Article:

Yu, Yi, Lu, Lu, Zakharov, Yury orcid.org/0000-0002-2193-4334 et al. (2 more authors) (2022) Robust sparsity-aware RLS algorithms with jointly-optimized parameters against Impulsive noise. IEEE Signal Processing Letters. pp. 1037-1041. ISSN 1070-9908

https://doi.org/10.1109/LSP.2022.3166395

Robust Sparsity-Aware RLS algorithms with Jointly-Optimized Parameters Against Impulsive Noise

Yi Yu, *Member, IEEE*, Lu Lu, *Member, IEEE*, Yuriy Zakharov, *Senior Member, IEEE*, Rodrigo C. de Lamare, *Senior Member, IEEE*, and Badong Chen, *Senior Member, IEEE*

***Abstract*—This paper proposes a unified sparsity-aware robust recursive least-squares RLS (S-RRLS) algorithm for the identification of sparse systems under impulsive noise. The proposed algorithm generalizes multiple algorithms only by replacing the specified criterion of robustness and sparsity-aware penalty. Furthermore, by jointly optimizing the forgetting factor and the sparsity penalty parameter, we develop the jointly-optimized S-RRLS (JO-S-RRLS) algorithm, which not only exhibits low misadjustment but also can track well sudden changes of a sparse system. Simulations in impulsive noise scenarios demonstrate that the proposed S-RRLS and JO-S-RRLS algorithms outperform existing techniques.**

***Index Terms*—Impulsive noises, variable parameters, robust RLS, sparse systems.**

I. Introduction

IN realistic environments, in addition to Gaussian noise, impulsive noise is often present such as in echo cancellation, underwater acoustics, audio processing, and communications [1]–[3]. Realizations of the impulsive noise are random and sparse in the time domain, and typically have much higher amplitudes than Gaussian noise. In these scenarios, the popular recursive least squares (RLS) algorithm experiences significant performance deterioration.

Aiming at impulsive noise cases, various strategies have been applied to develop robust RLS-type adaptive algorithms [4]–[9]. In particular, the recursive least M-estimate (RLM) algorithm was proposed by introducing the M-estimator of the error signal [4]; it not only shows good robustness against impulsive noises, but also almost retains the fast convergence which the RLS algorithm possesses in the Gaussian noise. The correntropy defines similarity between two variables so that it has the property to identify normal and abnormal samples [10]. Based on this property, the recursive maximum correntropy criterion (RMCC) algorithm was presented [9], which is capable of dealing with impulsive noises.

In studying adaptive algorithms, there is interest to exploit the system sparsity. A sparse vector signifies that it only has a few non-zero entries such as the impulse responses of propagation channels in underwater acoustic and terrestrial communications [11]–[13]. With different sparsity-aware penalties (e.g., the l_1-norm, l_0-norm, and logarithmic based penalties), several sparsity-aware variants of the standard RLS algorithm were reported [14], [15], which reduce the steady-state misadjustment. Likewise, in the impulsive noise environment, sparsity-aware robust RLS-type algorithms were proposed, such as the sparsity-aware RLM algorithm [16] and the sparsity-aware RMCC algorithm [17]. However, it is worth noting that these algorithms have two drawbacks. First, they use the constant forgetting factor that controls the trade-off between the steady-state misadjustment and tracking capability for sudden changes of systems. Second, their performance depends on the sparsity-penalty parameter, which is usually chosen in a trial-and-error way, thus limiting their usefulness. In the literature, these two problems were separately considered for Gaussian noise scenarios. As such, variable forgetting factor (VFF) [18]–[20] and variable sparsity-penalty parameter (VSPP) [14] schemes have been developed. Nevertheless, they have not been considered jointly and for impulsive noise scenarios.

In this paper, our contributions are as follows. 1) We propose a unified S-RRLS framework, which covers different algorithms by applying straightforwardly the specified robustness criterion and sparsity-aware penalty. 2) By decoupling the effect of forgetting factor and sparsity-penalty parameter from each other, we derive adaptive rules for adjusting online these two parameters, and therefore the resulting jointly-optimized S-RRLS (JO-S-RRLS) algorithm reaches a low steady-state misadjustment and good tracking capability in impulsive noise.

The remaining part of this paper is organized as follows. Section II reviews the signal model and introduces the S-RRLS framework. In Section III, we present the JO-S-RRLS algorithm. Section IV shows simulations results. Conclusions are presented in Section V.

