Joint gradient-based time-delay estimation and adaptive minimum mean-squared-error filtering

I. Introduction

The problem of estimating the time delay between two continuous versions of the same signal, each one corrupted by uncorrelated noise components, has been the subject of many research efforts in recent years. The maximum likelihood estimator for the unknown delay has been derived for a static [1]-[2] and a time-varying delay [3]. Closed-loop adaptive techniques using the minimum mean-squared-error (MMSE) or the least squares (LS) criteria have also been proposed. In these cases, the estimator structure is such that one signal is processed by an adaptive system for which the output is compared to the other signal, with the error being used to adapt a conventional adaptive transversal filter or an adaptive delay element.

In this paper, we consider a signal model that generalizes somewhat the conventional model by allowing frequency-dependent attenuation in the delayed path. We also specifically consider discrete-time signals and systems. This work finds some applications in systems modelling problems, where the unknown system often has an impulse response that can be modelled as a pure time delay in series with a linear filter. This can occur in noise or echo cancellation, digital communication or geophysical exploration.

We study a joint adaptive estimator which is composed of an adaptive delay element working in conjunction with an adaptive filter. The adaptive delay element attempts to model the reference delay and can take any real value. The addition of this adaptive delay to the usual adaptive filtering operations can improve the conventional adaptive parameter estimation techniques that would otherwise be of limited usefulness, especially in the case where the main adaptive filter input and its reference signal are uncorrelated with time. A simple adaptive filter has the potential to model both the reference delay and the reference filter, since the overall function can be approximated by an FIR filter with the proper number of taps. This approach is inefficient in the sense that the reference delay is modelled by a shift in the adaptive filter impulse response. For a fixed filter order, this shift may result in an error that is larger than the error corresponding to a perfect model. An additional adaptive delay estimation algorithm, specifically designed to track the reference delay variations, allows a better impulse response centering and the use of an adaptive filter with a smaller order.

In this paper, we present the results of an analysis for the joint adaptive delay and filter structure based on the MMSE performance index. A joint steepest-descent (SD) algorithm and a joint least-mean-square (LMS) algorithm are investigated. The principal contributions of this paper are the generalization of existing gradient-based time-delay estimation without reference filtering, as proposed in [4], and the analysis of a new joint algorithm for the synchronization of the input and the reference signals used by an adaptive filter. Our joint algorithms are not based on the assumption that the input signal and the reference signal fed to an adaptive filter are sampled in the same clock period. They also allow the tracking of time-varying delays in the reference path by a process separated from the adaptive filter.
which is itself free to perform the task of modelling the linear reference filter or its inverse.

The paper is organized as follows. In the next section, the minimum mean-squared-error function is considered in general terms, as a joint function of the two estimates. The form of this function allows one to draw some conclusions about the general convergence behaviour of the joint algorithms. The presence of a multitude of minima in the objective function is discussed. Then the joint steepest-descent algorithm is studied in section 111, where the conditions for convergence to a local minimum of the mean-squared-error function are given. The joint least-mean-square algorithm is investigated in section 112. Analytical results for the convergence in the mean and in the mean square, for both the adaptive delay estimator and the adaptive filter weight vector, are presented. Finally some experimental results are given, in order to complete the presentation.

II. General minimum mean-squared-error function

We consider a situation that generalizes the conventional model used in delay estimation by allowing frequency-dependent attenuation in the delayed path. We also specifically consider discrete-time signals and systems. The corresponding model, where \( z_1(n) \) and \( z_2(n) \) are the two observed signals, is of the form

\[
\begin{align*}
    z_1(n) &= s(n) + v_1(n), \\
    z_2(n) &= h(n) \ast s(nT - D_n) + v_2(n),
\end{align*}
\]

where \( n \) is the discrete-time index, \( s(n) \) is the transmitted signal, \( D_n \) is a time delay (possibly time-varying), and \( h(n) \) is the impulse response of a linear filter which is applied on a delayed-by-\( D_n \) version of the signal \( s(n) \). The discrete-time noise processes, \( v_1(n) \) and \( v_2(n) \), are zero-mean and stationary and are assumed to be uncorrelated with each other as well as with \( s(n) \). The operator \( \ast \) is the convolution operator.

Note that the time-varying reference delay, \( D_n \), is not limited to an integer number of sampling periods and can take any real value. All the discrete-time signals are assumed to be sampled versions, with sampling period \( T \) of continuous-time signals that are strictly band-limited to the frequency range \(-1/2T < f < 1/2T\). A block diagram corresponding to the mathematical model of (1.1)-(1.2) is illustrated in Fig. 1. Note that the case in which the delay, \( D_n \), follows the linear filter is also of interest, but is not considered in this paper. See [5] for more details.

In the joint estimation problem considered in this paper, it is required that both the time-varying delay, \( D_n \), and the reference filter, \( h(n) \), or its inverse, \( h^{-1}(n) \), be estimated\(^7\). The adaptive filter used to estimate \( h(n) \) or \( h^{-1}(n) \) is a transversal filter, with a weight vector \( w_n \) of length \( M \).

In joint MMSE delay estimation and adaptive filtering, the mean-squared-error surface is searched by both the adaptive filter estimation algorithm and the delay estimation algorithm. A system identification configuration takes the form given in Fig. 2.

