



HAL
open science

Speeding up VSLMS adaptation algorithms using dynamic adaptation gain: Analysis and Applications

Ioan Doré Landau, Dariusz Bismor, Tudor-Bogdan Airimitoai, Bernard Vau, Gabriel Buche

► **To cite this version:**

Ioan Doré Landau, Dariusz Bismor, Tudor-Bogdan Airimitoai, Bernard Vau, Gabriel Buche. Speeding up VSLMS adaptation algorithms using dynamic adaptation gain: Analysis and Applications. 2024. hal-04548552

HAL Id: hal-04548552

<https://hal.science/hal-04548552>

Preprint submitted on 18 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Speeding up VSLMS adaptation algorithms using dynamic adaptation gain: Analysis and Applications *

Ioan Doré Landau^a, Dariusz Bismor^b, Tudor-Bogdan Airimitoiaie^c, Bernard Vau^d, Gabriel Buche^a

Abstract—The paper explores the use of dynamic adaptation gain/step size (DAG) for improving the adaptation transient performance of variable step-size LMS (VS-LMS) adaptation algorithms. A generic form for the implementation of the DAG within the VS-LMS algorithms is provided. The properties of the VS-LMS algorithms using dynamic adaptation gain are discussed in detail. Stability issues in deterministic environment and convergence properties in stochastic environment are examined. A transient performance analysis is proposed. Criteria for the selection of the coefficients of the DAG filter are provided. The potential of the VS-LMS adaptation algorithms using a DAG is then illustrated by simulation results (adaptive line enhancer, filter identification) and experimental results obtained on a relevant adaptive active noise attenuation system.

I. INTRODUCTION

The modern development of adaptation techniques in automatic control and signal processing started at the end of the fifties and beginning of the sixties (20th century). The paper [30] introduced a gradient-based adaptation algorithm in the discrete-time, later named the least mean squares (LMS). While the choice of the adaptation gain/step size for assuring the stability of the system was an open problem, interesting applications in the field of signal processing have been done. The paper [29] gives an account of the applications of the LMS algorithm up to 1975.

In automatic control, the first attempt to synthesize adaptation algorithms has been probably the paper [28], where a continuous-time formulation of a gradient type algorithm has been proposed. Unfortunately, dealing with feedback control systems to which a non-linear/time-varying loop (the adaptation loop) is added, raised crucial stability issues. The problem of the choice of the adaptation gain (step size) assuring the stability of the full system is fundamental. Therefore, the research paradigm in control was directed toward synthesis of adaptive algorithms guaranteeing the stability of the full system for any (positive) value of the adaptation gain/step size. Discrete time adaptation algorithms assuring global asymptotic stability for any values of the adaptation gain were available since 1971 ([15], [16], [21]). The concepts of "a priori" and "a posteriori" adaptation error emerged as key points for understanding the stability issues in the discrete-time context. Use of this type

of algorithms in signal processing have been reported in [12], [23] among other references. These algorithms, derived from stability considerations, can be interpreted in the scalar case as gradient type algorithms trying to minimize a quadratic criterion in terms of the "a posteriori" prediction error ([13], [14]).

The signal processing community has concentrated its efforts in developing variable step size/adaptation gain algorithms in a scalar context (more exactly using a diagonal matrix adaptation gain) in order to improve the performance of the LMS algorithm. An exhaustive account of the various variable step-size LMS (VS-LMS) algorithms is provided by [6], where a unified presentation is done as well as an extensive comparison of the algorithms by applying them to a number of significant problems. It turns out that the adaptation algorithms developed in automatic control from a stability point of view can also be interpreted as "variable step size" LMS algorithms.

When using adaptive/learning recursive algorithms there is an important problem to be addressed: the compromise between alertness (with respect to environment changes – like plant or disturbance characteristics) and stationary performances when using a constant value for the adaptation gain/step size. Accelerating the adaptation transient without augmenting the value of the adaptation gain/step size is a challenging problem.

Recently, the concept of dynamic adaptation gain (DAG) has been introduced in [17], [19], [22] as a way to accelerate significantly the adaptation transients without modifying the steady state (asymptotic) properties of an algorithm for a given adaptation gain/step size. The correcting term in the adaptation algorithm is first filtered before its use for the estimation of a new parameter value. With an appropriate choice of the parameters of this filter, which should be characterized by a strictly positive real (SPR) transfer function, a significant improvement of the adaptation transient is obtained. The design of this filter is well understood and design tools are available.

The main objective of this paper is to show that the DAG introduced in the context of stability based adaptation algorithms can be successfully applied to VS-LMS adaptation algorithms leading to similar significant acceleration of the adaptation transients. This will be based on a theoretical analysis and will be illustrated by simulations (adaptive line enhancer, filter identification) and real-time experiments on an adaptive feedforward noise attenuator (a silencer).

The paper is organized as follows: Section II will introduce the concept of dynamic adaptation gain in the context of VS-LMS adaptation algorithms. Section III will focus on the analysis and design of the dynamic adaptation gain filter. Section IV will present simulations results for an adaptive line

*This work was not supported by any organization

^aIoan Doré Landau and Gabriel Buche are with the Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France `firstname.lastname@gipsa-lab.grenoble-inp.fr`

^bDariusz Bismor is with the Silesian University of Technology, ul. Akademicka 2A, 44-100 Gliwice, Poland `dariusz.bismor@polsl.pl`

^cTudor-Bogdan Airimitoiaie is with the Univ. Bordeaux, CNRS, Bordeaux INP, IMS, 33405 Talence, France `tudor-bogdan.airimitoiaie@u-bordeaux.fr`

^dBernard Vau is with Exail, 12 avenue des Coquelicots, 94385 Bonneuil sur Marne, France `bernard.vau@exail.com`

enhancer and for filter identification. Section V will present the experimental results obtained with three algorithms on an adaptive feedforward noise attenuator.

II. INTRODUCING THE DAG-VS-LMS ALGORITHMS

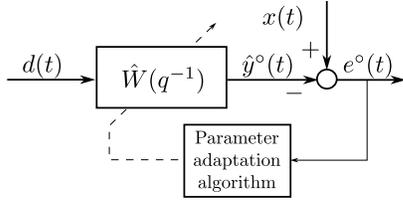


Fig. 1. Least mean squares (LMS) adaptive filtering problem.

Variable step size least mean squares (VS-LMS) algorithms are very popular in the field of signal processing and the field of active vibration and noise control. There is a strong similarity with some of the algorithms used in the adaptive control and recursive system identification. The VS-LMS algorithms (which are improvements of the original LMS algorithm) will be briefly reviewed in order to add the dynamic adaptation gain/step size introduced in [17]. The aim of the LMS parameter adaptation/learning algorithm (PALA) is to drive the parameters of an adjustable model in order to minimize a quadratic criterion in terms of the prediction error (difference between real data and the output of the model used for prediction).

The basic block diagram illustrating the LMS algorithm's operation is shown in Fig. 1. The adaptive filter $\hat{W}(q^{-1})$ is fed with the input sequence¹ $d(t)$. The output of the filter, $\hat{y}^o(t)$, is compared with the desired signal, $x(t)$, to compute the error signal $e^o(t)$. The LMS algorithm adjusts the weights of the $\hat{W}(q^{-1})$ filter to minimize the error.

Consider that the desired signal can be described by:

$$x(t) = \mathbf{w}^T \mathbf{r}(t), \quad (1)$$

where the *parameter vector* and the *measurement vector* are denoted by

$$\mathbf{w}^T = [w_0, w_1, \dots, w_{n_W}], \quad \text{and} \quad (2)$$

$$\mathbf{r}^T(t) = [d(t), d(t-1), \dots, d(t-n_W)], \quad (3)$$

respectively. The adjustable prediction model of the adaptive filter will be described by:

$$\hat{y}^o(t) = \hat{\mathbf{w}}^T(t-1) \mathbf{r}(t), \quad (4)$$

where $\hat{y}^o(t)$ is termed the *a priori* predicted output depending upon the values of the estimated parameter vector \mathbf{w} at instant $t-1$:

$$\hat{\mathbf{w}}^T(t-1) = [\hat{w}_0(t-1), \hat{w}_1(t-1), \dots, \hat{w}_{n_W}(t-1)]. \quad (5)$$

It is very useful to consider also the *a posteriori* predicted output computed on the basis of the new estimated parameter

¹Variables in discrete-time are denoted as $s(t)$, where t gives the integer number of sampling periods and it is related to the continuous-time by $\tau = t \cdot T_s$, where T_s is the sampling period.

vector at t , $\hat{\mathbf{w}}(t)$, which will be available somewhere between t and $t+1$. The *a posteriori* predicted output will be given by:

$$\hat{y}(t) = \hat{\mathbf{w}}^T(t) \mathbf{r}(t) \quad (6)$$

One defines an *a priori* prediction error as:

$$e^o(t) = x(t) - \hat{y}^o(t) = [\mathbf{w} - \hat{\mathbf{w}}(t-1)]^T \mathbf{r}(t) \quad (7)$$

and an *a posteriori* prediction error as:

$$e(t) = x(t) - \hat{y}(t) = [\mathbf{w} - \hat{\mathbf{w}}(t)]^T \mathbf{r}(t) \quad (8)$$

The VS-LMS algorithms update the filter taps according to the formula (see [6]):

$$\hat{\mathbf{w}}(t) = \hat{\mathbf{w}}(t-1) + \mu(t) \mathbf{r}(t) e^o(t), \quad (9)$$

where $\mu(t)$ is the variable step-size parameter. In the standard form of the LMS algorithm, the step-size has a constant value $\mu(t) = \mu$. Large values of μ allow for fast adaptation, but also give large excess mean square error (EMSE, see [6]). Too large step-sizes may lead to the loss of stability (see [13] for a discussion). On the other hand, too small step-sizes give slow convergence, which in many practical applications is not desirable. A variable step-size $\mu(t)$ can provide a compromise. The first VS-LMS algorithm was the Normalized LMS (NLMS) algorithm proposed in 1967 independently by [3], [25], which uses the following equation for the step-size:

$$\mu(t) = \frac{\mu}{\delta + \mathbf{r}(t)^T \mathbf{r}(t)}, \quad (10)$$

where δ is a very small scalar value used in order to avoid division by zero (typically of the order of 10^{-16}).²

In adaptive control, where the stability of the full adaptive control system is considered as a fundamental issue, VS-LMS algorithms have been developed from a stability point of view. However, when scalar type adaptation gain is used (i.e. diagonal matrix adaptation gain), these algorithms can be interpreted as gradient type algorithms for the minimization of a quadratic error in terms of the *a posteriori* prediction error defined in (8) - see [13] for details. See also [22] for a derivation of this type of algorithm and its application to active noise control. This algorithm will be termed PLMS³ (to distinguish it with respect to the LMS and the other VS-LMS that use the *a priori* prediction error). The algorithm has the form:

$$\hat{\mathbf{w}}(t) = \hat{\mathbf{w}}(t-1) + \mu \mathbf{r}(t) e(t) \quad (11)$$

$$= \hat{\mathbf{w}}(t-1) + \mu(t) \mathbf{r}(t) e^o(t), \quad (12)$$

where:

$$\mu(t) = \frac{\mu}{1 + \mu \mathbf{r}(t)^T \mathbf{r}(t)}, \quad (13)$$

$e(t)$ can be computed from $e^o(t)$. One has using Eq. (11):

$$e(t) = [\mathbf{w} - \hat{\mathbf{w}}(t)]^T \mathbf{r}(t) = e^o(t) - \mu \mathbf{r}(t)^T \mathbf{r}(t) e(t) \quad (14)$$

²The case of NLMS using larger values of δ (for example $\delta = 1$) is discussed in [9].

