REFERENCES

[1] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: A low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, Apr. 1996.

[2] L. Shue, S. Dey, B. Anderson, and F. De Bruyne, "On state-estimation of a two-state hidden Markov model with quantization," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 202–208, Jan. 2001.

[3] F. A. Galati and S. Dey, "Information theoretic quantiser design for decentralised estimation of hidden Markov models," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Hong Kong, Apr. 2003, vol. 6, pp. 749–752.

[4] M. Huang and S. Dey, "Dynamic quantizer design for hidden markov state estimation via multiple sensors with fusion center feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2887–2896, Aug. 2006.

[5] M. Huang and S. Dey, "Distributed state estimation for hidden Markov models by sensor networks with dynamic quantization," in *Proc. Intelligent Sensors, Sensor Networks, Information Processing Conf.*, Melbourne, Australia, Dec. 2004, pp. 355–360.

[6] M. Huang and S. Dey, "Dynamic quantization for multisensor estimation over bandlimited fading channels," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4696–4702, Sep. 2007.

[7] N. Ghasemi and S. Dey, "Power-efficient dynamic quantization for multisensor hmm state estimation over fading channels," in *Proc. IEEE Int. Symp. Communications, Control, Signal Processing (ISCCSP)*, Mar. 12–14, 2008, pp. 1553–1558.

[8] J. F. Chamberland and V. V. Veeravalli, "Decentralized dynamic power control for cellular CDMA systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 549–559, May 2003.

[9] D. V. Djonin and V. Krishnamurthy, "MIMO transmission control in fading channels-a constrained Markov decision process formulation with monotone randomized policies," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 5069–5083, Oct. 2007.

[10] A. K. Karmokar, D. V. Djonin, and V. K. Bhargava, "Optimal and suboptimal packet scheduling over correlated time varying flat fading channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 2, pp. 446–456, Feb. 2006.

[11] L. Tong, Q. Zhao, and S. Adireddy, "Sensor networks with mobile agents," in *Proc. IEEE Military Communications Conf. (MILCOM)*, Boston, MA, Oct. 2003, vol. 1, pp. 688–693.

[12] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: Wiley, 1994.

[13] J. C. Spall, "Adaptive stochastic approximation by the simultaneous perturbation method," *IEEE Trans. Autom. Control*, vol. 45, no. 10, pp. 1839–1853, Oct. 2000.

A Fast Robust Recursive Least-Squares Algorithm

Leonardo Rey Vega, Hernán Rey, Jacob Benesty, and Sara Tressens

***Abstract*—We present a fast robust recursive least-squares (FRRLS) algorithm based on a recently introduced new framework for designing robust adaptive filters. The algorithm is the result of minimizing a cost function subject to a time-dependent constraint on the norm of the filter update. Although the characteristics of the exact solution to this problem are known, there is no closed-form solution in general. However, the approximate solution we propose is very close to the optimal one. We also present some theoretical results regarding the asymptotic behavior of the algorithm. The FRRLS is then tested in different environments for system identification and acoustic echo cancellation applications.**

***Index Terms*—Acoustic echo cancellation, impulsive noise, recursive least-squares algorithm, robust filtering, system identification.**

I. INTRODUCTION

The recursive least-squares algorithm has the ability to solve the least-squares estimation problem recursively. Through its link with Kalman estimation [1], it can lead to the optimal estimate in the mean-square error sense. However, this is based on the assumption that the error signal e_i, which is by definition the difference between the system and model filter outputs, is Gaussian. In real-world environments, this assumption can be false. 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. The performance of the RLS can be significantly deteriorated in these cases.

Several algorithms have been proposed attempting to overcome this issue [2]–[5]. In this work, we use a recently introduced new framework for the construction of robust adaptive filters [6] in order to design a robust RLS algorithm. Throughout this correspondence, the term robust will be used as "slightly sensitive to large perturbations (outliers)".

Particularly, we use a universal cost function introduced in [7] that preserves the system estimate from the effect of impulsive noise (or double talk) through the memory factor in the classical RLS cost function. Then, we propose to optimize this function subject to a constraint on the norm of the adaptive filter update. However, the exact solution to this problem has no closed form. Therefore, we propose an approximate solution to the problem. This solution is actually very close to the optimal one.

Manuscript received February 25, 2008; revised November 07, 2008. First published December 09, 2008; current version published February 13, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Roberto Lopez-Valcarce. This work was supported in part by the Universidad de Buenos Aires under project UBACYT I005.

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

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

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2008.2010643

The result is a new algorithm that provides an automatic mechanism for switching between the standard RLS algorithm and another one that moves in the same direction but with a different magnitude.

We also present some theoretical results showing that the misalignment vector of the proposed robust RLS algorithm converges in the mean-square sense.

Since the RLS algorithm (and the robust version introduced here) has typically an $O(M^2)$ computational complexity, where M is the length of the adaptive filter, we introduce a fast version of it. However, practical issues regarding eventual nonstationary environments should be considered. Because of this, certain changes are made in the original algorithm, leading to the proposed *fast robust recursive least-squares* (FRRLS) algorithm. The performance of the algorithm is tested under several scenarios in system identification and acoustic echo cancellation applications.

Finally, we present certain definitions and the notation that is used throughout the correspondence. 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, $v_i = \vartheta_i + \eta_i$, where ϑ_i denotes for the background measurement noise and η_i is an impulsive noise or an undetected near-end signal in echo cancellation 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 $\hat{\mathbf{w}}_i$, an estimate of $\mathbf{w}_i$. This adaptive filter receives the same input, leading to the output filtering error $e_i = d_i - \mathbf{x}_i^T\hat{\mathbf{w}}_{i-1}$ and to the *a posteriori* error $e_{\mathrm{p},i} = d_i - \mathbf{x}_i^T\hat{\mathbf{w}}_i$. The misalignment vector is $\tilde{\mathbf{w}}_i = \mathbf{w}_i - \hat{\mathbf{w}}_i$. We also define the *a priori* error $e_{\mathrm{a},i} = \mathbf{x}^T\tilde{\mathbf{w}}_{i-1}$.

II. Recursive Least-Squares Algorithm

The recursive least-squares (RLS) algorithm is the result of the following optimization problem [8]:

$$\hat{\mathbf{w}}_i = \arg\min_{\hat{\mathbf{w}}_i\in\mathbb{R}^M}\sum_{n=1}^{i}\lambda^{i-n}\xi_n^2,\ \ \xi_n = d_n - \mathbf{x}_n^T\hat{\mathbf{w}}_i,\ \ 0<\lambda<1 \quad (1)$$

where λ is the forgetting factor. The solution is given by

$$\hat{\mathbf{w}}_i = \mathbf{\Phi}_i^{-1}\mathbf{z}_i \quad (2)$$

where

$$\mathbf{\Phi}_i = \lambda\mathbf{\Phi}_{i-1} + \mathbf{x}_i\mathbf{x}_i^T, \quad \mathbf{z}_i = \lambda\mathbf{z}_{i-1} + d_i\mathbf{x}_i \quad (3)$$

are the time-averaged correlation matrix and the time-averaged cross-correlation vector respectively. Using (3) we can write (2) as

$$\hat{\mathbf{w}}_i = \mathbf{\Phi}_{i-1}^{-1}\mathbf{z}_{i-1} + \mathbf{k}_i e_i \quad (4)$$

where

$$\mathbf{k}_i = \frac{\lambda^{-1}\mathbf{\Phi}_{i-1}^{-1}\mathbf{x}_i}{1+\lambda^{-1}\mathbf{x}_i^T\mathbf{\Phi}_{i-1}^{-1}\mathbf{x}_i} = \mathbf{\Phi}_i^{-1}\mathbf{x}_i \quad (5)$$

is the Kalman gain [8]. Let us analyze what happens when an impulsive noise sample is present at time index i. Since this affects d_i, it is clear that it will have a strong effect on $\mathbf{z}_i$. Based on (2) or the second term in (4) the estimation $\hat{\mathbf{w}}_i$ will be very poor. However, the main problem is that the effect of the impulsive noise at time index i will persist for several future time steps, severely affecting the estimation process. This is because of the memory induced by the cost function in (1), which is usually large as λ is close to 1 in typical implementations. So, the perturbation induced by an impulsive noise sample at time i will be hard to forget and will have a significant influence for almost $(1-\lambda)^{-1}$ time updates.