In general, the output of the adaptive branch can be defined as \( y(n,d_n) \), where the dependence on the adaptive delay is explicitly shown. The reference signal, \( r(n) \), is defined to be one of the two observed signals \( z_1(n) \) or \( z_2(n) \). Then the error signal, \( e(n,d_n) \), is defined as

\[
e(n,d_n) = r(n) - y(n,d_n),
\]

and the MSE function, at time \( n \), is

\[
\xi_n = E\left[ e(n,d_n)^2 \right].
\]

The joint estimation can be thought of as taking place in a vector space made of a weight vector subspace and a delay subspace. The two subspaces are not orthogonal, which implies that the two estimation processes are not independent (because the adaptive filter can model a reference delay). In order to obtain an expression for the MSE function, define as \( u(n) \) the input to the adaptive branch. The signal \( u(n) \) is therefore the generic representation of the observation that must be adaptively processed. It can be \( z_1(n) \), as in Fig. 2, or \( z_2(n) \) if it one wants to estimate the inverse of the reference branch (inverse filtering). The output of the adaptive branch, \( y(n,d_n) \), is assumed to be given by

\[
y(n,d_n) = w_n^H u_n,
\]

where the superscript \( H \) denotes complex conjugate transpose. The vector \( u_n \) is the vector of delayed input samples, stored at iteration \( n \), in the adaptive filter delay line; i.e.,

\[
u_n = \left[ u(nT - d_{n,1}), u(nT - T - d_{n,1}), \ldots, u(nT - MT + T - d_{n,M+1}) \right]^T.
\]

The input-signal autocorrelation matrix and the cross-correlation vector between this input and the reference signal are then expressed as

\[
R_u = E\left[ u_n u_n^H \right]
\]

and

\[
p_u = E\left[ u_n r^*(n) \right].
\]

The MSE function is represented by either one of the following equivalent equations.

---

\(^7\) Note that the inverse of any linear filtering operation \( h(n) \) is denoted as \( h^{-1}(n) \). Therefore \( h(n) \otimes h^{-1}(n) = \delta(n) \), where \( \delta(n) \) is the unit-sample sequence [6].
where the dot denotes a derivative with respect to the delay value \( d_n \). This approximation is used in order to linearize the delay estimation algorithm. The linearized SD algorithm is obtained by combining (10) and (11), and assuming that \( \theta_n \) is a minimum of \( \xi_n \{ \theta_n, w_n \} = 0 \).

It is given by

\[
d_{n+1} = \left(1 - \alpha d_n\right) d_n + \alpha \left[ \xi_n \{ \theta_n, w_n \} \right].
\]  

(12)

which models the behaviour of a first-order delay-lock loop [10].

A. Convergence of the joint SD algorithm

A necessary condition for a specific \( d_n \) and \( w_n \) to be a stationary solution of the joint SD algorithms is that both of the following equations be satisfied [11]:

\[
\nabla_{\theta_n} \xi_n = 0
\]

(13)

\[
\frac{d \xi_n}{d d_n} = 0.
\]

Note that the first equation of (13) is in fact a necessary and sufficient condition for convergence. This is so because \( \xi_n \) is quadratic with respect to \( w_n \), implying that there is a unique minimum in \( w_n \) for a given value \( d_n \). When the first equation of (13) is satisfied, this unique solution is attained, and any further modifications of \( d_n \) will increase \( \xi_n \). This is the case because the adaptive filter models both the relative delay and the reference filter in the minimum MSE sense. Then, this solution corresponds also to a minimum with respect to \( d_n \). The sufficiency of the condition is due to the uniqueness of the minimum with respect to \( w_n \).

If the adaptation factors \( \mu \) and \( \alpha \) are chosen sufficiently small, the process always reaches a limit point [12]. The next proposition gives a condition on \( \mu \) and \( \alpha \) that ensures convergence of the joint algorithms under specific conditions. This condition is derived in [11] for joint carrier phase acquisition and adaptive equalization as encountered in digital communications. It is reformulated here for the problem at hand. This condition is general in that it establishes the stability range for the two adaptation factors such that the MSE is reduced at each iteration, when the two adaptive processes are alternated. It is also important because it confirms that, with the right parameters, the joint SD algorithm converges eventually to a stationary point (i.e., (13) is satisfied).

**Proposition 1.** Let the set of positive integers be divided arbitrarily into two disjoint subsets \( \kappa_1 \) and \( \kappa_2 \) each containing an infinite number of positive integers. Let \( \alpha_n = 0 \) when \( n \in \kappa_1 \), and \( \mu_n = 0 \) when \( n \in \kappa_2 \). Let \( \lambda_{\text{max}}(n) \) be the maximum eigenvalue of the signal autocorrelation matrix \( \mathbf{R}_n \), and let \( d_n \) be the delay value closest to \( d_n \) for which \( \xi_n \{ d_n, w_n \} \) is minimum. The MSE will converge to a stationary point if

\[
0 < \mu_n < \frac{1}{\lambda_{\text{max}}(n)}
\]

(14)

for \( n \in \kappa_1 \), and

\[
0 < \alpha_n < 2 \left[ \frac{\partial}{\partial d_n} \xi_n \{ \theta_n, w_n \} \right]^{-1}
\]

(15)

for \( n \in \kappa_2 \).

