Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/XXXX.2022.1234567

Low cost variable step-size LMS with maximum similarity to the affine projection algorithm

Miguel Ferrer, Member,IEEE, María de Diego, Senior Member, IEEE, and Alberto Gonzalez, Senior Member, IEEE

[1]Institute of Telecommunications and Multimedia Applications (iTEAM), Universitat Politècnica de València (UPV), Valencia, 46022 SPAIN

Corresponding author: Alberto Gonzalez (email: agonzal@dcom.upv.es).

This work was supported in part by MICIN under grant PID2021-125736OB-I00 (MCIN/AEI/10.13039/501100011033/, "ERDF A way of making Europe"), by GVA under grant CIPROM/2022/20 and by EU under IN-NOVA: Active reduction of noise transmitted into and from enclosures through encapsulated structures, Marie Sklodowska-Curie grant agreement no. 101073037.

ABSTRACT

The LMS algorithm is widely employed in adaptive systems due to its robustness, simplicity, and reasonable performance. However, it is well known that this algorithm suffers from a slow convergence speed when dealing with colored reference signals. Numerous variants and alternative algorithms have been proposed to address this issue, though all of them entail an increase in computational cost. Among the proposed alternatives, the affine projection algorithm stands out. This algorithm has the peculiarity of starting from N data vectors of the reference signal. It transforms these vectors into as many data vectors suitably normalized in energy and mutually orthogonal. In this work, we propose a version of the LMS algorithm that, similar to the affine projection algorithm, starts from N data vectors of the reference signal but corrects them by using only a scalar factor that functions as a convergence step. Our goal is to align the behavior of this algorithm with the behavior of the affine projection algorithm without significantly increasing the computational cost of the LMS.

INDEX TERMS Adaptive filters, affine projection algorithm, variable step-size.

I. Introduction

THE LMS algorithm [1] is one of the most widely used adaptive algorithms in signal processing and other fields to iteratively reach the solution of a mimicking problem. This algorithm has become the reference algorithm for problems that allow an iterative solution, as it is a simple and efficient method. However, a notable drawback emerges when the input signal is colored, leading to a reduction in its convergence speed [2]. Consequently, alternative algorithms or modified LMS versions have arisen, aiming to enhance its robustness and speed in the presence of colored signals. Among these alternatives is the Affine Projection (AP) algorithm [3], which allows the algorithm's behaviour to be adjusted through the parameter projection order, denoted as N. As N increases, so does the computational cost, with the cost increasing in a cubic fashion with N. However, this increase in computational demand is justified by the significantly improved convergence speed, achieved by using a matrix correction in the coefficient fitting to orthogonalize the reference signal data, even when the reference signal is colored. Several AP-based algorithms have been proposed to overcome its computational burden, [4], [5], [6], [7], or other issues as the trade-off between convergence speed and misadjustment, [8], [9], or its robustness against impulsive noise, [10], [11], [12], [13], [14]. Our objective is to enhance the performance of the LMS algorithm in a similar way to the affine projection algorithm, but without incurring a significant increase in computational cost. To this end, we introduce an extended LMS algorithm that operates on N vectors from the reference signal. We achieve this with reduced computational cost by using only a single scalar to control its adaptation equation, which could be understood as a variable step-size algorithm. Several approaches of variable step-size algorithms have been proposed for the LMS, [15],

[16], [17], [18], and for the AP, [8], [19], [20], [21]. Our proposal integrates elements from both algorithms.

The outline of this paper is as follows: In Section II, we review the algorithms and versions of the algorithms most related to the new approach. The proposed algorithm is described in Section III, followed by a discussion on its performance in Section IV. The theoretical mean convergence analysis and its mean square performance are both presented in Sections V and VI, respectively. Finally, some simulation results and conclusions are included in Sections VII and VIII, respectively.

II. LMS, NLMS, Affine Projection and Affine Projection like algorithms

Given a system in which the goal is to estimate a desired signal $d(n)$ from the filtering of a reference signal $x(n)$ by an FIR filter with L coefficients ($\mathbf{w}(n)$), the LMS algorithm proposes to find the solution to this problem by iteratively adjusting the coefficients of the filter $\mathbf{w}(n)$ according to:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}_L(n) e^a(n), \tag{1}$$

where $\mathbf{x}_L(n)$ is a vector containing the last L samples of the reference signal, and $e^a(n) = d(n) - \mathbf{x_L}^T(n)\mathbf{w}(n-1)$ is usually called the a priori error signal. μ is a positive constant that controls how fast the algorithm can modify its current solution, related to the speed and convergence conditions of the algorithm, and is usually called the convergence step. This equation allows, under certain conditions, to estimate the optimal coefficients $\mathbf{w}^{opt}(n)$ that minimise the mean power of the error signal, defined as $e(n) = d(n) - \mathbf{x}_L^T(n)\mathbf{w}$, which is expressed as

$$\mathbf{w}^{opt} = \min_{\mathbf{w}} \left\{ \frac{1}{2} \mathrm{E}\left\{ (d(n) - \mathbf{x}_L^T(n)\mathbf{w})^2 \right\} \right\}. \tag{2}$$

The LMS algorithm avoids the direct solution of (2), which requires knowledge of signal statistics. Instead, it achieves an iterative solution through a stochastic gradient algorithm, relying only on the instantaneous data of the signals to iteratively approach the optimal solution of the problem. The use of instantaneous data to solve a statistical problem has the disadvantage of its variance, but it can be shown that on average the solution obtained by the LMS algorithm coincides with the optimal solution given by (2).

The LMS algorithm gives rise to a large family of variants, most notably its normalised version, NLMS (Normalised Least Mean Square Algorithm) [22]:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{1}{\mathbf{x}_L^T(n)\mathbf{x}_L(n)} \mathbf{x}_L(n) e^a(n). \tag{3}$$

The NLMS algorithm can be understood as the LMS, where the convergence step used in (1) is the one that minimises the instantaneous power of the a posteriori error signal, defined as $e^p(n) = d(n) - \mathbf{x}_L^T(n)\mathbf{w}(n)$. This version of the LMS algorithm makes it possible to automatically adjust the value of the convergence step, thereby maximising its convergence speed and enhancing the algorithm's adaptibility to potential changes in the energy of the reference signal. It can be shown that when the data between $\mathbf{x}_L(n)$ and $\mathbf{x}_L(n-1)$ are orthogonal ($\mathbf{x}_L^T(n)\mathbf{x}_L(n-1) \approx 0$), the convergence speed of this algorithm is maximal. However, when there is a correlation between them, the convergence speed decreases.

Among the many alternatives proposed to address the convergence speed issue in LMS-derived algorithms, especially when dealing with highly colored reference signals, the AP algorithm proposed in [3] stands out. This algorithm solves the problem of modeling the variation of the coefficients between iterations, denoted as $\|\mathbf{w}(n) - \mathbf{w}(n-1)\|^2$, subject to the N constraints given by $d(n-k) = \mathbf{x}_L^T(n-k)\mathbf{w}(n)$ (for $0 \leq k \leq N-1$). It is sometimes considered an extension of the NLMS algorithm, as when $N = 1$, it coincides with that algorithm. Both the AP algorithm and its variants are widely used in several applications, such as: echo cancellation [23], [24], active noise control (ANC) [25], [26], [27], [28], noise reduction [29], system identification [30], [31], beamforming [32], [33], [34] and acoustic feedback cancellation [35], [36], [37], among others.

