A Nonparametric VSS NLMS Algorithm

Jacob Benesty, Senior Member, IEEE, Hernán Rey, Leonardo Rey Vega, and Sara Tressens

Abstract—The aim of a variable step size normalized least-mean-square (VSS-NLMS) algorithm is to try to solve the conflicting requirement of fast convergence and low misadjustment of the NLMS algorithm. Numerous VSS-NLMS algorithms can be found in the literature with a common point for most of them: they may not work very reliably since they depend on several parameters that are not simple to tune in practice. The objective of this letter is twofold. First, we explain a simple and elegant way to derive VSS-NLMS-type algorithms. Second, a new nonparametric VSS-NLMS is proposed that is easy to control and gives good performances in the context of acoustic echo cancellation.

Index Terms—Acoustic echo cancellation, adaptive filters, least mean square (LMS), normalized LMS (NLMS), variable step size NLMS.

I. INTRODUCTION

DAPTIVE algorithms play a major role in many signal processing applications [1]–[4]. One of the most popular adaptive filters is the normalized least-mean-square (NLMS) algorithm [1]. Its popularity is due to the fact that it is robust and easy to implement. The stability of this algorithm is governed by a step-size parameter. It is well known that the choice of this parameter, within the stability conditions, reflects a tradeoff between fast convergence and good tracking ability on the one hand and low misadjustment on the other hand. To meet this conflicting requirement, the step size needs to be controlled. While the formulation of this problem is straightforward, a good and reliable solution is not that easy to find. Many different schemes have been proposed in the last two decades [5]-[12]. While some may perform better than others, it is not always obvious to compare them fairly since most of these algorithms require the tuning of many parameters. A fair comparison among several of these algorithms can be found in [12].

The objective of this letter is not to compare the numerous different variable step size NLMS (VSS-NLMS) algorithms that can be found in the literature but first to elaborate on a simple and elegant procedure to derive this kind of algorithms and second to propose a new nonparametric VSS-NLMS algorithm that is very easy to control in real-world applications. Simulations in the context of acoustic echo cancellation show that the proposed algorithm performs much better than NLMS, with fast convergence, good tracking, and low misadjustment.

Manuscript received March 8, 2006; revised March 31, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Vesa Valimaki.

Digital Object Identifier 10.1109/LSP.2006.876323

II. NONPARAMETRIC VSS-NLMS ALGORITHM

In this section, we describe the model, derive the proposed algorithm, and show an important link with the optimal variable regularized NLMS algorithm.

A. Model

In this letter, we consider the following system output signal:

$$y(n) = \mathbf{h}^T \mathbf{x}(n) + v(n) \tag{1}$$

where n is the time index, superscript T denotes transposition,

$$\mathbf{h} = [h_0 \quad h_1 \quad \cdots \quad h_{L-1}]^T$$

is the unknown system of length L that we will try to identify with an adaptive filter,

$$\mathbf{x}(n) = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-L+1) \end{bmatrix}^T$$

is a vector containing the L most recent samples of the system input signal, starting with and including x(n), and v(n) is the system noise that is independent of the input signal x(n). We assume that v(n) is stationary and that all signals are real-valued and zero-mean.

B. Derivation of the Algorithm

We define the *a priori* and *a posteriori* error signals as, respectively $e(n) = y(n) - \hat{\mathbf{h}}^T(n-1)\mathbf{x}(n)$

$$= \mathbf{x}^{T}(n) \left[\mathbf{h} - \hat{\mathbf{h}}(n-1) \right] + v(n)$$
 (2)

$$\varepsilon(n) = y(n) - \hat{\mathbf{h}}^{T}(n)\mathbf{x}(n)$$
$$= \mathbf{x}^{T}(n)\left[\mathbf{h} - \hat{\mathbf{h}}(n)\right] + v(n)$$
(3)

where the filters of length L, $\hat{\mathbf{h}}(n-1)$ and $\hat{\mathbf{h}}(n)$ are estimates of the system \mathbf{h} at time n-1 and n. Consider the linear update equation