The formal proof of this proposition is given in [5]. It is easily seen that when \( \alpha_n = 0 \), the usual SD adaptive filtering conditions apply and (14) is the conventional condition for convergence. When \( \mu_n = 0 \), the MSE function evaluated at \( d_n = \theta_n \) is constant and condition (15) guarantees the convergence of the linearized SD algorithm given in (12).
This proposition states that $d_n$ and $w_n$ may be adjusted in any alternating fashion, and the MSE will converge to a stationary point if $\mu_n$ satisfies (14) during the adjustment of $w_n$, and $\alpha_n$ satisfies (15) during the adjustment of $d_n$. The above condition is important because it confirms that, with the right parameters used in alternation, the MSE is reduced at each iteration and the joint SD algorithm converges eventually to a stationary point.

### B. Steady-state delay estimation properties of the algorithm

In this subsection, we briefly study the system and signal components that directly influence the stability and the delay tracking behaviour of the joint SD algorithm. In order to proceed, we assume that the reference filter $h(n)$ is time-invariant, that the signal-to-noise ratios (SNRs) are high and that the adaptive filter has fully adapted to $h(n)$ and is at least as long as this impulse response. These assumptions imply that in steady state, the $j$th filter coefficient, $w_{nj}$ at iteration $n$, is approximately

$$w_{nj} \approx \frac{h(j)}{h^{-1}(j)}$$

System identification (cancellation)

$$w_{nj} \approx \frac{h^{-1}(j)}{h(j)}$$

Inverse filtering (equalization),

where $h(j)$ is the $j$th weight of the reference path filter. In the analysis, we use the linearized delay adaptation algorithm of (12) with $\hat{E}(d_n, w_n) = \hat{e}_n$ and $\hat{d}_n = D_n$ for the cancellation configuration, and $\hat{d}_n = -D_n$ for the equalization structure. Furthermore, in steady state, we assume that $d_n = \pm D_n$ in which case the error is minimum and the corresponding MSE equals the MMSE. Then $\hat{e}_n||d_n = \pm D_n = \hat{e}_{\min}$ and is constant with time. From (12), the stability range for $\alpha$ is

$$0 < \alpha < \frac{1}{2\hat{e}^2_{\min}}.$$  

The time constant of delay adaptation can be defined by fitting the geometric ratio $1 - \alpha \hat{e}^2_{\min}$ to an exponential with time constant $\tau_{del}$.

$$1 - \alpha \hat{e}^2_{\min} = e^{-\tau_{del}/\tau_{del}} \approx 1 - 1/\tau_{del}$$

The time constant of delay adaptation is therefore

$$\tau_{del} = \frac{1}{\alpha \hat{e}^2_{\min}}.$$  

We assume a configuration in which the reference delay, $D_n$, varies slowly enough so that all the samples in the reference filter delay line are approximately affected by the same delay. Then it can be shown [5] that $\gamma_n$, the delay dependent term of the MSE function, takes the form

$$v^C_n = -2\text{Re} \left[ \sum_{t} \rho_h(t) \phi_h (-tT + D_n - d_n) \right].$$  

$$v^E_n = -2\text{Re} \left[ \phi_{\hat{w}} (D_n + d_n) \right].$$

where the superscripts $(C)$ and $(E)$ stand respectively for cancellation and equalization, and $\phi_h(k)$ is the deterministic autocorrelation of the reference filter impulse response and is defined as

$$\phi_h(k) = \sum_{i} h(i + k) h^*(i).$$

Note that $d_n$ is negative in the equalization case. Comparing (19) and (20), we note that the cancellation configuration is influenced by the form of both the deterministic autocorrelation $\phi_h(n)$ and the input signal autocorrelation $\phi_r(\tau)$, while the equalization configuration is a function of only $\phi_r(\tau)$. Since $\phi_r(\tau)$ exhibits a maximum at $\tau = 0$, $\psi_{\phi_r}(\tau)$ has a global minimum at $d_n = -D_n$. In the cancellation scenario, the characteristics of the delay tracking loop are functions of the reference filter, $h(n)$, but because $\phi_r(\psi)$ has a maximum at $n = 0$, there is a single global minimum corresponding to $d_n = D_n$.

Based on (19), (20) and (17), the following sufficient range of convergence can be computed for the delay gain factor:

$$0 < \alpha < \frac{1}{\Phi_{\phi_r} \text{Re}\left[ \phi_r(0) \right]}$$

Cancellation

and

$$0 < \alpha < \frac{(2\pi)^2}{\Phi_{\phi_r} \text{Im}\left[ \phi_r(0) \right]}$$

Equalization.

where $\Phi_{\phi_r}$ is the maximum value of the input signal power spectral density, $\Phi_{\phi_r}(\omega)$, and the prime denotes the derivative with respect to the continuous-time correlation argument. It is easy to show that $\phi_r(0)$ is proportional to the square of the reference filter bandwidth, as well as to $\phi_r(0)$. This implies that the convergence properties of the delay SD algorithm are related to the power distribution, across the total bandwidth, of the input signal and the reference filter in the cancellation case, and of the signal only in the equalization case. This is a behaviour essentially similar to the adaptive-filter convergence, which is related to the distribution of the eigenvalues of the input-signal autocorrelation matrix [7].

### IV. The joint least-mean-square algorithm

In order to implement the joint steepest-descent algorithm presented in the previous section, the MSE gradient with respect to the adaptive weight vector and the MSE derivative with respect to the adaptive delay both must be estimated. This can be accomplished in various ways; in particular, by approximating the derivatives with difference equations [13], or by approximating the MSE function, $\hat{e}_n = E[1 e(n, d_n)^2]$, with the instantaneous squared error, $\gamma_n = 1 e(n, d_n)^2$, and by applying the SD algorithm. This last option corresponds to the least-mean-square algorithm [14] and is the subject of this section.