A possible solution to this problem was presented in [4], where an M-estimate cost function is used to find the estimate

$$\hat{\mathbf{w}}_i = \arg\min_{\hat{\mathbf{w}}_i\in\mathbb{R}^M}\sum_{n=1}^{i}\lambda^{i-n}\rho(\xi_n), \quad \xi_n = d_n - \mathbf{x}_n^T\hat{\mathbf{w}}_i \quad (6)$$

with $\rho(\cdot)$ denoting for the Hampel's three-part redescending M-estimate function [9]. Although this approach provides insensitivity to impulsive noise samples since the derivative of the Hampel's function is bounded, it has two drawbacks. First, the parameters of the Hampel's function should depend on the noise and error filtering statistics, which might not be known, and second, the optimization problem (6) has no closed-form solution. In [4] and [5], the difficulty of this optimization is not addressed. In fact, the solution provided is not correct because the time-averaged correlation matrix and cross-correlation vector derived in those works depend on $\hat{\mathbf{w}}_i$. The implicit approximation in [4] and [5] is that for a span of time approximately equal to $(1-\lambda)^{-1}$ the estimate $\hat{\mathbf{w}}_i$ does not change significantly. This might not be true, for example, at the beginning of the adaptation process or when λ is very close to one. However, it seems that the proposed solution is close to the true solution and performs well. In [4] and [5], these issues were not discussed and the authors claimed that the proposed solution is the optimal one.

In order to make the algorithm robust to impulsive noise we will propose the following: modify the cost function in order to reduce the effect of the intrinsic memory of the RLS algorithm when an impulsive noise sample is present, and include a constraint in the optimization process in order to reduce the effect of an impulsive noise sample at time i.

III. Robust Recursive Least-Squares Algorithm

In [7] it has been shown that several classical adaptive algorithms and new ones can be obtained applying a common universal criterion. This criterion is formed by the sum of two terms: one of them is the square of the distance between the old and the new filter estimates and the other depends on the *a posteriori* error signal. Therefore, according to this principle one easy way to find adaptive filters is to minimize at every time index i the following cost function:

$$J(\hat{\mathbf{w}}_i) = d_i^2(\hat{\mathbf{w}}_i, \hat{\mathbf{w}}_{i-1}) + e_{\mathrm{p},i}^2 \quad (7)$$

where $d_i(\cdot)$ is a distance function. Different choices of the distance function should reflect our knowledge about the space where the true system is and could lead to different adaptive filters. The distance functions we will consider here are given by:

$$d_i^2(\hat{\mathbf{w}}_i, \hat{\mathbf{w}}_{i-1}) = (\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1})^T\mathbf{Q}_i(\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}) \quad (8)$$

with $\mathbf{Q}_i$ given by

$$\mathbf{Q}_i = \mathbf{\Phi}_i - \mathbf{x}_i\mathbf{x}_i^T. \quad (9)$$

More general distance functions could be considered in such a way that the parameter space would be a Riemmanian space, i.e., a curved manifold where the distance properties are not uniform along the space.

The unique solution to the problem (7) is given by

$$\mathbf{h}_i = \hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1} + \mathbf{\Phi}_i^{-1}\mathbf{x}_i e_i. \quad (10)$$

If we interpret $\hat{\mathbf{w}}_{i-1}$ in (10) as $\mathbf{\Phi}_{i-1}^{-1}\mathbf{z}_{i-1}$, we obtain the standard RLS algorithm as in (4). The central aspect here is that with the proposed cost function we are not explicitly identifying $\hat{\mathbf{w}}_{i-1}$ with $\mathbf{\Phi}_{i-1}^{-1}\mathbf{z}_{i-1}$. This is an important issue. The latter has the disadvantage that the effect of a single impulsive noise sample at time step $i-1$ would be propagated for several steps through the sequence $\mathbf{z}_i$ (given that λ is close to 1). The use of (10) will still have the advantage of the decorrelating property of the term $\mathbf{\Phi}_i^{-1}\mathbf{x}_i$, which increases the convergence speed of the algorithm.

Now, we focus in finding the way to make the algorithm more robust to the effect of an impulsive noise sample through the error signal. In doing so, we propose to use an additional constraint on the optimization problem:

$$\|\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}\|^2 \leq \delta_{i-1} \tag{11}$$

where $\{\delta_i\}$ is a positive sequence. Its choice will influence the dynamics of the algorithm but in any case, (11) guarantees that any noise sample can perturb the square norm of the filter update by at most the amount δ_{i-1}. This constraint was successfully used in the past to obtain a robust version of the normalized least-mean-square (NLMS) algorithm [6]. Then, the constrained problem is

$$\begin{aligned}\hat{\mathbf{w}}_i = \arg\min_{\hat{\mathbf{w}}_i \in \mathbb{R}^M} &\left\{(\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1})^T \mathbf{Q}_i (\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}) + e_{\mathrm{p},i}^2\right\} \\ \text{s.t.} \quad &\|\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}\|^2 \leq \delta_{i-1}. \end{aligned}\tag{12}$$

If the hypersphere (11) contains (10), the latter will be the solution. We will have this situation when

$$e_i^2 \mathbf{x}_i^T \mathbf{\Phi}_i^{-2} \mathbf{x}_i \leq \delta_{i-1}. \tag{13}$$

If (13) is not satisfied, defining $\mathbf{t}_i = \hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}$ and $\mathbf{r}_i = \mathbf{h}_i - \hat{\mathbf{w}}_{i-1}$, it is possible to formulate the optimization problem as

$$\mathbf{t}_i = \arg\min_{\mathbf{t}_i \in \mathbb{R}^M} (\mathbf{t}_i - \mathbf{r}_i)^T \mathbf{\Phi}_i (\mathbf{t}_i - \mathbf{r}_i) \quad \text{s.t.} \quad \|\mathbf{t}_i\|^2 = \delta_{i-1}. \tag{14}$$

It can be shown that its solution would involve the analytical calculation of the roots of a $2M$ degree polynomial which is not possible if $M > 2$. For this reason we should look for suboptimal solutions. We propose to normalize the update in (10) to satisfy the constraint (providing robustness to the algorithm), leading to

$$\mathbf{t}_i = \sqrt{\delta_{i-1}} \frac{\mathbf{\Phi}_i^{-1}\mathbf{x}_i}{\left\|\mathbf{\Phi}_i^{-1}\mathbf{x}_i\right\|} \operatorname{sign}(e_i). \tag{15}$$

If δ_{i-1} is small (i.e., all the points on the hypersphere are close to each other), (15) should be close to the optimal solution. This will be the most common situation when (13) is not satisfied. Actually, it is easy to show that the Euclidian distance between the suboptimal and the optimal solution is bounded by $2\sqrt{\delta_{i-1}}$. It should be clear that (15) is not the only possible suboptimal solution for (14). However, we expect it to perform well since it is in the same direction than the one corresponding to the update in (10).

