ANALYSIS OF A CLASS OF ADAPTIVE ROBUSTIFIED PREDICTORS IN THE PRESENCE OF NOISE UNCERTAINTY

Ivana Kostić Kovačević, Jelena Gavrilović, Branko Kovačević

Original scientific paper

A new class of adaptive robust predictors has been considered in the paper. First an optimal predictor is developed, based on the minimization of a generalized mean square prediction error criterion. Starting from the obtained result, an adaptive robust predictor is synthesized through minimization of a modified criterion in which a suitably chosen non-linear function of the prediction error is introduced instead of the quadratic one. Unknown parameters of the predictor are estimated at each step by applying a recursive algorithm of stochastic gradient type. The convergence of the proposed adaptive robustified prediction algorithm is established theoretically using the Martingale theory. It has been shown that the proposed adaptive robust prediction algorithm converges to the optimal systems output prediction. The feasibility of the proposed approach is demonstrated by solving a practical problem of designing a robust version of adaptive minimum variance controller.

Keywords: estimation, non-Gaussian noise, parameter estimation, recursive stochastic algorithms, robust adaptive prediction

Analiza jedne vrste adaptirano robusnog prediktora u prisutnosti nepoznatog šuma

Izvorni znanstveni članak

U radu je razmatrana nova klasa adaptiranih robusnih prediktora. Najprije je koncipiran optimalni prediktor utemeljen na minimizaciji generalizirane srednje kvadratne pogreške predikcije. Time je određena i struktura robusnog adaptiranog prediktora, čija je sinteza izvršena na temelju minimizacije modificiranog kriterija u kome je umjesto kvadratne uvedena proizvoljna nelinearna funkcija pogreške predikcije. Nepoznati parametri prediktora procjenjuju se u svakom koraku primjenom rekurzivnog algoritma tipa stohastičkog gradijenta. Konvergencija predloženog algoritma adaptirane robusne predikcije se teoretski utvrđuje koristeći teoriju Martingala. Time je pokazano da algoritam adaptirane robusne predikcije konvergira optimalnom sustavu izlazne predikcije. Dobiveni teorijski rezultati primijenjeni su na rješavanje praktičnog zadatka sinteze robusnog adaptivnog regulatora minimalne varijance.

Ključne riječi: procjena, parametarska procjena, rekurzivni algoritam, robusni adaptirani prediktor

1 Introduction

Tasks typically related to the modern systems theory are control, signal processing (filter design) and prediction. These tasks have been extensively studied in control theory, communication theory, signal processing and statistics [1 ÷ 6, 14, 22]. A wide variety of techniques have been developed for solving problems involving these tasks. The basic requirement of all such techniques is, on the one hand, maximum use of available a priori information about the properties of the system. This requirement involves the adoption of an appropriate presentation of the system (i.e. its mathematical model). On the other hand, an important practical requirement is robustness of the developed procedures in terms of insensitivity to departures from the assumptions from which they are derived (unmodeled dynamics, absence of full knowledge of noise statistics, nonstationarity, etc.). Moreover, in practice it is necessary to be concerned with outliers arising from many reasons, such as meter and communication errors, sensor failures, incomplete measurements, errors in mathematical models, etc. [7, 8, 9]. These have very detrimental effects on the statistical estimation schemes based on the Gaussian stochastic disturbance models [10, 11]. Therefore, from the practical point of view, it is very important to analyse robustness properties of adaptive estimation schemes in the presence of outliers. Robust alternatives abound in the robustness literature [7, 8, 9]. Although there are many meanings of the word "robust", its purely data – oriented version is the word "resistant" [10]. Namely, an estimate is called resistant if changing a small fraction of the data by large amounts of results in a small change to the estimate. This

requirement is one of insensitivity to outliers. In addition, one may also insist that small changes in most of the data result in only small changes in the estimates. This requirement is one of insensitivity to rounding, grouping and quantization errors, or patchy outliers. Furthermore, the term robustness also has a probabilistic meaning, and at last three distinct probabilistic notations of robustness can be perceived. The oldest and most accessible is that of efficiency robustness. Namely, an estimator is said to be efficiency robust if it has high efficiency, say greater than 90 %, at a nominal Gaussian model, and high efficiency at a variety of strategically chosen non-Gaussian distributions [10]. Another notion is that of min-max robustness over a family of distributions [7, 11]. Typically, the family of distributions is infinite, and asymptotic variances are used as quantifiable performance costs. Finally, the third form of robustness is qualitative robustness [8]. This is, in fact, a continuity requirement which is the probabilistic embodiment of the notion that small changes in the data should produce only small changes in the estimates, where small changes involve both large changes in a small fraction of the data and small changes in all the data. An important concept of this robustness is that of the influence curve, which measures the perturbation of the estimate caused by a single additional observation, the so-called contamination [8]. Unfortunately, the highly technical character of min-max robustness and qualitative robustness makes them relatively inaccessible to applied workers. However, one can make estimation procedures having readily apparent resistance properties, along with desirable efficiency robustness. The robust adaptive prediction algorithm proposed in this paper is based on the last approach,

involving resistant robustness property, along with efficiency robustness.

The paper first develops the optimal structure of the predictor, based on a generalized mean square prediction error [12]. Then, in continuation, a robust adaptive one-step predictor is synthesized based on the minimization of a corresponding non-linear criterion. This criterion is derived from the generalized mean square criterion through substitution of the quadratic function by a suitably chosen non-linear function. Starting from the efficiency robustness property and practical importance of achieving insensitivity to outliers contaminating the Gaussian disturbances, this nonlinearity should look like the quadratic function for small value of the argument, whereas it has to grow more slowly than the quadratic function for large values of the argument. Furthermore, the resistant robustness property requires that the criterion function derivative be bounded and continuous, since boundedness provides for no single observation to have an arbitrarily large influence, while continuity provides for patchy outliers not to have a major effects. The unknown parameters of the proposed adaptive robust predictor are estimated in each step by applying a recursive algorithm of the stochastic gradient type [11, 13]. The convergence of the adaptive robustified predictor algorithm is established theoretically using the Martingale theory [2, 15, 16, 17]. Starting from a linear single-input/single-output ARMAX system representation, it has been shown that the proposed adaptive robust one-step ahead prediction converges, in the Cesaro sense, to the optimal prediction of the system output. The derived theoretical results are used to design a robust adaptive minimum variance controller.