Consider a cancellation configuration where it is assumed that the delay, $d_n$, propagates instantaneously into the adaptive filter delay line. The adaptive branch output can be expressed as

$$y(n, d_n) = \mathbf{w}^T u(nT - d_n).$$

and $u(nT - d_n)$ is the delayed vector of input samples defined as

$$u(nT - d_n) = [u(nT - d_n), u(nT - T - d_n), \ldots, u(nT - MT + T - d_n)]^T.$$  

In the adaptive weight vector subspace, the LMS algorithm that we consider is then given by [7]

$$w_{n+1} = w_n + 2\mu e(n, d_n) u(nT - d_n).$$

The error $e(n, d_n)$ is represented by (2). In the adaptive delay subspace, the derivative estimate is given by

$$\hat{e}_n = \frac{\partial}{\partial d_n} e(n, d_n) = -2\text{Re} \left[ e(n, d_n) \frac{\partial y(n, d_n)}{\partial d_n} \right].$$

The LMS adaptive delay algorithm is obtained by using the derivative estimate of (25) in the SD adaptive delay algorithm, defined in (10).
The joint LMS algorithm is defined by

\[ w_{n+1} = w_n + 2\mu e^*(n,d_n) y(nT-d_n) \]  

(26)

and

\[ d_{n+1} = d_n + 2\alpha\Re\{e^*(n,d_n) w_n^*(nT-d_n)\} \]  

(27)

In order to ease the derivations, all signals and systems are considered 
real in the analyses. At this point, we are interested in the convergence 
of the joint LMS algorithm (26) and (27) from an arbitrary initial 
condition.

With the help of the ordinary differential equations (ODE) method 
[15], it is shown in [5] that the joint LMS algorithm, when the gain 
factors are of the form \( \mu = \alpha = 1/n \), converges to a local minimum of 
the MSE function, like the exact version of the joint SD algorithm. This 
result, even if it does not apply directly to algorithm (26)-(27), is 
important by itself since it shows that if the adaptation factors are 
chosen sufficiently small, the estimates produced by the algorithm will 
be, on average, close to a stable stationary point of the MSE function. 
Furthermore, the above result shows that if the gain factors are small 
but constant, convergence cannot be attained in the sense that there 
exists an integer \( N \) such that \( 0(n+1) = 0(n) \) for \( N \leq n \), but the difference 
between the parameter estimate and a stable stationary point will be 
small as \( n \) becomes large and can be made smaller by decreasing the 
gain factors.

A. The joint LMS algorithm in steady state

The quality of the joint LMS algorithm can be studied by considering the 
quality of the two estimates that it generates. Since the delay 
and weight vector estimates are random variables, the joint algorithm 
can be analysed in terms of convergence in the mean and in the mean 
square of either estimate. Because of the coupling between the two 
adaptive processes, the gradient error will affect the delay tracking and 
the derivative uncertainty will itself influence the adaptive filter. These 
mutual effects can be included in the delay variance and weight-noise 
vector correlation matrix, in steady-state conditions. The bounds for \( \mu \) 
and \( \alpha \) are determined for both types of convergence. The results for 
the delay estimator are given first. Then the weight vector estimator is 
considered and finally the two sets of results are combined, to obtain 
some misadjustment expressions for the joint LMS algorithm.

In the course of the analyses, in addition to the general real signals 
and systems assumption already mentioned, the following assumptions 
are used:

1) The input and noise signals are zero-mean Gaussian random pro-
cesses. The noise signals are also assumed to be white noise pro-
cesses.

2) The adaptive system is in steady state and the reference system is 
estationary; i.e., the reference delay is constant at \( D_0 = D \) and the 
reference filter is also fixed in time.

3) Independence theory holds; i.e., the zero-mean input data vectors 
are uncorrelated with each other and with \( r(k) \). Then

\[ E[u_k u_k^T] = 0 \quad \text{for} \quad k = 0, 1, \ldots, n - 1. \]  

(28.1)

\[ E[u_k r(k)] = 0 \quad \text{for} \quad k = 0, 1, \ldots, n - 1. \]  

(28.2)

The terminology independence theory is common in the analysis of adaptive algorithms [see [7] for example].

4) In steady state, the adaptive weight vector, \( w_n \), can be expressed as

\[ w_n = w_{opt} + \eta_n, \]  

(29)

where \( w_{opt} \) is the optimum Wiener solution given by

\[ w_{opt} = K^D p_{|n|a-D}. \]  

(30)

and \( \eta_n \) is the weight-noise vector.

5) In the analysis of the delay estimator, the vector \( \eta_n \) is a zero-mean 
stationary Gaussian vector, uncorrelated with the data vectors 
(because of (28.1)-(28.2)) and such that

\[ E[\eta_i \eta_j] = 0 \quad \text{for} \quad i \neq j. \]  

(31)

The noise vector correlation matrix, defined as

\[ K_\eta = E[\eta_n \eta_n^T], \]  

(32)

is therefore diagonal with the values \( E[\eta_i^2] \) on the main diagonal. In the analysis of the weight vector estimate, the delay estimate 
is assumed stationary.

6) The system is in cancellation configuration (see Fig. 2). The results 
can be extended in a straightforward manner to the equalization 
case.

7) When the signal-to-noise ratios are assumed high, the adaptive-
filter Wiener solution for \( d_0 = D \) is approximately equal to the 
reference filter (in practice, this amounts to SNRs greater than 
10 dB).