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu(n)\mathbf{x}(n)e(n)$$
(4)

where $\mu(n)$ is a positive scalar known as the step size, included to control the changes along the selected direction. One reasonable way to derive a $\mu(n)$ that makes (4) stable is to cancel the *a posteriori* error signal (see [13] and references therein). Replacing (4) in (3) with the requirement $\varepsilon(n)=0$, we easily find, assuming $e(n)\neq 0$, $\forall n$, that $\mu_{\rm NLMS}(n)=[{\bf x}^T(n){\bf x}(n)]^{-1}$. Therefore, the obtained algorithm is the classical NLMS [1], [2]

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \frac{\mathbf{x}(n)}{\mathbf{x}^{T}(n)\mathbf{x}(n)}e(n).$$
 (5)

While the above procedure makes sense in the absence of noise, finding the $\mu(n)$, in the presence of noise, that cancels (3) will introduce noise in $\hat{\mathbf{h}}(n)$ since $\mathbf{x}^T(n)[\mathbf{h}-\hat{\mathbf{h}}(n)] = -v(n) \neq$

J. Benesty is with the Université du Québec, INRS-EMT, Montréal, QC H5A 1K6, Canada (e-mail: benesty@emt.inrs.ca).

H. Rey, L. Rey Vega, and S. Tressens are with the Facultad de Ingeniería, Universidad de Buenos Aires, Buenos Aires, Argentina (e-mail: hrey@fi.uba.ar; lrey@fi.uba.ar; stres@fi.uba.ar).

 $0, \forall n$. What we would like, in fact, is to have $\mathbf{x}^T(n)[\mathbf{h} - \hat{\mathbf{h}}(n)] = 0, \forall n$, which implies that $\varepsilon(n) = v(n)$. Hence, in the new proposed procedure, we wish to find the step-size parameter $\mu(n)$ in such a way that

$$E\left\{\varepsilon^2(n)\right\} = \sigma_v^2, \,\forall n \tag{6}$$

where $E\{\cdot\}$ denotes mathematical expectation, and $\sigma_v^2 = E\{v^2(n)\}$ is the power of the system noise. Using the approximation $\mathbf{x}^T(n)\mathbf{x}(n) = L\sigma_x^2 = LE\{x^2(n)\}$ for $L\gg 1$, where σ_x^2 is the power of the input signal, knowing that $\mu(n)$ is deterministic in nature, substituting (4) into (3), using (2) to eliminate $\hat{\mathbf{h}}(n-1)$, and equating to (6), we find

$$E\left\{\varepsilon^{2}(n)\right\} = \left[1 - \mu(n)L\sigma_{x}^{2}\right]^{2}\sigma_{e}^{2}(n)$$

$$= \sigma_{v}^{2} \tag{7}$$

where $\sigma_e^2(n)=E\{e^2(n)\}$ is the power of the error signal. Developing (7), we obtain a quadratic equation

$$\mu^{2}(n) - \frac{2}{L\sigma_{x}^{2}}\mu(n) + \frac{1}{(L\sigma_{x}^{2})^{2}} \left[1 - \frac{\sigma_{v}^{2}}{\sigma_{e}^{2}(n)} \right] = 0$$
 (8)

for which the obvious solution is

$$\mu_{\text{NPVSS}}(n) = \frac{1}{\mathbf{x}^{T}(n)\mathbf{x}(n)} \left[1 - \frac{\sigma_{v}}{\sigma_{e}(n)} \right]$$
$$= \mu_{\text{NLMS}}(n)\alpha(n) \tag{9}$$

where $\alpha(n)[0 \le \alpha(n) \le 1]$ is the normalized step size. Therefore, the nonparametric VSS-NLMS (NPVSS-NLMS) algorithm is

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu_{\text{NPVSS}}(n)\mathbf{x}(n)e(n)$$
 (10)

where $\mu_{\text{NPVSS}}(n)$ is defined in (9).

