A STOCHASTIC MODEL FOR A NEW ROBUST NLMS ALGORITHM

Leonardo Rey Vega†, *Hernán Rey*†, *Jacob Benesty*‡ and *Sara Tressens*†

† Facultad de Ingeniería, Universidad de Buenos Aires
Paseo Colón 850, Buenos Aires, Argentina (email: lrey,hrey,stres@fi.uba.ar)
‡ INRS-EMT, Université du Québec
800 de la Gauchetiere Ouest, Montreal, Québec, Canada (email: benesty@emt.inrs.ca)

ABSTRACT

We present a stochastic model for a new recently proposed robust NLMS algorithm. Under very standard and reasonable assumptions we show that the algorithm converges to the true unknown system in a mean square sense. With the aggregate of more restrictive, but standard, assumptions we can build a model for the transient behavior of the algorithm. The model can take into account the presence of impulsive noise. Finally we also present simulations results which show the excellent agreement with the model.

1. 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 cancellation, 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. Most of them are directly or indirectly related with the optimization of a combination of L_1 and L_2 norms as the objective function. The former presents a low sensitivity against perturbations and the latter improves the convergence speed of the adaptive filter. Recently, a new robust NLMS-like algorithm has been introduced based on a novel design framework [1]. It provides an automatic mechanism for switching between the normalized least-mean-square (NLMS) and normalized sign algorithm (NSA).

Although its robust performance is guaranteed by design, 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. Although several (sometimes strong) assumptions are required, the predictions and the simulated results are in good agreement.

Finally, we present certain definitions and notations that are used in the paper. Let $\mathbf{w}_i = \left(w_{i,0}, w_{i,1}, \ldots, w_{i,M-1}\right)^T$ 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, $v_i = v_i^B + v_i^I$, where v_i^B stands for the background measurement noise and v_i^I is an impulsive noise. 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 $e_{\mathrm{p},i} = \mathbf{x}_i^T \tilde{\mathbf{w}}_i + v_i$.

This work was supported in part by Universidad de Buenos Aires, project UBACYT I005.

2. THE NEW ROBUST NLMS ALGORITHM

In this section we will briefly present the new robust NLMS algorithm [1]. A great number of robust adaptive filters were derived in the past, [2]-[5]. In the derivation of these adaptive filters robust statistic ideas were used. 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, 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. However for the design of these adaptive filters we need to have statistical information about the noise v_i or the error signal e_i, which can be difficult to have, specially for e_i which is clearly a non-stationary stochastic process. In [1] a different approach was proposed to solve the robust adaptive filtering problem. Suppose an adaptive filter has a given estimate of the true system at a certain time-step. Now, 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. This is the problem with the standard NLMS algorithm in an impulsive environment. Because impulsive noise samples can occur infinitely often, i.e. double-talk situations in echo cancelation, the algorithm will always present a very poor performance. 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:

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

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

Next, a cost function is required and the adaptive filter will be the result of optimizing this cost function subject to the constraint (1). Different choices of the cost function and the $\{\delta_i\}$ sequence will lead to different algorithms. Although other cost functions can be chosen, for simplicity and ease of treatment we choose to minimize the square of the *a posteriori* error signal. Then,

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

subject to the constraint (1). The details regarding this optimization problem can be found in [1]. Here we only present the final result:

$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1} + \min\left[\frac{|e_i|}{||\mathbf{x}_i||}, \sqrt{\delta_{i-1}}\right] \mathrm{sign}(e_i) \frac{\mathbf{x}_i}{||\mathbf{x}_i||}. \tag{3}$$

We see that the quantity $\frac{|e_i|}{||\mathbf{x}_i||}$ acts as a decision variable: if $\frac{|e_i|}{||\mathbf{x}_i||} \leq \sqrt{\delta_{i-1}}$ then the algorithm reduces to standard NLMS with unity step-size. If $\frac{|e_i|}{||\mathbf{x}_i||} > \sqrt{\delta_{i-1}}$, then the algorithm reduces to a normalized sign algorithm (NSA) with step-size $\sqrt{\delta_{i-1}}$. Then, depending on the values of $\frac{|e_i|}{||\mathbf{x}_i||}$ and δ_{i-1} the algorithm has two operation modes. The NLMS mode will give a fast convergence and the NSA mode will give the robust behavior.