Note that Assumption 3 can hardly be justified in practice, but has 
been used with success in the analysis of stochastic algorithms [7]. 
The Gaussian assumption about \( \eta_n \) is also commonly used in the analy-
sis of the LMS algorithm [16]-[17]. The noise vector properties put 
forth in Assumption 5 follow largely from these assumptions and will 
prove to be useful in the analyses. Note in particular, that \( K_\eta \) was 
found to be approximately equal to \( \mu e_{\eta D} J \) in [9], for the LMS 
algorithm. The validity of this approximation is directly related to the 
validity of Assumption 3. In most cases, it is only asymptotically valid as 
the adaptation constant \( \mu \) vanishes.

1. Results for the LMS delay estimator in steady state

The LMS delay tracking algorithm in (27) is analyzed in terms of 
convergence of the delay estimate in the mean and in the mean square. 
The analysis parallels and extends that of Messer [4] and can be found 
in [5].

For \( d_0 = D \), the output of the adaptive branch can be expressed as

\[ y(n, D) = w_{opt}^T u(nT-D) + \eta_n^T u(nT-D). \]  

(33)

The first term on the right is defined as the optimum output, \( \hat{r}(n) \), 
since it represents the adaptive branch output for perfect modelling 
in the MSE sense. The second term on the right is defined as the output 
steady-state noise, \( \chi(n, D) \). Define \( e_{\min}(n, D) \) as the error between 
the optimum adaptive branch and the reference branch; i.e.,

\[ e_{\min}(n, D) = \hat{r}(n) - \hat{r}(n). \]  

(34)
and the corresponding MSE as:

$$\xi_{\text{min}} = E\left[ e_{\text{min}}^2(n, D) \right].$$  (35)

Then, relying largely on Assumptions 3 and 5, we have the following two propositions:

**Proposition 2.** In steady-state conditions, the delay estimator, given by the LMS delay tracking algorithm operating jointly with an adaptive filter, is an unbiased estimator if

$$0 < \alpha < \frac{2}{\xi_{\text{min}}}.  \quad (36)$$

**Proposition 3.** In steady-state conditions, the delay estimator, given by the LMS delay tracking algorithm operating jointly with an adaptive filter, is convergent in the mean square if

$$0 < \alpha < \frac{\xi_{\text{min}}}{2\sigma_G^2}. \quad (37)$$

where the quantity $\sigma_G^2$ is given in (38).

The quantity $\sigma_G^2$ can be shown to be [5]

$$\sigma_G^2 = 3(\phi_{_R}^2(0))^2 + 6(\phi_{_R}(0)\phi_{_m}(0)\text{tr}[K_n]) + 3(\phi_{_m}(0)\text{tr}[K_n])^2$$

$$+ \left[ \phi_{_R}(0) - \phi_{_R}(0) + \phi_{_m}(0)\text{tr}[K_n] \right] \left[ 3(\phi_{_R}(0)\text{tr}[K_n]) + \phi_{_m}(0)\text{tr}[K_n] \right]$$

$$+ 2(\phi_{_R}(0) - \phi_{_R}(0) + \phi_{_m}(0)\text{tr}[K_n]) \right] (\phi_{_R}(0) + \phi_{_m}(0)\text{tr}[K_n]), \quad (38)$$

which, for high signal-to-noise ratios ($\phi_{_R}^2 = \phi_{_R}^2$), can be approximated by

$$\sigma_G^2 = 3(\phi_{_R}^2(0))^2 + 4\phi_{_R}(0)\phi_{_m}(0)\text{tr}[K_n] + 3(\phi_{_m}(0)\text{tr}[K_n])^2$$

$$+ \left[ \phi_{_R}(0) - \phi_{_R}(0) + \phi_{_m}(0)\text{tr}[K_n] \right] \left[ 3(\phi_{_R}(0)\text{tr}[K_n]) + \phi_{_m}(0)\text{tr}[K_n] \right] \right] (\phi_{_R}(0) + \phi_{_m}(0)\text{tr}[K_n]), \quad (39)$$

where $\text{tr}[*]$ is the trace operator, $K_n$ is the weight-noise correlation matrix defined in (32), $\phi(\tau)$ denotes a correlation between two random processes, the prime denotes a derivative with respect to $\tau$, and $\phi(0)(0)$ denotes $\phi(\tau)/\phi(0)\text{tr}[K_n]$ at $\tau = 0$.

Note that, in interpreting the propositions, it is important to keep in mind that the result is true if no false lock occurs; i.e., if no noise samples force the delay estimate to lock on a local solution, or if the adaptive filter does not compensate at all for the delay reference.

The steady-state delay estimate variance is given by

$$\nu_{\text{ss}} = \lim_{n \to \infty} E\left[ (d_n - D)^2 \right] = \frac{\alpha^2_n}{2\xi_{\text{min}} - 4\alpha^2_n}. \quad (40)$$

where $\alpha_n$ is given by

$$\alpha_n = -\left( \phi_{_R}(0) - \phi_{_R}(0) + \phi_{_m}(0)\text{tr}[K_n] \right) \left[ (\phi_{_R}(0) + \phi_{_m}(0)\text{tr}[K_n]) \right] \right] \right] (\phi_{_R}(0) + \phi_{_m}(0)\text{tr}[K_n]), \quad (41)$$

2. Results for the LMS adaptive filter in steady state

As in the case of the LMS delay tracking algorithm, the LMS weight vector adaptive algorithm of (26) can be analyzed in terms of convergence in the mean and the mean square of the weight vector estimate. That type of analysis has been performed by many authors, and the details concerning our problem can be found in [5]. Due to the assumptions made, in particular the instantaneous propagation of the adaptive delay value through the adaptive-filter delay line, the behaviour of the filter is not affected in many different ways by the delay element. The following two propositions characterize the convergence of the weight vector.

