A New Robust Variable Step-Size NLMS Algorithm

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

***Abstract*—A new framework for designing robust adaptive filters is introduced. It is based on the optimization of a certain cost function subject to a time-dependent constraint on the norm of the filter update. Particularly, we present a robust variable step-size NLMS algorithm which optimizes the square of the *a posteriori* error. We also show the link between the proposed algorithm and another one derived using a robust statistics approach. In addition, a theoretical model for predicting the transient and steady-state behavior and a proof of almost sure filter convergence are provided. The algorithm is then tested in different environments for system identification and acoustic echo cancelation applications.**

***Index Terms*—Acoustic echo cancelation, adaptive filtering, impulsive noise, normalized least-mean-square (NLMS) algorithm, robust filtering.**

I. Introduction

In real-world adaptive filtering applications, severe impairments may occur. Perturbations such as background and impulsive noise can deteriorate the performance of many adaptive filters under a system identification setup. In echo cancelation, double-talk situations can also be viewed as impulsive noise sources.

Many different approaches have been proposed in the literature to deal with this problem [1]–[8]. Most of them are either directly or indirectly related with the optimization of a combination of L_1 and L_2 norms as the objective function. The former provides a low sensitivity against perturbations and the latter improves the convergence speed of the adaptive filter. In this paper, we introduce a new framework for the construction of robust adaptive filters. Throughout this paper, the term robust will be used as "slightly sensitive to large perturbations (outliers)."

First, we introduce a special case of a previously presented algorithm [1]. Its derivation is based on an optimization using robust statistics [9]. Then, we introduce the new framework that is based on the optimization of a certain cost function subject to a constraint on the norm of the adaptive filter update. Particularly, when the square of the *a posteriori* error is used as the cost function, the result is another algorithm that provides an automatic mechanism for switching between the normalized least-mean-square (NLMS) and normalized sign algorithm (NSA). In principle, the algorithm derived with the new approach and the one with the robust statistics approach are equivalent. However, important differences exist with respect to the assumptions used on each framework.

We also introduce a theoretical model to predict the transient and steady-state behavior of the new algorithm operating in a possibly impulsive noise contaminated environment. Although several (sometimes strong) assumptions are required, the predictions and the simulated results are in good agreement. Also, a proof of almost-sure convergence of the adaptive filter is presented.

However, some practical issues should be considered, particularly in connection with eventual nonstationary environments. Because of this, certain changes are made in the original algorithm. The performance of the algorithm is tested under several scenarios in system identification and acoustic echo cancelation applications.

Part of the results concerning the derivation of the algorithm and its theoretical model have been presented in [10] and [11].

Finally, we present certain definitions and notations that are used in this paper. Let $\mathbf{w}_i = (w_{i,0}, w_{i,1}, \ldots, w_{i,M-1})^T$ be an unknown $M \times 1$ linear finite-impulse response system. The $M \times 1$ input vector at time $i, \mathbf{x}_i = (x_i, x_{i-1}, \ldots, x_{i-M+1})^T$, passes through the system giving an output $y_i = \mathbf{x}_i^T \mathbf{w}_i$. This output is observed, but it is usually corrupted by a noise, v_i, which will be considered additive. In many practical situations, $v_i = b_i + \eta_i$, where b_i stands for the background measurement noise and η_i is an impulsive noise or an undetected near-end signal in echo cancelation applications. Thus, each input $\mathbf{x}_i$ gives an output $d_i = \mathbf{x}_i^T \mathbf{w}_i + v_i$. We want to find $\hat{\mathbf{w}}_i$ to estimate $\mathbf{w}_i$. This adaptive filter receives the same input, leading to an output error $e_i = d_i - \mathbf{x}_i^T \hat{\mathbf{w}}_{i-1}$. We also define the misalignment vector $\tilde{\mathbf{w}}_i = \mathbf{w}_i - \hat{\mathbf{w}}_i$ and the *a posteriori* error signal $e_{\mathrm{p},i} = \mathbf{x}_i^T \tilde{\mathbf{w}}_i + v_i$.

II. A Combined NLMS-Sign Adaptive Filter Using Robust Statistics

Robust statistics have been applied several times in adaptive filter theory [1]–[4]. In this context, the term robust stands for insensitivity to small deviations of the real probability distribution from the assumed model distribution. Usually, a Gaussian distribution is assumed to model the additive noise, but in many situations this assumption proves to be false. This is the case in system identification in an impulsive noise-contaminated environment, and in echo cancelation with double-talk situations. For this reason, a long-tailed probability density function (pdf) is preferred for modeling the noise in those applications.

We introduce here a special case of the algorithm derived in [1]. The main idea is to optimize a nonlinear function of the

Manuscript received March 6, 2007; revised October 30, 2007. This work was supported in part by the Universidad de Buenos Aires under Project UBACYT I005. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. John J. Shynk.

L. Rey Vega is with the Department of Electronics and CONICET, Universidad de Buenos Aires, Paseo Colon 850 (1063), Buenos Aires, Argentina (e-mail: lrey@fi.uba.ar).

H. Rey is with the Instituto de Ingeniería Biomédica (FIUBA) and CONICET, Paseo Colon 850 (1063), Buenos Aires, Argentina (e-mail: hrey@fi.uba.ar).

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

S. Tressens is with the Department of Electronics, Universidad de Buenos Aires, Buenos Aires, Paseo Colon 850 (1063), Argentina (e-mail: stres@fi.uba.ar).

Digital Object Identifier 10.1109/TSP.2007.913142

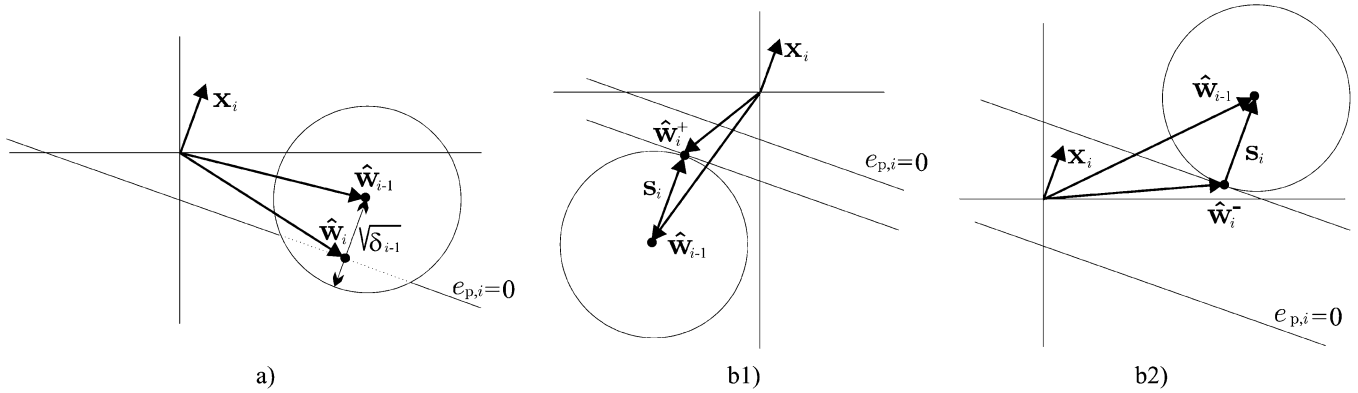

Fig. 1. (a) In this case, an infinite number of solutions exists (represented by the dotted line). Out of those solutions, we choose that which provides the lowest energy in the update. (b1) and (b2) Two updates, $\hat{\mathbf{w}}_i^+$ and $\hat{\mathbf{w}}_i^-$, are possible depending on the relative position between the hyperplane and the hypersphere.

error. This function depends on $p(v)$, the assumed pdf of the disturbing noise. Let us choose the cost function

$$J(\hat{\mathbf{w}}) = \rho\left[\frac{e_i}{||\mathbf{x}_i||}\right], \text{with} \quad \rho[g] = \begin{cases} g^2/2 & |g| \le \sqrt{\delta} \\ \sqrt{\delta}|g| - \delta/2 & |g| > \sqrt{\delta}. \end{cases} \tag{1}$$

This ρ function is associated with the least informative distribution [9]. Although its difference with the Gaussian pdf is small, having a bounded derivative of its logarithm is the key feature that provides robustness.

Although in [1] the ρ function was associated with the pdf of the (stationary) noise, it is also possible to associate it with the pdf of the (nonstationary) error signal, i.e., $p(e_i)$. By doing this, the cutoff $\sqrt{\delta}$ can be time-dependent.

Now, the update of the adaptive filter is found in the opposite direction of the gradient of (1). After computing this gradient with respect to $\hat{\mathbf{w}}_{i-1}$, the resulting algorithm is

$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1} + \frac{\mathbf{x}_i}{||\mathbf{x}_i||} \min\left[\frac{|e_i|}{||\mathbf{x}_i||}, \sqrt{\delta}\right] \operatorname{sign}\left[\frac{e_i}{||\mathbf{x}_i||}\right] \tag{2}$$

or in its alternative form

$$\hat{\mathbf{w}}_i = \begin{cases} \hat{\mathbf{w}}_{i-1} + \frac{\mathbf{x}_i}{||\mathbf{x}_i||^2} e_i & \frac{|e_i|}{||\mathbf{x}_i||} \le \sqrt{\delta} \\ \hat{\mathbf{w}}_{i-1} + \sqrt{\delta}\operatorname{sign}(e_i)\frac{\mathbf{x}_i}{||\mathbf{x}_i||} & \frac{|e_i|}{||\mathbf{x}_i||} > \sqrt{\delta} \end{cases} \tag{3}$$

This algorithm has been previously derived in [1] where in (1), a more general scale factor s was used instead of $||\mathbf{x}_i||$. Here it is very clear how this algorithm operates. When the normalized absolute value of the error is smaller than the pdf cutoff, the adaptive filter behaves as an NLMS with unit step-size. This is reasonable because in this region the pdf is very close to the Gaussian. On the other hand, when the pdf cutoff is exceeded, as the proposed one has longer tails, a large amplitude noise sample might be present. In this case the algorithm uses an NSA with a step-size equal to $\sqrt{\delta}$.

However, when using a fixed δ, there is a tradeoff between convergence speed and noise sensitivity. If δ is too small, the algorithm will behave well against noise, but its convergence might be extremely slow. On the contrary, if δ is large, a large noise sample might degrade the performance of the algorithm. This is a consequence of the transmission of large amplitude noise samples to the adaptive filter through the error signal in the NLMS update.

III. New Framework for Derivation of Robust Adaptive Filters

In this section we propose to solve the problem of robust adaptive filtering using an approach different from that of robust statistics. Suppose an adaptive filter has a given estimate of the true system at a certain time-step. If a large noise sample perturbs it, the result will be a large change in the system estimate, degrading the performance of the adaptive filter. To prevent these situations, the proposed approach is to constrain the energy of the filter update at each iteration. This can be formally stated as

$$||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2 \le \delta_{i-1} \tag{4}$$

where $\{\delta_i\}$ is a positive sequence. Its choice will influence the dynamics of the algorithm. Nevertheless, (4) guarantees that any noise sample can perturb the square norm of the filter update by at most the amount δ_{i-1}.

Next, a cost function is required and the adaptive filter will be the result of optimizing this cost function subject to the constraint (4). Different choices of the cost function and the $\{\delta_i\}$ sequence will lead to different algorithms. Although they will be robust in the sense of satisfying the constraint (4), no general *a priori* prediction can be made about their mean-square performance.

A. Special Case With the a Posteriori Error Cost Function

Now, we propose to find the updating strategy as

$$\hat{\mathbf{w}}_i = \arg\min_{\hat{\mathbf{w}}_i \in \mathbb{R}^M} e_{\mathrm{p},i}^2 \tag{5}$$

subject to the constraint (4).

To perform this optimization, we divide the problem into two cases: (a) the hypersphere defined by (4) has a *nonempty* intersection with the hyperplane defined by $e_{\mathrm{p},i} = 0$. (b) the hypersphere defined by (4) has an *empty* intersection with the hyperplane defined by $e_{\mathrm{p},i} = 0$. These situations are graphically shown in Fig. 1 for the case $M = 2$.

