Signal Processing 90 (2010) 2806–2810

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

A robust variable step-size affine projection algorithm

Leonardo Rey Vega [a,*], Hernan Rey [a], Jacob Benesty [b]

[a] Facultad de Ingeniería, Universidad de Buenos Aires and CONICET, Buenos Aires, Argentina
[b] INRS-EMT, Université du Québec, Montréal, Québec, Canada H5A 1K6

ARTICLE INFO

Article history:
Received 6 January 2010
Received in revised form
25 March 2010
Accepted 26 March 2010
Available online 1 April 2010

Keywords:
System identification
Affine projection algorithm
Impulsive noise
Robust filtering
Acoustic echo cancellation

ABSTRACT

We present a robust variable step-size affine projection algorithm (RVSS-APA) using a recently introduced new framework for designing robust adaptive filters. The algorithm is the result of minimizing the square norm of the *a posteriori* error vector subject to a time-dependent constraint on the norm of the filter update. The RVSS-APA is then successfully tested in different environments for system identification and acoustic echo cancellation applications.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The affine projection algorithm (APA) is an adaptive scheme that estimates an unknown system based on multiple input vectors [1]. In real-world adaptive filtering applications, 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 [2–4]. 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. Recently, a new framework for the construction of robust adaptive filters was introduced [5]. Throughout this paper, the term robust represents "slightly sensitive to large perturbations (outliers)". The main idea is to find the filter estimate as the result of the minimization of the square norm of the *a posteriori* error vector subject to a constraint on the norm of the adaptive filter update.

Finally, we present certain definitions and the notation used throughout the paper. Let $\mathbf{w}_i = (w_{i,0}\ w_{i,1}\ \ldots\ w_{i,M-1})^T$ be an unknown linear finite-impulse response system. The 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 stands 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 *a priori* error $e_i = d_i - \mathbf{x}_i^T \hat{\mathbf{w}}_{i-1}$. When data blocks are used, we can define the data matrix $\mathbf{X}_i = [\mathbf{x}_i\ \mathbf{x}_{i-1}\ \ldots\ \mathbf{x}_{i-K+1}]$, the desired output data vector $\mathbf{d}_i = [d_i\ d_{i-1}\ \ldots\ d_{i-K+1}]^T$, the *a priori* error vector $\mathbf{e}_i = \mathbf{d}_i - \mathbf{X}_i^T \hat{\mathbf{w}}_{i-1}$ and the *a posteriori* error vector $\mathbf{e}_{p,i} = \mathbf{d}_i - \mathbf{X}_i^T \hat{\mathbf{w}}_i$.

* Corresponding author.
E-mail addresses: lrey@fi.uba.ar (L. Rey Vega), hrey@fi.uba.ar (H. Rey), benesty@emt.inrs.ca (J. Benesty).

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.03.029

2. The robust affine projection algorithm

We propose to design a new adaptive filter using the framework introduced in [5]. In order to avoid a degradation of the system performance when a large noise sample is present, the energy of the filter update is constrained at each iteration. This can be formally stated as

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

It is clear that independently of the actual values of the input and the noise, condition (1) constrains $\hat{\mathbf{w}}_i$ to be inside the hypersphere centered in $\hat{\mathbf{w}}_{i-1}$ and radius $\sqrt{\delta_{i-1}}$. The choice of the positive sequence $\{\delta_i\}$ will influence the dynamics of the algorithm but in any case, (1) 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. In order to include the effect of past input regressors in the optimization problem we find the filter estimate as

$$\hat{\mathbf{w}}_i = \arg\min_{\hat{\mathbf{w}}_i\in\mathbb{R}^M} \|\mathbf{e}_{\mathrm{p},i}\|^2 \quad \text{s.t. } \|\hat{\mathbf{w}}_i-\hat{\mathbf{w}}_{i-1}\|^2 \le \delta_{i-1}. \tag{2}$$