II. Signal Model and S-RRLS Algorithm

Consider a system identification problem with the input signal x_k at time k, then the output d_k of the system is

This work was supported by National Science Foundation of P.R. China (Nos. 61901400 and 61901285) and Doctoral Research Fund of Southwest University of Science and Technology (No. 19zx7122). The work of Y. Zakharov was supported in part by the U.K. EPSRC through Grants EP/R003297/1 and EP/V009591/1. Corresponding author: Lu Lu.

Y. Yu is with School of Information Engineering, Robot Technology Used for Special Environment Key Laboratory of Sichuan Province, Southwest University of Science and Technology, Mianyang, 621010, China (e-mail: yuyi_xyuan@163.com).

L. Lu is with School of Electronics and Information Engineering, Sichuan University, Chengdu, 610065, China (e-mail: lulu19900303@126.com).

Y. Zakharov is with the Department of Electronics, University of York, York YO10 5DD, U.K. (e-mail: yury.zakharov@york.ac.uk).

R. C. de Lamare is with CETUC, PUC-Rio, Rio de Janeiro 22451-900, Brazil. (e-mail: delamare@cetuc.puc-rio.br).

B. Chen is with Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, 710049, Shaanxi Province, China. (e-mail: chenbd@mail.xjtu.edu.cn).

described by

$$d_k = \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{w}^o + v_k, \tag{1}$$

where $\boldsymbol{w}^o \in \mathbb{R}^{M\times 1}$ is the impulse response of the sparse system to be identified, $\boldsymbol{x}_k = [x_k, x_{k-1}, ..., x_{k-M+1}]^{\mathrm{T}} \in \mathbb{R}^{M\times 1}$ is the input vector, and v_k is the noise. An estimate of $\boldsymbol{w}^o$ at time k, the vector of coefficients of adaptive filter, is denoted by $\boldsymbol{w}_k \in \mathbb{R}^{M\times 1}$. For updating the filter coefficients in Gaussian noise, the sparsity-aware RLS algorithm that exploits the sparsity of $\boldsymbol{w}^o$ is popular. However, the surroundings noise may also include impulsive samples which deteriorate the algorithm performance. For such scenarios, we introduce the unified robust sparsity-aware minimization problem:

$$\boldsymbol{w}_k = \arg\min_{\boldsymbol{w}_k} \left\{ \sum_{j=0}^{k} \lambda^{k-j} \varphi\left(d_j - \boldsymbol{x}_j^{\mathrm{T}} \boldsymbol{w}_k\right) + \rho f(\boldsymbol{w}_k) \right\}, \tag{2}$$

where $0 < \lambda < 1$ is the forgetting factor and $\varphi(\cdot) : \mathbb{R} \to \mathbb{R}$ is a robustness function that can deal with impulsive noise. In (2), $f(\cdot) : \mathbb{R}^{M\times 1} \to \mathbb{R}$ is a general sparsity-aware penalty function, and $\rho > 0$ is the sparsity-aware penalty parameter defining the weight of this penalty term. By solving the optimization in (2), we obtain the following normal equation:

$$\boldsymbol{R}_k \boldsymbol{w}_k = \boldsymbol{z}_k - \rho \triangledown f(\boldsymbol{w}_k), \tag{3}$$

where

$$\begin{aligned} \boldsymbol{R}_k &= \sum_{j=0}^{k} \lambda^{k-j} q_j \boldsymbol{x}_j \boldsymbol{x}_j^{\mathrm{T}} = \lambda \boldsymbol{R}_{k-1} + q_k \boldsymbol{x}_k \boldsymbol{x}_k^{\mathrm{T}} \\ \boldsymbol{z}_k &= \sum_{j=0}^{k} \lambda^{k-j} q_j d_j \boldsymbol{x}_j = \lambda \boldsymbol{z}_{k-1} + q_k d_n \boldsymbol{x}_k \end{aligned} \tag{4}$$

are the time-averaged autocorrelation matrix of $\boldsymbol{x}_k$ and the time-averaged crosscorrelation vector of d_k and $\boldsymbol{x}_k$, respectively, and $\triangledown f(\boldsymbol{w}_k)$ is the subgradient of $f(\boldsymbol{w}_k)$ with respect to $\boldsymbol{w}_k$. Also, $q_k = \varphi'(\epsilon_k)/\epsilon_k$, where $\epsilon_k = d_k - \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{w}_k$ is the *a posteriori* error and $\varphi'(\epsilon_k)$ is the derivative of $\varphi(\epsilon_k)$.