In the first case, there is an infinite number of valid solutions. Among them, we particularly choose the one which provides the lowest energy in the update, i.e., $\min ||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2$. Another benefit of this solution is that the direction of update is the same

as the one of the input vector, so it will be easy to implement. As in this case the *a posteriori* error is zero, the solution would be the popular NLMS algorithm [12], i.e.,

$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1} + e_i \frac{\mathbf{x}_i}{||\mathbf{x}_i||^2}. \tag{6}$$

This will be the solution to our problem (and the NLMS update has to be used) at each time-step when the distance between the point $\hat{\mathbf{w}}_{i-1}$ and the hyperplane defined by $e_{\mathrm{p},i} = 0$ is smaller than $\sqrt{\delta_{i-1}}$. It can be easily shown that this is satisfied when

$$\frac{|e_i|}{||\mathbf{x}_i||} \leq \sqrt{\delta_{i-1}}. \tag{7}$$

In the second case, when the hyphersphere defined by (4) has an empty intersection with the hyperplane defined by $e_{\mathrm{p},i} = 0$, two different possibilities can be considered. As shown in Fig. 1, the hyperplane separates the space $\mathbb{R}^M$ in two halves and the constraint (4) will belong to one of them. As a consequence, the two possible updates are

$$\begin{aligned} &\hat{\mathbf{w}}_i^+ = \hat{\mathbf{w}}_{i-1} + \mathbf{s}_i \quad \text{or} \quad \hat{\mathbf{w}}_i^- = \hat{\mathbf{w}}_{i-1} - \mathbf{s}_i, \\ &\text{where} \quad \mathbf{s}_i = \sqrt{\delta_{i-1}} \frac{\mathbf{x}_i}{||\mathbf{x}_i||}. \end{aligned} \tag{8}$$

The one that leads to the minimum $e_{\mathrm{p},i}^2$ should be used as the update for a certain time-step, as long as (7) is not fulfilled. It can be easily shown that this could be checked with the sign of the estimation error, so that

$$\text{if } e_i > 0 \quad \hat{\mathbf{w}}_i = \hat{\mathbf{w}}_i^+, \quad \text{otherwise} \quad \hat{\mathbf{w}}_i = \hat{\mathbf{w}}_i^-.$$

Thus, with the initial condition $\hat{\mathbf{w}}_0$ and for some $\{\delta_i\}$, the new algorithm can be put in a compact way such as

$$\begin{aligned} e_i &= d_i - \mathbf{x}_i^T \hat{\mathbf{w}}_{i-1}, \\ \hat{\mathbf{w}}_i &= \hat{\mathbf{w}}_{i-1} + \min\left[\frac{|e_i|}{||\mathbf{x}_i||}, \sqrt{\delta_{i-1}}\right] \operatorname{sign}(e_i) \frac{\mathbf{x}_i}{||\mathbf{x}_i||}. \end{aligned} \tag{9}$$

B. Link With the Robust Statistics Approach

As pointed out before, (2) can be also used with a time-dependent δ. It can be seen that the former algorithm would be the same as the one derived in (9) with the proposed approach. However, the two algorithms were derived under very different frameworks. Moreover, the meaning of δ is quite different in both cases. In the first one, it is the cutoff of the noise (or the e_i error) pdf. Thus, the dynamics of the delta sequence will be dependent on the pdf assumed in the model. In the new algorithm, δ is the square of the radius of a certain hypersphere centered at $\hat{\mathbf{w}}_{i-1}$. At every time-step, the new adaptive filter is allowed to evolve only to a new point inside this hypersphere. As a consequence, the dynamics of the delta sequence could be completely arbitrary.

It should be noted that the new framework introduced allows the design of robust adaptive filters without requiring any statistical information of the noise v_i or the error signal e_i. This is probably the most important difference with the robust statistics approach. However, the choice of the delta sequence will influence the mean-square performance of the algorithm. The information of the pdf of the error could be used to find an optimal delta sequence. As this information is not available in practice, we follow in the next subsection a different approach for choosing the delta sequence.

C. Choice of the Delta Sequence

In principle, one would desire $\{\delta_i\}$ to have values as large as possible at the beginning of the adaptation. This will lead to a good initial speed of convergence. Still, it should not be so large that the robust performance against large noise samples is lost. On the other hand, when the algorithm is close to its steady state, lower values of δ_i will lead to a lower final error. This behavior can not be achieved using a fixed parameter δ.

A natural selection should make $\{\delta_i\}$ dependent on the convergence dynamics of the adaptive filter. Thus, we propose

$$\begin{aligned} \delta_i &= \alpha\delta_{i-1} + (1-\alpha)||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2 \\ &= \alpha\delta_{i-1} + (1-\alpha)\min\left[\frac{e_i^2}{||\mathbf{x}_i||^2}, \delta_{i-1}\right] \end{aligned} \tag{10}$$

where $0 < \alpha < 1$ is a memory factor. In Section V we will prove that this delta sequence goes to zero as time goes to infinity. As this property holds no matter the initial condition of the delta sequence, the choice of δ_0 and α will control the tradeoff between robustness and speed of convergence.

The new algorithm has two operation modes: an NLMS with $\mu = 1$ or the NSA with step-size $\sqrt{\delta_{i-1}}$. However, this last update is only used when $\sqrt{\delta_{i-1}} < |e_i|/||\mathbf{x}_i||$. So we can view the update of the new algorithm as another NLMS with step-size

$$\mu' = \min\left[\frac{\sqrt{\delta_{i-1}}}{|e_i|/||\mathbf{x}_i||}, 1\right]. \tag{11}$$

This fact allows us to interpret the new algorithm as a variable step-size NLMS algorithm. Its step-size belongs to $(0, 1]$, but with the property of going towards 0 as $i \to \infty$ (due to the asymptotic behavior of $\{\delta_i\}$ we will prove later). For this reason, we name the new algorithm as *robust variable step-size NLMS* (RVSS-NLMS).

Based on the dynamics of (10), δ_i is decreased only when $\sqrt{\delta_{i-1}} > |e_i|/||\mathbf{x}_i||$. In this case, the used update is the NLMS with $\mu = 1$. So if $\{\delta_i\}$ goes to zero as $i \to \infty$, the above condition should be satisfied an infinite number of times. This fact will be discussed again in Section V and in Appendix G. The reason why the proposed RVSS-NLMS algorithm can still have a robust performance against noise is that these (infinite number of) NLMS updates with $\mu = 1$ take place only at the iterations where the error is small enough. Thus, the performance of the proposed algorithm is not compromised even when the NLMS update is done using $\mu = 1$. This property allows the RVSS-NLMS to outperform the standard NLMS with fixed $\mu = 1$ in both impulsively and nonimpulsively perturbed environments.

We would like to emphasize that the proposed algorithm *switches* at each iteration between the NLMS and NSA solutions. The one chosen depends on the relation between $\sqrt{\delta_{i-1}}$ and $|e_i|/||\mathbf{x}_i||$. If the NLMS is associated with the L_2-norm solution and the NSA to the L_1-norm solution, the RVSS-NLMS algorithm could be seen as a "*switched-norm*" algorithm. This is in contrast with the previously proposed *mixed-norm* algorithms [5]–[8]. The *switched-norm* algorithm

tries to explode the idea of using the NSA solution only at the iterations when the NLMS performance could be compromised. On the other hand, *mixed-norm* algorithms use a weighted sum of the two solutions at each iteration. As a consequence, in certain scenarios with strong perturbations these algorithms might perform just as well as a NSA.

IV. Stochastic Model for the Proposed Algorithm

In the previous section we emphasized that the robust performance of the algorithm is guaranteed by design but no *a priori* statement can be made on its mean-square performance. In order to overcome this issue, we introduce here a theoretical model to predict the transient and steady-state behavior of the new algorithm.

Using (9) and (10), assuming a stationary system, i.e., $\mathbf{w}_i = \mathbf{w}_0\ \forall i$, and noting that the minimum functions in the two updates are logically equivalent, yields

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\operatorname{sign}(e_i)\frac{\mathbf{x}_i}{||\mathbf{x}_i||}. \tag{12}$$

We are interested in the mean-square behavior of $\tilde{\mathbf{w}}_i$. Defining the *a priori* error $e_{\mathrm{a},i} = \mathbf{x}_i^T\tilde{\mathbf{w}}_{i-1}$ and using (12) we can write

$$||\tilde{\mathbf{w}}_i||^2 = ||\tilde{\mathbf{w}}_{i-1}||^2 - 2\sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\frac{\operatorname{sign}(e_i)e_{\mathrm{a},i}}{||\mathbf{x}_i||} + \frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}. \tag{13}$$

We will assume that the noise sequence v_i is i.i.d., zero-mean and it is independent of the input regressors $\mathbf{x}_i$, which belong to a zero-mean stationary process. This is a reasonable and standard assumption. We also assume that the filter length is sufficiently large to assume that $e_{\mathrm{a},i}$ is Gaussian distributed. This assumption can be justified by central limit arguments and it was used and validated with simulations in [13].

The parameter α is typically close to one. This means that (10) is the result of a low-pass filtering of $||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2$. Then the variance of the random variable δ_i would be small enough to assume that

$$E\left[\sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\frac{\operatorname{sign}(e_i)e_{\mathrm{a},i}}{||\mathbf{x}_i||}\right] \approx \sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}}E\left[\frac{\operatorname{sign}(e_i)e_{\mathrm{a},i}}{||\mathbf{x}_i||}\right]. \tag{14}$$

This assumption on the variance of δ_i is very accurate as we will see later in the validation of the model. Taking expectation in (13) and using (14)

$$E[||\tilde{\mathbf{w}}_i||^2] = E[||\tilde{\mathbf{w}}_{i-1}||^2] - 2\sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}} \times E\left[\frac{\operatorname{sign}(e_i)e_{\mathrm{a},i}}{||\mathbf{x}_i||}\right] + \frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}. \tag{15}$$

It should be noted that for sufficiently long filters we can write

$$E\left[\frac{\operatorname{sign}(e_i)e_{\mathrm{a},i}}{||\mathbf{x}_i||}\right] \approx E\left[\frac{1}{||\mathbf{x}_i||}\right]E[\operatorname{sign}(e_i)e_{\mathrm{a},i}]. \tag{16}$$

This is justified by the variance of $1/||\mathbf{x}_i||$ decreasing at least as $1/M$ in many situations of interest (for white Gaussian regressors it actually decreases as $1/M^2$). Thus, the variations in $1/||\mathbf{x}_i||$ are very small for large M, and then we can assume that this quantity is uncorrelated with respect to $\operatorname{sign}(e_i)e_{\mathrm{a},i}$.

Defining $r = E[1/||\mathbf{x}_i||]$, and using (16), (15) can be put as

$$E[||\tilde{\mathbf{w}}_i||^2] = E[||\tilde{\mathbf{w}}_{i-1}||^2] - 2r\sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}} \times E[\operatorname{sign}(e_i)e_{\mathrm{a},i}] + \frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}. \tag{17}$$

Because $E[\delta_i]$ is time-variant we need a recursion for it. Again assuming that the variance of δ_i is small enough we can make the following approximation:

$$E\left[\min\left[\frac{e_i^2}{||\mathbf{x}_i||^2}, \delta_{i-1}\right]\right] \approx E[\delta_{i-1}]P_i[z > E[\delta_{i-1}]] + \int_0^{E[\delta_{i-1}]} z\, dF_i(z) \tag{18}$$

where $z \doteq e_i^2/||\mathbf{x}_i||^2$, i.e., z and $e_i^2/||\mathbf{x}_i||^2$ have the same distribution. $P_i[A]$ and $F_i(z)$ denote the probability of the event A and the distribution function of z at time-step i, respectively. Taking expectation in (10) and using (18) it can be shown that

$$E[\delta_i] = \alpha E[\delta_{i-1}] + (1-\alpha)\left\{E[\delta_{i-1}]P_i[z > E[\delta_{i-1}]] + \int_0^{E[\delta_{i-1}]} z\, dF_i(z)\right\}. \tag{19}$$

After (17) and (19), we can obtain results for the steady-state behavior of the algorithm.

A. Steady-State Behavior

Taking limits in (19) (this is possible because $E[\delta_i]$ is a positive and decreasing sequence) and assuming that $e_i^2/||\mathbf{x}_i||^2$ has a limiting distribution (see the remark below) leads to

$$E[\delta_\infty]P_\infty[z \le E[\delta_\infty]] = \int_0^{E[\delta_\infty]} z\, dF_\infty(z) \tag{20}$$

where $E[\delta_\infty] \equiv \lim_{i\to\infty} E[\delta_i]$, $P_\infty[z \le E[\delta_\infty]] \equiv \lim_{i\to\infty} P_i[z \le \delta_i]$ and $F_\infty(\cdot)$ denotes the limiting distribution of $e_i^2/||\mathbf{x}_i||^2$. In Appendix A we show that the solution to (20) is $E[\delta_\infty] = 0$. In order to obtain results regarding the steady-state behavior of $\sigma_{e_{\mathrm{a},i}}^2 \equiv E[e_{\mathrm{a},i}^2]$ or $E[||\tilde{\mathbf{w}}_i||^2]$ we will need some insights about the asymptotic behavior of $E[\delta_i]$. In Appendix B we show that $\exists i_0, c_1, c_2, n_1, n_2$ such that $\forall i \ge i_0$

$$\frac{c_2}{(n_2+i)^2} \le E[\delta_i] \le \frac{c_1}{(n_1+i)^2}. \tag{21}$$

This means that $E[\delta_i]$ behaves asymptotically as $1/i^2$ and that $\sqrt{E[\delta_i]}$ behaves as $1/i$. Therefore

$$\sum_{i=1}^{\infty} E[\delta_i] < \infty, \sum_{i=1}^{\infty}\sqrt{E[\delta_i]} = \infty. \tag{22}$$

Observing that $E[\|\tilde{\mathbf{w}}_i\|^2] - E[\|\tilde{\mathbf{w}}_{i-1}\|^2]$ constitutes a telescoping series and assuming that $\lim_{i\to\infty} E[\|\tilde{\mathbf{w}}_i\|^2]$ exists, we can write from (17)

$$\lim_{i\to\infty} E[\|\tilde{\mathbf{w}}_i\|^2] - \|\mathbf{w}_0\|^2 = \sum_{i=1}^{\infty}\left\{-2r\sqrt{\frac{E[\delta_i]-\alpha E[\delta_{i-1}]}{1-\alpha}} \cdot E\left[\operatorname{sign}(e_i)e_{\mathrm{a},i}\right] + \frac{E[\delta_i]-\alpha E[\delta_{i-1}]}{1-\alpha}\right\} \quad (23)$$

where we assume that the algorithm is initialized with $\hat{\mathbf{w}}_0 = \mathbf{0}$. However, from (22) it is clear that

$$\sum_{i=1}^{\infty}\frac{E[\delta_i]-\alpha E[\delta_{i-1}]}{1-\alpha} < \infty. \quad (24)$$

