An Adaptive Channel Parameter Estimation Using QQ-plot

Z. DJUROVIC*, I. KOVACEVIC**, B. KOVACEVIC*
*Faculty of Electrical Engineering, University of Belgrade, SERBIA
**High School of Electrical Engineering, SERBIA
zdjurovic@etf.bg.ac.yu; ivanak@vets.edu.yu; ,kovacevic_b@etf.bg.ac.yu

Abstract. New algorithm for estimation of parameters of communication channel in the circumstances of existence of intensive impulse noise within measurement sequence is proposed in this paper. Proceeding from the theory of robust estimation, a simple, adaptive, practically applicable algorithm is derived that in the circumstances of contaminated normal distribution of measurement noise demonstrates high level of efficiency. QQ-plot technique is used as a framework for estimation of contaminated measurements distribution providing the algorithm adaptation. Application of proposed algorithm is broad, both in the field of wireless communications, equalization of transmitting channels, suppressing of noise and in modeling communication and control systems.

Keywords: Channel identification, robustness, adaptation, QQ-plot.

1. Introduction

In wireless communication systems, the transmitted signals usually experience fading which either attenuates the received power or causes dispersion. Therefore, it is necessary to obtain the fading channel information, i.e. the channel gains of different resolvable paths. However, the channel state estimation is unknown to both the transmitter and the receiver in many practical applications, thus necessitating channel estimation at the front end of the coherent receiver [1]. A variety of channel estimation algorithms have been developed, based primarily on one of two aspects of random variables: the distribution and the moments. When the distribution of the received signal conditioned on the channel state information is known, the maximum likelihood (ML) criterion can be applied to yield asymptotically optimal performance [2]. Such ML channel estimation algorithms are suitable for training-symbol based systems, in which a subset of the transmitted symbols is known to both the transmitter and receiver [3]. The main concern with such a ML channel estimation is the necessary amount of training data [4]. When only the information symbols are available, this is usually called blind channel estimation. In the recent past, a large number of blind channel estimation algorithms have been developed using moment estimation, particularly the second-order statistics or SOS [5]. Usually, based on SOS estimation, the subspace method and moment matching are applied. However, the subspace technique is suitable only for stationary channels, since it requires the signal subspace to be time-invariant. However, in many practical code division multiple-access (CDMA) systems long codes are employed, thus making the overall channel nonstationary. On the other hand, SOS-based estimation techniques are commonly recognized as the natural tools to be used in the presence of Gaussian noise. Research efforts on higher-order statistics (HOS) have led to the development of improved estimation algorithms for non-Gaussian environments. Important non-Gaussian impulsive processes are found in a variety of practical problems that include wireless communications and teletraffic. These processes can be efficiently modeled by heavy-tailed distributions with a huge variance, for which neither the classical SOS theory nor the theory of HOS are well-defined [6]. Additionally, in many applications one expects that channel estimation can be made adaptively so as to accommodate time-varying environments and system parameter variation. The recursive least-squares (RLS) algorithms may be one of the most interesting and powerful techniques that implement adaptive parameter estimation [7]. However, when an impulsive noise in the system output exists, the RLS algorithm usually fails to yield an unbiased estimate of the system parameters, thereby causing the performance of adaptive filtering to be significantly degenerated. The robust estimation theory represents a suitable tool to cope with an impulsive noise environment. In this paper the problem of robustified adaptive parameter channel estimation in the min-max robust estimation context is considered.

2. Formulation of problem

This paper deals with the problem of identification of communication channel by estimation of parameter on the basis of measurable input and output signal. Therefore, let us represent a channel by the abstract, linear, dynamic, stationary, discrete-time system, which can be modeled by difference equation with constant parameters:

$$y(i) = -\sum_{k=1}^{n} a_k y(i-k) + \sum_{k=1}^{m} b_k u(i-k) + \xi(i) \quad (1)$$

whereas $y(i) \in R^1,\ u(i) \in R^1,\ \xi(i) \in R^1$ are the sequences of system output, measurable input signal and stochastic input, respectively, while the constants $a_i, i = 1,...,n$ and $b_j,\ j = 1,...,m$ represent system parameters. If backward operator is introduced $q^{-k} y(i) = y(i-k)$, relation (1) can be written in the following polynomial form

$$A(q^{-1})\, y(i) = B(q^{-1})\, u(i) + \xi(i), \quad (2)$$

where

$$A(q^{-1}) = 1 + \sum_{k=1}^{n} a_k q^{-k}, \quad B(q^{-1}) = \sum_{k=1}^{m} b_k q^{-k}, \quad (3)$$

are so called characteristic and control polynomials. Relation (1) can be also written in linear regression form

$$y(i) = Z^T(i)\,\Theta + \xi(i) \quad (4)$$

where $Z^T(i) = [-y(i-1),...,-y(i-n), u(i-1),...,u(i-m)]$ represents vector of input and output samples, and $\Theta^T = [a_1 \cdots a_n\ b_1 \cdots b_m]$ represents a vector of constant system parameters. The problem of recursive system identification described by relation (4) is actually the problem of estimation of unknown parameters included in vector Θ in real time, and on the basis of measuring signals at input and output of the system. On the basis of incomplete a priori information about the nature of stochastic input (disturbance), a min-max optimal robust identification algorithm can be constructed. It minimizes the adopted performance index $V(T,p)$ which represents the asymptotic estimation error covariance matrix whereas the form of estimator *T* comes from the class of possible estimators τ and for some particular probability density function *p* from a prespecified class *P*. Therefore, we choose the estimator form $T \in \tau$, while a particular probability density function (p.d.f) $p \in P$ is defined by the nature of the system and measuring, so that $V(T,p)$ represents a final result of a such strategy. A such selection has a saddle-point (T_0, p_0) if estimator T_0 and p.d.f p_0 satisfy the following equation

$$\min_{T\in\tau} \max_{p\in P} V(T,p) = V(T_0, p_0) = \max_{p\in P} \min_{T\in\tau} V(T,p) \quad (5)$$

