Robust Parameter and Scale Factor Estimation in Nonstationary and Impulsive Noise Environment

Zoran Đ. Banjac, Branko D. Kovačević

Abstract— The problem of FIR system parameter estimation has been considered in the paper. A new robust recursive algorithm for simultaneously estimation of parameters and scale factor of prediction residuals in non-stationary environment corrupted by impulsive noise has been proposed. The performance of derived algorithm has been tested by simulations.

Keywords—Adaptive filtering, Non-Gaussian filtering, Robust estimation, Scale factor estimation.

I. INTRODUCTION

ADAPTIVE filtering provides a powerful approach to many signal processing problems. Most signal processing algorithms, developed for adaptive filtering, are based on the assumption that the interfering noise is Gaussian distributed. Recursive least squares (RLS) algorithm is one of them, and has been used extensively for its good convergence property and small mean square error. However, sparse impulses arise frequently in many physical environments. Namely, impulsive noise is relatively short duration pulses, characterized by heavy-tailed non-Gaussian distribution. A such type of disturbances is typical for communications system, atmospheric environment, relay switching systems etc. Impulsive noise highly depends on physical environment and may be non-stationary, which makes it impossible to obtain accurate statistical description. It is also well known that the standard gradient based adaptive algorithms degrade badly when the filter is subjected to input signals, like additive noise corrupted by impulsive interference, represented by unexpected number of large noise values that are generated by non-Gaussian distribution. Namely, the standard estimation procedures minimize the sum of squared residuals, and weight all prediction residuals equally, so that small changes from assumed nominal normal model in noise distribution can produce great changes in estimates. Possible treatments for parameter estimation problem in impulsive noise environment

include nonlinear transformations of residuals, and such estimators are called robust. Some of the most popular robust estimators are M-estimators (the M stands for approximate maximum-likelihood) first introduced by Huber [1]. Robust version of RLS algorithm, named robust recursive least square (RRLS) algorithm minimizes the sum of weighted residuals [2, 3]. In the RRLS procedures the assigned weight is a function of prediction residuals, and a convex cost function is selected to assign more weight to the bulk of small residuals while down weighting the small number of large residuals.

One important part of the RRLS algorithm is an estimate of the scale factor s. Traditionally, the scale is used to make robust algorithm invariant to noise level. It should reflect the minimum mean square error, and provides robustness to shorter burst disturbances, together with adequate tracking of longer changes of the residual error. A popular robust estimate of s is the median of absolute median deviations [2, 4]

$$s = median\left(\frac{|d(i) - median(d(i))|}{0.6745}\right), \ i = 1, 2, ..., M, \qquad (1)$$

where M is the frame, or data window, size. However, RRLS algorithm using (1) found to be sensitive to abrupt noise level changes.

In order to overcome this difficulty, we proposed a RRLS based algorithm with the new scheme for scale factor estimation.

II. REVIEW OF ROBUST RECURSIVE LEAST SQUARE ALGORITHM

In FIR system identification applications, the error signal at time k, representing the difference between the unknown system and adaptive filter outputs is given by:

$$e(k) = y(k) - \hat{\mathbf{H}}(k)^T \mathbf{X}(k), \qquad (2)$$

where $\hat{\mathbf{H}} = [h_0 \ h_1 ... h_L]^T$ is the parameter vector, the superscript T denotes the transpose of vector, and $\mathbf{X}(k) = [x(k) \ x(k-1) \ ... \ x(k-L)]^T$ is a vector containing the last $L+1$ samples of input signal x [5, 6].

The unknown system parameter vector $\mathbf{H}$ is estimated at each step by minimizing the performance criterion J:

$$J(\hat{\mathbf{H}}) = E\left\{\varphi\left[e^2(k)\right]\right\}, \qquad (3)$$

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:1, No:2, 2007

where $\varphi(\cdot)$ is the Huber's robust score or loss function which has to suppress the influence of impulsive noise [1, 4]:

$$\varphi(x) = \begin{cases} \dfrac{m|x|}{\sigma} + c_1, & |x| \geq m\sigma \\[2mm] \dfrac{x^2}{2\sigma^2} + c_2, & |x| < m\sigma \end{cases} \qquad (4)$$