³In [12] a close algorithm is termed HARF.

yielding:

$$e(t) = \frac{e^\circ(t)}{1 + \mu \mathbf{r}(t)^T \mathbf{r}(t)} \quad (15)$$

When using the *dynamic adaptation gain/step size* (DAG), Eq. (9) of the VS-LMS algorithms will take the form:

$$\hat{\mathbf{w}}(t) = \hat{\mathbf{w}}(t-1) + \frac{C(q^{-1})}{D'(q^{-1})} [\mu(t) \mathbf{r}(t) e^\circ(t)] \quad (16)$$

where⁴ $\frac{C(q^{-1})}{D'(q^{-1})}$ is termed the “dynamic adaptation gain/step size” (DAG) and has the form:

$$H_{DAG}(q^{-1}) = \frac{C(q^{-1})}{D'(q^{-1})} = \frac{1 + c_1 q^{-1} + \dots + c_{n_C} q^{-n_C}}{1 - d'_1 q^{-1} - \dots - d'_{n_{D'}} q^{-n_{D'}}} \quad (17)$$

The effective implementation of the algorithm given in Eq. (16) leads to:

$$\begin{aligned} \hat{\mathbf{w}}(t) = & d_1 \hat{\mathbf{w}}(n-1) + d_2 \hat{\mathbf{w}}(n-2) + \dots + d_{n_D} \hat{\mathbf{w}}(n-n_D) \\ & + \mu(t) \mathbf{r}(t) e^\circ(t) + c_1 \mu(n-1) \mathbf{r}(n-1) e^\circ(n-1) + \\ & + \dots + c_{n_C} \mu(n-n_C) \mathbf{r}(n-n_C) e^\circ(n-n_C) \end{aligned} \quad (18)$$

where ($n_D = n_{D'} + 1$):

$$d_i = (d'_i - d'_{i-1}) \quad ; i = 1, \dots, n_D; d'_0 = -1, d'_{n_D} = 0 \quad (19)$$

To implement the algorithm, one needs a computable expression for $e^\circ(t)$. This is obtained by computing $\hat{y}^\circ(t)$ in (7) as⁵

$$\hat{y}^\circ(t) = \hat{\mathbf{w}}_0^T(t-1) \mathbf{r}(t), \quad (20)$$

where

$$\begin{aligned} \hat{\mathbf{w}}_0^T(t-1) = & d_1 \hat{\mathbf{w}}(t-1) + d_2 \hat{\mathbf{w}}(t-2) + \dots + d_{n_D} \hat{\mathbf{w}}(t-n_D) \\ & + c_1 \mu(t-1) \mathbf{r}(t-1) e^\circ(t-1) + \\ & + \dots + c_{n_C} \mu(t-n_C) \mathbf{r}(t-n_C) e^\circ(t-n_C). \end{aligned} \quad (21)$$

Relations with other algorithms

Many algorithms have been proposed for accelerating the speed of convergence of the adaptation algorithms derived using the “gradient rule”. The algorithm of (16) is termed ARIMA (Autoregressive with Integrator Moving Average). As discussed in [19, Section 8], a number of well known algorithms are particular cases of the ARIMA algorithm. The various algorithms described in the literature are of MAI (Moving Average with Integrator) form or ARI (Autoregressive with Integrator) form. The MAI form includes “Integral+Proportional” algorithm [1], [13] ($c_1 \neq 0, c_i = 0, \forall i > 1, d'_i = 0, \forall i > 0$), “Averaged gradient” ($c_i, i = 1, 2, \dots, d'_i = 0, \forall i > 0$) [26], [27]. The ARI form includes “Conjugate gradient” and “Nesterov” algorithms [8], [24] ($c_i = 0, i = 1, 2, \dots, d'_1 \neq 0, d'_i = 0, i > 1$) as well as the “Momentum back propagation” algorithm [11] which corresponds to the conjugate gradient plus a normalization of μ by $(1 - d'_1)$ ⁶. A

⁴The complex variable z^{-1} will be used for characterizing the system’s behaviour in the frequency domain and the delay operator q^{-1} will be used for describing the system’s behavior in the time domain.

⁵For low orders n_C and $n_{D'}$, $\hat{\mathbf{w}}_0(t-1)$ can efficiently be approximated by $\hat{\mathbf{w}}(t-1)$.

⁶There are very few indications how to choose the various coefficients in the above mentioned algorithms.

particular form of the ARIMA algorithms termed “ARIMA2” ($c_1, c_2 \neq 0, c_i = 0, \forall i > 2, d'_1 \neq 0, d'_i = 0, \forall i > 1$) will be studied subsequently and evaluated experimentally.⁷

III. ANALYSIS AND DESIGN OF THE DYNAMIC ADAPTATION GAIN/STEP SIZE

A. Performance issues

The dynamic adaptation gain/learning rate will introduce a frequency-dependent phase distortion on the gradient. Assume that the algorithms should operate for all frequencies in the range: 0 to $0.5f_s$ (f_s is the sampling frequency). Assume that the gradient of the criterion to be minimized contains a single frequency. In order to minimize the criterion, the phase distortion introduced by the dynamic adaptation gain/learning rate should be less than 90° at all the frequencies. In other terms, the transfer function $\frac{C(z^{-1})}{D'(z^{-1})}$ should be strictly positive real (SPR). In order that a transfer function be strictly positive real, it must first have its zeros and poles inside the unit circle. One has the following property:

Lemma 3.1: Assume that the polynomials $C(z^{-1})$ and $D'(z^{-1})$ have all their zeros inside the unit circle, then:

$$I = \int_0^\pi \log \left(\left| \frac{C(e^{-i\omega})}{D'(e^{-i\omega})} \right| \right) d\omega = 0. \quad (22)$$

The proof relies on the Cauchy Integral formula (see [19]).

This result allows to conclude that the average gain of a SPR filter over the frequency range 0 to $0.5f_s$ is 0 dB, i.e. on the average this filter will not modify the adaptation gain/step size. It is just a frequency weighting of the adaptation gain/step size. As it can be seen in Fig. 7, Fig. 11 (Section IV) and in Fig. 16 (Section V), these filters are SPR low band pass filters with an average gain of 0 dB over the frequency range 0 to $0.5f_s$. This means that if the frequency content of the gradient is in the low frequency range, the effective adaptation gain/learning rate will be larger than μ , which should have a positive effect upon the adaptation/learning transient. In particular, the DAG which has the largest gain in low frequencies should provide the best performance (This is indeed the case—see Sections IV and V).

Since we need to have a DAG which is SPR, we will provide subsequently the tools for design of a SPR DAG. We will consider the case of the ARIMA2 algorithm introduced in [17]. The DAG in this case will have the form:

$$H_{DAG} = \frac{C(q^{-1})}{D'(q^{-1})} = \frac{1 + c_1 q^{-1} + c_2 q^{-2}}{1 - d'_1 q^{-1}} \quad (23)$$

A criterion for the selection of c_1, c_2 and d'_1 in order that the DAG be SPR [19] is given in Appendix A.

From the conditions of Lemma A.1 (Appendix A), closed contours in the plane $c_2 - c_1$ can be defined for the different values of d'_1 allowing to select c_1 and c_2 for a given value of d'_1 such that the DAG be SPR. Note that a necessary condition for the selection of the parameters c_1, c_2, d'_1 is that both the denominator and the numerator of the H_{DAG} should be asymptotically stable.

⁷The algorithms mentioned above can be viewed as particular cases of the ARIMA2 algorithm.

B. Stability issues for unconstrained values of the adaptation gain $\mu > 0$

If one wants to use VS-LMS algorithms with large values of the adaptation gain μ , the stability analysis of the resulting scheme using a dynamic adaptation gain is an important issue. For the case of the PLMS algorithm, this analysis has been carried in detail in [19], [22]⁸. For the convenience of the reader, we will indicate the main result subsequently. Equation (16) can be expressed also as:

$$(1 - q^{-1})\hat{\mathbf{w}}(t+1) = + \frac{C(q^{-1})}{D'(q^{-1})} [\mu(t)\mathbf{r}(t)e^\circ(t+1)]$$

leading to:

$$\hat{\mathbf{w}}(t+1) = H_{PAA}(q^{-1})[\mu\mathbf{r}(t)e(t+1)], \quad (24)$$

where $e(t+1) = e^\circ(t+1)(1 + \mu\phi^T(t)\phi(t))^{-1}$ and $H_{PAA} = (1 - q^{-1})^{-1}H_{DAG}$ is a MIMO diagonal transfer operator having identical terms. All the diagonal terms are described by:

$$\begin{aligned} H_{PAA}^{ii}(q^{-1}) &= \frac{1 + c_1q^{-1} + \dots + c_{n_c}q^{-n_c}}{(1 - q^{-1})(1 - d'_1q^{-1} - \dots - d'_{n_D}q^{-n_D})} \\ &= \frac{C(q^{-1})}{(1 - q^{-1})D'(q^{-1})} = \frac{C(q^{-1})}{D(q^{-1})}. \end{aligned} \quad (25)$$

The relation between the coefficients of polynomials D and D' is given in (19).

One has the following result:

Theorem 3.2: For the system described by Equations (1) through (8) using the PLMS algorithm of (16) and (17) one has $\lim_{t \rightarrow \infty} e(t+1) = 0$ for any positive adaptation gain μ and any initial conditions $\hat{\mathbf{w}}(0), e(0)$, if $H_{PAA}^{ii}(z^{-1})$ given in (25) is a positive real (PR) transfer function with a pole at $z = 1$.

The proof is given in [19].

