with few computational efforts.

![Fig. 1. Block diagram of the « Pseudo AP » algorithm.](image)

A considerable reduction of the computational efforts is reached if the predictor coefficients are not newly computed at each sampling instant, but are adapted periodically to the short-term characteristics of the excitation signal. Additional robustness can be achieved by implementing the decorrelation filter in a lattice structure that enables the coefficients at a given stage to be computed independently of those following that stage. Thus, using the Gram-Schmidt type of orthogonality provided by the lattice/minimum MSE combination, the optimal number of decorrelation filter coefficients (which is related to the AP projection order) is periodically estimated from the final prediction-error (FPE) criteria as

$$\hat{P} = \text{Argmin}_{P=1,2, \ldots, P_{\text{max}}} \left[ \frac{N+P+1}{N-P-1} ||\tilde{u}(k)||^2 \right]$$

(10)

where $N$ denotes the block size, and $P_{\text{max}}$ the maximum prediction order set to a value of 10 according to the implementation constraints (for a sampling frequency of 8 kHz). The block diagram of the identification system is depicted in Figure 1.

2.4. Comparative performance

In the experiments reported in this section, the input signal is a Gaussian $AR(1)$ process, and the unknown system to be identified is a 128-coefficients FIR filter. Perfect modeling of the unknown system is assumed, i.e. $L = 128$. 

The block diagram of the proposed system is depicted in Figure 1. In relation with previous work on decorrelation filters for the NLMS algorithm, it turns out that the stochastic gradient technique defined by (9) does not correspond to classical decorrelation structures that have been already proposed in [3-6]. In [5], a symmetric coupled adaptive predictor-canceler system is proposed where the decorrelation filtering has to be carried twice (on the input sequence and on the identification error). In [6], several approaches were proposed for the application of the decorrelation filters either requiring the inverse of the decorrelation filter or an auxiliary loop for the adaptation of the filter coefficients. Again, these algorithms do not have the convergence and tracking characteristics of the algorithm defined by (9) which is more related to the interpretation of the AP algorithm presented by Rupp [7] (equivalent to the USWC RLS algorithm of Slock [2] with $\mu = 0$), and its statistical behavior analysis described in [8]. However, the gradient in [5-6] has the direction of the decorrelated vector given by (7). Our approximation (9) of this vector leads to a simple implementation with few computational efforts.
A zero-mean white Gaussian noise is added to the desired signal such that $SNR = 40\, \text{dB}$. Comparisons are made by observing the time evolution of the mean-square error (MSE) of respective errors computed on a sliding rectangular window of 128 samples. In all experiments the step-size of the algorithms is set to unity and the projection order of the AP and «Pseudo-AP» algorithms is set to $P = 9$ (thus, the estimation of the optimal number of decorrelation filter coefficients in relation (10) is frozen). In Figure 2, we compare the learning curves for the NLMS, AP, and «Pseudo-AP» algorithms for a time-varying context of great interest since convergence and tracking are two different entities. In fact the fine structure of the echo path remains the same, but a linear increase in the loudspeaker gain (approximately 15 dB) has been introduced at time $t = 10^4\, T$.\[\text{Fig. 2. Learning curves for (a) the NLMS, (b) «Pseudo-AP», and (c) AP algorithms. AR(1) with pole } \alpha = 0.9.\]

For the «Pseudo-AP» algorithm, the coefficients of the decorrelation filter are computed on blocks of 160 samples by the covariance method, and are updated every 40 samples. From the transient analysis, it is readily observed from the curves depicted in Figure 2 that the tracking performance of the «Pseudo-AP» and AP algorithms is far better than the NLMS algorithm. We can also notice the similar behavior of the «Pseudo-AP» and AP algorithms for both the transient and steady-state phases. After the initial convergence period, the steady-state MSE reaches the same level for both algorithms. Observe also that this level is increased (in comparison with the NLMS) because the noise sequence is filtered, thus increasing the noise power by a factor $\left(1 + a(k)^2\right)$ [see equation (8)]. For a pole $\alpha = 0.9$, this leads to an increase in the noise power of 2.6 dB. This theoretical value is in agreement with the experimental relative difference between the steady-state MSE of the «Pseudo-AP» and NLMS algorithms (see Figure 2).

