A Multichannel Affine Projection Algorithm with Applications to Multichannel Acoustic Echo Cancellation

Jacob Benesty, Pierre Duhamel, and Yves Grenier

Abstract — A straightforward generalization of the so-called affine projection algorithm (APA) to the multichannel (MC) case is easily obtained. However, due to the strong correlation between the input signals of the various channels, the resulting algorithm converges very slowly. This letter describes the way to overcome this problem and derives an efficient algorithm that turns out to make use of additional orthogonal projections.

I. INTRODUCTION

MULTICHANNEL (MC) sound pick-up, transmission, and diffusion are likely to be implemented in future teleconference systems to provide users with enhanced quality. Even the future MPEG4 standard plans to make up to eight sound channels available in order to enhance the sound localization effect. However, the echoes in such systems are even more important than in classical hands-free communication so that acoustic echo cancelers are necessary.

Conceptually, MC acoustic echo cancellation can be viewed as a straightforward generalization of the usual single-channel acoustic echo cancellation [1]–[3]. A basic scheme for two-channel (stereophonic) acoustic echo cancellation is sketched in Fig. 1. Note that only one microphone path is shown because the arrangement is fully symmetrical with respect to the two microphone signals in the local room. Clearly, according to this scheme, stereophonic acoustic echo cancellation fits within the framework of direct identification of a multinput, multioutput unknown linear system with inputs and outputs observed. The part of the system concerning microphone (M1) is formed by the parallel combination of the two acoustic paths (W1, W2) extending through the local room from the loudspeakers (HP1, HP2) to (M1). The stereophonic acoustic echo canceler tries to modelize this unknown system with a pair of adaptive filters (H1, H2). The same model applies to the other microphone, with the acoustic paths changed to ones that are appropriate to that microphone.

The input signals are strongly cross correlated. Therefore, an adaptive algorithm (such as the MC RLS algorithm) should take into account these cross-correlation statistics in order to have a good convergence rate [3]. The MC LMS algorithm is very attractive due to its simplicity; however, its behavior is severely degraded in comparison with the monochannel case because it fails to take these cross correlations into account [3].

The affine projection algorithm (APA) was first proposed in [4] and improved in [5] and [6]. It can be seen as a generalization of the normalized LMS (NLMS) algorithm, in which correlations of the input signal are involved. However, it is shown below that a straightforward derivation of the MC APA from the MC NLMS algorithm results in an algorithm in which the above described problem is not solved. Section II presents the mathematical derivation of an improved algorithm taking the input signals cross correlation into account.

II. THE PROPOSED ALGORITHM

Consider a length L echo path. A simple trick for obtaining the monochannel APA is to employ a stochastic gradient-type algorithm cancelling $N$ a posteriori errors. This requirement results in an underdetermined set of linear equations out of which the minimum-norm solution is chosen. In the following, this technique is extended in order to fit our problem. First, the new algorithm is derived for the two-channel case ($P = 2$).

A. Derivation of the Algorithm with $P = 2$

By definition, the $N$ a priori errors and the $N$ a posteriori errors are

$$E(n+1) = Y(n+1) - X(n+1)H(n-M+1)$$
$$E_a(n+1) = Y(n+1) - X(n+1)H(n+1)$$

where $H(n+1) = [H_1(n+1) \ H_2(n+1)]$, $H_1(n+1)$ and $H_2(n+1)$ are the two length $L$ filters at time $n+1$. $X(n+1) = [X_1(n+1) \ X_2(n+1)]$ is the $L \times N$ matrix $X_i(n+1) = [X_i(n+1), X_i(n), \cdots, X_i(n-N+2)]$ made from the $N$ last-input vectors $X_i(n+1)$ ($i = 1, 2$), and $Y(n+1)$ (resp. $E(n+1)$) is the vector of the $N$ last samples of the microphone signal $y(n+1)$ (resp. error signal $e(n+1)$).