Equation (4) represents the equation used to update the coefficients of the AP algorithm. It is worth noting that this expression entails much higher computational cost compared to the expressions of the LMS, as shown in (1), and the NLMS, as shown in (3):

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{X}(n)[\mathbf{X}^T(n)\mathbf{X}(n)]^{-1}\mathbf{e}^a(n). \tag{4}$$

The a priori error vector used in (4) is defined as $\mathbf{e}^a(n) = \mathbf{d}(n) - \mathbf{X}^T(n)\mathbf{w}(n-1)$, being $\mathbf{d}(n)$ a vector with the last N ($N \leq L$) samples of the signal $d(n)$ and $\mathbf{X}(n) = [\mathbf{x}_L(n), \mathbf{x}_L(n-1), \cdots, \mathbf{x}_L(n-N+1)]$ a data matrix of size $L \times N$. The convergence speed of the AP algorithm increases with N, but so does the computational complexity and numerical instability of the computation $(\mathbf{X}^T(n)\mathbf{X}(n))^{-1}$. Equation (4) is usually modified by introducing two parameters with practical meaning: a convergence step $0 < \mu \leq 1$ and a regularisation factor δ. The modified equation is then given by

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{X}(n)[\mathbf{X}^T(n)\mathbf{X}(n) + \delta \mathbf{I}_N]^{-1}\mathbf{e}^a(n), \tag{5}$$

where $\mathbf{I}_N$ is the identity matrix of size $N \times N$. The regularisation factor is necessary to avoid instability in the matrix inversion, especially when the data in the columns of $\mathbf{X}(n)$ are closely related to linear combinations of other columns, resulting in an ill-conditioned matrix that is challenging to invert numerically (or even non-invertible) in practice. This may happen when dealing with highly correlated signals. If the value of δ required to guarantee matrix inversion is large, regularised matrix inversion introduces a bias in the optimal solution for the filter coefficients and also a deviation in the behaviour of the algorithm during convergence, which may require the use of a value of $\mu < 1$ to avoid divergence. For small δ, the algorithm usually converges for $\mu = 1$. The consequences of using the regularisation factor are analysed in [38]–[41]. These works also proposed several strategies to

identify the regularisation factor by optimising the behaviour of the AP algorithm.

Considering that the most significant computational cost in (4) and (5) lies in the matrix inversion, it is desirable to either avoid or optimise this process. There are a number of proposals in this regard, including methods for obtaining the exact matrix inversion [42]–[46] as well as for introducing certain approximations [47]–[51]. In [52], an alternative class of algorithms called affine projection-like (APL) is proposed. In these algorithms, the update equation avoids matrix inversion and is given by

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu(n)\mathbf{X}(n)\mathbf{e}^a(n). \tag{6}$$

This update expression is obtained by solving a minimisation problem similar to (2), but considering instantaneous values and a number of samples of the signal $d(n)$ and vectors of the reference signal equal to N, i.e:

$$\mathbf{w}_{APL}^{opt} = \min_{\mathbf{w}} \left\{ \frac{1}{2}[\mathbf{d}(n) - \mathbf{X}^T(n)\mathbf{w}]^T[\mathbf{d}(n) - \mathbf{X}^T(n)\mathbf{w}] \right\}. \tag{7}$$

We can interpret this solution as an extension of the LMS algorithm using more than one dimension ($N > 1$), or as a version of the unconstrained affine projection algorithm. It can also be considered as an approximate version of the AP algorithm by replacing the matrix $[\mathbf{X}^T(n)\mathbf{X}(n)]^{-1}$ with the identity matrix $\mathbf{I}_N$. This approximation is quite accurate when the reference signal is white noise of unit average power.

Although (6) can converge faster than the NLMS, its convergence speed is limited by the dispersion of the N non-zero (and positive) eigenvalues of the $[L \times L]$-size matrix, $\mathbf{X}(n)\mathbf{X}^T(n)$. It is suggested in [52] that:

$$0 < \mu(n) < \mu_{\max}(n) = \frac{2}{\lambda_{\max}(n)}, \tag{8}$$

where $\lambda_{\max}(n)$ is the maximum eigenvalue of $\mathbf{X}(n)\mathbf{X}^T(n)$, which can be bounded by the sum of the non-zero eigenvalues of $\mathbf{X}(n)\mathbf{X}^T(n)$, given by $\lambda_\Sigma(n) = \sum_{l=1}^{N} \lambda_l(n)$. Although satisfying the upper bound of (8) is not a necessary condition to guarantee convergence, it is a sufficient condition.

When the signal is highly colored, $\lambda_{\max}(n) \to \lambda_\Sigma(n)$, while when the signal is white, $\lambda_{\max}(n) \to \lambda_\Sigma(n)/N$. That is, the maximum convergence step that would give the highest convergence speed while satisfying a sufficient stability condition must be in the range:

$$\frac{2}{\lambda_\Sigma(n)} \leq \mu_{\max}(n) \leq \frac{2N}{\lambda_\Sigma(n)}. \tag{9}$$

Therefore, the APL reduces the computational cost of the AP at the expense of slower convergence as the reference signal becomes more colored (up to a maximum of N times).

The calculation of $\lambda_\Sigma(n)$ can be performed using the reference signal data as:

$$\lambda_\Sigma(n) = \mathrm{Tr}\left\{\mathbf{X}^T(n)\mathbf{X}(n)\right\} = \sum_{p=0}^{N-1} \mathbf{x}_L^T(n-p)\mathbf{x}_L(n-p), \tag{10}$$

where the operator denoted by $\mathrm{Tr}\{\mathbf{A}\}$ gives the trace of a given matrix $\mathbf{A}$.

The APL algorithm applies a $\mu(n)$ value of $\mu_{max}(n)/2$ when $N = 1$, which aligns it with NLMS (when dealing with white signals, $\lambda_\Sigma(n) \approx N\mathbf{x}_L^T(n)\mathbf{x}_L(n)$, with $\mu_{\max}(n) = 2/\mathbf{x}_L^T(n)\mathbf{x}_L(n)$). In the event that the nature of the reference signal is unknown, it is imperative to limit the convergence step to the worst-case scenario. This leads to a slower convergence rate compared to its maximum speed when the reference signal lacks coloration, as well as a slower rate compared to the AP algorithm. It is suggested in [52] to use the value of $\mu(n)$ that minimises the squared 2-norm of the a posteriori error vector ($\|\mathbf{e}^p(n)\|^2$) to address this reduction in convergence speed (the a posteriori error vector is equal to zero in the AP algorithm). This means

$$\mu_I(n) = \min_{\mu(n)} \left\{ \frac{1}{2}\|\mathbf{d}(n) - \mathbf{X}^T(n)\mathbf{w}(n)\|^2 \right\}, \tag{11}$$

which is given by

$$\mu_I(n) = \frac{\|\mathbf{X}(n)\mathbf{e}^a(n)\|^2}{\|\mathbf{X}^T(n)\mathbf{X}(n)\mathbf{e}^a(n)\|^2}, \tag{12}$$

and defines the algorithm called affine-like-I (APL-I) [52].

The affine-like-I algorithm requires $2LN + 3N$ multiplications to update the coefficients, which is an intermediary cost between the NLMS algorithm ($2L$) and the exact AP ($N^3 + N^2(L+1) + LN$). It is considered that the calculation for matrix inversion with a size of $N \times N$ requires N^3 multiplications.

III. Variable step-size LMS with maximum similarity to the AP algorithm

It may be inferred that an adaptive algorithm would exhibit similar behaviour to a given one if its coefficients were very close at each algorithm iteration. Therefore, we propose the use of a variable convergence step that minimises the squared 2-norm of the difference between the coefficients of the exact AP algorithm (with $\mu = 1$ for clarity and without loss of generality), denoted as $\mathbf{w}_{AP}(n)$ and shown in (4), and the approximate version, denoted as $\mathbf{w}_{APL}(n)$ and shown in (6). This means

$$\tilde{\mu}(n) = \arg\min_{\mu(n)} \left\{ \|\mathbf{w}_{AP}(n) - \mathbf{w}_{APL}(n)\|^2 \right\}, \tag{13}$$

or equivalently

$$\tilde{\mu}(n) = \arg\min_{\mu(n)} \left\{ \|\mathbf{X}(n)(\mathbf{X}(n)^T\mathbf{X}(n))^{-1} - \mu(n)\mathbf{X}(n))\mathbf{e}^a(n)\|^2 \right\}, \tag{14}$$

which is given by

$$\tilde{\mu}(n) = \frac{(\mathbf{e}^a(n))^T\mathbf{e}^a(n)}{[\mathbf{X}(n)\mathbf{e}^a(n)]^T\mathbf{X}(n)\mathbf{e}^a(n)} = \frac{\|\mathbf{e}^a(n)\|^2}{\|\mathbf{e}^a(n)\|^2_{\mathbf{\Sigma}(n)}}, \tag{15}$$

where $\mathbf{\Sigma}(n) = \mathbf{X}(n)^T\mathbf{X}(n)$. Thus, the proposed approach uses the update equation in (6), just like the AP and the APL-I, except that the convergence step is obtained by solving the minimization problem in (13). This approach would require $LN + 3N$ multiplications for updating the coefficients, which is a lower count compared to the AP and the APL-I algorithms.