**Proposition 4.** In steady-state conditions, the weight vector estimator, given by the adaptive filter LMS algorithm operating jointly with a mean-square convergent delay tracking algorithm, converges in the mean if

$$0 < \mu < \frac{1}{\lambda_{\text{max}}} \quad (42)$$

where $\lambda_{\text{max}}$ denotes the maximum eigenvalue of the input signal autocorrelation matrix $R$. The weight vector estimate experiences a bias given by

$$b = 1/2 \nu_{\text{ss}} R^{-1} \phi(0), \quad (43)$$

where $\phi(0)$ represents the second derivative of the cross-correlation vector with respect to the delay $d_n$.

**Proposition 5.** In steady-state conditions, the weight vector estimator, given by the adaptive filter LMS algorithm operating jointly with a mean-square convergent delay tracking algorithm, is convergent in the mean if

$$0 < \mu < \frac{1}{\sum_i M_i \lambda_i} \quad (44)$$

where $\lambda_i$ is the $i^{th}$ eigenvalue of the $M \times M$ input signal autocorrelation matrix $R$.

Note that the convergence condition of (42) and (44) are identical to the usual conditions for convergence of an LMS adaptive filter [7], but that the effect of the delay estimator on the adaptive filter is to add a bias to the weight vector estimate.

3. Excess mean squared error and misadjustment with the joint LMS algorithm

From (8), the steady-state MSE function is

$$\xi_{\text{ss}} = \phi_{_R}(0) + E[w_n^T R w_n] - 2E[w_n^T p_n], \quad (45)$$

where the values of the estimates take on their steady-state form. Neglecting some terms involving the square of $\nu_{\text{ss}}$, we can transform (45) into

$$\xi_{\text{SS}} = \xi_{\text{min}} + \nu_{\text{ss}} \xi_{\text{min}} / 2 + \nu_{\text{ss}} \xi_{\text{min}} / 2 \text{tr}[R] \quad (46)$$

The excess MSE is given by the expression $\xi_{\text{ex}} = \xi_{\text{ss}} - \xi_{\text{min}}$, which can be transformed into

$$\xi_{\text{ex}} = \phi_{_R}(0) + \nu_{\text{ss}} \xi_{\text{min}} / 2 + \nu_{\text{ss}} \xi_{\text{min}} / 2 \text{tr}[R] \quad (47)$$

where the excess MSE specific to the adaptive delay element is defined as

$$\xi_{\text{ex}} = \frac{\nu_{\text{ss}} \xi_{\text{min}}}{2}. \quad (48)$$
the excess MSE specific to the adaptive filter is defined as
\[ \xi_{ef} = \frac{\mu \xi_{min} v_{ss} \text{tr} [R]}{1 - \mu \text{tr} [R]} \] (49)
and the cross-product excess MSE is defined as
\[ \xi_{eff} = \frac{\mu v_{ss} \text{tr} [R]}{2(1 - \mu \text{tr} [R])} \] (50)
The expression for \( \xi_{ef} \) is valid for pure LMS delay estimation [4], and the expression for \( \xi_{ef} \) is valid for an adaptive LMS filter operating without an adaptive delay [7].

The misadjustment is defined as the ratio of the excess MSE to \( \xi_{min} \). Therefore, the misadjustment expression can be shown to be
\[ M = M_d + M_f + M_d M_f, \] (51)
where \( M_f \) and \( M_d \) are the misadjustments specific to the adaptive delay element and to the adaptive filter respectively. They are obtained by dividing \( \xi_{ef} \) and \( \xi_{eff} \) by \( \xi_{min} \).

B. Discussion of the LMS algorithm algorithm

The joint steepest-descent algorithm and its stochastic counterpart, the joint LMS algorithm, represent the generalizations of either the conventional SD (LMS) delay tracking algorithm [4] or the conventional SD (LMS) adaptive transversal filter algorithm [14]. It is therefore not surprising to all the results concerning the delay algorithm degenerate to those of [4] when the signals are properly interpreted, and that the adaptive-filter derivations come down to the LMS adaptive-filter results when the delay, \( D \), and the variance are set equal to zero.

Another point to note is that, as long as the delay estimation algorithm is convergent in the mean square (the steady-state delay variance \( \nu_{ss} \) is finite), the conditions for convergence of the LMS adaptive filter are identical to the usual conditions for a similar adaptive filter operating alone or with a fixed delay element. The convergence depends on the eigenvalues of the input-signal autocorrelation matrix. Note also that, because of the adaptive delay element, the weight vector estimate is biased.