After some manipulation, using (1) and (2) plus the requirement that $E_a(n+1) = 0$, we obtain

$$\Delta H(n+1) = E(n+1)$$

where

$$\Delta H(n+1) = \begin{bmatrix} H_1(n+1) - H_1(n-M+1) \\ H_2(n+1) - H_2(n-M+1) \end{bmatrix}.$$
After some calculations, the minimum-norm solution of (7) is given by
\[
\Delta H(n+1) = Z(n+1)[Z(n+1)Z(n+1)]^{-1} E(n+1)
\]
(8)
where \(Z(n+1) = [Z_1(n+1) Z_2(n+1)]^T\) and
\[
Z_1(n+1) = (I - X_2(n+1)[X_2(n+1)X_2(n+1)]^{-1} X_1(n+1) X_1(n+1) \quad \text{and}
\]
\[
Z_2(n+1) = (I - X_1(n+1)[X_1(n+1)X_1(n+1)]^{-1} X_2(n+1) X_2(n+1) \quad \text{and}
\]
(9)
(10)
The actual algorithm is
\[
H(n+1) = H(n - M + 1) + \alpha_0 Z(n+1)
\]
\[
\cdot [Z(n+1)Z(n+1)]^{-1} E(n+1)
\]
(11)
\[
Z_i(n+1) = (I - \alpha_i P_{-x_i}(n+1) X_i(n+1)); \quad i = 1, 2
\]
where \(0 < \alpha_i < 1; \quad j = 0, 1, 2\) are constants that are used to obtain various tradeoffs between convergence rate and residual error. Note that the straightforward APA is a special case of this algorithm, with \(\alpha_1 = \alpha_2 = 0\).

B. Generalization to Arbitrary Number of Channels

The algorithm explained for two channels is easily generalized to an arbitrary number of channels \(P\). Define the following matrix of size \(L \times (P - 1)N\):
\[
\begin{bmatrix}
X_1(n+1) & \cdots & X_j(n+1) & \cdots & X_P(n+1)
\end{bmatrix},
\]
(13)
where \(P \) orthogonality constraints are
\[
C_i(n+1) = [X_i^{P-1}(n+1)]^T \Delta H_i(n+1) = 0_{(P-1)N \times 1}, \quad i = 1, 2, \ldots, P.
\]
(14)
By using the same steps as for \(P = 2\), a solution similar to (8) is easily obtained, and the resulting algorithm is given by (11). The main difference lies in the definition of the projection matrix \(P_{-x_i}(n+1)\) of size \(L \times L\), which is now defined as
\[
P_{-x_i}(n+1) = [X_i^{P-1}(n+1)][X_i^{P-1}(n+1)]^T[\Delta H_i(n+1)]^{-1} [X_i^{P-1}(n+1)]^T,
\]
(15)
\quad \text{for } i = 1, 2, \ldots, P.

Note that this equation holds only under the condition \(L \geq (P - 1)N\), so that the matrix that appears in (15) is invertible. The proposed algorithm is still basically a projection
algorithm, and the convergence proofs derived under this context should still hold.

III. SIMULATION Results

This section compares by simulation the previous algorithms in the two-channel case ($P = 2$). The impulse responses ($W_1$, $W_2$) to be identified are truncated to 256 points. They were obtained from measurements taken in an actual teleconference room and from the two input signals. The length of the filters ($H_1$, $H_2$) is $L = 256$. The signal source is a correlated noise with a spectrum equivalent to the average spectrum of speech. This signal is stationary. A white noise is added to the output. All plots show the mean-squared modeling error versus the number of iterations. Fig. 2 compares the two-channel (T-C) NLMS algorithm to the optimal T-C APA when $N = 1$ and $N = 8$. It is observed that the proposed algorithm converges much faster than the T-C NLMS algorithm and performance improves when $N$ increases. Fig. 3 shows that the straightforward T-C APA converges very slowly, which is a behavior that could be predicted based on the above considerations.

IV. CONCLUSION

This letter has shown how to generalize efficiently the single-channel APA to the multichannel case. This original approach forces a second projection to appear in the algorithm so that the strong cross correlation between the various input signals is taken into account.

REFERENCES