UDC 681.513

Azarskov V. M. Doct. of Sci.
Zhiteckij L. S. Cand. of Sci.
Sushchenko O. A. Cand. of Sci.

ADAPTIVE PREDICTION OF A CLASS OF STOCHASTIC PROCESSES

National Aviation University, fsu@nau.edu.ua

The problem of the asymptotically optimal prediction of some stochastic processes described by difference equations whose parameters are known is considered. To solve this problem, a simple adaptive estimation algorithm is proposed and analyzed. The asymptotical properties of this algorithm are established. The simulation results are given.

Introduction

The prediction of stochastic processes is an important problem in the area of modern control and signal processing due to its practical applicability. For example, such a problem arises in process control if the plant to be controlled comprises an unit with pure time delay and its external signals are random.

During the last three decades, significant progress has been achieved in the direction of designing predictors with acceptable performance. In particular, substantial breakthroughs in the solution to the above problem have been made by Box and Jenkins who have advanced various approaches to predict the future values of some discrete-time stochastic processes [1]. To implement the predictors suggested in their book [1], the coefficients of difference equations describing these processes must be known a priori. However, such knowledge can hardly be obtained in practice.

To overcome difficulties associated with initial uncertainty about the parameters of process to be predicted, adaptive approaches may be utilized (see the books [2 – 5]). Namely, the adaptive prediction algorithm based on the stochastic approximation method is proposed in [2, subsection 2.2.4°], where its ultimate properties are strictly proved. In [3, subsection 4.2.3°], another adaptive prediction algorithm exploiting the well-known recursive least squares method is derived to update the estimates of unknown parameters. A common disadvantage of these algorithms is that they are complex enough.

This paper deals with the adaptive prediction of some stochastic processes via employing a simple recursive estimation of their parameters. To this end, basic ideas of previous works [6, 7] are extended to the stochastic case. The main effort is focused on establishing the asymptotical properties of the adaptive prediction procedure.

Problem statement

We consider the class of so-called autoregressive stochastic processes (AR processes) [1, p. 24] caused by the discrete-time model whose output sequence $\{y_t\}$ takes values in $\mathbf{R}$ and satisfies the difference equation

$$A(z^{-1})y_t = \zeta_t, \quad t = 1,2,..., \tag{1}$$

where $\{\zeta_t\}$ represents the sequence of independent random variables with zero mean and variance $\sigma_\zeta^2 < \infty$ (white noise) and

$$A(z^{-1}) = 1 + a_1 z^{-1} + ... + a_N z^{-N} \tag{2}$$

is the polynomial of degree N in the inverse shift operator z^{-1}.

Define the N-dimensional parameter vector of this model as

$$\theta^{\mathrm{T}} = [a_1, ..., a_N]$$

and introduce the vector

$$\varphi_{t-1}^{\mathrm{T}} = [-y_{t-1}, ..., -y_{t-N}]$$

containing the N past outputs $y_{t-1}, ..., y_{t-N}$ taken with opposite sign. Equation (1) can then be rewritten in the regression form as follows:

$$y_t = \theta^{\mathrm{T}}\varphi_{t-1} + \zeta_t. \qquad (3)$$

Let $p_t(\zeta)$ denote the probability distribution density of the random variable ζ_t. From the definition of $\{\zeta_t\}$ it can be concluded that

$$p_t\{\zeta \mid \zeta_{t-1},\ldots,\zeta_0\} \equiv p_t\{\zeta\} \overset{\text{def}}{=} p\{\zeta\},$$

where $p_t\{\zeta \mid \bullet\}$ is the conditional distribution density of ζ_t.

The following assumptions about the AR process generated by (1) are made.

1) The polynomial $A(z^{-1})$ described in (2) is asymptotically stable, i.e., it has all zeros inside the closed unit disk: $A(z^{-1}) \neq 0$ for all z: $1 \leq |z| < \infty$.

