

## JOINT GRADIENT-BASED TIME DELAY ESTIMATION AND ADAPTIVE FILTERING

Daniel Boudreau<sup>1,3</sup> and Peter Kabal<sup>1,2</sup><sup>1</sup>Electrical Engineering  
McGill University  
Montreal, Quebec H3A 2A7<sup>2</sup>INRS-Télécommunications  
Université du Québec  
Verdun, Quebec H3E 1H6<sup>3</sup>Communications Research Centre  
Department of Communications  
Ottawa, Ontario K2H 8S2**Abstract**

A general estimation model is defined in which two observations are available; one being a noisy version of the transmitted signal, while the other is a noisy filtered and delayed version of the same transmitted signal. The time-varying delay and the filter are unknown quantities that must be estimated. A joint estimator is proposed. It is composed of an adaptive delay element in conjunction with a transversal adaptive filter. The same error signal is used by the two adaptive algorithms to adjust the delay element and the filter such that the minimum mean squared error is attained. Two joint gradient-based adaptation algorithms are studied. The joint steepest-descent (SD) algorithm is first investigated. The possibility of convergence to a multitude of solutions is established and a condition of convergence is presented. A stochastic implementation of the joint SD algorithm, under the form of a joint least mean square (LMS) algorithm, is then investigated. It is analyzed in terms of convergence in the mean and in the mean square of both the delay estimate and the adaptive filter weight vector estimate. The conditions of convergence of the joint LMS algorithm are established as functions of the power spectral densities of the observed signals and the minimum mean squared error.

**1. Introduction**

Estimating the time delay between two continuous versions of the same signal, each corrupted by uncorrelated noise components, has been the subject of many research efforts over recent years. In this paper, we consider the more difficult model in which the delayed path is subject to frequency dependent attenuation. The corresponding discrete-time model is of the form

$$\begin{aligned} y_1(n) &= s(n) + v_1(n) \\ y_2(n) &= L_{D_n, h(n)}[s(n)] + v_2(n), \end{aligned} \quad (1)$$

where  $s(n)$  is the transmitted signal,  $D_n$  is a delay, possibly time-varying, and  $L_{D_n, h(n)}[s(n)]$  is an unknown linear operator, taking the form of a filtering operation, with the filter impulse response  $h(n)$ , of a delayed by  $D_n$  version of  $s(n)$ . The signals  $v_1(n)$  and  $v_2(n)$  are zero-mean noise signals, assumed uncorrelated with each other as well as with  $s(n)$  and  $L_{D_n, h(n)}[s(n)]$ . Examples of such systems are encountered in system 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, as is the case in geophysical exploration, echo cancelling or multipath communication channels.

This paper presents an analysis of joint delay estimation and system estimation based on the minimum mean squared error (MMSE) performance index, when the system estimation is specifically performed by an adaptive transversal filter and the delay estimation is accomplished independently from this filter, by an adaptive delay element. A joint steepest-descent algorithm and a joint LMS algorithm are investigated in tracking mode. The assumption of tracking mode is important because it is then assumed acquisition was initially performed, in order to bring the estimate close to its optimum value.

The principal contributions of this paper are the generalization of existing gradient-based time delay estimation without reference filtering  $h(n)$ , as proposed in [1] and the analysis of a new joint algorithm for the synchronization of the input and the reference signals used by an adaptive filter. The joint steepest-descent and LMS algorithms that we investigate in this paper are generalizations of joint clock phase recovery and adaptive equalization based on LMS phase tracking, as proposed in [2] (see also [3]), since we allow not only the sampling phase, but also the sampling period to be tracked and since we consider general signals. Our joint algorithms are not based on the assumption that the input signal and the reference signal fed to an adaptive filter are sampled with 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 itself is free to perform the task of modeling the linear filter  $h(n)$  or its inverse.

**2. The General Theory**

Depending on the problem at hand, the operator  $L_{D_n, h(n)}[s(n)]$  can take the form of the filtering of a delayed version of  $s(n)$  or the form of a filter followed by a delay. The former configuration is defined as a Type I system and the latter as a Type II system. These two definitions also apply to the joint adaptive estimator. Note that the two types of systems are equivalent if the corresponding delay is constant with time.

The adaptive filter is a finite length transversal filter, with weight vector  $w_n$ . It is desired that the reference delay value  $D_n$  be estimated separately from the adaptive filter, by an adaptive delay element  $d_n$  cascaded with the filter in Type I or Type II form. In joint MMSE delay estimation and adaptive filtering, the mean squared error (MSE) surface is searched by both the adaptive filter estimation algorithm and the delay estimation one. In system identification (cancellation) scenarios,  $y_1(n)$  is filtered by an estimate of  $L_{D_n, h(n)}[s(n)]$  and the resulting signal is subtracted from  $y_2(n)$  in order to form the error signal. In inverse filtering (equalization),  $y_2(n)$  is passed through an estimate of  $L_{-D_n, h^{-1}(n)}[s(n)]$  and compared to  $y_1(n)$ . This illustrated in Figs. 1 and 2. Figs. 3 and 4 give respectively a detailed form of Type I and Type II systems in cancellation configuration.