Combining (10), (13), and (15), and defining $\Delta\hat{\mathbf{w}}_i^{\mathrm{RLS}} = \mathbf{\Phi}_i^{-1}\mathbf{x}_i e_i$, we put the proposed algorithm in the following way:

$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1} + \min\left\{\sqrt{\delta_{i-1}}, \left\|\Delta\hat{\mathbf{w}}_i^{\mathrm{RLS}}\right\|\right\} \frac{\Delta\hat{\mathbf{w}}_i^{\mathrm{RLS}}}{\|\Delta\hat{\mathbf{w}}_i^{\mathrm{RLS}}\|}. \tag{16}$$

Simply, at time-step i, if the squared norm of the RLS update is smaller than δ_{i-1}, the RLS update is performed; if not, it is normalized to have a norm $\sqrt{\delta_{i-1}}$ and then performed. So the new algorithm has two operation modes: at the beginning, if the values of δ_i are not too small the algorithm will act as an RLS algorithm, providing fast convergence but being sufficiently robust against noise. If a large noise sample occurs, then the algorithm will act as an RLS algorithm with a "step-size" given by $\sqrt{\delta_{i-1}}/\|\mathbf{k}_i\||e_i|$, avoiding the divergence of the filter. So, we could think that this algorithm acts as an RLS algorithm with a switching "step-size" between the values 1 and $\sqrt{\delta_{i-1}}/\|\mathbf{k}_i\||e_i|$.

The only thing that remains is the choice of the delta sequence. We use the one proposed in [6], which in this case has the form

$$\delta_i = \alpha\delta_{i-1} + (1-\alpha)\min\left\{\delta_{i-1}, \|\mathbf{k}_i\|^2 e_i^2\right\} \tag{17}$$

where $0 < \alpha < 1$ is a memory factor. This selection should make δ_i dependent on the convergence dynamics of the adaptive filter allowing it (without compromising the robust behavior) to take large values at the beginning of the adaptation (increasing the speed of convergence) and lower values at the end (improving the final misadjustment). The memory factor of the RLS and the parameter α are chosen as

$$\lambda = 1 - \frac{1}{\kappa M}, \quad \alpha = 1 - \frac{1}{\kappa_\delta M} \tag{18}$$

where κ and κ_δ are positive integers, typically between 1 and 6. The delta sequence is initialized as

$$\delta_0 = E_{\mathrm{c}} \frac{\sigma_d^2}{\sigma_x^2 M}, \tag{19}$$

with σ_x^2 and σ_d^2 standing for the power of the input and observed output signals respectively, and E_{c} is an integer.

In order to show how well the proposed algorithm behaves with respect to the optimal scheme following the solution of (12), we compared them in a numerical simulation. In addition to the background noise ϑ_i (with power σ_b^2), an impulsive noise η_i is also added to the output signal y_i. This noise is generated as $\eta_i = \omega_i N_i$, where ω_i is a Bernoulli process with $P[\omega_i = 1] = p_{\mathrm{imp}}$ and N_i is a zero-mean Gaussian with power $\sigma_N^2 = 1000\sigma_y^2$, where σ_y^2 is the power of the uncorrupted output signal. Although this noise has finite variance it is useful to test the robustness of an adaptive algorithm, and it has been previously used in the literature [4], [5]. We use the *mismatch* as a measure of performance, which is defined as

$$10\log_{10}\left[\frac{\|\tilde{\mathbf{w}}_i\|^2}{\|\mathbf{w}_i\|^2}\right]. \tag{20}$$

We found that both algorithms present an almost identical mismatch (not shown). For the solution of (14) we implemented a gradient-based numerical algorithm at any time where the condition (13) is not satisfied. This supports the use of the approximate update that can be satisfactorily implemented in practice, which is not the case of the optimal solution.

IV. Mean-Square Steady-State Behavior

We are interested in the mean-square steady-state behavior of $\tilde{\mathbf{w}}_i$. 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. Assuming that the true system is stationary, $\mathbf{w}_i = \mathbf{w}_0, \forall i$ and using the definition of the misalignment vector, (16) and (17), it is easy to show that

$$\tilde{\mathbf{w}}_i = \tilde{\mathbf{w}}_{i-1} - \sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}} \frac{\mathbf{k}_i}{\|\mathbf{k}_i\|} \operatorname{sign}(e_i). \tag{21}$$

Taking the expectation of the squared norm on both sides

$$\begin{aligned} E\left[\|\tilde{\mathbf{w}}_i\|^2\right] = E\left[\|\tilde{\mathbf{w}}_{i-1}\|^2\right] - 2E&\left[\sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\right. \\ &\left.\cdot \frac{\tilde{\mathbf{w}}_{i-1}^T \mathbf{k}_i}{\|\mathbf{k}_i\|} \operatorname{sign}(e_i)\right] + E\left[\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}\right]. \end{aligned}\tag{22}$$

The parameter α is typically close to one. This means that (17) is the result of low-pass filtering $\|\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}\|^2$. Then the variance of the random variable δ_i would be small enough to assume that

$$E\left[\sqrt{\frac{\delta_i - \alpha\delta_{i-1}}{1-\alpha}}\frac{\tilde{\mathbf{w}}_{i-1}^T\mathbf{k}_i}{\|\mathbf{k}_i\|}\mathrm{sign}(e_i)\right] \approx \sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}} \cdot E\left[\frac{\tilde{\mathbf{w}}_{i-1}^T\mathbf{k}_i}{\|\mathbf{k}_i\|}\mathrm{sign}(e_i)\right], \quad (23)$$

$$E\left[\min\left\{\delta_{i-1}, \|\mathbf{k}_i\|^2 e_i^2\right\}\right] \approx E[\delta_{i-1}]P_i\left[z > E[\delta_{i-1}]\right] + \int_0^{E[\delta_{i-1}]} z dF_z^i(z) \quad (24)$$

where $z \doteq e_i^2\|\mathbf{k}_i\|^2$, i.e., z and $e_i^2\|\mathbf{k}_i\|^2$ have the same distribution. $P_i[A]$ and $F_z^i(z)$ denote the probability of the event A and the distribution function of z at time-step i respectively. This assumption on the variance of δ_i is very accurate and was successfully used and validated in [6]. Observing that $E[\|\tilde{\mathbf{w}}_i\|^2] - E[\|\tilde{\mathbf{w}}_{i-1}\|^2]$ constitutes a telescoping series, using (23), $\hat{\mathbf{w}}_0 = \mathbf{0}$, and assuming that $\lim_{i\to\infty} E[\|\tilde{\mathbf{w}}_i\|^2]$ exists,

$$\lim_{i\to\infty} E\left[\|\tilde{\mathbf{w}}_i\|^2\right] - \|\mathbf{w}_0\|^2 = \sum_{i=1}^{\infty}\left\{-2\sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}} \cdot E\left[\frac{\tilde{\mathbf{w}}_{i-1}^T\mathbf{k}_i}{\|\mathbf{k}_i\|}\mathrm{sign}(e_i)\right] + \frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}\right\}. \quad (25)$$

From Appendix A, we know that

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

This implies that we can split the series of the right hand of (25) in two. Then we should have

$$\sum_{i=1}^{\infty}\left\{\sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}} E\left[\frac{\tilde{\mathbf{w}}_{i-1}^T\mathbf{k}_i}{\|\mathbf{k}_i\|}\mathrm{sign}(e_i)\right]\right\} < \infty. \quad (27)$$

However, from Appendix A, we have

$$\sum_{i=1}^{\infty}\left\{\sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}}\right\} = \infty. \quad (28)$$

Using (28) and in order to satisfy (27), we should have

$$\lim_{i\to\infty} E\left[\frac{\tilde{\mathbf{w}}_{i-1}^T\mathbf{k}_i}{\|\mathbf{k}_i\|}\mathrm{sign}(e_i)\right] = 0 \quad (29)$$

where we assume the existence of the limit. Under the additional assumptions that the input regressors are independent and come from a spherically invariant random process (SIRP), the adaptive filter is long enough and the value of λ is very close to one, it is shown in Appendix B that (29) implies

$$\lim_{i\to\infty} E\left[\|\tilde{\mathbf{w}}_i\|^2\right] = 0. \quad (30)$$

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. This is not the case with other robust RLS approaches proposed in the literature, like [5], where the analysis is done with stronger assumptions. This result was verified through simulations. It was observed that after a sufficiently long time the norm of the misalignment vector was in the order of the machine precision.