IV. Convergence discussion

Equation (15) is a generalised Rayleigh quotient [53]. When $\mathbf{\Sigma}(n)$ is positive definite, its maximum and minimum bounds are determined by the maximum and minimum eigenvalues of $(\mathbf{\Sigma}(n))^{-1}$, which are equal to the inverses of the eigenvalues of $\mathbf{\Sigma}(n)$. This ensures that the following boundaries are satisfied

$$\frac{1}{\lambda_{\max}(n)} = \lambda_{\min(\mathbf{\Sigma}^{-1})} \leq \tilde{\mu}(n) \leq \lambda_{\max(\mathbf{\Sigma}^{-1})} = \frac{1}{\lambda_{\min}(n)}. \tag{16}$$

Equation (16) is also fulfilled by $\mu_I(n)$ suggested by APL-I in (12), since the same relation can also be obtained through the use of the generalized Rayleigh quotient with the eigenvalues of $\left(\mathbf{\Sigma}^2\right)^{-1}\mathbf{\Sigma}$ (which is equal to $\mathbf{\Sigma}^{-1}$).

The proposed convergence step, $\tilde{\mu}(n)$, cannot ensure that (8) is fulfilled unless $\lambda_{\max}(n)/\lambda_{\min}(n) < 2$. Therefore, we can only guarantee that (8) is fulfilled when the eigenvalues of the matrix $\mathbf{X}^T(n)\mathbf{X}(n)$ are not sparse. Nevertheless the nature of the reference signal and the values of L and N influence these eigenvalues. This influence arises because the eigenvalues are computed from the N initial values of the temporal autocorrelation of the reference signal, which is windowed with a sliding rectangular window of size L. To reduce the time dependency of (16) for stationary signals, a large L value should be used.

However, the breach of $\lambda_{\max}(n)/\lambda_{\min}(n) < 2$ does not indicate whether the convergence step satisfies (8) or not. Furthermore, failing to satisfy (8) does not necessarily indicate algorithm divergence. Consequently, analysing the long-term convergence of the algorithm is challenging.

On the other hand, the similarity condition, given by the difference between the coefficients of the proposed algorithm and the AP (see (13)), can be considered as an indicator of the deviation in the convergence behaviour of the proposed algorithm and the AP. It can be shown that this difference can be bounded by substituting (15) into (14) and calculating the energy of this difference as

$$\|\mathbf{w}_{AP}(n) - \mathbf{w}_{APL}(n)\|^2 = (\mathbf{e}^a(n))^T[\mathbf{X}^T(n)\mathbf{X}(n)]^{-1}\mathbf{e}^a(n) - \frac{(\|\mathbf{e}^a(n)\|^2)^2}{(\mathbf{e}^a(n))^T\mathbf{X}^T(n)\mathbf{X}(n)\mathbf{e}^a(n)} \leq 2\frac{\|\mathbf{e}^a(n)\|^2}{\lambda_{\min}(n)}. \tag{17}$$

Therefore, as the minimum eigenvalue $\lambda_{\min}(n)$ increases, the difference between the coefficient vectors decreases. Since the energy of the a priori error signal decreases over time (indicating algorithm convergence), the algorithm becomes more sensitive to the appearance of small $\lambda_{\min}(n)$ values in the initial stages (during the transient period). When the a priori error energy becomes small, the proposed algorithm tends to exhibit behavior similar to the AP, regardless of the reference signal. That is, the largest differences can occur during the transient period when small $\lambda_{\min}(n)$ values appear. Note that for colored signals, increasing the value of N does not necessarily improve the behavior of the proposed algorithm. In such cases, the dispersion of the eigenvalues may increase with higher values of N.

Equation (15) can be rewritten as

$$\tilde{\mu}(n) = \frac{\|\mathbf{e}^a(n)\|^2}{\sum_{k=1}^{N}\lambda_k(n)|\mathbf{u}_k^T(n)\mathbf{e}^a(n)|^2} = \frac{1}{\sum_{k=1}^{N}\lambda_k(n)(\cos\theta_k(n))^2}, \tag{18}$$

where $\lambda_k(n)$ and $\mathbf{u}_k(n)$ $(1 \leq k \leq N)$ are respectively the eigenvalues and eigenvectors of the $\mathbf{\Sigma}(n)$ matrix, which depend only on the reference signal data at n time instant. The angle $\theta_k(n)$ (for $1 \leq k \leq N$) is the angle that creates the a priori error vector with each of the eigenvectors of matrix $\mathbf{\Sigma}(n)$.

When $N = 1$, the proposed convergence parameter coincides with the NLMS and consequently only depends on the reference signal data. In contrast, as N increases, its dependence on the a priori error signal increases. This dependence will be greater if the reference signal is highly colored because the eigenvalues of $\mathbf{\Sigma}(n)$ will be more sparse. In the particular case of a low colored signal, it is satisfied that $\lambda_k(n) \approx \lambda_{\max}(n), \forall k$, and therefore:

$$\tilde{\mu}(n) \approx \frac{1}{\lambda_{\max}(n)\sum_{k=1}^{N}(\cos\theta_k(n))^2} = \frac{1}{\lambda_{\max}(n)}, \tag{19}$$

being $\tilde{\mu}(n)$ dominated only by the data of the reference signal. On the other hand, in the case of a highly colored signal, it holds for small values of N that $\sum_{k=1}^{N}\lambda_k(n)(\cos\theta_k)^2 \approx \lambda_{\max}(n)(\cos\theta_{k_{\max}})^2$, where $k_{\max}$ is the index that determines the maximum eigenvalue and therefore:

$$\tilde{\mu}(n) \approx \frac{1}{\lambda_{\max}(n)(\cos\theta_{k_{\max}}(n))^2}. \tag{20}$$

In this case, the value of $\tilde{\mu}(n)$ is found to be more dependent on the a priori error, which can lead to undesired results in the behaviour of the algorithm in time periods where the a priori error is not small (during transients). This dependence on the a priori error will be greater as N increases, since more angle-dependent terms $\theta_k(n)$ will be included in the denominator of (20). This circumstance can be minimised by pre-whitening the reference signal or by adding a regularisation factor to the calculation of the denominator of (15).

Note that this dependence of the convergence step on the a priori error is not as significant in the APL-I algorithm, where:

$$\mu_I(n) = \frac{\sum_{k=1}^{N}\lambda_k(n)(\cos\theta_k)^2}{\sum_{k=1}^{N}(\lambda_k(n))^2(\cos\theta_k)^2}, \tag{21}$$

that for both low colored and highly colored signals provides a convergence step value that depends only on the reference

signal as

$$\mu_I(n) \approx \frac{1}{\lambda_{\max}(n)}. \tag{22}$$

Therefore, it is recommended to incorporate a regularization factor $0 < \alpha < 1$ in the product $\mathbf{X}^T(n)\mathbf{X}(n)$ of the denominator of (15) for highly colored signals and high values of N. This would give rise to the following equation:

$$\tilde{\mu}_\alpha(n) = \frac{\|\mathbf{e}^a(n)\|^2}{\|\mathbf{e}^a(n)\|^2_{\mathbf{\Sigma}+\alpha\mathbf{I}}} = \frac{\|\mathbf{e}^a(n)\|^2}{\|\mathbf{e}^a(n)\|^2_{\mathbf{\Sigma}} + \alpha\|\mathbf{e}^a(n)\|^2}, \tag{23}$$

and the convergence parameter boundaries in (16) can be rewritten as

$$\frac{1}{\lambda_{\max}(n)+\alpha} \leq \tilde{\mu}_\alpha(n) \leq \frac{1}{\lambda_{\min}(n)+\alpha}. \tag{24}$$

Hence the difference between the algorithms given by (17) becomes less dependent on $\lambda_{min}(n)$ at the expense of decreasing the convergence speed.