In general, the output of the adaptive branch can be defined as  $y(n)$  and the reference signal as  $r(n)$ . Then the error signal is defined as

$$e(n) = r(n) - y(n), \quad (2)$$

and the MSE function, at time  $n$ , as

$$\xi_n = E[|e(n)|^2]. \quad (3)$$

The joint estimation can be thought of as taking place into a vector space made of a weight vector subspace and a delay subspace. The MSE function can take the following two forms, depending on the type of system

$$\xi_n = \begin{cases} \phi_{rr}(n, n) + w_n^H R_n w_n - 2\text{Re}[w_n^H p_n] & \text{Type I} \\ \phi_{rr}(nT + d_n, nT + d_n) + w_n^H R w_n - 2\text{Re}[w_n^H p_n] & \text{Type II,} \end{cases} \quad (4)$$

where the superscript  $H$  denotes complex conjugate transpose and  $\text{Re}(\cdot)$  is the real value operator,  $T$  is the sampling period,  $\phi_{rr}(n, m)$  is the autocorrelation function of the reference signal for the specific problem at hand. In a Type I system,  $R_n$  is the autocorrelation matrix of  $u(nT - d_n)$ , the delayed adaptive branch input. In a Type II system,  $R$  is the autocorrelation matrix of the adaptive filter input  $u(n)$ . The vector  $p_n$  is the cross-correlation vector between the same inputs and the reference signal. In a Type I system, each component of the adaptive filter input vector can experience a different delay, and the autocorrelation matrix is time-varying. In a Type II system, the delay  $d_n$  is assumed transferred to the reference branch, such that there is no delay between the adaptive filter output and the error signal. Note also that the vector  $p_n$  depends on  $d_n$  in both types of systems.

In the weight vector subspace, the MSE function is a quadratic surface [4]. The delay subspace is one-dimensional and, in general, the MSE function is not unimodal with respect to  $d_n$ . In order to see this, note that  $\xi_n$  depends on two correlation functions that vary according to the adaptive filter and the operator  $L[s(n)]$ , as well as the autocorrelation function of the signal  $u(n)$ . All of these time functions are multimodal with respect to  $d_n$ , which in turn causes the MSE function to behave similarly and produce a multitude of local extrema.

**3. The Joint Steepest-Descent Algorithm**

The simplest joint gradient algorithm is the first order one, which takes the general form

$$\begin{aligned} w_{n+1} &= w_n - \mu \nabla_{w_n} \xi_n \\ d_{n+1} &= d_n - \alpha \frac{\partial \xi_n}{\partial d_n}, \end{aligned} \quad (5)$$

where  $\mu$  and  $\alpha$  are small positive adaptation constants. This algorithm is referred to as the joint steepest-descent (SD) algorithm. It is composed of the usual SD

adaptive filter and the SD adaptive delay element, each one adapted, at iteration  $n+1$ , using the MSE function at iteration  $n$ . The two adaptive systems can be cascaded in either Type I or Type II, and a system identification or an inverse filtering configuration can be used.

### 3.1 Convergence of the Algorithm

A necessary condition for a specific  $d_n$  and  $\mathbf{w}_n$  to be a stationary solution of (5) is that both of the following equations be satisfied [5]

$$\begin{cases} \nabla_{\mathbf{w}_n} \xi_n = 0 \\ \frac{\partial \xi_n}{\partial d_n} = 0. \end{cases} \quad (6)$$

Assuming that both the adaptive and the reference systems are Type II, a necessary condition for  $(d_n, \mathbf{w}_n)$  to be a stationary solution of (5) is

$$\begin{cases} \mathbf{w}_n = \mathbf{R}^{-1} \mathbf{p}_n \\ \text{Re}\{\mathbf{w}_n^H \mathbf{p}_n\} = 0, \end{cases} \quad (7)$$

i.e. the stationary weight vector solution is the Wiener solution when the delay  $d_n$  is such that  $\mathbf{w}_n$  is orthogonal to  $\mathbf{p}_n$  or the product  $\mathbf{w}_n \mathbf{p}_n$  is purely imaginary. Note that the solution of (7) is not unique.

Assume, for general systems, that the adaptation constants can be time-varying and denote them as  $\mu_n$  and  $\alpha_n$ . Express also the MSE as an explicit function of  $d_n$  and  $\mathbf{w}_n$ , i.e. as  $\xi\{d_n, \mathbf{w}_n\}$ . Define a stationary point of  $\xi\{d_n, \mathbf{w}_n\}$  as a solution  $(d_n, \mathbf{w}_n)$  of the necessary condition (6). Define  $\vartheta_n$  as the delay value closest to  $d_n$ , for which  $\xi\{d_n, \mathbf{w}_n\}$  is minimum. Then a truncated Taylor expansion of the MSE function around  $\vartheta_n$  allows one to approximately express the SD delay algorithm as

$$d_{n+1} = d_n - \alpha_n (d_n - \vartheta_n) \ddot{\xi}\{d_n, \mathbf{w}_n\}, \quad (8)$$

where  $\ddot{\xi}\{d_n, \mathbf{w}_n\}$  denotes the second derivative of  $\xi\{d_n, \mathbf{w}_n\}$ , with respect to  $d_n$ , evaluated at  $d_n = \vartheta_n$ . The next proposition is a modification of proposition 2 of [5] and establishes the adaptation constants range for a stationary point to be attained by the joint SD algorithm, when  $d_n$  and  $\mathbf{w}_n$  are modified in some alternate fashion.

