A Practical Regularized Recursive Least-Squares Algorithm for Robust System Identification

Radu-Andrei Otopeleanu
Dept. Telecomm., POLITEHNICA Bucharest
Academy of Romanian Scientists
Ilfov 3, 050044 Bucharest, Romania
radu.otopeleanu@upb.ro

Jacob Benesty
INRS-EMT
University of Quebec
Montreal, QC H5A 1K6, Canada
jacob.benesty@inrs.ca

Constantin Paleologu
Dept. Telecomm.
POLITEHNICA Bucharest
060042 Bucharest, Romania
pale@comm.pub.ro

Cristian-Lucian Stanciu
Dept. Telecomm., POLITEHNICA Bucharest
Academy of Romanian Scientists
Ilfov 3, 050044 Bucharest, Romania
cristian@comm.pub.ro

Laura-Maria Dogariu
Dept. Telecomm., POLITEHNICA Bucharest
Academy of Romanian Scientists
Ilfov 3, 050044 Bucharest, Romania
ldogariu@comm.pub.ro

Silviu Ciochină
Dept. Telecomm.
POLITEHNICA Bucharest
060042 Bucharest, Romania
silviu@comm.pub.ro

***Abstract*—The recursive least-squares (RLS) adaptive filtering algorithm is frequently used in system identification problems. The popularity of this algorithm is mainly related to its fast convergence rate. In this context, the main parameter that controls the convergence features of the RLS filter is the forgetting factor. On the other hand, in noisy environments, the robustness of the algorithm can be improved by using an appropriate regularization term. In this paper, we propose a regularized RLS-type algorithm, by considering a linear state model and following the weighted least-squares optimization criterion. The resulting optimal regularization parameter also includes a specific term related to the model uncertainties, which is estimated in a practical manner within the algorithm. Simulation results obtained in the framework of echo cancellation support the performance features of the regularized RLS algorithm, which could represent an appealing solution for robust system identification.**

***Index Terms*—Adaptive filter, recursive least-squares (RLS) algorithm, regularization, robustness, system identification.**

I. Introduction

The recursive least-squares (RLS) algorithm [1] is a popular tool in many adaptive filtering applications. Due to its fast convergence rate and robustness to the character of input data (in terms of correlation), it is widely involved in a variety of frameworks, e.g., see [2]–[5], among many others. In this context, an important category of applications is related to system identification problems, where echo cancellation represents one of the most challenging cases. This is an interesting combination between a system identification scheme (aiming to model the echo path) and an interference cancellation configuration (targeting to recover the near-end signal) [6].

There are many interesting versions of the RLS algorithm, using different theoretical approaches and filter structures. Among them, the exponentially weighted RLS algorithm implemented using a transversal filter structure [1] is usually considered as the conventional version and common benchmark. Its overall performance is mainly controlled by the so-called forgetting factor, which is a positive term (smaller than one) that "weights" the contribution of the error signal involved into the cost function. Setting the forgetting factor leads to a compromise between the main performance criteria. For example, in echo cancellation, a large value leads to good accuracy of the echo path estimate, but affects the tracking capabilities. Nevertheless, even when using very large values of the forgetting factor, there is an inherent limitation of the algorithm related to its robustness in the presence of the near-end signal. Thus, a robust system identification is problematic in noisy environments (with high and nonstationary background noise) or in double-talk scenarios (when the speakers talk simultaneously) [2], [6], and it cannot be achieved using only the forgetting factor as the control parameter.

In order to improve the robustness of the RLS algorithm in such challenging scenarios, a natural approach is to consider a proper regularization term within the cost function. There are many interesting solutions to this problem, which were designed in the framework of different applications, e.g., see [3], [5], [7]–[13], and the references therein. However, most of the so-called regularized RLS algorithms developed in this context require some a priori information about the system/environment or need some additional parameters that are not always easy to estimate/tune in real-world scenarios.

In this paper, we derive a more practical regularized RLS-type algorithm, which relies on a linear state model and exploits the weighted least-squares optimization criterion. The regularization parameter of the proposed algorithm includes the contribution of both the external noise and the model uncertainties, thus leading to improved robustness in terms of system identification. In addition, it does not require any additional information related to the environment.