The only thing that we have to do is to choose the values of $\{\delta_i\}$ which will affect the performance of the algorithm and the switching between the two modes of operation. 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 too large, so that the robust performance against large noise samples is not 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:

$$\begin{aligned}\delta_i &= \alpha\delta_{i-1} + (1-\alpha)||\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}||^2 \\ &= \alpha\delta_{i-1} + (1-\alpha)\min\left[\frac{e_i^2}{||\mathbf{x}_i||^2}, \delta_{i-1}\right],\end{aligned} \tag{4}$$

where $0 < \alpha < 1$ is a memory factor and δ_0 is the initial value in (4). Equations (3) and (4) define the RVSS-NLMS algorithm.

3. STOCHASTIC MODEL

In this section we will present some results concerning the stochastic behavior of the algorithm. Using (3) and (4), 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:

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\,\mathrm{sign}(e_i)\frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}. \tag{5}$$

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 (5) we can write:

$$\|\tilde{\mathbf{w}}_i\|^2 = \|\tilde{\mathbf{w}}_{i-1}\|^2 - 2\sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\frac{\mathrm{sign}(e_i)e_{\mathrm{a},i}}{\|\mathbf{x}_i\|} + \frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}. \tag{6}$$

Now we make the following assumptions:

A1): The noise sequence can be put as $v_i = v_i^B + v_i^I$, where v_i^B is a background noise assumed to be Gaussian with zero mean and $E[(v_i^B)^2] = \sigma_B^2$. v_i^I is the impulsive part of the noise and can be written as $v_i^I = \omega_i N_i$, 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:

$$p_{v_i}(v_i) = p\mathcal{N}\left(0,(K+1)\sigma_B^2\right) + (1-p)\mathcal{N}\left(0,\sigma_B^2\right) \tag{7}$$

The noise sequence is independent of the input regressors $\mathbf{x}_i$, which belong to a zero-mean stationary process as well.

A2): The variance of δ_i is sufficiently small to approximate:

$$E\left[\sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\right] \approx \sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}}, \tag{8}$$

$$E\left[\min\left[\frac{e_i^2}{\|\mathbf{x}_i\|^2}, \delta_{i-1}\right]\right] \approx E[\delta_{i-1}]P_i[z \geq E[\delta_{i-1}]] + \int_0^{E[\delta_{i-1}]} z\, dF_i(z), \tag{9}$$

where $z \doteq e_i^2/\|\mathbf{x}_i\|^2$, i.e., they have the same distribution, $P_i[A]$ denotes the probability of the event A at time-step i and $F_i(z)$ denotes the distribution function of z. We will also assume that δ_i is statistically independent of $\mathbf{x}_j$ and $v_i\ \forall i,j$.

A3): The filter length is sufficiently large to assume that $e_{\mathrm{a},i}$ is Gaussian distributed.

A1) is a reasonable assumption. A2) is necessary to account for the behavior of $\{\delta_i\}$ and its influence on equation (6) in a simple way. It seems to be a strong assumption but in fact it is very accurate as we will see later in the simulations. A3) can be justified by central limit arguments and it was used and verified with simulations in [6].