2) The coefficients of $A(z^{-1})$ are unknown.

3) The integer N defining the dimension of θ is assumed to be known.

4) The random sequence $\{\zeta_t\}$ is upper bounded, i.e., there exists a finite C_ζ such that

$$\sup_{0\leq t<\infty} |\zeta_t| \leq C_\zeta < \infty. \qquad (4)$$

5) One knows the upper bound C_ζ on the absolute values of ζ_t s.

6) $p(\zeta)$ is a continuous function of ζ which may become zero at isolated points on $[-C_\zeta, C_\zeta]$ only.

7) The variables y_t are measurable.

Comments. Assumption 1) is necessary to ensure that the AR process will be stationary [1, subsection 3.1.3]. Assumption 4) implies

$$\int_{-C_\zeta}^{C_\zeta} P_\zeta(\zeta)d\zeta = 0.$$

This assumption is employed in [2, theorem 2.2.3] for establishing the properties of the adaptive prediction algorithm. Assumption 6) together with 4) essentially means that

$$\mathrm{P}\{\varepsilon' \leq \zeta_t \leq \varepsilon''\} \overset{\text{def}}{=}$$

$$\overset{\text{def}}{=} \int_{-\varepsilon'}^{\varepsilon''} P(\zeta)d\zeta = P(\varepsilon', \varepsilon'') > 0 \qquad (5)$$

for any t and arbitrary $\varepsilon', \varepsilon''$ satisfying

$$-C_\zeta \leq \varepsilon' < \varepsilon'' \leq C_\zeta,$$

where $\mathrm{P}\{\bullet\}$ denotes the probability of a random event included in brackets.

Let $\hat{y}_{t+1}$ be an estimate of the future y_{t+1} to be predicted at each time instant t employing current and past measurements $y_t, y_{t-1},\ldots$ which are available. Further, introduce the prediction quality index as

$$J_{t+1} = \mathrm{M}\{(y_{t+1} - \hat{y}_{t+1})^2 \mid y_t,\ldots,y_0\}, \qquad (6)$$

which is the conditional mean square of the prediction error

$$e_{t+1} = y_{t+1} - \hat{y}_{t+1}. \qquad (7)$$

The problem is to devise a simple adaptive prediction algorithm for minimizing the upper bounds on $\{J_{t+1}\}$ with probability 1 (w.p.1) as t goes to infinity:

$$\lim_{t\to\infty} \mathrm{M}\{(y_{t+1} - \hat{y}_{t+1})^2 \mid y_t,\ldots,y_0\} =$$

$$= \min \qquad \text{(w.p.1).} \qquad (8)$$

Preliminaries

It is shown in [1, subsection 5.1.2] that is the true coefficients of $A(z^{-1})$ in (1) are all known then the algorithm for determining $\hat{y}_{t+1}$ of the form

$$\hat{y}_{t+1} = \theta^{\mathrm{T}}\varphi_t \qquad (9)$$

minimizes (6) for each time instant t. In this case, minimum J_{t+1} is given by

$$\min J_{t+1} = \sigma_\zeta^2.$$

Substituting (9) into expression (7) and using (3), we get

$$e_{t+1} = \zeta_{t+1}. \qquad (10)$$

Since $\mathrm{M}\{\zeta_t\} = 0$, it follows from (10) that the prediction errors e_{t+1} will not be biased: $\mathrm{M}\{e_{t+1}\} = 0$.

Clearly, estimation scheme (9) cannot be used to predict the stochastic process generated by (1) if one does not know the coefficients of $A(z^{-1})$ which are the components of $\theta \in \mathbf{R}^N$. To do this, we need to design an adaptive procedure for estimating unknown θ. Then, instead of (9), we suppose to exploit the equation

$$\hat{y}_{t+1} = \theta_t^{\mathrm{T}} \varphi_t \tag{11}$$

replacing unknown θ in (9) by its suitable estimate

$$\theta_t^{\mathrm{T}} = [a_1(t),\ldots,a_N(t)]$$

to be updated by using some adaptation algorithm.