Let

$$\mathbf{a}_i = (\mathbf{X}_i^T)^\dagger \mathbf{d}_i+\mathbf{b}_i = \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{d}_i+\mathbf{b}_i, \quad \mathbf{b}_i\in \mathrm{Ker}(\mathbf{X}_i^T), \tag{3}$$

where $(\cdot)^\dagger$ denote Moore–Penrose pseudoinverse and $\mathrm{Ker}(\cdot)$ denote the kernel of a given linear transformation. Using (3), $\|\mathbf{e}_{\mathrm{p},i}\|^2$ can be written as

$$\|\mathbf{e}_{\mathrm{p},i}\|^2 = (\hat{\mathbf{w}}_i-\mathbf{a}_i)^T\mathbf{X}_i\mathbf{X}_i^T(\hat{\mathbf{w}}_i-\mathbf{a}_i), \tag{4}$$

where $\mathbf{a}_i$ is any vector given by (3). Given that $\mathrm{rank}(\mathbf{X}_i\mathbf{X}_i^T)=K<M$, (4) defines a degenerate hyperellipsoid and $\mathbf{a}_i$ can be seen as any point in its *axis* (for different choices of $\mathbf{b}_i\in\mathrm{Ker}(\mathbf{X}_i^T)$). When $K>2$, this axis could be a general affine space of dimension $M-K$. All points in this axis are given by (3) which coincide with all values of $\hat{\mathbf{w}}_i$ giving $\mathbf{e}_{\mathrm{p},i}=\mathbf{0}$. For the optimization problem in (2) we have to analyze two situations:

Case 1: The hypersphere (1) intersects the affine space $\mathbf{e}_{\mathrm{p},i}=\mathbf{0}$.

The minimum value of $\|\mathbf{e}_{\mathrm{p},i}\|^2$ is zero. We see that there are infinite solutions lying on the axis of the hyperellipsoid. All these solutions can be put as

$$\hat{\mathbf{w}}_i = \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{d}_i+\mathbf{b}_i, \quad \mathbf{b}_i\in\mathrm{Ker}(\mathbf{X}_i^T), \tag{5}$$

where as a consequence of (1), $\mathbf{b}_i$ has to satisfy the condition

$$\|\mathbf{b}_i-\hat{\mathbf{w}}_{i-1}\|^2 \le \delta_{i-1}-\mathbf{d}_i^T(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{d}_i+2\mathbf{d}_i^T(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T\hat{\mathbf{w}}_{i-1}. \tag{6}$$

Eqs. (5) and (6) define all the solutions. However, putting them in an easy parametric way that would allow us to pick different values of $\mathbf{b}_i$ might be quite difficult. Consider the particular choice:

$$\mathbf{b}_i = (\mathbf{I}-\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{X}_i^T)\hat{\mathbf{w}}_{i-1}. \tag{7}$$

It is easy to see that $\mathbf{b}_i\in\mathrm{Ker}(\mathbf{X}_i^T)$ and satisfies (6). Replacing (7) in (5) leads to the estimate:

$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1}+\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i, \tag{8}$$

which is the standard APA with step-size equal to 1. This solution must be used *only* when the hypersphere (1) intersects the affine space $\mathbf{e}_{\mathrm{p},i}=\mathbf{0}$. By computing the norm of the filter update, this condition can be expressed as

$$\mathbf{e}_i^T(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i \le \delta_{i-1}. \tag{9}$$

Case 2: The hypersphere (1) does not intersect the affine space $\mathbf{e}_{\mathrm{p},i}=\mathbf{0}$.