This implies that we can split the series of the right-hand side (RHS) of (23) in two. Then we should have

$$\sum_{i=1}^{\infty}\left\{E\left[\operatorname{sign}(e_i)e_{\mathrm{a},i}\right]\sqrt{\frac{E[\delta_i]-\alpha E[\delta_{i-1}]}{1-\alpha}}\right\} < \infty. \quad (25)$$

However, as pointed out in the Appendix B, the following series is divergent

$$\sum_{i=1}^{\infty}\left\{\sqrt{\frac{E[\delta_i]-\alpha E[\delta_{i-1}]}{1-\alpha}}\right\} = \infty. \quad (26)$$

It is not difficult to show that under the Gaussianity assumption of $e_{\mathrm{a},i}$, $E[\operatorname{sign}(e_i)e_{\mathrm{a},i}] \geq 0\ \forall i$ which allows us to conclude

$$\lim_{i\to\infty} E\left[\operatorname{sign}(e_i)e_{\mathrm{a},i}\right] = 0 \quad (27)$$

because if that limit is a number different from zero then (25) would not be valid. Using the Gaussianity of $e_{\mathrm{a},i}$ and the statistical independence with respect to v_i we can write $E[\operatorname{sign}(e_i)e_{\mathrm{a},i}]$ as

$$E[\operatorname{sign}(e_i)e_{\mathrm{a},i}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2_{e_{\mathrm{a},i}}}}e_{\mathrm{a},i}\operatorname{sign}(e_{\mathrm{a},i}+v_i) \times \exp\left[-\frac{e^2_{\mathrm{a},i}}{2\sigma^2_{e_{\mathrm{a},i}}}\right]p_v(v_i)de_{\mathrm{a},i}dv_i \quad (28)$$

where $e_i = e_{\mathrm{a},i} + v_i$. It is not difficult to show

$$E[\operatorname{sign}(e_i)e_{\mathrm{a},i}] = 2\sigma^2_{e_{\mathrm{a},i}}E\left[\frac{1}{\sqrt{2\pi\sigma^2_{e_{\mathrm{a},i}}}}\exp\left[-\frac{v^2}{2\sigma^2_{e_{\mathrm{a},i}}}\right]\right]. \quad (29)$$

where on the RHS the expectation operator is taken with respect the noise pdf. Then defining $\sigma^2_{e_{\mathrm{a},\infty}} \equiv \lim_{i\to\infty}\sigma^2_{e_{\mathrm{a},i}}$, (27) can be written as

$$\sigma^2_{e_{\mathrm{a},\infty}}E\left[\frac{1}{\sqrt{2\pi\sigma^2_{e_{\mathrm{a},\infty}}}}\exp\left[-\frac{v^2}{2\sigma^2_{e_{\mathrm{a},\infty}}}\right]\right] = 0. \quad (30)$$

Equation (30) cannot be satisfied if $\sigma_{e_{\mathrm{a},\infty}} \neq 0$ for any noise pdf because the function inside the expectation operator is strictly positive over the entire real axis. Assuming that the noise pdf $p_v(v)$ is continuous at $v = 0$

$$\lim_{\sigma_{e_{\mathrm{a},\infty}}\to 0}E\left[\frac{1}{\sqrt{2\pi\sigma^2_{e_{\mathrm{a},\infty}}}}\exp\left[-\frac{v^2}{2\sigma^2_{e_{\mathrm{a},\infty}}}\right]\right] = p_v(0). \quad (31)$$

The reason for this is that $1/\sqrt{2\pi\sigma^2_{e_{\mathrm{a},\infty}}}\exp[-v^2/2\sigma^2_{e_{\mathrm{a},\infty}}]$ approaches the Dirac delta function as $\sigma^2_{e_{\mathrm{a},\infty}} \to 0$. Equation (31) together with (30) implies

$$\sigma^2_{e_{\mathrm{a},\infty}} = 0. \quad (32)$$

In Appendix C, under the additional assumption of $\tilde{\mathbf{w}}_{i-1}$ and $\mathbf{x}_i$ being statistically independent, we prove that

$$\lim_{i\to\infty} E[\|\tilde{\mathbf{w}}_i\|^2] = 0. \quad (33)$$

This is a very interesting result which states that under the hypotheses taken, after a sufficiently long time and independently of α and δ_0 the adaptive filter converges to the true system in a mean-square sense.

Remark: Regarding the assumption on the existence of a limiting distribution for $e_i^2/\|\mathbf{x}_i\|^2$, we would expect the noise v_i to dominate significantly over the *a priori* error $e_{\mathrm{a},i}$ when i is large. Then, under the assumption that the noise v_i and the input regressor $\mathbf{x}_i$ are stationary in the strict sense and independent or that $v_i^2/\|\mathbf{x}_i\|^2$ converges to a limiting distribution, the assumption on the existence of a limiting distribution for $e_i^2/\|\mathbf{x}_i\|^2$ is reasonable.

B. Transient Behavior

Equations (17) and (19) can be used to study the transient behavior of the algorithm. However, these equations are not self-contained. We need one more equation linking $\sigma^2_{e_{\mathrm{a},i}}$ and $E[\|\tilde{\mathbf{w}}_{i-1}\|^2]$. For that reason we will work with the classical approach of obtaining a recursion for the covariance matrix of $\tilde{\mathbf{w}}_i$. We define $\mathbf{K}_i = E[\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^T]$. Using (12) and the assumptions on the small variance of δ_i and $1/\|\mathbf{x}_i\|$

$$\mathbf{K}_i = \mathbf{K}_{i-1} - r\sqrt{\frac{E[\delta_i]-\alpha E[\delta_{i-1}]}{1-\alpha}} \times E\left[\operatorname{sign}(e_i)\left(\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T + \mathbf{x}_i\tilde{\mathbf{w}}_{i-1}^T\right)\right] + \frac{E[\delta_i]-\alpha E[\delta_{i-1}]}{1-\alpha}\mathbf{G} \quad (34)$$

where $\mathbf{G} = E[\mathbf{x}_i\mathbf{x}_i^T/\|\mathbf{x}_i\|^2]$. Assuming that the input regressors $\{\mathbf{x}_i\}_{i=0}^{\infty}$ are statistically independent with zero-mean

$$\sigma^2_{e_{\mathrm{a},i}} = \operatorname{Tr}[\mathbf{K}_{i-1}\mathbf{R}] \quad (35)$$

where $E[\mathbf{x}_i\mathbf{x}_i^T] = \mathbf{R}$.

Before proceeding we need to obtain an expression for $E[\operatorname{sign}(e_i)(\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T + \mathbf{x}_i\tilde{\mathbf{w}}_{i-1}^T)]$. A solution for that term can be difficult to obtain in a general situation. For that reason, and with the only purpose of obtaining an useful expression of that

term we will assume that the input regressors are Gaussian. We do this because in order to solve that term we could use Price's theorem [14] which permits to evaluate expectations of nonlinear functions in the Gaussian case. We should note that this is the only place where we take that assumption, and the rest of the analysis is valid even if the input regressors are not Gaussian. We will also assume that the noise sequence can be put as $v_i = b_i + \eta_i$, where b_i is a background noise assumed to be Gaussian with zero-mean and $E[b_i^2] = \sigma_b^2$. η_i is the impulsive part of the noise and can be written as $\eta_i = \omega_i N_i$, where ω_i is Bernoulli with $P(\omega = 1) = p$ and N_i is Gaussian with zero-mean and $E[N_i^2] = K\sigma_b^2, K \gg 1$. Clearly the probability of impulse is p. Then the noise pdf is a mixture of Gaussians

$$p_{v_i}(v_i) = p\mathcal{N}\left(0, (K+1)\sigma_b^2\right) + (1-p)\mathcal{N}\left(0, \sigma_b^2\right). \quad (36)$$

Although impulsive noise is better represented by an α-stable distribution (for acoustic applications the parameter α typically takes the value 1.6 [15]) it is difficult to obtain useful mathematical expressions with them. The previous model for the noise is important because it provides mathematical tractability and was used to approximate situations with impulsive noise [5]–[8], [16]. Using the Gaussian assumption on the input and (36) we proved in Appendix D that

$$\begin{aligned} &E\left[\operatorname{sign}(e_i)\left(\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T + \mathbf{x}_i\tilde{\mathbf{w}}_{i-1}^T\right)\right] \\ &\quad = \sqrt{\frac{2}{\pi}}\left\{\frac{1-p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + \sigma_b^2}} + \frac{p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + (K+1)\sigma_b^2}}\right\} \\ &\qquad \times [\mathbf{K}_{i-1}\mathbf{R} + \mathbf{R}\mathbf{K}_{i-1}]. \end{aligned} \quad (37)$$

The final model is

$$\begin{aligned} \mathbf{K}_i = \mathbf{K}_{i-1} &- r\sqrt{\frac{2}{\pi}}\frac{E[\delta_i - \alpha\delta_{i-1}]}{1-\alpha}[\mathbf{R}\mathbf{K}_{i-1} + \mathbf{K}_{i-1}\mathbf{R}] \\ &\cdot\left\{\frac{p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + (K+1)\sigma_b^2}} + \frac{1-p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + \sigma_b^2}}\right\} \\ &+ \frac{E[\delta_i - \alpha\delta_{i-1}]}{1-\alpha}\mathbf{G} \end{aligned} \quad (38)$$

Then with (19), (35), and (38) we have a self-contained set of equations that gives the transient behavior of the algorithm under the hypotheses taken. To solve them, we require the distribution of $e_i^2/\|\mathbf{x}_i\|^2$ and the values of $\mathbf{G}$ and r. To compute the exact quantities of $\mathbf{G}$ and r, we could take a sample of the input process and calculate the corresponding ensemble averages. In Appendix E the calculation of r is treated when the input is white Gaussian and the calculation of $\mathbf{G}$ when the input is Gaussian but with an arbitrary covariance matrix.

In order to use (19) for obtaining the dynamics of $E[\delta_i]$ we assume that we can take e_i^2 and $\|\mathbf{x}_i\|^2$ as independent variables. This is more accurate as the algorithm is closer to its steady state. In this situation, $e_{\mathrm{a},i}$ can be small compared with the background noise part of v_i. Then, v_i will dominate e_i, and because of the independence between v_i and $\mathbf{x}_i$ the independence assumption between e_i^2 and $\|\mathbf{x}_i\|^2$ will hold approximately. Then defining $t \doteq e_i^2$ and $u \doteq \|\mathbf{x}_i\|^2$ it is not difficult to show that $z \doteq e_i^2/\|\mathbf{x}_i\|^2$ has the pdf

$$p_z(z) = \int_0^\infty u p_t(zu) p_u(u) du. \quad (39)$$

The Gaussianity assumption on $e_{\mathrm{a},i}$ together with (36) allows us to characterize e_i as a mixture of two Gaussian variables with mixing parameters p and $1-p$ and variances $\sigma_{e_{\mathrm{a},i}}^2 + (K+1)\sigma_b^2$ and $\sigma_{e_{\mathrm{a},i}}^2 + \sigma_b^2$, respectively. We want to obtain the pdf of $e_i^2/\|\mathbf{x}_i\|^2$. It is not difficult to show that e_i^2 is a mixture of two nonstandard χ^2 distributions with one degree of freedom and mixing parameters p and $1-p$. In general the expression for the pdf of $\|\mathbf{x}_i\|^2$ could be very difficult to obtain, even in the case when $\mathbf{x}_i$ is Gaussian with arbitrary correlation matrix. The only tractable case is when $\mathbf{x}_i$ is Gaussian with $\mathbf{R} = \sigma_x^2\mathbf{I}$. In that case it can be proved that $p_z(z)$ is a mixture of two nonstandard F distributions. This result was obtained in [11].

However, in order to account for arbitrary correlation matrices and probability density functions for $\mathbf{x}_i$ we assume that we only have the characteristic function of $\|\mathbf{x}_i\|^2$

$$\Psi_u(\omega) = \int_{-\infty}^{\infty} p_u(u) e^{j\omega u} du, \quad p_u(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Psi_u(\omega) e^{-j\omega u} d\omega \quad (40)$$

where $j = \sqrt{-1}$. In many cases it is easier to specify the characteristic function of $\|\mathbf{x}_i\|^2$ instead of its pdf. If the components of $\mathbf{x}_i$ are independent and $\Psi_i(\omega)$ is the characteristic function of $(x_i^n)^2$ (the squared nth component of $\mathbf{x}_i$) it is easy to show that the characteristic function of $\|\mathbf{x}_i\|^2$ is

$$\Psi_u(\omega) = \prod_{n=1}^{M} \Psi_i(\omega). \quad (41)$$

In general, we only need to specify the characteristic function of $\|\tilde{\mathbf{x}}_i\|^2$, with $\tilde{\mathbf{x}}_i = \mathbf{Q}^T\mathbf{x}_i$ where $\mathbf{Q}$ is given by the eigendecomposition $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$. $\tilde{\mathbf{x}}_i$ is a decorrelated version of $\mathbf{x}_i$, and the reason why we can consider $\|\tilde{\mathbf{x}}_i\|^2$ instead of $\|\mathbf{x}_i\|^2$ is that $\mathbf{Q}$ is orthonormal and norm preserving. This could allow simplifying the obtention of the characteristic function at least for the Gaussian case. For example, if $\mathbf{x}_i$ Gaussian with autocorrelation matrix $\mathbf{R}$ the characteristic function of $\|\mathbf{x}_i\|^2$ is difficult to handle. However, for $\|\tilde{\mathbf{x}}_i\|^2$ it is

$$\Psi_u(\omega) = \prod_{n=1}^{M} (1 - 2j\lambda_{nn}\omega)^{-1/2} \quad (42)$$

where λ_{nn} are the eigenvalues of $\mathbf{R}$. Equation (42) is easy to handle for numerical integration. We define

$$g(x, y, z, \omega) = \sqrt{x + 2j\left[(y+1)\sigma_b^2 + z\right]\omega} \quad (43)$$

and the event $C_i = \{z : z > E[\delta_{i-1}]\}$. In Appendix F we show that the terms $P_i[C_i]$ and $\int_0^{E[\delta_{i-1}]} z\, dF_i(z)$ in (19) can be expressed as functions of $\Psi_u(\omega)$ and are given by (44), and

(45), shown at the bottom of the page. In general, these equations have to be integrated numerically and we have to note that although integrals (44) and (45) are well defined their numerical integration has to be done carefully. Then (19), (35), (38), (44), and (45) constitutes the complete model for the transient behavior of the algorithm.