Applying the matrix inversion lemma [21], we have

$$\boldsymbol{P}_k \triangleq \boldsymbol{R}_k^{-1} = \frac{1}{\lambda}\left(\boldsymbol{P}_{k-1} - \boldsymbol{K}_k \boldsymbol{x}_k^T \boldsymbol{P}_{k-1}\right), \tag{5}$$

where

$$\boldsymbol{K}_k \triangleq \frac{q_k \boldsymbol{P}_{k-1} \boldsymbol{x}_k}{\lambda + q_k \boldsymbol{x}_k^T \boldsymbol{P}_{k-1} \boldsymbol{x}_k} \tag{6}$$

is the Kalman gain vector, and $\boldsymbol{P}_k$ is initialized by $\boldsymbol{P}_0 = \delta^{-1} \boldsymbol{I}_M$ with $\delta > 0$ being the regularization parameter. Then, substituting (5) and (6) into (3), the recursion for $\boldsymbol{w}_k$ is established:

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \boldsymbol{K}_k e_k - \rho \boldsymbol{P}_k [\triangledown f(\boldsymbol{w}_k) - \lambda \triangledown f(\boldsymbol{w}_{k-1})], \tag{7}$$

where $e_k = d_k - \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{w}_{k-1}$ is the *a priori* error. Naturally, the above recursion is not realizable, since q_k and $\triangledown f(\boldsymbol{w}_k)$ require knowing $\boldsymbol{w}_k$ at time k beforehand. To overcome this problem, it is assumed that q_k and $\triangledown f(\boldsymbol{w}_k)$ do not change considerably at adjacent time. Hence, we can rewrite (7) as

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \boldsymbol{K}_k e_k - \rho \boldsymbol{P}_k \triangledown f(\boldsymbol{w}_{k-1}), \tag{8}$$

where $(1 - \lambda)$ is absorbed into ρ that is $\rho \leftarrow \rho(1 - \lambda)$, and we can approximate q_k as

$$q_k \approx \varphi'(e_k)/e_k. \tag{9}$$

III. Proposed JO-S-RRLS Algorithm

Clearly, the S-RRLS algorithm performance depends on the parameters λ and ρ. Specifically, λ governs the trade-off between steady-state error and tracking capability through the Kalman gain vector. The parameter ρ must be properly chosen to ensure that the S-RRLS algorithm fully exploits the system's sparsity and thus outperforms the RRLS algorithm. In this section, we design adaptive schemes for λ and ρ, producing time-varying parameters λ_k and ρ_k. For this purpose, we rearrange (8) in two steps as

$$\boldsymbol{\psi}_k = \boldsymbol{w}_{k-1} + \boldsymbol{K}_k e_k, \tag{10a}$$

$$\boldsymbol{w}_k = \boldsymbol{\psi}_k - \rho_k \boldsymbol{P}_k \triangledown f(\boldsymbol{\psi}_k). \tag{10b}$$

The step (10a) plays the adaptive learning role of the RRLS algorithm; the step (10b) drives inactive coefficients in $\boldsymbol{\psi}_k$ to zero, thereby improving the performance of identifying sparse $\boldsymbol{w}^o$. Importantly, we also replace the original $\triangledown f(\boldsymbol{w}_{k-1})$ in (10b) with $\triangledown f(\boldsymbol{\psi}_k)$. In doing so, λ_k and ρ_k will be designed independently according to (10a) and (10b), respectively.

A. Design of ρ_k

By subtracting (10b) from $\boldsymbol{w}^o$, we obtain

$$\widetilde{\boldsymbol{w}}_k = \widetilde{\boldsymbol{\psi}}_k + \rho_k \boldsymbol{P}_k \triangledown f(\boldsymbol{\psi}_k), \tag{11}$$

where $\widetilde{\boldsymbol{w}}_k \triangleq \boldsymbol{w}^o - \boldsymbol{w}_k$ and $\widetilde{\boldsymbol{\psi}}_k \triangleq \boldsymbol{w}^o - \boldsymbol{\psi}_k$ represent the deviation vectors for the estimates $\boldsymbol{w}_k$ and $\boldsymbol{\psi}_k$. Taking the l_2-norm on both sides of (11), it is found that

$$\begin{aligned} ||\widetilde{\boldsymbol{w}}_k||_2^2 = ||\widetilde{\boldsymbol{\psi}}_k||_2^2 + \\ \underbrace{2\rho_k \widetilde{\boldsymbol{\psi}}_k^T \left(\boldsymbol{P}_k \triangledown f(\boldsymbol{\psi}_k)\right) + \rho_k^2 ||\boldsymbol{P}_k \triangledown f(\boldsymbol{\psi}_k)||_2^2}_{\triangle_k}. \end{aligned} \tag{12}$$