Using the first part of A2), and taking expectation in (6):

$$\begin{aligned}E[\|\tilde{\mathbf{w}}_i\|^2] = E[\|\tilde{\mathbf{w}}_{i-1}\|^2] - 2\sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}}\,\cdot \\ E\left[\frac{\mathrm{sign}(e_i)e_{\mathrm{a},i}}{\|\mathbf{x}_i\|}\right] + \frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}.\end{aligned} \tag{10}$$

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

$$E\left[\frac{\mathrm{sign}(e_i)e_{\mathrm{a},i}}{\|\mathbf{x}_i\|}\right] \approx E\left[\frac{1}{\|\mathbf{x}_i\|}\right]E\left[\mathrm{sign}(e_i)e_{\mathrm{a},i}\right]. \tag{11}$$

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 $\mathrm{sign}(e_i)e_{\mathrm{a},i}$. Now, before proceeding we need the following lemma:

Lemma 1 *Let b be a Gaussian zero-mean random variable with variance σ_b^2, and let $y = b + v$, with the PDF of v given in (7), and v independent of b. Let $z_1 = b + h_1$ and $z_2 = b + h_2$ where h_1 and h_2 are independent of b and are zero-mean Gaussian variables with variances $\sigma_{h_1}^2 = (K+1)\sigma_B^2$ and $\sigma_{h_2}^2 = \sigma_B^2$. Then:*

$$E\left[\mathrm{sign}(y)b\right] = pE\left[\mathrm{sign}(z_1)b\right] + (1-p)E\left[\mathrm{sign}(z_2)b\right] \tag{12}$$

This lemma is a special case of Lemma 1 in [7] whose proof can be found there. Using Lemma 1 and Price‘s theorem [8] we can show:

$$E[\mathrm{sign}(e_i)e_{\mathrm{a},i}] = \sqrt{\frac{2}{\pi}}\sigma_{e_{\mathrm{a},i}}^2\left\{\frac{p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + (K+1)\sigma_B^2}} + \frac{1-p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + \sigma_B^2}}\right\} \tag{13}$$

where $\sigma_{e_{\mathrm{a},i}}^2 = E[e_{\mathrm{a},i}^2]$. Defining $r = E[1/\|\mathbf{x}_i\|]$, (10) can be put as:

$$\begin{aligned}E[\|\tilde{\mathbf{w}}_i\|^2] = E[\|\tilde{\mathbf{w}}_{i-1}\|^2] - 2r\sqrt{\frac{2}{\pi}}\sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}}\sigma_{e_{\mathrm{a},i}}^2\,\cdot \\ \left\{\frac{p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + (K+1)\sigma_B^2}} + \frac{1-p}{\sqrt{\sigma_{e_{\mathrm{a},i}}^2 + \sigma_B^2}}\right\} + \frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}.\end{aligned} \tag{14}$$

Because $E[\delta_i]$ is time-variant we need a recursion for it. Using the second part of A2) and (4) yields:

$$E[\delta_i] = \alpha E[\delta_{i-1}] + (1-\alpha)\left\{E[\delta_{i-1}]P_i[z \geq E[\delta_{i-1}]] + \int_0^{E[\delta_{i-1}]} z\, dF_i(z)\right\} \tag{15}$$

After (14) and (15) we can obtain results for the steady-state behavior of the algorithm.

3.1 Steady-State Behavior

Assuming that the limit of (14) exists, we can take limits as $i \to \infty$ to obtain:

$$\sqrt{E[\delta_\infty]} = 2r\sqrt{\frac{2}{\pi}}\sigma_{e_{\mathrm{a},\infty}}^2\left\{\frac{p}{\sqrt{\sigma_{e_{\mathrm{a},\infty}}^2 + (K+1)\sigma_B^2}} + \frac{1-p}{\sqrt{\sigma_{e_{\mathrm{a},\infty}}^2 + \sigma_B^2}}\right\} \tag{16}$$

where $E[\delta_\infty] \equiv \lim_{i\to\infty} E[\delta_i]$ and $\sigma^2_{e_{a,\infty}} \equiv \lim_{i\to\infty} \sigma^2_{e_{a,i}}$. Taking limits in (15) (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 (which is reasonable in for the assumptions considered) leads to:

$$E[\delta_\infty] P_\infty[z < E[\delta_\infty]] = \int_0^{E[\delta_\infty]} z\, dF_\infty(z), \tag{17}$$