**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  $\lambda_{\max}(n)$  be the maximum eigenvalue of the signal autocorrelation matrix  $\mathbf{R}_n$  and  $\vartheta_n$ , the delay value closest to  $d_n$ , for which  $\xi\{d_n, \mathbf{w}_n\}$  is minimum. Assuming that  $d_n$  is sufficiently close to  $\vartheta_n$ , the MSE will converge to a stationary point if

$$0 < \mu_n < \frac{1}{\lambda_{\max}(n)}, \quad (9)$$

for  $n \in \kappa_1$ , and

$$0 < \alpha_n < 2 \left[ \frac{\partial^2}{\partial d_n^2} \xi\{d_n, \mathbf{w}_n\} \right]^{-1}, \quad (10)$$

for  $n \in \kappa_2$ . ■

This proposition states that  $d_n$  and  $\mathbf{w}_n$  may be adjusted in any alternating fashion and the MSE will converge to a stationary point if  $\mu_n$  satisfies (9) during the adjustment of  $\mathbf{w}_n$ , and  $\alpha_n$  satisfies (10) during the adjustment of  $d_n$ . For example,  $d_n$  may be fixed and  $\mathbf{w}_n$  adjusted until  $\nabla_{\mathbf{w}_n} \xi = 0$ , then  $\mathbf{w}_n$  may be fixed and  $d_n$  adjusted until  $\partial \xi / \partial d_n = 0$ . The cycle can be repeated until both  $\nabla_{\mathbf{w}_n} \xi$  and  $\partial \xi / \partial d_n$  approach zero. 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.

### 3.2 The Steepest-Descent Delay Tracking Algorithm

The particular case treated in this section is such that the adaptive filter has fully adapted to the characteristics of  $h(n)$  and is at least as long as the impulse response  $h(n)$ . For high signal-to-noise ratios, the  $i^{\text{th}}$  adaptive filter coefficients  $w_{ni}$ , at iteration  $n$ , is approximately of the form

$$w_{ni} \approx \begin{cases} h(i) & \text{System identification (cancellation)} \\ h^{-1}(i) & \text{Inverse filtering (equalization)}. \end{cases} \quad (11)$$

where  $h(i)$  is the  $i^{\text{th}}$  weight of the reference path filter. Since we assume the algorithm in tracking mode, the delay tracking algorithm can be linearized by using a truncated Taylor expansion of the MSE function around  $d_n = D_n$ . The small error SD delay algorithm is then of the form

$$d_{n+1} = d_n - \alpha \ddot{\xi}_n (d_n - D_n), \quad (12)$$

where  $\ddot{\xi}_n$  is equal to  $\partial^2 \xi_n / \partial d_n^2$  evaluated at  $d_n = D_n$ . Assume a restricted class of system in which  $\ddot{\xi}_n$  is time-independent and equal to  $\ddot{\xi}$ . This category is not too restrictive and is applicable to systems in which the reference filter  $h(n)$  varies slowly. The stability range for  $\alpha$  is

$$0 < \alpha < 2/\ddot{\xi}, \quad (13)$$

and the time constant of delay adaptation is

$$\tau_{del} \approx \frac{1}{\alpha \ddot{\xi}}. \quad (14)$$

Note that condition (13) is equal to condition (10) with  $\vartheta_n = D_n$ .

For Type II, we can obtain a tighter bound on  $\alpha$ . In this case, the MSE function has only one delay dependent term equal to  $\phi_{yr}(n, n)$ , the correlation between the adaptive branch output and the reference signal. Then  $\ddot{\xi} = \phi_{yr}(n, n)$  and for the cancellation and equalization configurations, this cross-correlation is given respectively by

$$\phi_{yr}^{(C)}(n, n) = -2\text{Re} \left[ \sum_{\ell} \rho(\ell) \phi_{ss}(-\ell T + D_n - d_n) \right] \quad (15)$$

and

$$\phi_{yr}^{(E)}(n, n) = -2\text{Re}[\phi_{ss}(D_n + d_n)], \quad (16)$$

where  $\phi_{ss}(\tau)$  is the autocorrelation function of the signal  $s(n)$  and  $\rho(k)$  is the deterministic autocorrelation of the reference filter impulse response and is defined as

$$\rho(k) = \sum_i h(k+i)h^*(i). \quad (17)$$

Defining the maximum value of the input signal power spectral density  $\Phi_{ss}(e^{j\omega})$  as  $\Phi_{\max}$ , the following sufficient bounds can be derived

$$0 < \alpha < -1/\Phi_{\max} \text{Re}\{\rho''(0)\} \quad \text{Cancellation} \quad (18)$$

and

$$0 < \alpha < \frac{3}{\Phi_{\max} \pi^2} \quad \text{Equalization}. \quad (19)$$

The prime in (18) denotes the derivative with respect to the continuous-time correlation argument.

## 4. The Joint LMS Algorithm

In practice, the gradient and derivative functions used in the SD algorithm have to be estimated. In the joint LMS algorithm, the MSE function  $\xi_n$  is estimated by the magnitude of the squared error  $|e(n)|^2$ . For a Type I configuration, the joint SD algorithm then becomes the joint LMS algorithm,

$$\mathbf{w}_{n+1} = \mathbf{w}_n + 2\mu e^*(n) \mathbf{u}_n \quad (20)$$

$$d_{n+1} = d_n + 2\alpha \text{Re} \left[ e^*(n) \frac{\partial y(n)}{\partial d_n} \right], \quad (21)$$

where  $\mathbf{u}_n$  is the vector of delayed samples at the adaptive filter input.