We see from (9) that before the algorithm converges, $\sigma_e(n)$ is large compared to σ_v ; thus, $\mu_{\rm NPVSS}(n) \approx \mu_{\rm NLMS}(n)$. On the other hand, when the algorithm starts to converge to the true solution, $\sigma_e(n) \approx \sigma_v$ and $\mu_{\rm NPVSS}(n) \approx 0$. This is exactly what we desire to have: both good convergence and low misadjustment. As we can notice, this approach was derived with almost no assumptions compared to all other algorithms belonging to the same family.

An algorithm, called set membership (SM) NLMS, that looks similar at first glance to (10) was proposed in [14], where $\alpha(n)$ is replaced by

$$\alpha_{\rm SM}(n) = \begin{cases} 1 - \frac{\gamma}{|e(n)|}, & \text{if } |e(n)| > \gamma \\ 0, & \text{otherwise} \end{cases}$$
 (11)

with γ being a bound on the noise. Since there is no averaging on |e(n)|, it is clear that we cannot expect a low misadjustment as for the NPVSS-NLMS. In fact, simulations in [14] show only a slight performance improvement of the SM-NLMS compared to the NLMS.

C. Convergence of the Misalignment

We suppose that the system is perfectly stationary. We define the misalignment vector at time n as $\mathbf{m}(n) = \mathbf{h} - \hat{\mathbf{h}}(n)$.

Rewriting the NPVSS-NLMS algorithm in terms of the misalignment gives

$$\mathbf{m}(n) = \mathbf{m}(n-1) - \mu_{\text{NPVSS}}(n)\mathbf{x}(n)e(n). \tag{12}$$

Taking the l_2 norm and then mathematical expectation of both sides and assuming that $E\{v(n)\mathbf{x}^T(n)\mathbf{m}(n-1)\}=0$ (which is true if the noise signal, v(n), is white), we obtain

$$E\left\{ \|\mathbf{m}(n)\|_{2}^{2} \right\} - E\left\{ \|\mathbf{m}(n-1)\|_{2}^{2} \right\}$$

$$= -\mu_{\text{NPVSS}}(n) \left[\sigma_{e}^{2}(n) + \sigma_{e}(n)\sigma_{v} - 2\sigma_{v}^{2} \right]$$

$$= -\mu_{\text{NPVSS}}(n) \left[\sigma_{e}(n) - \sigma_{v} \right] \left[\sigma_{e}(n) + 2\sigma_{v} \right]$$

$$\leq 0. \tag{13}$$

This shows that for $\mu_{\text{NPVSS}}(n)$ defined in (9), the length of the misalignment vector, $\mathbf{m}(n)$, is nonincreasing, which implies that

$$\lim_{n \to \infty} \sigma_e^2(n) = \sigma_e^2(\infty) = \sigma_v^2. \tag{14}$$

Expression (14) does not imply that $E\{||\mathbf{m}(\infty)||_2^2\} = 0$. However, under the independence assumption, we can show the equivalence. Indeed, from $e(n) = \mathbf{x}^T(n)\mathbf{m}(n-1) + v(n)$, it can be shown that $E\{e^2(n)\} = \sigma_v^2 + \operatorname{tr}[\mathbf{R}\mathbf{K}(n-1)]$ if $\mathbf{x}(n)$ are independent, where $\mathbf{R} = E\{\mathbf{x}(n)\mathbf{x}^T(n)\}$ and $\mathbf{K}(n-1) = E\{\mathbf{m}(n-1)\mathbf{m}^T(n-1)\}$. Since $\sigma_e^2(\infty) = \sigma_v^2$, $\operatorname{tr}[\mathbf{R}\mathbf{K}(\infty)] = 0$, which implies that $\mathbf{K}(\infty) = \mathbf{0}$, assuming that $\mathbf{R} > 0$. As a result, $E\{||\mathbf{m}(\infty)||_2^2\} = 0$.

D. Practical Considerations

In practice, all adaptive algorithms need to be regularized in order to avoid divisions by small numbers. This implies that a positive constant δ needs to be added to the denominator of both step sizes $\mu_{\rm NLMS}(n)$ and $\mu_{\rm NPVSS}(n)$. This is the first important practical consideration.