2 Problem statement

The optimized one-step predictor minimizes the criterion based on the generalized mean square prediction error of the form [12]

$$\Im_p = E\left\{\left[\frac{P^o(q^{-1})}{Q^o(q^{-1})}\nu\left(y(i),\hat{y}(i)\right)\right]^2\right\}, \qquad (1)$$

where the prediction error is

$$\nu\left(y(i),\hat{y}(i)\right) = y(i) - \hat{y}(i) = \nu(i|i-1), \qquad (2)$$

and $\hat{y}(i)$ is the one-step prediction of the output from the system $y(i)$. Polynomials $P^o(q^{-1})$ and $Q^o(q^{-1})$ of suitable properties are pre-selected. Specifically, for a dynamic system described by the ARMAX model the relations between the systems output, $y(i)$, systems input, $u(i)$, and noise or disturbance, $e(i)$, are expressed by [1÷4,15, 22].

$$A(q^{-1})y(i) = q^{-k}B(q^{-1})u(i) + C(q^{-1})e(i), \qquad (3)$$

where q^{-1} is the unit delay operator, $q^{-1}y(i) = y(i-1)$, and k is the process delay. Here A, B and C are polynomials defined by expressions

$$A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n}$$
$$B(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_m q^{-m} \qquad (4)$$
$$C(q^{-1}) = 1 + c_1 q^{-1} + \ldots + c_\ell q^{-\ell}$$

Noise $e(i)$ is a white discrete random sequence of zero mean value and variance σ^2.

For the system defined by Eq. (3), the one-step optimized predictor $\hat{y}^*(i|i-1)$, which minimizes the performance index (1), is defined by the equations (more details are in Appendix A)

$$P^o(q^{-1})C(q^{-1})\hat{y}*(i|i-1) = $$
$$= q^{-k}Q^o(q^{-1})B(q^{-1})u(i) + q^{-1}G(q^{-1})y(i) \qquad (5)$$

where the polynomial $G(\bullet)$ satisfies the prespecified equation

$$P^o(q^{-1})C(q^{-1}) = Q^o(q^{-1})A(q^{-1}) + q^{-1}G(q^{-1}) \qquad (6)$$

Here the polynomial $G(\bullet)$ is defined by

$$G(q^{-1}) = g_0 + g_1 q^{-1} + \ldots + g_{\bar{n}-1}q^{-\bar{n}+1}, \bar{n} = \max(n,\ell) \qquad (7)$$

The generalized prediction error is defined by the relation

$$\frac{P^o(q^{-1})}{Q^o(q^{-1})}\nu(i|i-1) = e(i) \qquad (8)$$

and the criterion (1) reaches the minimum

$$\Im_p^* = \min \Im_p = E\left\{e^2(i)\right\} = \sigma^2 \qquad (9)$$

Relation (5), which defines the optimized predictor, may be written in a more compact form as

$$C(q^{-1})\hat{y}*(i|i-1) = $$
$$= q^{-k}\frac{Q^o(q^{-1})}{P^o(q^{-1})}B(q^{-1})u(i) + q^{-1}G(q^{-1})\frac{y(i)}{Q^o(q^{-1})} \qquad (10)$$

If one denotes the filtered quantities of the input, $\tilde{u}(i)$, and output from the system, $\tilde{y}(i)$, as

$$\tilde{u}(i) = \frac{Q^o(q^{-1})}{P^o(q^{-1})}u(i), \quad \tilde{y}(i) = \frac{y(i)}{Q^o(q^{-1})}, \qquad (11)$$

then the relation (10) reduces to

$$C\left(q^{-1}\right)\hat{y}*\left(i|i-1\right) =$$
$$= q^{-k}B\left(q^{-1}\right)\tilde{u}(i) + q^{-1}G\left(q^{-1}\right)\tilde{y}(i) \qquad (12)$$

Using (4) and (7), relation (12) may be written in the linear regression form as

$$\hat{y}*\left(i|i-1\right) = Z^T(i)\theta \qquad (13)$$

where the regression vector is

$$Z^T(i) = \left[\tilde{y}(i-1)...\tilde{y}(i-\overline{n})\,\tilde{u}(i-k)...\right.$$
$$\left....\tilde{u}(i-k-m) - \hat{y}*\left(i-1|i-2\right)...-\hat{y}*\left(i-\ell+1|i-\ell\right)\right] \qquad (14)$$

and θ is the predictor parameter vector, which contains the coefficients of the polynomials $C(q^{-1})$, $B(q^{-1})$, $G(q^{-1})$, respectively, that is

$$\theta^T = \left[q_0\ \ q_1...q_{\overline{n}-1}\ b_0\ b_1...b_m\ c_1\ c_2...c_l\right].$$

In practice, parameters θ of the optimized predictor (13), (14) are generally unknown. The basic requirement is to define a parameter estimation procedure based on available measured data on the input to and output from the system, arriving at the adaptive form of the predictor.

3 A new robust adaptive predictor

Since the criterion (1) weights all prediction errors (2) equally, one can expect that it will be susceptible to outliers and hence be non-robust. One simple way to robustify estimation procedure is to use a non-linear criterion

$$\Im(\theta) = \mathrm{E}\left\{H\left(i, \nu\left(i|i-1\right)\right)\right\} = \mathrm{E}\left\{H\left(i, \nu(i,\theta)\right)\right\}, \qquad (15)$$

instead of the quadratic one in (1), where $H(\bullet)$ is a robust score or loss function that has to suppress the influence of outliers. Let us now show why criterion (1) can be generalized by relation (15). Namely, the expression for the mean square criterion gradient follows from relation (1), and by equating it to zero we derive the condition for the minimum of the criterion

$$\nabla\Im_p = 2\mathrm{E}\left\{\frac{P^o\left(q^{-1}\right)}{Q^o\left(q^{-1}\right)}\nu\left(i|i-1\right)\cdot\frac{\partial\nu\left(i|i-1\right)}{\partial\theta}\right\} = 0, \qquad (16)$$

Given that Eq. (8) applies to the optimized solution in (16) for $\theta = \theta*$, which corresponds to the criterion minimum, and since $e(i)$ and $\dfrac{d\nu\left(i|i-1\right)}{\partial\theta}$ are uncorrelated random quantities, in the general case it is possible to introduce a non-linear function of the generalized

prediction error, so that the optimality condition (16) is still satisfied. In order to achieve this requirement, this function has to be equal to zero for zero-value arguments and features an even property. Moreover, as mentioned before, with regard to the practical importance of achieving insensitivity to outliers contaminating the Gaussian disturbances, an even loss function $H(\bullet)$ in (15) should look like the quadratic one for small values of the argument, whereas it has to grow more slowly than the quadratic function for large values of the argument. In addition, a resistant robustness property requires that the loss function derivative $\psi(\bullet) = H'(\bullet)$ be bonded and continuous. This corresponds, for example, to the choice of the Huber's loss function [7, 9]

$$H(\nu) = \begin{cases} c_1 + \dfrac{\Delta|\nu|}{\sigma_n}, & if\ |\nu| \ge \Delta\sigma_n \\[2mm] \dfrac{\nu^2}{2\sigma_n^2} + c_2, & if\ |\nu| < \Delta\sigma_n \end{cases}, \qquad (17)$$