V. Mean convergence analysis

Considering that $\mathbf{d}(n) = \mathbf{X}^T(n)\mathbf{w}^o$, and defining $\tilde{\mathbf{w}}(n) = \mathbf{w}^o - \mathbf{w}(n)$, (6) can be rewritten for the proposed variable convergence step as

$$\tilde{\mathbf{w}}(n) = (\mathbf{I} - \tilde{\mu}(n)\mathbf{X}(n)\mathbf{X}^T(n))\tilde{\mathbf{w}}(n-1). \tag{25}$$

Taking mean values and assuming statistical independence between the reference signal data and the shifted coefficient vectors, $\tilde{\mathbf{w}}(n)$ and $\tilde{\mathbf{w}}(n-1)$, we obtain

$$\mathrm{E}\{\tilde{\mathbf{w}}(n)\} = (\mathbf{I} - \mathrm{E}\{\tilde{\mu}(n)\mathbf{X}(n)\mathbf{X}^T(n)\})\mathrm{E}\{\tilde{\mathbf{w}}(n-1)\}, \tag{26}$$

where the mathematical expectation is denoted by $\mathrm{E}\{\cdot\}$.

We define $\mathrm{E}\{\tilde{\mu}(n)\mathbf{X}(n)\mathbf{X}^T(n)\} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, being $\mathbf{\Lambda}$ a diagonal matrix with the eigenvalues of $\mathrm{E}\{\tilde{\mu}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}$ along its diagonal. Then the rotation of (26) by the unitary matrix $\mathbf{Q}^T$ gives

$$\tilde{\mathbf{w}}'(n) = (\mathbf{I} - \mathbf{\Lambda})\,\tilde{\mathbf{w}}'(n-1), \tag{27}$$

which can be recursively expressed starting from an arbitrary initialisation of the weights as

$$\tilde{\mathbf{w}}'(n) = (\mathbf{I} - \mathbf{\Lambda})^n\,\tilde{\mathbf{w}}'(0). \tag{28}$$

As discussed in Section II, (8) presents a sufficient condition for the convergence of APL algorithms, which is the type proposed here. This is because the eigenvalues of $\mathrm{E}\{\tilde{\mu}(n)\mathbf{X}(n)\mathbf{X}^T(n)\}$ will all be less than 2, and $\lim_{n\to\infty}\tilde{\mathbf{w}}'(n) = 0$ in (28). Using (16), the sufficient condition in this case can be expressed as the quotient $\lambda_{max}(n)/\lambda_{min}(n) \leq 2$, or $\lambda_{max}(n)/(\lambda_{min}(n)+\alpha) \leq 2$ with regularization. This condition is met for low colored signals, since its eigenvalue ratio is close to one regardless of the value of N, provided that L is sufficiently large. For colored signals, it is not possible to approximate $\tilde{\mu}$ using the equation given in (19) and the sufficient condition cannot be guaranteed for all values of n.

For small values of $\mathbf{e}^a(n)$, it is possible to consider the error as independent of the data, so that a statistical independence can be assumed between the reference signal data and the convergence step $\tilde{\mu}(n)$, that is

$$\mathrm{E}\left\{\tilde{\mu}(n)\mathbf{X}(n)\mathbf{X}^T(n)\right\} = \mathrm{E}\left\{\tilde{\mu}(n)\right\}\mathrm{E}\left\{\mathbf{X}(n)\mathbf{X}^T(n)\right\}. \tag{29}$$

For these small values of $\mathbf{e}^a(n)$ the values of $\tilde{\mu}(n)$ are small and very close (always positive), so it is possible to make the following approximation using (18) [54]:

$$\mathrm{E}\left\{\tilde{\mu}(n)\right\} \approx \frac{1}{\mathrm{E}\left\{\sum_{k=1}^N \lambda_k(n)[\cos(\theta_k(n))]^2\right\}}. \tag{30}$$

In this case of independence between the a priori error and the reference signal data, $\theta_k(n)$ can be considered a random variable uniformly distributed between $-\pi$ and π and independent of the reference signal, hence it follows that:

$$\mathrm{E}\left\{\tilde{\mu}(n)\right\} \approx \frac{2}{\sum_{k=1}^N \mathrm{E}\{\lambda_k(n)\}} < \frac{2}{\lambda_{\max}}, \tag{31}$$

where λ_{max} is the maximum eigenvalue of $\mathrm{E}\left\{\mathbf{X}(n)\mathbf{X}^T(n)\right\}$. In this way, the convergence in mean of the proposed algorithm is proved in this case, since (28) tends to zero with $n \to \infty$ when $\mathrm{E}\left\{\tilde{\mu}(n)\right\} < 2/\lambda_{max}$.

In the case of large values of $\mathbf{e}^a(n)$ and small values of N, the parameter $\tilde{\mu}(n)$ is governed by (20). Then the a priori error can be considered to be highly data-dependent and is aligned with the eigenvalue corresponding to the largest eigenvalue of $\mathbf{X}(n)\mathbf{X}^T(n)$, accordingly:

$$\tilde{\mu}(n) \approx \frac{1}{\lambda_{\max}(n)}, \tag{32}$$

which ensures convergence in this case.

Finally, when the signal is highly colored and N is large, the algorithm's behaviour may deviate from the desired outcome during convergence. Unlike the AP algorithm, whose behaviour remains stable as N increases for colored signals, this algorithm may worsen its convergence during the transient phase as N increases above the value that saturates the convergence speed of the AP algorithm. This effect is less pronounced when regularisation is employed.

VI. Mean square error performance

The steady-state mean squared error (MSE) of the proposed algorithm is defined as: $\mathrm{MSE} = \lim_{n\to\infty}\mathrm{E}\left\{(e(n))^2\right\}$, where $e(n)$ is the first element of the error vector in (6), given by $e(n) = d(n) + v(n) - \mathbf{x}(n)^T\mathbf{w}(n-1)$. Here, $v(n)$ is an additive zero-mean measurement noise that accounts for modeling errors. The error vector and the convergence parameter become, respectively:

$$\mathbf{e}(n) = \mathbf{d}(n) - \mathbf{X}^T(n)\mathbf{w}(n-1) + \mathbf{v}(n) \tag{33}$$

and

$$\tilde{\mu}(n) = \frac{\mathbf{e}^T(n)\mathbf{e}(n)}{\mathbf{e}^T(n)\mathbf{X}^T(n)\mathbf{X}(n)\mathbf{e}(n)}. \tag{34}$$

Equation (6) can be expressed as a function of the coefficients $\tilde{\mathbf{w}}(n)$, resulting in:

$$\tilde{\mathbf{w}}(n) = \tilde{\mathbf{w}}(n-1) - \tilde{\mu}(n)\mathbf{X}(n)\mathbf{e}(n). \tag{35}$$

By defining $\tilde{\mathbf{e}}^p(n) = \mathbf{X}^T(n)\tilde{\mathbf{w}}(n)$ and $\tilde{\mathbf{e}}^a(n) = \mathbf{X}^T(n)\tilde{\mathbf{w}}(n-1)$, the relationship between the error vectors becomes:

$$\tilde{\mathbf{e}}^p(n) = \tilde{\mathbf{e}}^a(n) - \tilde{\mu}(n)\mathbf{X}^T(n)\mathbf{X}(n)\mathbf{e}(n), \tag{36}$$

and

$$\mathbf{e}(n) = \tilde{\mathbf{e}}^a(n) + \mathbf{v}(n). \tag{37}$$

After inserting the cleared value of $\mathbf{e}(n)$ from (36) into (35), the following expression is obtained:

$$\begin{aligned} &\tilde{\mathbf{w}}(n) + \mathbf{X}(n)(\mathbf{X}^T(n)\mathbf{X}(n))^{-1}\tilde{\mathbf{e}}^a(n) \\ &= \tilde{\mathbf{w}}(n-1) + \mathbf{X}(n)(\mathbf{X}^T(n)\mathbf{X}(n))^{-1}\tilde{\mathbf{e}}^p(n), \end{aligned} \tag{38}$$

and then, the following energy equality [55] should hold:

$$\begin{aligned} &||\tilde{\mathbf{w}}(n)||^2 + (\tilde{\mathbf{e}}^a(n))^T(\mathbf{X}^T(n)\mathbf{X}(n))^{-1}\tilde{\mathbf{e}}^a(n) \\ &= ||\tilde{\mathbf{w}}(n-1)||^2 + (\tilde{\mathbf{e}}^p(n))^T(\mathbf{X}^T(n)\mathbf{X}(n))^{-1}\tilde{\mathbf{e}}^p(n). \end{aligned} \tag{39}$$