Following this introduction, the proposed regularized RLS algorithm is presented in Section II. Its performance is supported by the experimental results provided in Section III, in the framework of echo cancellation. Finally, Section IV concludes this work.

II. Proposed Regularized RLS Algorithm

In the general framework of a single-input single-output (SISO) system identification problem, the main goal is to model/estimate an unknown system characterized by the time-varying impulse response $\mathbf{h}(t)$, with L coefficients, where t is the discrete-time index. To this purpose, at each time index t, a reference signal, $d(t)$, obtained at the system output, is considered to be available, together with the last L samples of the input sequence, $x(t)$, which are grouped into the vector $\mathbf{x}(t) = \begin{bmatrix} x(t) & x(t-1) & \dots & x(t-L+1) \end{bmatrix}^T$, where the superscript ^{T} stands for transposition. In this context, let us consider the linear state model:

$$d(t) = \mathbf{h}^T(t)\mathbf{x}(t) + v(t), \quad v(t) \sim \mathcal{N}\left[0, r_v(t)\right], \tag{1}$$

$$\mathbf{h}(t) = \mathbf{h}(t-1) + \mathbf{u}(t), \quad \mathbf{u}(t) \sim \mathcal{N}\left[\mathbf{0}, \mathbf{R}_{\mathbf{u}}(t)\right], \tag{2}$$

where $v(t)$ is a zero-mean additive noise (uncorrelated to the input signal), which corrupts the output of the unknown system, and $\mathbf{h}(t)$ follows a simplified first-order Markov model, with $\mathbf{u}(t)$ being a zero-mean white Gaussian noise signal vector [uncorrelated to $\mathbf{h}(t-1)$ and $v(t)$]. Related to the model from (1) and (2), $r_v(t)$ and $\mathbf{R}_{\mathbf{u}}(t)$ are the variance and covariance matrix of $v(t)$ and $\mathbf{u}(t)$, respectively, while $\mathbf{h}(t)$ is the state system (of length L).

One possible way to solve the problem of estimating $\mathbf{h}(t)$ is via the least-squares (LS) approach. In this context, denoting by $\widehat{\mathbf{h}}(t)$ the estimate of $\mathbf{h}(t)$, provided by an adaptive filter, the weighted LS criterion can be written as

$$\begin{aligned} \mathcal{J}_0\left[\widehat{\mathbf{h}}(t)\right] &= \sum_{i=1}^{t} \lambda^{t-i} \frac{\left[d(i) - \widehat{\mathbf{h}}^T(t)\mathbf{x}(i)\right]^2}{r_v(t)} \\ &+ \frac{1}{L}\sum_{i=1}^{t} \lambda^{t-i} \left[\widehat{\mathbf{h}}(t) - \mathbf{h}(i-1)\right]^T \mathbf{R}_{\mathbf{u}}^{-1}(t) \left[\widehat{\mathbf{h}}(t) - \mathbf{h}(i-1)\right], \end{aligned} \tag{3}$$

where λ $(0 \ll \lambda < 1)$ is known as the forgetting factor. The second term from the right-hand side of (3) acts like a regularization component, which is related to the model uncertainties from (2). Assuming that $\mathbf{R}_{\mathbf{u}}(t) = r_u(t)\mathbf{I}$, where the variance $r_u(t)$ captures the uncertainties in $\mathbf{h}(t)$ and $\mathbf{I}$ denotes the identity matrix (of size $L \times L$), the cost function from (3) can be equivalently expressed as

$$\begin{aligned} \mathcal{J}\left[\widehat{\mathbf{h}}(t)\right] &= \sum_{i=1}^{t} \lambda^{t-i} \left[d(i) - \widehat{\mathbf{h}}^T(t)\mathbf{x}(i)\right]^2 \\ &+ \frac{r_v(t)}{L r_u(t)} \sum_{i=1}^{t} \lambda^{t-i} \left[\widehat{\mathbf{h}}(t) - \mathbf{h}(i-1)\right]^T \left[\widehat{\mathbf{h}}(t) - \mathbf{h}(i-1)\right] \\ &= r_d(t) - 2\widehat{\mathbf{h}}^T(t)\mathbf{r}_{\mathbf{x}d}(t) + \widehat{\mathbf{h}}^T(t)\mathbf{R}_{\mathbf{x}}(t)\widehat{\mathbf{h}}(t) \\ &+ \frac{r_v(t)}{L r_u(t)} \left[\ell(t)\widehat{\mathbf{h}}^T(t)\widehat{\mathbf{h}}(t) - 2\widehat{\mathbf{h}}^T(t)\sum_{i=1}^{t}\lambda^{t-i}\mathbf{h}(i-1) \right. \\ &\left. + \sum_{i=1}^{t}\lambda^{t-i}\mathbf{h}^T(i-1)\mathbf{h}(i-1)\right], \end{aligned} \tag{4}$$