Minimizing $||\widetilde{\boldsymbol{w}}_k||_2^2$ with respect to ρ_k, the optimal ρ_k is obtained as

$$\rho_k^{\mathrm{opt}} = \frac{[\boldsymbol{\psi}_k - \boldsymbol{w}^o]^T \left(\boldsymbol{P}_k \triangledown f(\boldsymbol{\psi}_k)\right)}{||\boldsymbol{P}_k \triangledown f(\boldsymbol{\psi}_k)||_2^2}. \tag{13}$$

At time k, although we do not know the true $\boldsymbol{w}^o$ in (13), it can be approximated by the previous estimate $\boldsymbol{w}_{k-1}$ from (10b). Thus, we modify (13) to the following rule for choosing ρ_k:

$$\hat{\rho}_k^{\mathrm{opt}} = \max\left[\frac{[\boldsymbol{\psi}_k - \boldsymbol{w}_{k-1}]^T \left(\boldsymbol{P}_k \triangledown f(\boldsymbol{\psi}_k)\right)}{||\boldsymbol{P}_k \triangledown f(\boldsymbol{\psi}_k)||_2^2},\ 0\right]. \tag{14}$$

Also, since the estimate $\boldsymbol{w}_k$ at the early stage of the adaptation has a larger deviation as compared to $\boldsymbol{w}^o$, we enforce $\hat{\rho}_k^{\mathrm{opt}} = 0$ when $k \leq M/2$, to ensure the algorithm's convergence.

B. Design of λ_k

To derive the VFF scheme, we insert (6) into (10a) to yield

$$\boldsymbol{\psi}_k = \boldsymbol{w}_{k-1} + \frac{q_k \boldsymbol{P}_{k-1} \boldsymbol{x}_k}{\lambda_k + \theta_k} e_k, \tag{15}$$

where $\theta_k = q_k \boldsymbol{x}_k^T \boldsymbol{P}_{k-1} \boldsymbol{x}_k$. By defining the intermediate error $\xi_k = d_k - \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{\psi}_k$, we can find from (15) that

$$\xi_k = \left(1 - \frac{\theta_k}{\lambda_k + \theta_k}\right) e_k. \tag{16}$$

Inspired by [18], we propose to impose the condition

$$E\{\xi_k^2\} = \sigma_v^2 \tag{17}$$

on (16), which helps to recover the system noise in the intermediate error signal, where $E\{\cdot\}$ denotes the mathematical expectation. However, it is emphasized that σ_v^2 in (17) denotes the variance of the noise excluding impulsive noise samples, due to the fact that the negative influence of impulsive noises can be transferred to q_k given in (9) (see the following subsection C for more explanation). For solving (17), we introduce two assumptions: 1) the *a priori* error and input signals are uncorrelated, i.e., the orthogonality principle [21]; 2) the forgetting factor is deterministic at each time [18], [20]. As a result, the solution of (17) with respect to λ_k is given as

$$\lambda_k = \frac{\sigma_{\theta,k}\sigma_v}{\sigma_{e,k} - \sigma_v}, \tag{18}$$

where $\sigma_{e,k}^2 = E\{e_k^2\}$ and $\sigma_{\theta,k}^2 = E\{\theta_k^2\}$ are the variances of the corresponding signals. Both $\sigma_{e,k}^2$ and $\sigma_{\theta,k}^2$ could be estimated in a recursive way:

$$\hat{\sigma}_{e,k}^2 = \chi\hat{\sigma}_{e,k-1}^2 + (1-\chi)q_k^2 e_k^2, \tag{19a}$$
$$\hat{\sigma}_{\theta,k}^2 = \chi\hat{\sigma}_{\theta,k-1}^2 + (1-\chi)\theta_k^2, \tag{19b}$$

where q_k^2 aims to reduce the negative influence of the impulsive noise on the estimated variance $\hat{\sigma}_{e,k}^2$ and $\chi \in [0.9, 1)$ is a the smoothing factor. To estimate the variance σ_v^2 of the background noise, we extend the approach in [20] to impulsive noise environments, formulated as

$$\hat{\sigma}_{v,k}^2 = \frac{\hat{\sigma}_{d,k}^2 \hat{\sigma}_{e,k}^2}{\hat{\sigma}_{e,k}^2 + \hat{\sigma}_{y,k}^2}, \tag{20}$$