From (9), this case will take place only when $\delta_{i-1}<\mathbf{e}_i^T(\mathbf{X}_i^T\mathbf{X}_1)^{-1}\mathbf{e}_i$. Replacing (4) in (2) and after some lengthy but straightforward calculations we see that the optimum value of $\hat{\mathbf{w}}_i$ can be calculated as

$$\mathbf{t}_i = \arg\min_{\mathbf{t}_i\in\mathbb{R}^M}(\mathbf{t}_i-(\mathbf{X}_i^T)^\dagger\mathbf{e}_i)^T\mathbf{X}_i\mathbf{X}_i^T(\mathbf{t}_i-(\mathbf{X}_i^T)^\dagger\mathbf{e}_i) \quad \text{s.t. } \|\mathbf{t}_i\|^2=\delta_{i-1}, \tag{10}$$

where $\mathbf{t}_i=\hat{\mathbf{w}}_i-\hat{\mathbf{w}}_{i-1}$. This is a quadratic optimization problem with a quadratic constraint. Although the characteristics of the optimal solution are well known [6], there is no known closed-form solution to these kind of problems in general. For this reason we should look for suboptimal solutions to the problem. If δ_{i-1} is small enough the following equation satisfies the constraint in (10) and should be close to the optimal solution:

$$\mathbf{t}_i = \sqrt{\delta_{i-1}}\,\frac{\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i}{\|\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i\|}. \tag{11}$$

This is because if δ_{i-1} is small all the points on the hypersphere are closer to each other. Combining (8), (9) and (11) we obtain that the proposed algorithm can be written as

$$\hat{\mathbf{w}}_i = \hat{\mathbf{w}}_{i-1}+\min\left\{1,\frac{\sqrt{\delta_{i-1}}}{\|\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i\|}\right\}\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i. \tag{12}$$

The only thing that remains is the choice of the delta sequence. Similarly to the one proposed in [5]:

$$\delta_i = \alpha\delta_{i-1}+(1-\alpha)\min\left\{\frac{e_i^2}{\|\mathbf{x}_i\|^2},\delta_{i-1}\right\}. \tag{13}$$

The memory factor $\alpha\in(0,1)$ can be chosen as $\alpha=1-K/\kappa M$, where κ is an integer typically between 1 and 10. Although the recursion in (13) could be implemented using $\|\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i\|^2$ instead of $e_i^2/\|\mathbf{x}_i\|^2$, the second possibility is chosen because it gives better results. This is specially true if there is impulsive noise present. In that situation one sample of impulsive noise extends its influence during K samples through the vector $\mathbf{e}_i$. This could affect the dynamics of the sequence $\{\delta_i\}$. Using (13) for the calculation of δ_i allows us to bound the influence of one sample of impulsive noise to only one time step. The initial condition of $\{\delta_i\}$ can be set as $\delta_0=\sigma_d^2/(\sigma_x^2 M)$, with σ_x^2 and σ_d^2 standing for the power of the input and observed output signals, respectively. From (12), the proposed algorithm can be interpreted as a variable step-size APA with a step-size given by

$$\mu_i = \min\left\{1,\frac{\sqrt{\delta_{i-1}}}{\|\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i\|}\right\}. \tag{14}$$

When the condition (9) is satisfied, i.e., there is a low chance of having large noise samples contaminating the error vector (as long as δ_{i-1} is not too large), $\mu_i=1$ and the standard APA is performed. If this is not the case, the APA is performed with $\mu_i < 1$. In [5], it was proved that the sequence (13) is strictly decreasing towards zero, regardless of the values of δ_0 and α. Therefore, the adaptive step-size is not only providing the algorithm with a switching mechanism in order to perform robustly against the noise, but as the error decreases during the adaptation, the variable step-size will also let the algorithm to further decrease the error when the standard APA is no longer capable of doing so. We will show this effect later in the Simulation results. For these reasons, the proposed algorithm is named as *robust variable step-size affine projection algorithm* (RVSS-APA). It should be noticed that if K=1 the RVSS-APA reduces to the RVSS-NLMS presented in [5].