where

$$r_d(t) = \sum_{i=1}^{t} \lambda^{t-i} d^2(i) = \lambda r_d(t-1) + d^2(t), \tag{5}$$

$$\mathbf{r}_{\mathbf{x}d}(t) = \sum_{i=1}^{t} \lambda^{t-i} \mathbf{x}(i) d(i) = \lambda \mathbf{r}_{\mathbf{x}d}(t-1) + \mathbf{x}(t)d(t), \tag{6}$$

$$\mathbf{R}_{\mathbf{x}}(t) = \sum_{i=1}^{t} \lambda^{t-i} \mathbf{x}(i)\mathbf{x}^T(i) = \lambda \mathbf{R}_{\mathbf{x}}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t), \tag{7}$$

$$\ell(t) = \sum_{i=1}^{t} \lambda^{t-i} = \lambda \ell(t-1) + 1. \tag{8}$$

In (4), for t large enough, the term $\sum_{i=1}^{t} \lambda^{t-i}\mathbf{h}(i-1)$ can be approximated to zero. Then, from the minimization of $\mathcal{J}\left[\widehat{\mathbf{h}}(t)\right]$ with respect to $\widehat{\mathbf{h}}(t)$, we obtain the normal equations:

$$\left[\mathbf{R}_{\mathbf{x}}(t) + \delta(t)\mathbf{I}\right]\widehat{\mathbf{h}}(t) = \mathbf{r}_{\mathbf{x}d}(t), \tag{9}$$

where

$$\delta(t) = \frac{\ell(t)}{L}\eta(t) \tag{10}$$

acts like a regularization parameter, which includes the noise-to-uncertainty ratio (NUR), defined as

$$\eta(t) = \frac{r_v(t)}{r_u(t)}. \tag{11}$$

The optimal solution of (9) results in

$$\widehat{\mathbf{h}}(t) = \left[\mathbf{R}_{\mathbf{x}}(t) + \delta(t)\mathbf{I}\right]^{-1}\mathbf{r}_{\mathbf{x}d}(t), \tag{12}$$

which can be recursively obtained using an RLS-type algorithm. To this purpose, similar to (9), let us consider the normal equations from time index $t-1$, i.e.,

$$\left[\mathbf{R}_{\mathbf{x}}(t-1) + \delta(t-1)\mathbf{I}\right]\widehat{\mathbf{h}}(t-1) = \mathbf{r}_{\mathbf{x}d}(t-1). \tag{13}$$

Next, multiplying by λ on both sides of (13) and using the updates (6) and (7), the previous normal equations become

$$\left[\mathbf{R}_{\mathbf{x}}(t) - \mathbf{x}(t)\mathbf{x}^T(t) + \delta(t)\mathbf{I}\right]\widehat{\mathbf{h}}(t-1) = \mathbf{r}_{\mathbf{x}d}(t) - \mathbf{x}(t)d(t), \tag{14}$$

also assuming that the regularization parameter is slowly varying from one time index to another, so that, for λ close to 1, we can use the approximation $\delta(t) \approx \lambda\delta(t-1)$ at this point. Thus, (14) can be further developed as

$$\left[\mathbf{R}_{\mathbf{x}}(t) + \delta(t)\mathbf{I}\right]\widehat{\mathbf{h}}(t-1) + \mathbf{x}(t)e(t) = \mathbf{r}_{\mathbf{x}d}(t), \tag{15}$$

where

$$e(t) = d(t) - \mathbf{x}^T(t)\widehat{\mathbf{h}}(t-1) \tag{16}$$

is the a priori error signal. Multiplying (to the left) both sides of (15) with $\left[\mathbf{R}_{\mathbf{x}}(t) + \delta(t)\mathbf{I}\right]^{-1}$ (considering that this matrix is invertible) and using (12), together with (10) and (11), the update of the regularized RLS-type algorithm becomes

$$\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}(t-1) + \left[\mathbf{R}_{\mathbf{x}}(t) + \frac{\ell(t)}{L} \cdot \frac{r_v(t)}{r_u(t)}\mathbf{I}\right]^{-1} \mathbf{x}(t)e(t). \quad (17)$$