where the estimated powers $\hat{\sigma}_{d,k}^2$ and $\hat{\sigma}_{y,k}^2$ are calculated similar to (19a) as

$$\hat{\sigma}_{d,k}^2 = \chi\hat{\sigma}_{d,k-1}^2 + (1-\chi)q_k^2 d_k^2, \tag{21a}$$
$$\hat{\sigma}_{y,k}^2 = \chi\hat{\sigma}_{y,k-1}^2 + (1-\chi)y_k^2, \tag{21b}$$

with $y_k = \boldsymbol{x}_k^{\mathrm{T}} \boldsymbol{w}_{k-1}$ being the output of the adaptive filter. Due to using the power estimates, it could happen that $\hat{\sigma}_{e,k}^2 < \hat{\sigma}_{v,k}^2$; however, (18) shows that this situation has to be prevented. As such, we could set λ_k to $\lambda_{\max}$ which is close to one. Furthermore, in the steady-state $\hat{\sigma}_{e,k}^2$ may vary around $\hat{\sigma}_{v,k}^2$. Therefore, we propose a more practical solution for λ_k by imposing the convergence state $\hat{\sigma}_{e,k}^2 < \tau\hat{\sigma}_{v,k}^2$ with $\tau \in [1, 2]$, as follows:

$$\lambda_k = \begin{cases} \lambda_{\max}, \text{ if } \hat{\sigma}_{e,k}^2 < \tau\hat{\sigma}_{v,k}^2, \\ \min\left\{ \dfrac{\hat{\sigma}_{\theta,k}\hat{\sigma}_v}{|\hat{\sigma}_{e,k} - \hat{\sigma}_v| + \kappa}, \lambda_{\max} \right\}, \text{ otherwise}, \end{cases} \tag{22}$$

where κ is a small positive constant. This completes the VFF's derivation for the S-RRLS algorithm.

By equipping S-RRLS with the proposed $\hat{\rho}_k^{\text{opt}}$ and λ_k adaptations, we arrive at the JO-S-RRLS algorithm. In fact, the proposed $\hat{\rho}_k^{\text{opt}}$ and λ_k originate from the alternating optimization idea which is a powerful way to solve the challenging global optimization problems [22], [23].

Remark 1: As can be seen in (12), the term $\triangle_k$ results from the sparsity-aware step (10b). Only when $\triangle_k < 0$, the S-RRLS algorithm will work better than the RRLS algorithm. Accordingly, ρ must satisfy the inequality $0 < \rho < 2\rho_k^{\text{opt}}$. Following a similar derivation in Appendix D in [24], $\triangle_k < 0$ is likely to be true when identifying a sparse vector $\boldsymbol{w}^o$. It follows that the optimal ρ_k^{opt} given in (14) may exist. In the future, we will analyze the mean and mean-square behaviors of the JO-S-RRLS algorithm.

C. Practical considerations

For implementing the S-RRLS and JO-S-RRLS algorithms, two problems should be addressed.

Firstly, the robustness of the algorithms against impulsive noises relies on how to design the robustness function $\varphi(e_k)$ to further obtain q_k in (9). As an example, the modified M-estimator is used [4], i.e., $\varphi(e_k) = \begin{cases} e_k^2/2, \text{ if } |e_k| \le \xi, \\ \xi^2/2, \text{ if } |e_k| > \xi, \end{cases}$ such that $q_k = \begin{cases} 1, \text{ if } |e_k| \le \xi, \\ 0, \text{ if } |e_k| > \xi, \end{cases}$. It reveals that when $|e_k| > \xi$ holds (generally, when the impulsive noise happens), the Kalman gain $\boldsymbol{K}_k$ will be a zero vector due to $q_k = 0$, thus stopping the update of the adaptive filter. The threshold ξ is chosen as $\xi = \vartheta\hat{\sigma}_{\varepsilon,k}$, $\hat{\sigma}_{\varepsilon,k}^2 = \zeta\hat{\sigma}_{\varepsilon,k-1}^2 + c_\sigma(1-\zeta)\text{med}(\boldsymbol{a}_k^\varepsilon)$, where $\zeta \in [0.9, 1)$ is an exponentially weighting factor (except $\zeta = 0$ at time $k = 0$), the median operator $\text{med}(\cdot)$ is to remove data in the window $\boldsymbol{a}_k^\varepsilon = [e_k^2, e_{k-1}^2, ..., e_{k-N_w+1}^2]$ disturbed by impulsive noises, and $c_\sigma = 1.483(1 + 5/(N_w - 1))$ is the correction factor. Note that the window length N_w needs to be properly chosen. Larger N_w provide a more robust estimate $\hat{\varepsilon}_{e,k}^2$ but require a higher complexity [4]. Also, the value of ϑ is often chosen as 2.576 [4].