Such T_0 estimator is denoted in the literature as the min-max optimal robust estimator, while p_0 is named the least favorable p.d.f [8,14]. In particular, if the class of τ estimators was chosen in the set of stochastic gradient recursive estimators, then a recursive relation by which parameters of Θ vector are estimated can be written in the form [9,14]

$$\hat{\Theta}(i) = \hat{\Theta}(i-1) + \Gamma(i)\, Z(i)\, \psi\left(\nu\left(i, \hat{\Theta}(i-1)\right)\right) \quad (6)$$

where $\nu\left(i,\hat{\Theta}\right) = y(i) - \hat{\Theta}^T Z(i)$ is the predication error or the residual, while the min-max robust estimator T_0 is defined by the following relation

$$\psi(\cdot) = \psi_0(\cdot) = -[\log p_0(\cdot)]' \quad (7)$$

$$\Gamma(i) = \Gamma_0(i) = \left[iE\left\{\psi_0^{'}(\cdot)\right\} E\left\{Z(i) Z^T(i)\right\}\right]^{-1} \quad (8)$$

The least favorable probability density function $p_0(\cdot)$ is the one which minimizes the Cramer-Rao bound within the considered distribution class *P*. Thus $\psi_0(\cdot)$ in relation (7) represents a maximum likelihood (ML) type function which corresponds to the specific p.d.f p_0, and the estimation procedure (6) is led down to the ML identification method. However, the problem is how to define the least favorable p.d.f $p_0(\cdot)$, since for that purpose a nonclassical variation problem must be solved, mostly by numerical methods. This problem can be solved analytically only in a case of final memory system ($A(\cdot) = 1$ in relation (2)) where the problem is led down to the minimization of the Fisher information $I(p) = E\left\{(p'/p)^2\right\}$ within the adopted distribution class *P*. A lot of different examples for *P* class and corresponding p.d.f $p_0(\cdot)$ which minimizes the Fisher information $I(p)$ within *P* class, can be found in the literature [9]. Moreover, practical selection of the weighted matrix $\Gamma_0(\cdot)$ in relation (8) is not possible without the information about the real distribution of stochastic input. Therefore, the robust estimation practice requires a such estimator that will be efficient in a case when the distribution of disturbance is the normal one, but at the same time it should have good features even in a case when the disturbance distribution significantly deviates from the normal one, possessing the so called „ heavy tails" which result in extremely high values in the output sequence. This is the so-called efficiency robustness. For that reason, aiming to design a practically applicable recursive robust estimator, an additional effort must be made in approximation of the min-max optimal solution. One of possible approaches is based on the method of weighted least squares, which can reduce the impact of extremely bad measurements [10, 15]. An alternative approach is stemmed on the stochastic gradient algorithm, and this method is presented in the paper. Recursive estimator assigned by the relation (6), which for the particular function ψ does not use a ML type function assigned by the relation (7), is named an approximate maximum likelihood recursive estimator or M-estimator in short.

3. Robust stochastic gradient recursive algorithm

In order to realize the min-max optimal robust recursive algorithm (6)-(8), it is necessary to make selection of nonlinear function $\psi(\cdot)$ with the basic aim to recognize somehow irregular measurements, the so called outliers, and to belittle their influence to the identification quality of system parameters. As mentioned before, nonlinear $\psi(\cdot)$ function, which is also named the influence function, must be selected so to provide the efficiency robustness property. In addition, it is suitable this function to be bounded and continuous [11]. Boundedness of function $\psi(\cdot)$ provides that any individual observation or outlier cannot have unlimited influence to the quality of estimation, while the continuity of function enables that effects of rounding and quantization errors or patchy outliers are minor. This is the so-called resistant robustness. Thus, the function $\psi(z)$ must be linear for low values of arguments, and to increases slower than the linear one for large absolute values of arguments, which corresponds, for example, to

$$\psi(z) = \min\left(z/\sigma^2, k/\sigma^2\right)\operatorname{sgn}(z) \tag{9}$$

where σ denotes the standard deviation of measurement noise, and parameter k represents a constant which should be selected with the aim to reach the wanted efficiency robustness. Since the value of variance σ^2 is mostly unknown, it is indispensable to estimate it somehow. A popular and frequently used form of robust variance evaluation σ^2 is the median [11]:

$$d = \frac{median\left\{\left|y(i) - median\left\{y(i)\right\}\right|\right\}}{0.6745} \tag{10}$$

The other, applicable types of nonlinear function $\psi(\cdot)$ can be found in literature [9,11]. The proposed recursive robust estimator has the following characteristics. Namely, if the following ML form is chosen for non-linear influence function:

$$\psi(\cdot) = \psi^*(\cdot) = -\left[\log\left(p^*(\cdot)\right)\right]'; \quad p^* = \arg\min_{p\in P} I(p) \tag{11}$$

such a selection does not satisfy optimum in regard to the min-max condition (5), but it minimizes the conditional estimation error covariance matrix $E\left\{\left[\hat{\Theta}(i) - \hat{\Theta}(i-1)\right]\left[\hat{\Theta}(i) - \hat{\Theta}(i-1)\right]^T \middle| F_{i-1}\right\}$ for the worst case p.d.f p^* in (11). Furthermore, the idea of introducing nonlinear influence function $\psi(\cdot)$ in the form of relation (9) corresponds to the function form $\psi^*(\cdot)$ from relation (11) in the case when *P* class defines the so called $\varepsilon-$contaminated p.d.f family, defined as:

$$P = P_\varepsilon = \left\{p \middle| p = (1-\varepsilon)N\left(0,\sigma^2\right) + \varepsilon h\right\},\ \varepsilon\in[0,1) \tag{12}$$

where $h(\cdot)$ is a symmetrical p.d.f with zero mean, while $N\left(0,\sigma^2\right)$ represents the Gaussian distribution with zero mean and variance σ^2. The least favorable p.d.f p^* from this class is zero mean normal with exponential „heavy tails", resulting in $I^{-1}\left(p^*\right) = 2(1-\varepsilon)erf(k)\sigma^2$, where *erf* denotes the error function.

