ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2009, Vol.38, No.1

PARAMETER IDENTIFICATION FOR ASYMMETRICAL POLYNOMIAL LOSS FUNCTION

Piotr Kulczycki * **, **Aleksander Mazgaj** **

* *Polish Academy of Sciences, Systems Research Institute*
ul. Newelska 6, PL-01-447 Warsaw, Poland
e-mail: kulczycki@ibspan.waw.pl
** *Cracow University of Technology, Department of Automatic Control*
ul. Warszawska 24, PL-31-155 Cracow, Poland

Abstract. The parameter identification for problems where losses arising from overestimation and underestimation are different and can be described by an asymmetrical and polynomial function is investigated in this paper. The Bayes decision rule allowing to minimize potential losses is used. Calculation algorithms are based on the nonparametric methodology of statistical kernel estimators, which releases the method from dependence on distribution type. Three basic cases are considered in detail: a linear, a quadratic, and finally a general concept for a higher degree polynomial – here the cube-case is described in detail as an example. For each of them, the final result constitutes a numerical procedure enabling to effectively calculate the optimal value of a parameter in question, presented in its complete form which demands neither detailed knowledge of the theoretical aspects nor laborious research of the user. Although the above method was investigated from the point of view of automatic control problems, it is universal in character and can be applied to a wide range of tasks, also outside the realm of engineering.

Keywords: parameter identification, Bayes estimation, asymmetrical polynomial loss function, nonparametric estimation, kernel estimators, numerical algorithm.

1. Introduction

The dynamic development of computer science and metrological technology now means obtaining more and more precise measurements, as well as collecting received information with the aim of further processing forever more complex control algorithms. However, more precise and faster instrumentation only manage to lessen the error in measurements, not eliminate it entirely. As an example: one of the main parameters of mechanical systems – mass, is often not measured at all, or only grossly assumed on the basis of general conditioning. Moreover, it often changes with consumption of fuel or other substances used in the production process.

Quality of control is influenced not only by errors in measurement but also the structure of the model itself, which directly determines the range of possibilities for the reflection of properties in real processes. A full consideration of all phenomena occurring in an object is not just impossible, but even pointless, as the complexity of such a model would make it unusable. Thus, it is not achievable to accurately measure parameters occurring in a model, not only due to metrological limitations, but also because of structural reasons, where many parameters represent a wide range of phenomena, and so, often do not possess "exact" values. For further details in this topic see [3, 4, 17-19, 21], where subject literature can also be found, which in the case of identification is particularly diverse regarding wealth of methodology as well as the variety of tasks it is applied to.

As identification is in practice always subject to a higher goal (usually conditioned by the control algorithm), very valuable results can be obtained thanks to the consideration – during estimation of the parameters' values – of the losses implied through errors, as mentioned earlier, unchecked in practice. In control engineering applications such losses can often be described by the function assuming the following asymmetrical and polynomial form:

$$l(\hat{\lambda},\lambda)=\begin{cases}(-1)^k\, a\,(\hat{\lambda}-\lambda)^k & \text{for } \hat{\lambda}-\lambda\le 0\\ b\,(\hat{\lambda}-\lambda)^k & \text{for } \hat{\lambda}-\lambda\ge 0\end{cases}, \qquad (1)$$

with $k\in \mathrm{N}\setminus\{0\}$, where the coefficients a and b are positive, and may differ, while λ and $\hat{\lambda}$ mean the values of the parameter under consideration and its estimator, respectively. Assuming the above form of the loss function constitutes a comfortable compromise between accuracy of description of losses resulting from modeling errors, and complexity, and in consequence usefulness of the approach proposed.

Similar conditioning can also be shown for many problems outside the area of automatic control, or even broadly understood engineering. For example, as stated by Kahneman – a Nobel laureate in the field of economics – behavior in business is not completely rational. According to his theory, a human reacts strongly to extreme stimuli and is disposed to exaggerating losses as well as undervaluing gains. This fear of large losses enables animals to survive in nature, however, in the economy it leads to an illogical dread of change. Therefore, if one describes the psychological preferences of the ordinary person, then it can be defined by formula (1), for example with $k = 2$, i.e. in the quadratic case. Here an inverse relationship to losses and gains is represented by nonsymmetry, and fear of extremes by quadratic form.