Secondly, to effectively characterize the sparsity of systems, how to design $f(\cdot)$ in (2) is also a key factor. Here we choose the popular log-penalty $f(\boldsymbol{w}_k) = \sum_{m=1}^{M} \ln(1 + |w_{m,k}|/\mu)$ [24], where $w_{m,k}$ denotes the m-th entry of $\boldsymbol{w}_k$, and $\mu > 0$ denotes the shrinkage factor that helps to distinguish non-zero and zero entries. Thus, the entries of $\nabla(\boldsymbol{\psi}_{m,k})$ in (10b) are given by $\nabla(\psi_{m,k}) = \frac{\text{sgn}(\psi_{m,k})}{\mu + |\psi_{m,k}|}, m = 1, ..., M$.

It needs to point out that other robust criteria [5]–[9], [25] and sparsity-aware penalties [13], [14], [23], [24] can also be applied to present different JO-S-RRLS algorithms. However, discussing the effects of different choices of $\varphi(e_k)$ or $f(\boldsymbol{w}_k)$ is not the focus of this paper.

Remark 2: Compared with the original S-RRLS algorithm, the extra computational complexity of the JO-S-RRLS algorithm stems from the adaptations of ρ_k and λ_k, which requires $M^2 + 3M + 13$ multiplications, $M^2 + 2M + 6$ additions, 3 divisions, and 3 square-roots per iteration.

IV. Simulation Results

In this section, simulations are presented to evaluate the proposed JO-S-RRLS algorithm. It is assumed that the length of the adaptive filter is the same as that of the unknown sparse vector $\boldsymbol{w}^o$. The input signal x_k is generated by filtering a zero-mean white Gaussian random process ν_k with unit variance through a second-order autoregressive model $x_k = 0.4x_{k-1} - 0.4x_{k-2} + \nu_k$. The normalized mean square deviation is used as the performance measure, defined as $\text{NMSD}_k = 10\log_{10}(||\boldsymbol{w}_k - \boldsymbol{w}^o||_2^2/||\boldsymbol{w}^o||_2^2)$. All the curves are obtained by averaging the results over 100 independent runs.

Case 1: The sparse vector $\boldsymbol{w}^o$ has $M = 64$ elements, and its non-zero elements follow from a zero-mean Gaussian distribution with variance $1/\sqrt{Q}$ and their positions are randomly selected from the binomial distribution, with Q being the number of non-zero elements. A smaller Q means sparser $\boldsymbol{w}^o$. The vector $\boldsymbol{w}^o$ cardinality changes from $Q = 4$ to $Q = 8$ at time $k = 1501$. The noise with impulsive behavior v_k is drawn from the contaminated-Gaussian (CG) process, i.e., $v_k = v_k^{\text{g}} + v_k^{\text{i}}$. Specifically, v_k^{g} is zero-mean white Gaussian noise, with variance σ_{g}^2 given by the signal-to-noise ratio of 30 dB. The impulsive noise v_k^{i} is described as $v_k^{\text{i}} = b_k\eta_k$, where b_k follows from the Bernoulli distribution with the probabilities $P\{b_k = 1\} = p$ and $P\{b_k = 0\} = 1 - p$, and η_k is also zero-mean white Gaussian noise with a large variance of $\sigma_\eta^2 = 1000E\{(\boldsymbol{x}_k^{\text{T}}\boldsymbol{w}^o)^2\}$. Fig. 1 compares the proposed S-RRLS and JO-S-RRLS algorithms with the existing RLS, S-RLS, and RLM algorithms. To fairly evaluate them, we set $\lambda = 0.995$ and $\delta = 0.5$ for all the algorithms, the M-estimate parameter $N_w = 9$ and $\zeta = 0.99$ for all the robust algorithms, the log-penalty parameter $\mu = 0.01$ for all the sparsity-aware algorithms. As expected, the RLS and S-RLS algorithms are not suitable for impulsive noise scenarios due to the performance degradation, while other algorithms show good robustness. Benefited from the sparsity-aware step, S-RRLS algorithm reduces the steady-state misadjustment in sparse systems as compared to the RLM algorithm. Also, by using the proposed adaptation of ρ, the S-RRLS algorithm with ρ_k^{opt} avoids the choice problem of ρ. To deal with the sudden change of $\boldsymbol{w}^o$, we may also resort to the reset (Rs) technique[1], namely, the S-RRLS with ρ_k^{opt} and Rs algorithm recovers the tracking capability. Importantly, by jointly optimizing ρ and λ, the proposed JO-S-RRLS algorithm not only obtains a low steady-state misadjustment but also good tracking capability.