4. Adaptation of robust recursive algorithm

The $\varepsilon-$contaminated p.d.f family is extremely important for the field of robust estimation, since it models a number of various applications in which there is a sporadic phenomenon of high intensity irregular measurements. Moreover, in a such situation it is appropriate for contaminant $h(\cdot)$ from relation (12) to adopt the Gaussian distribution of zero mean and variance which is considerably higher than the nominal one. In accordance to that, the model of p.d.f of measurement noise in the following form is adopted:

$$p = (1-\varepsilon)N\left(0,\sigma^2\right) + \varepsilon N\left(0,\sigma_o^2\right) \tag{13}$$

which is described with three unknown parameters: contamination intensity or degree $\varepsilon\in[0,1)$, nominal variance of regular measurement noise σ^2

and outlier variance $\sigma_o^2 \gg \sigma^2$. Due to no a priori knowledge on these parameters, a robust estimator defined by (6,8) is usually adopted with influence function (9), where the unknown variance is replaced with median (10) and parameter k is selected so to be $1.42 \times d$ for the sake of providing 95% of efficiency of estimator. Such a selection is reasonable in lack of a priori information on parameters of real measurement noise p.d.f, however it also has considerable shortcomings. The first one is that it is impossible to prove properties of estimator (10) and therefore the expected efficiency is jeopardized. The other important shortcoming is that it is intuitively clear that the influence function must be susceptible to intensity of contamination and variance of outlier changes, what is not provided with the selection of function (9). In order to illustrate the influence of parameters ε, σ^2 and σ_o^2, in the noise model (13), to the selection of influence function, let's look at Figures 1.a,b and c obtained by using the method of ML likelihood ($\psi = \left[-\ln(p)\right]'$, where p is given by (13)).

These ideal ML influence functions are sketched in these figures, where in the Figure 1.a, $\varepsilon = 5\%$, $\sigma_o^2 = 10$ is adopted, while the variance of nominal model has different values $\sigma^2 \in \{1,2,3,4\}$. In the Figure 1.b the variances of nominal model and outliers are constant $\sigma^2 = 1$, $\sigma_o^2 = 10$, while intensity of contamination alters $\varepsilon[\%] \in \left\{2.5, 5, 7.5, 10\right\}$. In the Figure 1.c, variance of nominal model $\sigma^2 = 1$ and intensity of contamination $\varepsilon = 5\%$ are constant, while outlier variance is changed $\sigma_o^2 \in \{5, 10, 15, 20\}$.

The following conclusions may be derived on the basis of presented influence functions. Each of the presented function forms has three distinct regions. The first region is the field for low value of arguments (from both sides of the origin) in which the influence function is almost ideally linear. The variance of nominal model σ^2 is exclusively responsible for the slope of linear segment in this region. Third region is in charge of presence of outliers in the structure of measuring signal (13) and it is also linear, but within the ranges of intensive residuals of big, positive or negative values.

On the basis of presented figures, it is evident that the slope of linear segments in the third region comes as a consequence of outlier variance σ_o^2. Finally, there is also another region, transitional field between the first one and the third one, which mostly depends on the intensity of contamination ε and the variance of nominal model σ^2, while it is quite insensitive to the variance of outliers σ_o^2.

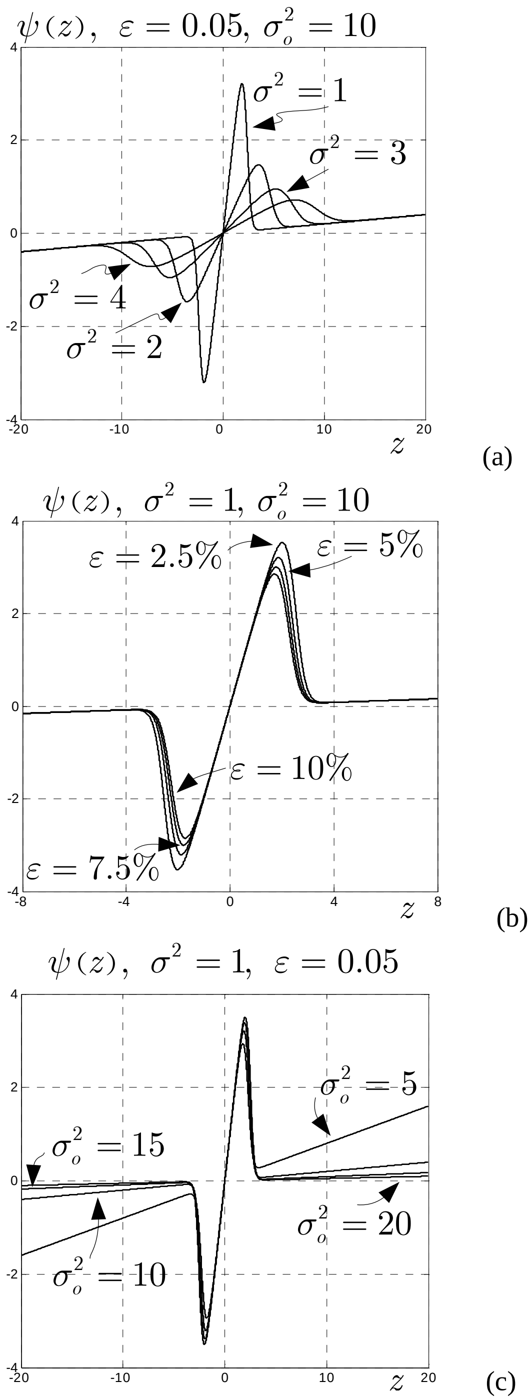

Figure 1: ML influence functions for different parameter values ε, σ^2 and σ_o^2 in noise model (13): a) for different nominal variance σ^2, b) for different contamination intensity ε, c) for different outlier variance σ_o^2