3. INTEGRATING SPEECH CODING FUNCTIONS INTO ECHO CANCELLERS

The previous results demonstrate the practical interest of the proposed approach. In this section, we are interested in demonstrating the advantages of integrating speech coding functions into echo cancellation filters. For this purpose, we have selected the 8 kbit/s ITU-T G.729 codec which is recommended for digital simultaneous voice and data systems, and for use in a voice over frame relay system. Furthermore, extensions at 6.4 and 11.8 kbit/s are available for use in digital (DCME) and packet (PCME) circuit multiplexion equipments.

3.1. The ITU-T 8 kbit/s G.729

The ITU-T G.729 CS-ACELP coder is based on the Code-Excited Linear-Prediction (CELP) coding model. The coder operates on speech frames of 10 ms ($f_s = 8\, \text{kHz}$). For every 10 ms frame, the speech signal is analyzed to extract the parameters of the CELP model (tenth order linear-prediction filter coefficients, adaptive and fixed-codebook indices and gains). These parameters are encoded and transmitted.

At the decoder, these parameters are used to retrieve the excitation and synthesis filter parameters. The speech is reconstructed by filtering the excitation through the short-term synthesis filter, as it is represented in Figure 3. The long-term synthesis filter is implemented using the so-called adaptive codebook approach. The post-filter consists of three filters: a long-term post filter, a short-term post filter, a tilt compensation filter, and an adaptive gain control.

3.2. Integrating coding functions into echo cancellers

Let us now consider the problem encountered in hands-free telephony applications where echo cancelling and speech coding algorithms are both integrated on the same DSP board. In such situations, the excitation signal generated by the speech decoder can be efficiently used by the echo canceller since this signal is representative of the decorrelated input vector $\hat{u}(k)$ used in the update relation (8) of the «Pseudo-AP» algorithm.

A possible solution for this integration is given in Figure 3. The decorrelated input sequence corresponds to the reconstructed excitation including the fixed and adaptive codebook contributions. Before using this signal into the echo canceller part, a copy of the post-filter is applied on the reconstructed excitation to remove non-linear effects introduced by the postfilter such as the adaptive gain control which is used in the decoder to compensate for gain differences between the reconstructed speech signal and the postfiltered signal.

3.3. Comparative performance

In the previous section, artificially simulated signals were used to demonstrate the behavior of the proposed algorithm. In this section the performance of the «Pseudo-
Post filtering

From far-end

to far-end

Fig. 3. Simplified block diagram of the « Pseudo-AP » echo canceller including G.729 CELP decoding functions.

Fig. 4. MSE curves for (a) the NLMS, (b) « Pseudo-AP », and (c) AP algorithms. Input sequence: real speech.

Experiments are conducted with speech encoded with the ITU-T G.729 codec according to the system depicted in Figure 3. Simulations results are given in Figure 4. Also shown on this figure are the echo envelope together with the AP, « Pseudo-AP », and NLMS MSE curves in the noiseless case, i.e. \( n(k) = 0 \) in Figure 3. One can easily observe a) the enhanced performance of the former algorithms compared to the NLMS and b) the close behavior obtained between the « Pseudo-AP » and AP algorithms. The echo cancellation obtained via the « Pseudo-AP » algorithm is clearly superior to that obtained via the NLMS for approximately the same complexity. These results demonstrate the practical interest of the proposed system, since echo cancellation and speech coding are mandatory functions for any hands-free application over mobile, or IP networks.

4. CONCLUSIONS

This paper has presented a simple echo cancelling algorithm derived from the decorrelating properties of the original AP algorithm for its fastest convergence rate. Compared to the NLMS algorithm, the « Pseudo-AP » algorithm exhibits superior performance with respect to convergence rate and tracking, while requiring only a slight increase in the arithmetic operations for computing the decorrelated input samples. Monte-Carlo simulations have shown that the improvement is obtained for any stationary signal even in the presence of a variation of the system to be modeled. Simulations on both synthetic (autoregressive) and real (speech) inputs demonstrate that the proposed algorithm converges as fast as the original AP algorithm. Finally, implementation details have been discussed in the context of hands-free telephony where speech coding functions can be efficiently re-used into the « Pseudo-AP » algorithm to provide an optimized solution with reduced computational complexity, and with tracking and steady-state behavior close to the original AP algorithm.

5. REFERENCE