where $c_{1,2}$ are appropriately defined constants and Δ is chosen to give the desired efficiency at the nominal zero-mean Gaussian noise sequence $\{e(i)\}$ in (3), with variance σ_n^2. Therefore, the tuning parameter Δ in (17) has to be chosen so as to provide the desired efficiency at the nominal Gaussian model. A common choice is $\Delta = 1{,}5$, known as the 1,5-Hiber's robust procedure [7]. This is, in fact, an efficient robustness requirement. On the other hand, the derivative Ψ of the H-function in (17), the so-called influence function in qualitative robustness [13, 14], is given by

$$\Psi(x) = H'(x) = \min\left\{\max\left[\frac{x}{\sigma_n^2} - \frac{\Delta}{\sigma_n^2}\right], \frac{\Delta}{\sigma_n^2}\right\}, \qquad (18)$$

and is bounded and continuous. This is, in turn, a resistant robustness requirement. Thus, the choice of the Huber's loss function in (17) provides for resistant robustness, along with an efficient robustness property.

Application of the Robbins-Monro approach [11, 13] results in a stochastic gradient algorithm for estimating unknown vectors of parameters θ of an adaptive robust predictor, where the gradient of criterion defined by Eq. (15) is given by

$$\frac{\partial\Im(\theta)}{\partial\theta} = -\mathrm{E}\left\{\ell(i,\theta)\,\Psi\left(i, \nu(i,\theta)\right)\right\};\ \ell(i,\theta) = \frac{d\,\hat{y}\left(i|\theta\right)}{\partial\theta} \qquad (19)$$

In relation (19) the non-linear function $\psi(\bullet) = H'(\bullet)$, where $H(\bullet)$ is given by (15), (17).

The form of the stochastic gradient algorithm is [11, 13, 17]

$$\hat{\theta}(i) = \hat{\theta}(i-1) + \gamma(i)\,\ell\left(i, \hat{\theta}(i-1)\right)\Psi\left(i, \nu\left(i, \hat{\theta}(i-1)\right)\right) \qquad (20)$$

Here $\gamma(i)$ is a sequence of nonnegative numbers which control the speed of parameter estimates convergence, and $\psi(\bullet) = H'(\bullet)$ is given by (18), (19).

To accelerate convergence of algorithm (20) in the neighbourhood of the minimum of functional (15), gain $\gamma(i)$ can be multiplied by the positively definitive matrix, resulting in a Newton – Raphson type algorithm [13, 16]

$$\hat{\theta}(i) = \hat{\theta}(i-1) + \gamma(i) R^{-1}(i) \ell(i, \hat{\theta}(i-1)) \cdot$$
$$\cdot \Psi(i, \nu(i, \hat{\theta}(i-1))) \tag{21}$$

Matrix $R(i)$ is Hessian and is determined by recursive relation [13, 16, 17]

$$R(i) = R(i-1) + \gamma(i) \Big[\ell(i, \hat{\theta}(i-1)) \cdot$$
$$\Psi_\nu'(i, \nu(i, \hat{\theta}(i-1))) \ell^T(i, \hat{\theta}(i-1)) - R(i-1) \Big] \tag{22}$$

where $\Psi'(\bullet) = \dfrac{\partial \psi(\bullet)}{\partial \nu}$.

Eq. (22) follows from the fact that the second derivative of criterion (15), defining the Hessian matrix, is given by

$$\frac{\partial^2 \Im(\theta)}{\partial \theta^2} = E \Big\{ -\frac{\partial}{\partial \theta} \ell(i, \theta) \Psi(i, \nu(i, \theta)) +$$
$$+ \ell(i, \hat{\theta}) \Psi_\nu'(i, \nu(i, \theta)) \ell^T(i, \theta) \Big\} \tag{23}$$

Moreover, in the vicinity of optimal solution $\theta \approx \theta^*$, the Eq. (23) can be approximated by the relation

$$\frac{\partial^2 \Im(\theta)}{\partial \theta^2} \cong E \Big\{ \ell(i, \theta) \Psi_\nu'(i, \nu(i, \theta)) \ell^T(i, \hat{\theta}) \Big\} \tag{24}$$

since then $\nu(i, \theta) \approx e(i)$ and $\Psi(e(i)) \approx 0$, due to the fact that mean value of $e(i)$ is zero.

Finally, let us approximate the mathematical expectation (24) by the corresponding arithmetic mean,

$$\frac{\partial^2 \Im(\theta)}{\partial \theta} \approx R(i) = \frac{1}{i} \sum_{k=1}^{i} l(k, \hat{\theta}(k-1))$$
$$\Psi_\nu'(k, \hat{\theta}(k-1)) l^T(k, \hat{\theta}(k-1))$$

from which it follows

$$iR(i) = (i-1)R(i-1) + l(i, \hat{\theta}(i-1))$$
$$\Psi_\nu'(i, \hat{\theta}(i-1)) l^T(i, \hat{\theta}(i-1))$$

The last relation represents the Eq. (22) with $\gamma(i) = \dfrac{1}{i}$.

Let us assume further that in algorithm (21), (22) the scalar factor is $\gamma(i) = \dfrac{1}{i}$ and let us introduce the matrix

$\overline{R}(i) = [iR(i)]$. Let us also note that the algorithm (21), (22) was derived for the general model. Since this research considers an ARMAX model (3), it is possible to determine the prediction derivative $\ell(i, \hat{\theta}(i-1))$ more specifically. Moreover, one often resorts to approximation [11, 13]

$$\frac{\partial}{\partial \theta} \hat{y}(i, \theta) = Z(i) \tag{25}$$

The approximation (25) reflects that the implicit dependency of vector $Z(i)$ on predictor parameter θ is disregarded. If one adopts the matrix trace symbols $r(i) = tr\overline{R}(i)$, and replaces the matrix gain factor $R(i)$ in (21) with the scalar gain $r(i)$, the definitive form of the algorithm of stochastic approximation type for adaptive robust predictor parameter estimation is given by

$$\hat{\theta}(i) = \hat{\theta}(i-1) + \frac{1}{r(i)} Z(i) \Psi(i, \nu(i, \hat{\theta}(i-1))) \tag{26}$$

$$r(i) = r(i-1) + \Psi_\nu'(i, \nu(i, \hat{\theta}(i-1))) Z^T(i) Z(i),$$
$$r(0) = 1 \tag{27}$$

$$Z^T(i) = \big[\tilde{y}(i-1) \ldots \tilde{y}(i-n) \ \tilde{u}(i-k) \ldots$$
$$\tilde{u}(i-k-m) \ -\hat{y}(i-1) \ldots -\hat{y}(i-\ell) \big] \tag{28}$$

$$\hat{y}(i) = Z^T(i) \hat{\theta}(i-1), \ \nu(i, \hat{\theta}(i-1)) = y(i) - \hat{y}(i) \tag{29}$$

The Eq. (27) follows from (22) after introducing $\gamma(i) = \dfrac{1}{i}$ and replacing the matrix $R(i)$ with the previously defined matrix $\overline{R}(i)$, as well as by taking the matrix trace operation on the so obtained relation. The Eq. (28) and (29) are like before derived equation (2), (13) and (14) respectively.