C. Validation of the Model

To test the validity of the model, we compare its predictions with simulated results. We used white and AR1(0.8) input signals. The forgetting factor α is chosen according to the rule of thumb

$$\alpha = 1 - \frac{1}{\kappa M} \tag{46}$$

where κ is a parameter that depends on the color of the input signal and typically ranges in between 1 and 6. If σ_y^2 and σ_b^2 are the power of the uncorrupted output signal and the background noise, respectively, then the signal-to-background-noise ratio (SBNR) is defined as

$$\text{SBNR} = 10\log_{10}\left[\frac{\sigma_y^2}{\sigma_b^2}\right]. \tag{47}$$

In the simulations, σ_b^2 is chosen in such a way that SBNR=10 or 40 dB. The length of the true system is fixed to $M = 32$ and its gain is scaled so that the input and output powers are equal, i.e., $\sigma_x^2 = \sigma_y^2$. The measure of performance considered is the *system mismatch* defined as

$$10\log_{10}\left[\frac{\|\tilde{\mathbf{w}}_i\|^2}{\|\mathbf{w}_i\|^2}\right]. \tag{48}$$

We will also consider the presence of impulsive noise, with $p = 0.1$ and $K = 1000$. It should be noted that good results were also found for p up to 0.5 (not shown) although this extreme condition might be of no practical interest.

In Fig. 2, we can see that the model is in good agreement with the simulated results concerning the system mismatch. However, the transient behavior agreement decreases as the correlation of the input signal increases. The fact that the assumption

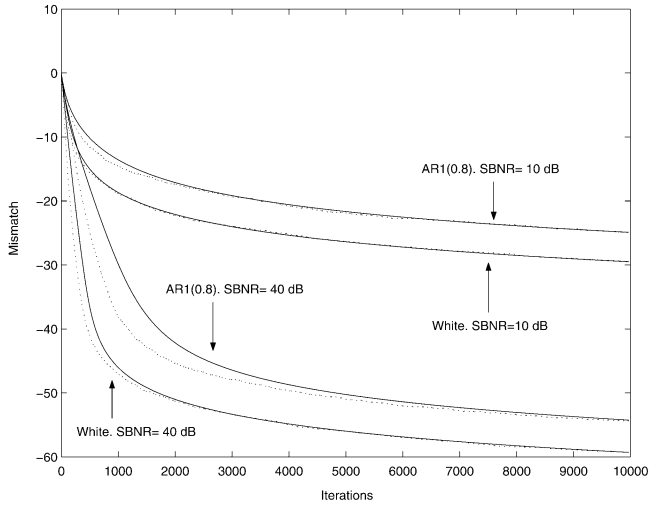

Fig. 2. Mismatch (in dB) under different conditions. Inputs: white and AR1(0.8). SBNR: low (10 dB) and high (40 dB). The theoretical predictions are in solid line while the simulated results are in dotted line. $M = 32$. $\kappa = 1$ for white input and $\kappa = 3$ for AR1(0.8) input. $p = 0.1$ and $K = 1000$. The curves are the result of ensemble averaging over 100 independent runs.

of independent input regressors becomes less accurate, is the main reason for this phenomenon. We see that the system mismatch is monotonically decreasing and, although not shown, its limit is in the order of the machine precision. This confirms the discussion above.

In Fig. 3, a single realization of the simulated sequence $\{\delta_i\}$ is compared to the (ensemble average) prediction of the model. The theoretical results fits very well, indicating that δ_i has a very low variance, and thus confirming the hypothesis taken. Again, its limit is in the order of the machine precision.

Fig. 4 shows the proportion of NLMS updates with $\mu = 1$. According to the model the probability of executing the NLMS update is the one associated with the complement of the event computed in (44). At the beginning of the adaptation, specially when the SBNR is high, the predictions of the model differ from the experimental results. This is due to the hypothesis of independence between e_i^2 and $\|\mathbf{x}_i\|^2$ considered for obtaining a closed-form expression for the distribution of $e_i^2/\|\mathbf{x}_i\|^2$.

$$P_i[C_i] = \frac{1-p}{2\pi j}\int_{-\infty}^{\infty}\Psi_u(\omega)\frac{g\left(E[\delta_{i-1}],0,\sigma^2_{e_{\mathrm{a},i}},\omega\right)-\sqrt{E[\delta_{i-1}]}}{\omega g\left(E[\delta_{i-1}],0,\sigma^2_{e_{\mathrm{a},i}},\omega\right)}d\omega + \frac{p}{2\pi j}\int_{-\infty}^{\infty}\Psi_u(\omega)\frac{g\left(E[\delta_{i-1}],K,\sigma^2_{e_{\mathrm{a},i}},\omega\right)-\sqrt{E[\delta_{i-1}]}}{\omega g\left(E[\delta_{i-1}],K,\sigma^2_{e_{\mathrm{a},i}},\omega\right)}d\omega \tag{44}$$

$$\begin{aligned}\int_0^{E[\delta_{i-1}]} z p_z(z)dz &= (1-p)\frac{\left(\sigma_b^2+\sigma^2_{e_{\mathrm{a},i}}\right)}{\pi}\\ &\times\int_{-\infty}^{\infty}\Psi_u(\omega)\left\{\log\left[\frac{\sqrt{E[\delta_{i-1}]}+g\left(E[\delta_{i-1}],0,\sigma^2_{e_{\mathrm{a},i}},\omega\right)}{g\left(0,0,\sigma^2_{e_{\mathrm{a},i}},\omega\right)}\right]-\frac{\sqrt{E[\delta_{i-1}]}}{g\left(E[\delta_{i-1}],0,\sigma^2_{e_{\mathrm{a},i}},\omega\right)}\right\}d\omega + p\frac{\left((K+1)\sigma_b^2+\sigma^2_{e_{\mathrm{a},i}}\right)}{\pi}\\ &\int_{-\infty}^{\infty}\Psi_u(\omega)\left\{\log\left[\frac{\sqrt{E[\delta_{i-1}]}+g\left(E[\delta_{i-1}],K,\sigma^2_{e_{\mathrm{a},i}},\omega\right)}{g\left(0,K,\sigma^2_{e_{\mathrm{a},i}},\omega\right)}\right]-\frac{\sqrt{E[\delta_{i-1}]}}{g\left(E[\delta_{i-1}],K,\sigma^2_{e_{\mathrm{a},i}},\omega\right)}\right\}d\omega.\end{aligned} \tag{45}$$

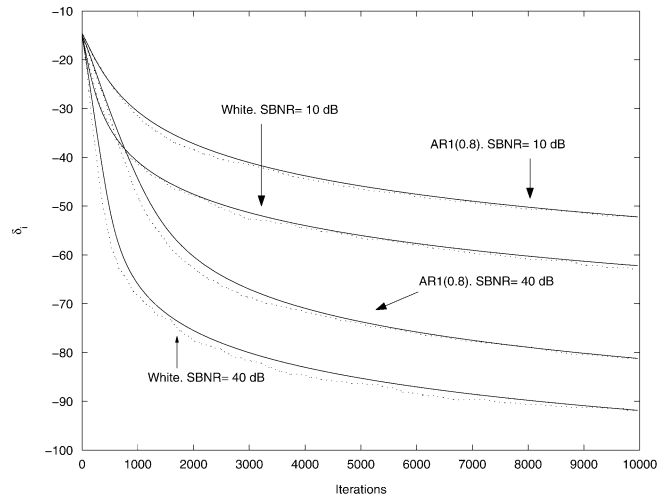

Fig. 3. Evolution of $\{\delta_i\}$ (in dB). The setup is the same as in Fig. 2. Although the theoretical result is an ensemble average, the experimental one corresponds to a single run.

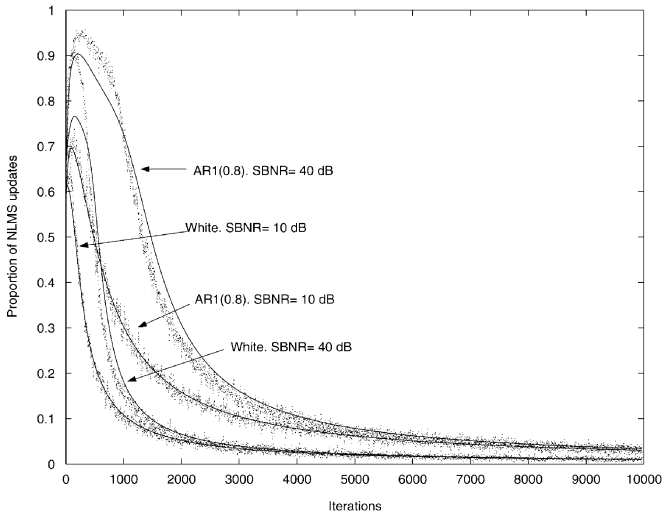

Fig. 4. Evolution of the probability of executing the NLMS update. The setup is the same as in Fig. 2. The experimental plots were low-pass filtered with a moving average with a span of 7 samples.

However, as the adaptation proceeds for a long time interval, the agreement improves. This validates the previous discussion about the independence between e_i^2 and $||\mathbf{x}_i||^2$ for large iteration numbers. It is clear from the figure that, at the beginning of the adaptation, the NLMS update is mostly used (as long as the error is not large, so that the algorithm remains robust) leading to a fast convergence speed. As time progresses, the NSA is used more often (with a decreasing step-size) which allows the algorithm to estimate the true system with negligible variance.

The overall result is that even with very strong (and sometimes unrealistic) assumptions, the predictions of the model are quite accurate, especially for large number of iterations.

V. Almost Sure Convergence of the Proposed Algorithm

From (10), it is easy to see that any realization of the sequence $\{\delta_i\}$ is decreasing. As it is also nonnegative, its limit exists. We denote this limit as δ_∞. In fact, under certain mixing conditions on $e_i^2/||\mathbf{x}_i||^2$ (see Appendix G) it can be proved that

$$\delta_\infty = 0 \; a.e. \tag{49}$$

where a.e. denotes *almost everywhere*. In Appendix G we give a proof of this fact.

This result has an important implication: we can guarantee the almost sure convergence of the adaptive filter. This is stated in the following corollary.

Corollary 1: As $\delta_\infty = 0$, $\hat{\mathbf{w}}_i$ converges almost surely.

Proof: From (10) it can be seen that

$$\lim_{i\to\infty} ||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2 = \delta_\infty = 0 \; a.e. \tag{50}$$

The existence of the limit $\lim_{i\to\infty} \hat{\mathbf{w}}_i$ then follows. □

If we assume that the system is stationary, we can also conclude that the misalignment vector converges. This implies the existence of $\lim_{i\to\infty} E[||\tilde{\mathbf{w}}_i||^2]$, which was assumed in the preceding section, and the mean square stability of the algorithm. It should be also noted that this convergence is guaranteed regardless of the distribution of the noise (it can be impulsive or not) and without requiring the assumptions used in the previous section. However, from this fact we cannot make predictions about its limiting value (although the prediction from the model introduced in the previous section is that it has negligible variance).

VI. Practical Considerations

In this section we discuss certain algorithm implementation issues. First of all, when the NLMS update is used, a regularization constant $\beta = 20\sigma_x^2$ [17] is added to the denominator. Also, every time a division by $||\mathbf{x}_i||$ is computed, a small constant ϵ is added to the denominator to avoid division by zero. These changes still guarantee that the constraint (4) is satisfied. The forgetting factor α is chosen according to (46). A possible initial condition δ_0 can be obtained as

$$\delta_0 = \frac{\sigma_d^2}{\sigma_x^2 M} \tag{51}$$

where σ_d^2 is the power of the observed output.

Finally, a major issue should be considered carefully. As the proposed delta sequence has the decreasing property shown in Section V, although the algorithm becomes more robust against perturbations, it also loses its tracking capacity. For this reason, if there is a chance of being in a nonstationary environment, an *ad hoc* control should be included. The objective is to detect changes in the true system. It should be noted that this control was not included in the analysis performed in Sections IV and V. This means that as the control is less used, the actual performance of the algorithm will be closer to that predicted in Sections IV and V.