Some practical considerations should be considered. In order to update the algorithm the most computational expensive term is $\mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i)^{-1}\mathbf{e}_i$ (see (12)). It is known that this quantity can be computed using the guidelines given in [7] leading easily to a fast implementation of the proposed algorithm. A regularization constant $\beta = 20\sigma_x^2$ [7] is added to the matrix $\mathbf{X}_i^T\mathbf{X}_i$ before being inverted. It should be noted that even with this change, the constraint (1) will still be satisfied. Finally, a major issue should be regarded carefully. As the proposed delta sequence has the decreasing property shown in [5], although the algorithm becomes more robust against perturbations, it also loses its tracking ability. 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. We use similar controls as the ones proposed in [5], 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. While control 1 is better suited for system identification, control 2 is preferred for echo cancellation. See [5] for a detailed description of the parameters and their role. To avoid undesirable effects caused by the use of multiple regressors, the only difference from the methods in [5] is that the check on Δ_i is

$\Delta_i = (ctrl_{\text{new}} - ctrl_{\text{old}})/\delta_{i-1}$

if $\Delta_i > \xi$
$\delta_i = \delta_0$
elseif $ctrl_{\text{new}} > ctrl_{\text{old}}$
$\delta_i = \delta_{i-1} + (ctrl_{\text{new}} - ctrl_{\text{old}})/K$
else
$\delta_i = \alpha\delta_{i-1} + (1-\alpha)\min\left\{\frac{e_i^2}{\|\mathbf{x}_i\|^2}, \delta_{i-1}\right\}$
end

3. Simulation results

The system is taken from a measured acoustic impulse response truncated to M=512. The adaptive filter length is set to M. We choose this length because it is clearly a very harsh situation. Smaller lengths have been tested and the compared performance with the other algorithm remains qualitatively the same. We use the *mismatch* in dB, defined as $10\log_{10}(\|\mathbf{w}_i - \hat{\mathbf{w}}_i\|^2/\|\mathbf{w}_i\|^2)$, as a measure of performance. The plots are the result of single realizations of all the algorithms without any additional smoothing (except in Fig. 1(a) where 5 independent runs were averaged). A zero-mean Gaussian white noise ϑ_i is added to the system output to achieve a certain signal to background noise ratio, defined as SBNR $= 10\log_{10}(\sigma_y^2/\sigma_\vartheta^2)$, where σ_y^2 and σ_ϑ^2 stand for the power of the system output and background noise, respectively. All the algorithms are regularized with $\beta = 20\sigma_x^2$.

The behavior of the proposed RVSS-APA is compared with other strategies. We simulate a standard APA ($\mu = 1$), the robust APA (RAPA) introduced in [4] (which has a parameter k_0), and the gradient-limited APA (GL-APA) [3]. For the latter, by choosing the parameters k_{T_1}, k_{T_2}, k_{S_1} and k_{S_2}, the remaining parameters are calculated as

$$T_1 = \frac{10^{(k_{T_1}/20)}}{\sqrt{M}}, \quad T_2 = \frac{10^{(k_{T_2}/20)}}{\sqrt{M}}, \quad S_1 = k_{S_1}T_1, \quad S_2 = k_{S_2}S_1.$$

As in [5], in order to measure the performance of the nonstationary control methods, the quantities $\mathcal{M} = \max\Delta_i$ and $\mathcal{R} = \mathcal{M}/\mathcal{N}$ (where $\mathcal{N}$ is the second largest value of Δ_i) are computed. In every simulation (except in Fig. 1(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 detection threshold while that of $\mathcal{R}$ gives an idea on the reliability of the detection of a sudden change.

3.1. System identification under impulsive noise

The input is highly correlated AR1 process with pole in 0.95. The nonstationary control 1 is used in this application. 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 $P[\omega_i = 1] = p_{\text{imp}}$ and N_i is a zero-mean Gaussian with power $\sigma_N^2 = 1000\sigma_y^2$.