V. Practical Considerations: Fast Version and Nonstationary Control

We can implement the proposed algorithm using an RLS algorithm to compute $\Delta\hat{\mathbf{w}}_i^{\mathrm{RLS}}$ at each time step and then apply (16) to perform the update.

Since the standard RLS algorithm presents a large computational cost, specially if the length of the adaptive filter is large, we will use a fast transversal filter (FTF) implementation [10]. The reason for using this implementation instead of a least-squares lattice implementation [11] is that the former permits the coefficients of the filter to be obtained directly, which are the desired information in applications that involve system identification. Moreover, as the coefficients of the filter are necessary to compute the value of δ_i it is clear that an FTF implementation is better suited for the proposed algorithm.

Finally, a major issue should be considered carefully. As the proposed delta sequence is decreasing, although the algorithm becomes more robust against perturbations, it also loses its tracking ability. For this reason, if there is evidence supporting the possibility of being in a nonstationary environment, an *ad hoc* control should be included. The objective is to detect changes in the true system. We use the same controls proposed in [6], although other schemes might be used. The advantage of the proposed schemes is that the parameters are not coupled to each other as in other previously proposed algorithms. Each parameter is used to deal with a specific feature of the environment. Therefore, this set of parameters allows the algorithm to work well under many different scenarios. See [6] for a detailed description of the methods and their parameters. The only difference with the methods used in [6] is that when a large change in the system is detected, in addition to re-initialize the delta sequence, we also re-initialize the parameters used in the prediction part of the fast RLS algorithm.

Table I summarizes the resulting *fast robust recursive least-squares* (FRRLS) algorithm. The prediction part is the same as in any FTF implementation. It has a computational complexity of $O(M)$. The filtering or joint-process estimation part has also a complexity of $O(M)$. The calculation of δ_i only requires two extra multiplications.

VI. 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 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 ϑ_i is added to the system output to achieve a certain SBNR, which is defined as

$$\mathrm{SBNR} = 10\log_{10}\left[\frac{\sigma_y^2}{\sigma_b^2}\right]. \quad (31)$$

In addition to the standard FRLS algorithm, we include two more schemes for comparing the performance with the proposed algorithm.

The robust FRLS algorithm is the one presented in [3]. The scale factor s is initialized with the standard deviation of the input signal and it is never allowed to go below the standard deviation of the background noise.

The other scheme is based on [5]. It uses M-estimates with a Huber function to end up in a new robust RLS filter. It has been shown that its performance is similar to the one of the RLM algorithm [4], which uses the (more complicated) Hampel's three-part redescending M-estimate function. In [5], the authors propose a fast version using a least-squares lattice scheme. However, we chose here the FTF scheme so we modified their algorithm and arrive finally to the Huber fast transversal filter (HFTF). It requires the estimation of the power of the impulse-free error signal ($\hat{\sigma}_e^2$) which is done by averaging (with a memory factor of 0.99)

TABLE I
THE FRRLS ALGORITHM

Parameters:	κ, κ_δ, $E_{\rm c}$, β, $V_{\rm T}$, ζ, ($V_{\rm D}$) or (V_θ, C_1, τ)
Initialization:	$\hat{\mathbf{w}}_0 = \mathbf{f}_0 = \mathbf{b}_0 = \mathbf{k}'_0 = \mathbf{0}$, $\lambda = 1 - 1/(\kappa M)$, δ_0, $l_c = 0$
	$\alpha = 1 - 1/(\kappa_\delta M)$, $\phi_0 = \lambda$, $E_0 = \sigma_x^2$, $E_{{\rm f},0} = ME_0/E_{\rm c}$
	$E_{{\rm b},0} = E_{{\rm f},0}\lambda^{-M}$, $E_{{\rm e},0} = 10ME_0$, $\lambda_{\rm e} = 1 - 1/M$
Prediction:	if $l_c = 0$
	$x_{{\rm b},i} = \mathbf{x}_{i,M}$
	else
	$x_{{\rm b},i} = 0$, $l_c = l_c - 1$
	$E_i = \lambda E_{i-1} + (1-\lambda)x_i^2$
	$e_{{\rm f},i} = x_i - \mathbf{f}_{i-1}^T \mathbf{x}_{i-1}$
	$E_{{\rm e},i} = \lambda_{\rm e} E_{{\rm e},i-1} + e_{{\rm f},i} x_i$
	$\tilde{\phi}_i = \phi_{i-1} + e_{{\rm f},i}^2 / E_{{\rm f},i-1}$
	$\begin{bmatrix} \mathbf{g}_i \\ m_i \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{k}'_{i-1} \end{bmatrix} + \begin{bmatrix} 1 \\ -\mathbf{f}_{i-1} \end{bmatrix} e_{{\rm f},i}/E_{{\rm f},i-1}$
	$\mathbf{f}_i = \mathbf{f}_{i-1} + \mathbf{k}'_{i-1} e_{{\rm f},i}/\phi_{i-1}$
	$E_{{\rm f},i} = (E_{{\rm f},i-1} + e_{{\rm f},i}^2/\phi_{i-1})\lambda$
	$\mathbf{k}'_i = \mathbf{g}_i + \mathbf{b}_{i-1} m_i$
	$e_{{\rm b},i} = \beta(x_{{\rm b},i} - \mathbf{b}_{i-1}\mathbf{x}_i) + (1-\beta)(E_{{\rm b},i-1} m_i)$
	$\phi_i = \tilde{\phi}_i - e_{{\rm b},i} m_i$
	$\gamma_i = \lambda/\phi_i$
	$\mathbf{b}_i = \mathbf{b}_{i-1} + \mathbf{k}'_i e_{{\rm b},i}/\phi_i$
	$E_{{\rm b},i} = (E_{{\rm b},i-1} + e_{{\rm b},i}^2/\phi_i)\lambda$
	if $(E_{{\rm e},i} < 0) \mid (\gamma_i \le 0) \mid (\gamma_i > 1)$
Re-initialization:	$l_c = M$, re-initialize $\mathbf{f}$, $\mathbf{b}$, $\mathbf{k}'$, ϕ, $E_{\rm f}$, $E_{\rm b}$, $E_{\rm e}$
Filtering:	$e_i = d_i - \mathbf{x}_i^T \hat{\mathbf{w}}_{i-1}$
	$\Delta\hat{\mathbf{w}}_i^{\rm RLS} = e_i \mathbf{k}'_i/\phi_i$
	$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1} + \min\left\{\sqrt{\delta_{i-1}}, \lVert\Delta\hat{\mathbf{w}}_i^{\rm RLS}\rVert\right\} \frac{\Delta\hat{\mathbf{w}}_i^{\rm RLS}}{\lVert\Delta\hat{\mathbf{w}}_i^{\rm RLS}\rVert}$
Delta sequence:	$\delta_i = \alpha\delta_{i-1} + (1-\alpha)\lVert\hat{\mathbf{w}}_i - \hat{\mathbf{w}}_{i-1}\rVert^2$
NS control 1:	if $\mathrm{mod}(i, V_{\rm T}) = 0$
(system	$\mathbf{M} = \mathrm{diag}(\mathbf{1}_{V_{\rm T}-V_{\rm D}}, \mathbf{0}_{V_{\rm D}})$
identification)	$\mathbf{c}^T = \mathcal{O}(\lvert e_i\rvert/\lVert\mathbf{x}_i\rVert, \ldots, \lvert e_{i-V_{\rm T}+1}\rvert/\lVert\mathbf{x}_{i-V_{\rm T}+1}\rVert)^\dagger$
	$ctrl_{\rm new} = \mathbf{c}^T\mathbf{M}\mathbf{c}/(V_{\rm T} - V_{\rm D})$
	$\Delta_i = (ctrl_{\rm new} - ctrl_{\rm old})/\delta_{i-1}$
	if $\Delta_i > \zeta$
	$\delta_i = \delta_0$, Re-initialization
	elseif $ctrl_{\rm new} > ctrl_{\rm old}$
	$\delta_i = \delta_{i-1} + (ctrl_{\rm new} - ctrl_{\rm old})$
	else
	Delta sequence
	$ctrl_{\rm old} = ctrl_{\rm new}$
NS control 2:	if (double-talk is not declared)
(acoustic echo	$r_i = e_i/\lVert\mathbf{x}_i\rVert$
cancellation)	else
	$r_i = 0$
	$q_i = \mathrm{median}[r_i, \ldots, r_{i-V_\theta+1}]$
	$\theta_i = \tau\theta_{i-1} + C_1(1-\tau)q_i^2$
	Then, same as control 1 but with:
	$ctrl_{\rm new} = \mathrm{mean}(\theta_i, \ldots, \theta_{i-V_{\rm T}+1})$