In (4) standard deviation is denoted by σ, m is an efficiency-tuning constant, which has to be chosen so to give the desired efficiency at the assumed nominal Gaussian model, and $c_{1,2}$ are some constants. The derivate of loss function is known in the literature as influence function [1, 4]

$$\psi(x) = \min\left\{\frac{|x|}{\sigma^2}, \frac{m}{\sigma}\right\}\operatorname{sign}(x); \quad \operatorname{sign}(x) = \begin{cases} 1; & x > 0 \\ 0; & x = 0 \\ -1; & x < 0 \end{cases} \qquad (5)$$

To derive a recursive form of the proposed robust LS procedure, instead of criterion (3) the empirical criterion can be used:

$$J_i(\hat{\mathbf{H}}) = i^{-1}\sum_{k=1}^{i}\varphi\big[e(k,\hat{\mathbf{H}})\big]. \qquad (6)$$

Under certain conditions, with i growing J_i converges to J in (3). Furthermore, one can resort to the approximate Newton-Raphson type method for solving a set of nonlinear equations, which result from the optimality condition $\partial J_i(\mathbf{H})/\partial\mathbf{H} = 0$, where $\partial(\cdot)/\partial\mathbf{H}$ denotes the partial derivative operator, [5, 6]. Using this notation, and applying the Newton-Raphson algorithm for solving iteratively the mentioned equation, one can obtain the robust recursive least-squares (RRLS) algorithm

$$\hat{\mathbf{H}}(k) = \hat{\mathbf{H}}(k-1) + \mathbf{P}(k)\mathbf{X}(k)\psi\big[e(k)\big];$$

$$\mathbf{P}(k) = \mathbf{P}(k-1) - \frac{\mathbf{P}(k-1)\mathbf{X}(k)\mathbf{X}^T(k)\mathbf{P}(k-1)\psi'[e(k)]}{1+\psi'[e(k)]\mathbf{X}^T(k)\mathbf{P}(k-1)\mathbf{X}(k)}. \qquad (7)$$

However, the standard deviation in (5) is usually unknown and must be estimated.

A commonly used sample variance cannot be used for this purpose, since it is highly sensitive to impulsive noise. A better choice is to use the median of absolute median deviation in (1). An alternative approach that provides the simultaneous estimation of FIR filter parameters and scaling factor has been proposed in the next section.

III. Recursive Parameter and Scale Factor Estimation

Assume that probability density function (pdf) of stochastic disturbance $n(k)$ is known with accuracy up to scale factor s. If we denote the pdf for $s=1$ by $\bar{p}(n)$, then the pdf for arbitrarily chosen scale factor is given by:

$$p(n) = \frac{1}{s}\bar{p}\left(\frac{n}{s}\right). \qquad (8)$$

In the case when a real noise pdf belongs to the given class $\mathscr{P}$, it is necessary to minimize the Cramer-Rao bound in order to determine the least favorable pdf, $\bar{p}_*(n)$ [1÷4]. The choice of the optimal loss function for the worst-case pdf $\bar{p}_*(n)$ is given by the maximum-likelihood score function:

$$F^*\big(e(k,\mathbf{H}),s\big) = -\ln\frac{1}{s}\bar{p}_*\left(\frac{e(k,\mathbf{H})}{s}\right) = \ln s + \varphi\left(\frac{e(k,\mathbf{H})}{s}\right) \qquad (9)$$

and the robust performance index is given by the mathematical expectation of (9), that is

$$J(\mathbf{H},s) = \ln s + E\left\{\varphi\left(\frac{e(k,\mathbf{H})}{s}\right)\right\} \qquad (10)$$

Starting from (10), one can obtain an empirical approximation:

$$J_k(\mathbf{H},s) = \ln s + \frac{1}{k}\sum_{i=1}^{k}\varphi\left(\frac{e(i,\mathbf{H})}{s}\right). \qquad (11)$$