The delay and weight vector estimates being random variables, we can analyze the joint algorithm 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 noise will affect the delay tracking and the derivative noise will itself influence the adaptive filter. These mutual effects can be included in the delay variance and weight vector covariance matrix, in steady-state conditions. The bounds for  $\mu$  and  $\alpha$  are determined, for both types of convergence. In the course of the analyses, it is assumed that all the signals and systems are real, that the input signals are zero-mean Gaussian processes, that the adaptive system is Type I and in steady-state and that the reference system is stationary ( $D_n = D$ ). It is also assumed that  $\mathbf{u}_n = \mathbf{u}(nT - d_n)$ , i.e. that any adaptive delay modification is reflected on every sample of the adaptive filter delay line. This simplifies the analyses by making the input signal autocorrelation matrix time-invariant. It is further assumed that independence theory holds, i.e. (see [4] for example)  $E[\mathbf{u}(n)\mathbf{u}^T(k)] = 0$  for  $k = 0, 1, \dots, n-1$  and  $E[\mathbf{u}(n)r(k)] = 0$  for  $k = 0, 1, \dots, n-1$ . We also assume that in steady-state, the adaptive weight vector  $\mathbf{w}_n$  can be expressed as  $\mathbf{w}_n = \mathbf{w}_{\text{opt}} + \boldsymbol{\eta}_n$ , where  $\mathbf{w}_{\text{opt}}$  is the optimum Wiener solution given by  $\mathbf{w}_{\text{opt}} = \mathbf{R}^{-1} \mathbf{p}_n|_{d_n=D}$ , and  $\boldsymbol{\eta}_n$  is a zero-mean Gaussian noise vector independent of the data and such that  $E[\boldsymbol{\eta}_n \boldsymbol{\eta}_k^T] = 0$  for  $n \neq k$ , and  $E[\boldsymbol{\eta}_i \boldsymbol{\eta}_j] = 0$  for  $i \neq j$ . The noise vector covariance matrix, defined as  $\mathbf{K}_\eta = E[\boldsymbol{\eta}_n \boldsymbol{\eta}_n^T]$ , is therefore diagonal with the values  $E[\eta_i^2(n)]$  on the main diagonal.

We present the analysis for the delay estimator in Section 4.1, the analysis for the weight vector estimator in Section 4.2 and we combine the results in Section 4.3, in order to obtain the total excess MSE and misadjustment for the joint LMS algorithm.

### 4.1 Analysis for the LMS Delay Estimator in Tracking Mode

The LMS delay tracking algorithm is analyzed in terms of convergence of the delay estimate, in the mean and in the mean square. The following analysis parallels and extends that of Messer [1].

For  $d_n = D$ , the output of the adaptive branch can be expressed as

$$y(n, D) = \mathbf{w}_{\text{opt}}^T \mathbf{u}(nT - D) + \boldsymbol{\eta}_n^T \mathbf{u}(nT - D). \quad (22)$$

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. Define

$e_{\min}(n, D)$  as the error between the optimum adaptive branch and the reference branch, i.e.

$$e_{\min}(n, D) = r(n) - \hat{r}(n), \quad (23)$$

and the corresponding MSE as

$$\xi_{\min} = E[e_{\min}^2(n, D)]. \quad (24)$$

In tracking conditions, the squared error can be approximated by the first three terms of its Taylor expansion around  $d_n = D$ . Then, for real signals,

$$e^2(n, d_n) \approx e^2(n, D) - 2(d_n - D)e(n, D)\dot{y}(n, D) + (d_n - D)^2[\dot{y}^2(n, D) - e(n, D)\ddot{y}(n, D)]. \quad (25)$$

Defining the derivative noise as

$$N_n = -2e(n, D)\dot{y}(n, D) \quad (26)$$

and the quantity  $G_n$  as

$$G_n = \dot{y}^2(n, D) - e(n, D)\ddot{y}(n, D), \quad (27)$$

the LMS algorithm is therefore approximately expressed as

$$d_{n+1} = d_n - 2\alpha(d_n - D)G_n - \alpha N_n. \quad (28)$$

#### 4.1.1 Convergence in the Mean of the Delay Estimate

Take the expected value with respect to the input signals, on both sides of (28), and rearrange. The result is

$$E[d_{n+1}] = E[(1 - 2\alpha G_n)d_n] + 2D\alpha E[G_n] - \alpha E[N_n]. \quad (29)$$

It can be shown, under the independence assumption, that the random variable  $d_n$  is uncorrelated with  $G_n$  and  $N_n$ . Furthermore, we have

$$\begin{aligned} E[G_n] &= -\phi''_{rr}(0) \\ E[N_n] &= 0, \end{aligned} \quad (30)$$

where the prime denotes a derivative with respect to the correlation argument. Equation (29) simplifies to

$$E[d_{n+1}] = (1 + 2\alpha\phi''_{rr}(0))E[d_n] - 2D\alpha\phi''_{rr}(0). \quad (31)$$

It can be shown that  $\xi_{\min} = -2\phi''_{rr}(0)$  and (31) can be written as

$$E[d_{n+1}] = (1 - \alpha\xi_{\min})E[d_n] + D\alpha\xi_{\min}, \quad (32)$$

which exhibits a form similar to the SD delay tracking algorithm of (12), with  $\xi_n = \xi_{\min}$ . Equation (32) converges if  $|1 - \alpha\xi_{\min}| \leq 1$ , and from the above derivations, the following proposition emerges.