First, in Fig. 1(a) we compare the RVSS-APA with the standard APA for different projection orders. In this scenario there is a high SBNR and no impulsive noise. As it is well known, the standard APA shows faster convergence for larger K but with a worse steady-state. The RVSS-APA shows the same speed of convergence as the APA with a much lower steady-state. Moreover, the performance of the RVSS-APA with K=2,4,8 is very similar, even with respect to the steady-state error. This is because of the intrinsic variable step-size mechanism of the proposed algorithm. Therefore, there is no gain in using higher projection orders as they show similar performance but with increasing computational complexity.

Next, we include the other schemes. The parameters of the RAPA and the GL-APA were chosen to match the tracking performance of the RVSS-APA when no impulsive

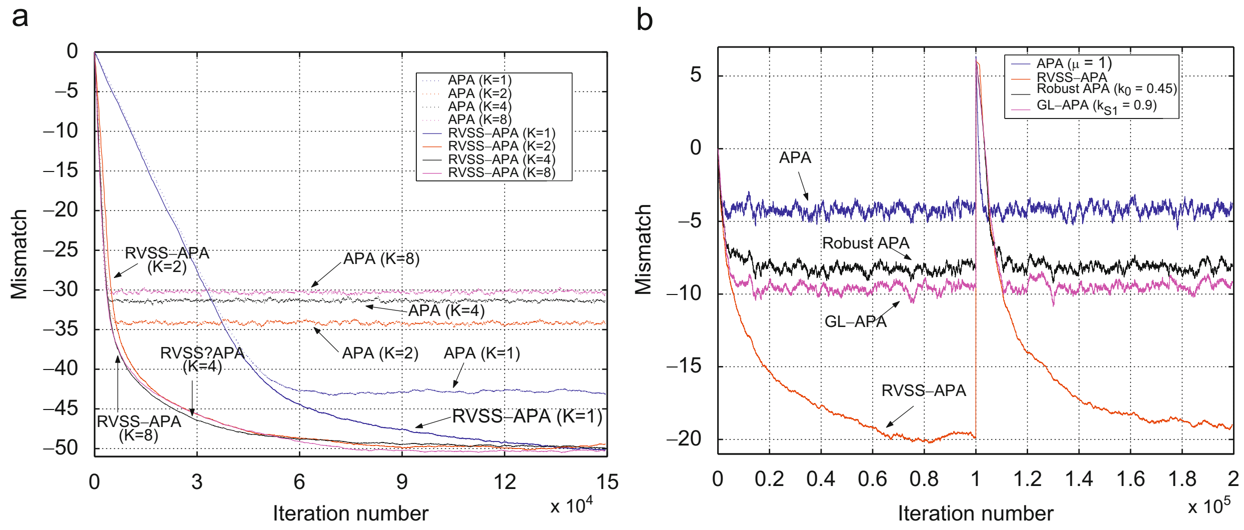

Fig. 1. Mismatch (in dB). AR1(0.95) input. No impulsive noise. $\kappa_{APA}=3$. $V_{\rm T}=3M$. $V_{\rm D}=0.75V_{\rm T}$. $\xi=50$. (a) SBNR=40 dB. $\kappa_{NLMS}=9$. (b) SBNR=10 dB. K=2. k_0=0.45. $k_{T_1}=-8$. $k_{T_2}=10$. $k_{S_1}=0.9$. $k_{S_2}=0.3$. $\mathcal{M}=62$. $\mathcal{R}=31$.