Linearizing $J_k(\mathbf{H},s)$ in the vicinity of the preceding estimate $\hat{\mathbf{H}}(k-1)$ and using Newton-Raphson procedure to minimize iteratively the obtained linearised criterion, it can be written

$$\hat{\mathbf{H}}(k) = \hat{\mathbf{H}}(k-1) - \left[k\frac{\partial^2 J_k\big(\hat{\mathbf{H}}(k-1),s\big)}{\partial\mathbf{H}^2}\right]^{-1} \times$$
$$\times\left[k\frac{\partial J_k\big(\hat{\mathbf{H}}(k-1),s\big)}{\partial\mathbf{H}}\right] + O\big(\|\hat{\mathbf{H}}(k)-\hat{\mathbf{H}}(k-1)\|\big). \qquad (12)$$

Here $\lim\limits_{\|x\|\to\infty}\dfrac{O(\|x\|)}{\|x\|}=0$, and $\|\cdot\|$ denotes the Euclidean norm. With respect to (11), we have

$$J_k(\mathbf{H},s) = \ln s + \frac{1}{k}\left[(k-1)J_{k-1}(\mathbf{H},s) - (k-1)\ln s + \varphi\left(\frac{e(k,\mathbf{H})}{s}\right)\right]. \qquad (13)$$

From (11) further follows:

$$kJ_k(\mathbf{H},s) = \ln s + (k-1)J_{k-1}(\mathbf{H},s) + \varphi\left(\frac{e(k,\mathbf{H})}{s}\right) \qquad (14)$$

By differentiating (15) twice with respect to $\mathbf{H}$, one obtains

$$k\frac{\partial J_k(\mathbf{H},s)}{\partial\mathbf{H}} = (k-1)\frac{\partial J_{k-1}(\mathbf{H},s)}{\partial\mathbf{H}} - \frac{\mathbf{X}(k)}{s}\Psi\left(\frac{e(k,\mathbf{H})}{s}\right) \qquad (15)$$

$$k\frac{\partial^2 J_k(\mathbf{H},s)}{\partial\mathbf{H}^2} = (k-1)\frac{\partial^2 J_{k-1}(\mathbf{H},s)}{\partial\mathbf{H}^2} + \frac{\mathbf{X}(k)\mathbf{X}^T(k)}{s^2}\psi'\left(\frac{e(k,\mathbf{H})}{s}\right) \qquad (16)$$

where $\psi = \varphi'$.

Taking $\mathbf{H} = \hat{\mathbf{H}}(k-1)$ and $s = \hat{s}(k-1)$, we have

$$k\frac{\partial^2 J_k\big(\hat{\mathbf{H}}(k-1),\hat{s}(k-1)\big)}{\partial\mathbf{H}^2} = (k-1)\bar{R}(k) + \frac{\mathbf{X}(k)\mathbf{X}^T(k)}{\hat{s}^2(k-1)}\psi'\left(\frac{e(k,\mathbf{H}(k-1))}{\hat{s}(k-1)}\right) \qquad (17)$$

World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:1, No:2, 2007

where $\bar{R}(k) = k \dfrac{\partial^2 J_k \left(\hat{\mathbf{H}}(k-1), \hat{s}(k-1) \right)}{\partial \mathbf{H}^2}$.

By introducing the following hypothesis:

H1: The estimate $\hat{\mathbf{H}}(k)$ is in the vicinity of the estimate $\hat{\mathbf{H}}(k-1)$, yielding: $O\left(\left\| \hat{\mathbf{H}}(k) - \hat{\mathbf{H}}(k-1) \right\| \right) = 0$

H2: The estimate $\hat{\mathbf{H}}(k-1)$ is optimal at the instant $k-1$, yielding $\dfrac{\partial J_{k-1} \left(\hat{\mathbf{H}}(k-1), \ \hat{s}(k-1) \right)}{\partial \mathbf{H}} = 0$;

the relation (12) becomes

$$\hat{\mathbf{H}}(k) = \hat{\mathbf{H}}(k-1) - \bar{R}(k)^{-1} \left[k \frac{\partial J_k \left(\hat{\mathbf{H}}(k-1), \hat{s}(k-1) \right)}{\partial \mathbf{H}} \right] , \quad (18)$$

$$\bar{R}(k) = \bar{R}(k-1) + \psi' \left(\frac{e(k, \hat{\mathbf{H}}(k-1))}{\hat{s}(k-1)} \right) \frac{\mathbf{X}(k) \mathbf{X}^T(k)}{\hat{s}^2(k-1)} . \quad (19)$$