**Proposition 2.** *In steady-state conditions and under our assumptions, the delay estimator, given by the LMS delay tracking algorithm operating jointly with an adaptive filter in Type I configuration, is an unbiased estimator if*

$$0 < \alpha < \frac{2}{\xi_{\min}} = -\frac{1}{\phi''_{rr}(0)}. \quad (33)$$

#### 4.1.2 Convergence in the Mean Square of the Delay Estimate

Subtract the value  $D$  from each side of (28), square each side of the resulting equation and take the expected value. Defining the time-varying delay estimate variance  $v_n$  as

$$v_n = E[(d_n - D)^2], \quad (34)$$

the final result is

$$v_{n+1} = E[(1 - 2\alpha G_n)^2]v_n + \alpha^2 E[N_n^2]. \quad (35)$$

Equation (35) indicates that there is convergence in the mean square sense if

$$|E[(1 - 2\alpha G_n)^2]| \leq 1. \quad (36)$$

Using the result of (30), the expected value is equal to

$$\begin{aligned} E[(1 - 2\alpha G_n)^2] &= 1 - 4\alpha E[G_n] + 4\alpha^2 E[G_n^2] \\ &= 1 + 4\alpha\phi''_{rr}(0) + 4\alpha^2 E[G_n^2]. \end{aligned} \quad (37)$$

The value of  $E[G_n^2]$  can be shown to be

$$\begin{aligned} E[G_n^2] &= 3(\phi''_{rr}(0))^2 + 4\phi''_{rr}(0)\phi''_{uu}(0)\text{tr}[\mathbf{K}_\eta] + 9(\phi''_{uu}(0))^2\text{tr}[\mathbf{K}_\eta^2] \\ &\quad + (\phi_{rr}(0) - \phi_{rr}(0) + \phi_{uu}(0)\text{tr}[\mathbf{K}_\eta])(\phi_{rr}^{(4)}(0) + \phi_{uu}^{(4)}(0)\text{tr}[\mathbf{K}_\eta]) \\ &\quad + 3(\phi''_{uu}(0)\text{tr}[\mathbf{K}_\eta])^2, \end{aligned} \quad (38)$$

where  $\text{tr}[\cdot]$  is the trace operator and  $\mathbf{K}_\eta$  is the weight noise covariance matrix and  $\phi^{(4)}(0)$  denotes  $\partial^4\phi(\tau)/\partial\tau^4$  at  $\tau = 0$ .

The condition (36) leads to the following proposition.

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

$$0 < \alpha < \frac{-\phi''_{rr}(0)}{E[G_n^2]} = \frac{\xi_{\min}}{2E[G_n^2]}, \quad (39)$$

where the quantity  $E[G_n^2]$  is given in (38). ■

Because the expected values in (35) are time-invariant, the steady-state delay estimate variance is given by

$$\begin{aligned} v_{\text{ss}} &= \lim_{n \rightarrow \infty} v_n \\ &= \frac{\alpha^2 E[N_n^2]}{1 - E[(1 - 2\alpha G_n)^2]} \\ &= \frac{-\alpha E[N_n^2]}{4\phi''_{rr}(0) + 4\alpha E[G_n^2]}, \end{aligned} \quad (40)$$

where  $E[N_n^2]$  can be shown to be

$$E[N_n^2] = -4(\phi_{rr}(0) - \phi_{rr}(0) + \phi_{uu}(0)\text{tr}[\mathbf{K}_\eta])(\phi''_{rr}(0) + \phi''_{uu}(0)\text{tr}[\mathbf{K}_\eta]). \quad (41)$$

Note that the steady-state variance is approached at the fastest rate when the quantity  $E[(1 - 2\alpha G_n)^2]$  in (35) is minimum. This happens when the adaptation constant is

$$\alpha_{\text{opt}} = \frac{-\phi''_{rr}(0)}{2E[G_n^2]}, \quad (42)$$

which is one half the maximum adaptation constant allowed by (39).

#### 4.2 Analysis for the LMS Adaptive Filter in Tracking Mode

As with the LMS delay tracking algorithm, the LMS weight vector adaptive algorithm can be analyzed in terms of convergence in the mean and the mean square of the weight vector estimate.

##### 4.2.1 Convergence in the Mean of the Weight Vector Estimate

Take the expected value of each side of (20) (for real signals). The result is

$$\begin{aligned} E[\mathbf{w}_{n+1}] &= E[\mathbf{w}_n] + 2\mu E[e(n, d_n)\mathbf{u}_n] \\ &= E[\mathbf{w}_n] + 2\mu(E[r(n)\mathbf{u}_n] \\ &\quad - E[\mathbf{u}_n\mathbf{u}_n^T\mathbf{w}_n]). \end{aligned} \quad (43)$$

Making use of independence theory, (43) can be expressed as [6]

$$E[\mathbf{w}_{n+1}] = E[\mathbf{w}_n] + 2\mu(E[\mathbf{p}_n] - \mathbf{R}E[\mathbf{w}_n]), \quad (44)$$

where  $\mathbf{p}_n$  is now a conditional expectation, conditioned on  $d_n$  and  $E[\mathbf{p}_n]$  denotes the expectation with respect to  $d_n$ . It can be shown that, in steady-state,

$$E[\mathbf{p}_n] \approx \mathbf{R}\mathbf{w}_{\text{opt}} + 1/2v_{\text{ss}}\ddot{\mathbf{p}}(D), \quad (45)$$

where  $\ddot{\mathbf{p}}(D)$  denotes the second derivative with respect to  $d_n$  of  $\mathbf{p}_n$ , when  $d_n = D$ . Equations (44) and (45) indicate that the weight vector is biased and we have the following proposition:

**Proposition 4.** *In steady-state conditions and under our assumptions, the weight vector estimator, given by the adaptive filter LMS algorithm operating jointly with a delay tracking algorithm in Type I configuration, converges in the mean if*

$$0 < \mu < \frac{1}{\lambda_{\text{max}}}, \quad (46)$$

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

$$\mathbf{b} = v_{\text{ss}}\mathbf{R}^{-1}\ddot{\mathbf{p}}(D). \quad (47)$$

Note that the convergence condition of (46) is identical to the usual condition for convergence in the mean of an LMS adaptive filter [4].