This is a domain in which both regular measurements and bad measurements may occur with almost equal probability; so that the influence function in this domain has a negative gradient, since with the increase of the residual value the probability of regular measurement is constantly decreased.

Having in mind the aforesaid, it becomes clear that estimation of noise parameters σ^2, ε i σ_o^2 help much in the right choice of influence function, what would ultimately result in an estimator of system parameters with efficiency and resistant robustness properties. A suitable theoretical scope for estimation of these noise parameters is QQ-plot, which is usually used as a validity test that p.d.f of random variable corresponds to the adopted one [12,13]. If we assume to have at disposal the sequence of the last *N* residuals:

$$\nu\left(k,\hat{\Theta}\right) = y(k) - \hat{\Theta}^T Z(k), k = i - N + 1, \ldots, i \quad (14)$$

where *i* denotes the current instant, $y(k)$ represents measuring in the *k-th i*nstant, while $\hat{\Theta}$ and $Z(k)$ are defined by (4) and (7).

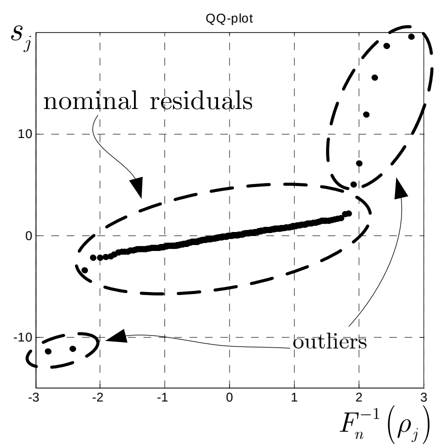

Figure 2: An example of QQ-plot for residual sequence belonging to contaminated normal p.d.f of measurement noise (13)

If for the residual sequence obtained in that way, we form QQ-plot (typical illustration of QQ-plot is given in the Figure 2) where values $F_n^{-1}\left(\rho_j\right)$, $j = 1, \ldots, N$ are put on the horizontal axis with $F_n^{-1}(\cdot)$ being the inversed distribution function for the Gaussian distribution with zero mean and unit variance and

$$\rho_j = (j-1)/(N-1),$$

and on the vertical axis values $s_j,\ j = 1, \ldots, N$ which represent sorted non-decreasing residuals $\nu\left(k,\hat{\Theta}\right),\ k = i - N + 1, \ldots, i$ we will get a graphic design with information about all three required and unknown noise parameters $\left(\varepsilon,\ \sigma^2,\ \sigma_o^2\right)$.

It can be seen in the Figure 2 that the most of residual samples (theoretically $(1-\varepsilon)N$ of them) will be concentrated along the linear segment, while the rare outliers (theoretically εN of them) will considerably deviate from this linear segment.

The upper and lower limit of QQ-plot, within which measurements can be considered regular can be found by minimizing an additional criteria, as is suggested below:

$$\begin{aligned} &\beta_i = \arg\min_{\beta} J_i(\beta); \\ &J_i(\beta) = \left|\beta - F_n^{-1}\left(-p + F_n(\beta / i)\right)\right|; \quad (15) \\ &\alpha_i = F_n^{-1}\left(F_n(\beta_i / i) - p / i\right);\ i = 1, \ldots, N \end{aligned}$$

where parameter *p* is taken in the range $[0.7, 0.8]$. A suitable value of this free parameter in the proposed optimization procedure is determined by simulations. In this procedure α_i and β_i are chosen so that the probability mass in the region $\left(\alpha_i, \beta_i\right)$ is equal to a prespecified limit *p*:

$$p = \int_{\alpha_i}^{\beta_i} f(y/i)dy = F(\beta_i/i) - F(\alpha_i/i)$$

The value of *p* determines the desired efficiency under the nominal Gaussian model. However, this does not define two parameters uniquely, so we need additional requirement. A reasonable choice is to find a such solution for which the distance

$$d_i = \left|\beta_i - \alpha_i\right|,\ i = 1, \ldots, N$$

is minimal. In this way, it is found that the most probable observations, i.e. the observations which are not outliers, propagate unchanged through the parameter estimation algorithm. The posed minimization problem (15) is nonlinear and iterative methods are required to solve it. The minimization of the adopted criterion J_i in (15) can be done by using the Newton-Raphson type method:

$$\begin{aligned} &\hat{\beta}_i(k+1) = \hat{\beta}_i(k) - \Delta\hat{\beta}_i(k) \\ &\Delta\hat{\beta}_i(k) = \left[J_i^{''}\left(\hat{\beta}_i(k)\right)\right]^{-1}\left[J_i^{'}\left(\hat{\beta}_i(k)\right)\right];\ k = 0, 1, 2, \ldots \end{aligned}$$

starting from some initial value $\hat{\beta}_i(0)$.