In order to further simplify (17), for t large enough, we can use the approximation $\ell(t) \approx 1/(1-\lambda)$, which results based on (8). Also, if we commonly set $\lambda = 1 - 1/(KL)$, with $K \geq 1$, for $L \gg 1$ (which is usually the case in echo cancellation), we obtain $\ell(t)/L \approx K$.

Most important, the variances required in (17) should be estimated in a simple and practical manner. First, we focus on the additive noise variance, $r_v(t)$. Of course, $v(t)$ is not available; however, some related information can be extracted from the error signal, $e(t)$. The reason is that in system identification scenarios, the goal of the adaptive algorithm is not to drive the error signal to zero, since this would introduce noise into the filter estimate. Instead, we should recover the noise signal from the error of the adaptive filter, after this one converges to its solution, so that $r_v(t) \approx r_e(t)$, where $r_e(t) = E[e^2(t)]$, with $E[\cdot]$ denoting the mathematical expectation. Consequently, we can use the recursive estimation:

$$\widehat{r}_v(t) = \lambda \widehat{r}_v(t-1) + (1-\lambda)e^2(t), \quad (18)$$

with $\widehat{r}_v(0) = 0$. Similar estimation strategies were previously used in the context of RLS algorithms with variable forgetting factors [14], [15].

Second, the model uncertainties, $r_u(t)$, should be estimated. Using the adaptive filter estimates from time indices t and $t-1$ in (2), we can write $\widehat{\mathbf{h}}(t) - \widehat{\mathbf{h}}(t-1) \approx \mathbf{u}(t)$, while $\|\mathbf{u}(t)\|^2 \approx L r_u(t)$, for $L \gg 1$, with $\|\cdot\|$ denoting the Euclidean norm. As a result, using a similar estimator as in (18), the model uncertainties can be recursively evaluated as

$$\widehat{r}_u(t) = \lambda \widehat{r}_u(t-1) + (1-\lambda)\frac{\left\|\widehat{\mathbf{h}}(t) - \widehat{\mathbf{h}}(t-1)\right\|^2}{L}, \quad (19)$$

with $\widehat{r}_u(0) = \epsilon$, where ϵ is a small positive constant, which should be used for the initialization [since this estimate appears at the denominator in (17)]. At this point, we should notice that a different weighting factor could be used (instead of λ) in (18) and (19), e.g., $\gamma = 1 - 1/(QL)$, with $Q \geq 1$ and $Q \neq K$. Nevertheless, for the sake of simplicity, the forgetting factor λ is also used to this purpose. Finally, for numerical reasons [when dealing with very small values of $\widehat{r}_u(t)$], it is recommended to add a very small positive constant ($\varepsilon > 0$) to the denominator of the ratio from (17).

Concluding, under these circumstances, the filter update from (17) results in

$$\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}(t-1) + \left[\mathbf{R}_{\mathbf{x}}(t) + K\frac{\widehat{r}_v(t)}{\varepsilon + \widehat{r}_u(t-1)}\mathbf{I}\right]^{-1} \mathbf{x}(t)e(t). \quad (20)$$

It can be noticed that different time indices are used in the estimated NUR, as compared to (17), i.e., $\widehat{r}_v(t)$ and $\widehat{r}_u(t-1)$, respectively. The explanation relies on the fact that $e(t)$ is available before the filter update, so that we can use $\widehat{r}_v(t)$ from (18), while $\widehat{\mathbf{h}}(t)$ is not yet available for the estimation of $\widehat{r}_u(t)$ in (19), which requires using the estimate from the previous time index, $\widehat{r}_u(t-1)$. Nevertheless, $\widehat{r}_v(t)$ also depends on $\widehat{\mathbf{h}}(t-1)$, which appears in $e(t)$ from (16). Summarizing, the proposed regularized RLS algorithm based on the weighted LS criterion, namely WR-RLS, is provided in Table I.