##### 4.2.2 Convergence in the Mean Square of the Weight Vector Estimate

The weight noise vector covariance matrix  $\mathbf{K}_\eta(n+1)$ , at iteration  $n+1$ , is computed in this section and a condition for its convergence, in the matrix norm sense, to a finite steady-state value is established. From (20) and the definition of  $\boldsymbol{\eta}_n$ , the noise vector can be written as

$$\begin{aligned} \boldsymbol{\eta}_{n+1} &= \mathbf{w}_{n+1} - \mathbf{w}_{\text{opt}} \\ &= \mathbf{w}_n + 2\mu e(n, d_n)\mathbf{u}_n - \mathbf{w}_{\text{opt}} \\ &= [\mathbf{I} - 2\mu\mathbf{u}_n\mathbf{u}_n^T]\boldsymbol{\eta}_n \\ &\quad + 2\mu[\mathbf{u}_n r(n) - \mathbf{u}_n\mathbf{u}_n^T\mathbf{w}_{\text{opt}}]. \end{aligned} \quad (48)$$

Then, proceeding as in [4],  $\mathbf{K}_\eta(n+1)$  can be computed to be

$$\begin{aligned} \mathbf{K}_\eta(n+1) &= \mathbf{K}_\eta(n) - 2\mu[\mathbf{K}_\eta(n)\mathbf{R} + \mathbf{R}\mathbf{K}_\eta(n)] \\ &\quad + 4\mu^2\mathbf{R}\text{tr}[\mathbf{R}\mathbf{K}_\eta(n)] + 4\mu^2\mathbf{R}[\xi_{\min} + \xi_{\min}v_n/2]. \end{aligned} \quad (49)$$

Except for the term involving the delay estimate variance, Eq. (49) is identical to the one for an adaptive filter operating alone ([4], Eq. (5.74)). Proceeding as in [4], the input signal autocorrelation matrix  $\mathbf{R}$  can be transformed to normal form, using a unitary similarity transformation  $\mathbf{Q}$ . Defining the transformed covariance matrix  $\mathbf{X}(n)$  as

$$\mathbf{X}(n) = \mathbf{Q}^T \mathbf{K}_\eta(n) \mathbf{Q},$$

the norm of the covariance matrix  $\mathbf{K}_\eta(n)$  converges to a finite value if and only if the trace of  $\mathbf{X}(n)$  converges to a finite value. For a convergent in the mean square delay estimator, it can be shown that  $\text{tr}[\mathbf{X}(n)]$  converges to the steady-state value

$$\lim_{n \rightarrow \infty} \text{tr}[\mathbf{X}(n)] = \text{tr}[\mathbf{K}_\eta] = \mu(L+1) \frac{\xi_{\min} + \xi_{\min} v_{ss}/2}{1 - \mu \text{tr}[\mathbf{R}]} \quad (50)$$

if and only if the parameter  $\mu$  satisfies the condition  $0 < \mu < 1/\text{tr}[\mathbf{R}]$ . Therefore, if the delay estimate variance  $v_n$  is finite, the trace of the weight-error covariance matrix  $\mathbf{K}_\eta$  is finite and the condition for convergence in the mean square is given in the following proposition.

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

$$0 < \mu < \frac{1}{\sum_{i=1}^{L+1} \lambda_i}, \quad (51)$$

where  $\lambda_i$  is the  $i^{\text{th}}$  eigenvalue of the  $(L+1) \times (L+1)$  input signal autocorrelation matrix  $\mathbf{R}$ . ■

This condition for convergence in the mean square sense is identical to the one for an adaptive filter operating alone.

If the adaptation constant  $\mu$  is small enough to make  $\mu \text{tr}[\mathbf{R}] \ll 1$ , then (50) can be written as

$$\begin{aligned} \text{tr}[\mathbf{K}_\eta] &\approx \mu(L+1)(\xi_{\min} + \xi_{\min} v_{ss}/2) \\ &= \text{tr}[\mathbf{K}'_\eta] + \mu(L+1)\xi_{\min} v_{ss}/2, \end{aligned} \quad (52)$$

where  $\text{tr}[\mathbf{K}'_\eta]$  is defined as the trace of the weight-error covariance matrix specific to the adaptive filter and is given as

$$\text{tr}[\mathbf{K}'_\eta] = \mu(L+1)\xi_{\min}. \quad (53)$$

Combining (38), (41) and (52) in (40), one obtains (with great pain) a closed-form expression for the steady-state delay estimate variance  $v_{ss}$ . This expression can in turn be used to determine the adaptation constants as explicit functions of the input signals statistics. An easier and more practical approach is to fix *a priori* the steady-state delay estimate variance  $v_{ss}$  at a certain acceptable level. Then, the quantities  $E[G_n^2]$  and  $E[N_n^2]$  are available as functions of  $\mu$ . Using these results in (40) gives  $\alpha$  as a function of  $\mu$ . The adaptation constants can then be determined, with the help of (39) and (51), to insure convergence in the mean square.