Eqs. (26) ÷ (29) define the adaptive robust predictor. This algorithm represents a compromise between rate of convergence and computational complexity. The next task is to analyse the convergence of algorithm (26) ÷ (29). It is important because of the definition of stringent conditions in which the algorithm is applicable.

4 Convergence analysis

The convergence property of the proposed adaptive robust predictor can be investigated using the Martingale theory [2, 16, 17, 18]. The basic convergence result is the lemma of Neveu [18]. The result is restated in a number of forms that suite better in specific theoretical analysis. A unified treatment of a number of almost sure convergence theorems, based on fact that the processes involved possess a common "almost super-martingale" properties, has been proposed in the literature [19]. To be precise, let $\{\Omega, F, P\}$ be a probability space and $F_1 \subset F_2 \subset \ldots$ a

sequence of increasing sub-sigma-algebras of F. For each n let z_n, β_n, ξ_n and ζ_n be non-negative F_n-measurable random variables such that the conditional mathematical expectation

$$\mathrm{E}\{z_{n+1}|F_n\} \le z_n(1+\beta_n) + \xi_n - \zeta_n \qquad (30)$$

Then, the following theorem can be proven [19].

Theorem 1: $\lim_{n\to\infty} z_n$ exists and is finite, and

$$\sum_{n=1}^{\infty} \zeta_n < \infty \quad \text{with probability one} \quad \text{(w.p.1.)} \quad \text{on}$$

$$\left\{ \sum_{n=1}^{\infty} \beta_n < \infty, \quad \sum_{n=1}^{\infty} \xi_n < \infty \right\}.$$

In addition, the following propositions on sequences are frequently used in establishing convergence results [20].

Lemma 1. Let the sequences $\sum_{n=1}^{\infty} x_n y_n$ and $\sum_{n=1}^{\infty} y_n^{k_1}$ converge. Then the sequence $\sum_{n=1}^{\infty} x_n^{k}$ also converges under the condition $\frac{1}{k} + \frac{1}{k_1} = 1$.

The results of Theorem 1 and Lemma 1 can be used to prove the convergence of the proposed adaptive robust predictor (26) ÷ (29). However, first we need to prove the following auxiliary lemmas.

Lemma 2. Consider the model (3) and the algorithm (26) – (29). Let us assume further that the first derivative $\Psi_v'(i, \bullet)$ of the $\Psi(\bullet)$ function satisfies $\Psi_v'(i, \bullet) \in (0, k_p]$, where $0 < k_p < \infty$. Then the sequences $\sum_{i=1}^{\infty} \frac{Z^T(i) \cdot Z(i)}{r^2(i)}$ converges on $\left\{ \lim_{i\to\infty} r(i) = \infty \right\}$.

The proof is given in Appendix B.

Lemma 3. The sequence $\sum_{i=1}^{\infty} \frac{Z^T(i) \cdot Z(i)}{r(i)}$ diverges on $\left\{ \lim_{i\to\infty} r(i) = \infty \right\}$ under the assumptions of Lemma 1.

The proof is given in Appendix C.

Starting from the results of Theorem 1 and Lemmas 1, 2 and 3 one can prove the following convergence theorem.

Theorem 2: Consider the model (3) and the algorithm (26) ÷ (29) subject to the conditions:

C1: All zeros of the polynomial $C\left(q^{-1}\right)$ are inside the unit circle.

C2: $\{e(i)\}$ is a sequence of bounded, independent and identically distributed (i.i.d) random variables, such that the probability distribution function $P(\bullet)$ is symmetric, and $\sup_N \frac{1}{N} \sum_{i=1}^{N} e^2(i) < \infty$ $w.p.1$, while the conditional

mean and variance

$$\mathrm{E}\{e(i)|F_{i-1}\} = 0, \quad \mathrm{E}\{e^2(i)|F_{i-1}\} = \sigma^2 < \infty$$

C3: The influence function $\Psi(\bullet)$ is odd and continuous almost everywhere.

C4: The influence function $\Psi(\bullet)$ and the probability distribution function $P(\bullet)$ have a common raising point, i.e. $\Psi(z+\varepsilon) > \Psi(z-\varepsilon)$, $P(z+\varepsilon) > P(z-\varepsilon)$ for $\varepsilon > 0$.

C5: The function $\Psi(\bullet)$ is bounded, i.e. $|\Psi(z)| < k, \ k \in (0, \infty)$

C6: The derivative $\Psi_v'(i, \bullet)$ of the $\Psi(\bullet)$ function satisfies $\Psi_v'(i, \bullet) \in (0, k_p]$, where $0 < k_p < \infty$.

C7: The polynomial $C\left(q^{-1}\right)$ in (3) is strictly positive real, i.e. the real part $\mathrm{Re}\left\{ \frac{1}{C\left(q^{-1}\right)} \right\} > 0$.

C8: The system input and output signals satisfy

$$\sup_N \frac{1}{N} \sum_{i=1}^{N} u^2(i) < \infty \quad w.p.1.,$$

$$\sup_N \frac{1}{N} \sum_{i=1}^{N} y^2(i) < \infty \quad w.p.1.$$

Then the adaptive robust prediction $\hat{y}(i)$ in (29) converges, in the Cesaro sense, to the optimal one-step ahead predictor $\hat{y}^*(i|i-1)$ in (13) with probability one (w.p.1), i.e.

$$P\left\{ \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left[\hat{y}^*(i|i-1) - \hat{y}(i) \right]^2 = 0 \right\} = 1 \quad \text{w.p.1.}$$

The proof is given in Appendix D.

The conditions C1 and C7 are commonly used when the standard martingale results are applied for convergence analysis [16, 17]. The assumption C2 represents a standard noise condition in the robust estimation [11]. The conditions C3 ÷ C6 define a class of nonlinearities $\Psi(\bullet)$, or influence functions, that have to cut off the outliers. Many $\Psi(\bullet)$ functions that are commonly used in robust estimation, except the Huber's influence function in (18), such as Hampel's, Tukey's or Andrew's nonlinearity, satisfy the above assumptions [7 ÷ 9]. Furthermore, it is fairly obvious that some condition on the input sequence must be introduced in order to secure a reasonable result. Clearly, an input that is identically zero will not be able to yield full information about the system input-output properties. The condition C8 represents a reasonable practical assumption that input-output sequences are discrete-time signals with finite energy.