Counting the number of points N_o on QQ-plot which do not satisfy the condition

$$\alpha_i < s_i < \beta_i,\ i = 1, \ldots, N \quad (16)$$

and dividing by overall number of data *N*, we get the estimate of the contamination intensity ε:

$$\hat{\varepsilon} = N_o / N \quad (17)$$

According to that, the set of all sorted residuals $s_j,\ j = 1,\dots,N$ can be divided into two subsets. The first one is $s_j^r,\ j = 1,\dots,N - N_o$, which includes the regular samples satisfying the condition (16). The second one is $s_j^o,\ j = 1,\dots,N_o$ which covers the samples under the influence of outliers and do not satisfy the condition (16). It can be shown that the slope of linear segment modeling the most of the points of QQ-plot is an estimation of the standard deviation of nominal noise model. This estimate can be reached by using the method of least squares [12]:

$$\begin{bmatrix}\hat{m}\\ \hat{\sigma}\end{bmatrix} = \left(\Sigma^T\Sigma\right)^{-1}\Sigma^T S;$$
$$S^T = \left[s_1^r\ s_2^r\ \cdots s_{N-N_o}^r\right] \quad (18)$$
$$\Sigma^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ F_n^{-1}(\rho_1) & F_n^{-1}(\rho_2) & \cdots & F_n^{-1}\left(\rho_{N-N_o}\right)\end{bmatrix}$$

where $\hat{m}$ denotes the estimation of mean of the nominal noise model, and it should be approximately equal to zero, while $\hat{\sigma}$ denotes the estimation of standard deviation of the nominal noise model.

Moreover, it can be demonstrated that such an estimation of standard deviation of the nominal noise mode is biased, whereas the relative bias depends on the sequence size *N*, and that for $N > 100$ a relative bias is less than few percents. Finally, separating those residuals which are under undoubted influence of outliers, and with estimation of their variance by the arithmetic mean

$$\hat{\sigma}_o^2 = \frac{1}{N_o}\sum_{j=1}^{N_o}\left(s_j^o - \frac{1}{N_o}\sum_{k=1}^{N_o} s_k^o\right)^2 \quad (19)$$

we get the required estimate of parameter σ_o^2.

With periodical estimation of these three noise parameters and their substitution into the ML influence function

$$\psi(\cdot) = -\left[\log\left(p(\cdot)\right)\right]';$$
$$p = (1-\varepsilon)N\left(0,\sigma^2\right) + \varepsilon N\left(0,\sigma_o^2\right) \quad (20)$$

the robust recursive identification algorithm (6-9) becomes adaptive, thus increasing significantly the quality of system parameters identification, what will be illustrated by the following numerical example.

5. Results of simulation

Aiming to analyze practical robustness of the proposed algorithm, the following experiment was carried out. Let us consider the fourth-order system model which is presented in the form of relation (4):

$$Z^T(i) = \left[-y(i-1)\dots - y(i-4)\ \ u(i-1)\right],$$

with

$$\Theta^T = [\text{-}1.48\ 1.28\ \text{-}1.09\ \ 0.37\ 2]$$

vector of parameters. Sequence $\{u(i)\}$ is generated as a white noise sequence with Gaussian p.d.f of zero mean and unit variance, while disturbance $\xi(i)$ is generated as the ε - contaminated p.d.f in (13), with intensity $\varepsilon = 10\%$, nominal variance $\sigma^2 = 1$ and outlier variance $\sigma_o^2 = 10$. The following algorithms are tested: 1) recursive least-squares algorithm denoted as RLS; 2) recursive robust algorithm defined by relations (6)-(9) denoted as RRA. 3) adaptive robust recursive algorithm defined by relations (6)-(8) and adaptation mechanism on the basis of relations (14)-(20), denoted as ARA. A cumulative relative error of parameters estimation is introduced as a measure of algorithms goodness:

$$CPEE(i) = \frac{1}{i}\sum_{k=1}^{i}\sum_{l=1}^{5}\left|\left(\hat{\Theta}_l(k) - \Theta_l\right)/\Theta_l\right| \quad (21)$$

where $\Theta_l,\ l = 1,\dots,5$ denotes the *l*-th element in the vector of unknown parameters Θ, while $\hat{\Theta}_l(k)$ denotes the estimate of this parameter in the *k*-th iteration. Typical value trajectories of these criteria for the proposed algorithms are presented in the Figure 3.

This figure illustrates the efficiency of the suggested ARA algorithm when an impulsive noise in the measurement signal is present. It is noticeable that this algorithm is less susceptible to sporadic phenomena of outliers, what results in minor cumulative error in total. It is also noticeable that RRA algorithm is less susceptible to such kind of measurement noise, in comparison to classical RLS algorithm. However, since RRA does not have the corresponding information about the statistics of contaminated p.d.f of measurement noise, parameter *k* in the relation (9) is the fixed one and ,therefore, RRA is less efficient in comparison to ARA.

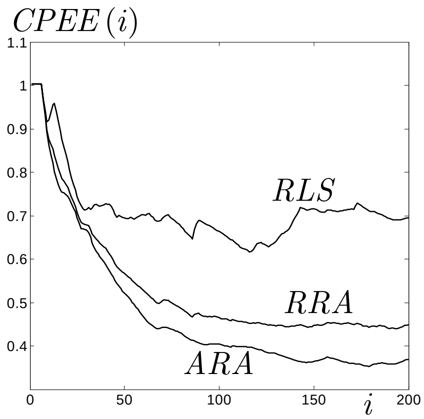

Figure 3: Values of cumulative criterion CPEE for different algorithms

Suggested ARA algorithm demonstrates high efficiency for the values of contamination intensity $\varepsilon < 20\%$, disregarding the variance of outlier σ_o^2. In case when this parameters is higher than 0.2, the assumption on sporadic phenomenon of outliers is not tenable anymore, and for that reason the estimations of parameters are degraded.