For the particular case of the ARIMA2 algorithm, the coefficients c_1, c_2 and d'_1 should be chosen such that the DAG is SPR and the H_{PAA}^{ii} is positive real (PR), i.e.

$$H_{PAA}^{ii} = \frac{1 + c_1q^{-1} + c_2q^{-2}}{1 - d_1q^{-1} - d_2q^{-2}} = \frac{1 + c_1q^{-1} + c_2q^{-2}}{(1 - q^{-1})(1 - d'_1q^{-1})} \quad (26)$$

should be PR. A criterion for the selection of the coefficients c_1, c_2 and d'_1 can be found in [19]. Closed contours can be defined in the $c_2 - c_1$ plane for given values of d'_1 .

The adaptive/learning system considered in the Theorem 3.2, leads to an equivalent feedback representation where the equivalent feedforward path is a constant positive gain and the equivalent feedback path features the H_{PAA} (see [19]). The feedback path is passive under the condition that H_{PAA}^{ii} is positive real and the feedforward pass is strictly passive allowing to conclude upon the stability of the closed loop. However, in a number of cases (like output error configurations) the equivalent feedforward path may be a transfer operator. In such situations in addition to the PR condition upon the H_{PAA} , there will be an additional SPR condition upon the transfer operator characterizing the equivalent feedforward path.

⁸For $\mu = 1$ and $\delta = 1$, PLMS and NLMS are identical

It is interesting to see intersections of the contours ensuring the SPR of the H_{DAG}^{ii} with the contours assuring that H_{PAA}^{ii} is PR. Such an intersection is shown in Fig.2. From this figure one can conclude that not all the SPR H_{DAG} will lead to a PR H_{PAA} . In such cases, the performance is improved for low adaptation gains, but asymptotic stability cannot be guaranteed for high adaptation gain values. Fig. 2 shows also that there is a region where despite that H_{PAA} is PR, H_{DAG} is not SPR. For such configurations, large adaptation gains can be used but the adaptation transient is slower than for the basic gradient algorithm.

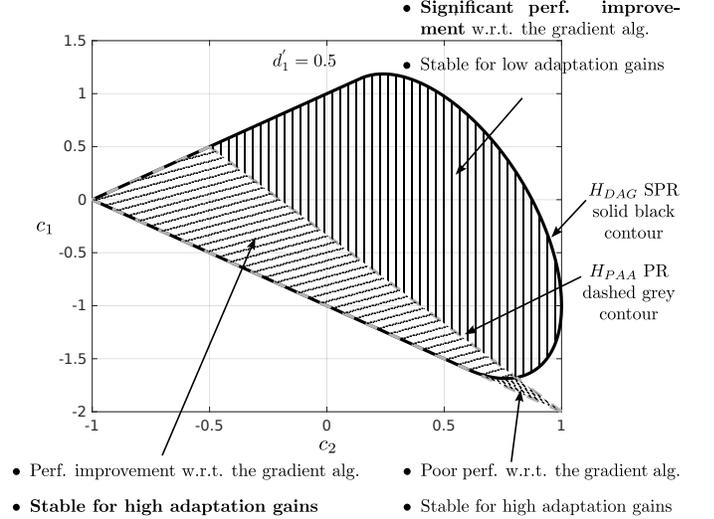


Fig. 2. Intersection in the plane $c_1 - c_2$ of the contour $H_{PAA} = PR$ with the contour $H_{DAG} = SPR$ for $d'_1 = 0.5$.

C. Stability issues for low values of the adaptation gain/step size $\mu > 0$

For small values of the adaptation gains/learning rates, the passivity/stability condition can be relaxed using *averaging* [4]. Using the results of [20], under the hypothesis of an input signal spanning all the frequencies up to half of the sampling frequency, *passivity in the average* will be assured if the frequency interval where H_{PAA}^{ii} is not positive real is smaller than the frequency interval where H_{PAA}^{ii} is positive real. In fact, the most important is that the H_{PAA}^{ii} is PR in the frequency region of operation (mainly defined by the spectrum of the input signals to the systems). However, more specific results can be obtained using *averaging* as shown next⁹. Defining the parameter error as:

$$\tilde{\mathbf{w}}(t) = \hat{\mathbf{w}}(t) - \mathbf{w} \quad (27)$$

from Eq. (24) on gets:

$$\tilde{\mathbf{w}}(t+1) = \tilde{\mathbf{w}}(t) + FH_{DAG}(q^{-1})[\mathbf{r}(t)e(t+1)]; \quad F = \mu I \quad (28)$$

Taking into account that :

$$e(t+1) = \mathbf{r}^T[\mathbf{w} - \hat{\mathbf{w}}(t+1)] = -\mathbf{r}^T\tilde{\mathbf{w}}(t+1) \quad (29)$$

⁹The subsequent analysis is valid for most of the VS-LMS algorithms.

Eq. (28) becomes:

$$\tilde{\mathbf{w}}(t+1) = \tilde{\mathbf{w}}(t) - FH_{DAG}(q^{-1})[\mathbf{r}(t)\mathbf{r}^T\tilde{\mathbf{w}}(t+1)] \quad (30)$$

We are now interested on the average of the correcting term $\mathbf{r}(t)\mathbf{r}^T\tilde{\mathbf{w}}(t+1)$. To do averaging, one assumes that the adaptation gain is small enough such that the estimated parameters evolve slowly. This means that:

$$\begin{aligned} \frac{1}{N+1} \sum_{i=0}^N (\mathbf{r}(t-i)\mathbf{r}(t-i)^T\tilde{\mathbf{w}}(t-i+1)) \\ \approx \frac{1}{N+1} \sum_{i=0}^N (\mathbf{r}(t-i)\mathbf{r}(t-i)^T)\tilde{\mathbf{w}}(t) \end{aligned} \quad (31)$$

However, for large enough N :

$$\frac{1}{N+1} \sum_{i=0}^N (\mathbf{r}(t-i)\mathbf{r}(t-i)^T) \approx \mathbf{E}\{\mathbf{r}(t)\mathbf{r}(t)^T\} = \mathbf{E}_r \quad (32)$$

and the evolution of the parameter error on the average will be given by:

$$\tilde{\mathbf{w}}(t+1) = \tilde{\mathbf{w}}(t) - FH_{DAG}(q^{-1})[\mathbf{E}_r\tilde{\mathbf{w}}(t+1)] \quad (33)$$

$$= \tilde{\mathbf{w}}(t) - F\mathbf{E}_rH_{DAG}(q^{-1})[\tilde{\mathbf{w}}(t+1)] \quad (34)$$

leading to the equivalent feedback representation shown in Fig.3. If in addition one makes the assumption that $\mathbf{r}(t)$ is a stationary persistently exciting signal:

$$\sigma_1 I < \frac{1}{N+1} \sum_{i=0}^N (\mathbf{r}(t-i)\mathbf{r}(t-i)^T) < \sigma_2 I; \quad \sigma_1, \sigma_2 > 0 \quad (35)$$

it results that $\mathbf{E}_r > 0$ (equivalent to the condition of persistent excitation, see [13]). As a consequence, the equivalent feedback system will be asymptotically stable (i.e. the parameter error will go to zero as well as the adaptation error) under the sufficient condition that $H_{DAG}(z^{-1})$ is a strictly positive real (SPR) transfer function (which is indeed the case) since the feedforward path is passive (an integrator) [13]. Therefore, the stability conditions are relaxed when working with low values of the adaptation gain/step size $\mu > 0$.

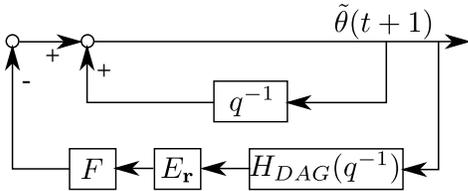


Fig. 3. Equivalent feedback representation of the averaged dynamic VS-LMS algorithm.

D. Adaptation transient analysis

Under the assumptions of persistence of excitation and low adaptation gains (slow adaptation) one can push further the approximation of the equation describing the behavior of the averaged adaptation algorithm via linearization. This allows to understand the effect of the DAG and provides further hints

for its design.

Consider the case of a single parameter to adapt. Linearization corresponds in Fig. 3 to the replacement of the product $F\mathbf{E}_r$ by a constant positive definite matrix, which for $\dim(\tilde{\mathbf{w}}) = 1$ is a positive scalar denoted g .

The linearized approximation of the algorithm will be described by:

$$\tilde{\mathbf{w}}(t+1) = \tilde{\mathbf{w}}(t) - gH_{DAG}(q^{-1})[\tilde{\mathbf{w}}(t+1)] \quad (36)$$

which corresponds to a linear feedback system whose output is $\tilde{\mathbf{w}}(t+1)$ (in Fig. 3 one replaces $F\mathbf{E}_r$ by g). The adaptation transient behavior will be described by the output sensitivity function of this feedback system. For the particular case of an ARIMA2 dynamic adaptation gain, the output sensitivity function will be given by:

$$\begin{aligned} S &= \frac{1 - q^{-1}}{1 + gH_{DAG}(q^{-1})} \\ &= \frac{(1 - q^{-1})(1 - d'_1 q^{-1})}{(1 - q^{-1})(1 - d'_1 q^{-1}) + g(1 + c_1 q^{-1} + c_2 q^{-2})} \end{aligned} \quad (37)$$

We are interested in the response of this transfer function with respect to a step parameter error. Fig.4 shows the step response for various values of the DAG coefficients and two selected values of the adaptation gain g .

As it can be observed for $g = 0.01$ the convergence time

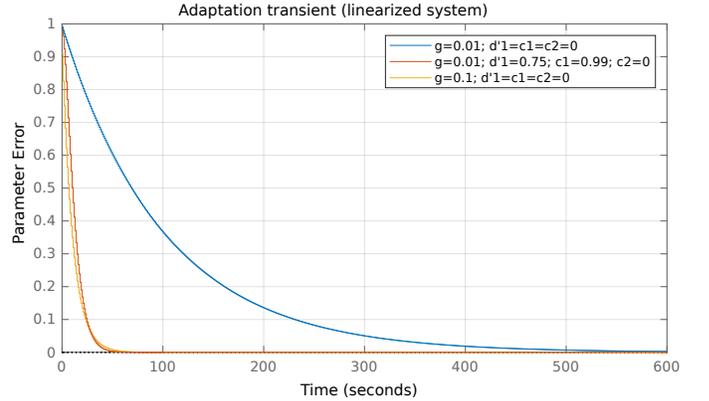


Fig. 4. Adaptation transient on the linearized system.

for the basic algorithm is approximately 600 s. By adding the DAG ($d'_1 = 0.75, c_1 = 0.99; c_2 = 0$) the convergence time is reduced to approximately 70 s. Note that a similar performance can be obtained with the basic algorithm by multiplying the gain g by 10 (i.e. the acceleration obtained with the DAG without augmenting the adaptation gain is equivalent to using an adaptation gain 10 times larger on the basic algorithm).