Remark: The convergence, in Cesaro sense, allows the number of departures of adaptive robust prediction from the optimal one to be infinite. Therefore, a stronger result is to prove almost surely convergence, or convergence with probability one, for which

$$P\left\{\lim_{i\to\infty}\left[\hat{y}^*\left(i\mid i-1\right)-\hat{y}\left(i\right)\right]=0\right\}=1$$

5 Practical example

To demonstrate the feasibility of derived theoretical results, the proposed approach will be applied to the problem of designing a robust adaptive minimum-variance controller [1, 2, 21]. Let the dynamic plant under consideration be represented by ARMAX model (3) with unit delay for which parameter k is equal to 1. The dynamic plant has to be controlled to make the behaviour of the entire control system with a stationary random set-point approach the desired behaviour of the pre-specified reference system [1, 2, 21]. In other words, the system output, y, should differ as little as possible, in some sense, from the desired output y^*, with the given set point. The measure of this difference can be specified by performance index (1) with $P^o = Q^o = 1$, where the misalignment $v(i)$ is defined as $v(i) = y(i) - y^*(i)$. Here the one-step prediction of the output, $\hat{y}$, is replaced by the desired output, y^*. With the system parameters in (3) known, what we have is the problem of designing an optimal controller, minimizing the adopted criterion (1).

The optimal control is given by (5), with $P^o = Q^o = 1$ and $k=1$, that is

$$B\left(q^{-1}\right)u(i) = C\left(q^{-1}\right)y^*\left(i+1\right) - G\left(q^{-1}\right)y\left(i\right) \qquad (31)$$

where the polynomial $G(\bullet)$ is defined by (7), and satisfies the equation (6), i.e.

$$C\left(q^{-1}\right) = A\left(q^{-1}\right) + q^{-1}G\left(q^{-1}\right) \qquad (32)$$

Taking into account (8) and (9), one concludes that for the optimal control the misalignment, v, is equal to the white noise, e. Moreover, the minimal value of the criterion (1) is equal to the noise variance. The controller equation (31) may be represented in an explicitly recursive form, which is more convenient in adaptive systems. Namely, by introducing the notation of the controller parameters vector

$$\theta_0^T = \left[q_0\ q_1...q_{n_g}\quad b_0\ b_1...b_m\quad c_1\ c_2...c_l\right] \qquad (33)$$

and the observation vector $Z(i)$ in (14), one can rewrite (31) as

$$u(i) = \frac{1}{b_0}\left[y^*\left(i+1\right) - Z^T{}_0\left(i,y^*\right)\right] \qquad (34)$$

where from (14) follows

$$Z_0^T\left(i,s\right) = \left[y(i)...y\left(i-n_g\right)\ ru(i)\ u\left(i-1\right)...\right.$$
$$\left....u\left(i-m\right)\ -s(i)...-s\left(i+1-l\right)\right] \qquad (35)$$

Here the parameter r is either 0 or 1, and $Z(i)$ in (14) corresponds to the value of r equal to one.

However, with an unknown system parameters in (3), the need for an adaptive system arises. This system employs an adaption algorithm which changes controller parameters (33) to make the entire control system meet the requirements. Adaptive control system can be obtained in several ways [1,2]. One possibility is to determine directly the unknown controller parameters (33) [21]. This can be done by robust algorithms predicting the desired reference value y^*, and minimizing the functional of the prediction error, v, in (15). The solution of the prediction problem relies on one –step ahead prediction, $\hat{y}$, of y^*, and the controller equation (34). This leads to the recursive algorithm (26) – (29) with $Z(i)$ being equal to $Z_i(i,\hat{y})$ in (35), and $v(i) = y(i) - \hat{y}(i)$. It should be noted that the relation (34) can be rewritten in the form of equation (13), that is

$$Z^T\left(i\right)\theta\left(i-1\right) = y^*\left(i\right);\quad Z\left(i\right) = Z_1\left(i;\hat{y}\right)$$
(36)

Thus, the minimum variance strategy is obtained by predicting robustly one-step ahead the output, y, with (29) and then choosing a control, u, that makes the prediction, $\hat{y}$, equal to the desired output, y^*, as is show in (36). The performances of this algorithm, compared with the convenient non-robust minimum variance type adaptive controller, are analysed by simulations in the literature [21].

6 Conclusion

A new adaptive robust one-step ahead predictor was synthesized by minimizing a suitable chosen non-linear prediction error criterion. Given the importance of occurrence of pulse noise, or outliers, within the Gaussian samples of the measurement noise population, this nonlinearity should look like a quadratic function for small values of the argument, whereas it has to grow more slowly than the quadratic one for large value of the argument. In addition, the non-linear loss function derivative, named the influence function, has resistant robustness property, along with the efficiency robustness. The unknown parameters of the proposed adaptive robust predictor are estimated in each step by applying a recursive algorithm of the stochastic gradient type. The convergence of the adaptive robust prediction algorithm, in the Cesaro sense, is established theoretically using standard Martingale theory. It has been shown that the proposed adaptive robust prediction converges to the optimal systems output predictions. The obtained theoretical results are used to solve the problem of designing a robust version of an adaptive minimum variance type controller.

Further problems in the robust prediction context, that are of practical interest, include a multi-step predictor that plays a significant role in processes involving delay. It is well-known from engineering practice that the delay phenomenon renders the generation of adequate control action rather difficult. It would also be of interest, because

of convergence speed, to consider a robust version of the parameter estimation algorithm where the scalar gain factor is replaced with a suitable matrix. This, in turn, increases the computing complexity of the parameter estimation algorithm.