Consider therefore the typical situation where one has m values of the investigated parameter x_1, $x_2, \ldots, x_m$ obtained directly by measuring or with the aid of auxiliary quantities. In this paper, the uncertainty of the examined parameter is considered with a probabilistic approach. For identification of characteristics of probabilistic measure the statistical kernel estimators methodology [8, 20, 22] will be used. This is the current leading concept of nonparametric estimation, the present development of which is connected with a dynamic growth of possibilities and in particular the universal availability of computer systems. As opposed to classical parametric estimation, where firstly one arbitrarily assumes a typical probability distribution type, and next calculates the values of its parameters, in the case of kernel estimators practically no assumptions are made, and atypical, complex and multimodal distributions can be treated exactly the same as simple, even textbook cases. Finally, a Bayes estimator, optimal in the sense of minimization of expectation value of losses, will be found according to principles of the Bayes decision rule [2].

Three basic cases will be investigated in the following: linear (Section 3.1), quadratic (Section 3.2), and higher degree polynomial (Section 3.3) – here the cube-case will be described in detail. In every case, the final result will be a procedure for the calculation of values for an optimal estimator. Thanks to the presence of complete established algorithms applied here, as well as clear analytical forms of quantities used, its practical implementation will consist of only routine introduction dependencies. The proposed procedure is universal and can be applied in a wide range of tasks, not only in the field of engineering. Furthermore, the method worked out can be used for other uncertainty approaches apart from that of probability, e.g. fuzzy logic [7].

The preliminary version of this article was presented as the conference-paper [15]. Its main theses have also been included in the synopses [10-12].

List of Notations

In order of appearance:

l – loss function
λ – parameter
$\hat{\lambda}$ – estimator of the parameter λ
k – degree of asymmetrical polynomial loss function
a – coefficient of the asymmetrical polynomial loss function regarding underestimation
b – coefficient of the asymmetrical polynomial loss function regarding overestimation
N – set of natural numbers
x_1, $x_2, \ldots, x_m$ – random sample, interpreted here as the measurements of the estimated parameter
m – size of random sample
R – set of real numbers
Z – set of states of nature
D – set of possible decisions
f – density of probability distribution
z – state of nature
d – decision
l_B – Bayes loss function
d_B – Bayes decision
X – random variable
$\hat{f}$ – kernel estimator of density of probability distribution
x – independent variable, interpreted here as estimated parameter
i, j – natural indices
K – kernel
h – smoothing parameter
r – degree of the plug-in method
$\hat{\sigma}$ – estimator of standard deviation
$\tilde{\psi}_{10}$, ψ_s, g_1, g_2, g_3, g, p, P, R, L – auxiliary parameters and functions used in the plug-in method
s_1, $s_2, \ldots, s_m$ – modifying parameters
c – intensity of modifying procedure
$\bar{s}$ – auxiliary parameter used for modifying procedure
M – mass submitted to control
t – time
X_1 – first coordinate of state of a dynamic system
X_2 – second coordinate of state of a dynamic system
U – control
$J_{\text{t-o}}$ – minimum-time performance index for optimal control
T_U – time to reach the origin when the control U is used
$\hat{M}$ – estimator of the parameter M
$\hat{x}$ – estimator of the parameter x

I – primitive function of the kernel K
y – auxiliary variable
U_i, V_i, W_i – auxiliary functions used in the investigated algorithm
L – auxiliary function for Newton's algorithm
J_q – quadratic performance index for optimal control
$\boldsymbol{Q}$, $\boldsymbol{R}$ – matrices of losses of the quadratic performance index
$\boldsymbol{X}$ – state vector of a dynamic system
$\boldsymbol{U}$ – control vector
Λ – parameter
$\hat{\Lambda}$ – estimator of the parameter Λ
$N(\mu,\sigma)$ – normal distribution with the expectation value μ and the standard deviation σ