Nonstationary Control Method 1

This is a low cost method, suitable for system identification under impulsive noise, where the probability of long impulse

bursts is usually low. Define the positive integers V_{T} and $V_{\mathrm{D}} < V_{\mathrm{T}}$. Then

$$\text{if } \mathrm{mod}(i, V_{\mathrm{T}}) = 0,\ \mathrm{ctrl}_{\mathrm{new}} = \frac{\mathbf{c}^T \mathbf{M} \mathbf{c}}{V_{\mathrm{T}} - V_{\mathrm{D}}} \tag{52}$$

where $\mathrm{mod}(a, b)$ stands for the remainder of the division between integers a and b, $\mathbf{M} = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$ is a diagonal matrix with its first $V_{\mathrm{T}} - V_{\mathrm{D}}$ elements set to one, $\mathbf{c}^T = \mathcal{O}([|e_i|/(\epsilon + \|\mathbf{x}_i\|), \ldots, |e_{i-V_{\mathrm{T}}+1}|/(\epsilon + \|\mathbf{x}_{i-V_{\mathrm{T}}+1}\|)])$ and $\mathcal{O}(\cdot)$ is the ascending order operator. Now, with $\mathrm{ctrl}_{\mathrm{old}}$ being the value of the estimator during the previous V_{T} samples and ξ being some threshold, the following rule is applied every V_{T} iterations:

$$\Delta_i = (\mathrm{ctrl}_{\mathrm{new}} - \mathrm{ctrl}_{\mathrm{old}})/\delta_{i-1}. \tag{53}$$

if $\Delta_i > \xi$

$$\delta_i = \delta_0$$

elseif $\mathrm{ctrl}_{\mathrm{new}} > \mathrm{ctrl}_{\mathrm{old}}$

$$\delta_i = \delta_{i-1} + (\mathrm{ctrl}_{\mathrm{new}} - \mathrm{ctrl}_{\mathrm{old}})$$

else

$$\delta_i = \alpha \delta_{i-1} + (1 - \alpha) \|\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}\|^2$$

end

$$\mathrm{ctrl}_{\mathrm{old}} = \mathrm{ctrl}_{\mathrm{new}}$$

The sorting of V_T numbers has a complexity of $V_T \log V_T$. However, this sorting is performed only every V_T iterations. Moreover, at each iteration where the quantity $e_i^2/(\epsilon + \|\mathbf{x}_i\|^2)$ must be stored, it could be done efficiently in order to provide a better situation when the final sorting is required.

The larger V_T, the smaller the steady-state value but with the cost of increasing complexity and delaying the tracking ability. When the noise level is large, $V_T - V_D$ should be increased in order to discard the large noise samples that will otherwise contribute as outliers for the estimation of a change in the system. The parameter ξ is a threshold to detect a major change in the system that cannot be followed by a small δ_{i-1}, and then perform a reinitialization of the delta sequence in order to track it rapidly. If the change is not too large, the second line of the "if" sentence can solve the tracking problem without restarting the delta sequence. The sentence after the "else" is the delta update presented in the previous sections, which is the only one that should be used in stationary systems.

Nonstationary Control Method 2

In echo cancelation, the double-talk detector may miss several impulses with the result that long impulse bursts can corrupt the output of the system. With the first control method, large values of V_{T} and V_{D} would be required, which introduces undesired delays. For this reason we propose another method for this type of application. The only difference is the way of computing the estimator ctrl. So (52) should be replaced by the following procedure. At each time-step

if (double-talk is not declared)

$$r_i = e_i/(\epsilon + \|\mathbf{x}_i\|)$$

else

$$r_i = 0$$

end

$$\theta_i = \tau \theta_{i-1} + C_1 (1 - \tau) k_i^2$$

where τ is a forgetting factor, C_1 is a positive gain, $k_i = \mathrm{median}[r_i, \ldots, r_{i-V_\theta+1}]$, and V_θ is the length of the sliding window. Then, the new estimator is computed as

$$\text{if} \quad \mathrm{mod}(i, V_{\mathrm{T}}) = 0, \quad \mathrm{ctrl}_{\mathrm{new}} = \mathrm{mean}(\theta_i, \ldots, \theta_{i-V_{\mathrm{T}}+1}). \tag{54}$$

Finally, the same comparison between $\mathrm{ctrl}_{\mathrm{new}}$ and $\mathrm{ctrl}_{\mathrm{old}}$ is performed, as explained in the last part of the first method.

In this method, long impulse bursts should be discarded in order to have a "clean" estimation of the system changes. τ is a memory factor that should be small, so that it can quickly track the nonstationarities of the perturbations. C_1 is a gain that weights the changes in the normalized error sequence. V_θ and V_T are, respectively, the length of the window where the noise bursts and changes in the system are searched.

It should be noted that these proposed controls are *ad hoc*. Other schemes might be used. The advantage of the proposed schemes is that the parameters are not coupled to each other as in other previous proposed algorithms. Each parameter is used to deal with an specific feature of the environment. Therefore, this set of parameters allows the algorithm to work well under many different scenarios. Although certain values for the parameters are used in the next section, we swept the parameter space (in some cases up to one order of magnitude) and the changes in performance were not significant.

VII. Simulation Results

The system is taken from a measured acoustic impulse response and it was truncated to $M = 512$. Its gain is scaled so that the input and output powers are equal, i.e., $\sigma_x^2 = \sigma_y^2$. The adaptive filter length is set to M in each case. We use the *mismatch*, as defined in (48), as a measure of performance. The plots are the result of single realizations of all the algorithms without any additional smoothing. A zero-mean Gaussian white noise b_i is added to the system output such that a certain SBNR, that was defined in (47), is achieved.

The behavior of the proposed algorithm is compared with other strategies. We simulate an NLMS algorithm with $\mu_1 = 1$ and a regularization factor $\beta = 20\sigma_x^2$. The proposed algorithm does not need a regularization factor (if the norm of the input vector is too small, it will perform the NSA update). However, it was included in the simulations to keep the comparison with the NLMS algorithm fair. The second scheme used is the NSA

$$\hat{\mathbf{w}}_i^{\mathrm{sign}} = \hat{\mathbf{w}}_{i-1}^{\mathrm{sign}} + \mu_2 \,\mathrm{sign}(e_i) \frac{\mathbf{x}_i}{\|\mathbf{x}_i\| + \epsilon} \tag{55}$$

with the value of μ_2 chosen to give the same steady-state mismatch as the one in the proposed RVSS-NLMS.

We also tried different schemes proposed in the literature [3]–[8]. Unfortunately, their performance under the experimental setup chosen here was just as good as the one of the NSA, and thus, we decided not to show them. A possible reason for this was discussed at the end of Section III.C. Another possibility is to include the algorithm derived in Section II with a fixed time-independent δ. However, it should be clear that this will lead to a slow robust performance or a fast nonrobust performance (this was already discussed at the end of Section II). Finally, we also tested the median LMS algorithm. Even though its performance is better than the NSA, it is more complex (the median operator is used in every iteration) and with worse performance than that of the proposed algorithm. Actually, in [3], the robust Huber filter was shown to outperform the median LMS. The RVSS-NLMS is indirectly related to the Huber distribution as shown in Sections II and III. Because of these we decided to compare the proposed algorithm only with the NLMS and NSA described earlier.

The value of α is set using $\kappa = 3$ in (46). The initial value for the delta sequence is chosen according to (51). We also want to test the nonstationary controls discussed in the previous section. As a measure of their performance, we compute for each simulation

$$\mathcal{M} = \max_{i:\mathrm{mod}(i,V_T)=0} \Delta_i \quad \text{and} \quad \mathcal{R} = \frac{\mathcal{M}}{\mathcal{N}} \tag{56}$$

where $\mathcal{N}$ is the second largest value of Δ_i. In every simulation a sudden change is introduced at a certain time-step by multiplying the system coefficients by -1. In all the cases, $\mathcal{M}$ is accomplished when the sudden change is introduced, while $\mathcal{N}$ is accomplished at any other time-step. The value of $\mathcal{M}$ is related to the threshold ξ while that of $\mathcal{R}$ gives an idea of the reliability of detection of a sudden change.

A. System Identification Under Impulsive Noise

The input process is an AR1 with pole in 0.8. The nonstationary control 1 is used in this application. In addition to the background noise b_i, an impulsive noise η_i could also be added to the output signal y_i. The impulsive noise is generated as $\eta_i = \omega_i N_i$, where ω_i is a Bernoulli process with probability of success $P[\omega_i = 1] = p$ and N_i is a zero-mean Gaussian with power $\sigma_N^2 = 1000\sigma_y^2$.

First we study the case of high SBNR (40 dB). In Fig. 5, no impulse noise is present. As it can be seen, the NLMS has a good initial convergence while the NSA has a dramatically slow performance. However, the RVSS-NLMS can extract the good properties of each algorithm. It shows the same speed of convergence of the NLMS but with a 10 dB lower steady-state error. This is possible because when δ_i is small enough, the algorithm employs the NLMS update with $\mu' < 1$, which allows it to reach a lower steady-state level.

Then we include the impulse noise with $p = 0.1$. In Fig. 6, the NLMS with $\mu = 1$ has a positive mismatch (in dB), while the performance of the other two algorithm remains barely unchanged. The interesting thing is that the NLMS update with

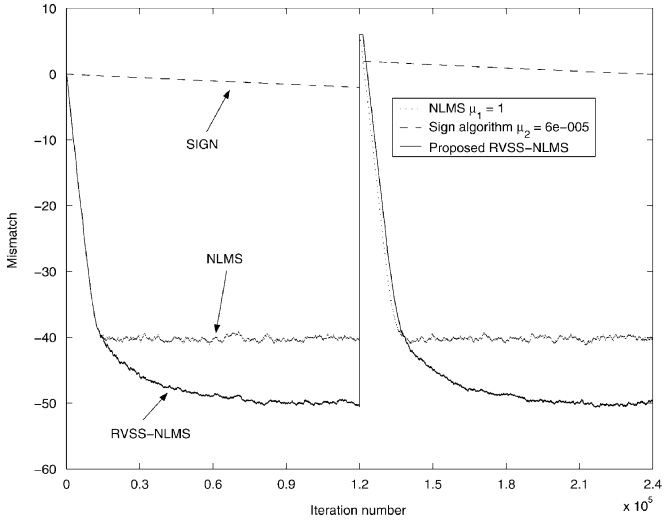

Fig. 5. Mismatch (in dB). AR1(0.8) input. No impulsive noise. $\mathrm{SBNR} = 40$ dB. $\xi = 25$. $V_{\mathrm{T}} = 3M$. $V_{\mathrm{D}} = 0.75V_{\mathrm{T}}$. The curves are the result of a single realization of the algorithms. $\mathcal{M} = 26972$. $\mathcal{R} = 41496$.

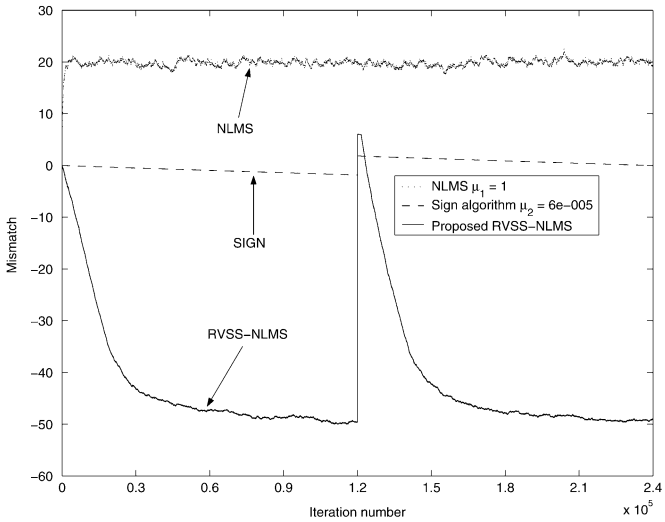

Fig. 6. Mismatch (in dB). AR1(0.8) input. $p = 0.1$. The other parameters are the same as in Fig. 5. $\mathcal{M} = 29005$. $\mathcal{R} = 70745$.

$\mu = 1$ is performed by the proposed algorithm only at the iterations when the error is small enough. On the other hand, as it can be seen from (11), the larger the error (probably due to the appearance of a large noise sample), the smaller the value of μ' used in the update. This allows an NLMS algorithm to perform robustly against impulsive noise.

Now we repeat the experiments but with a low SBNR (10 dB). In Figs. 7 and 8 it can be seen that the qualitative results are similar to the case when the SBNR is high. Here, as the background noise power is larger, so is that of the estimation error. As a consequence, μ' decreases and the algorithm shows less sensitivity to the noise.

The proposed algorithm performed well even in tests with $p = 0.5$ (not shown here), although this might not be a reasonable condition in a system identification application.

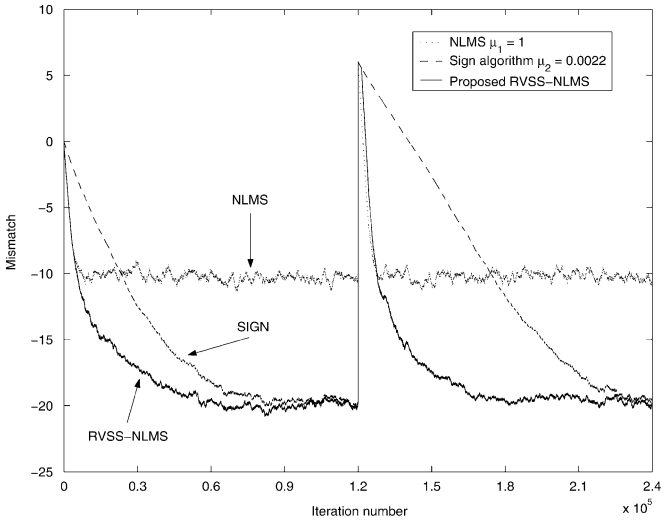

Fig. 7. Mismatch (in dB). AR1(0.8) input. $\mathrm{SBNR} = 10$ dB. No impulsive noise. The other parameters are the same as in Fig. 5. $\mathcal{M} = 32$. $\mathcal{R} = 63$.