where $P_\infty[z < \delta_\infty] \equiv \lim_{i\to\infty} P_i[z < \delta_i]$ and $F_\infty(\cdot)$ denotes the limiting distribution of $e_i^2/\|\mathbf{x}_i\|^2$. It should be clear that (17) is satisfied when $E[\delta_\infty] = 0$. If the limiting distribution of $e_i^2/\|\mathbf{x}_i\|^2$ is such that $P_\infty[z < c] = 0$ for some $c > 0$ then it could happen that $E[\delta_\infty] = a$ with $a \leq c$. But because of the presence of the noise v_i, which is a continuous random variable whose probability density extends over the entire real axis, this is not a realistic assumption. So we will have $E[\delta_\infty] = 0$. Replacing this in (16) we obtain:

$$\sigma^2_{e_{a,\infty}} = 0. \tag{18}$$

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

$$\lim_{i\to\infty} E[\|\tilde{\mathbf{w}}_i\|^2] = 0. \tag{19}$$

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.

3.2 Transient Behavior

Equations (14) and (15) can be used to study the transient behavior of the algorithm. However these equations are not self-contained. We need one more equation linking $\sigma^2_{e_{a,i}}$ 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 will also need two more assumptions:

B1): The input regressors $\{\mathbf{x}_i\}_{i=0}^{\infty}$ are statistically independent and Gaussian with zero mean and covariance matrix $E\left[\mathbf{x}_i\mathbf{x}_i^T\right] = \mathbf{R}$.

B2): We can approximate the mean squared value of e_i conditioned on the misalignment vector $\tilde{\mathbf{w}}_{i-1}$, by the unconditional mean squared error:

$$E\left[e_i^2|\tilde{\mathbf{w}}_{i-1}\right] \approx E\left[e_i^2\right] = \sigma^2_{e_i}. \tag{20}$$

B1) is a classical assumption and it has been used many times in the adaptive filter literature [9], [10]. B2) was succesfully used in the past for analyzing the sign algorithm [9], and it is valid when δ_i is small enough. Although this is strictly valid 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. Defining $\mathbf{K}_i = E[\tilde{\mathbf{w}}_i\tilde{\mathbf{w}}_i^T]$, using (5), A1), B1), A2), B2), Lemma 1 and Price‘s theorem we prove in the appendix that:

$$\begin{aligned}\mathbf{K}_i = \mathbf{K}_{i-1} - r\sqrt{\frac{2}{\pi}\frac{E[\delta_i - \alpha\delta_{i-1}]}{1-\alpha}}\,[\mathbf{R}\mathbf{K}_{i-1} + \mathbf{K}_{i-1}\mathbf{R}]\cdot \\ \left\{\frac{p}{\sqrt{\sigma^2_{e_{a,i}} + (K+1)\sigma_B^2}} + \frac{1-p}{\sqrt{\sigma^2_{e_{a,i}} + \sigma_B^2}}\right\} + \frac{E[\delta_i - \alpha\delta_{i-1}]}{1-\alpha}\mathbf{G},\end{aligned} \tag{21}$$

$$\sigma^2_{e_{a,i}} = \mathrm{Tr}\left[\mathbf{K}_{i-1}\mathbf{R}\right], \tag{22}$$

where $\mathbf{G} = E\left[(\mathbf{x}_i\mathbf{x}_i^T)/\|\mathbf{x}_i\|^2\right]$. It is interesting to observe that if we take trace to both sides of equation (21) and use (22), we obtain (14). This proves that B1) and B2) give results consistent with (14) which was based on weaker assumptions. Then with (15), (21) and (22) 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. The calculation of $\mathbf{G}$ is based on the ideas introduced in [10]. The calculation of $r = E\left[1/\|\mathbf{x}_i\|\right]$ with $\mathbf{x}_i$ Gaussian and covariance matrix $\mathbf{R}$ is very difficult and there is no known closed form. For this reason we compute it for $\mathbf{x}_i$ Gaussian with covariance matrix $\sigma_x^2\mathbf{I}$. Using generalized spherical coordinates it is not difficult to show that in this case:

$$E\left[\frac{1}{\|\mathbf{x}_i\|}\right] = \frac{\Gamma\left(\frac{M-1}{2}\right)}{\sqrt{2}\sigma_x\Gamma\left(\frac{M}{2}\right)} = r, \;\; M \geq 2, \tag{23}$$