6. Conclusion

New algorithm for estimation of communication channel parameters in the presence of intensive impulse noise within measurements is proposed in this paper. Proceeding from the robust estimation theory, an adaptive, practically applicable algorithm is derived, that in the cases of contaminated normal distribution of measurement noise demonstrates high efficiency. Characteristics of adaptation are derived on the basis of estimation of intensity of contamination and variances of both nominal and contaminated noise. Estimations of these parameters are generated using QQ-plot, which when applied to the sequence of residual gives the form that directly depends on the unknown noise parameters. In order to analyze the performances of the algorithm, its simulation is carried out and comparison with the classical method of least squares and Huber robust estimator is performed. Results of simulation have shown that the proposed algorithms is more efficient for typical values of contamination level, regardless the variance of contaminated distribution. Application of proposed algorithm is broad, both in the field of wireless communications, transmitting channels equalization, noise suppressing and in communication and control systems modeling.

References

[1] H. Li, H.V. Poor, Performance Analysis of Semiblind Channel Estimation in Long-Code DS-CDMA Systems, *IEEE Transactions on Signal Processing*, vol. 54, No. 9, 2006, pp. 3383-3399.

[2] H.V. Poor, *An Introduction to Signal Detection and Estimation*, New York: Springer Verlag, 1994.

[3] S. Buzzi and H.V. Poor, Channel estimation and multiuser detection in long-code DS/CDMA systems, *IEEE J. Sel. Areas Commun.* , vol. 19, No. 8, 2001, pp. 1476-1487.

[4] B. Hassibi and B. M. Hochwald, How much training is needed inmultiple-antenna wireless links, *IEEE Trans. Inf. Theory*, vol. 49, No. 4, 2003, pp. 951-963.

[5] H.H. Zeng and L. Tong, Blind channel estimation using the second order statistics algorithms, *IEEE Trans. Signal Proc*., vol. 45, No. 8, 1997, pp. 1919-1930.

[6] J. G. Gonzales, J.L. Paredes and G.R. Arce, Zero-order statistics: A mathematical framework for the processing and characterization of very impulsive signals, *IEEE Trans. Sig. Proc*., vol. 54, No. 10, 2006, pp. 3839-3851.

[7] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, N. J. Prentice Hall, 1996.

[8] Ya. Z. Tsypkin, *Foundations of informational identification theory*, Nauka, Moscow, 1984.

[9] C. H. Lee, On robust linear prediction, *IEEE Transaction on Acoustics, Speech and Signal Processing,* Vol. 36, No. 2, 1988, pp. 642-650.

[10] G. A. Williamson, P.M. Clarkson, W.A. Sethares, Performance characteristics of median LMS adaptive filter, *IEEE Transaction on Signal Processing*, Vol. 41, No.2, 1993, pp. 667-680.

[11] P. J. Huber, *Robust Statistics*, John Wiley, 1981.

[12] Ž. M. Đurović, B. D. Kovačević, QQ-plot approach to robust Kalman fitlering, *Int. J. Control*, vol. 61, No. 4, 1995, pp. 837-857.

[13] Ž. M. Đurović, B. D. Kovačević B., Bandwith selection for kernel density estimation based on QQ-plot, *WSEAS Trans. on Circuits and Systems*, Vol.3, No. 3, 2004, pp. 679-687.

[14] I. Kovačević, B. Kovačević, Ž. Đurović, On Strong Consistency of a Class of Robust Stochastic Gradient Type Systems Identification Algorithms, *WSEAS Transaction on Circuits and Systems*, Issue 8, Vol. 5, 2006, pp.1244-1253.

[15] Z. Banjac, Kovačevi'ć B., Milosavljević M., Veinović M., Robust adaptive filtering with variable forgetting factor, *WSEAS Trans. Circuits and Systems*, Vol. 3, No. 2, 2004, pp. 223-229.

Appendix A: Review of QQ-plot procedure

Let us consider the case of the random samples $\{z_i\}, i=1,...,n$ from a distribution $F(z)$ having the corresponding probability density function (p.d.f.) $f(z)$. By ranking the samples $\{z_i\}$ we obtain the non-decreasing sequence $\{y_i\}$, $i=1,...,n$ such that $y_i < y_j$ for $i<j$. The probability that some observation y will have rank i in the ordered sequence $\{y_i\}$ follows directly from the Bernoulli experiment [7,8]

$$P(i/y)=\binom{n-1}{i-1}F^{i-1}(y)\big(1-F(y)\big)^{n-i} \quad \text{(A1)}$$

Starting from relation (A1), the mathematical expectation of the random variable *i*, assuming the observation *y*, is given by

$$E\{i/y\}=\big(1-F(y)\big)^{n-1}\times \times\sum_{i=1}^{n}\frac{i(n-1)!}{(i-1)!(n-1)!}\left(\frac{F(y)}{1-F(y)}\right)^{i-1} \quad \text{(A2)}$$

By introducing the notation

$$\alpha=\left(\frac{F(y)}{1-F(y)}\right) \quad \text{(A3)}$$

one obtains

$$\begin{aligned} m_{i/y} &=\big(1-F(y)\big)^{n-1}\times \\ &\times\frac{d}{d\alpha}\sum_{i=1}^{n}\frac{(n-1)!}{(i-1)!(n-i)!}\int i\alpha^{i-1}d\alpha \quad \text{(A4)} \\ &=\big(1-F(y)\big)^{n-1}\frac{d}{d\alpha}\Big(\alpha(1+\alpha)^{n-1}\Big) \end{aligned}$$