Appendix A: Derivation of Eqs. (5) ÷ (9)

Algorithm (5) ÷ (9) is a modification of the results presented in [12]. Namely, it follows from Eq. (6) that

$$
\begin{aligned}
&P^o\left(q^{-1}\right)C\left(q^{-1}\right)v\left(i|i-1\right)= \\
&Q^o\left(q^{-1}\right)A\left(q^{-1}\right)y(i)+q^{-1}G\left(q^{-1}\right)y(i)- \\
&-Q^o\left(q^{-1}\right)A\left(q^{-1}\right)\hat{y}\left(i|i-1\right)-q^{-1}G\left(q^{-1}\right)\hat{y}\left(i|i-1\right)
\end{aligned}
\tag{A1}
$$

Considering Eq. (3), and replacing the term Ay from (3) into (A1), equation (A1) acquires the form

$$
\begin{aligned}
&P^o\left(q^{-1}\right)C\left(q^{-1}\right)v\left(i|i-1\right)=Q^o\left(q^{-1}\right)C\left(q^{-1}\right)e(i)+ \\
&+\left[q^{-k}Q^o\left(q^{-1}\right)B\left(q^{-1}\right)u(i)+q^{-1}G\left(q^{-1}\right)y(i)-\right. \\
&\left.-Q^o\left(q^{-1}\right)A\left(q^{-1}\right)\hat{y}^*\left(i|i-1\right)-q^{-1}G\left(q^{-1}\right)\hat{y}^*\left(i|i-1\right)\right]
\end{aligned}
\tag{A2}
$$

If the expression in square brackets in (A2) is equated to zero, the predictor optimality equation becomes

$$
\begin{aligned}
&Q^o\left(q^{-1}\right)A\left(q^{-1}\right)\hat{y}^*\left(i|i-1\right)+ \\
&+q^{-1}G\left(q^{-1}\right)\hat{y}^*\left(i|i-1\right)= \\
&=q^{-k}Q^o\left(q^{-1}\right)B\left(q^{-1}\right)u(i)+ \\
&+q^{-1}G\left(q^{-1}\right)y(i)
\end{aligned}
\tag{A3}
$$

Finally, by expressing $Q_0 A$ from the adopted equation (6) and substituting this result into (A3), one obtains

$$
\begin{aligned}
&P^o\left(q^{-1}\right)C\left(q^{-1}\right)\hat{y}*\left(i|i-1\right)= \\
&=q^{-k}Q^o\left(q^{-1}\right)B\left(q^{-1}\right)u(i)+q^{-1}G\left(q^{-1}\right)y(i)
\end{aligned}
\tag{A4}
$$

representing the Eq. (5).

Eq. (8) follows from (A2) and (A3), that is

$$
\frac{P^o\left(q^{-1}\right)}{Q^o\left(q^{-1}\right)} \; v\left(i|i-1\right)=e(i)
\tag{A5}
$$

while the Eq. (9) follows directly from (1) and (8), thus completing the proof.

Appendix B: Prof of Lemma 2

Starting from the assumption of Lemma 2, one concludes in $f\ \Psi_v{}'\left(i,\bullet\right)\geq\delta>0$. Therefore, there exists a

finite positive constant k_1 such that $k_1\ \Psi_v{}'\left(i,v\left(i,\hat{\theta}(i-1)\right)\right)\geq1$. Taking into account (27), one can write

$$
\begin{aligned}
&r(i)=r(0)+\sum_{k=1}^{i}\Psi_v{}'\left(k,\hat{\theta}(k-1)\right)Z^T(k)\cdot Z(k), \\
&r(0)=1
\end{aligned}
\tag{B1}
$$

and

$$
\begin{aligned}
&\sum_{i=1}^{\infty}\frac{Z^T(i)\cdot Z(i)}{r^2(i)}\leq \\
&\sum_{i=1}^{\infty}\frac{k_1\Psi_v{}'\left(i,\hat{\theta}(i-1)\right)Z^T(i)\cdot Z(i)}{\left[1+\sum_{k=1}^{i}\Psi_v{}'\left(k,\hat{\theta}(k-1)\right)Z^T(k)\cdot Z(k)\right]^2}\leq \\
&k_1\sum_{i=1}^{\infty}\frac{\Psi_v{}'\left(i,\hat{\theta}(i-1)\right)Z^T(i)\cdot Z(i)}{\left[\sum_{k=1}^{i}\Psi_v{}'\left(k,\hat{\theta}(k-1)\right)Z^T(k)\cdot Z(k)\right]^2}<\infty
\end{aligned}
\tag{B2}
$$

The last inequality in (B2) results from Abel-Dini's theorem [19], which can be applied since the sequence $\sum_{i=1}^{\infty}\Psi_v{}'\left(i,v\left(i,\hat{\theta}(i-1)\right)\right)Z^T(i)Z(i)$ diverges, which completes the proof.

Appendix C: Proof of Lemma 3

Similarly as in the proof of Lemma 2, one concludes that there exists a finite positive constant k_2, such that $k_2\Psi_v{}'\left(i,v\left(i,\hat{\theta}(i-1)\right)\right)\leq1$. Thus, one can write further using (B1)

$$
\begin{aligned}
&\sum_{i=1}^{\infty}\frac{Z^T(i)\cdot Z(i)}{r(i)}\geq k_2\cdot \\
&\sum_{i=1}^{\infty}\frac{\Psi_v{}'\left(i,\hat{\theta}(i-1)\right)Z^T(i)\cdot Z(i)}{\left[1+\sum_{k=1}^{i}\Psi_v{}'\left(k,\hat{\theta}(k-1)\right)Z^T(k)\cdot Z(k)\right]}=\infty
\end{aligned}
\tag{C1}
$$

The right hand side of relation (C1) is a consequence of the Abel-Dini's theorem [19], which completes the proof.

Appendix D: Prof of Theorem 2

Let us introduce Lyapunov's stochastic function

$$
V(i)=\|\tilde{\theta}(i)\|^2,\quad\tilde{\theta}(i)=\hat{\theta}(i)-\theta
\tag{D1}
$$

where $\hat{\theta}(i)$ is the predictor parameter vector estimate, generated by (26), while θ is the true unknown predictor

parameter vector to be estimated, and $\hat{\theta}(i)$ is the estimated error in the i-th step. Symbol $\| \cdot \|$ denotes the Euclidean norm. Following the methodology presented in [12], one obtains for the prediction error the relation

$$v(i|i-1) = -\tilde{\theta}^T(i-1)\frac{Z(i)}{C(q^{-1})} + e(i) \tag{D2}$$

Taking into account (D2), the relation (26) can be rewritten

$$\tilde{\theta}(i) = \tilde{\theta}(i-1) - \frac{1}{r(i)}Z(i)\cdot$$
$$\Psi\left(\tilde{\theta}^T(i-1)\frac{Z(i)}{C(q^{-1})} - e(i)\right) \tag{D3}$$

from which it follows

$$\left\|\tilde{\theta}(i)\right\|^2 = \tilde{\theta}^T(i)\tilde{\theta}(i) = \left\|\tilde{\theta}(i-1)\right\|^2 - \frac{2}{r(i)}\tilde{\theta}^T(i-1)\cdot$$
$$Z(i)\Psi\left(\tilde{\theta}^T(i-1)\frac{Z(i)}{C(q^{-1})} - e(i)\right) + \frac{1}{r^2(i)}\cdot \tag{D4}$$
$$Z^T(i)Z(i)\Psi^2\left(\tilde{\theta}^T(i-1)\frac{Z(i)}{C(q^{-1})} - e(i)\right)$$