It is clear that $\sigma_e(n) \geq \sigma_v$, which implies that $\mu_{\rm NPVSS}(n) \geq 0$. In practice, the quantity $\sigma_e^2(n)$ is estimated as follows:

$$\hat{\sigma}_e^2(n) = \lambda \hat{\sigma}_e^2(n-1) + (1-\lambda)e^2(n)$$
 (15)

where λ is an exponential window. This estimation could result in a lower magnitude than σ_v^2 , which would make $\mu_{\rm NPVSS}(n)$ negative. The simple solution to this problem when it occurs is to set $\mu_{\rm NPVSS}(n)=0$.

Table I summarizes the proposed NPVSS-NLMS algorithm. If we take $\lambda = 0$ and $\gamma = \sigma_v$, we find the SM-NLMS algorithm [14].

E. Link With the Optimal Variable Regularized NLMS Algorithm

Let us now consider the following variable regularized NLMS algorithm:

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \frac{\mathbf{x}(n)}{\delta(n) + \mathbf{x}^{T}(n)\mathbf{x}(n)} e(n)$$
 (16)

where $\delta(n)$ is a variable regularization parameter. We can follow the same procedure explained in Section II-B to derive a regu-

TABLE I NPVSS-NLMS ALGORITHM

Initialization:
$$\hat{\mathbf{h}}(0) = \mathbf{0}$$

$$\hat{\sigma}_e^2(0) = 0$$

Parameters:
$$\lambda = 1 - \frac{1}{KL}$$
, exponential window with $K \ge 2$

 σ_n^2 , noise power known or estimated

 $\delta = \text{constant} \cdot \sigma_x^2$, regularization

 $\epsilon > 0$, very small number to avoid division by zero

Error:
$$e(n) = y(n) - \hat{\mathbf{h}}^T(n-1)\mathbf{x}(n)$$

$$\begin{split} \text{Update:} \qquad & \hat{\sigma}_e^2(n) = \lambda \hat{\sigma}_e^2(n-1) + (1-\lambda)e^2(n) \\ \beta(n) &= \left[\delta + \mathbf{x}^T(n)\mathbf{x}(n)\right]^{-1} \left[1 - \frac{\sigma_v}{\epsilon + \hat{\sigma}_e(n)}\right] \\ \\ \mu_{\text{NPVSS}}(n) &= \begin{cases} \beta(n) & \text{if } \hat{\sigma}_e(n) \geq \sigma_v \\ 0 & \text{otherwise} \end{cases} \end{split}$$

$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mu_{\text{NPVSS}}(n)\mathbf{x}(n)e(n)$$

larization factor in such a way that $E\{\varepsilon^2(n)\}=\sigma_v^2, \forall n$. After straightforward calculations, we find

$$\delta_{\rm VR}(n) = \frac{L\sigma_x^2 \sigma_v}{\sigma_e^2(n) - \sigma_v^2} \left[\sigma_v + \sigma_e(n) \right]. \tag{17}$$

We can compare (17) to the optimal regularization parameter derived in [15] and under strong assumptions

$$\delta_{\text{OVR}}(n) = \frac{L\sigma_x^2 \sigma_v^2}{\sigma_e^2(n) - \sigma_v^2}.$$
 (18)

We see that (17) is larger than (18) by at least a factor of two, but the two variable regularization factors have the same effect for good convergence and low misadjustment.

Furthermore, we can easily verify the equality, $[\delta_{\rm VR}(n) + {\bf x}^T(n){\bf x}(n)]^{-1} = \mu_{\rm NPVSS}(n)$, so that the NPVSS-NLMS and variable regularized NLMS algorithms are strictly equivalent. However, we believe that the former algorithm as shown in Table I is more robust and more easy to control in real-world applications.