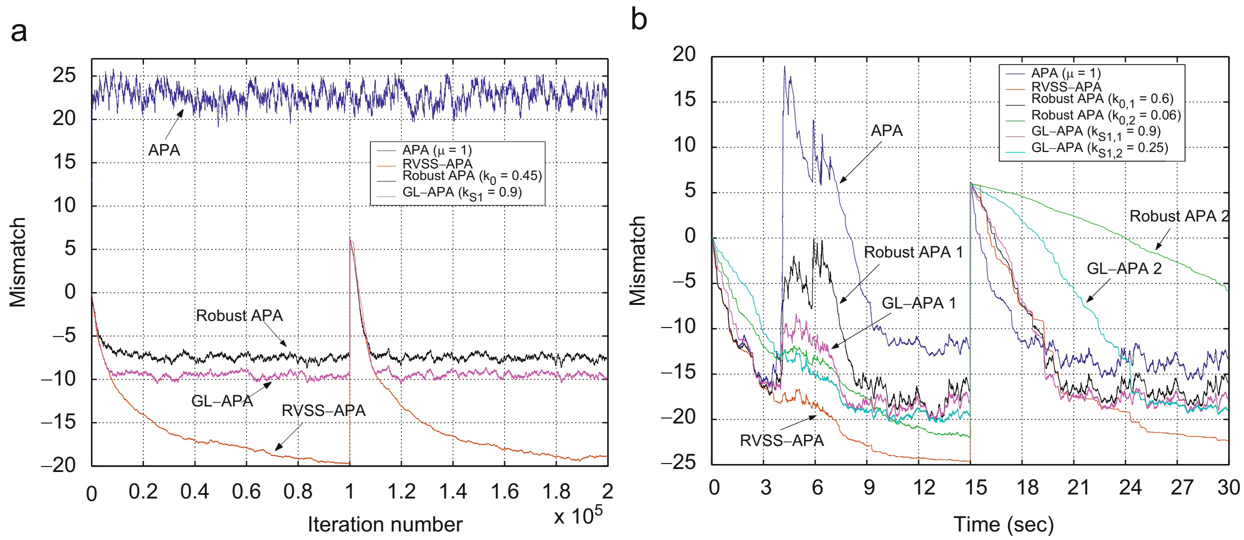

Fig. 2. Mismatch (in dB). (a) $p_{\rm imp}$=0.05. The other parameters are the same as in Fig. 1(b). $\mathcal{M}=105$. $\mathcal{R}=54$. (b) SBNR=25 dB. STNR=0 dB. K=2. $\kappa=5$. T=1.25. $\xi=50$. C_1=3. $V_{\rm T}=3M$. $V_\theta=6M$. $\tau=0.95$. $k_{0,1}=0.6$. $k_{T_1}=-18$. $k_{T_2}=0$. $k_{S_1,1}=0.9$. $k_{S_2}=0.3$. $k_{0,2}=0.06$. $k_{S_1,2}=0.25$. $\mathcal{M}=498$. $\mathcal{R}=668$.

noise was present (they were not able to match the low steady-state of the RVSS-APA without severely compromising their tracking performance).

In Fig. 1(b) it can be seen that the standard APA presents a large sensitivity to the low SBNR condition. Although the RAPA and the GL-APA perform better, the RVSS-APA has a steady-state more than one order of magnitude lower, without deteriorating the speed of convergence nor the tracking ability.

When the impulsive noise with $p_{\rm imp}=0.05$ is included, Fig. 2(a) shows that the standard APA presents a very poor performance. The remaining algorithms have the same performance as in Fig. 1(b).

We see that the switching mechanism of the RVSS-APA shows the advantage of not using the standard APA with $\mu=1$ when an impulse of noise is present or when the mismatch is low enough (and further gain can be accomplished). We also performed simulations under a high SBNR of 40 dB (not shown). Although the parameters of the GL-APA and RVSS-APA remained the same, the value k_0 of the RAPA had to be increased. When no impulsive noise was present, the RAPA and the GL-APA performed similarly to the standard APA. In contrast, the RVSS-APA reached a steady-state 15 dB lower without deteriorating the speed of convergence nor the tracking ability. When impulsive noise was added, the results were qualitatively similar to the ones in the low SBNR condition.

Regarding the performance of the nonstationary control, the results show that the RVSS-APA can recover well from a sudden change, with a speed close to the one of the standard APA with $\mu = 1$.