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

In order to obtain a more explicit recursion for (15) we need A3). This assumption, together with A1) allows us to characterize e_i as a mixture of two Gaussian variables with mixing parameters p and $1-p$ and variances $\sigma^2_{e_{a,i}} + (K+1)\sigma_B^2$ and $\sigma^2_{e_{a,i}} + \sigma_B^2$ respectively. We want to obtain the PDF of $\frac{e_i^2}{\|\mathbf{x}_i\|^2}$. It is not difficult to show that e_i^2 is a mixture of two non-standard χ^2 distributions with mixing parameters p and $1-p$. In the general case with covariance matrix $\mathbf{R}$, there is no simple expression for the PDF of $\|\mathbf{x}_i\|^2$. Because of this, we consider that the distribution of $\|\mathbf{x}_i\|^2$ is χ^2 as if the input regressors had a diagonal covariance matrix, i.e. $\sigma_x^2\mathbf{I}$. 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, $e_{a,i}$ can be small compared with the background noise part of v_i. Then, v_i will dominate e_i, and because of A1) the independence assumption between e_i^2 and $\|\mathbf{x}_i\|^2$ will hold approximately. In the simulations section we will see that these assumptions give good results. Making the corresponding change of variables the PDF of $z \doteq e_i^2/\|\mathbf{x}_i\|^2$ can be show to be a mixture of two non-standard F distributions. More precisely, for $z \geq 0$:

$$\begin{aligned}p_z(z) = \frac{\Gamma[(M+1)/2]}{\Gamma[1/2]\Gamma[M/2]\sigma_x^M}\left\{\frac{p\left((K+1)\sigma_B^2 + \sigma^2_{e_{a,i}}\right)^{\frac{M}{2}} z^{-1/2}}{\left(z + \frac{(K+1)\sigma_B^2 + \sigma^2_{e_{a,i}}}{\sigma_x^2}\right)^{\frac{M+1}{2}}} + \right. \\ \left. \frac{(1-p)\left(\sigma_B^2 + \sigma^2_{e_{a,i}}\right)^{\frac{M}{2}} z^{-1/2}}{\left(z + \frac{\sigma_B^2 + \sigma^2_{e_{a,i}}}{\sigma_x^2}\right)^{\frac{M+1}{2}}}\right\}.\end{aligned} \tag{24}$$

Now we can show that:

$$\begin{aligned}P_i\left[z \geq E\left[\delta_{i-1}\right]\right] = 1 - p\,\mathrm{cdf}\left[\frac{M\sigma_x^2 E\left[\delta_{i-1}\right]}{\sigma^2_{e_{a,i}} + (K+1)\sigma_B^2}, 1, M\right] - \\ (1-p)\mathrm{cdf}\left[\frac{M\sigma_x^2 E\left[\delta_{i-1}\right]}{\sigma^2_{e_{a,i}} + \sigma_B^2}, 1, M\right],\end{aligned} \tag{25}$$

where we define the cumulative density function:

$$\mathrm{cdf}[\gamma, m_1, m_2] = \frac{1}{\beta\left(\frac{m_1}{2}, \frac{m_2}{2}\right)}\left(\frac{m_1}{m_2}\right)^{\frac{m_1}{2}}\int_0^\gamma \frac{\psi^{\frac{m_1}{2}-1}}{\left(1 + \frac{m_1}{m_2}\psi\right)^{\frac{m_1+m_2}{2}}}\,d\psi, \tag{26}$$