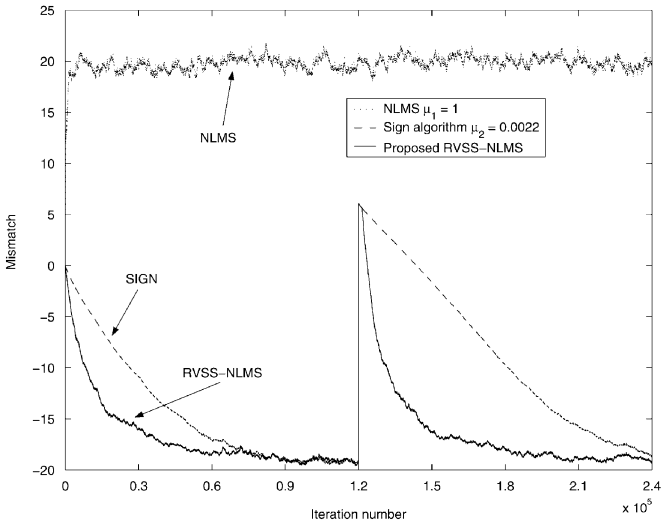

Fig. 8. Mismatch (in dB). AR1(0.8) input. $p = 0.1$. The other parameters are the same as in Fig. 7. $\mathcal{M} = 48$. $\mathcal{R} = 60$.

In the four cases studied so far, the performance of the nonstationary control can also be analyzed. As shown in the figures, the RVSS-NLMS can recover from a sudden change with nearly the same speed as the NLMS with $\mu = 1$. On the other hand, the NSA has a poor tracking performance. The values of $\mathcal{M}$ and $\mathcal{R}$ appear on the caption of each figure. Both of them increase with the SBNR but in every case the sudden change can be reliably identified due to the large values of $\mathcal{R}$.

B. Acoustic Echo Cancelation With Double-Talk Situations

In echo cancelation applications, a double-talk detector (DTD) is used to suppress adaptation during periods of simultaneous far- and near-end signals. A simple and efficient way of detecting double-talk is to compare the magnitude of the input and output signals and declare double-talk if the output magnitude is larger than a value set by the input signal. A proven algorithm that has been in commercial use for many years is the Geigel DTD [25]. Although newer algorithms are available today, presenting a lower probability of miss, we believe that having good results with the simple Geigel DTD will also imply good performance with better DTDs.

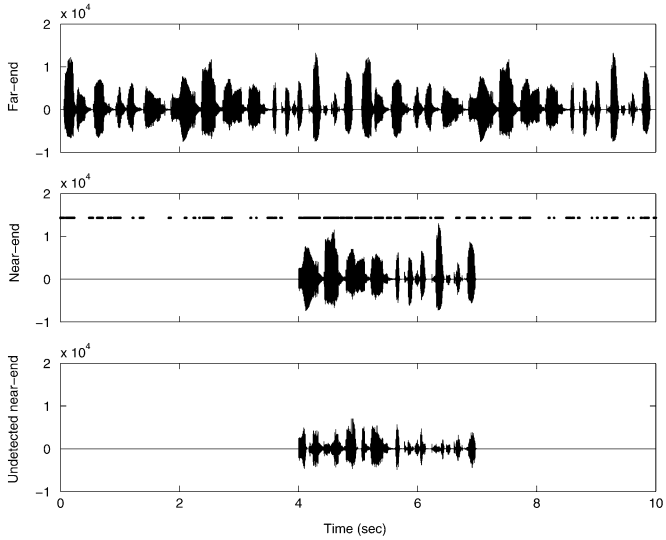

Fig. 9. Acoustic echo cancelation. *Top*. Far-end signal. It was used twice for a total of 20 s. *Middle*. After 4 s of adaptation, the near-end signal appears for a period of 3 s. In the upper part, each dot stands for a double-talk declaration from the DTD. The proportion of detections during the double-talk situation was 22% while the proportion of false alarms when no double-talk was present was 3%. *Bottom*. The undetected near-end signal becomes the impulsive noise for the echo path identification. After passing through the DTD, the power of the near-end signal was reduced about 2.5 times.

The Geigel DTD declares double-talk if

$$\frac{\max(|x_i|, |x_{i-1}|, \ldots, |x_{i-D+1}|)}{|d_i|} < T, \tag{57}$$

where x_i are the samples of the far-end signal and d_i are the samples of the far-end signal filtered by the acoustic impulse response and possibly contaminated by a background noise and a near-end signal. Although the adaptation is usually inhibited for a certain period of time when double-talk is declared, we choose to stop it just for the current time-step. If the algorithm is robust enough to deal with the undetected near-end signal, this decision will avoid an unnecessary decrease in the speed of convergence.

In Fig. 9, the far-end, near-end, and undetected near-end signals are shown. The far-end and near-end signals are speech sampled at 8 kHz. The SBNR is 20 dB while the *signal-to-total-noise ratio* (STNR), i.e.,

$$\mathrm{STNR} = 10 \log_{10} \left[\frac{\sigma_y^2}{\sigma_b^2 + \sigma_\eta^2} \right]$$

is set to 0 dB, where σ_η^2 is the power of the near-end signal before passing through the DTD. The parameters of the DTD are $D = M$ and $T = 1.25$. Under these conditions, the DTD detected only 22% of the near-end signal which causes long bursts of impulsive noise. Increasing the value of T would also lead to an increase of the false alarms and the convergence speed will be undesirably slowed down. For these reasons, the nonstationary control 2 is used for the proposed algorithm.

As shown in Fig. 10, the NLMS with $\mu = 1$ is highly affected by the double-talk situation. This is not the case for the other two algorithms but the RVSS-NLMS performs faster than the NSA. When the sudden change occurs, the RVSS-NLMS can track

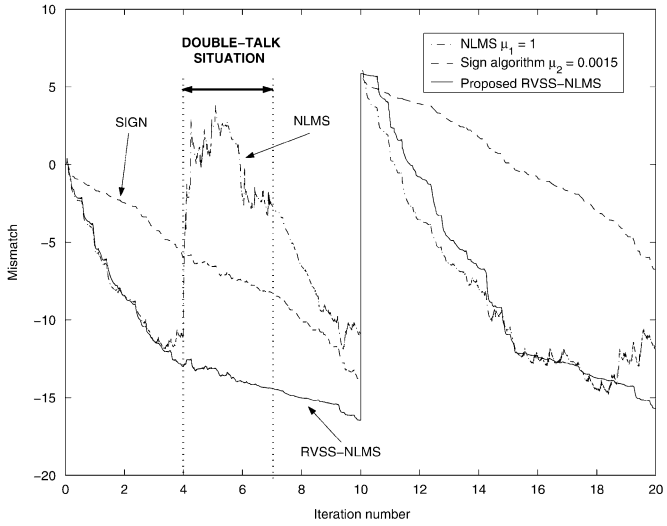

Fig. 10. Mismatch (in dB) for speech input. $\mathrm{SBNR} = 20$ dB. $\mathrm{STNR} = 0$ dB. $\xi = 25$. $T = 1.25$. $C_1 = 3$. $V_{\mathrm{T}} = 2\,M$. $V_\theta = 5\,M$. $\tau = 0.95$. $\mathcal{M} = 105$. $\mathcal{R} = 188$.

the system with a small delay with respect to the NLMS while the NSA shows a very poor performance. The large values of $\mathcal{M}$ and $\mathcal{R}$ indicate again that the sudden change can be reliably detected.

VIII. Conclusion

A new framework for designing robust adaptive filters was introduced. It is based on the optimization of a certain cost function subject to a time-dependent constraint ($\{\delta_i\}$) on the norm of the filter update. Particularly, we derived the RVSS-NLMS algorithm by optimizing the square of the *a posteriori* error. We showed that the new algorithm is equivalent to another one derived using a robust statistics approach. However, the new framework requires no statistical information about the probability density functions of the noise nor the error signal. Then, we proposed a certain dynamics for $\{\delta_i\}$ which provides the algorithm with fast initial convergence as the NLMS with $\mu = 1$ but also a performance against noise as robust as the one of the NSA.

By assuming certain mixing conditions on the normalized square error, we proved the almost sure convergence of the adaptive filter (regardless whether the noise is impulsive or not). In addition, we introduced a theoretical model for the RVSS-NLMS that under certain reasonable assumptions allowed us to prove that the limiting mean-square misalignment is zero. Finally, we included other assumptions in order to predict the transient behavior of the algorithm. Overall, the predicted results are in good agreement with the simulated ones in Section IV-C.

The decreasing property of the delta sequence provides the algorithm with robustness against large noise samples and good steady-state performance. However, an *ad hoc* nonstationary control was necessary in order to deal with system changes. We introduced two methods and tested them in system identification and acoustic echo cancelation applications. As shown in the simulations, even under severe conditions, their performance was very good.

Appendix A
Solution to (20)

Suppose that $E[\delta_\infty] \neq 0$. We denote as $C[0, E[\delta_\infty]]$ the space of continuous functions in $[0, E[\delta_\infty]]$ with norm $\|f\| = \max_{t\in[0,E[\delta_\infty]]} |f(t)|$. Clearly $z(t) = t$ belongs to $C[0, E[\delta_\infty]]$ and $\|z\| = E[\delta_\infty]$. At the same time $F_\infty(t) \in NBV[0, E[\delta_\infty]]$, the space of bounded variation functions which vanish at zero and are right-continuous in $(0, E[\delta_\infty])$ with norm $\|F\| = \int_0^{E[\delta_\infty]} |dF(t)|$. It is known that $NBV[0, E[\delta_\infty]]$ is the dual space of $C[0, E[\delta_\infty]]$ [18]. For that reason we can look at F_∞ as a linear functional on $C[0, E[\delta_\infty]]$ whose expression when it is evaluated in $z(t) = t$ is $\int_0^{E[\delta_\infty]} z\, dF_\infty(z)$. Clearly $\|F_\infty\| = P_\infty[z \leq E[\delta_\infty]]$. Then (20) can be put as

$$\|z\|\|F_\infty\| = \int_0^{E[\delta_\infty]} z\, dF_\infty(z). \tag{58}$$

This means that z and $F_\infty(z)$ are *aligned*. However, this happens if and only if $F_\infty(z)$ varies only at the points where z reach $\|z\|$ [18]. The only point where this happen is $E[\delta_\infty]$. Then it must be $F_\infty(z) = 0\,\forall z < E[\delta_\infty]$ with a discontinuity in $E[\delta_\infty]$ of height $P_\infty[z \leq E[\delta_\infty]]$. Of course, this is not reasonable for $F_\infty(z)$, especially if we take into account the presence in e_i of the noise v_i whose probability density extends over the entire real axis. Then we can not have (20) with $E[\delta_\infty] \neq 0$. It must be $E[\delta_\infty] = 0$.

Appendix B
Asymptotic Behavior of $E[\delta_i]$

We begin with (19), which can be put as

$$E[\delta_i] = E[\delta_{i-1}] - (1-\alpha) \int_0^{E[\delta_{i-1}]} \{E[\delta_{i-1}] - z\}\, p_z^i(z)\, dz \tag{59}$$

where $dF_i(z) = p_z^i(z)dz$ is the pdf at time i of $z \doteq e_i^2/\|\mathbf{x}_i\|^2$. Integrating the second term on the RHS of (59) it can be obtained:

$$\int_0^{E[\delta_{i-1}]} \{E[\delta_{i-1}] - z\}\, p_z^i(z)\, dz = [\{E[\delta_{i-1}] - z\}F_i(z)]_0^{E[\delta_{i-1}]} + \int_0^{E[\delta_{i-1}]} F_i(z)\, dz, \tag{60}$$

Taking into account that $F_i(0) = 0$ then

$$E[\delta_i] = E[\delta_{i-1}] - (1-\alpha) \int_0^{E[\delta_{i-1}]} F_i(z)\, dz. \tag{61}$$

Defining $p_{s,u}^i(s,u)$ as the joint pdf of $s \doteq e_i$ and $u \doteq \|\mathbf{x}_i\|^2$ at time i and assuming without loss of generality that it is symmetric on s, i.e., $p_{s,u}^i(s,u) = p_{s,u}^i(-s,u)\ \forall i,u$, we can obtain the pdf of $z = t/u \doteq e_i^2/\|\mathbf{x}_i\|^2$:

$$p_z^i(z) = z^{-1/2} g^i(z), g^i(z) = \int_0^\infty u^{1/2} p_{s,u}^i(\sqrt{zu}, u)\, du. \tag{62}$$

We have that $p_u^i(u) = p_u(u)\ \forall i$ because it was assumed that the input forms a stationary process. Then we can write

$$p_{s,u}^i(\sqrt{zu}, u) = p_{s|u}^i(\sqrt{zu}|u)p_u(u). \tag{63}$$

Before continuing we will make some assumptions about $p_{s|u}^i(\sqrt{zu}|u)$ and $p_u(u)$

A1) $\exists M > 0$ in such a way that $p_{s|u}^i(s|u) \leq M\ \forall s, u, i$.

A2) $p_u(u)$ decreases fast enough in order to have

$$\int_0^\infty u^{1/2} p_u(u) du = K < \infty. \tag{64}$$

A3) $\exists\ s_0, u_0, m > 0$ in such a way that for all $s \in [0, s_0]$ and $u \in [0, u_0]$ we have $p_{s|u}^i(s|u) > m > 0\ \forall i$.