Adaptive estimation algorithm

It follows from equation (3) together with (4) that

$$| y_t - \theta^{\mathrm{T}} \varphi_{t-1} | \le C_\zeta$$

holds and causes the set of compatible inequalities

$$| y_t - \hat{\theta}^{\mathrm{T}} \varphi_{t-1} | \le C_\zeta, \quad t = 1,2,\ldots \tag{12}$$

with respect to the unknown vector $\hat{\theta}$.

Now, the adaptive estimation algorithm is derived as a recursive procedure for solving inequalities (12) in the form

$$\theta_t = \theta_{t-1} + \gamma_t \frac{f(e_t)}{\| \varphi_{t-1} \|^2} \varphi_{t-1}, \tag{13}$$

where θ_t is the current solution of (12). The prediction error e_t, by virtue of (11) together with (7), may be determined as

$$e_t = y_t - \theta_{t-1}^{\mathrm{T}} \varphi_{t-1}. \tag{14}$$

The variable

$$f(e) = \begin{cases} e - C_\zeta & \text{if} \quad e > C_\zeta, \\ 0 & \text{if} \quad | e | \le C_\zeta, \\ e + C_\zeta & \text{if} \quad e < -C_\zeta \end{cases} \tag{15}$$

represents the dead-zone function depicted in fig.1. γ_t is the scalar variable chosen arbitrarily from

$$0 < \gamma' \le \gamma_t \le \gamma'' < 2 \tag{16}$$

to ensure $a_i(t) \ne 0$ for all $i \in [1, N]$. $\| \bullet \|$ denotes the Euclidean vector norm.

Equation (13) is the stochastic analogy of a gradient projection type procedure for updating θ_t which can be found in [2, 4 – 7] and other works. It is similar to that in the paper [6]. The difference is that equation (13) exploits the continuous dead-zone function $f(e)$ given by (15) whereas [7] deals with discontinuous one.

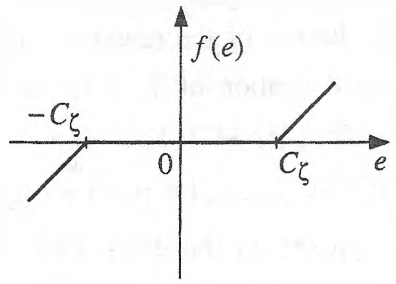

Fig.1. Dead-zone function

According to (13) together with (15), the estimate θ_t is updated only when the absolute value of the current prediction error e_t determined by (14) exceeds the threshold C_ζ. To implement the algorithms described in equations (11), (13) and (14), the adaptive prediction system has to be designed as shown in fig. 2.

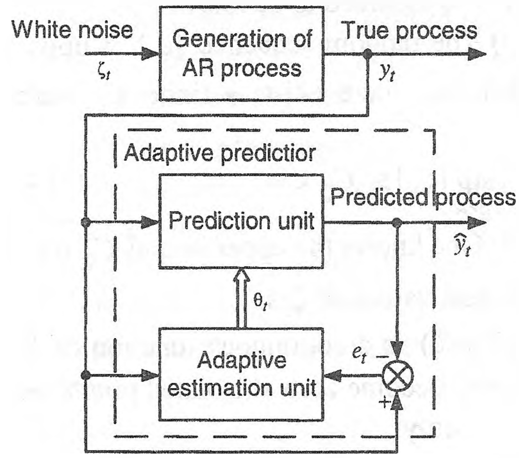

Fig.2. Adaptive prediction system

Convergence analysis

The key convergence property of the adaptive estimation procedure proposed above is given in the next lemma.