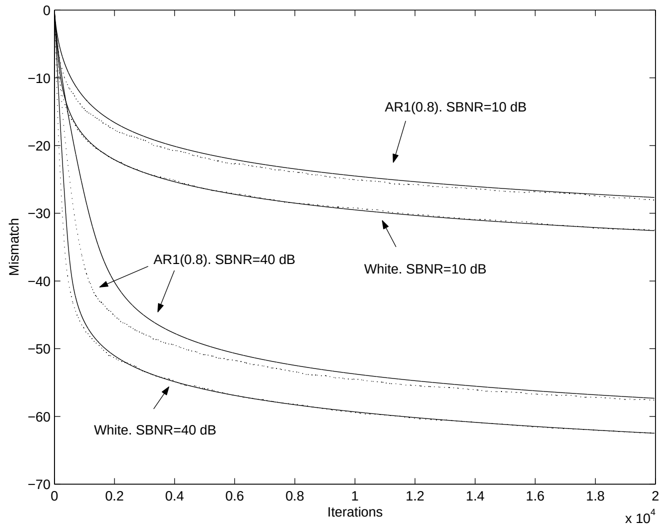

Figure 1: Mismatch (in dB) under different conditions. Inputs: white and AR1(0.8). SBNR: low (10 dB) and high (40 dB). The theoretical predictions are in solid line while the simulated results are in dotted line. $M = 32$. $\kappa = 1$ for white input and $\kappa = 3$ for AR1(0.8) input. $p = 0.1$ and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

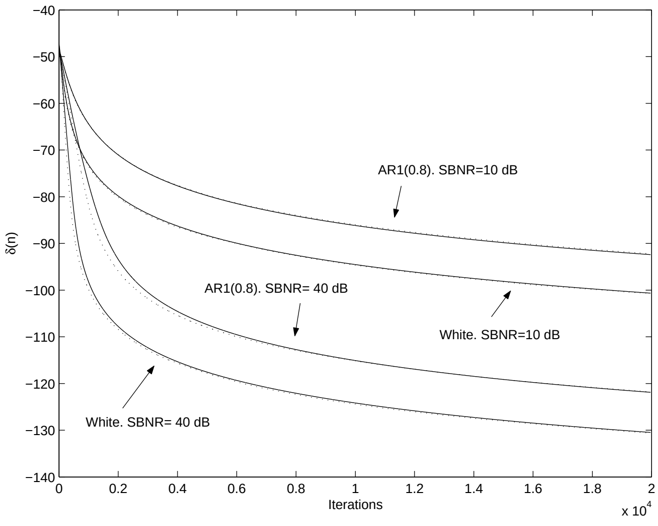

Figure 2: Evolution of $\{\delta_i\}$ (in dB). The setup is the same as in Fig. 1. Although the theoretical result is an ensamble average, the experimental one corresponds to a single run.

where $\beta(n,m) = \int_0^1 x^{n-1}(1-x)^{m-1}dx$. In the same way:

$$\int_0^{E[\delta_{i-1}]} zp(z)\,dz = \frac{\Gamma(3/2)\Gamma[(M-2)/2]}{\Gamma(1/2)\Gamma(M/2)}\cdot$$

$$\left\{ p\left(\frac{\sigma^2_{e_{a,i}} + (K+1)\sigma^2_B}{\sigma^2_x}\right) \mathrm{cdf}\left[\frac{(M-2)\sigma^2_x E[\delta_{i-1}]}{3\left(\sigma^2_{e_{a,i}} + (K+1)\sigma^2_B\right)}, 3, M-2\right] + \right.$$
$$\left. (1-p)\left(\frac{\sigma^2_{e_{a,i}} + \sigma^2_B}{\sigma^2_x}\right) \mathrm{cdf}\left[\frac{(M-2)\sigma^2_x E[\delta_{i-1}]}{3\left(\sigma^2_{e_{a,i}} + \sigma^2_B\right)}, 3, M-2\right] \right\}. \tag{27}$$

Then we have the complete model for the transient behavior of the algorithm.