#### 4.3 Excess Mean-Squared Error and Misadjustment with the Joint LMS Algorithm

From (20), the MSE function is, for real signals and under our usual assumptions,

$$\xi = \phi_{rr}(0) + E[\mathbf{w}_n^T \mathbf{R} \mathbf{w}_n] - 2E[\mathbf{w}_n^T \mathbf{p}_n]. \quad (54)$$

The use of the definitions of  $\eta_n$  gives the steady-state MSE

$$\xi_{ss} = \xi_{\min} + \frac{v_{ss} \xi_{\min}}{2} + \frac{\mu(\xi_{\min} + \xi_{\min} v_{ss}/2) \text{tr}[\mathbf{R}]}{1 - \mu \text{tr}[\mathbf{R}]} \quad (55)$$

The excess MSE is then

$$\begin{aligned} \xi_{\text{ex}} &= \xi_{ss} - \xi_{\min} \\ &= \frac{v_{ss} \xi_{\min}}{2} + \frac{\mu(\xi_{\min} + \xi_{\min} v_{ss}/2) \text{tr}[\mathbf{R}]}{1 - \mu \text{tr}[\mathbf{R}]} \\ &= \xi_{\text{ex}}^d + \xi_{\text{ex}}^f + \xi_{\text{ex}}^{\text{df}}, \end{aligned} \quad (56)$$

where

$$\xi_{\text{ex}}^d = \frac{v_{ss} \xi_{\min}}{2}, \quad (57)$$

$$\xi_{\text{ex}}^f = \frac{\mu \xi_{\min} \text{tr}[\mathbf{R}]}{1 - \mu \text{tr}[\mathbf{R}]}, \quad (58)$$

$$\xi_{\text{ex}}^{\text{df}} = \frac{\mu \xi_{\min} v_{ss} \text{tr}[\mathbf{R}]}{2(1 - \mu \text{tr}[\mathbf{R}])}. \quad (59)$$

The misadjustment is defined as the ratio of the excess MSE to  $\xi_{\min}$ . From (56) to (59), we find that the misadjustment expression is

$$\begin{aligned} M &= M^d + M^f + M^{\text{df}} \\ &= M^d + M^f + M^d M^f, \end{aligned} \quad (60)$$

where the different forms of misadjustments are defined in relation to the above excess MSE definitions.

#### 4.4 Discussion

The main results of Section 4 are expressed by (33), (38), (39), (40), (41), (46), (51), (52), (56) and (60).

The first point to note is the fact that, as long as the delay estimation algorithm is convergent in the mean square ( $v_{ss}$  is finite), the conditions for convergence of the LMS adaptive filter (46) and (51) are identical to the usual conditions for a similar adaptive filter operating alone, i.e. the convergence depends solely on the eigenvalues of the input signal autocorrelation matrix. Note also that, because of the adaptive delay element, the weight vector estimate is biased.

Another point to note is that the convergence of the LMS adaptive delay element depends on the input signal power  $\phi_{uu}(0)$  and the minimum MSE  $\xi_{\min}$ .

The expression (40) for the delay estimate variance is complicated by the presence of the adaptive filter-related terms. The delay estimate variance is also encountered in the excess MSE and misadjustment expressions of (56) and (60). Once the delay variance is computed, these two quantities are seen to be function of two terms specific to the adaptive delay element and to the adaptive filter, respectively, and of a cross-product term (note that the delay specific term being function of  $v_{ss}$ , it is indirectly function of the adaptive filter). Note that the expressions for  $\xi_{\text{ex}}^d$  and  $\xi_{\text{ex}}^f$  are identical to those obtained for the respective adaptive algorithms operating alone [1], [4]. The cross-product terms  $\xi_{\text{ex}}^{\text{df}}$  and  $M^{\text{df}}$  are essentially the result of gradient and derivative estimation noise in the two joint 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 expressions  $M^d$  and  $M^f$ , but also includes a term due to the joint estimation noise.

It is important to keep in mind that all the results of Section 4 were derived for an adaptive Type I system in which the delay  $d_n$  applies to every sample of  $\mathbf{u}_n$ . If the delay adaptation is slow enough, the results in the previous sections can be applied to the case where each sample of  $\mathbf{u}_n$  experiences a different delay.

#### 4.5 Simulation Results

The performance of a system identification Type I configuration has been simulated with a spectrally white input signal  $s(n)$ . Table 1 shows the measured excess MSE and misadjustment for different adaptation conditions and a SNR of 10 dB. The theoretical misadjustment, as given in (60), is also given as  $M_{\text{th}}$ . The cross-product term  $M^d M^f$  being a second order component, its effect is therefore small or negligible, as can be seen from the fact that  $\xi_{\text{ex}}^{\text{df}}$  is always approximately equal to the sum of  $\xi_{\text{ex}}^f$  and  $\xi_{\text{ex}}^d$ . Fig. 5 illustrates the type of learning curve obtained after a unit step occurs in the reference delay.

$\mu$	$\alpha$	$\xi_{\text{ex}}^f$	$\xi_{\text{ex}}^d$	$\xi_{\text{ex}}^{\text{df}}$	$M$	$M_{\text{th}}$
0.1	0.5	0.00312	0.00193	0.00563	40.5%	39.4%
0.05	0.5	0.00141	0.00193	0.00308	22.1%	25.4%
0.1	0.1	0.00312	0.00010	0.00313	22.5%	23.3%
0.01	0.5	0.00026	0.00193	0.00195	14.0%	16.0%
0.05	0.25	0.00141	0.00051	0.00163	11.7%	14.2%

Table 1

It is noted in the simulations that if the adaptive filter has a time constant of adaptation on the same order of the adaptive delay time constant, the reference delay tends to be compensated by both adaptive systems, and the delay element converges to a value different than the reference value. This effect is illustrated in Fig. 6 for a delay element with  $\alpha = 0.5$  and for one with  $\alpha = 0.1$ , when a unit delay step is applied in the reference branch.