TABLE I
WR-RLS ALGORITHM

Parameters:
$\lambda = 1 - \frac{1}{KL}, \; K \geq 1, \; \varepsilon > 0$
Initialization:
$\mathbf{R}_{\mathbf{x}}(0) = \mathbf{0}_{L\times L}, \; \widehat{\mathbf{h}}(0) = \mathbf{0}_{L\times 1}, \; \widehat{r}_v(0) = 0, \; \widehat{r}_u(0) = \epsilon, \; \epsilon > 0$
For time-index $t = 1, 2, \ldots$:
$\mathbf{x}(t) = [\; x(t) \quad x(t-1) \quad \cdots \quad x(t-L+1) \;]^T$
$\mathbf{R}_{\mathbf{x}}(t) = \lambda \mathbf{R}_{\mathbf{x}}(t-1) + \mathbf{x}(t)\mathbf{x}^T(t)$
$e(t) = d(t) - \mathbf{x}^T(t)\widehat{\mathbf{h}}(t-1)$
$\widehat{r}_v(t) = \lambda \widehat{r}_v(t-1) + (1-\lambda)e^2(t)$
$\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}(t-1) + \left[\mathbf{R}_{\mathbf{x}}(t) + K\frac{\widehat{r}_v(t)}{\varepsilon + \widehat{r}_u(t-1)}\mathbf{I}\right]^{-1} \mathbf{x}(t)e(t)$
$\widehat{r}_u(t) = \lambda \widehat{r}_u(t-1) + (1-\lambda)\frac{\left\|\widehat{\mathbf{h}}(t) - \widehat{\mathbf{h}}(t-1)\right\|^2}{L}$

Several remarks can be outlined related to the computational complexity of this algorithm, which contains several challenging operations. First, the update of the matrix $\mathbf{R}_{\mathbf{x}}(t)$ can be efficiently performed by taking advantage of its symmetry and the time-shift property of the input vector, $\mathbf{x}(t)$. As a result, only the first column (and the first row) of the matrix $\mathbf{x}(t)\mathbf{x}^T(t)$ should be evaluated in each iteration, i.e., $x(t)\mathbf{x}(t)$ (and its transpose), while its bottom-right $(L-1)\times(L-1)$ block is identical to the top-left $(L-1)\times(L-1)$ block of $\mathbf{x}(t-1)\mathbf{x}^T(t-1)$. Second, the direct matrix inversion required within the filter update is computationally expensive, especially for large values of L. Alternatively, efficient iterative techniques could be involved to solve this issue, like the dichotomous coordinate descent (DCD) method [16], [17], with a computational complexity proportional to $\mathcal{O}(L)$. In other words, the filter update is rewritten as $\widehat{\mathbf{h}}(t) = \widehat{\mathbf{h}}(t-1) + \Delta\widehat{\mathbf{h}}(t)$, while the update term (namely the increment) results as the iterative solution of an auxiliary set of normal equations, i.e., $\left[\mathbf{R}_{\mathbf{x}}(t) + K\frac{\widehat{r}_v(t)}{\varepsilon+\widehat{r}_u(t-1)}\mathbf{I}\right]\Delta\widehat{\mathbf{h}}(t) = \mathbf{r}_{\mathbf{x}e}(t)$, where $\mathbf{r}_{\mathbf{x}e}(t) = \lambda\mathbf{r}_{\mathbf{x}e}(t-1) + \mathbf{x}(t)e(t)$ is the so-called residual vector [16]. The derivation of the DCD-based version of the WR-RLS algorithm is beyond the scope of this paper and it is presented in a subsequent publication [18]. The main goal of the current work remains the development and investigation of the basic version of the proposed WR-RLS algorithm, together with its performance related to robust system identification.

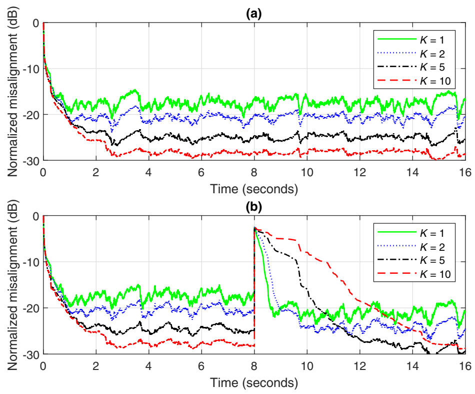

Fig. 1. Normalized misalignment of the proposed WR-RLS algorithm using $\lambda = 1 - 1/(KL)$, with different values of K and $L = 128$: (a) without echo path change and (b) with echo path change (after 8 seconds). The input signal is a speech sequence and SNR = 20 dB.