Case 2: The sparse vector $\boldsymbol{w}^o$ is the echo channel 1 from the ITUT G.168 standard, with $M = 256$ taps [27]. The noise with impulsive behavior v_k follows from the symmetric α-stable random process, called the α-stable noise. Its characteristic function [1] is given as $\phi(t) = \exp(-\gamma|t|^\alpha)$, where $\alpha \in (0, 2]$ describes the impulsiveness of the noise (smaller α corresponds to stronger impulsive noises) and $\gamma > 0$ is similar to the variance of the noise. Note that when $\alpha = 2$, it reduces to the Gaussian distribution. In Fig 2, we set $\alpha = 1.65$ and $\gamma = 0.02$. As one can see, the proposed JO-S-RRLS algorithm achieves the best performance among these algorithms, since it optimizes the parameters ρ and λ simultaneously.

[1]Here the Rs technique modifies the one in [26] due to the presence of impulsive noises, i.e., if $\log(q_k^2 e_k^2/e_{\text{avr},k-1}^2) > 1.5$, we reset $\boldsymbol{P}_k$ with $\boldsymbol{P}_0$, where $e_{\text{avr},k-1}^2 = te_{\text{avr},k-1}^2 + (1-t)q_k^2 e_k^2$ with $t = 0.98$.

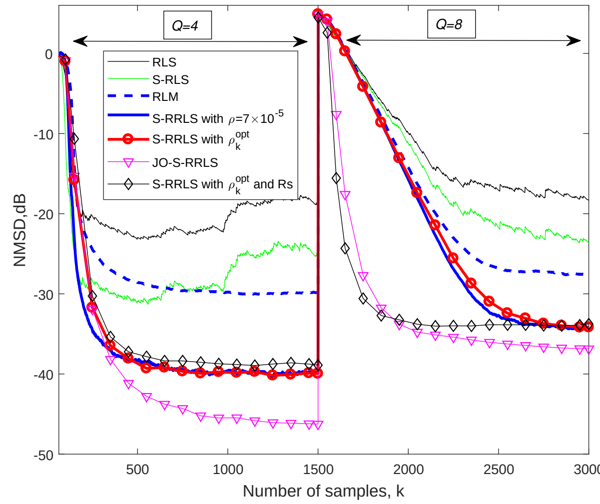

Fig. 1. NMSD performance of RLS-type algorithms in the CG noise with $p = 0.001$. The VFF's parameters are set to $\lambda_{\max} = 0.99999, \chi = 0.96$, and $\tau = 1.5$.

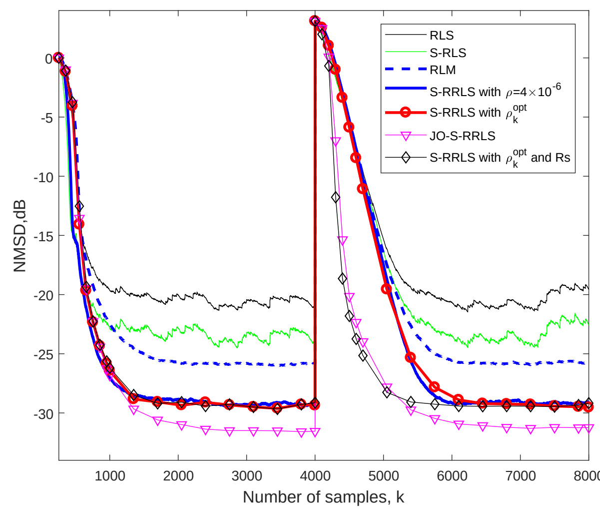

Fig. 2. NMSD performance of RLS-type algorithms in the α-stable noise. Some parameters are retuned as: $\lambda = 0.977$, $N_w = 16$, and $\mu = 0.001$ for all the algorithms.

V. Conclusion

In this work, a unified S-RRLS algorithm update was presented, which aims at identifying sparse systems in impulsive noise. By replacing straightforwardly the specified robustness criterion and sparsity-aware penalty, it can lead to different S-RRLS algorithms. We then developed adaptive schemes for both the forgetting factor and the sparsity penalty parameter in the S-RRLS algorithm, thus arriving at the JO-S-RRLS algorithm with a further improved performance in terms of the steady-state misadjustment and tracking capability. Simulations in various impulsive noise scenarios have been conducted to verify the effectiveness of the proposed algorithms.