$^\dagger\mathcal{O}(\cdot)$ is the ascending order operator

the median of the error sequence over a sliding window of length $N_{\rm w}$. Then, it performs the FRLS update if and only if $|e_i| < 2.576\hat{\sigma}_e$; otherwise, the filter estimate is not updated.

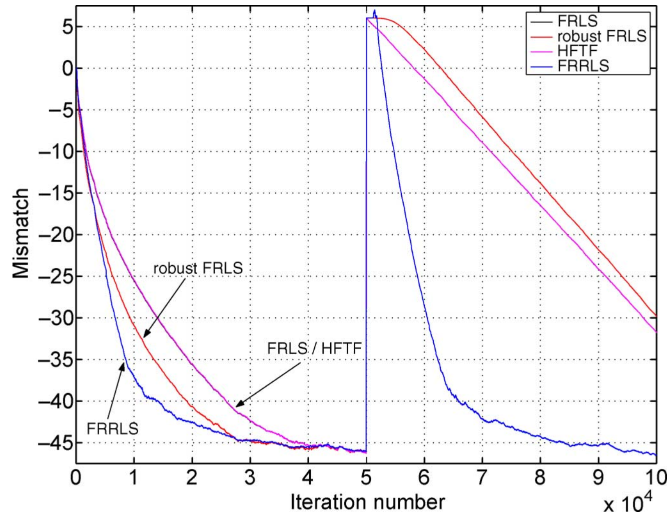

Fig. 1. Mismatch (in dB). AR1(0.95) input. $\mathrm{SBNR} = 40$ dB. No impulsive noise. $\kappa_{\rm FRRLS} = 5$. $\kappa_{\rm robust\ FRLS} = 25$. $\kappa_{\rm FRLS} = \kappa_{\rm HFTF} = 22$. $\kappa_\delta = \kappa s = 2$. $E_{\rm c} = 10$. $\beta = 0.5$. $V_{\rm T} = 2$ M. $V_{\rm D} = 0.75V_{\rm T}$. $\zeta = 20$. $N_{\rm w} = 5$. $\mathcal{M} = 38107$. $\mathcal{R} = 50768$.

The memory factor for the delta and scale factor sequences is chosen from (18), with κ_δ and κ_s respectively.

We also want to test the nonstationary controls. As a measure of their performance, we compute for each simulation

$$\mathcal{M} = \max_{i:\mathrm{mod}(i,V_T)=0} \Delta_i \quad \text{and} \quad \mathcal{R} = \frac{\mathcal{M}}{\mathcal{N}} \tag{32}$$

where $\mathcal{N}$ is the second largest value of Δ_i. In every simulation (except the one in Fig. 2) 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 ζ 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 a highly correlated AR1 with pole at 0.95. The nonstationary control 1 is used in this application. In addition to the background noise ϑ_i, an impulsive noise η_i as the one included in Section III, could also be added to the output signal y_i. In Figs. 1 and 2, the performance of the FRLS and HFTF algorithms is the same.

The tradeoff between good tracking and low steady-state mismatch is well known. In Fig. 1, we modified the memory factors (associated with the cost function) of the other algorithms so they can reach the same steady-state performance as the one of the FRRLS. The initial convergence of the other algorithms is slowed down but more importantly, their tracking performance is severely compromised.

In Fig. 2, we test the low SBNR case. The FRRLS shows the same initial convergence as the other algorithms but with 7 or 8 dB less steady-state mismatch. There is also an interesting effect that can be observed. Due to numerical error accumulation, the FTF implementation can present unstable behavior. The use of a rescue variable allows the re-initialization of the prediction part when an instability is taking place [10]. However, some residuals from the instability can affect the system update as shown in Fig. 2. But this is not the case for the proposed algorithm. In fact, at that time step its rescue variable was also out of range and the re-initialization took place. The reason for not seeing the same effect in the mismatch is due to the restriction on the filter update. As $\lVert\Delta\hat{\mathbf{w}}_i^{\rm RLS}\rVert^2$ becomes larger than δ_i during the instability, the

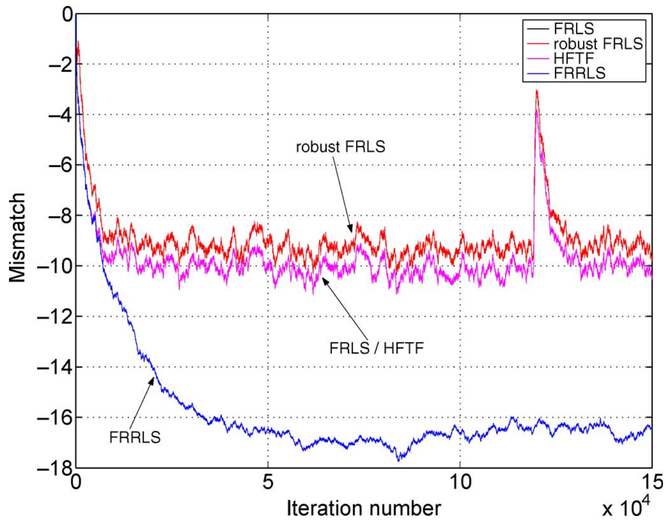

Fig. 2. Mismatch (in dB). $\mathrm{SBNR} = 10$ dB. $E_c = 50$. The other parameters are the same as in Fig. 1. $\mathcal{M} = 0.84$. $\mathcal{R} = 1.27$.