Lemma. Subject to assumptions 1) – 5) and 7); if adaptive estimation algorithm (13) together with (14) is applied to model (3), then

$$\overline{\lim_{t\to\infty}}\, e_t \le C_\zeta, \quad \underline{\lim_{t\to\infty}}\, e_t \ge -C_\zeta \tag{17}$$

provided γ_t satisfies (16).

Proof. Let $\tilde{\theta}_t = \theta - \theta_t$. Using

$$V_t = \tilde{\theta}_t^{\mathrm{T}} \tilde{\theta} = \| \theta - \theta_t \|^2$$

as a Lyapunov function candidate [2, 4, 5] and taking into account (3) and (16), from (13) it can finally be written

$$V_0 \geq \gamma'(2-\gamma'')\sum_{t=0}^{\infty}\frac{f^2(e_t)}{\|\varphi_{t-1}\|^2}. \qquad (18)$$

Since $V_t = \tilde{\theta}_t^{\mathrm{T}}\tilde{\theta}$ is finite for any initial θ_0 satisfying $\|\theta\| < \infty$, inequality (18) gives that the series

$$\sum_{t=0}^{\infty}\frac{f^2(e_t)}{\|\varphi_{t-1}\|^2}$$

converges. This means

$$f(e_t)\|\varphi_{t-1}\|^{-1} \to \infty \text{ as } t \to \infty. \qquad (19)$$

In view of assumptions 1) and 4), there exists a finite C_φ such that

$$\|\varphi_t\| \leq C_\varphi < \infty \text{ for all } t. \qquad (20)$$

Applying (20) to (19) gives

$$\lim_{t\to\infty} f(e_t) = 0. \qquad (21)$$

Recalling the definition (15) of the dead-zone function $f(e)$, from (21) result (17) follows. Q.E.D.

The question we now need to answer is as follows. Are assumptions 1) – 7) sufficient to guarantee that the adaptive prediction algorithm synthesized above is capable to achieve goal (8)? It turns out that with additional assumption 6), the following result can be shown to be valid.

Theorem. Under assumptions 1) – 7), the adaptive prediction algorithm described in equations (11), (13) – (16) is asymptotically optimal in the sense that requirement (8) is satisfied.

Proof. By combining equations (3) and (14) and using the notation of $\tilde{\theta}_t$ we get

$$e_t = \tilde{\theta}_{t-1}^{\mathrm{T}}\varphi_{t-1} + \zeta_t,$$

which can be substituted into (17) to obtain

$$\overline{\lim_{t\to\infty}}\ \tilde{\theta}_{t-1}^{\mathrm{T}}\varphi_{t-1} \leq C_\zeta - \overline{\lim_{t\to\infty}}\zeta_t,$$

$$\underline{\lim_{t\to\infty}}\ \tilde{\theta}_{t-1}^{\mathrm{T}}\varphi_{t-1} \geq -C_\zeta + \underline{\lim_{t\to\infty}}\ \zeta_t. \qquad (22)$$

Choose a sufficiently small positive δ and consider the random events

$$B_\delta^+(t) \overset{\mathrm{def}}{=} \{C_\zeta - \delta \leq \zeta_t \leq C_\zeta\}$$

and

$$B_\delta^-(t) \overset{\mathrm{def}}{=} \{-C_\zeta \leq \zeta_t \leq -C_\zeta + \delta\}$$

that are always possible for arbitrary $\delta: 0 < \delta << C_\zeta$ (due to (4)). Setting $\varepsilon' = C_\zeta - \delta$, $\varepsilon'' = C_\zeta$ and $\varepsilon' = -C_\zeta$, $\varepsilon'' = -C_\zeta + \delta$, it can be concluded from (5) that

$$\mathrm{P}[B_\delta^+(t)] = P_\delta^+ > 0$$

and

$$\mathrm{P}[B_\delta^-(t)] = P_\delta^- > 0,$$

respectively. Summing each of these probabilities in t, we have

$$\sum_{t=0}^{\infty}\mathrm{P}[B_\delta^+(t)] = \infty, \ \sum_{t=0}^{\infty}\mathrm{P}[B_\delta^-(t)] = \infty. \qquad (23)$$