3.2. 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 activity. In this work, we used it partially. In [3], the GL-APA was successfully tested without a DTD so we did not include it. In the case of the RAPA and the RVSS-APA, long bursts of double-talk will affect the update of the scaling factor [4] and nonstationary control, respectively. To prevent this situation, the simple Geigel DTD [8] controls at each time step whether those updates should be performed. However, it should be emphasized that the update of the filter coefficients is performed at every iteration no matter the result of the DTD. This will avoid an unnecessary decrease in the speed of convergence (due to false alarms).

The far-end and near-end signals are speech sampled at 8 kHz, and they were both used previously in [5]. The SBNR is 25 dB while the *signal to total noise ratio* (STNR), defined as $\mathrm{STNR} = 10\log_{10}[\sigma_y^2/(\sigma_\vartheta^2+\sigma_\eta^2)]$, is set to 0 dB, where σ_η^2 is the power of the near-end signal before passing through the DTD. The span of the DTD is $D=M$ and the threshold is $T=1.25$. Under these conditions, the DTD detected only 22% of the near-end signal which, causing 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 nonstationary control 2 of [5], which was used under this environment.

In the first half of Fig. 2(b) we show that the RVSS-APA, the RAPA with $k_{0,2}$ and the GL-APA with $k_{S_1,2}$ present a similar behavior during the double-talk situation. The other three schemes suffer large mismatch deviations. In the second half, the RVSS-APA, the RAPA with $k_{0,1}$ and the GL-APA with $k_{S_1,1}$ show a similar tracking performance after the sudden change, while the other schemes that were robust during the double-talk situation present now a poor tracking performance. Moreover, the RAPA1 and the GL-APA1 present a steady-state just slightly better than the standard APA. In contrast, the RVSS-APA shows an extra 10 dB in steady-state (with a SBNR of 25 dB) with respect to the APA. This is another evidence of the superiority of the proposed algorithm over the other strategies. The large values of $\mathcal{M}$ and $\mathcal{R}$ indicate again that the sudden change was reliably detected.

4. Conclusions

In this work we derived a new robust version of the APA, the RVSS-APA, based on the framework introduced in [5]. It follows from optimizing the square norm of the *a posteriori* error vector subject to a time-dependent constraint (δ_i) on the norm of the filter update. The proposed dynamics for δ_i provide the algorithm with fast initial convergence as the standard APA with $\mu = 1$ but also a robust performance against noise. As shown in the simulations under system identification and acoustic echo cancellation scenarios, even under severe conditions, the performance of the proposed algorithm was very good.

References

[1] K. Ozeki, T. Umeda, An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties, Electron. Commun. Jpn. 67-A (5) (May 1984) 19–27.
[2] E.V. Papoulis, T. Stathaki, A normalized robust mixed-norm adaptive algorithm for system identification, IEEE Signal Process. Lett. 11 (1) (January 2004) 56–59.
[3] S. Shimauchi, Y. Haneda, A. Kataoka, A. Nishihara, Gradient-limited affine projection algorithm for double-talk robust and fast-converging acoustic echo cancellation, IEICE Trans. Fundam. E 90-A (3) (March 2007) 633–641.
[4] T. Gänsler, S.L. Gay, M.M. Sondhi, J. Benesty, Double-talk robust fast converging algorithms for network echo cancellation, IEEE Trans. Speech Audio Process. 8 (6) (November 2000) 656–663.
[5] L. Rey Vega, H. Rey, J. Benesty, S. Tressens, A new robust variable step-size NLMS algorithm, IEEE Trans. Signal Process. 56 (5) (May 2008) 1878–1893.
[6] S. Lucidi, L. Palagi, M. Roma, On some properties of quadratic programs with a convex quadratic constraint, SIAM J. Optim. 8 (1998) 105–122.
[7] S. Gay, S. Tavathia, The fast affine projection algorithm, in: Proceedings of the ICASSP-95, Detroit, USA, 1995, pp. 3023–3026.
[8] D.L. Duttweiler, A twelve channel digital echo canceler, IEEE Trans. Commun. 26 (5) (May 1978) 647–653.