E. Stochastic environment - convergence analysis

We will consider the I/O model given in (1) where the output is disturbed by a noise $\mathbf{n}(t+1)$:

$$x(t+1) = \mathbf{w}^T \mathbf{r}(t) + \mathbf{n}(t+1) \quad (38)$$

In this equation, $\mathbf{n}(t)$ is a zero mean stationary stochastic disturbance with finite moments. Using Eqs. (13), (15) and (16), the PLMS algorithm can be expressed as:

$$\tilde{\mathbf{w}}(t+1) = \tilde{\mathbf{w}}(t) + \mu H_{DAG}(q^{-1})[\mathbf{r}(t)e(t+1)] \quad (39)$$

For the analysis of this algorithm in a stochastic context, we will use the ODE (ordinary differential equation) approach of Benveniste-Lam [5], [7]. Convergence results will be extracted from an ODE which approximate the algorithm by using an average of the correcting term. We will make the following assumptions: (1) *Stationary processes* $\mathbf{r}(t, \hat{\mathbf{w}})$ and $e(t+1, \hat{\mathbf{w}})$ can be defined for $\hat{\mathbf{w}}(t) \equiv \hat{\mathbf{w}}$, (2) $\hat{\mathbf{w}}(t)$ generated by the algorithm belongs infinitely often to the domain (D_s) for which the stationary processes $\mathbf{r}(t, \hat{\mathbf{w}})$ and $\mathbf{n}(t+1, \hat{\mathbf{w}})$ can be defined. For the adaptation algorithm given in Eq. (39) one has the following result:

Lemma 3.3: Assume that the Benveniste's smoothness, boundness and mixing conditions [5] are satisfied.

1) Assume that:

$$\sigma_2 I \geq \mathbf{E} \{ \mathbf{r}(t, \hat{\mathbf{w}}) \mathbf{r}^T(t, \hat{\mathbf{w}}) \} \geq \sigma_1 I; \quad \sigma_1, \sigma_2 > 0 \quad (40)$$

2) Assume that $\mathbf{r}(n, \hat{\mathbf{w}})$ is such (persistence of excitation):

$$\mathbf{r}(n, \hat{\mathbf{w}})^T (\hat{\mathbf{w}} - \mathbf{w}) = 0 \Rightarrow \hat{\mathbf{w}} = \mathbf{w} \quad (41)$$

3) Assume that either:

- $\mathbf{n}(t+1)$ is a sequence of independent equally distributed normal random variables $(0, \sigma)$ or
-

$$\mathbf{E} \{ \mathbf{r}(t, \hat{\mathbf{w}}), \mathbf{n}(t+1, \hat{\mathbf{w}}) \} = 0 \quad (42)$$

4) Assume that $H_{DAG}(z^{-1})$ is strictly positive real

For a sufficiently small μ one has:

$$\mathbf{P} \{ \| \hat{\mathbf{w}}([T/\mu]) - \mathbf{w} \| \geq \eta + C\zeta(\mu) \} < C'\zeta(\mu) \quad (43)$$

where: $\zeta(\mu)$ is a positive decreasing function, C, C' are positive constants, $\bar{\mathbf{w}}(t)$ is the average of $\mathbf{w}(t)$, $T < \infty$ is such that $\{ \| \bar{\mathbf{w}}(T) - \mathbf{w} \| \} < \eta$ and $[T/\mu]$ is the closest integer to T/μ .

The proof of this lemma is given in Appendix B.

The interpretation of this result is as follows: the parametric error (distance) is bounded in *probability* and this bound depends upon the magnitude of μ ¹⁰.

F. Convergence rate

For estimating the convergence rate, one can consider the ODE equation associated with the algorithm and the Lyapunov function candidate used for stability analysis. The ratio $\frac{|\dot{V}|}{V}$ is an estimation of the convergence rate for large t . From Eqs. (62) and (63) one gets:

$$\frac{|\dot{V}|}{V} = \frac{(\hat{\mathbf{w}} - \mathbf{w})^T [G_{\mathbf{w}} + G_{\mathbf{w}}^T] (\hat{\mathbf{w}} - \mathbf{w})}{(\hat{\mathbf{w}} - \mathbf{w})^T \mu^{-1} (\hat{\mathbf{w}} - \mathbf{w})} \quad (44)$$

where $G_{\mathbf{w}} = \mathbf{E} \{ H_{DAG}(q^{-1}) \mathbf{E}_{\mathbf{r}} \}$. Assuming that the components of the $\mathbf{r}(t)$ are in the low frequency range (this is often the case), then $H_{DAG}(q^{-1})$ can be approximated by its static gain and one has from Eq. (44):

$$\frac{|\dot{V}|}{V} \approx \frac{1 + \sum_{j=1}^{n_C} c_j (\hat{\mathbf{w}} - \mathbf{w})^T [\mathbf{E}_{\mathbf{r}} + \mathbf{E}_{\mathbf{r}}^T] (\hat{\mathbf{w}} - \mathbf{w})}{1 - \sum_{j=1}^{n_{D'}} d'_j (\hat{\mathbf{w}} - \mathbf{w})^T \mu^{-1} (\hat{\mathbf{w}} - \mathbf{w})} \quad (45)$$

¹⁰This is a stronger result than convergence on the average only

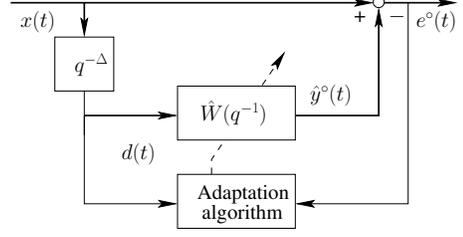


Fig. 5. Block diagram of the Adaptive Line Enhancer.

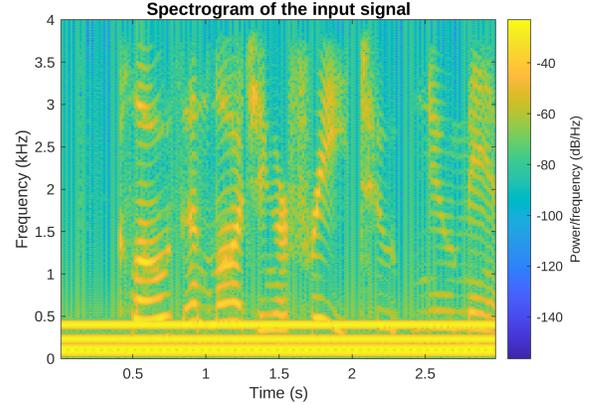


Fig. 6. The spectrogram of the input signal.

where $\mathbf{E}_{\mathbf{r}} = \mathbf{E}[\mathbf{r}(t, \hat{\mathbf{w}}), \mathbf{r}(t, \hat{\mathbf{w}})^T]$. Since $H_{DAG}(q^{-1})$ is strictly positive real, its static gain is positive. Provided that the steady state gain $S SG = \frac{1 + \sum_{j=1}^{n_C} c_j}{1 - \sum_{j=1}^{n_{D'}} d'_j} > 1$ an acceleration with respect to the standard VS-LMS will be obtained and this will be illustrated in the simulation and experimental sections.

IV. SIMULATION RESULTS

A. Adaptive Line Enhancer

The properties of several different Dynamic Adaptation Gains (DAGs) were tested in simulations of the Adaptive Line Enhancer (ALE) – an adaptive filtering system introduced by Widrow, capable of detecting and removing highly correlated signals embedded in wide-band signals [29], [10]. The block diagram of the ALE is presented in Fig. 5. The input signal to the ALE, $x(t)$, was constructed of a speech recording with four contaminating sine signals. The speech recording was one of the Matlab's Audio Toolbox exemplary files entitled *FemaleSpeech-16-8-mono-3secs.wav* (sampling frequency: 8 kHz). A contaminating signal formed by four sines with frequencies of 80, 125, 230 and 400 Hz has been considered. The signal thus constructed is non-stationary, with the spectrogram presented in Fig. 6.

The adaptive filter length and the decorrelation delay (Δ) used in the simulations were both 100. The FIR filter was longer than can be expected, provided that only four sine signals are to be detected. However, it is known that when the sines are close to each other, the required filter length

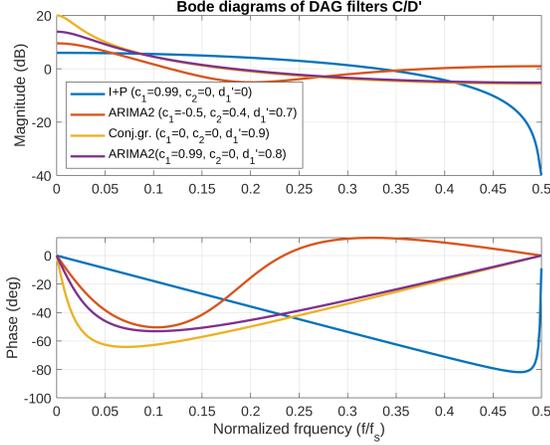


Fig. 7. Bode diagrams of the DAGs used in simulations of the ALE.

to detect and cancel them efficiently is usually higher.¹¹ The adaptive filter was always started with zero initial conditions.

Both the NLMS and PLMS algorithms, incorporating a DAG, were tested and compared in the simulations, while the NLMS and PLMS without the DAG served as a reference for comparison. The step size for the NLMS was experimentally selected as 0.02. The step size for the PLMS ($5 \cdot 10^{-4}$) was adjusted to achieve the same initial convergence speed for both the NLMS and PLMS algorithms without the DAG.

The performance of different DAG settings were evaluated by the means of two quantities. The first was the convergence speed calculated as time (in samples) after which the mean squared error (MSE) achieved the level of -40 dB for the first time. The second was the sum of the MSE during the first 3200 iterations (0.4 s) of the simulation. Please note that excess MSE could not have been used due to the lack of the Wiener filter model for this nonstationary input signal, which implies impossibility to calculate the system noise. Fig. 7 summarizes the frequency characteristics of the various DAG used. All are SPR.

Figure 8 presents the results obtained with the NLMS algorithm for the selected set of coefficients of the $H_{DAG}(q^{-1})$. The figure is zoomed to show the MSE during the initial convergence of the ALE filter. The plots clearly show that all the selected DAGs speed up the initial convergence significantly, compared to the NLMS without the DAG. This is confirmed by numerical values of the selected performance indicators, presented in Table I. In case of the NLMS algorithm, the fastest convergence has been obtained for set no. 5 ($c_1 = -0.5$, $c_2 = 0.4$ and $d'_1 = 0.7$). In the same time, the smallest sum of MSE is obtained for set no. 6 ($c_1 = 0.99$, $c_2 = 0$, $d'_1 = 0.8$).