Assumptions A1), A2), A3) are not too strong. A1) is asking for a bounded family of conditional pdfs. Asking for A2) is less than asking for the existence of the mean of $u \doteq \|\mathbf{x}_i\|^2$. A3) can be justified if the noise pdf is such that $p_v(0) \neq 0$ and $p_{s|u}^i(s|u)$ is continuous in s and u at the origin. With these assumptions we will obtain some insights into the asymptotic behaviour of $E[\delta_i]$. We have the following:

Theorem 1: Given (61) and using assumptions A1), A2), A3) we can bound $E[\delta_i]$

$$E[\delta_{i-1}] - \frac{4}{3}(1-\alpha)MK(E[\delta_{i-1}])^{3/2} \leq E[\delta_i] \leq E[\delta_{i-1}] - \frac{4}{3}(1-\alpha)mK'(E[\delta_{i-1}])^{3/2} \tag{65}$$

for $i \geq i_0$, where i_0 is such that $E[\delta_{i_0-1}] \leq z_0$ and K' and z_0 are positive numbers to be defined.

Proof: Consider the definition of $g^i(z)$ in (62). Using A1) and A2) we can write

$$\begin{aligned} g^i(z) &= \int_0^\infty u^{1/2} p_{s|u}^i(\sqrt{zu}|u)p_u(u)\, du \\ &\leq M \int_0^\infty u^{1/2} p_u(u)\, du \\ &= MK. \end{aligned} \tag{66}$$

Now, let us choose z_0, u_0' in such a way that $\sqrt{z_0 u_0'} = s_0$ and $u_0'' = \min\{u_0, u_0'\}$, where s_0 and u_0 are those given in A3). Equation (62) can be written as

$$g^i(z) = \int_0^{u_0''} u^{1/2} p_{s|u}^i(\sqrt{zu}|u)p_u(u)\, du + \int_{u_0''}^\infty u^{1/2} p_{s|u}^i(\sqrt{zu}|u)p_u(u)\, du. \tag{67}$$

Using assumption A3)

$$g^i(z) \geq m \int_0^{u_0''} u^{1/2} p_u(u)\, du. \tag{68}$$

Defining $K' = \int_0^{u_0''} u^{1/2} p_u(u)\, du$,

$$g^i(z) \geq mK' > 0 \quad \forall z \in [0, z_0]. \tag{69}$$

Using (66) and (69),

$$0 < mK' z^{-1/2} \leq p_z^i(z) \leq MKz^{-1/2}, \quad \forall i, \quad \forall z \in (0, z_0] \tag{70}$$

which implies

$$2mK' z^{1/2} \leq F_i(z) \leq 2MKz^{1/2}, \quad \forall i, \quad \forall z \in (0, z_0]. \tag{71}$$

Using (61) and (71) we obtain the desired result. □

Remember that $E[\delta_i]$ is a monotone decreasing sequence converging to zero, so given z_0 we can always find a value i_0 such that $E[\delta_{i_0-1}] \leq z_0$. The interesting fact about Theorem 1 is that we can study the behavior of $E[\delta_i]$ by studying the following sequence:

$$a_i = a_{i-1} - b(a_{i-1})^{3/2}. \tag{72}$$

It is not difficult to show that choosing i_0 such that

$$E[\delta_{i_0-1}] < \beta = \min\left\{\frac{1}{4(1-\alpha)^2 M^2 K^2}, z_0\right\} \tag{73}$$

and defining the sequences beginning at $i = i_0$

$$\begin{aligned} a_i' &= a_{i-1}' - \frac{4}{3}(1-\alpha)mK'(a_{i-1}')^{3/2}, \\ E[\delta_{i_0-1}] &\leq a_{i_0-1}' \leq \beta, \end{aligned} \tag{74}$$

$$\begin{aligned} a_i'' &= a_{i-1}'' - \frac{4}{3}(1-\alpha)MK(a_{i-1}'')^{3/2} \\ a_{i_0-1}'' &\leq E[\delta_{i_0-1}] < \beta \end{aligned} \tag{75}$$

we can write next

$$a_i'' = a_{i-1}'' - \frac{4}{3}(1-\alpha)MK(a_{i-1}'')^{3/2} \leq E[\delta_i] \leq a_{i-1}' - \frac{4}{3}(1-\alpha)mK'(a_{i-1}')^{3/2} = a_i' \quad \forall i \geq i_0. \tag{76}$$

It is not difficult to show that if we choose $c_1 \geq 9/(4(1-\alpha)^2 m^2 K'^2)$, $c_2 \leq 9/(64(1-\alpha)^2 K^2 M^2)$ and n_1 and n_2 in such a way that

$$a_{i_0-1}' = \frac{c_1}{(n_1 + i_0 - 1)^2} \leq \beta, \tag{77}$$

$$a_{i_0-1}'' = \frac{c_2}{(n_2 + i_0 - 1)^2} \leq E[\delta_{i_0-1}] \tag{78}$$

we can obtain

$$a_i' \leq \frac{c_1}{(n_1 + i)^2}, \quad a_i'' \geq \frac{c_2}{(n_2 + i)^2} \quad \forall i \geq i_0, \tag{79}$$

which implies (21). Considering

$$\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha} = E[\delta_{i-1}] - \int_0^{E[\delta_{i-1}]} F_i(z)\, dz, \tag{80}$$

we can follow the same procedure used to prove (21) in order to show that $(E[\delta_i] - \alpha E[\delta_{i-1}])/(1-\alpha)$ behaves also as $1/i^2$, which implies (26).

APPENDIX C
PROOF OF (33)

Clearly, $\sigma_{e_{a,i}}^2 = E[\tilde{\mathbf{w}}_{i-1}^T \mathbf{x}_i \mathbf{x}_i^T \tilde{\mathbf{w}}_{i-1}]$. If we assume that $\tilde{\mathbf{w}}_{i-1}$ and $\mathbf{x}_i$ are statistically independent it can be shown that $\sigma_{e_{a,i}}^2 =$

$E[\tilde{\mathbf{w}}_{i-1}^T\mathbf{R}\tilde{\mathbf{w}}_{i-1}]$ where $\mathbf{R} = E[\mathbf{x}_i\mathbf{x}_i^T]$. Using the eigenvalue decomposition of the autocorrelation matrix of the input regressors, $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, and defining $\mathbf{c}_{i-1} = \mathbf{Q}^T\tilde{\mathbf{w}}_{i-1}$, we can write

$$\sigma_{e_{\mathrm{a},\infty}}^2 = \lim_{i\to\infty}\sum_{j=1}^{M}\lambda_j E[(c_i^j)^2] \tag{81}$$

where c_i^j denotes the jth component of $\mathbf{c}_i$. Because of (32) and the positive definiteness of $\mathbf{R}$, it is clear that $\lim_{i\to\infty} E[(c_i^j)^2] = 0\ j = 1,\ldots,M$. This implies that $\lim_{i\to\infty} E[||\mathbf{c}_i||^2] = 0$. As $\mathbf{Q}$ is a unitary matrix, (33) follows.

Appendix D
Proof of (37)

We need the following lemma:

Lemma 1: Let s_1 and s_2 be jointly Gaussian zero-mean random variables with variances σ_1^2, σ_2^2 and let $y = s_2 + v$, with the pdf of v given in (36), and v independent of s_1 and s_2. Let $z_1 = s_2 + h_1$ and $z_2 = s_2 + h_2$ where h_1 and h_2 are zero-mean Gaussian variables with variances $\sigma_{h_1}^2 = (K+1)\sigma_b^2$ and $\sigma_{h_2}^2 = \sigma_b^2$ and independent of s_1 and s_2. Then

$$E[\mathrm{sign}(y)s_1] = pE[\mathrm{sign}(z_1)s_1] + (1-p)E[\mathrm{sign}(z_2)s_1]. \tag{82}$$

This lemma is a special case of Lemma 1 in [16] whose proof can be found there. We will make the following assumption:

$$E\left[e_i^2 \mid \tilde{\mathbf{w}}_{i-1}\right] \approx E\left[e_i^2\right] = \sigma_{e_i}^2. \tag{83}$$

This assumption was succesfully used in the past for analyzing the sign algorithm [19], and it is valid when δ_i is small enough. Although this is satisfied for the steady-state we take it only for mathematical tractability. However we will see that the results obtained are in agreement with the simulation results. It is clear that we only need to calculate $E[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T]$. We can write

$$E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T\right] = E\left\{E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T \mid \tilde{\mathbf{w}}_{i-1}\right]\right\}. \tag{84}$$

Using (36), Lemma 1, Price's theorem [14], (83) and (84) it holds

$$E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T \mid \tilde{\mathbf{w}}_{i-1}\right] = \tilde{\mathbf{w}}_{i-1}\sqrt{\frac{2}{\pi}}E\left[\mathbf{x}_i^T e_i \mid \tilde{\mathbf{w}}_{i-1}\right] \left\{\frac{p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + (K+1)\sigma_b^2}} + \frac{1-p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + \sigma_b^2}}\right\}. \tag{85}$$

Using that $E[\mathbf{x}_i^T e_i \mid \tilde{\mathbf{w}}_{i-1}] = \tilde{\mathbf{w}}_{i-1}^T\mathbf{R}$ and taking expectation with respect to $\tilde{\mathbf{w}}_{i-1}$ we obtain

$$E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T\right] = \sqrt{\frac{2}{\pi}}\mathbf{K}_{i-1}\mathbf{R} \times \left\{\frac{p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + (K+1)\sigma_b^2}} + \frac{1-p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + \sigma_b^2}}\right\}. \tag{86}$$

The term $E[\mathrm{sign}(e_i)\mathbf{x}_i\tilde{\mathbf{w}}_{i-1}^T]$ can be obtained with the same procedure.

Appendix E
Calculation of r and $\mathbf{G}$ for the Gaussian Case

The calculation of $E[1/||\mathbf{x}_i||]$ with $\mathbf{x}_i$ Gaussian and covariance matrix $\mathbf{R}$ is very difficult and there is no known closed form. In order to have an analytical approximation, we compute it for $\mathbf{x}_i$ Gaussian with covariance matrix $\sigma_x^2\mathbf{I}$. Using generalized spherical coordinates [20] it is not difficult to show that in this case

$$E\left[\frac{1}{||\mathbf{x}_i||}\right] = \frac{\Gamma\left(\frac{M-1}{2}\right)}{\sqrt{2}\sigma_x\Gamma\left(\frac{M}{2}\right)} = r, M \geq 2 \tag{87}$$

where $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ is the *complete gamma function*.

The term $E[(\mathbf{x}_i\mathbf{x}_i^T)/||\mathbf{x}_i||^2]$ can be evaluated exactly for any covariance matrix $\mathbf{R}$. The result is in terms of an integral that must be solved numerically. The ideas used for obtaining this result were presented in [21]

$$\mathbf{L} = \int_0^\infty \left(\prod_{j=1}^{M}\sqrt{\frac{1}{1+2\varepsilon\lambda_j}}\right)\mathrm{diag}\left(\frac{\lambda_i}{1+2\varepsilon\lambda_i}\right)d\varepsilon \tag{88}$$

where $\{\lambda_i\}_{i=1}^M$ are the eigenvalues of $\mathbf{R}$ and $\mathrm{diag}(\beta_i)$ is a diagonal matrix with its ith element equal to β_i. With $\mathbf{L}$ as in (88) and $\mathbf{Q}$ the matrix of eigenvectors of $\mathbf{R}$:

$$E\left[\frac{\mathbf{x}_i\mathbf{x}_i^T}{||\mathbf{x}_i||^2}\right] = \mathbf{Q}\mathbf{L}\mathbf{Q}^T = \mathbf{G}. \tag{89}$$

Appendix F
Calculation of (44) and (45)

Using

$$p_u(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Psi_u(\omega)e^{-j\omega u}d\omega \tag{90}$$

we can write (39) as

$$p_z(z) = \int_0^\infty up_t(zu)\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty}\Psi_u(\omega)e^{-j\omega u}d\omega\right\}du. \tag{91}$$

Assuming that we can interchange the order of integration (a sufficient condition for doing this is that $\Psi_u(\omega)$ is absolutely integrable)

$$p_z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Psi_u(\omega)\left\{\int_0^\infty up_t(zu)e^{-j\omega u}du\right\}d\omega. \tag{92}$$

The pdf of e_i^2 is

$$p_t(t) = p\,t^{-1/2}\frac{\exp\left[-\frac{t}{2\left[(K+1)\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2\right]}\right]}{\sqrt{(K+1)\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2}\sqrt{(2)}\Gamma\left(\frac{1}{2}\right)} + (1-p)\,t^{-1/2}\frac{\exp\left[-\frac{t}{2\left[\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2\right]}\right]}{\sqrt{\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2}\sqrt{(2)}\Gamma\left(\frac{1}{2}\right)}. \tag{93}$$

Using (93) it is not difficult to show that

$$\int_0^\infty up_t(zu)e^{-j\omega u}du = p\frac{z^{-1/2}\left[(K+1)\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2\right]}{\left(z+2i\left[(K+1)\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2\right]\omega\right)^{3/2}} + (1-p)\frac{z^{-1/2}\left[\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2\right]}{\left(z+2i\left[\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2\right]\omega\right)^{3/2}}. \quad (94)$$

Then, using (43), (92) can be put as

$$p_z(z) = p\frac{z^{-1/2}}{2\pi}\int_{-\infty}^{\infty}\frac{\Psi_u(\omega)\left[(K+1)\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2\right]}{\left[g\left(z,K,\sigma_{e_{\mathrm{a},i}}^2,\omega\right)\right]^3}d\omega + (1-p)\frac{z^{-1/2}}{2\pi}\int_{-\infty}^{\infty}\frac{\Psi_u(\omega)\left[\sigma_b^2+\sigma_{e_{\mathrm{a},i}}^2\right]}{\left[g\left(z,0,\sigma_{e_{\mathrm{a},i}}^2,\omega\right)\right]^3}d\omega. \quad (95)$$

Clearly

$$P_i[z > E[\delta_{i-1}]] = \int_{E[\delta_{i-1}]}^{\infty} p_z(z)dz. \quad (96)$$

Using (95) in (96) and interchanging the order of integration we arrive at the result in (44). The calculation of (45) follows the same steps. We should mention that the integrals in (44) and (45) are well defined in the vecinity of $\omega = 0$, because the integrands in (44) remain finite when $\omega \to 0$ and the integrands of (45) tends to $\log(\omega^{-1/2})$ as $\omega \to 0$ which is an integrable function.