Taking into account (18) and (15) for $\mathbf{H} = \hat{\mathbf{H}}(k-1)$ and $s = \hat{s}(k-1)$, and using hypothesis H2, it follows

$$\hat{\mathbf{H}}(k) = \hat{\mathbf{H}}(k-1) + \bar{R}(k)^{-1} \left[\psi \left(\frac{e(k, \hat{\mathbf{H}}(k-1))}{\hat{s}(k-1)} \right) \frac{\mathbf{X}(k)}{\hat{s}(k-1)} \right] . \quad (20)$$

Similarly to (12), one can define the recurrent Newton-Raphson estimation procedure.

$$\hat{s}(k) = \hat{s}(k-1) - \left[k \frac{\partial^2 J_k \left(\hat{\mathbf{H}}(k-1), s(k-1) \right)}{\partial s^2} \right]^{-1} \times$$
$$\times \left[k \frac{\partial J_k \left(\hat{\mathbf{H}}(k-1), s(k-1) \right)}{\partial s} \right]. \quad (21)$$

By differentiating (14) twice with respect to s, for $\mathbf{H} = \hat{\mathbf{H}}(k-1)$ and $s = \hat{s}(k-1)$, as using the hypothesis that $\hat{s}(k-1)$ is optimal at the instant $k-1$, and introducing $\gamma(k) = k \dfrac{\partial^2 J_k \left(\hat{\mathbf{H}}(k-1), \hat{s}(k-1) \right)}{\partial s^2}$ one obtains further:

$$\hat{s}(k) = \hat{s}(k-1) - \frac{1}{\gamma(k) \hat{s}^2(k-1)} \times$$
$$\times \left[\hat{s}(k-1) - e\left(k, \hat{\mathbf{H}}(k-1) \right) \psi \left(\frac{\left(k, \hat{\mathbf{H}}(k-1) \right)}{\hat{s}(k-1)} \right) \right] \quad (22)$$

$$\gamma(k) = \gamma(k-1) - \frac{1}{\hat{s}^2(k-1)} + \frac{2e(k, \hat{\mathbf{H}}(k-1))}{\hat{s}^3(k-1)} \psi \left(\frac{e(k, \hat{\mathbf{H}}(k-1))}{\hat{s}(k-1)} \right)$$
$$+ \frac{e^2(k, \hat{\mathbf{H}}(k-1))}{\hat{s}^4(k-1)} \psi' \left(\frac{e(k, \hat{\mathbf{H}}(k-1))}{\hat{s}(k-1)} \right) \quad (23)$$

The proposed algorithm is very difficult to be implemented and we need further approximation in order to obtain simple and efficient algorithm. Thus, if one assumes that the estimate $\hat{\mathbf{H}}(k-1)$ and $\hat{s}(k-1)$, for k large enough, are close to the optimal solution $\mathbf{H}_{opt}$ and s_{opt}, and that:

$$\lim_{k \to \infty} \frac{\partial^2 J_k \left(\hat{\mathbf{H}}(k-1), \hat{s}(k-1) \right)}{\partial s^2} = \frac{\partial^2 J \left(\mathbf{H}_{opt}, s_{opt} \right)}{\partial s^2} , \quad (24)$$

we have

$$\hat{s}(k) = \hat{s}(k-1) - \left[k \frac{\partial^2 J \left(\mathbf{H}_{opt}, s_{opt} \right)}{\partial s^2} \right]^{-1} \left[k \frac{\partial J_k \left(\mathbf{H}_{opt}, s_{opt} \right)}{\partial s} \right] \quad (25)$$

where:

$$\frac{\partial J \left(\mathbf{H}_{opt}, s_{opt} \right)}{\partial s} = \frac{1}{s_{opt}} - E\left\{ \frac{e(k, \mathbf{H}_{opt})}{s_{opt}^2} \psi \left(\frac{e(k, \mathbf{H}_{opt})}{s_{opt}} \right) \right\} ; \quad (26)$$

$$\frac{\partial^2 J \left(\mathbf{H}_{opt}, s_{opt} \right)}{\partial s^2} = \frac{-1}{s_{opt}^2} + E\left\{ \psi' \left(\frac{e(k, \mathbf{H}_{opt})}{s_{opt}} \right) \frac{e^2(k, \mathbf{H}_{opt})}{s^4} \right\} +$$
$$+ 2E\left\{ \psi \left(\frac{e(k, \mathbf{H}_{opt})}{s_{opt}} \right) \frac{e(k, \mathbf{H}_{opt})}{s^3} \right\} \quad (27)$$