Figure 9 presents similar results obtained for the PLMS algorithm, and the corresponding values of the performance indicators are presented in Table II. In the case of this algorithm, set no. 6 provides both the shortest convergence time and the smallest sum of MSE.

¹¹This effect was confirmed in simulations with shorter filters, which were not capable of cancelling all the four sines

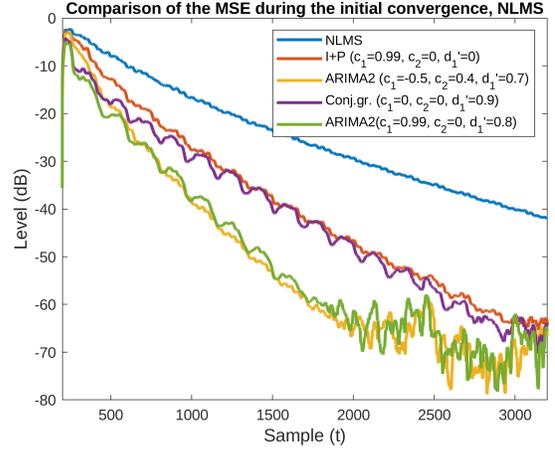


Fig. 8. The spectrogram of the input signal.

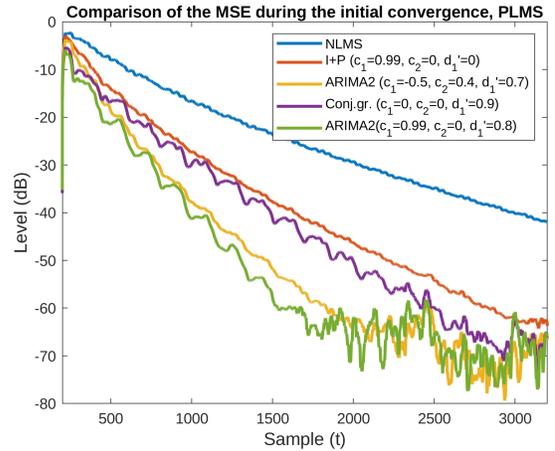


Fig. 9. The spectrogram of the input signal.

TABLE I
RESULTS OF THE ALE EXPERIMENTS AVERAGED OVER 250 RUNS, ALE LENGTH: 100

No	Algorithm / DAG settings	Conv.time (samples)	Sum of MSE
1	NLMS	2791	145.2
3	$c_1 = 0.99, c_2 = 0, d'_1 = 0$	1426	74.4
4	$c_1 = 0, c_2 = 0, d'_1 = 0.9$	1296	33.7
5	$c_1 = -0.5, c_2 = 0.4, d'_1 = 0.7$	843	49.8
6	$c_1 = 0.99, c_2 = 0, d'_1 = 0.8$	916	24.9

TABLE II
RESULTS OF THE ALE EXPERIMENTS AVERAGED OVER 250 RUNS, PLMS ALGORITHMS.

No	Algorithm / DAG settings	Conv.time (samples)	Sum of MSE
1	NLMS	2791	145.2
2	PLMS	2796	132.0
3	$c_1 = 0.99, c_2 = 0, d'_1 = 0$	1429	62.1
4	$c_1 = 0, c_2 = 0, d'_1 = 0.9$	1286	31.5
5	$c_1 = -0.5, c_2 = 0.4, d'_1 = 0.7$	914	41.5
6	$c_1 = 0.99, c_2 = 0, d'_1 = 0.8$	770	20.5

B. Filter identification

Assume that the signal $x(t)$ in Fig. 1 is the output of an IIR filter whose input is a PRBS (pseudo random binary sequence) with $N = 2^8 - 1 = 255$ samples. The IIR filter is characterized by the transfer operator (unknown).

$$S_{ys} = \frac{q^{-2} + 0.5q^{-3}}{1 - 1.5q^{-1} + 0.7q^{-2}} \quad (46)$$

The objective was to estimate the parameters of this IIR filter as well as the parameters of an FIR filter able to approximate this IIR filter. The PLMS algorithm has been used to illustrate the properties of the various DAGs.

1) *IIR identification*: For the adaptation gain/learning rate $\mu = 0.02$, the performance of the adaptation algorithms was evaluated with respect to the choice of the coefficients c_1, c_2, d'_1 . To assess the performance, the following indicators were used: (i) the sum of the squared *a posteriori* prediction errors: $J_\epsilon(t) = \sum_{t=0}^N e^2(t+1)$, (ii) the square of the parametric distance: $D^2(t) = \{[\mathbf{w} - \hat{\mathbf{w}}(t)]^T [\mathbf{w} - \hat{\mathbf{w}}(t)]\}$, and (iii) the sum of the squared parametric distance: $J_D(t) = \sum_{t=0}^N D^2(t)$. Table III summarizes the performance of the 2nd order

TABLE III
PERFORMANCE OF 2ND ORDER ARIMA ALGORITHMS.

Algorithm	c_1	c_2	d'_1	$J_D(N)$	$J_\epsilon(N)$	$J_D(N)$	$J_\epsilon(N)$
				IIR	IIR	FIR	FIR
Gradient	0	0	0	91.23	81.44	589.1	396.7
Conj.Gr/Nest..	0	0	0.5	66.95	60.27	350.3	249.6
I+P+D	0	0.99	0	68.86	74.15	351.7	255
I+P	0.9	0	0	69.71	61.66	356.8	247.6
ARIMA 2	0.65	0	0.3	64.45	59.93	340.9	222

0 dB axis). One can notice that the phase lag is less than 90 degrees at all the frequencies. It was verified that the average gain over the all frequency range was 0 dB. This means that the improvement in performance is related to the frequency distribution of the adaptation gain/learning rate. Now examining the magnitude, one observes that both are low pass filters amplifying low frequencies. This means that if the frequency content of the gradient is in the low frequency range, the effective adaptation gain/learning rate will be larger than the gradient adaptation gain (0 dB), which should have a positive effect upon the adaptation/learning transient. In particular, the DAG which has a larger gain in low frequencies (ARIMA2) should provide better performance than the (I+P) DAG (which was indeed the case). This observation is also coherent with the estimated asymptotic convergence rate.

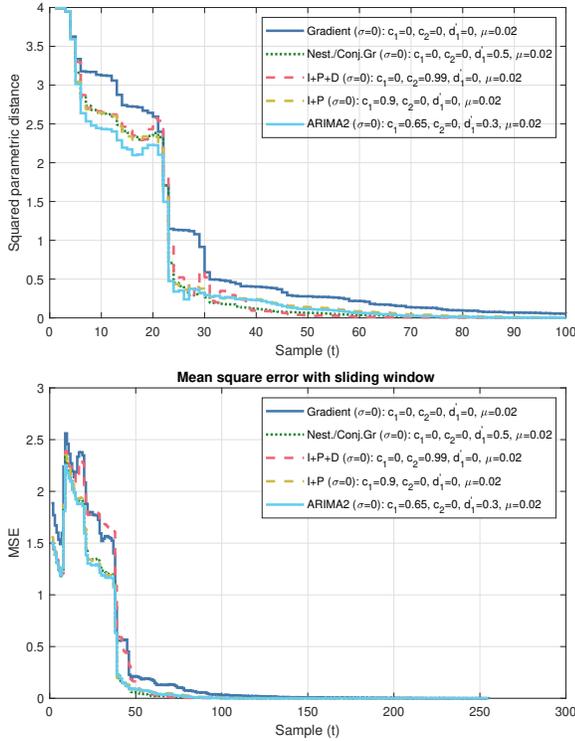


Fig. 10. Evolution of the squared parametric distance $D^2(t)$ (top) and of the MSE (mean square error)(bottom)(IIR identification).

ARIMA algorithm (ARIMA 2) and of the various particular cases. All the DAGs in Table III are SPR. Fig. 10 shows the evolution of the squared parametric distance (top) and of the MSE (mean square error)(bottom). Clearly the use of a DAG provided a significant performance improvement with respect to the basic PLMS algorithm. The ARIMA 2 DAG gave the best results.

Fig. 11 gives the Bode diagram for the ARIMA 2 and I+P algorithms (the gradient algorithm corresponds to the

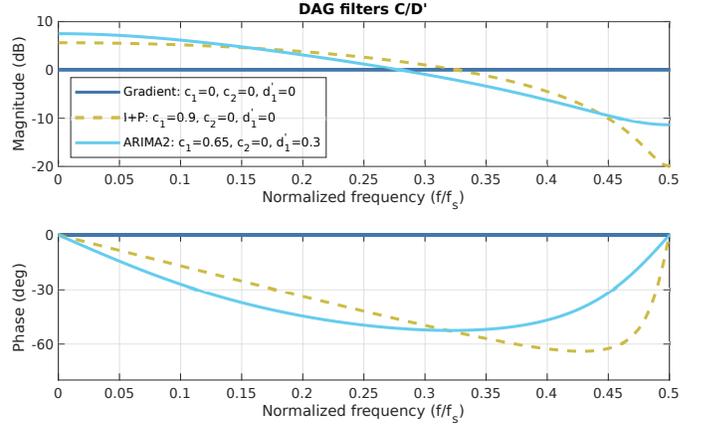


Fig. 11. Bode diagram of the dynamic adaptation gain/learning rate H_{DAG} for ARIMA2, I+P and Gradient algorithms (from Table IV).

2) *FIR identification*: A FIR filter with 30 parameters allows to well approximate the IIR filter given in Eq. (46). To evaluate the adaptation transient for the estimated parameters an estimation of "good" parameters was done first. Figure 12 gives the time evolution of the squared parametric distance (top) and of the MSE (bottom). Performance of the various DAGs are summarized in Table III (last two columns). The conclusions drawn for the IIR filter identification hold also for FIR filter identification.

3) *Stochastic case*: To the same simulation example, a white noise was added on the output (signal/noise ratio (standard deviation): 33 dB). The adaptation gain/step size $\mu = 0.01$ was used. The input in this case was a PRBS with $N = 2^{11} - 1 = 20427$ samples. Figure 13 shows a zoom of the

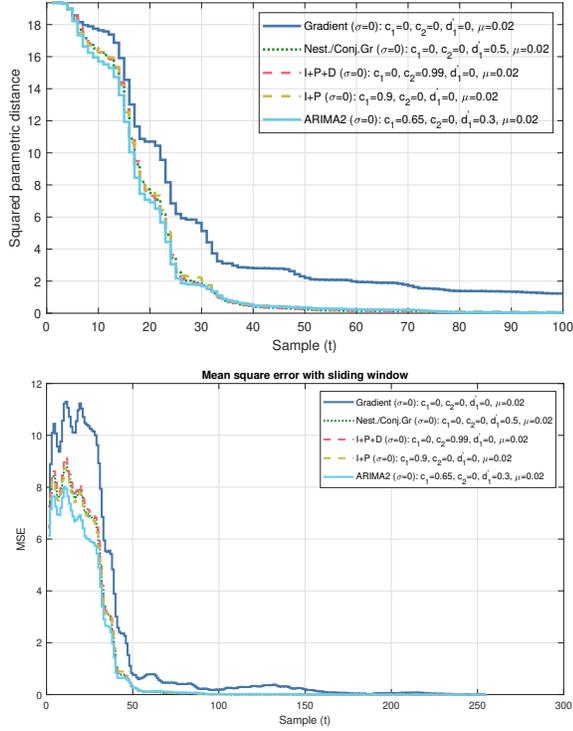


Fig. 12. Evolution of the squared parametric distance $D^2(t)$ (top) and of the MSE (bottom) for the identification of a FIR with 30 parameters.

evolution of the squared parametric distance (average over 100 noise realizations) for IIR filter identification. One gets almost asymptotic unbiased parameters estimates (initial value of the squared parametric distance is 4) and the improvement of the transient performances with respect to the basic algorithm is obvious¹².