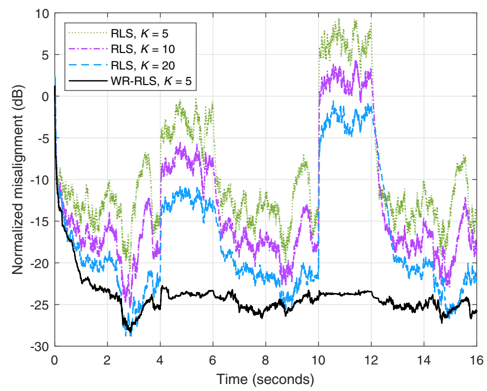

Fig. 2. Normalized misalignment of the conventional RLS algorithm using $\lambda = 1-1/(KL)$, with different values of K and $L = 128$, and the proposed WR-RLS algorithm using $\lambda = 1 - 1/(5L)$. The input signal is a speech sequence and the SNR varies from 20 dB to 10 dB between time 4 to 6 (seconds), and from 20 dB to 0 dB between time 10 to 12 (seconds).

III. Simulation Results

The experimental framework is based on an echo cancellation scenario, aiming to identify a network echo path from the ITU-T G168 Recommendation [19]. To this purpose, the fourth cluster of coefficients ($\mathbf{b}_4$) from [19] is selected, with the length $L = 128$. The impulse response of the echo path, $\mathbf{h}(t)$, is obtained by adding a white Gaussian noise to $\mathbf{b}_4$, with the variance $\xi \left\|\mathbf{b}_4\right\|^2$, using $\xi = 10^{-4}$. The input signal (i.e., the far-end signal) is a speech sequence from a female voice, with a sampling rate of 8 kHz. The background noise that corrupts the output of the echo path is white and Gaussian, with the signal-to-noise ratio (SNR) set to 20 dB. The normalized misalignment (in dB) is used as performance measure, being computed as $20\log_{10} \left\|\mathbf{h}(t) - \widehat{\mathbf{h}}(t)\right\| / \left\|\mathbf{h}(t)\right\|$.

In the first set of experiments, the influence of the forgetting factor on the performance of the proposed WR-RLS algorithm is assessed. This is performed by varying the value of K, since $\lambda = 1 - 1/(KL)$. Clearly, larger values of K lead to larger values of λ (i.e., closer to one). As we can notice in Fig. 1(a), larger values of the forgetting factor are suitable for a better accuracy of the estimate provided by the adaptive filter, since the misalignment is decreasing when the value of K (or λ) increases. This is an expected result, according to the common knowledge about the exponentially weighted RLS-type algorithms. On the other hand, higher values of λ slow down the tracking capabilities of these algorithms, as also supported in Fig. 1(b). Here, an echo path change is introduced in the middle of simulation, by changing the value of ξ to 10^{-2}. As expected, the higher the forgetting factor, the slower the tracking reaction. Nevertheless, the WR-RLS algorithm can achieve a reasonable compromise in terms of these two performance criteria, i.e., accuracy versus tracking.

The second experiment is dedicated to a common scenario in echo cancellation applications, where the background noise could vary in time. The conventional (exponentially weighted) RLS algorithm [1] is used as a benchmark. Its performance is basically influenced by the value of the forgetting factor, so different values of λ are considered in this experiment (also by varying the value of K). The WR-RLS algorithm uses $\lambda = 1 - 1/(5L)$. Two bursts of background noise variation are simulated, by first decreasing the SNR from 20 dB to 10 dB between time 4 to 6 seconds, then from 20 dB to 0 dB between time 10 to 12 seconds. The results are reported in Fig. 2, where we can notice the good robustness of the proposed WR-RLS algorithm in both situations. On the other hand, the conventional RLS algorithm is significantly affected by the SNR variations, for all the values of the forgetting factor used in this experiment. In order to improve its accuracy and robustness, the value of λ should be increased, which can result in reducing the tracking capabilities of the algorithm.