III. SIMULATIONS IN ACOUSTIC ECHO CANCELLATION CONTEXT

The acoustic coupling between the loudspeaker and the microphone in hands-free telephones generates echoes [16]. To cancel these echoes, we need to identify the impulse response due to this acoustic coupling. In our simulations, we use an acoustic impulse response of length L=512. The same length is used for all the adaptive filters. The sampling rate is 8 kHz. The input signal, x(n), is either a white Gaussian signal, an AR(1) process generated by filtering a white Gaussian noise

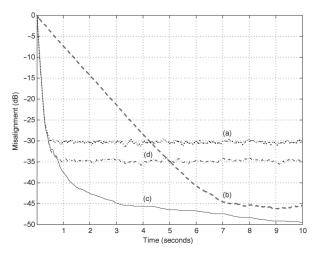


Fig. 1. Misalignment of the NLMS algorithm, with two different step sizes (a) $[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$ and (b) $0.055[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$, and misalignment of the (c) NPVSS-NLMS and (d) SM-NLMS algorithms. The input signal is a white Gaussian noise, L = 512, $\lambda = 1 - 1/(2L)$, and SNR = 30 dB.

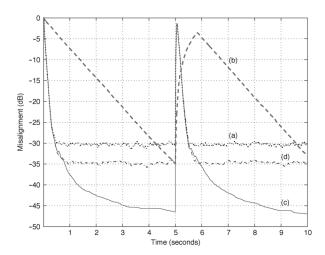


Fig. 2. Misalignment during impulse response change. The impulse response changes at time 5. Other conditions are the same as in Fig. 1. (a) NLMS with step size $[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$. (b) NLMS with step size $0.055[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$. (c) NPVSS-NLMS. (d) SM-NLMS.

through a first-order system $1/(1-0.8z^{-1})$, or a speech signal. An independent white Gaussian noise signal is added to the system output, y(n), with 30-dB signal-to-noise ratio (SNR) for the white and AR(1) input signals and with 15-dB SNR for the speech input signal. We assume that the power of the system noise, σ_v^2 , is known. In practice, it can be easily estimated during silences. The parameter settings chosen for all the simulations are $\hat{h}_l(0)=0,\ l=0,1,\cdots,L-1,\ \hat{\sigma}_e^2(0)=0,\ \delta=20\sigma_x^2,\ \text{and}\ \lambda=1-(1/KL),\ \text{with}\ K=2\ \text{for the white input signal,}\ \text{and}\ K=6\ \text{for the AR}(1)\ \text{and speech input signals.}$ We use the convergence of the normalized misalignment (in dB), $20\log_{10}(||\mathbf{h}-\hat{\mathbf{h}}(n)||_2/||\mathbf{h}||_2)$, for all the algorithms as a measure of performance.

Fig. 1 compares the convergence of the NLMS algorithm, with two different step sizes $[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$ and $0.055[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$, with the convergence of the NPVSS-NLMS and SM-NLMS algorithms. In this simulation, the input signal is a white Gaussian noise. It is clear from this figure that when the convergence rate of the NLMS is the same as the NPVSS-

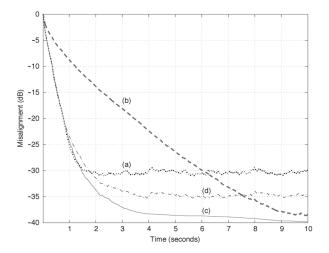


Fig. 3. Misalignment of the NLMS algorithm, with two different step sizes (a) $[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$ and (b) $0.2[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$, and misalignment of the (c) NPVSS-NLMS and (d) SM-NLMS algorithms. The input signal is an AR(1) process, L = 512, $\lambda = 1 - 1/(6L)$, and SNR = 30 dB.

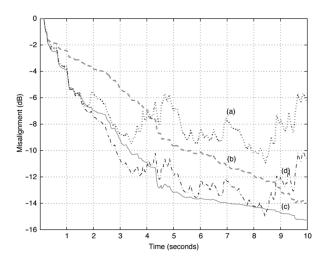


Fig. 4. Misalignment of the NLMS algorithm, with two different step sizes (a) $[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$ and (b) $0.1[\delta + \mathbf{x}^T(n)\mathbf{x}(n)]^{-1}$, and misalignment of the (c) NPVSS-NLMS and (d) SM-NLMS algorithms. The input signal is speech, L=512, $\lambda=1-1/(6L)$, and SNR = 15 dB.