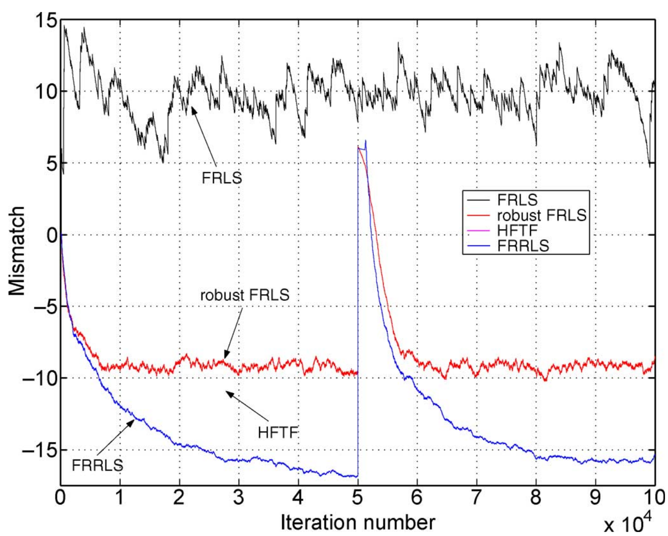

Fig. 3. Mismatch (in dB). $\mathrm{SBNR} = 10$ dB. $p_{\mathrm{imp}} = 0.01$. The other parameters are the same as in Fig. 1. $\mathcal{M} = 42$. $\mathcal{R} = 59$.

normalized update is performed instead of the one of the FRLS, preventing the increase of the mismatch. In this sense, the robust behavior of the algorithm also covers (at least to some extent) the numerical errors, leading to a more stable implementation.

In Fig. 3, impulsive noise is added. The FRLS cannot perform well under this scenario while the other algorithms present a similar performance to the case with no impulsive noise.

B. Acoustic Echo Cancellation With Double-Talk Situations

In echo cancellation applications, a double-talk detector (DTD) is used to suppress adaptation during periods of simultaneous far- and near-end signals. We use the simple Geigel DTD [12]. The Geigel DTD declares double-talk if

$$\frac{\max\left(|x_i|, |x_{i-1}|, \ldots, |x_{i-D+1}|\right)}{|d_i|} < T \tag{33}$$

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. An

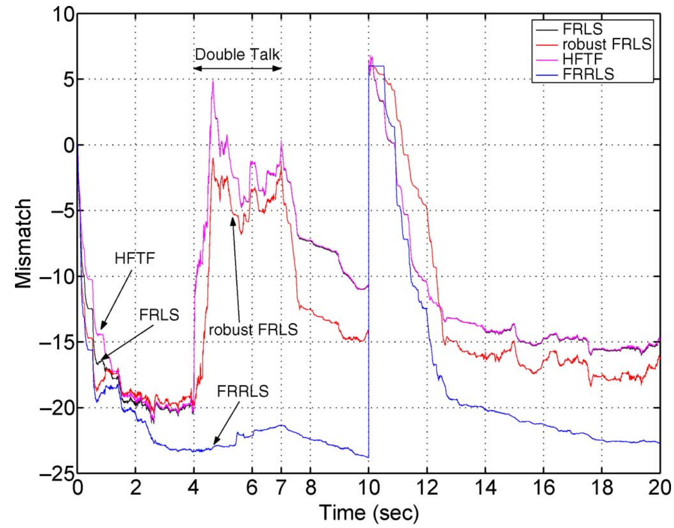

Fig. 4. Mismatch (in dB) for speech input. $\mathrm{SBNR} = 25$ dB. $\mathrm{STNR} = 0$ dB. $D = M$. $T = 1.25$. $\kappa = 8$. $\kappa_\delta = 3$. $\kappa s = 3$. $E_c = 10$. $\beta = 0.5$. $\zeta = 20$. $C_1 = 5$. $V_T = 2M$. $V_\theta = 7M$. $\tau = 0.95$. $N_w = 3$. $\mathcal{M} = 509$. $\mathcal{R} = 202$.

important detail is that except on the proposed FRRLS, the filter update is not performed when double talk is detected. Then, the FRRLS will not suffer from the undesirable effect of the false alarms but will be more vulnerable (in principle) to the presence of double talk. However, the double talk detections are considered in the nonstationary control 2 used under this application. In the robust FRLS, the scale factor is not updated when double talk is detected. In the HFTF, the estimation of $\hat{\sigma}_e^2$ is done over the last N_w samples of the error sequence where double talk was not declared.

The far-end and near-end signals are speech sampled at 8 kHz, and they were both used previously in [6]. The SBNR is 25 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. Under these conditions, the DTD detected only 20% of the near-end signal which causes long bursts of impulsive noise. Although this might seem a small percentage of detection, the remaining nondetected samples are small enough not to disturb the filter estimate and the nonstationary control 2. The proportion of false alarms when no double-talk was present was 1.5%. After passing through the DTD, the power of the near-end signal was reduced about 2.6 times.

For the HFTF we set $N_w = 3$ because larger values slow down the initial convergence (the number of RLS updates decreases) and do not increase significantly the robust behavior of the algorithm.

In Fig. 4, all the algorithms have a similar initial convergence with a lower steady-state mismatch for the FRRLS. When double talk appears, with the exception of the proposed algorithm, the mismatch grows significantly. The FRRLS can deal with the double talk through the restriction on the filter update. After the sudden change, the FRRLS recovers faster and better than the other algorithms.

VII. Conclusion

In this work, we derived a new robust version of the RLS algorithm and proposed a fast implementation, leading to the FRRLS algorithm. It is based on the novel framework introduced in [6]. It follows from optimizing a certain cost function subject to a time-dependent constraint ($\{\delta_i\}$) on the norm of the filter update. Although the exact solution to the optimization problem does not have a closed-form, we proposed a practical approximation and it is found to be very close to the optimal one and more important it is easy to implement. Then, we proposed certain dynamics for $\{\delta_i\}$. These dynamics provide the algorithm with fast initial convergence as the standard FRLS but also a robust performance against noise (impulsive, numerical error accumulation and double-talk situations). We also presented theoretical results for the convergence in the mean-square of the misalignment vector, which are valid for a large variety of input processes and noise distributions. The simulations presented provide evidence of the good performance behavior and robustness of the proposed algorithm.

APPENDIX A
BEHAVIOR OF $E[\delta_i]$

It is interesting to see from (17) that δ_i is a positive nonincreasing sequence. This means that its limit exists for every realization. For the same reason, the limit of $E[\delta_i]$ exists. Assuming the existence of the limit of the distribution functions $F_i(z)$ and $z \doteq e_i^2\|\mathbf{k}_i\|^2$, and using the result in [6], we have

$$\lim_{i\to\infty} E[\delta_i] = 0. \tag{34}$$

Using (24), we can write

$$E[\delta_i] = \alpha E[\delta_{i-1}] + (1-\alpha)\left\{ E[\delta_{i-1}]P_i\left[z > E[\delta_{i-1}]\right] + \int_0^{E[\delta_{i-1}]} z p_z^i(z) \right\} \tag{35}$$

where we assumed the existence of the PDF of z, $dF_z^i(z) = p_z^i(z)dz$, $\forall i$. Using the results of [6] it is straightforward to put (35) as

$$E[\delta_i] = E[\delta_{i-1}] - (1-\alpha)\int_0^{E[\delta_{i-1}]} F_z^i(z)dz. \tag{36}$$