Maybe the most challenging scenario in echo cancellation is the double-talk situation, when the speakers talk at the same time. In this case, the near-end speech acts like a large level of nonstationary disturbance, which can significantly bias the adaptive filter estimate. Usually a double-talk detector (DTD) is involved in such scenarios [2], [6], detecting the presence of the near-end speech and freezing (or slowing down) the adaptation during the double-talk periods. However, this is not an easy task in practice, since the algorithm behind the DTD has some inherent latencies, while it can also trigger false alarms or miss detections. Due to these reasons, the adaptive algorithms used for echo cancellation should be as robust as possible in the presence of double-talk.

In this context, the last experiment considers a double-talk scenario without using any DTD, which represents a very challenging situation for the adaptive algorithm. The proposed WR-RLS is compared to the conventional RLS algorithm, but also to the variable-regularized RLS (VR-RLS) algorithm from [10], which evaluates its time-varying regularization parameter as a function of the estimated SNR. This is also considered a practical algorithm, since it does not require any additional parameter related to the system or the environment. All the algorithms use different values of their forgetting

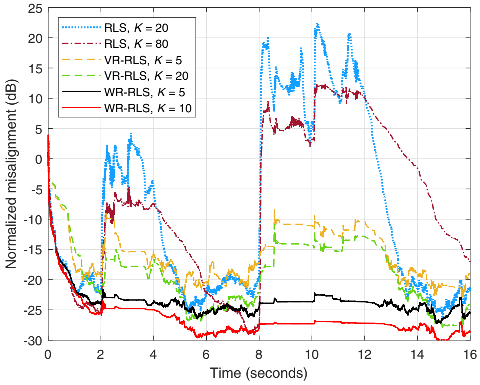

Fig. 3. Normalized misalignment of the conventional RLS, VR-RLS [10], and proposed WR-RLS algorithms, using $\lambda = 1 - 1/(KL)$, with different values of K and $L = 128$. The far-end (input) signal is a speech sequence, SNR = 20 dB, and the near-end speech (double-talk scenario) appears between time 2 to 4 (seconds) and between time 8 to 12 (seconds).

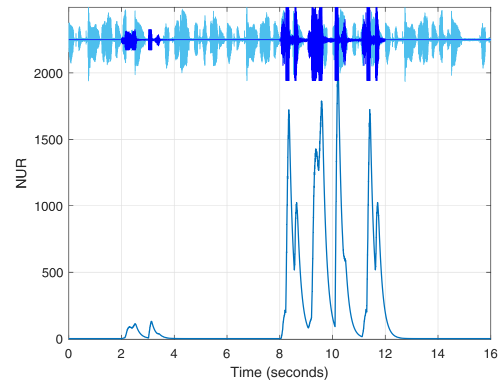

Fig. 4. The estimated NUR of the proposed WR-RLS algorithm [with $\lambda = 1 - 1/(10L)$] for the double-talk scenario reported in Fig. 3. The far-end speech is depicted in the top part of the figure (in light blue), together with the near-end speech (in dark blue).

factor, $\lambda = 1 - 1/(KL)$, which are selected by varying the values of K. Two periods of double-talk are simulated, by using a male voice at the near-end (with different intensities), between time 2 to 4 seconds and between time 8 to 12 seconds, respectively. The results are provided in Fig. 3. As we can notice, the WR-RLS algorithm is very robust during both double-talk periods, while its accuracy is improving for larger values of the forgetting factor. The VR-RLS algorithm [10] also has a reasonable robust behavior during double-talk, outperforming the conventional RLS algorithm. Despite the large values of the forgetting factor, this conventional benchmark cannot cope with the double-talk scenario, while also facing a slower recovery when λ (or K) increases. Overall, the estimated NUR within the regularization parameter of the WR-RLS algorithm provides a reliable measure during double-talk periods, as shown in Fig. 4. Basically, the ratio $\widehat{r}_v(t)/\widehat{r}_u(t-1)$ increases in the presence of the near-end speech, thus reducing the update term of the algorithm and, consequently, slowing down its adaptation, which represents the desired behavior.