NLMS, the final misalignment is worse by 20 dB. To make the final misalignment of the NLMS closer to that of the NPVSS-NLMS, the step size has to be reduced significantly, and as a consequence, the convergence rate is strongly reduced. We also notice that the NPVSS-NLMS is 15 dB better than the SM-NLMS.

Tracking is a very important issue in adaptive algorithms. In applications like acoustic echo cancellation, it is essential that an adaptive filter tracks fast since impulse responses are not very stationary. Fig. 2 compares the algorithms in a tracking situation when after 5, the acoustic impulse response is shifted to the right by 12 samples. The other conditions of Fig. 2 are the same as that in Fig. 1. According to this simulation, the NPVSS-NLMS tracks as fast as the NLMS with its maximum step size.

Figs. 3 and 4 compare again the NLMS, NPVSS-NLMS, and SM-NLMS algorithms this time with AR(1) and speech signals as inputs. The NPVSS-NLMS still performs much better than the NLMS and SM-NLMS algorithms. For the speech signal, we can notice that the SM-NLMS becomes unstable.

IV. CONCLUSIONS

In the NLMS algorithm, we need to find a compromise between fast convergence and low final misadjustment. However, in many applications, this compromise may not be satisfactory so a VSS-NLMS-type algorithm may be required. In this letter, we have explained a simple method for the derivation of algorithms that are less sensitive to this compromise. From this approach, we have proposed a new nonparametric VSS-NLMS algorithm. Simulations in the context of acoustic echo cancellation have shown that this new adaptive filter outperforms the classical NLMS algorithm.

REFERENCES

- [1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [2] S. Haykin, Adaptive Filter Theory, Fourth ed. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [3] J. Benesty and Y. Huang, Eds., Adaptive Signal Processing—Applications to Real-World Problems. Berlin, Germany: Springer-Verlag, 2003
- [4] Y. Huang, J. Benesty, and J. Chen, Acoustic MIMO Signal Processing. Boston, MA: Springer, 2006.
- [5] R. W. Harris, D. M. Chabries, and F. A. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 309–316, Apr. 1986.
- [6] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1633–1642, Jul. 1992.
- [7] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2075–2087, Jun. 1993.
- [8] J. B. Evans, P. Xue, and B. Liu, "Analysis and implementation of variable step size adaptive algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 8, pp. 2517–2535, Aug. 1993.
- [9] T. Aboulnasr and K. Mayyas, "A robust variable step-size LMS-type algorithm: analysis and simulations," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 631–639, Mar. 1997.
- [10] D. I. Pazaitis and A. G. Constantinides, "A novel kurtosis driven variable step-size adaptive algorithm," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 864–872, Mar. 1999.
- [11] A. Mader, H. Puder, and G. U. Schmidt, "Step-size control for acoustic echo cancellation filters—an overview," *Signal Process.*, vol. 80, pp. 1697–1719, Sep. 2000.
- [12] H.-C. Shin, A. H. Sayed, and W.-J. Song, "Variable step-size NLMS and affine projection algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 132–135, Feb. 2004.
- [13] D. R. Morgan and S. G. Kratzer, "On a class of computationally efficient, rapidly converging, generalized NLMS algorithms," *IEEE Signal Process. Lett.*, vol. 3, no. 8, pp. 245–247, Aug. 1996.
- [14] S. Gollamudi, S. Nagaraj, S. Kapoor, and Y.-F. Huang, "Set-member-ship filtering and a set-membership normalized LMS algorithm with an adaptive step size," *IEEE Signal Process. Lett.*, vol. 5, no. 5, pp. 111–114, May 1998.
- [15] H. Rey, L. Rey Vega, S. Tressens, and J. Benesty, "Optimum variable explicit regularized affine projection algorithm," in *Proc. IEEE ICASSP*, 2006, pp. III-197–III-200.
- [16] J. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, Advances in Network and Acoustic Echo Cancellation. Berlin, Germany: Springer-Verlag, 2001.