Denote by t_k^+ and t_k^- the time instants when

$$\zeta_{t_k^+} \in [C_\zeta - \delta, \ C_\zeta]$$

and

$$\zeta_{t_k^-} \in [-C_\zeta, \ -C_\zeta + \delta],$$

respectively. According to the Borel–Cantelli lemma [8, section 15.3] it follows from divergence of series defined in (23) that the subsequences $\{t_k^+\}$ and $\{t_k^-\}$ will both be infinite (w.p.1) because $\{B_\delta^+(t)\}$ and $\{B_\delta^-(t)\}$ are the sequences of independent events. By the definition of the upper and lower limits [9] this yields

$$\overline{\lim_{t\to\infty}}\ \zeta_t = C_\zeta, \ \underline{\lim_{t\to\infty}}\ \zeta_t = -C_\zeta \quad \text{(w.p.1)}. \qquad (24)$$

Taking into account (24), from (22) we obtain

$$\overline{\lim_{t\to\infty}}\ \tilde{\theta}_{t-1}^{\mathrm{T}}\varphi_{t-1} = 0, \ \underline{\lim_{t\to\infty}}\ \tilde{\theta}_{t-1}^{\mathrm{T}}\varphi_{t-1} = 0 \text{ (w.p.1)}$$

meaning

$$\tilde{\theta}_{t-1}^{\mathrm{T}}\varphi_{t-1} \to 0 \text{ as } t \to \infty \quad \text{(w.p.1)}. \qquad (25)$$

For understanding, introduce the notations $\hat{y}_{t+1}(\theta)$ and $\hat{y}_{t+1}(\theta_t)$ of the prediction estimates caused by equations (9) and (11), respectively. Then, due to (25), we can write

$$\theta^{\mathrm{T}}\varphi_t - \theta_t^{\mathrm{T}}\varphi_t \to 0 \text{ as } t \to \infty \quad \text{(w.p.1)}$$

to see

$$\hat{y}_{t+1}(\theta) - \hat{y}_{t+1}(\theta_t) \to 0 \text{ as } t \to \infty \text{ (w.p.1)}. (26)$$

Since equation (9) as shown minimizes J_{t+1} for $\hat{y}_{t+1} = \hat{y}_{t+1}(\theta)$ and each t, it can now be concluded from (26) that (8) will be satisfied for

$$\hat{y}_{t+1} = \hat{y}_{t+1}(\theta_t).$$

This proves the theorem. Q.E.D.

Corollary. Adaptive prediction algorithm (11), (13) – (16) leads to

$$\overline{\lim_{t\to\infty}}\ J_{t+1} = \sigma_\zeta^2$$

with J_{t+1} defined by (6) for

$$\hat{y}_{t+1} = y_{t+1}(\theta_t).$$

Remark 1. No one establishes the finite convergence of (13). This contrasts with the algorithms advanced in [2].

Remark 2. The theorem does not establish the convergence of estimates θ s to θ in the sense that

$$\lim_{t\to\infty} \theta_t = \theta \qquad \text{(w.p.1).} \quad (27)$$

Nevertheless, requirement (27) is quite not necessary in order to achieve (8).

Thus, we can observe that the performance of the adaptive prediction algorithm which has been proposed is satisfactory.

Simulation results

To illustrate some features of the proposed scheme, the simulations of system (1), (11), (13) with

$$A(z^{-1}) = 1 + a_1 z^{-1} + ... + a_N z^{-N}$$

were carried out setting $a_1 = -1{,}4$, $a_2 = 0{,}7$. Initial estimates were chosen as $a_1(0) = 2{,}0$, $a_2(0) = -1{,}0$ to observe the worst case of parameter uncertainty.