Since $\dfrac{\partial J \left(\mathbf{H}_{opt}, s_{opt} \right)}{\partial s} = 0$ for $\mathbf{H} = \mathbf{H}_{opt}$ and $s = s_{opt}$,

$$\frac{\partial^2 J \left(\mathbf{H}_{opt}, s_{opt} \right)}{\partial s^2} = \frac{1}{s^2} + \frac{1}{s^4} E\left\{ \psi' \left(\frac{e(k, \mathbf{H}_{opt})}{s_{opt}} \right) e^2(k, \mathbf{H}_{opt}) \right\} \quad (28)$$

yielding

$$\hat{s}(k) = \hat{s}(k-1) - \frac{a}{k} \left[\hat{s}(k-1) - e\left(k, \hat{\mathbf{H}}(k-1) \right) \Psi \left(\frac{e\left(k, \hat{\mathbf{H}}(k-1) \right)}{\hat{s}(k-1)} \right) \right] \quad (29)$$

where $a = \dfrac{1}{\left[I_d(\bar{p}_*) - 1 \right]}$, with $E\left\{ \left(\dfrac{\bar{p}_*'(n)}{\bar{p}_*(n)} \right)^2 n^2 \right\} = I_d(\bar{p}_*)$

Finally, the algorithm for simultaneous estimation of FIR filter parameters and scale factor is derived by (19), (20) and (29).

IV. SIMULATION RESULTS

To demonstrate the performance of proposed algorithm, a FIR system identification simulation reported in [7] is used. The desired filter response signal $d(k)$ is formed by inputting white normal noise $x(k)$ of unit power to the ninth order FIR filter with the coefficients $\mathbf{H} = [0.1, 0.2, 0.3, 0.4, 0.5, 0.4, 0.3, 0.2, 0.1]$, and independent normal noise $n(k)$ of variable variance is added to its output. The variance of $n(k)$ is chosen in such a way that in the first segment the SNR is 30 dB, then 10 dB, and at the last one 20 dB, before the addition of the impulsive noise component (Fig 1.a). The impulsive noise is generated from the model: $n(k) = a(k)A(k)$ where $a(k)$ is a binary independent identically distributed occurrence process with the probabilities $P(a(k) = 1) = 0.01$ and $P(a(k) = 0) = 0.99$, whereas $\text{var}\{A(k)\} = 10^4 / 12$. Fig. 1 depicts one realization of this additive noise without (Fig 1.a) and with (Fig 1.b) impulsive contamination.
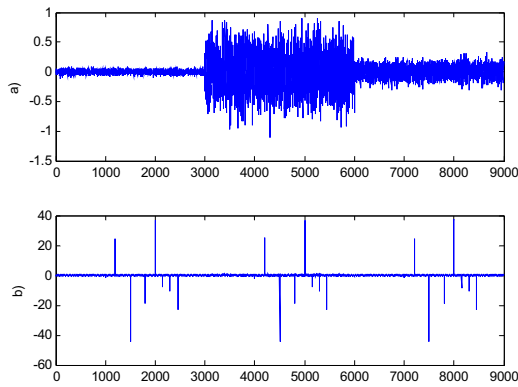
World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:1, No:2, 2007

Fig.1 Additive noise a) pure Gaussian b) corrupted by impulsive noise

The simulation results are compared in terms of normalized estimation error (*NEE*), defined by [6]

$$10\log\frac{\left\|\hat{\mathbf{H}}(n)-\mathbf{H}\right\|^2}{\left\|\mathbf{H}\right\|^2} \tag{30}$$
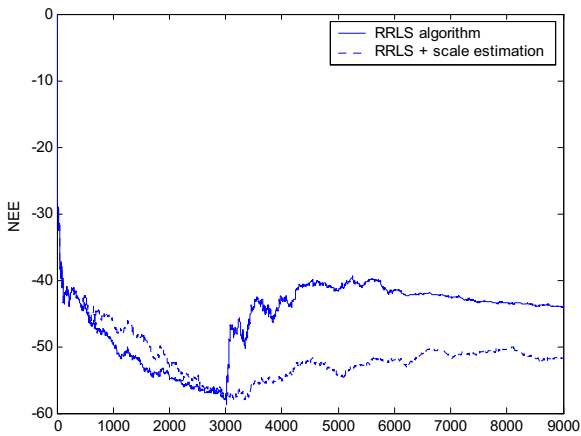