References

[1] C. L. Nikias and M. Shao, *Signal processing with Alpha-stable distributions and applications*. Wiley-Interscience, 1995.

[2] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Transactions on Electromagnetic Compatibility*, vol. 44, no. 1, pp. 249–258, 2002.

[3] 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 Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 7, pp. 1359–1363, 2020.

[4] 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 Transactions on Signal Processing*, vol. 52, no. 4, pp. 975–991, 2004.

[5] Á. Navia-Vazquez and J. Arenas-Garcia, "Combination of recursive least p-norm algorithms for robust adaptive filtering in alpha-stable noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1478–1482, 2012.

[6] H. Zayyani, "Continuous mixed p-norm adaptive algorithm for system identification," *IEEE Signal Processing Letters*, vol. 21, no. 9, pp. 1108–1110, 2014.

[7] L. Lu, W. Wang, X. Yang, W. Wu, and G. Zhu, "Recursive Geman-Mcclure estimator for implementing second-order Volterra filter," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 7, pp. 1272–1276, 2019.

[8] L. Lu, H. Zhao, and B. Chen, "Improved-variable-forgetting-factor recursive algorithm based on the logarithmic cost for Volterra system identification," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 63, no. 6, pp. 588–592, 2016.

[9] H. Radmanesh and M. Hajiabadi, "Recursive maximum correntropy learning algorithm with adaptive kernel size," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 7, pp. 958–962, 2018.

[10] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Príncipe, "Generalized correntropy for robust adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3376–3387, 2016.

[11] J. Radecki, Z. Zilic, and K. Radecka, "Echo cancellation in IP networks," in *The 2002 45th Midwest Symposium on Circuits and Systems, 2002. MWSCAS-2002.*, vol. 2, 2002, pp. II–II.

[12] W. F. Schreiber, "Advanced television systems for terrestrial broadcasting: Some problems and some proposed solutions," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 958–981, 1995.

[13] Y. V. Zakharov and V. H. Nascimento, "DCD-RLS adaptive filters with penalties for sparse identification," *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3198–3213, 2013.

[14] E. M. Eksioglu and A. K. Tanc, "RLS algorithm with convex regularization," *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 470–473, 2011.

[15] E. M. Eksioglu, "Sparsity regularised recursive least squares adaptive filtering," *IET Signal Processing*, vol. 5, no. 5, pp. 480–487, 2011.

[16] K. Pelekanakis and M. Chitre, "Adaptive sparse channel estimation under symmetric alpha-stable noise," *IEEE Transactions on Wireless Communications*, vol. 13, no. 6, pp. 3183–3195, 2014.

[17] W. Ma, J. Duan, B. Chen, G. Gui, and W. Man, "Recursive generalized maximum correntropy criterion algorithm with sparse penalty constraints for system identification," *Asian Journal of Control*, vol. 19, no. 3, pp. 1164–1172, 2017.

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

[19] M. Z. A. Bhotto and A. Antoniou, "New improved recursive least-squares adaptive-filtering algorithms," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 6, pp. 1548–1558, 2012.

[20] C. Paleologu, J. Benesty, and S. Ciochină, "A practical variable forgetting factor recursive least-squares algorithm," in *The 11th International Symposium on Electronics and Telecommunications (ISETC)*. IEEE, 2014, pp. 1–4.

[21] A. H. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, 2003.

[22] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.

[23] R. C. de Lamare and R. Sampaio-Neto, "Sparsity-aware adaptive algorithms based on alternating optimization and shrinkage," *IEEE Signal Processing Letters*, vol. 21, no. 2, pp. 225–229, 2014.

[24] Y. Yu, T. Yang, H. Chen, R. C. de Lamare, and Y. Li, "Sparsity-aware ssaf algorithm with individual weighting factors: Performance analysis and improvements in acoustic echo cancellation," *Signal Processing*, vol. 178, p. 107806, 2021.

[25] V. Roth, "The generalized LASSO," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 16–28, 2004.

[26] L. Shi, H. Zhao, W. Wang, and L. Lu, "Combined regularization parameter for normalized LMS algorithm and its performance analysis," *Signal Processing*, vol. 162, pp. 75–82, 2019.

[27] *Digital Network Echo Cancellers Recommendation*, Std. ITU-TG.168 (V8), 2015.