As (37) and (40) suggest, the convergence of the LMS adaptive delay element depends on \( \xi_{min} \), \( \sigma_v^2 \) and \( \sigma_r^2 \). Using the high SNR assumption \( \Phi_{rr} = \Phi_{rr} \) and the fact that
\[ \xi_{min} = \Phi_{rr}(0) - \Phi_{rr} \] (52)
(39) and (41) can take the form
\[ \sigma_C^2 = 3/4 \xi_{min}^2 - 1/2 \xi_{min} \xi_{eff} \]
\[ + \left[ \xi_{min}^2 \Phi_{ss} (0) - \xi_{min} \Phi_{ss} (0) - 2 \xi_{min}^2 \Phi_{ss} (0) \text{tr} [K_n] \right] \] (53)
and
\[ \sigma_r^2 = 2 \xi_{min} \xi_{eff} \]
\[ + \left[ \xi_{eff}^2 \Phi_{ss} (0) - 4 \xi_{min}^2 \Phi_{ss} (0) \text{tr} [K_n] \right] \] (54)

Equations (53) and (54) indicate that the convergence of the LMS adaptive delay element depends on the input signal power \( \Phi_{ss}(0) \) and the minimum MSE \( \xi_{min} \). If \( \nu_{ss} \) is small, \( \text{tr} [K_n] \) is small and we have \( \sigma_C^2 = 3/4 \xi_{min}^2 \) and \( \sigma_r^2 = 2 \xi_{min} \xi_{eff} \). We therefore see, from (36) and (37), that the upper bound for convergence in the mean square is about one-third of the upper bound for convergence in the mean. The steady-state delay variance is also approximately given by \( \nu_{ss} = \alpha \xi_{min} \).

The delay estimate variance is encountered in the excess MSE and misadjustment expressions, such as (47) and (51). Once the delay variance is computed or fixed, these two quantities are seen to be functions of two terms specific to the adaptive delay element and to the adaptive filter respectively, and of a cross-product term (note that, since the delay-specific term is a function of \( \nu_{ss} \), it is indirectly a function of the adaptive filter). The expressions for \( \xi_{ef} \) and \( \xi_{eff} \) are identical to those obtained for the respective adaptation algorithms operating alone [4], [7]. The cross-product terms, \( \xi_{ef} \) and \( \xi_{eff} \), are essentially the result of gradient and derivative estimation errors in the two adaptation processes. For stationary input and reference processes, the estimation noise in one adaptive algorithm is increased by the gradient estimation noise present in the other adaptive system. Therefore, the total misadjustment, \( M \), is not merely the sum of the adaptive delay element and adaptive filter misadjustment expression \( M_f \) and \( M_d \), but also includes a term due to the joint estimation noise. Note, however, that the cross-product misadjustment, \( \xi_{eff} \), is equal to the product of \( M_d \) and \( M_f \), making it a second-order term that, in practical situations, can be one order of magnitude smaller than the individual terms.

As a final remark, note that the key quantities in the analyses are \( \xi_{min} \) and its second derivative. These quantities can be estimated from a priori knowledge of the transmitted signal and from the estimation of the received signal's autocorrelation functions. Some possible estimation procedures are given in [5].

C. Experimental results with the joint LMS algorithm

Using the analysis results, it is possible to compute the adaptive delay gain factor, \( \alpha \), as a function of the adaptive filter gain factor, \( \mu \). In order to perform this task, we combine the expression for \( \nu_{ss} \) given in (40), with equations (53) and (54) and the expression for \( \text{tr} [K_n] \) given by [5]:
\[ \text{tr} [K_n] = \mu M_d \xi_{min} + \xi_{eff} / 2. \]

![Figure 3: Theoretical curve of a versus \( \mu \); SNR = 10 dB; small-dash curve: \( \nu_{ss} = 0.001 \); large-dash curve: \( \nu_{ss} = 0.01 \); continuous curve: \( \nu_{ss} = 0.1 \).](Note: The image contains a graph showing theoretical curves for various conditions.)

† Note that these expressions are exact for white input and noise signals.
Table 1
Excess mean squared errors and misadjustments for different combinations of α's and μ's.
The signal-to-noise ratio is 10 dB.

<table>
<thead>
<tr>
<th>μ</th>
<th>α</th>
<th>ξEE</th>
<th>ξRE</th>
<th>ξMSE</th>
<th>M</th>
<th>Mth</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.5</td>
<td>0.00312</td>
<td>0.00193</td>
<td>0.00563</td>
<td>40.5%</td>
<td>39.4%</td>
</tr>
<tr>
<td>0.05</td>
<td>0.5</td>
<td>0.00141</td>
<td>0.00193</td>
<td>0.00308</td>
<td>22.1%</td>
<td>25.4%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
<td>0.00141</td>
<td>0.00010</td>
<td>0.00313</td>
<td>22.5%</td>
<td>23.3%</td>
</tr>
<tr>
<td>0.01</td>
<td>0.5</td>
<td>0.00026</td>
<td>0.00193</td>
<td>0.00195</td>
<td>14.0%</td>
<td>16.0%</td>
</tr>
<tr>
<td>0.05</td>
<td>0.25</td>
<td>0.00141</td>
<td>0.00051</td>
<td>0.00163</td>
<td>11.7%</td>
<td>14.2%</td>
</tr>
</tbody>
</table>

Figure 4: Impulse response of the reverberant room.

Figure 5: LMS adaptive delay response to a reference delay ramp of 0.01 samples/sampling period and for a 200-s ramp impulse response: dashed curve: reference delay; μ = 0.01, α = 0.02. White Gaussian input.

Figure 6: LMS adaptive delay response to a sinusoidal reference delay variation and for a 200-s ramp impulse response: dashed curve: reference delay; μ = 0.01, α = 0.02. White Gaussian input.