Fig.2 Normalized estimation error in Gaussian noise environment for different algorithms

In the situations without impulsive noise (see Fig. 2) the two mentioned robust algorithms start estimation process on the similar way. At the point of additive noise variance changing the standard RRLS algorithm degrades its performance, producing poor variance estimation. On the other hand, proposed robust algorithm with scale factor estimation is less sensitive to the changes of variance values.

Fig 3 has shown the obtained results in the case of simulation with impulsive noise component added to the Gaussian sequence.

The presented results have shown that the classical RLS algorithm is very sensitive to the impulsive noise presence. The RRLS and RRLS algorithm with adaptive scale factor estimation are insensitive to impulsive noise component, but the last algorithm has better performance in a non-stationary noise environment.
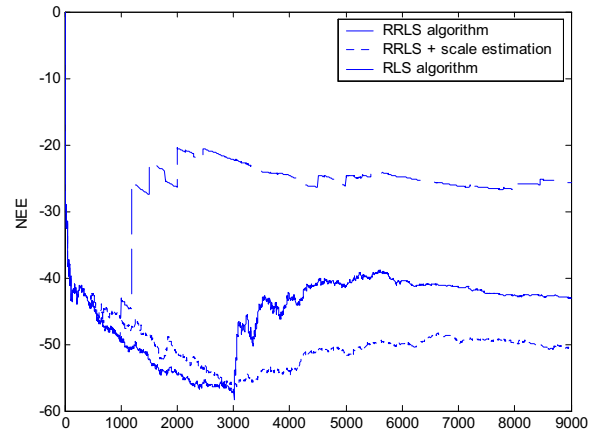

Fig.3 Normalized estimation error in impulsive noise environment for different algorithms

V. CONCLUSION

The FIR system identification problem in the presence of non-stationary additive noise contaminated by impulsive noise component has been considered in the paper. The new algorithm for simultaneously robust estimation of filter parameters and the scale factor has been proposed. This algorithm represents a combination of the well-known robust recursive least square algorithm and adaptive scheme for scale factor estimation. The performance of the proposed algorithm has been tested by simulation. The obtained results have shown that the proposed algorithm gives similar performance to the commonly used recursive LS algorithm in the case of stationary pure Gaussian additive noise. On the other hand, in the presence of impulsive noise the proposed algorithm behaves similarly to the robust recursive LS algorithm, derived in the literature. However, the proposed algorithm is more efficient in the case of non-stationary additive noise, compared to the robust recursive LS algorithm.

REFERENCES

[1] P. J. Huber, *Robust Statistic,* New York, Willey, 1981
[2] Z. Banjac, B. Kovačević, M. Veinović, M. Milosavljević, "Robust least mean square adaptive FIR filter algorithm", *IEE Proceedings - Vision, Image, and Signal Processing*, Vol.148, No.5, 2001, pp. 332-336.
[3] Z. Banjac, B. Kovačević, M. Milosavljević, M. Veinović, "Local echo canceller with optimal input design for true full-duplex speech scrambling system", *IEEE Transaction on Signal Processing*; Vol.50, No.8, 2002, pp.1877-1882.
[4] Barnet, T. Lewis, *Outliers in Statistical Data,* New York, Willey, 1978
[5] L. Ljung, T. Soderstrom, Theory and Practice of Recursive identification, MA: MIT Press, 1983
[6] S. Hykin, *Adaptive Filter Theory,* NY, Prentice Hall, 1991
[7] J. Chambers, A. Avlonitis, "A Robust Mixed Norm Adaptive Filter Algorithm," *IEEE Signal Processing. Letters*, Vol.4, No.2, 1997, pp. 46 − 48.