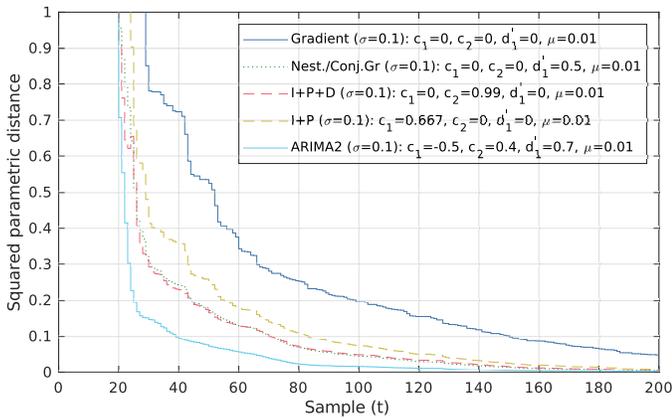


Fig. 13. Evolution of the squared parametric distance in the presence of noise (zoom).

V. EXPERIMENTAL RESULTS

The objective of this section is to show that the dynamic adaptation gain can be implemented on various VS-LMS

¹²The transient performance can be related to the asymptotic convergence rate given in Section III-F.

algorithms and this will lead to a significant acceleration of the adaptation transient. Specifically, for this paper the LMS, NLMS and PLMS algorithms were implemented and tested experimentally on an active noise control test-bench (adaptive feedforward noise attenuation). Figure 14 shows the view of the test-bench used for experiments. Detailed description can be found in [22]. The speakers and the microphones were connected to a target computer with Simulink Real-Time[®]. A second computer is used for development and operation with Matlab/Simulink. The sampling frequency was $f_s = 2500$ Hz.

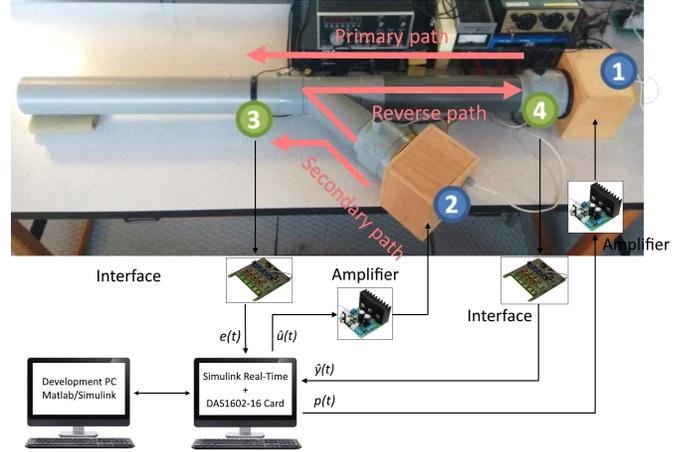


Fig. 14. Duct active noise control test-bench.

The various paths are described by models of the form: $X(q^{-1}) = q^{-d_x} \frac{B_X(q^{-1})}{A_X(q^{-1})} = q^{-d_x} \frac{b_1^X q^{-1} + \dots + b_n^X q^{-n}}{1 + a_1^X q^{-1} + \dots + a_n^X q^{-n}}$, with $B_X = q^{-1} B_X^*$ for any $X \in \{G, M, D\}$. $\hat{G} = q^{-d_G} \frac{\hat{B}_G}{\hat{A}_G}$, $\hat{M} = q^{-d_M} \frac{\hat{B}_M}{\hat{A}_M}$, and $\hat{D} = q^{-d_D} \frac{\hat{B}_D}{\hat{A}_D}$ denote the identified (estimated) models of the secondary (G), reverse (M), and primary (D) paths. The system's order is defined as (the indexes G , M , and D have been omitted): $n = \max(n_A, n_B + d)$.

The models of the above paths are characterized by the presence of many pairs of very low damped poles and zeros. These models were identified experimentally. The orders of the various identified models were: $n_G = 33$, $n_M = 27$ and $n_D = 27$.

The objective is to attenuate an incoming unknown broadband noise disturbance eventually mixed with several tonal disturbances. The corresponding block diagram for the adaptive feedforward noise compensation using FIR Youla-Kucera (FIR-YK) parametrization of the feedforward compensator (introduced in [18] for active vibration control and in [2] for active noise control) is shown in Figure 15.

The adjustable filter \hat{Q} has the FIR structure:

$$\hat{Q}(q^{-1}) = \hat{q}_0 + \hat{q}_1 q^{-1} + \dots + \hat{q}_{n_Q} q^{-n_Q} \quad (47)$$

and the parameters \hat{q}_i will be adapted in order to minimize the residual noise.

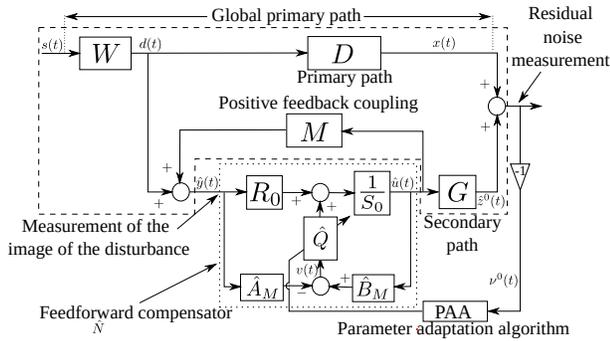


Fig. 15. Feedforward ANC with FIR-YK adaptive feedforward compensator.

The algorithm which was used (introduced in [18]) can be summarized as follows. One defines:

$$\mathbf{w}^T = [q_0, q_1, q_2, \dots, q_{n_Q}] \quad (48)$$

$$\hat{\mathbf{w}}^T = [\hat{q}_0, \hat{q}_1, \hat{q}_2, \dots, \hat{q}_{n_Q}] \quad (49)$$

$$\mathbf{v}^T(t) = [v(t+1), v(t), \dots, v(t-n_Q+1)] \quad (50)$$

where:

$$v(t+1) = B_M \hat{u}(t+1) - A_M \hat{y}(t+1) = B_M^* \hat{u}(t) - A_M \hat{y}(t+1) \quad (51)$$

One defines also the regressor vector (a filtered observation vector) as:

$$\mathbf{r}(t) = L(q^{-1})\mathbf{v}(t) = [v_f(t+1), v_f(t), \dots, v_f(t-n_Q+1)] \quad (52)$$

where

$$v_f(t+1) = L(q^{-1})v(t+1) \quad (53)$$

Using $R_0 = 0$ and $S_0 = 1$, the poles of the internal positive closed loop are defined by A_M and they will remain unchanged. The filter used in (53) becomes $L = \hat{G}$ and the associated linear transfer operator appearing in the feedforward path of the equivalent feedback system is:

$$H(q^{-1}) = \frac{G(q^{-1})}{\hat{G}(q^{-1})} \quad (54)$$

(the algorithm uses an approximate gradient). The transfer function associated to $H(q^{-1})$ should be SPR in order to assure asymptotic stability in the case of perfect matching [18]. This is a very mild condition as long as a good experimental identification of the models is done.

The VS-LMS algorithms which were used are of the form given in (9), where $e^o(t)$ is the measured residual noise with minus sign ($\nu^o(t)$ in Fig. 15), $\hat{\mathbf{w}}$ is given by Eq. (49) and \mathbf{r} is given in Eq. (52)¹³. The adjustable filter \hat{Q} had 60 parameters. The values of the adaptation gain μ for the three algorithms were tuned such that in the absence of the DAG, performance was close for the three algorithms. A specificity of this application is also the low value of the average of $\mathbf{r}(t)^T \mathbf{r}(t)$ ($\ll 1$). This means that the LMS and PLMS show a very close behavior for a given μ and that the adaptation gain μ for the NLMS should be much lower in order to get similar

¹³In signal processing literature when using a filtering of the observation vector, the algorithms are termed FU-VSLMS.

performances. The choices of μ were 0.2 for LMS, 0.0003 for NLMS and 0.22 for the PLMS.

TABLE IV
PARAMETERS OF ARIMA2 DYNAMIC ADAPTATION GAIN.

Algorithm	H_{PAA-PR}	$H_{DAG-SPR}$	c_1	c_2	d'_1
Integral (gradient)	Y	Y	0	0	0
Conj.Gr/Nest..	N	Y	0	0	0.9
I+P+D	N	Y	1.4	0.5	0
I+P	Y	Y	0.99	0	0
ARIMA 2	N	Y	0.99	0	0.9

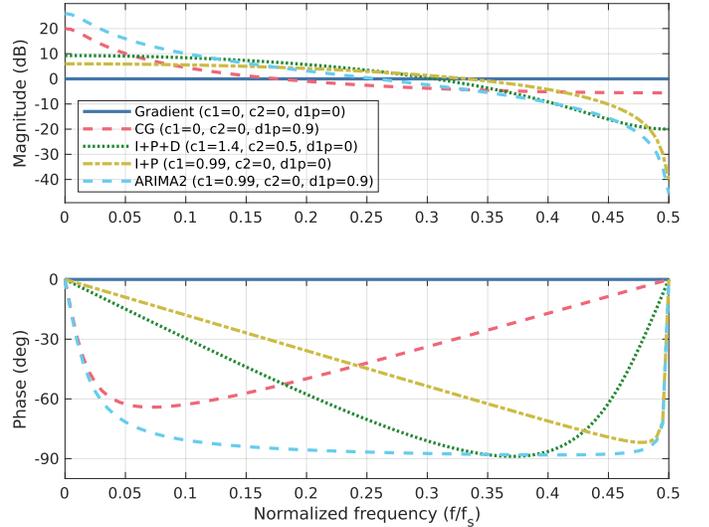


Fig. 16. Frequency characteristics of DAGs used in the experiments (see also Table IV).