4. SIMULATIONS

We will test the preceding model. We use white and AR1(0.8) input signals. The forgetting factor α is chosen according to the rule of thumb:

$$\alpha = 1 - \frac{1}{\kappa M}, \tag{28}$$

where κ is a parameter that depends on the color of the input signal and tipically is between 1 and 6. The signal to background noise ratio is defined as:

$$\mathrm{SBNR} = 10\log_{10}\left[\frac{\sigma^2_y}{\sigma^2_b}\right]. \tag{29}$$

where σ^2_y and σ^2_B are the power of the uncorrupted output signal and the background noise respectively. In the simulations, σ^2_b is chosen in such a way that SBNR=10 dB or 40 dB. The length of the true system is fixed to $M = 32$. The measure of performance considered is the *system mismatch* defined as:

$$10\log_{10}\left[\frac{\|\tilde{\mathbf{w}}_i\|^2}{\|\mathbf{w}_i\|^2}\right]. \tag{30}$$

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. 1 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 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 that the filter converges in a mean-square sense.

In Fig. 2 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 hypothesis A2). Again, its limit is in the order of the machine precision.

Fig. 3 shows the probability of executing the NLMS update with $\mu = 1$. We only show the result of the first 10000 iterations to appreciate the details at the beginning of the adaptation. According to the model this probability is given by $1 - P_i[z \geq E[\delta_{i-1}]]$. 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 considered for obtaining a closed-form expression for the distribution of $\frac{e_i^2}{\|\mathbf{x}_i\|^2}$. However, as the adaptation proceeds for a long time interval, the agreement improves. This validates the previous discussion where the independence between e_i^2 and $\|\mathbf{x}_i\|^2$ was introduced. 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 have zero misalignment.

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

5. CONCLUSIONS

A new framework for designing robust adaptive filters was recently introduced. Particularly, the RVSS-NLMS algorithm was derived. Here, we introduced a theoretical model for predicting its mean-

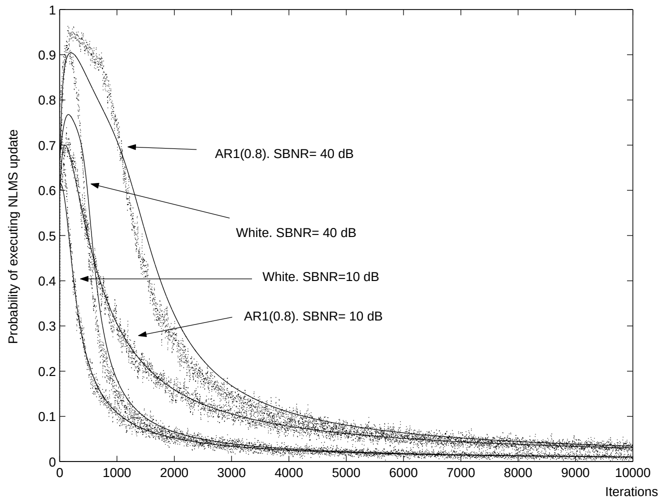

Figure 3: Evolution of the probability of executing the NLMS update. The setup is the same as in Fig. 1. The experimental plots were low pass-filtered with a moving average with a span of 7 samples.

square behavior. Under certain reasonable assumptions we proved that the limiting mean-square misalignment is zero. Then, 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.

6. APPENDICES

6.1 Proof of (19)

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

$$\sigma^2_{e_{a,\infty}} = \lim_{i\to\infty} \sum_{j=1}^{M} \lambda_j E[(c_i^j)^2], \tag{31}$$

where c_i^j denotes the j-th component of $\mathbf{c}_i$. Because of (18) 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, (19) follows.

6.2 Proof of (21) and (22)

Using the definition of $\mathbf{K}_i$ in (5) we can write:

$$\mathbf{K}_i = \mathbf{K}_{i-1} - E\left[\sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\,\mathrm{sign}(e_i)\left(\tilde{\mathbf{w}}_{i-1}\frac{\mathbf{x}_i^T}{\|\mathbf{x}_i\|} + \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}\tilde{\mathbf{w}}_{i-1}^T\right)\right] + E\left[\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}\,\frac{\mathbf{x}_i\mathbf{x}_i^T}{\|\mathbf{x}_i\|^2}\right]. \tag{32}$$