With this choosing, the results of 200-step long simulation when ζ_t was a pseudo-random digital signal (RPDS) distributed uniformly in [-1,0, 1,0] are presented in figs. 3–5.

Fig. 3 shows the behavior of signal y_t and corresponding prediction estimate $\hat{y}_t(\theta_{t-1})$ founded before at the $(t-1)$ th time instant. It can be seen how

$$d_t \stackrel{\text{def}}{=} \hat{y}_t(\theta) - \hat{y}_t(\theta_{t-1})$$

defining the derivation $\hat{y}_t(\theta_{t-1})$ from optimal $\hat{y}_t(\theta)$ tends to zero as t increases (fig. 4).

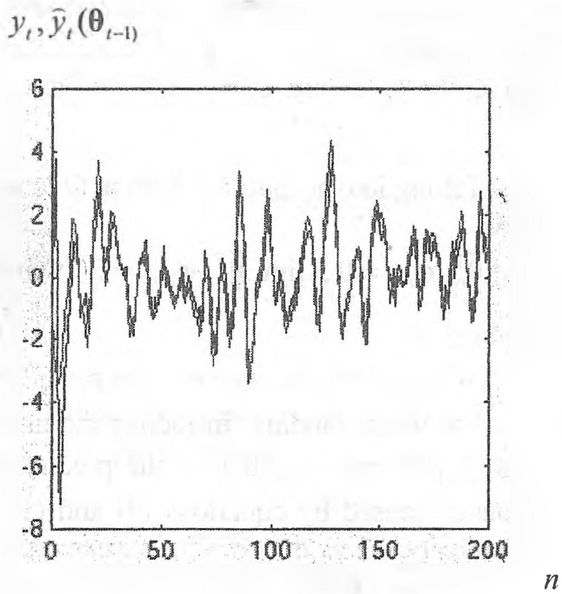

Fig. 3. True variable y_t and its adaptive prediction estimate $\hat{y}_t(\theta_{t-1})$

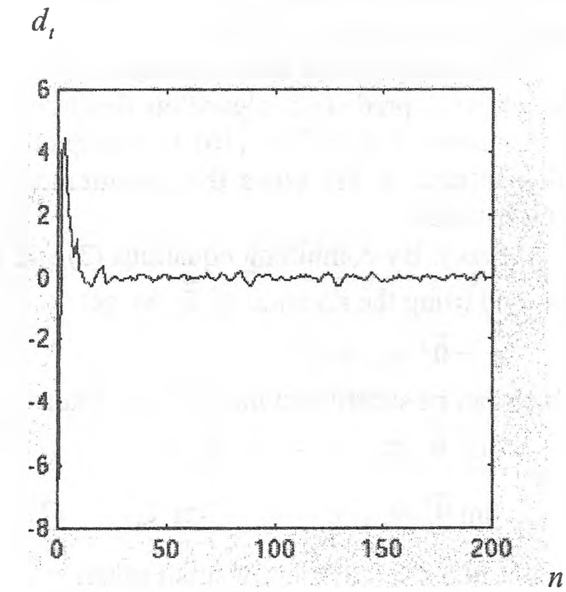

Fig. 4. Derivation of adaptive prediction estimate $\hat{y}_t(\theta_{t-1})$ from optimal $\hat{y}_t(\theta_t)$

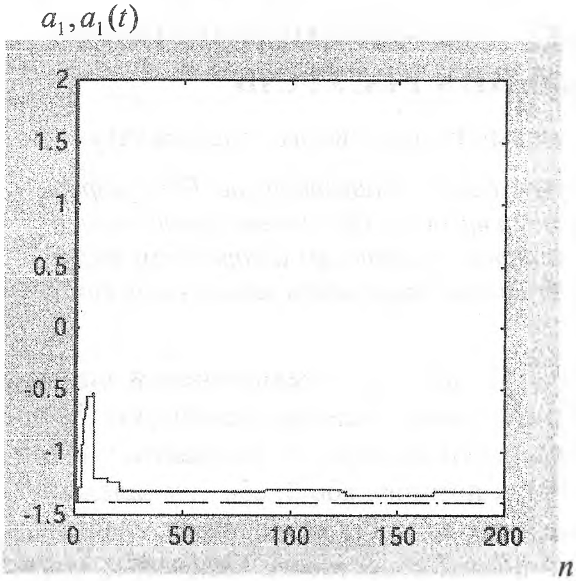

Fig. 5. a)