A signal consisting of a broad-band disturbance 70–170 Hz plus two tonal disturbances located at 100 Hz and 140 Hz was used as an unknown disturbance acting on the system. The steady state and transient attenuation¹⁴ were evaluated for the various values of the parameters c_1 , c_2 and d'_1 given in Table IV. The frequency characteristics of the various DAG given in Table IV are shown in Fig. 16. One observes that all are SPR (Phase lag between 0 and -90 degrees with a 0 dB average gain). The system was operated in open-loop during the first 15 s. Figure 17 shows the time response of the system as well as the evolution of the global attenuation when using the standard NLMS algorithm (top) and when the ARIMA 2 DAG is incorporated (bottom) with $c_1 = 0.99$, $c_2 = 0$, $d'_1 = 0.9$ (last row of Table IV). One observes a significant acceleration of the adaptation transient. Figure 18 shows for the same experiments a comparison of the PSD of the residual noise in the absence of the control and the PSD of the residual noise at 300 s under the control effect. One can observe that the ARIMA 2 algorithm has improved the attenuation in the low frequency range with respect to the standard NLMS algorithm which is coherent

¹⁴The attenuation is defined as the ratio between the variance of the residual noise in the absence of the control and the variance of the residual noise in the presence of the adaptive feedforward compensation. The variance is evaluated over an horizon of 3 seconds.

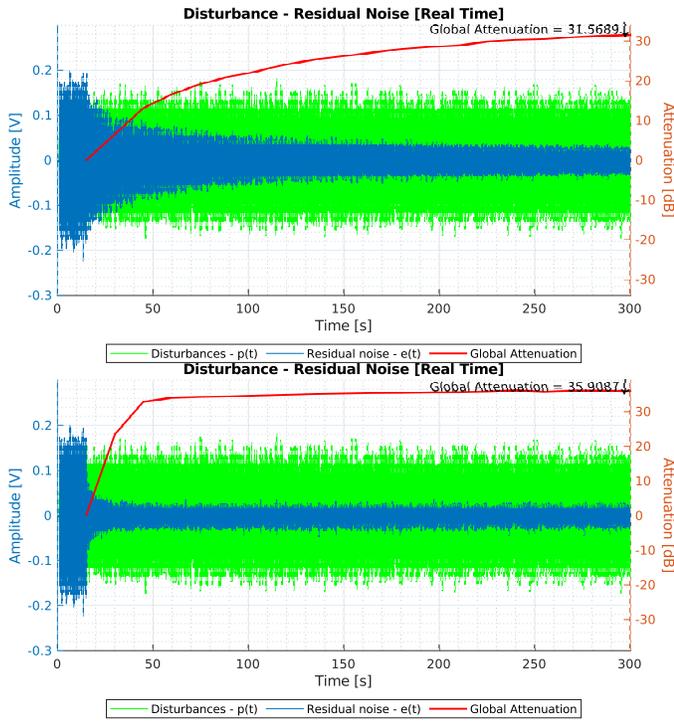


Fig. 17. Time evolution of the residual noise and of the global attenuation using the standard NLMS adaptation algorithm (top) and using the NLMS + ARIMA2 algorithm (bottom), for $\mu = 0.0003$.

with the frequency characteristic of the ARIMA 2 filter shown in Fig. 16.

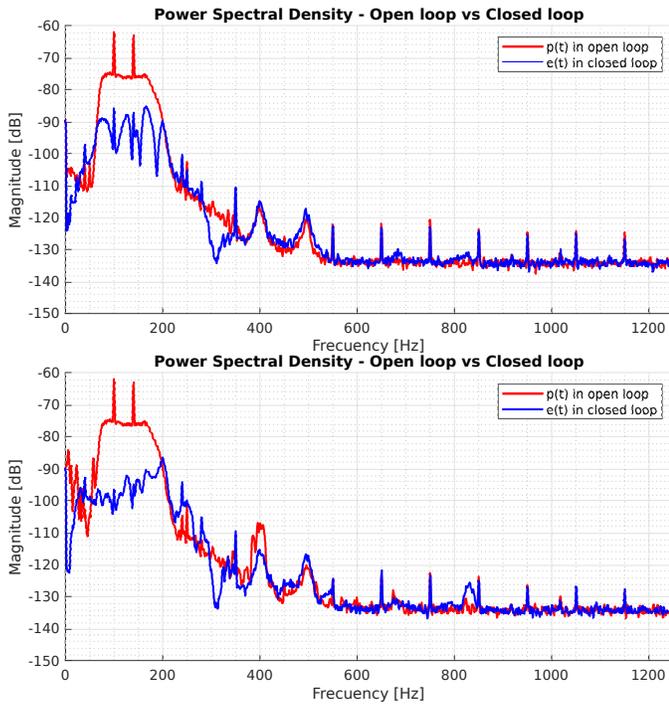


Fig. 18. Comparison of the residual noise PSD in open loop and in closed loop ($t = 300$ s) using the standard NLMS algorithm (top) and using the NLMS+ARIMA 2 algorithm (bottom). Disturbance: 70 - 170 Hz + 2 tonal disturbances (100 Hz, 140 Hz).

The acceleration obtained is equivalent to that obtained with

an adaptation gain 25 times higher is used in the standard NLMS algorithms. This is illustrated in Figure 19. Similar results were obtained for the LMS and the PLMS algorithms.

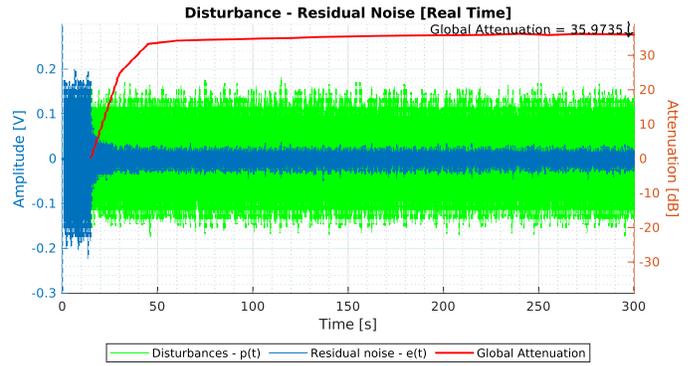


Fig. 19. Time evolution of the residual noise and global attenuation using the standard NLMS adaptation algorithm with $\mu = 0.0075$.

Figures 20, 21, and 22 show the time evolution of the attenuation for the LMS, NLMS and PLMS algorithms when using the various DAG given in Table IV. As one can observe, the effect of the DAG is similar for the three algorithms. The DAG which has the highest steady state gain (ARIMA 2) gives the best result and the performance of the various DAGs are related to the value of their steady state gain (as predicted by Eq. (45)).

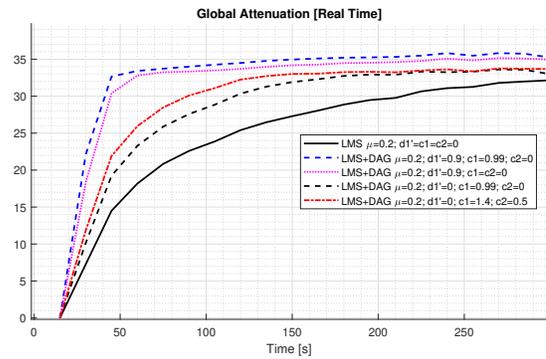


Fig. 20. Time evolution of the global attenuation for the LMS algorithm with various DAGs, using $\mu = 0.2$.

Similar results were obtained for broad – band disturbances with a wider spectrum.

VI. CONCLUSION

The paper emphasizes the potential of the dynamic adaptation gain/step size (DAG) for improving the adaptation transients of VS-LMS adaptation algorithms. The main point is that the DAG should be characterized by a *strictly positive real* (SPR) transfer function for a correct operation over the full frequency range from 0 to 0.5 of the sampling frequency. Simulations results on relevant adaptive filtering problems and experimental results on an adaptive active noise control system have illustrated the feasibility and the performance improvement achieved when a DAG is incorporated in VS-LMS adaptation algorithms.

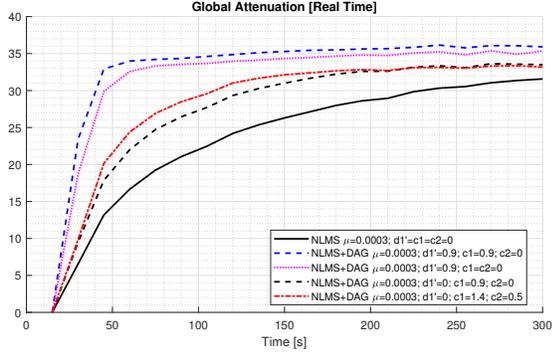


Fig. 21. Time evolution of the global attenuation for the NLMS algorithm with various DAGs, using $\mu = 0.0003$.

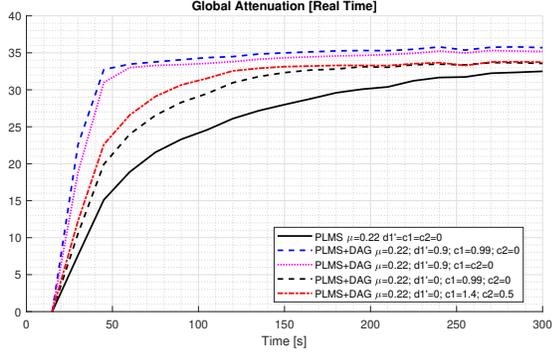


Fig. 22. Time evolution of the global attenuation for the PLMS algorithm with various DAGs, using $\mu = 0.22$.

APPENDIX A

CRITERION FOR THE SELECTION OF THE COEFFICIENTS OF THE ARIMA2 DAG

Lemma A.1: The conditions assuring that $H_{DAG}(z^{-1}) = \frac{1+c_1z^{-1}+c_2z^{-2}}{1-d_1'z^{-1}}$ is strictly positive real (SPR) are:

- for $c_2 \leq 0$, c_1 must be such that

$$-1 - c_2 < c_1 < 1 + c_2$$

- for $c_2 \geq 0$

- if the following condition is satisfied

$$2(d_1' - c_2) < \sqrt{2(c_2 - c_2^2)(1 - d_1'^2)} < 2(d_1' + c_2)$$

the maximum bound on c_1 is given by

$$c_1 < d_1' - 3d_1'c_2 + 2\sqrt{2(c_2 - c_2^2)(1 - d_1'^2)}$$

otherwise the maximum bound on c_1 is given by

$$c_1 < 1 + c_2$$

- if the following condition is satisfied

$$2(d_1' - c_2) < -\sqrt{2(c_2 - c_2^2)(1 - d_1'^2)} < 2(d_1' + c_2)$$

the minimum bound on c_1 is given by

$$c_1 > d_1' - 3d_1'c_2 - \sqrt{2(c_2 - c_2^2)(1 - d_1'^2)}$$

otherwise the minimum bound on c_1 is given by

$$c_1 > -1 - c_2$$

The proof of this result is given in [19].