#### 5. Conclusion

In this paper, we have studied in some details the joint SD and joint LMS algorithms for time delay estimation and adaptive filtering. The convergence to a multitude of stationary points was established and the tracking behaviour of the two algorithms was investigated. The coupling between the two LMS adaptive algorithms was shown to give a misadjustment expression equal to the sum of the individual misadjustments and a cross-product term. These expressions can be used to obtain a theoretical view of the application of such algorithms in specific environments.

#### References

1. H. Messer and Y. Bar-Ness, "Closed-loop least mean square time-delay estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 413-424, April 1987.
2. S. U. H. Qureshi, "Timing recovery for equalized partial-response systems," *IEEE Trans. on Communications*, pp. 1326-1331, December 1976.

3. E. A. Lee and D. G. Messerschmitt, *Digital Communication*, Kluwer Academic Publishers, 1988.
4. S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1986.
5. R. W. Chang, "Joint optimization of automatic equalization and carrier acquisition for digital communication," *Bell System Technical Journal*, pp. 1069-1105, July-August 1970.
6. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, 1985.

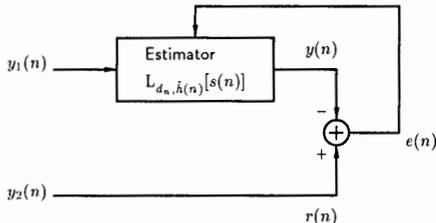


Fig. 1 System identification (cancellation) configuration

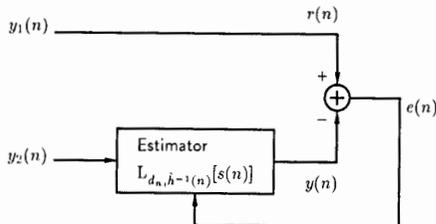


Fig. 2 Inverse filtering (equalization) configuration

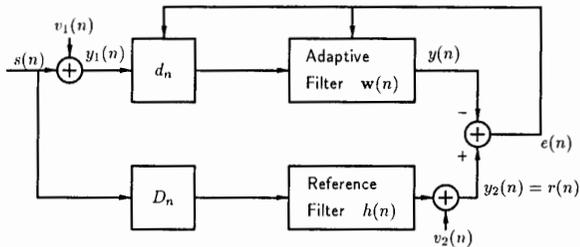


Fig. 3 Type I systems in cancellation configuration

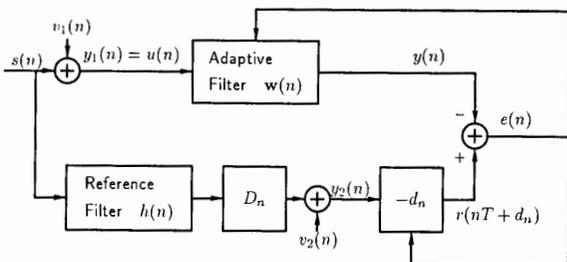


Fig. 4 Type II systems in cancellation configuration

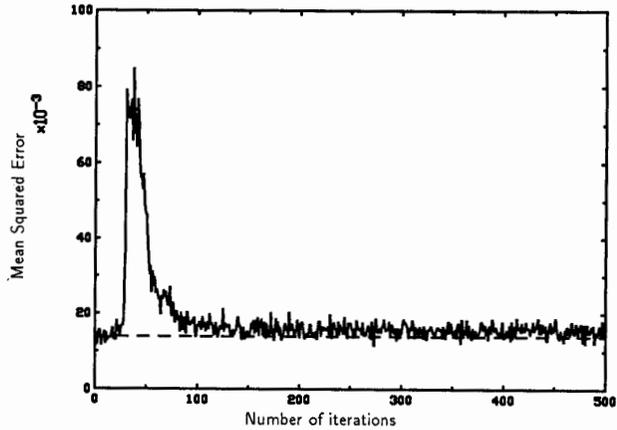


Fig. 5 Learning curve for a unit step reference path delay variation;  $\mu = 0.01$  and  $\alpha = 0.5$ . The dashed line indicates the minimum MSE,  $\xi_{\min} = 0.0139$

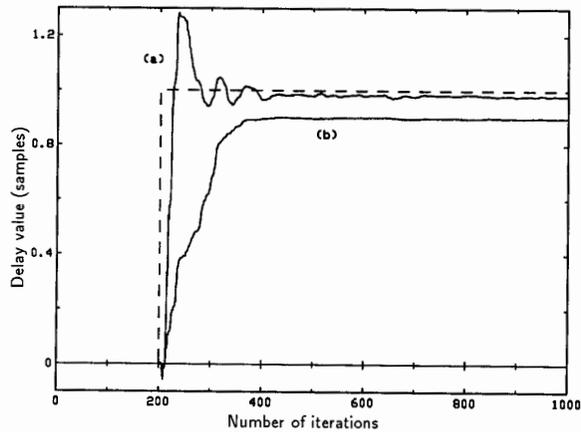


Fig. 6 Delay element response for a unit step reference path delay variation;  $\mu=0.01$ ; (a)  $\alpha = 0.5$ , (b)  $\alpha=0.1$