Assuming after convergence that $\lim_{n\to\infty}||\tilde{\mathbf{w}}(n)||^2 = \lim_{n\to\infty}||\tilde{\mathbf{w}}(n-1)||^2$, (39) becomes

$$\begin{aligned} &\lim_{n\to\infty}(\tilde{\mathbf{e}}^a(n))^T(\mathbf{X}^T(n)\mathbf{X}(n))^{-1}\tilde{\mathbf{e}}^a(n) \\ &= \lim_{n\to\infty}(\tilde{\mathbf{e}}^p(n))^T(\mathbf{X}^T(n)\mathbf{X}(n))^{-1}\tilde{\mathbf{e}}^p(n) \end{aligned}. \tag{40}$$

Substituting (36) into (40) gives

$$\lim_{n\to\infty}\tilde{\mu}(n) = \lim_{n\to\infty}\frac{2(\tilde{\mathbf{e}}^a(n))^T\mathbf{e}(n)}{\mathbf{e}(n)^T\mathbf{X}^T(n)\mathbf{X}(n)\mathbf{e}(n)}, \tag{41}$$

which, after specifying $\tilde{\mu}(n)$ to the value given by (34) and taking mean values, yields

$$\lim_{n\to\infty}\mathrm{E}\left\{\mathbf{e}^T(n)\mathbf{e}(n)\right\} = \lim_{n\to\infty}\mathrm{E}\left\{2(\tilde{\mathbf{e}}^a(n))^T\mathbf{e}(n)\right\}. \tag{42}$$

For convenience, we express $\mathbf{e}^T(n)\mathbf{e}(n) = \mathrm{Tr}\left\{\mathbf{e}(n)\mathbf{e}^T(n)\right\}$ as in [55], and use (37) in (42) obtaining

$$\begin{aligned} &\lim_{n\to\infty}\mathrm{Tr}\left\{\mathrm{E}\left\{\tilde{\mathbf{e}}^a(n)(\tilde{\mathbf{e}}^a(n))^T\right\}\right\} \\ &+\lim_{n\to\infty}\mathrm{Tr}\left\{\mathrm{E}\left\{\mathbf{v}(n)\mathbf{v}^T(n)\right\}\right\} = \lim_{n\to\infty}\mathrm{E}\left\{2(\tilde{\mathbf{e}}^a(n))^T\tilde{\mathbf{e}}^a(n)\right\}, \end{aligned} \tag{43}$$

where independence between the measurement noise, $v(n)$, and $e^a(n)$ has been considered. It can be assumed during the steady-state that $\mathrm{Tr}\left\{\mathrm{E}\left\{\tilde{\mathbf{e}}^a(n)(\tilde{\mathbf{e}}^a(n))^T\right\}\right\} \approx \mathrm{E}\left\{(\tilde{e}^a(n))^2\right\}$ for large convergence steps, whereas $\mathrm{Tr}\left\{\mathrm{E}\left\{\tilde{\mathbf{e}}^a(n)(\tilde{\mathbf{e}}^a(n))^T\right\}\right\}$ tends to $N\mathrm{E}\left\{(\tilde{e}^a(n))^2\right\}$ for small values of the convergence steps, as it is shown in [55]. Therefore (43) can be expressed as

$$\lim_{n\to\infty}\mathrm{E}\left\{(\tilde{e}^a(n))^2\right\} + N\sigma_v^2 \approx \lim_{n\to\infty}(2N)\mathrm{E}\left\{(\tilde{e}^a(n))^2\right\}, \tag{44}$$

where σ_v^2 is the mean power of the measurement noise.

The expression $\lim_{n\to\infty}\mathrm{E}\left\{(\tilde{e}^a(n))^2\right\}$ denotes the excess mean squared error (EMSE) and, since $e(n) = \tilde{e}^a(n) + v(n)$, its relation with the MSE is given by: $\mathrm{MSE} = \mathrm{EMSE} + \sigma_v^2$. Hence, the MSE can be approximated finally as

$$\mathrm{MSE} \approx \sigma_\mathrm{v}^2 + \sigma_\mathrm{v}^2\frac{\mathrm{N}}{2\mathrm{N}-1} = \sigma_\mathrm{v}^2\frac{3\mathrm{N}-1}{2\mathrm{N}-1}, \tag{45}$$

and $\mathrm{MSE} < 2\sigma_v^2$. This bound arises when only very small values of the convergence steps are considered.

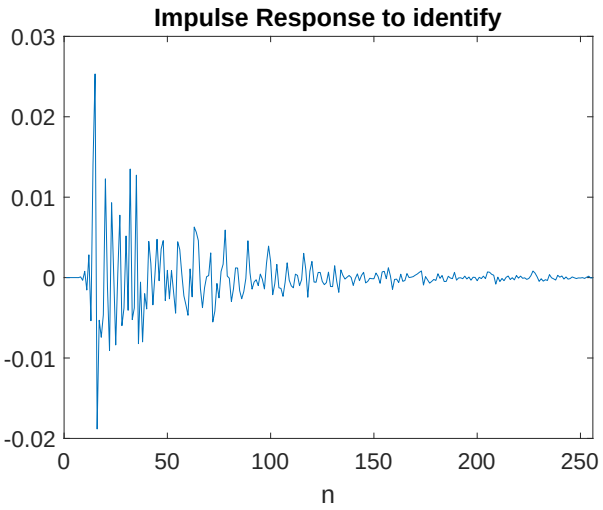

FIGURE 1. Impulse response of the system to be identified. Measured in a $9.36m \times 4.78m \times 2.63m$ listening room.

VII. Results

The proposed algorithm has been employed on a system identification application. The system (shown in Figure VII) has been modelled with a FIR filter of 256 coefficients (although its energy is concentrated within the first 200 coefficients) and an adaptive filter of $L = 250$ coefficients is used for its identification.

The performance of the proposed algorithm has been evaluated against the exact AP method (including projection order $N = 1$, equivalent to NLMS) and APL-I algorithms. Simulations were conducted with projection orders of $N \in \{1,\ 2,\ 4,\ 6,\ 8,\ 10,\ 20\}$. The reference signal, $x(n)$, was white Gaussian noise of unit power, $n(n)$, filtered using the following AR model: $x(n) = n(n) - \gamma x(n-1)$. The values of γ used were: 0 (white signal), 0.9, 0.99, 0.999 (low-pass signals, becoming more low-pass as γ approaches unity), -0.9, -0.99, -0.999 (high-pass signals, becoming more high-pass as γ approaches unity).

The learning curves of the algorithms have been calculated by $L(n) = 10\log_{10}\left[e_f^2(n)/d_f^2(n)\right]$. Power estimate values of the signals have been obtained through exponential windowing given by $e_f^2(n) = \beta e_f^2(n-1) + (1-\beta)e^2(n)$ and $d_f^2(n) = \beta d_f^2(n-1) + (1-\beta)d^2(n)$, where $\beta = 0.999$ has been used. Furthermore, 50 independent trials have been averaged to smooth these learning curves.