Appendix G
Proof of (49)

Suppose we have the probability space $(\Omega, \mathcal{F}, P)$, where Ω denotes an arbitrary space, $\mathcal{F}$ denotes a σ-field of events in Ω, and P is a probability measure on Ω. It is not difficult to see that

$$P\left[\omega \in \Omega : \exists \lim_{i\to\infty}\delta_i(\omega)\right] = 1. \quad (97)$$

This follows from the fact that for almost every $\omega \in \Omega$ the sequence $\{\delta_i(\omega)\}$ is decreasing and nonnegative. We denote that limit as $\delta_\infty(\omega)$ (in the following we will drop the explicit dependence on ω in all random variables). From (10), it follows immediately that a sufficient condition for having $\delta_\infty = 0$ is $\lim_{i\to\infty}\inf(e_i^2/\|\mathbf{x}_i\|^2) = 0$ a.e. This condition can be put in the following alternative manner:

$$P\left[\omega \in \Omega : \frac{e_i^2}{\|\mathbf{x}_i\|^2} < \epsilon \text{ infinitely often}\right] = 1 \quad \forall\epsilon > 0. \quad (98)$$

Defining $A_i^\epsilon = \{\omega \in \Omega : e_i^2/\|\mathbf{x}_i\|^2 < \epsilon\}\ \forall i > 0, \forall\epsilon > 0$ we can write (98) as

$$P\left[\bigcap_{i=1}^{\infty}\bigcup_{m=i}^{\infty}A_m^\epsilon\right] = 1 \quad \forall\epsilon > 0. \quad (99)$$

Assuming that $e_i^2/\|\mathbf{x}_i\|^2$ has a limiting distribution, and $\forall i$ (including the limiting distribution) the distributions are nonzero in an interval $[0,\epsilon)$, $\forall\epsilon > 0$, it is reasonable to suppose that $\sum_{i=1}^{\infty}P[A_i^\epsilon] = \infty\ \forall\epsilon > 0$. Note that if the events $\{A_i^\epsilon\}$ are independent, (99) can be established using the second Borel-Cantelli lemma [22]. If this is not the situation, the second Borel-Cantelli lemma can fail. For this reason we will impose on the process $\{e_i^2/\|\mathbf{x}_i\|^2\}$ the ψ-mixing condition [23], i.e., $\lim_{n\to\infty}\psi(n) = 0$, where

$$\psi(n) = \sup_{j>0}\sup_{\substack{A\in\mathcal{F}_0^j, B\in\mathcal{F}_{j+n}^\infty \\ P[A]P[B]>0}}\left\{\left|\frac{P[A\bigcap B]}{P[A]P[B]} - 1\right|\right\} \quad (100)$$

and $\mathcal{F}_m^n$ stands for the σ-field spanned by $\{e_i^2/\|\mathbf{x}_i\|^2, m \le i \le n\}$. This is one of the various measures of dependence for stochastic processes [23]. We see that the ψ-mixing condition says that events corresponding to a given stochastic process that are distant in time are almost independent. Of course, if the process is m-dependent, it is ψ-mixing too. The reverse implication is not true in general. Next, we present the following proposition that was proved in [22].

Proposition 1: Let $\{A_n\}$ be a sequence of events such that $\sum_{n=1}^{\infty}P[A_n] = \infty$ and the indicator functions

$$I_{A_n}(\omega) = \begin{cases} 1 & \text{if} \quad \omega \in A_n \\ 0 & \text{if} \quad \omega \notin A_n. \end{cases}$$

If they form a ψ-mixing process, i.e., $\lim_{n\to\infty}\psi(n) = 0$, then

$$\lim_{n\to\infty}\frac{\sum_{i=1}^n I_{A_i}}{\sum_{i=1}^n P[A_i]} = 1 \ a.e. \quad (101)$$

Now we can introduce the following theorem.

Theorem 2: Given (10) and (101), assuming the process $\{e_i^2/\|\mathbf{x}_i\|^2\}$ is ψ-mixing and $\sum_{i=1}^{\infty}P[A_i^\epsilon] = \infty\ \forall\epsilon > 0$, then (99) holds. Thus, $\delta_\infty = 0\ a.e.$

Proof: Let $S_n = \sum_{i=1}^n I_{A_i^\epsilon}(\omega)$ and $T_n = \sum_{i=1}^n P[A_i^\epsilon]$. Suppose that $P[\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}A_m^\epsilon] < 1$. This is equivalent to $P[\omega \in \Omega : \exists n, \forall m \ge n, \omega \notin A_m^\epsilon] > 0$. Let $B = \{\omega \in \Omega : \sum_{n=1}^{\infty}I_{A_n^\epsilon}(\omega) < \infty\}$. It should be clear that we must have $P[B] > 0$. This means that $\forall\omega \in B$ we have $\lim_{n\to\infty}(S_n/T_n) = 0$ because $\lim_{n\to\infty}T_n = \infty$.
Now it is evident that (101) cannot be true because $P[B] = P[\omega \in \Omega : \lim_{n\to\infty}(S_n)/(T_n) = 0] > 0$. Then it must be true that $P[\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}A_m^\epsilon] = 1$. □

Now we can see that the conditions of Theorem 2 can be relaxed by only asking the ψ-mixing property for the process $\{I_{A_i^\epsilon}\}\ \forall\epsilon > 0$. In fact, instead of the ψ-mixing property we can set weaker conditions on the dependence of the process $\{I_{A_i^\epsilon}\}\ \forall\epsilon > 0$ (α-mixing property, ρ-mixing property, θ-mixing property, etc. [22]–[24]) if we assume a certain rate of divergence for the series $\sum_{i=1}^{\infty}P[A_i^\epsilon]$.

Acknowledgment

The authors would like to thank Prof. S. Grynberg and Prof. C. Belaustegui Goitia from the School of Engineering, University of Buenos Aires, for their very helpful remarks.

References

[1] T. Gänsler, S. L. Gay, M. M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancelation," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 656–663, Nov. 2000.
[2] T. Gänsler and J. Benesty, "The fast normalized cross-correlation double-talk detector," *Signal Process.*, vol. 86, no. 6, pp. 1124–1139, Jun. 2006.

[3] P. Petrus, "Robust Huber adaptive filter," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 1129–1133, Apr. 1999.
[4] Y. Zou, S. C. Chan, and T. S. Ng, "Least mean M-estimate algorithms for robust adaptive filtering in impulse noise," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 12, pp. 1564–1569, Dec. 2000.
[5] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive adaptive filter algorithm," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 46–48, Feb. 1997.
[6] E. V. Papoulis and T. Stathaki, "A normalized robust mixed-norm adaptive algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 11, no. 1, pp. 56–59, Jan. 2004.
[7] E. V. Papoulis and T. Stathaki, "Improved performance adaptive algorithm for system identification under impulsive noise," *Electron. Lett.*, vol. 39, no. 11, pp. 878–880, May 2003.
[8] J. Arenas-García and A. R. Figueiras-Vidal, "Adaptive combination of normalized filters for robust system identification," *Electron. Lett.*, vol. 41, no. 15, pp. 874–875, Jul. 2005.
[9] P. J. Huber, *Robust Statistics*. New York: Wiley , 1981.
[10] L. R. Vega, H. G. Rey, J. Benesty, and S. Tressens, "A robust adaptive filtering algorithm against impulsive noise," in *Proc. ICASSP-2007*, Honolulu, HI, Apr. 2007.
[11] L. R. Vega, H. G. Rey, J. Benesty, and S. Tressens, "A stochastic model for a new robust NLMS algorithm," in *Proc. EUSIPCO-2007*, Poznan, Poland, Sep. 2007.
[12] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2001.
[13] T. Al-Naffouri and A. Sayed, "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 653–663, Mar. 2003.
[14] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inf. Theory*, vol. IT-4, pp. 69–72, Jun. 1958.
[15] P. G. Georgiou, P. Tsakalides, and C. Kyriakakis, "Alpha-stable modeling of noise and robust time- delay estimation in the presence of impulsive noise," *IEEE Trans. Multimedia*, vol. 1, no. 3, pp. 291–301, Sep. 1999.
[16] S. Chan Bang, S. Ann, and I. Song, "Performance analysis of the dual sign algorithm for additive contaminated-Gaussian noise," *IEEE Signal Process. Lett.*, vol. 1, no. 12, pp. 196–198, Dec. 1994.
[17] S. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. ICASSP-95*, Detroit, MI, 1995, pp. 3023–3026.
[18] D. Luenberger, *Optmization by Vector Space Methods*. New York: Wiley, 1969.
[19] V. J. Mathews and S. H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 4, pp. 450–454, Apr. 1987.
[20] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithm with uncorrelated data," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 680–691, Jul. 1988.
[21] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 4, pp. 793–806, Aug. 1986.
[22] D. Tasche, "On the second Borel-Cantelli lemma for strongly mixing sequences of events," *J. Appl. Probabil.*, vol. 34, pp. 381–394, 1997.
[23] R. C. Bradley, "Basic properties of strong mixing conditions. A survey and some open questions," *Probabil. Surveys*, vol. 2, pp. 107–144, 2005.
[24] M. Iosifescu and R. Theodorescu, *Random Processes and Learning*. New York: Springer-Verlag, 1969.
[25] D. L. Duttweiler, "A twelve channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647–653, May 1978.

Leonardo Rey Vega was born in Buenos Aires, Argentina, in 1979. He received the B.Eng. degree in electronic engineering from University of Buenos Aires in 2004.

Since 2004, he has been a Research Assistant with the Department of Electronics, University of Buenos Aires, where he is currently pursuing the Ph.D. degree. His research interests include adaptive filtering theory and statistical signal processing.

Hernán Rey was born in Buenos Aires, Argentina, in 1978. He received the B.Eng. degree in electronic engineering from the University of Buenos Aires, in 2002.

Since 2002, he has been a Research Assistant with the Department of Electronics, University of Buenos Aires. In 2004, he joined the Institute of Biomedical Engineering, University of Buenos Aires, where he is currently pursuing the Ph.D. degree. His research interests include adaptive filtering theory, neural networks, and computational neuroscience.

Jacob Benesty (M'97–SM'04) was born in 1963. He received the Masters degree in microwaves from Pierre and Marie Curie University, France, in 1987, and the Ph.D. degree in control and signal processing from Orsay University, France, in April 1991. During his Ph.D. degree program (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris, France.

From January 1994 to July 1995, he was with Telecom Paris University working on multichannel adaptive filters and acoustic echo cancelation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff of Bell Laboratories, Murray Hill, NJ. In May 2003, he joined the University of Quebec, INRS-EMT, Montreal, Quebec, Canada, as an associate professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He coauthored the books *Acoustic MIMO Signal Processing* (Boston, MA: Springer-Verlag, 2006) and *Advances in Network and Acoustic Echo Cancelation* (Berlin, Germany: Springer-Verlag, 2001). He is also a coeditor/coauthor of the books Speech Enhancement (Berlin, Germany: Springer-Verlag, 2005), *Audio Signal Processing for Next Generation Multimedia Communication Systems* (Boston, MA: Kluwer Academic, 2004), *Adaptive Signal Processing: Applications to Real-World Problems* (Berlin, Germany: Springer-Verlag, 2003), and *Acoustic Signal Processing for Telecommunication* (Boston, MA: Kluwer Academic, 2000).

Dr. Benesty received the 2001 Best Paper Award from the IEEE Signal Processing Society. He was a member of the editorial board of the EURASIP *Journal on Applied Signal Processing* and was the Co-Chair of the 1999 International Workshop on Acoustic Echo and Noise Control.

Sara Tressens was born in Buenos Aires, Argentina. She received the degree of electrical engineering from the University of Buenos Aires in 1967.

From 1967 to 1982, she was with the Institute of Biomedical Engineering, University of Buenos Aires, where she became Assistant Professor in 1977 and worked in the area of speech recognition, digital communication and adaptive filtering. During 1980–1981, she worked in the Laboratoire des Signaux et Systemes, Gif-sur Yvette, France. From 1982 to 1993, she was with the National Research Center in Telecommunications (CNET), Issy les Moulineaux, France. Her research interests were in the areas of spectral estimation and spatial arrays processing. From 1994 to 2006, she was an Associate Professor with the University de Buenos Aires. Since 2006, she has been a Full Professor. Her primary research interests include adaptive filtering, communications, and spectral analysis.

Professor Tressens received the 1990 Best Paper Award from the IEEE Signal Processing Society.