Variables: a) – a_1 (dotted line), $a_1(t)$

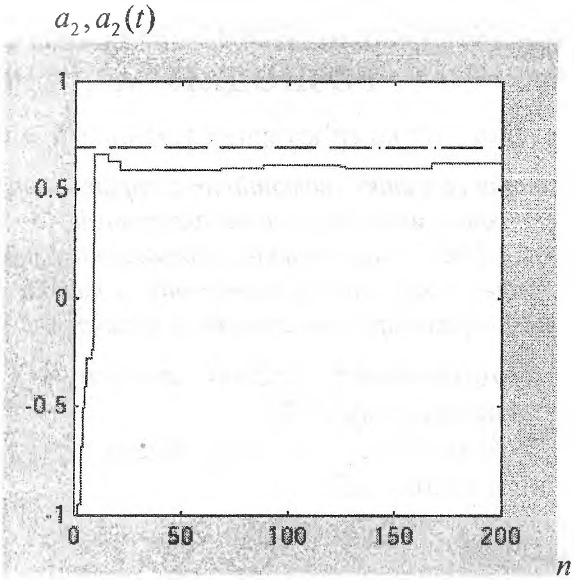

Fig.5. b)

Variables b) – a_2 (dotted line), $a_2(t)$

The evolution of the parameter estimates $a_1(t)$ and $a_2(t)$ is depicted in fig.5. It shows that they converge fast enough to a neighborhood of their true values a_1 and a_2, respectively. The number of the corrections of these estimates was 36.

Conclusion

A simple adaptive estimation algorithm based on projection type procedure may be applied for the prediction of stochastic AR process caused by the white noise whose values are upper bounded. It has been proved that this algorithm ensures the asymptotical optimality of the predicted estimates. Simulation results demonstrate its effectiveness. From a practical point of view, it is attractive that the above mentioned algorithm requires small computational effort for its implementation.

References

1. *Бокс Дж. и Дженкинс Г.* Анализ временных рядов. Прогноз и управление. Вып.1. – М.: Мир, 1974. – 406 с.

2. *Фомин В.Н., Фрадков А.Л., Якубович В.Я.* Адаптивное управление динамическими объектами. – М.: Наука, 1981. – 447 с.

3. *Фомин В.Н.* Рекуррентное оценивание и адаптивная фильтрация. –М.: Наука, 1984. – 288 с.

4. *Goodwin G. C., Sin K. S.* Adaptive filtering, prediction and control. – Engewood Cliffs, NJ.: Prentice-Hall, 1984. – 572 p.

5. *Landau I.D., Lozano R., M'Saad M.* Adaptive control. – London: Springer, 1997. – 562 p.

6. *Zhiteckij L. S.* Adaptive control of systems subjected to bounded disturbances //Bounding approaches to system identification /M. Milanese etc., eds. – New York–London: Plenum Press, 1996. – Chapt. 24. – P. 383–407.

7. *Житецкий Л.С.* Обучение управлению технологическим процессом по наблюдению за действиями опытного оператора в замкнутой человеко-машинной системе // Кибернетика и вычисл. техника. – 1998. – Вып. 116. – С. 14–24.

8. *Лоэв М.* Теория вероятностей /Пер. с англ. – М.: Изд-во иностранной лит., 1962. – 719 с.

9. *Корн Г., Корн Т.* Справочник по математике. М.: Наука, 1974. – 831 с.