Figure 7: Learning curve for the joint algorithm coping with a reference delay ramp of 0.01 samples/sampling period (corresponding to Fig. 5); μ = 0.01, α = 0.02.

Figure 8: Learning curve for the single adaptive filter coping with a reference delay ramp of 0.01 samples/sampling period (note the scale difference compared to Fig. 7); μ = 0.01. White Gaussian input.
Using 21-coefficient adaptive and reference filters with white input and noise signals and for a signal-to-noise ratio of 10 dB, we obtain the plots of Fig. 3.

The gain factor $\alpha$ increases with $v_n$ and for a typical variance of 0.01 the value of $\alpha$ is approximately constant with $\mu$, and is around 0.5. This indicates that, for low variance, the adaptive filter does not significantly influence the behaviour of the adaptive delay. The upper bound on $\alpha$ for convergence in the mean square (36) is not significantly influenced by the delay variance and is approximately constant for $\mu \leq 0.01$ (see [5]). These critical values for $\alpha$ and $\mu$ are retained as indicators of the values that should be used in the simulations.

An important result from the previous sections is the expression for the excess MSE at the output of the joint LMS algorithm given by (47). We verify these results by computing the theoretical value of $\xi_{\text{res}}$ using (49), and by obtaining $\xi_{\text{res}}$ as well as $\xi_{\text{r}}$ through simulations. The results, for five different combinations of $\alpha$ and $\mu$, are presented in Table 1 for low-order adaptive and reference filters (21 coefficients). The corresponding measured total misadjustment, $M$, is obtained from $\xi_{\text{res}}$ through division by $\text{SNR}_{\text{r}}$, while the theoretical total misadjustment, $M_{\text{th}}$, is obtained using (51). This table shows the good agreement between the measured and the theoretical quantities. Note that since the cross-product term $M'M$ is a second-order component, its effect is small or negligible, as can be seen from the fact that $\xi_{\text{res}}$ is always approximately equal to the sum of $\xi_{\text{r}}$ and $\xi_{\text{d}}$.

D. Results with a long reference impulse response

In practice, the reference-filter impulse response can exhibit a fairly large number of coefficients. For example, the typical impulse response associated with a reverberant room, and encountered in audio surveillance [18], has more than 200 coefficients. We generated such an impulse response using the method proposed by Allen and Berkley [19]. The response that we obtained simulates the audio channel between a source of sound and a microphone located in a closed room, with specific wall-reflection coefficients. We arbitrarily selected the parameters to simulate the behaviour of a room measuring 6 m by 6 m, with a height of 3 m. The reflection coefficient for each wall is 0.8, the source sound is assumed to be located about 0.5 m away from one of the corners, and the location of the receiver is about 1 m from the same corner. The corresponding impulse response is given in Fig. 4. Note that the response is not symmetrical with respect to any point, and that it exhibits three large reflection peaks as well as five smaller ones. This reference impulse response is used in a system identification configuration (see Fig. 2), with a 200-coefficient adaptive filter and with spectrally white Gaussian and audio input signals.

With a white Gaussian input, the delay tracking of the joint algorithm is shown in Figs. 5 and 6, for a reference delay ramp and a sinusoidal reference delay in noiseless conditions. The rate of change of the linear delay is higher than what is typically encountered in audio surveillance [18].

The inaccuracies in the delay estimation are related to the excess MSE, which is proportional to the number of coefficients in the adaptive filter (see [49]). Note the different behaviour of positive and negative delay tracking, especially in Fig. 6. This difference is related to the fact that the reference impulse response is not symmetrical with respect to any of its points. In order to appreciate the effectiveness of the joint algorithm, the learning curve corresponding to the joint algorithm coping with a linearly changing delay (corresponding to Fig. 5) is illustrated in Fig. 7. The learning curve, corresponding to a system identification configuration in which there is no adaptive delay, i.e., a configuration in which the adaptive filter alone copes with the modeling of both the linear reference delay and the reference filter, is illustrated in Fig. 8. These curves were obtained by averaging 10 error curves. Note the scale difference between Fig. 7 and Fig. 8. It is obvious from these figures that the joint algorithm generates an MSE lower than that for the single adaptive filter. Similar results can be obtained with a speech input, although the joint LMS algorithm must be normalized in this case to take into account the power variations in the input signal [5].

V. Conclusions

In this article, we have studied the joint SD and joint LMS algorithms for time-delay estimation and adaptive filtering. The presence of a multitude of stationary points in the objective function was established, and the steady-state behaviour of the two algorithms was investigated. The coupling between the two LMS adaptive algorithms was shown to give a better behaviour with a squared error term equal to the sum of the individual misadjustments plus a cross-product term. The analyses were used to obtain a theoretical view of the application of such algorithms in specific environments. The simulation results that we provided give a flavour of the sort of behaviour that one can expect when using an adaptive delay element with a conventional adaptive filter. When a long impulse response filter has to be estimated, the additional computations incurred by the LMS adaptive delay algorithm are not significant, especially given the reduction in excess MSE that is attainable. This conclusion supports the use of the joint algorithm when the main input and the reference signal are believed to exhibit some form of non-synchronous behaviour.

Acknowledgements

This work was performed with the financial assistance of the Canadian Department of Communications and of the Natural Sciences and Engineering Research Council of Canada. The provision of computing equipment by Bell-Northern Research, l'Institut National de la Recherche Scientifique and the Communications Research Centre is also greatly acknowledged. The comments of the various reviewers are appreciated.

References