Starting from (A2) and (A3), one obtains, after differentiation, the following expression for the conditional expectation of the random variable *i* assuming the observation *y:*

$$m_{i/y}=E\{i/y\}=1+(n-1)F(y) \quad \text{(A5)}$$

The *a posteriori* conditional variance of the r.v. *i* assuming *y* is defined by

$$\begin{aligned} \sigma^2_{i/y} &=E\{i^2/y\}-E^2\{i/y\} \\ &=\sum_{i=1}^{n}i^2\binom{n-1}{i-1}F^{i-1}(y)\big(1-F(y)\big)^{2n-1}-m^2_{i/y} \end{aligned} \quad \text{(A6)}$$

Under (A1) and (A3), one further concludes

$$\begin{aligned} \sigma^2_{i/y} &=\big(1-F(y)\big)^{n-1}\frac{d}{d\alpha}\sum_{i=1}^{n}i\binom{n-1}{i-1}\alpha^{i}-m^2_{i/y} \\ &=\big(1-F(y)\big)^{n-1}\times \quad \text{(A7)} \\ &\times\frac{d}{d\alpha}\left[\alpha\frac{d}{d\alpha}\Big(\alpha(1-\alpha)^{n-1}\Big)\right]-m^2_{i/y} \end{aligned}$$

Starting directly from (A5), (A3) and (A7), one obtains the expression for the conditional variance of the random variable *i* assuming the observation *y*

$$\begin{aligned} \sigma^2_{i/y} &=E\{i^2/y\}-E^2\{i/y\} \\ &=(n-1)F(y)\big(1-F(y)\big) \end{aligned} \quad \text{(A8)}$$

It is worthwhile noting that $\sigma^2_{i/y}$ is rather small for such a data point *y* generated from the tails of the distribution $F(y)$, since either $F(y)$ or $\big(1-F(y)\big)$ is small. On the other hand, the maximum of $\sigma^2_{i/y}$ in (A8) corresponds to the value $F(y)=0.5$ and is equal to $(n-1)/4$. Thus, one concludes that $\sigma^2_{i/y}$ can be bounded from the above by a corresponding choice of the sample size, or the so-called data window size, *n*. Moreover, a small value of $\sigma^2_{i/y}$ denotes that the rank *i* of the random variable *y* in the ordered sequence $\{y_i\}$ is in the vicinity of the mean-value $m_{i/y}$, i.e. $i\sim m_{i/y}$, and it follows from (A5) that the relation between *i* and $F(y)$ is approximately linear. Of course, this approximation is better for smaller *n* and for the values of $F(y)$ far away from 0.5. A plot of the ordered data y_i versus the quantity

$$F^{-1}(\rho_i)=F^{-1}\big((i-1)/(n-1)\big)$$

s named the QQ-plot. By taking $m_{i/y}\sim i$ in (A5) one obtains the QQ-plot expressions

$$\begin{aligned} y_i &=F^{-1}(\rho_i)\approx F^{-1}(r_i) \\ \rho_i &=\frac{i-1}{n-1}\approx r_i=\frac{i-0.5}{n};\ i=1,2,...,n \end{aligned} \quad \text{(A9)}$$

where $F^{-1}(\cdot)$ is the inverse of the distribution function $F(\cdot)$. It is more convenient to use r_i than ρ_i since it assigns the finite values $F^{-1}(r_i)$ to the first and the last sample y_1 and y_n, respectively, in the case of commonly used Gaussian, Cauchy or Laplace distribution $F(\cdot)$. Thus, if the QQ-plot in (A9) is fairly linear, then it indicates that the observations have the same distribution function $F(\cdot)$, even in the tails. Moreover, if the observations y_i are in a strict sense white noise, i.e. they are independent and identically distributed (i.i.d.) with the distribution function $F(\cdot)$, the relation (A9) can be expressed in the linear regression form

$$y_i = m + \sigma F_n^{-1}(r_i) = m + \sigma \tilde{r}_i \qquad \text{(A10)}$$

Here $F_n(\cdot)$ is the standard distribution function generating the random variables $\tilde{r}_i = (y_i - m)/\sigma$ with zero-mean and unit variance, while $m = E\{y\}$ and $\sigma^2 = \text{var}\{y\}$. Starting from (A10) one can estimate the unknown parameters m and σ by using the least-squares (LSQ) algorithm

$$\begin{bmatrix} \hat{m} \\ \hat{\sigma} \end{bmatrix} = \left(\Sigma^T \Sigma\right)^{-1} \Sigma^T Y \qquad \text{(A11)}$$

where the $2 \times n$ matrix Σ^T and $n \times 1$ vector Y are defined by

$$\Sigma^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \tilde{r}_1 & \tilde{r}_2 & \dots & \tilde{r}_n \end{bmatrix}; \qquad Y^T = \left[y_1 \; y_2 \; \dots \; y_n\right] \qquad \text{(A12)}$$

Appendix B:

The performance analysis of the mean value and variance least-squares estimates based on QQ-plot data

Starting from (A11) and A(12), one obtains the mean value estimate in the form

$$\hat{m} = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} \tilde{r}_i^2 - \sum_{i=1}^{n} \tilde{r}_i \sum_{i=1}^{n} y_i \tilde{r}_i}{n \sum_{i=1}^{n} \tilde{r}_i^2 - \left[\sum_{i=1}^{n} \tilde{r}_i\right]^2} \qquad \text{(B1)}$$

so that the estimate bias is given by

$$b(\hat{m}) = E\{\hat{m}\} - m = \frac{\sum_{i=1}^{n} \tilde{r}_i \sum_{i=1}^{n} \tilde{r}_i \left(m - m_{y/i}\right)}{n \sum_{i=1}^{n} \tilde{r}_i^2 - \left[\sum_{i=1}^{n} \tilde{r}_i\right]^2} \qquad \text{(B2)}$$

where $m = E\{y\}$, while $m_{y/i}$ denotes the conditional expectation of an observation y assuming its rank i. In addition, by using (A9) and (A10), one obtains

$$\sum_{i=1}^{n} \tilde{r}_i = \frac{1}{\sigma} \sum_{i=1}^{n} \left[F^{-1}\left(\frac{i-1}{n-1}\right) - m\right] = -\frac{nm}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{n} F^{-1}\left(\frac{i-1}{n-1}\right) \qquad \text{(B3)}$$