Figure 2 displays the results for the different algorithms when the reference signal is slightly colored ($\gamma = 0.9$) with a medium-low projection order ((a) $N = 4$) and a high projection order ((b) $N = 10$). It can be observed that the proposed algorithm performs similarly to the AP, substantially enhancing the convergence speed of the NLMS and, to a lesser extent, that of the APL-I for $N = 4$. Furthermore, the proposed algorithm improves the steady-state behaviour in both cases. This behaviour becomes more evident when the reference signal is highly colored, as depicted in Figure

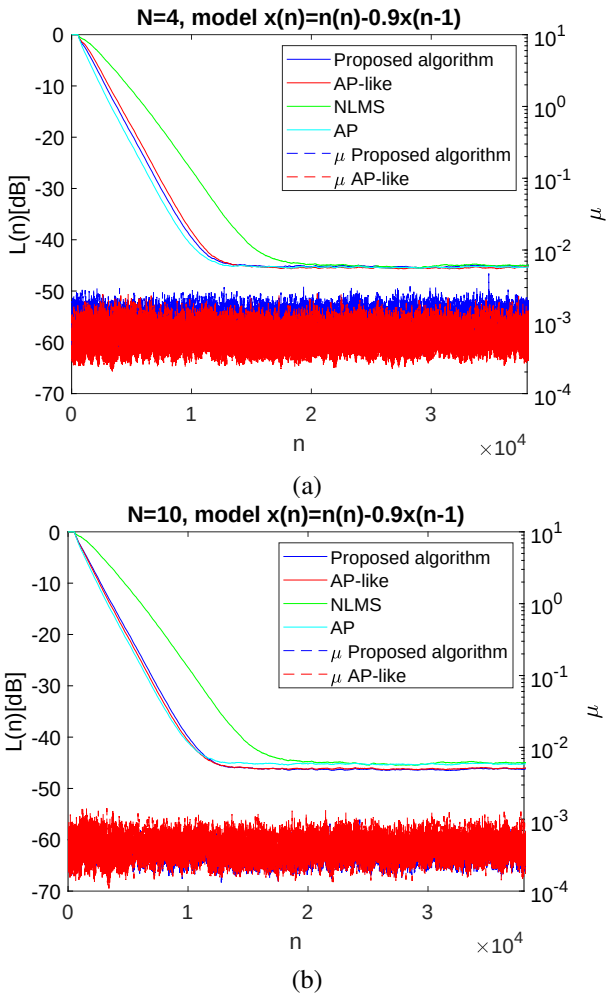

FIGURE 2. **Comparative learning curves and step-size values for: AP, APL-I and the proposed algorithm using low colored reference signals, with different projection orders: (a) $N = 4$ and (b) $N = 10$.**

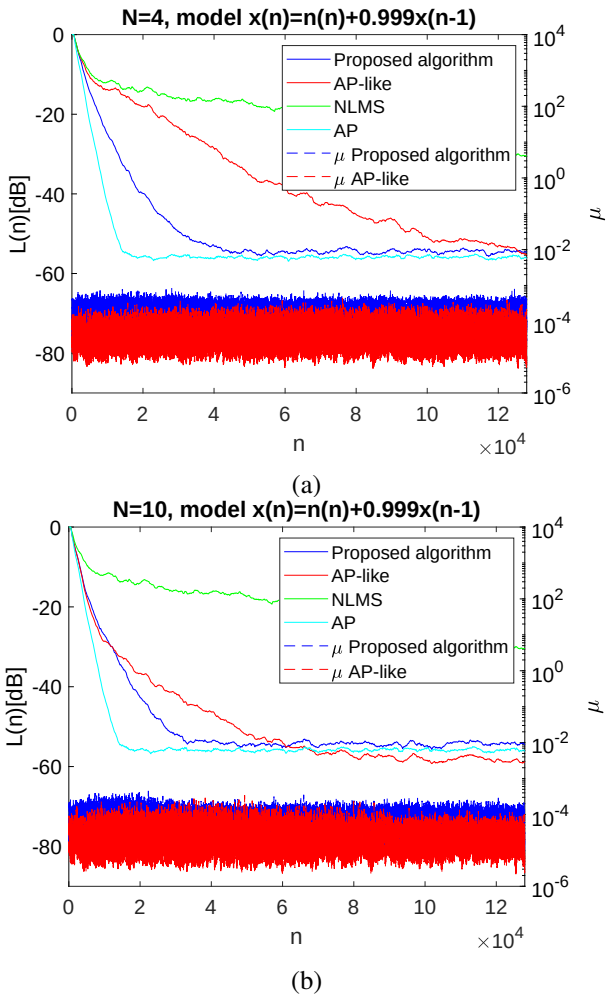

FIGURE 3. **Comparative learning curves and step-size values for the AP, APL-I and the proposed algorithm using high colored reference signals, for: (a) $N = 4$ and (b) $N = 10$.**

3, which presents the same data as in Figure 2 but for a reference signal with $\gamma = -0.999$. Figure 3 illustrates that for $N = 4$, the proposed algorithm outperforms APL-I. However, for $N = 10$, APL-I accelerates its convergence speed, while the proposed algorithm does not (similar to the behavior of AP). This difference arises because the proposed algorithm is more reactive to the a priori error signal for high projection orders and extremely colored signals than APL-I. Nonetheless, the performance of the proposed algorithm remains satisfactory, clearly outperforming the NLMS and APL-I and matching the performance of the AP without requiring high values of N to achieve results close to the best results of the AP.

Figure 4 illustrates the learning curves of the proposed algorithm for various projection orders for (a) a slightly colored signal ($\gamma = -0.9$) and (b) a highly colored signal ($\gamma = 0.999$), thus showing the behaviour of the new approach based on the nature of the reference signal. It is shown that the proposed algorithm behaves similarly to the AP when the reference signal is slightly colored. It improves the convergence speed with the projection order or saturates the convergence speed from a certain order. Thus, in Figure 4-(a) (slightly colored signal), the proposed algorithm achieves its maximum speed for $N = 4$. Subsequently, it exhibits no visible improvement in performance up to $N = 20$, but it does not deteriorate either.

However, the previous sections has shown that, as the signal becomes highly colored, the sensitivity of the algorithm to the a priori error signal becomes more critical, and the convergence speed may deteriorate for high projection orders. This is illustrated in Figure 4-(b) for a highly colored signal, where the proposed algorithm accelerates its convergence up to $N = 8$. Nevertheless, its performance slows down for large N values, as explained in Sections IV and V.

Figures 2 and 3 display the variable step-size for the APL-I and the proposed algorithms. It can be observed that the step-size of the proposed algorithm is slightly higher

This article has been accepted for publication in IEEE Open Journal of Signal Processing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/OJSP.2023.3340106

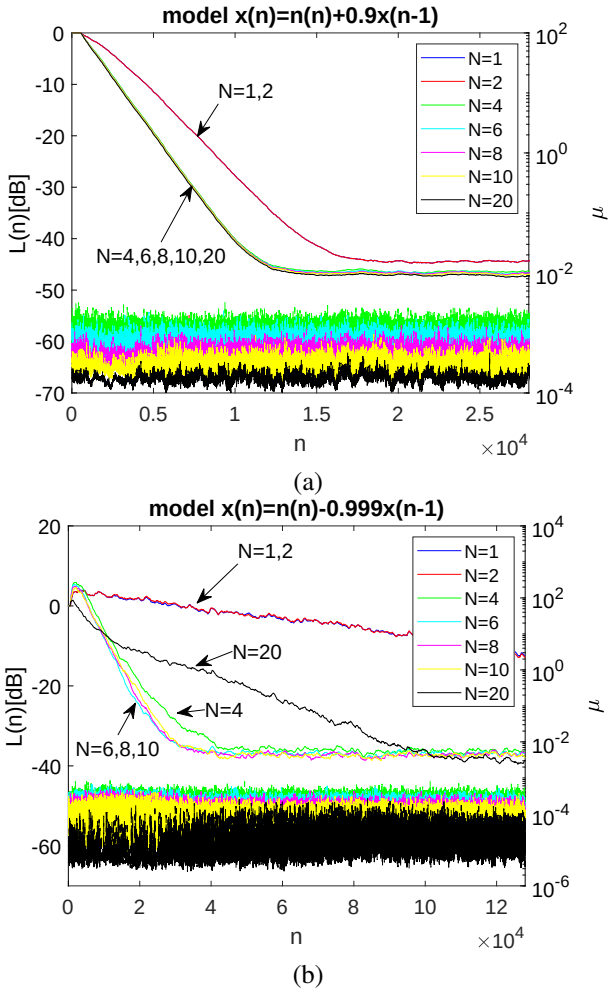

FIGURE 4. **Learning curves and step-size values of the proposed algorithm and different projection orders for: (a) slightly colored reference signal and (b) highly colored reference signal.**