Additionally, let us define the functions

$$\Phi\left(\tilde{\theta}^T(i-1)\frac{Z(i)}{C(q^{-1})}\right) =$$
$$= E\left\{\Psi\left(\tilde{\theta}^T(i)\frac{Z(i)}{C(q^{-1})} - e(i)\right)\middle|F_{i-1}\right\} \tag{D5}$$

$$\Phi_1\left(\tilde{\theta}^T(i-1)\frac{Z(i)}{C(q^{-1})}\right) =$$
$$= E\left\{\Psi^2\left(\tilde{\theta}^T(i)\frac{Z(i)}{C(q^{-1})} - e(i)\right)\middle|F_{i-1}\right\} \tag{D6}$$

Under the hypothesis C2 and C5 of Theorem 2, one concludes

$$\Phi_1\left(\tilde{\theta}^T(i-1)\frac{Z(i)}{C(q^{-1})}\right) \leq k, \quad k \in (0,\infty) \tag{D7}$$

In addition, the hypothesis C4 of Theorem 2 assumes the function $\Psi(\bullet)$ in (D5) to be monotone increasing and positive (negative) for positive (negative) arguments.

Moreover, under the hypothesis C3, C4 and C7 of the Theorem 2, one obtains

$$\tilde{\theta}^T(i-1)Z(i)\Phi\left(\tilde{\theta}^T(i-1)\frac{Z(i)}{C(q^{-1})}\right) > 0 \text{ for}$$
$$\tilde{\theta}^T(i-1)Z(i) \neq 0 \tag{D8}$$

Using the relations (D4), (D7) and (D8), one can write

$$E\left\{\left\|\tilde{\theta}(i)\right\|^2\middle|F_{i-1}\right\} \leq \left\|\tilde{\theta}(i-1)\right\|^2 -$$
$$\frac{2}{r(i)}\left|\tilde{\theta}^T(i-1)Z(i)\right|\left|\Phi\left(\frac{\tilde{\theta}^T(i-1)Z(i)}{C(q^{-1})}\right)\right| + \tag{D9}$$
$$\frac{kZ^T(i)Z(i)}{r^2(i)}$$

In order to apply the results of Theorem 1 on the relation (D9), with $z_i' = \left\|\tilde{\theta}(i)\right\|^2$ and $\beta_i = 0$, the following conditions have to be fulfilled:
i) The random quantities

$$\xi_i = k\frac{Z^T(i)Z(i)}{r^2(i)},$$

$$\zeta_i = \frac{2}{r(i)}\left|\tilde{\theta}^T(i-1)Z(i)\right|\left|\Phi\left(\frac{\tilde{\theta}^T(i-1)Z(i)}{C(q^{-1})}\right)\right|$$

have to be non-negative $F_{i-1}-$ measurable random variables.

ii) $\displaystyle\sum_{i=1}^{\infty}\frac{kZ^T(i)Z(i)}{r^2(i)} = k\sum_{i=1}^{\infty}\frac{Z^T(i)Z(i)}{r^2(i)} < \infty \quad$ w.p.1.

The condition i) is fulfilled obviously, while the condition ii) is satisfied due to the Lemma2. Thus, by applying the Theorem 1 on the relation (D9), one concludes

$$\lim_{i\to\infty}\left\|\tilde{\theta}(i)\right\|^2 = \theta^* \qquad \text{w.p.1.} \tag{D10}$$

$$\sum_{i=1}^{\infty}\frac{\left|Z^T(i)\tilde{\theta}(i-1)\right|}{r(i)}\cdot\left|\Phi\left(\frac{\tilde{\theta}^T(i-1)Z(i)}{C(q^{-1})}\right)\right| < \infty \text{ w.p.1.} \tag{D11}$$

Furthermore, let us analyse the relation (D11) in more details. Taking into account (B1) and (27), one concludes

$$\frac{\psi_v'\left(i,v\left(i,\hat{\theta}(i-1)\right)\right)\|Z(i)\|^2}{r(i)} \le 1 \tag{D12}$$

Moreover, using (D11) and (D12), one can write

$$\sum_{i=1}^{\infty} \frac{\left|Z^T(i)\tilde{\theta}(i-1)\right|}{\left|r^{\frac{1}{2}}(i)\right|} \cdot \frac{\left|\left[\Psi_v'\left(i,v\left(i,\hat{\theta}(i-1)\right)\right)\|Z(i)\|^2\right]^{\frac{1}{2}}\right|}{\left|r^{\frac{1}{2}}(i)\right|} \cdot \tag{D13}$$

$$\frac{1}{\left|r^{\frac{1}{2}}(i)\right|}\left|\Phi\left(\frac{\tilde{\theta}^T(i-1)Z(i)}{C(q^{-1})}\right)\right| < \infty$$

Bearing in mind hypotheses C4 and C5 of Theorem 2, and by applying Lemma 2, one obtains

$$\sum_{i=1}^{\infty} \frac{\Psi_v'\left(i,v\left(i,\hat{\theta}(i-1)\right)\right)\|Z(i)\|^2}{r^2(i)}\Phi^2\left(\frac{\tilde{\theta}^T(i-1)Z(i)}{C(q^{-1})}\right)$$

$$\le k\sum_{i=1}^{\infty} \frac{\Psi_v'\left(i,v\left(i,\hat{\theta}(i-1)\right)\right)\|Z(i)\|^2}{r^2(i)} < \infty \tag{D14}$$

Furthermore from D1 and Lemma 1, in which $k_1 = k = 2$, and

$$x_i = \frac{\left[Z^T(i)\tilde{\theta}(i-1)\right]^2}{\sqrt{r(i)}},$$

$$y_i = \frac{\left|\left[\Psi_v'\left(i,\hat{\theta}(i-1)\right)\|Z(i)\|^2\right]^{\frac{1}{2}}\right|}{r(i)}\cdot\left|\Phi\left(\frac{\tilde{\theta}^T(i-1)Z(i)}{C(q^{-1})}\right)\right|$$

one concludes

$$\sum_{i=1}^{\infty} \frac{\left[Z^T(i)\tilde{\theta}(i-1)\right]^2}{r(i)} < \infty \tag{D15}$$