Defining $p_{s,u}^i(s,u)$ as the joint PDF of $s \doteq e_i$ and $u \doteq \|\mathbf{k}_i\|^2$ 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 $z \doteq e_i^2\|\mathbf{k}_i\|^2$

$$p_z^i(z) = z^{-1/2}g^i(z),\; g^i(z) = \int_0^\infty u^{-1/2}p_{s,u}^i(\sqrt{z/u},u)du. \tag{37}$$

We will write $p_{s,u}^i(\sqrt{z/u},u)$ as

$$p_{s,u}^i(\sqrt{z/u},u) = p_{s|u}^i(\sqrt{z/u}|u)p_u^i(u). \tag{38}$$

Before continuing we will make some assumptions about the PDF of the noise, $p_v(v)$, $p_{s|u}^i(\sqrt{z/u}|u)$ and $p_u^i(u)$:

A1) $\exists B > 0$ in such a way that the PDF of the noise $p_v(v) \le B$, $\forall v$;

A2) $\exists \gamma > 0$ in such a way that $p_u^i(u)$ is continuous in $[0,\gamma]$, $\forall i$;

A3) $\exists s_0, u_0', u_0'', m$ all strictly greater than zero, such that $\forall s \in [0,s_0]$, $\forall u \in [u_0', u_0'']$, $p_{s|u}^i(s|u) > m > 0$, $\forall i$ and

$$\int_{u_0'}^{u_0''} u^{-1/2}p_u^i(u)du > K' > 0, \quad \forall i. \tag{39}$$

Assumptions A1), A2), A3) are not too strong. A1) is asking for a bounded noise PDF. This is not very restrictive because this assumption is fulfilled for several important noise distributions. Asking for A2) is a very mild condition which assures that $E[u^{-1/2}] = K_i < \infty$. This is valid $\forall i$ and with expectation taken with respect to $p_u^i(u)$. A3) asks basically for a lower bound strictly greater than zero for $p_{s|u}^i(s|u)$, valid in given intervals of s and u and $\forall i$. With these assumptions we could obtain some insights into the asymptotic behavior of $E[\delta_i]$. These assumptions are very similar to the ones in [6]. In that paper we have that z is the ratio of two positive random variables and here z is the product of two positive random variables. The little differences between the assumptions in this correspondence and the ones in [6] take that fact into consideration. Assuming also that $\exists K = \sup_i K_i$ and that $K < \infty$, we have the following theorem.

Theorem 1: Given (36) and using assumptions A1), A2), and A3), we can bound $E[\delta_i]$:

$$E[\delta_{i-1}] - \frac{4}{3}(1-\alpha)B\left(E[\delta_{i-1}]\right)^{3/2} \le E[\delta_i] \le E[\delta_{i-1}] - \frac{4}{3}(1-\alpha)C\left(E[\delta_{i-1}]\right)^{3/2} \tag{40}$$

for $i \ge i_0$, where i_0 is such that $E[\delta_{i_0-1}] \le z_0$ where z_0 is a positive number to be defined. The constants B and C are positive numbers.

The proof of this theorem is very similar to the one of Theorem 1 in [6], and for that reason we do not include it in this correspondence. From this point, and in the same way as in [6] we could prove that for large i

$$\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha} \sim \frac{1}{i^2}, \quad \sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}} \sim \frac{1}{i} \tag{41}$$

which imply

$$\sum_{i=0}^\infty \frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha} < \infty, \quad \sum_{i=0}^\infty \sqrt{\frac{E[\delta_i] - \alpha E[\delta_{i-1}]}{1-\alpha}} = \infty. \tag{42}$$

APPENDIX B
PROOF OF $\lim_{i\to\infty} E[\|\tilde{\mathbf{w}}_i\|^2] = 0$

We have $\mathbf{k}_i = \mathbf{\Phi}_i^{-1}\mathbf{x}_i$, where $\mathbf{\Phi}_i = \lambda\mathbf{\Phi}_{i-1} + \mathbf{x}_i\mathbf{x}_i^T$. We will consider $\hat{\mathbf{\Phi}}_i = (1-\lambda)\mathbf{\Phi}_i$ instead of $\mathbf{\Phi}_i$. Using this we define $\hat{\mathbf{k}}_i = \hat{\mathbf{\Phi}}_i^{-1}\mathbf{x}_i$. Obviously, we have

$$E\left[\frac{\tilde{\mathbf{w}}_{i-1}^T\hat{\mathbf{k}}_i}{\|\hat{\mathbf{k}}_i\|}\operatorname{sign}(e_i)\right] = E\left[\frac{\tilde{\mathbf{w}}_{i-1}^T\mathbf{k}_i}{\|\mathbf{k}_i\|}\operatorname{sign}(e_i)\right]. \tag{43}$$

The rationale for doing this is that $\hat{\mathbf{\Phi}}_i$ satisfies

$$\lim_{i\to\infty} E[\mathbf{\Phi}_i] = \mathbf{R}, \quad \lim_{i\to\infty} E\left[\|\hat{\mathbf{\Phi}}_i - \mathbf{R}\|_F^2\right] \le C\frac{1-\lambda}{1+\lambda} \tag{44}$$

if the input regressors are independent and the fourth moment of the input exists, i.e., $E[x_i^4] < \infty$. In (44), $\mathbf{R} = E[\mathbf{x}_i\mathbf{x}_i^T]$ is the correlation matrix of the input regressors, $\|\cdot\|_F$ denotes the Frobenius norm of a matrix and C is positive constant. We see that after a sufficiently long time and if λ is close to one, the variance in estimating $\mathbf{R}$ through $\hat{\mathbf{\Phi}}_i$ will be very small. Then, we will make the following approximation:

$$\hat{\mathbf{k}}_i = \hat{\mathbf{\Phi}}_i^{-1}\mathbf{x}_i \approx \mathbf{R}^{-1}\mathbf{x}_i. \tag{45}$$

Now consider:

$$E\left[\frac{1}{\|\hat{\mathbf{k}}_i\|^2}\right] = E\left[\frac{1}{\mathbf{x}_i^T\mathbf{R}^{-2}\mathbf{x}_i}\right]. \tag{46}$$

It could be very difficult to obtain a closed-form expression for (46). However, it is possible to obtain an idea of the behavior of this quantity with M. In a great number of cases of practical interest it can be shown that this quantity decreases at least as $1/M$, which implies that the variance of the quantity $1/\|\hat{\mathbf{k}}_i\|$ decreases at least as $1/M$. So, if M is large, as in many important applications of adaptive filters, the variance of that term will be small. In the class of processes that present this behavior are the spherically invariant random process (SIRP) [13].

SIRP are very important random processes that have been shown relevant to wireless applications [14] and speech processes [15]. In [13] it was shown that the M-dimensional PDF for that class of processes could be written

$$p_{\mathbf{x}}(\mathbf{x}_i)=|2\pi\mathbf{A}|^{-1/2}\int_0^{\infty}\frac{1}{w^M}\exp\left(-\frac{1}{2}\mathbf{x}_i^T(w^2\mathbf{A})^{-1}\mathbf{x}_i\right)dF_w(w) \tag{47}$$

where $\mathbf{A}$ is a positive definite symmetric matrix and $F_w(w)$ is, in principle, an arbitrary probability distribution function in $[0,\infty)$. Then, we have the following result, whose proof is not included for lack of space.