However, since

$$\begin{aligned}
\sum_{i=1}^{n} F^{-1}\left(\frac{i-1}{n-1}\right) &= \sum_{i=1}^{[n/2]} F^{-1}\left(\frac{i-1}{n-1}\right) + \sum_{i=[n/2]+1}^{[(n+1)/2]} F^{-1}\left(\frac{i-1}{n-1}\right) + \\
&\quad + \sum_{i=[(n+1)/2]+1}^{n} F^{-1}\left(\frac{i-1}{n-1}\right) \\
&= \sum_{i=1}^{[n/2]} F^{-1}\left(\frac{i-1}{n-1}\right) + \sum_{i=[n/2]+1}^{[(n+1)/2]} F^{-1}\left(\frac{i-1}{n-1}\right) \\
&\quad + \sum_{i=1}^{[n/2]} F^{-1}\left(\frac{n-i}{n-1}\right) \\
&= \sum_{i=1}^{[n/2]} \left\{F^{-1}\left(\frac{i-1}{n-1}\right) + F^{-1}\left(1 - \frac{i-1}{n-1}\right)\right\} + \\
&\quad + \sum_{i=[n/2]+1}^{[(n+1)/2]} F^{-1}(0.5)
\end{aligned} \qquad \text{(B4)}$$

where $[\cdot]$ is the integer part, and for a symmetric distribution $F(\cdot)$

$$F^{-1}(x) - m = m - F^{-1}(1-x) \qquad \text{(B5)}$$

one concludes from (B3)-(B5)

$$\sum_{i=1}^{n} \tilde{r}_i = 0 \qquad \text{(B6)}$$

Here, the fact that the last term on the right-hand side in (B4) is equal either to zero (for even *n*) or to *m* (for odd *n*) is used. Taking into account (B2) and (B6), it follows further that $b(\hat{m}) = 0$, i.e. the estimator (B1) is unbiased. In addition, from (B1) and (B6), one obtains

$$\hat{m} - m = \frac{\sum_{i=1}^{n} (y_i - m) \sum_{i=1}^{n} \tilde{r}_i^2}{n \sum_{i=1}^{n} \tilde{r}_i^2} \qquad \text{(B7)}$$

so that the variance of the mean value estimate (B1) is given by

$$\sigma_{\hat{m}}^2 = E\left\{(\hat{m} - m)^2\right\} = \frac{\sigma^2}{n} \qquad \text{(B8)}$$

where $\sigma^2 = E\left\{(y - m)^2\right\}$, from which one concludes that the estimator (B1) is consistent. Furthermore, by using (A11) and (A12), one obtains for the standard deviation estimate

$$\hat{\sigma} = \frac{n \sum_{i=1}^{n} y_i \tilde{r}_i - \sum_{i=1}^{n} \tilde{r}_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} \tilde{r}_i^2 - \left[\sum_{i=1}^{n} \tilde{r}_i\right]^2} \qquad \text{(B9)}$$

By applying the expectation to (B9), it follows further that

$$E\{\hat{\sigma}\} = \frac{n\sum_{i=1}^{n}\tilde{r}_i E\{y_i\} - \sum_{i=1}^{n}\tilde{r}_i \sum_{i=1}^{n} E\{y_i\}}{n\sum_{i=1}^{n}\tilde{r}_i^2 - \left[\sum_{i=1}^{n}\tilde{r}_i\right]^2} = \frac{n\sum_{i=1}^{n}\tilde{r}_i m_{y/i} - \sum_{i=1}^{n}\tilde{r}_i \sum_{i=1}^{n} m_{y/i}}{n\sum_{i=1}^{n}\tilde{r}_i^2 - \left[\sum_{i=1}^{n}\tilde{r}_i\right]^2} = \frac{n\sum_{i=1}^{n}\tilde{r}_i \left(m_{y/i} - m\right)}{n\sum_{i=1}^{n}\tilde{r}_i^2 - \left[\sum_{i=1}^{n}\tilde{r}_i\right]^2} \quad \text{(B10)}$$

Here, the notation $E\{y_i\} = m_{y/i}$, and the fact that $\sum_{i=1}^{n} m_{y/i} = nm$, are used. Taking into account (A9), (A10), (B6) and (B10) one obtains:

$$E\{\hat{\sigma}\} = \sigma \frac{\sum_{i=1}^{n}\left[F^{-1}\left(\frac{i-1}{n-1}\right) - m\right]\left[m_{y/i} - m\right]}{\sum_{i=1}^{n}\left[F^{-1}\left(\frac{i-1}{n-1}\right) - m\right]^2} \quad \text{(B11)}$$

On the other hand, the relation (A5) is valid for each *y*, so that it is satisfied for the particular value $y = m_{y/i}$, yielding

$$m_{i/m_{y/j}} = 1 + (n-1)F\left(m_{y/j}\right) \quad \text{(B12)}$$

Taking into account (B12), one obtains further

$$m_{y/j} = F^{-1}\left(\frac{m_{i/m_{y/j}} - 1}{n-1}\right) \quad \text{(B13)}$$

Finally, if for a large data size *n*

$$m_{i/m_{y/j}} \approx j \quad \text{(B14)}$$

one concludes from (B12)-(B14) that $E\{\hat{\sigma}\} \approx \sigma$, i.e. the estimator bias is negligible.