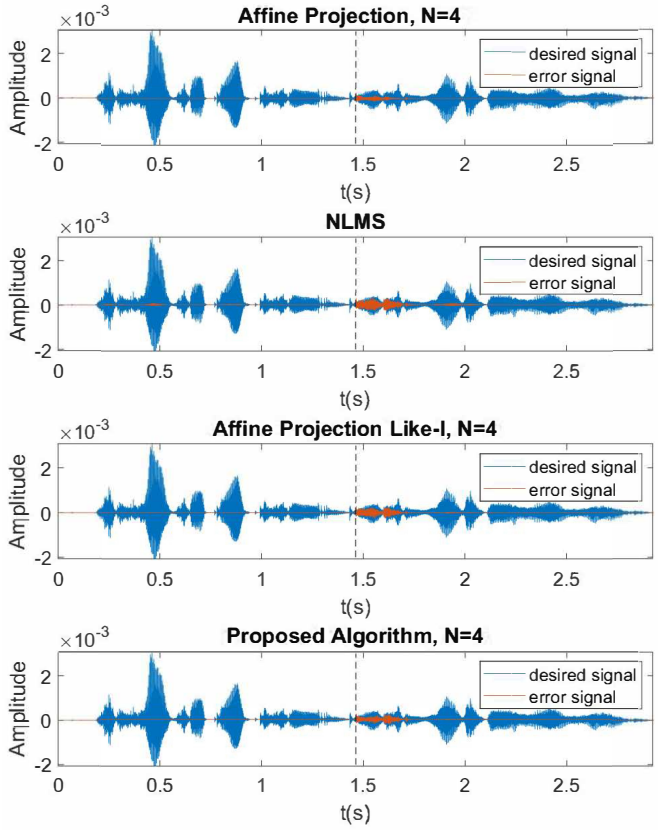

FIGURE 5. **Behaviour of the AP, APL, NLMS and proposed algorithm when the reference signal is a speech signal and the unknown system is suddenly modified at t=1.4s.**

for low projection order and high colored signals, whereas both step-sizes are similar for high projection order and low colored signals. Figure 4 also illustrates the temporal evolution of the variable step-size of the proposed algorithm. As expected, the step-size exhibits a lower range between its highest and lowest values when the projection order increases. This behavior is more evident for low colored signals. It should be noted that the step-size shows different ranges during transient and steady states for high colored signal and high projection order in accordance with the convergence behavior of the algorithm, which differs from its behavior for lower orders.

A speech signal has been used as reference in a last experiment, since one of the most common applications of system identification by adaptive filtering arises in echo cancellers. A sudden change in the acoustic path halfway through the experiment has been brought about to assess the tracking performance of the proposed algorithm. This change consisted of the reversal of the sign of the acoustic path together with an additional delay of two samples. The speech signal sample was taken from the TIMIT database [56], which utters: 'She had your dark suit in greasy wash water all year'. Figure 5 shows the desired signal (the speech signal filtered through the primary path) versus the error signal (the difference between the desired signal and the signal filtered through the adaptive filter, which is the estimated path) for the NLMS, AP, APL, and the proposed algorithm (the last three with $N = 4$). The proposed algorithm exhibits better performance than the NLMS and the APL, and its performance is close to the AP algorithm. All algorithms are robust against the sudden change, but the NLMS transient is slower.

Regarding the steady state performance of the proposed algorithm, the experimental MSE and the approximated MSE according to the model of (45) have been depicted in Figure 6 for different values of the measurement noise variance, σ_v^2, and several projection orders. The results have been obtained using a reference signal based on the AR model with $\gamma = 0.9$. The experimental MSE has been estimated as the average of the squares of the last 100 samples of the error signal. The number of samples for each simulation has been $132,300$ to guarantee reaching the steady state. It can be observed that the model provided by (45) accurately predicts the final MSE.

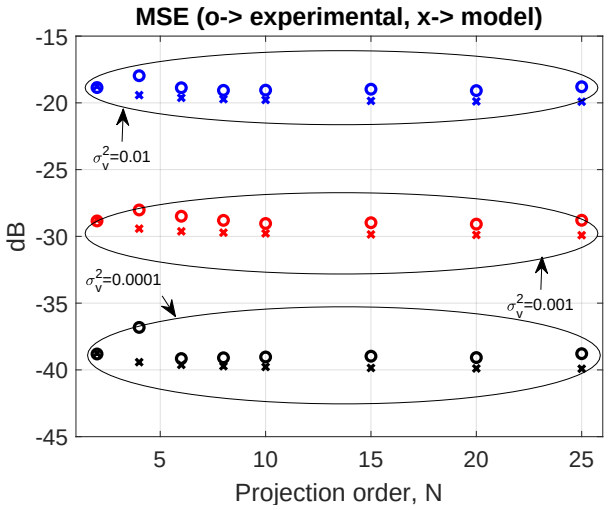

FIGURE 6. Experimental and theoretical MSE versus projection order for the system identification problem.

VIII. Conclusions

In this work, a modification of the LMS algorithm has been proposed that simultaneously uses N vectors of the reference signal and a correction factor that, under certain conditions, allows its behaviour to resemble that of the AP algorithm with projection order N. The main advantage of this algorithm is that it avoids the need for matrix inversion, resulting in a computationally efficient approach with very low correction calculation costs. In addition, it enhances the performance of similar methods like APL-I while maintaining lower computational demands. However, in cases where the reference signal is highly colored, and high projection orders are used, the performance of the proposed algorithm may deviate from that of AP and may even be inferior to APL-I. Therefore, it is advisable, in such cases, to employ regularisation of the variable step size or limit the value of N to that of the best algorithm performance. Nevertheless, it remains a robust algorithm that significantly improves the performance of NLMS with minimal additional computational cost and enhances the performance of APL-I with lower computational requirements.

It should be noted that the proposed algorithm approaches the behavior of the AP in terms of convergence when the reference signal is slightly colored and exhibits excellent performance for colored signals up to the projection orders where the convergence behavior of the AP cannot be improved either. Therefore, the algorithm's performance is significant, and the trade-off between convergence speed and computational cost is, in most cases, much superior to that of other similar algorithms such as NLMS or other APL proposals.

REFERENCES

[1] B. Widrow and M. E. Hoff, "Adaptive switching circuits," in *1960 IRE WESCON Convention Record, Part 4*. New York: IRE, 1960, pp. 96–104.

[2] P. S. R. Diniz, *Adaptive Filtering*, 4th ed. Springer New York, NY, 2012.

[3] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan (Part I: Communications)*, vol. 67, no. 5, pp. 19–27, 1984. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ecja.4400670503

[4] S. L. Gay, "The fast affine projection algorithm," *Acoustic signal processing for telecommunication*, pp. 23–45, 2000.

[5] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, "Fast projection algorithm and its step size control," in *1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 2. IEEE, 1995, pp. 945–948.

[6] F. Albu, M. Bouchard, and Y. Zakharov, "Pseudo-affine projection algorithms for multichannel active noise control," *IEEE transactions on audio, speech, and language processing*, vol. 15, no. 3, pp. 1044–1052, 2007.

[7] M. Bekrani, R. Bibak, and M. Lotfizad, "Improved clipped affine projection adaptive algorithm," *IET Signal Processing*, vol. 13, no. 1, pp. 103–111, 2019.

[8] C. Paleologu, J. Benesty, and S. Ciochina, "A variable step-size affine projection algorithm designed for acoustic echo cancellation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1466–1478, 2008.

[9] S.-E. Kim, S.-J. Kong, and W.-J. Song, "An affine projection algorithm with evolving order," *IEEE Signal Processing Letters*, vol. 16, no. 11, pp. 937–940, 2009.

[10] T. Shao, Y. R. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 327–330, 2010.

[11] Y.-R. Chien, "Variable regularization affine projection sign algorithm in impulsive noisy environment," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 102, no. 5, pp. 725–728, 2019.

[12] J. J. Jeong, "A robust affine projection algorithm against impulsive noise," *IEEE Signal Processing Letters*, vol. 27, pp. 1530–1534, 2020.

[13] H. Zhao, W. Xiang, and X. He, "Bias-compensated affine-projection-like algorithm based on maximum correntropy criterion for robust filtering," *Journal of the Franklin Institute*, vol. 359, no. 3, pp. 1274–1302, 2022.