Theorem 2: Given a SIRP with M-dimensional correlation matrix $\mathbf{R}$ and with $\int w^2 dF_w(w)=a_0<\infty$ and $\int w^{-2}dF_w(w)=b_0<\infty$, we have

$$E\left[\frac{1}{\mathbf{x}_i^T\mathbf{R}^{-2}\mathbf{x}_i}\right]\leq\frac{\lambda_{\max}^M a_0 b_0}{M-2} \tag{48}$$

where $\lambda_{\max}^M$ is the maximum eigenvalue of the correlation matrix $\mathbf{R}$.

Assume that $\lambda_{\max}^M$ remains bounded as $M\to\infty$. This is the case if the input is stationary and its power spectral density (PSD) is bounded, $\max_{\omega\in[-\pi,\pi]}S_{xx}(\omega)<\infty$, with the maximum eigenvalue of the correlation matrix being bounded by the maximum of the PSD $\forall\, M$. So, it is clear that the variance of $1/\|\hat{\mathbf{k}}_i\|$ decreases at least as $1/M$ and will be very small for large M. In this way, if the filter is very long, we could make the following approximation:

$$E\left[\frac{\tilde{\mathbf{w}}_{i-1}^T\hat{\mathbf{k}}_i}{\|\hat{\mathbf{k}}_i\|}\mathrm{sign}(e_i)\right]\approx rE\left[\tilde{\mathbf{w}}_{i-1}^T\hat{\mathbf{k}}_i\mathrm{sign}(e_i)\right],\ r=E\left[\frac{1}{\|\hat{\mathbf{k}}_i\|}\right]. \tag{49}$$

In [16], it was pointed out that if the length of the filter is long enough and if certain mixing conditions on the input are satisfied, then by using central limits arguments it can be considered that $e_{\mathrm{a},i}$ and $e_{\mathrm{a},i}^{\mathbf{\Sigma}}=\tilde{\mathbf{w}}_{i-1}^T\mathbf{\Sigma}\mathbf{x}_i$ are zero-mean Gaussian variables, for every constant matrix $\mathbf{\Sigma}$. Defining $e_{\mathrm{a},i}^{\mathbf{R^{-1}}}=\tilde{\mathbf{w}}_{i-1}^T\hat{\mathbf{k}}_i=\tilde{\mathbf{w}}_{i-1}^T\mathbf{R}^{-1}\mathbf{x}_i$ and using Price's theorem [17], we can write

$$E\left[\tilde{\mathbf{w}}_{i-1}^T\hat{\mathbf{k}}_i\mathrm{sign}(e_i)\right]=E\left[e_{\mathrm{a},i}^{\mathbf{R^{-1}}}e_{\mathrm{a},i}\right]\frac{E\left[e_{\mathrm{a},i}\mathrm{sign}(e_{\mathrm{a},i}+v_i)\right]}{\sigma_{e_{\mathrm{a},i}}^2} \tag{50}$$

where $\sigma_{e_{\mathrm{a},i}}^2\doteq E[e_{\mathrm{a},i}^2]$. It is easy to show that

$$\frac{E\left[e_{\mathrm{a},i}\mathrm{sign}(e_{\mathrm{a},i}+v_i)\right]}{\sigma_{e_{\mathrm{a},i}}^2}=2E\left[\frac{1}{\sqrt{2\pi}\sigma_{e_{\mathrm{a},i}}}e^{-\frac{v^2}{2\sigma_{e_{\mathrm{a},i}}^2}}\right] \tag{51}$$

where the expectation is taken with respect to the noise distribution. It remains to compute the term $E[e_{\mathrm{a},i}^{\mathbf{R^{-1}}}e_{\mathrm{a},i}]$. Assuming that the input regressors are independent

$$E\left[e_{\mathrm{a},i}^{\mathbf{R^{-1}}}e_{\mathrm{a},i}\right]=E\left[\|\tilde{\mathbf{w}}_{i-1}\|^2\right]. \tag{52}$$

Putting all these results together

$$E\left[\frac{\tilde{\mathbf{w}}_{i-1}^T\hat{\mathbf{k}}_i}{\|\hat{\mathbf{k}}_i\|}\mathrm{sign}(e_i)\right]\approx 2rE\left[\|\tilde{\mathbf{w}}_{i-1}\|^2\right]E\left[\frac{1}{\sqrt{2\pi}\sigma_{e_{\mathrm{a},i}}}e^{-\frac{v^2}{2\sigma_{e_{\mathrm{a},i}}^2}}\right]. \tag{53}$$

It is important to note that for all the noise distributions of interest the term in (51) is greater than zero for every value of $\sigma_{e_{\mathrm{a},i}}^2$. In fact, for $\sigma_{e_{\mathrm{a},i}}^2=0$ that term reduces to the value of the PDF of the noise at the origin, i.e., $p_v(0)$, if it is continuous at that point. This means that the term in the right vanishes as $i\to\infty$ only if $\lim_{i\to\infty}E[\|\tilde{\mathbf{w}}_i\|^2]=0$, which also implies $\lim_{i\to\infty}\sigma_{e_{\mathrm{a},i}}^2=0$. We should mention that if the independence assumption between the input regressors is not true, (52) would be almost fulfilled when $i\to\infty$ if the variance of the term $\|\tilde{\mathbf{w}}_i\|^2$, although not zero, is very small compared with the variance of $\mathbf{x}_i^T\mathbf{R}^{-1}\mathbf{x}_i$. This is very reasonable for an adaptive filter where the norm of the updates is very small. Remember that the norm of the updates in this algorithm, for large i, is basically equal to $\sqrt{\delta_i}$ which decreases towards zero. This is some kind of a small "step-size" argument. It is interesting to note that this result, which is valid under the considered assumptions, does not depend on the existence of moments of any order of the noise distribution. This means that this result would be valid for the important case where the noise comes from an α-stable distribution.

REFERENCES

[1] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Process. Mag.*, vol. 11, no. 3, pp. 18–60, 1994.
[2] P. Petrus, "Robust Huber adaptive filter," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 1129–1133, Apr. 1999.
[3] J. Benesty and T. Gaensler, "A robust fast recursive least squares adaptive algorithm," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Salt Lake City, UT, 2001, pp. 3785–3788.
[4] Y. Zou, S. C. Chan, and T. S. Ng, "A recursive least M-estimate (RLM) adaptive filter for robust filtering in impulse noise," *IEEE Signal Process. Lett.*, vol. 7, no. 11, pp. 324–326, Nov. 2000.
[5] S.-C. Chan and Y.-X. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulsive noise: Fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975–991, Apr. 2004.
[6] L. Rey Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1878–1893, May 2008.
[7] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Berlin, Germany: Springer-Verlag, 2006.
[8] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2001.
[9] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.
[10] J. M. Cioffi and T. Kailath, "Fast recursive-least-squares transversal filters for adaptive filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-32, pp. 304–337, Apr. 1984.
[11] D. T. Lee, M. Morf, and B. Friedlander, "Recursive least-squares ladder estimation algorithm," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, pp. 627–641, Jun. 1981.
[12] D. L. Duttweiler, "A twelve channel digital echo canceler," *IEEE Trans. Commun.*, vol. 26, no. 5, pp. 647–653, May 1978.
[13] K. Yao, "A representation theorem and its applications to spherically-invariant random processes," *IEEE Trans. Trans. Inf. Theory*, vol. 19, no. 5, pp. 600–608, Sep. 1973.
[14] K. Yao, M. Simon, and E. Biglieri, "A unified theory on wireless communication fading statistics based on SIRP," presented at the SPAWC 2004, Lisboa, Portugal, Jul. 2004.
[15] H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Process.*, vol. 12, pp. 119–141, Mar. 1987.
[16] 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.
[17] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Trans. Inf. Theory*, vol. IT-4, pp. 69–72, Jun. 1958.