APPENDIX B

PROOF OF LEMMA 3.3

The verification of the Benveniste smoothness, boundness and mixing conditions is similar to the verification done in [7] for the SHARF algorithm and it is omitted. The focus is on the analysis of the Benveniste's ODE (ordinary differential equation) associated to the algorithm. The algorithm of Eq. (39) can be rewritten as:

$$\hat{\mathbf{w}}(t+1) - \hat{\mathbf{w}}(t) = \mu H_{DAG}(q^{-1})[\mathbf{r}(t, \hat{\mathbf{w}})e(t+1, \hat{\mathbf{w}})] \quad (55)$$

i.e. one takes into account that $\mathbf{r}(t)$ may depend upon $\hat{\mathbf{w}}$ and that clearly $e(t+1)$ depends upon $\hat{\mathbf{w}}$. The associated ODE equation is:

$$\frac{d\hat{\mathbf{w}}}{d\tau} = \mu \mathbf{E}\{H_{DAG}[\mathbf{r}(t, \hat{\mathbf{w}})e(t+1, \hat{\mathbf{w}})]\} \quad (56)$$

But from Eqs. (38) and (8) one gets:

$$e(t+1, \hat{\mathbf{w}}) = \mathbf{r}^T(\mathbf{w} - \hat{\mathbf{w}}) + \mathbf{n}(t+1) \quad (57)$$

which leads to:

$$\begin{aligned} \frac{d\hat{\mathbf{w}}}{dt} = & -\mu \mathbf{E}\{H_{DAG}(q^{-1})[\mathbf{r}(n, \hat{\mathbf{w}})\mathbf{r}(n, \hat{\mathbf{w}})^T]\}(\hat{\mathbf{w}} - \mathbf{w}) \\ & + \mu \mathbf{E}\{H_{DAG}(q^{-1})[\mathbf{r}(n, \hat{\mathbf{w}})\mathbf{n}(t+1)]\} \end{aligned} \quad (58)$$

Since $\mathbf{n}(t+1)$ is a white noise or is a stochastic process independent of $\mathbf{r}(t, \hat{\mathbf{w}})$, the second term of the right hand side of Eq. (58) will be null. Therefore, the properties of the algorithm for large τ and small μ will be determined by the stability of:

$$\frac{d\hat{\mathbf{w}}}{d\tau} = -\mu G_{\mathbf{w}}(\hat{\mathbf{w}} - \mathbf{w}) \quad (59)$$

where:

$$G_{\mathbf{w}} = \mathbf{E}\{H_{DAG}(q^{-1})[\mathbf{r}(n, \hat{\mathbf{w}}), \mathbf{r}(n, \hat{\mathbf{w}})^T]\} \quad (60)$$

$$= H_{DAG}(q^{-1})\mathbf{E}_{\mathbf{r}} \quad (61)$$

where $\mathbf{E}_{\mathbf{r}} = \mathbf{E}[\mathbf{r}(n, \hat{\mathbf{w}}), \mathbf{r}(n, \hat{\mathbf{w}})^T]$. The asymptotic stability of the above equation can be assessed using a Lyapunov function candidate:

$$V(\hat{\mathbf{w}}) = (\hat{\mathbf{w}} - \mathbf{w})^T \mu^{-1}(\hat{\mathbf{w}} - \mathbf{w}) \quad (62)$$

The derivative of the Lyapunov function candidate along the trajectories of the system has the form:

$$\dot{V}(\hat{\mathbf{w}}) = -(\hat{\mathbf{w}} - \mathbf{w})^T [G_{\mathbf{w}} + G_{\mathbf{w}}^T](\hat{\mathbf{w}} - \mathbf{w}) \quad (63)$$

It remains to show that $G_{\mathbf{w}} + G_{\mathbf{w}}^T$ is positive definite. To do this, we will use the integration on the unit circle. One obtains:

$$G_{\mathbf{w}} + G_{\mathbf{w}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} [H_{DAG}(e^{j\omega})\Psi(\omega) \quad (64)$$

$$+ \Psi(\omega)^T H_{DAG}(e^{-j\omega})] d\omega \quad (65)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} 2\text{Re}[H_{DAG}^{ii}(e^{j\omega})]\Psi(\omega) d\omega \quad (66)$$

taking into account that $\Psi(\omega)$, the power spectral density of $\mathbf{r}(t, \mathbf{w})$ is symmetric. Since H_{DAG}^{ii} is SPR and $\Psi(\omega)$ is positive definite (as a consequence of condition of Eq. (40)), one concludes that $G_{\mathbf{w}} + G_{\mathbf{w}}^T$ is positive definite.

REFERENCES

- [1] T.-B. Airimitoie and I. D. Landau, "Improving adaptive feedforward vibration compensation by using integral-proportional adaptation," *Automatica*, vol. 49, no. 5, pp. 1501–1505, 2013.
- [2] T.-B. Airimitoie, I. D. Landau, R. Melendez, and L. Dugard, "Algorithms for Adaptive Feedforward Noise Attenuation—A Unified Approach and Experimental Evaluation," *IEEE Transactions on Control Systems Technology*, vol. 29, no. 5, pp. 1850–1862, Sep. 2021.
- [3] A. E. Albert and L. A. Gardner, *Stochastic approximation and nonlinear regression*. Cambridge, Mass.: M.I.T. Press, 1967, oCLC: 973152.
- [4] B. Anderson, R. Bitmead, C. Johnson, P. Kokotovic, R. Kosut, I. Mareels, L. Praly, and B. Riedle, *Stability of adaptive systems*. Cambridge Massachusetts, London, England: The M.I.T Press, 1986.
- [5] A. Benveniste, M. Goursat, and G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithms," *IEEE Transactions on Automatic Control*, vol. 25, no. 6, pp. 1042–1058, 1980.
- [6] D. Bismor, K. Cyz, and Z. Ogonowski, "Review and Comparison of Variable Step-Size LMS Algorithms," *The International Journal of Acoustics and Vibration*, vol. 21, no. 1, 2016.
- [7] H. Fan, "Application of benveniste's convergence results in the study of adaptive iir filtering algorithms," *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 692–709, 1988.
- [8] J. E. Gaudio, T. E. Gibson, A. M. Annaswamy, and M. A. B. E. Lavretsky, "Connections between adaptive control and optimization in machine learning," in *Proceedings of the IEEE CDC Conference, Dec. 2019, Nice, France*, 2019, pp. 4563–4568.
- [9] G. Goodwin and K. Sin, *Adaptive filtering prediction and control*. N. J.: Prentice Hall, 1984.
- [10] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs N.J.: Prentice Hall, 1996.
- [11] R. Jacobs, "Increased rates of convergence through learning rate adaptation," *Neural Networks*, vol. 27, no. 1, pp. 295–307, Winter 1988.
- [12] J. Johnson, C., "A convergence proof for a hyperstable adaptive recursive filter (corresp.)," *Information Theory, IEEE Transactions on*, vol. 25, no. 6, pp. 745 – 749, nov 1979.
- [13] I. D. Landau, R. Lozano, M. M'Saad, and A. Karimi, *Adaptive control*, 2nd ed. London: Springer, 2011.
- [14] I. Landau, *System Identification and Control Design: Using P.I.M.+ Software*, ser. Information and System Sciences Series. Prentice Hall, 1990.
- [15] I. D. Landau, "Synthesis of hyperstable discrete model reference adaptive systems," in *Fifth Asilomar Conference on Circuits and Systems*, Asilomar, California, USA, Nov. 1971.
- [16] —, "Design of discrete model reference adaptive systems using the positivity concept," in *Proceedings 3rd IFAC Symp. on Sensitivity, Adaptivity and Optimality*, Ischia, Italy, June 1973, pp. 307–314.
- [17] I. D. Landau and T.-B. Airimitoie, "Does a general structure exist for adaptation/learning algorithms?" in *CDC 2022 - 61st IEEE Conference on Decision and Control*, Cancun, Mexico, Dec. 2022. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-03738223>
- [18] I. D. Landau, T.-B. Airimitoie, and M. Alma, "A Youla–Kučera parametrized adaptive feedforward compensator for active vibration control with mechanical coupling," *Automatica*, vol. 48, no. 9, pp. 2152 – 2158, 2012.
- [19] I. D. Landau, T.-B. Airimitoie, B. Vau, and G. Buche, "On a general structure for adaptation/learning algorithms - stability and performance issues," *Automatica*, vol. 156, p. 111193, Mar. 2023.
- [20] I. D. Landau, M. Alma, and T.-B. Airimitoie, "Adaptive feedforward compensation algorithms for active vibration control with mechanical coupling," *Automatica*, vol. 47, no. 10, pp. 2185 – 2196, 2011.
- [21] I. D. Landau and G. Béthoux, "Algorithms for discrete time model reference adaptive systems," in *6th IFAC World Congress (IFAC 1975)*, Boston, USA, August 1975, pp. 682–692.
- [22] I. D. Landau, B. Vau, T.-B. Airimitoie, and G. Buche, "Improving performance of adaptive feedforward noise attenuators using a dynamic adaptation gain," *Journal of Sound and Vibration*, vol. 560, p. 117790, 2023.
- [23] M. Larimore, J. Treichler, and J. Johnson, C., "SHARF: An algorithm for adapting IIR digital filters," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 4, pp. 428 – 440, aug 1980.
- [24] I. Livieris and R. Pintelas, "A survey on algorithms for training artificial neural networks," *University of Patras Report*, vol. 455, no. TR08-01, 2014.
- [25] J. Nagumo and A. Noda, "A learning method for system identification," *IEEE Transactions on Automatic Control*, vol. 12, no. 3, pp. 282–287, 1967.
- [26] S. Pouyanfar, S. Sadiq, Y. Yan, and H. Tian, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys, Vol. 1, No. 1, Article 1. Publication date: January 2017.*, vol. 1, January 2017.
- [27] M. Schmidt, "Stochastic average gradient," *University of British Columbia.*, Winter 2018.
- [28] H. P. Whitaker, J. Yarmon, and A. Keezer, *Synthesis of model reference adaptive control systems by Lyapunov's second method*. Cambridge, USA: M.I.T. Instrumentation Laboratory, Report R-164, September 1958.
- [29] B. Widrow, J. Glover, J.R., J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, J. Eugene Dong, and R. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692 – 1716, dec. 1975.
- [30] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *1960 IRE WESCON Convention Record, Part 4*. New York: IRE, 1960, pp. 96–104.