[14] Y.-R. Chien, C.-H. Yu, and H.-W. Tsao, "Affine-projection-like maximum correntropy criteria algorithm for robust active noise control," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2255–2266, 2022.

[15] R. H. Kwong and E. W. Johnston, "A variable step size lms algorithm," *IEEE Transactions on signal processing*, vol. 40, no. 7, pp. 1633–1642, 1992.

[16] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size nlms algorithm," *IEEE transactions on signal processing*, vol. 56, no. 5, pp. 1878–1893, 2008.

[17] H.-C. Huang and J. Lee, "A new variable step-size nlms algorithm and its performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 2055–2060, 2011.

[18] Z. Mohagheghian Bidgoli and M. Bekrani, "A switching-based variable step-size pnlms adaptive filter for sparse system identification," *Circuits, Systems, and Signal Processing*, pp. 1–25, 2023.

[19] J. Shin, J. Yoo, and P. Park, "Variable step-size affine projection sign algorithm," *Electronics letters*, vol. 48, no. 9, p. 1, 2012.

[20] L. Shi, H. Zhao, Y. Zakharov, B. Chen, and Y. Yang, "Variable step-size widely linear complex-valued affine projection algorithm and performance analysis," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5940–5953, 2020.

[21] H. Zhao, B. Liu, and P. Song, "Variable step-size affine projection maximum correntropy criterion adaptive filter with correntropy induced metric for sparse system identification," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 11, pp. 2782–2786, 2020.

[22] D. Slock, "On the convergence behavior of the lms and the normalized lms algorithms," *IEEE Transactions on Signal Processing*, vol. 41, no. 9, pp. 2811–2825, 1993.

[23] C. Breining, P. Dreiscitel, E. Hansler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp, "Acoustic echo control. an application of very-high-order adaptive filters," *IEEE Signal Processing Magazine*, vol. 16, no. 4, pp. 42–69, 1999.

[24] E. Hänsler and G. Schmidt, *Acoustic echo and noise control: a practical approach*. John Wiley & Sons, 2005.

[25] S. Douglas, "The fast affine projection algorithm for active noise control," in *Conference Record of The Twenty-Ninth Asilomar Conference on Signals, Systems and Computers*, vol. 2, 1995, pp. 1245–1249 vol.2.

[26] M. Bouchard, "Multichannel affine and fast affine projection algorithms for active noise control and acoustic equalization systems," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 1, pp. 54–60, 2003.

[27] M. Ferrer, A. Gonzalez, M. De Diego, and G. Pinero, "Fast affine projection algorithms for filtered-x multichannel active noise control," *IEEE transactions on audio, speech, and language processing*, vol. 16, no. 8, pp. 1396–1408, 2008.

[28] F. Albu, Y. Zakharov, and C. Paleologu, "Modified filtered-x dichotomous coordinate descent recursive affine projection algorithm," in *2009 IEEE International conference on acoustics, speech and signal processing*. IEEE, 2009, pp. 257–260.

[29] R. A. Dobre, C. Paleologu, S. Ciochină, C. Negrescu, and D. Stanomir, "Investigation on the performances of apa in forensic noise reduction," in *2017 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*. IEEE, 2017, pp. 1–6.

[30] M. V. Lima, W. A. Martins, and P. S. Diniz, "Affine projection algorithms for sparse system identification," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 5666–5670.

[31] G. Piñero and P. A. Naylor, "Channel estimation for crosstalk cancellation in wireless acoustic networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 586–590.

[32] V. Yoganathan and T. Moir, "Switched griffiths-jim beamformer using the affine projection algorithm," in *2008 15th International Conference on Mechatronics and Machine Vision in Practice*. IEEE, 2008, pp. 391–396.

[33] Y. W. Jennifer, J. Homer, G. Rombouts, and M. Moonen, "Detection-guided fast affine projection channel estimator for speech applications," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2007, pp. 1–13, 2007.

[34] M. Rotaru, S. Ciochină, and F. Albu, "An efficient gsc vss-apa beamformer with integrated log-energy based vad for noise reduction in speech reinforcement systems," in *International Symposium on Signals, Circuits and Systems ISSCS2013*. IEEE, 2013, pp. 1–4.

[35] L. T. T. Tran, S. Nordholm, H. H. Dam, W.-Y. Yan, and C. Nakagawa, "Acoustic feedback cancellation in hearing aids using two microphones employing variable step size affine projection algorithms," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*. IEEE, 2015, pp. 1191–1195.

[36] L. T. Tran, H. H. Dam, and S. E. Nordholm, "Affine projection algorithm for acoustic feedback cancellation using prediction error method in hearing aids," in *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2016, pp. 1–5.

[37] T. Van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proceedings of the IEEE*, vol. 99, no. 2, pp. 288–327, 2010.

[38] Y.-S. Choi, H.-C. Shin, and W.-J. Song, "Adaptive regularization matrix for affine projection algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 12, pp. 1087–1091, 2007.

[39] A. Carini and G. L. Sicuranza, "Optimal regularization parameter of the multichannel filtered- x affine projection algorithm," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 4882–4895, 2007.

[40] H. Rey, L. R. Vega, S. Tressens, and J. Benesty, "Variable explicit regularization in affine projection algorithm: Robustness issues and optimal choice," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2096–2109, 2007.

[41] V. Myllylä and G. Schmidt, "Pseudo-optimal regularization for affine projection algorithms," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2002, pp. II–1917–II–1920.

[42] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. JHU Press, 2013. [Online]. Available: http://www.cs.cornell.edu/cv/GVL4/golubandvanloan.htm

[43] M. Ferrer, M. de Diego, G. Piñero, and A. Gonzalez, "Affine projection algorithm over acoustic sensor networks for active noise control," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 448–461, 2021.

[44] F. Albu, "Fast block exact gauss-seidel pseudo affine projection algorithm," *Electronics Letters*, vol. 40, pp. 1451–1453(2), October 2004. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/el_20046320

[45] M. C. Tsakiris and P. A. Naylor, "Fast exact affine projection algorithm using displacement structure theory," in *2009 16th International Conference on Digital Signal Processing*, 2009, pp. 1–6.

[46] G.-O. Glentis, "An efficient implementation of the memory improved proportionate affine projection algorithm," *Signal Processing*, vol. 118, pp. 25–35, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168415002042

[47] H. Ding, "Fast affine projection adaptation algorithms with stable and robust symmetric linear system slovers," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1730–1740, 2007.

[48] F. Yang and J. Yang, "A fast affine projection algorithm based on a modified toeplitz matrix," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2017, pp. 1378–1381.

[49] M. Bouchard and F. Albu, "The gauss-seidel fast affine projection algorithm for multichannel active noise control and sound reproduction systems," *International Journal of Adaptive Control and Signal Processing*, vol. 19, no. 2-3, pp. 107–123, 2005. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/acs.846

[50] Y. Zakharov and F. Albu, "Coordinate descent iterations in fast affine projection algorithm," *IEEE Signal Processing Letters*, vol. 12, no. 5, pp. 353–356, 2005.

[51] H. Ding, "A stable fast affine projection adaptation algorithm suitable for low-cost processors," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, vol. 1, 2000, pp. 360–363 vol.1.

[52] M. Z. A. Bhotto and A. Antoniou, "Affine-projection-like adaptive-filtering algorithms using gradient-based step size," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 7, pp. 2048–2056, 2014.

[53] G. Strang, *Linear Algebra and Its Applications*, 4th ed. Brooks/Cole, 2015.

[54] C. L. Chiang, "On the expectation of the reciprocal of a random variable," *The American Statistician*, vol. 20, no. 4, pp. 33–34, 1966.

[55] H.-C. Shin and A. Sayed, "Mean-square performance of a family of affine projection algorithms," *IEEE Transactions on Signal Processing*, vol. 52, no. 1, pp. 90–102, 2004.

[56] (1993) Timit acoustic-phonetic continuous speech corpus. (Last checked on 17 November 2023). [Online]. Available: https://catalog.ldc.upenn.edu/LDC93s1