Assuming that the filter is long enough we can make the approximation

$$E\left[\mathrm{sign}(e_i)\left(\tilde{\mathbf{w}}_{i-1}\frac{\mathbf{x}_i^T}{\|\mathbf{x}_i\|} + \mathbf{x}_i\frac{\tilde{\mathbf{w}}_{i-1}^T}{\|\mathbf{x}_i\|}\right)\right] \approx E\left[\frac{1}{\|\mathbf{x}_i\|}\right]\cdot E\left[\mathrm{sign}(e_i)\left(\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T + \mathbf{x}_i\tilde{\mathbf{w}}_{i-1}^T\right)\right]. \tag{33}$$

This approximation is justified with the same argument given for (11). Using A2) and (33) in (32) yields:

$$\mathbf{K}_i = \mathbf{K}_{i-1} - E\left[\frac{1}{\|\mathbf{x}_i\|}\right]\sqrt{\frac{E\left[\delta_i - \alpha\delta_{i-1}\right]}{1-\alpha}}\,E\left[\mathrm{sign}(e_i)\cdot\left(\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T + \mathbf{x}_i\tilde{\mathbf{w}}_{i-1}^T\right)\right] + \frac{E\left[\delta_i - \alpha\delta_{i-1}\right]}{1-\alpha}E\left[\frac{\mathbf{x}_i\mathbf{x}_i^T}{\|\mathbf{x}_i\|^2}\right]. \tag{34}$$

We need to calculate $E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T\right]$. We can write:

$$E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T\right] = E\left\{E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T|\tilde{\mathbf{w}}_{i-1}\right]\right\}. \tag{35}$$

By means of B1) we can apply Lemma 1 to each entry of $E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T|\tilde{\mathbf{w}}_{i-1}\right]$. Using Price's theorem and B1), B2) and Lemma 1 it holds:

$$E\left[\mathrm{sign}(e_i)\tilde{\mathbf{w}}_{i-1}\mathbf{x}_i^T\right] = \sqrt{\frac{2}{\pi}}\mathbf{K}_{i-1}\mathbf{R}\left\{\frac{p}{\sqrt{\sigma^2_{e_{a,i}} + (K+1)\sigma_B^2}} + \frac{1-p}{\sqrt{\sigma^2_{e_{a,i}} + \sigma_B^2}}\right\}. \tag{36}$$

The term $E\left[\mathrm{sign}(e_i)\mathbf{x}_i\tilde{\mathbf{w}}_{i-1}^T|\tilde{\mathbf{w}}_{i-1}\right]$ can be obtained with the same procedure. Then (21) follows. Equation (22) is a straightforward application of the independence between the inputs regressors stated in B1).

REFERENCES

[1] L. Rey Vega, H. Rey, J. Benesty, and S. Tressens "A robust adaptive filtering algorithm against impulsive noise," in *Proc. ICASSP 2007*, Honolulu, Hawaii, USA, April 15-20, 2007.

[2] T. Gänsler, S.L. Gay, M.M. Sondhi, and J. Benesty, "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Processing*, vol. 8, no. 6, pp. 656–663, Nov. 2000.

[3] T. Gänsler and J. Benesty, "The fast normalized cross-correlation double-talk detector," *Signal Processing*, vol. 86, no. 6, pp. 1124–1139, Jun. 2006.

[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 Processing Lett.*, vol. 4, no. 2, pp. 46–48, Feb. 1997.

[6] T. Al-Naffouri and A. Sayed "Transient analysis of adaptive filters with error nonlinearities," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 653–663, Mar. 2003.

[7] S. Chan Bang, S. Ann, and I. Song "Performance analysis of the dual sign algorithm for additive contaminated-gaussian noise," *IEEE Signal Processing Letters*, vol. 1, no. 12, pp. 196–198, Dec. 1994.

[8] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 69–72, Jun. 1958.

[9] V.J. Mathews and S.H. Cho, "Improved convergence analysis of stochastic gradient adaptive filters using the sign algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, no. 4, pp. 450-454, Apr. 1987.

[10] N.J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 4, pp. 793-806, Aug. 1986.