IV. Conclusions

In this paper, we have developed a regularized RLS-type algorithm, using the weighted LS optimization criterion and including a regularization component (based on the model uncertainties) into the cost function. The regularization term of the proposed WR-RLS algorithm contains the NUR, which reflects the influence of both the external noise and the model uncertainties. This is estimated in a practical manner within the algorithm, without any a priori information regarding the system/environment. Thus, the WR-RLS algorithm is efficient for robust system identification, as indicated by the echo cancellation experiments, including the double-talk scenario.

References

[1] S. Haykin, *Adaptive Filter Theory*. Fifth Edition, Upper Saddle River, NJ, USA: Pearson, 2014.

[2] J. Benesty and Y. Huang, Eds., *Adaptive Signal Processing–Applications to Real-World Problems*. Berlin, Germany: Springer-Verlag, 2003.

[3] N. Zhou, D. J. Trudnowski, J. W. Pierre, and W. A. Mittelstadt, "Electromechanical mode online estimation using regularized robust RLS methods," *IEEE Trans. Power Systems*, vol. 23, pp. 1670–1680, Nov. 2008.

[4] N. Iqbal and A. Zerguine, "AFD-DFE using constraint-based RLS and phase noise compensation for uplink SC-FDMA," *IEEE Trans. Vehicular Technology*, vol. 66, pp. 4435–4443, May 2017.

[5] Y. Zhong, C. Yu, X. Xiang, and L. Lian, "Proximal policy-optimized regularized least squares algorithm for noise-resilient motion prediction of UMVs," *IEEE J. Oceanic Eng.*, vol. 49, pp. 1397–1410, Oct. 2024.

[6] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control–A Practical Approach*. Hoboken, NJ, USA: Wiley, 2004.

[7] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Process.*, vol. 88, pp. 594–611, Mar. 2008.

[8] J. Benesty, C. Paleologu, and S. Ciochină, "Regularization of the RLS algorithm," *IEICE Trans. Fundamentals*, vol. E94-A, pp. 1628–1629, Aug. 2011.

[9] E. M. Eksioglu and A. K. Tanc, "RLS algorithm with convex regularization," *IEEE Signal Process. Lett.*, vol. 18, pp. 470–473, Aug. 2011.

[10] C. Elisei-Iliescu, C. Stanciu, C. Paleologu, J. Benesty, C. Anghel, and S. Ciochină, "Robust variable regularized RLS algorithms," in *Proc. IEEE HSCMA*, 2017, pp. 171–175.

[11] F. Yang, J. Yang, and F. Albu, "An alternative solution to the dynamically regularized RLS algorithm," in *Proc. APSIPA*, 2019, pp. 1072–1075.

[12] A. Lawal, K. Abed-Meraim, Q. Mayyala, N. Iqbal, and A. Zerguine, "Blind adaptive channel estimation using structure subspace tracking," in *Proc. IEEE Asilomar*, 2021, pp. 1372–1376.

[13] B. Li, S. Wu, E. E. Tripp, A. Pezeshki, and V. Tarokh, "Recursive least squares with minimax concave penalty regularization for adaptive system identification," *IEEE Access*, vol. 12, pp. 66993–67004, May 2024.

[14] S.-H. Leung and C. F. So, "Gradient-based variable forgetting factor RLS algorithm in time-varying environments," *IEEE Trans. Signal Process.*, vol. 53, pp. 3141–3150, Aug. 2005.

[15] C. Paleologu, J. Benesty, and S. Ciochină, "A robust variable forgetting factor recursive least-squares algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 15, pp. 597–600, 2008.

[16] Y. V. Zakharov, G. P. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Trans. Signal Process.*, vol. 56, pp. 3150–3161, July 2008.

[17] Y. Yu, L. Lu, Z. Zheng, W. Wang, Y. Zakharov, and R. C. de Lamare, "DCD-based recursive adaptive algorithms robust against impulsive noise," *IEEE Trans. Circuits Systems II: Express Briefs*, vol. 67, pp. 1359–1363, July 2020.

[18] R. A. Otopeleanu, C. L. Stanciu, C. Paleologu, J. Benesty, L. M. Dogariu, and S. Ciochină, "A robust RLS-DCD adaptive algorithm with variable regularization parameter," in *Proc. IEEE ISSCS*, 2025, 4 pages.

[19] *Digital Network Echo Cancellers*, ITU-T Recommendation G.168, 2015. [https://www.itu.int/rec/T-REC-G.168 (accessed on May 30th, 2025)]