In addition, let us define the quantity

$$p(i) = v(i|i-1) - e(i) \tag{D16}$$

Similarly as in [12], one can show

$$C(q^{-1})p(i) = -Z^T(i)\tilde{\theta}(i-1) \tag{D17}$$

Bearing in mind the hypothesis C1 of the Theorem 2 and the relations (D15) and (D17), one further concludes

$$\sum_{i=1}^{\infty} \frac{p^2(i)}{r(i)} < \infty \tag{D18}$$

Now, next step is to show that

$$\lim_{i\to\infty} \frac{1}{r(i)}\sum_{k=1}^{i} p^2(k) = 0 \tag{D19}$$

Therefore, let us define a quantity

$$T(i) = \frac{1}{r(i)}\sum_{k=1}^{i} p^2(k) \tag{D20}$$

from which it follows

$$T(i) = \frac{1}{r(i)}\sum_{k=1}^{i-1} p^2(k) + \frac{1}{r(i)}p^2(i) + \frac{1}{r(i-1)}\cdot$$

$$\sum_{k=1}^{i-1} p^2(k) - \frac{1}{r(i-1)}\sum_{k=1}^{i-1} p^2(k)$$

$$= T(i-1) - \frac{r(i)-r(i-1)}{r(i)}T(i-1) + \frac{p^2(i)}{r(i)} \tag{D21}$$

$$\le T(i-1) - \frac{k_m\Psi_v'\left(i,\hat{\theta}(i-1)\right)Z^T(i)Z(i)}{r(i)}\cdot$$

$$T(i-1) + \frac{p^2(i)}{r(i)}$$

where $k_m \in (1-\varepsilon,1]$, $\varepsilon > 0$. By applying the conditional expectation $E\{\bullet|F_{i-1}\}$ on (D21), one concludes, under the relation (D18), that $T(i)$ is a discrete super-martingale, satisfying Theorem 2 with

$$z_i = T(i),\ \beta_i = 0,\ \xi_i = \frac{p^2(i)}{r(i)},$$

$$\zeta_i = \frac{k_m\Psi_v'\left(i,\hat{\theta}(i-1)\right)Z^T(i)Z(i)}{r(i)}\cdot T(i-1)$$

This, in turn, results in

$$\lim_{i\to\infty} T(i) = T^* \tag{D22}$$

$$\sum_{i=1}^{\infty} \frac{k_m\Psi_v'\left(i,\hat{\theta}(i-1)\right)Z^T(i)Z(i)}{r(i)}\cdot T(i-1) < \infty \tag{D23}$$

However, using C6 of Theorem 2 and Lemma 3, it follows from (D22) and (D23) that $T^*=0$, so that the relation (D19) is proven. Following the methodology exposed in [12], one can also show from (D18) that

$$\lim_{i\to\infty} \frac{1}{i}\sum_{k=1}^{i} p^2(k) = 0 \tag{D24}$$

and since

$$p(i) = y(i) - \hat{y}(i|i-1) - e(i) =$$
$$\hat{y}^*(i|i-1) - \hat{y}(i)$$

(D25)

one completes the proof.

7 References

[1] Astrom, K. J.; Wittenmark, B. Adaptive Control. // Addison Wesley, New York, 1989.

[2] Goodwin, G. C.; Sin, K. C. Adaptive Filtering, Prediction and Control. // Prentice Hall, New Jersey, 1984.

[3] Bosch van der, P. P. J.; Klauw van der, A. C. Modelling, Identification and Simulation of Dynamical Systems. // CRC Press, Boca Raton, Florida, 1994.

[4] Verhacger, M.; Verdult, V. Filtering and System Identification. // Cambridge University Press, Cambridge, 2009.

[5] Mitra, S. W. Digital Signal Processing: A Computer based Approach. // McGraw-Hill, New York, 2002.

[6] Smith, S. W. Digital Signal Processing: A Practical Guide for Engineers and Scientists. // Newness: an imprint of Elsevier, New York, 2006.

[7] Huber, P. J.; Ronchetti, E. M. Robust Statistics. // John Wiley, New York, 2009. DOI: 10.1002/9780470434697

[8] Hampel, F. R.; Ronchetti, E. M.; Rousseew, P. J.; Stahel, W. A. Robust Statistics: The Approach based on Influence Function. // John Wiley, New York, 1986.

[9] Venables, W. N.; Riply, B. D. Modern Applied Statistics with S. // Springer, Berlin, 2002. DOI: 10.1007/978-0-387-21706-2

[10] Barnet, V.; Lewis, T. Outliers in Statistical Data. // John Wiley, New York 1978.

[11] Tsypkin, Ya. Z. Foundations of Informational Theory of Identification. // Nauka, Moscow, 1984.

[12] Goodwin, C. G.; Ramadge, P. J.; Caines, P. E. A globally convergent adaptive predictor. // Automatica. 17, 1(1981), pp. 135-140. DOI: 10.1016/0005-1098(81)90089-3

[13] Ljung, L.; Soderstrom, T. Theory and Practice of Recursive Identification. // MIT Press, Cambridge, 1983.

[14] Candy, J. V. Model-based Signal Processing. // John Wiley, New York, 2006.

[15] Kovačević, B.; Đurović, Ž. Fundamentals of Stochastic Signals, Systems and Estimation Theory. // Springer-Verlag, Berlin, 2008. DOI: 10.1007/978-3-540-70991-6

[16] Kovačević, I.; Kovačević, B.; Đurović, Ž. On strong consistency of a class of robust stochastic gradient type systems identification algorithms. // WSEAS Transaction on Circuits and Systems, WSEAS Press, Athens. 5, 8(2006), pp. 1244-1253.

[17] Kovačević, I.; Kovačević, B.; Đurović, Ž. On strong consistency of a class of recursive stochastic Newton-Raphson type algorithms with applications to robust linear dynamic system Identification. // Facta Universities Series: Electronics and Energetics, University Press, Niš. 21, 1(2008), pp. 1-22.

[18] Neveu, J. Discrete Parameter Martingals. // North-Holland, Amsterdam, 1975.

[19] Rustagi, J. S. Optimizing Methods in Statistic. // Academic Press, New York, 1971.

[20] Hardy, G. H.; Polya, G.; Littlewood, J. E. Inequalities.// Cambridge University Press, Cambridge, 1934.

[21] Filipović, V., Kovačević, B., On robustified adaptive minimum-variance controller. // Int. J. Control. 65, 1(1996), pp. 117-129. DOI: 10.1080/00207179608921689

[22] Kovačević, B.; Banjac, Z.; Milosavljević, M. Adaptive Digital Filters. // Springer-Verlag, Berlin, 2013. DOI: 10.1007/978-3-642-33561-7

Authors' addresses

Ivana Kostić Kovačević, Ph. D., Associate professor
Faculty of Informatics and Computing, Singidunum University
Danijelova 32, 11000 Belgrade, Serbia
E-mail: ikkovacevic@singidunum.ac.rs

Jelena Gavrilović, M.Sc., Teaching Assistant
Faculty of Informatics and Computing, Singidunum University
Danijelova 32, 11000 Belgrade, Serbia
E-mail: jgavrilovic@singidunum.ac.rs

Branko Kovačević, Ph. D., Full professor
School of Electrical Engineering, University Belgrade
Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia
E-mail: dekanat@etf.bg.ac.rs