2. Mathematical Preliminaries

2.1. Bayes Decision Rule

The main aim of decision theory [2] is the selection of a concrete decision based only on a representation of measure characterizing the imprecision of states of nature. Let there be given a nonempty set of states of nature $Z = \mathrm{R}$, and a nonempty set of possible decisions $D \subset \mathrm{R}$. Assume that the imprecision of states of nature is of probability type and its distribution is described by the density $f: \mathrm{R} \to [0,\infty)$. Let there be given also the loss function $l: D \times Z \to \mathrm{R}$, while its values $l(d,z)$ can be interpreted as losses occurring in a hypothetical case, when the state of nature is z and the decision d is taken. If for every $d \in D$ the integral $\int_{\mathrm{R}} l(d,z) f(z)\,\mathrm{d}z$ exists, then the Bayes loss function $l_B: D \to \mathrm{R} \cup \{\pm\infty\}$ can be defined as

$$l_B(d) = \int_{\mathrm{R}} l(d,z)\, f(z)\,\mathrm{d}z \quad . \tag{2}$$

Every element $d_B \in D$ such that $l_B(d_B) = \min_{d \in D} l_B(d)$ is called a Bayes decision, and the above procedure – a Bayes decision rule. As the above definition shows, the Bayes decision d_B is chosen from the elements of the set of possible decisions so as to minimize the value of the Bayes loss function – from the probabilistic point of view: the expectation value of losses after the decision d was made. Further details can be found in the book [2].

2.2. Statistical Kernel Estimators

Let the one-dimensional random variable X, with a distribution having the density f, be given. Its kernel estimator $f: \mathrm{R} \to [0,\infty)$ is calculated on the basis of the m-element simple random sample x_1, $x_2, \ldots, x_m$ acquired experimentally from the variable X, and is defined in its basic form by the formula

$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right) \quad , \tag{3}$$

where the function $K: \mathrm{R} \to [0,\infty)$, which is measurable, symmetrical relative to zero, and has a weak global maximum at this point, fulfilling the condition $\int_{\mathrm{R}} K(x)\,\mathrm{d}x = 1$, and is called a kernel, whereas the positive coefficient h is known as a smoothing parameter [8, 20, 22].

Fixing values introduced in definition (3), i.e. choosing the form of the kernel K and calculating the smoothing parameter h value, is most often carried out using the mean square criterion.

Thus, from the statistical point of view, the form of the kernel seems not to have essential meaning, thanks to which it becomes possible for the choice of the function K to be arbitrary, taking into account above all required properties of the estimator obtained, e.g. class of regularity, positive values, or other qualities important in the case of a particular problem, especially the convenience of calculations.

As opposed to the form of the kernel, the value of the smoothing parameter h has significant influence on the quality of the estimator obtained. In any case, convenient algorithms have been developed in order to calculate this value on the basis of a random sample. For the one-dimensional case considered here, the most convenient is the so-called plug-in method. Its concept consists of the calculation of this parameter using an approximate method, and after r steps improving the result, one obtains a value close to optimal. In practice, it is taken that $r \geq 2$, with the lowest possible value recommended. On the basis of simulation research carried out for the needs of the task worked out in this paper, $r = 3$ was assumed. In this case the plug-in method consists of the application of the following formulas:

$$\tilde{\psi}_{10} = \frac{-945}{64\pi^{1/2}\hat{\sigma}^{11}} \quad , \tag{4}$$

while $\hat{\sigma}$ denotes the estimator of a standard deviation

$$\hat{\sigma} = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m} x_i^{\,2} - \frac{1}{m(m-1)}(\sum_{i=1}^{m} x_i)^2} \tag{5}$$

and

$$g_1 = \left(\frac{-2L^{(8)}(0)}{mP(L)\tilde{\psi}_{10}}\right)^{1/11} \tag{6}$$

$$g_2 = \left(\frac{-2L^{(6)}(0)}{mP(L)\psi_8(g_1)}\right)^{1/9} \tag{7}$$

$$g_3 = \left(\frac{-2L^{(4)}(0)}{mP(L)\psi_6(g_2)}\right)^{1/7} \quad ; \tag{8}$$

finally

$$h = \left(\frac{R(K)}{mP(K)^2 \psi_4(g_3)} \right)^{1/5} \quad , \tag{9}$$

where the finite quantities are defined as

$$R(K) = \int_{\mathrm{R}} K(x)^2 \,\mathrm{d}x \tag{10}$$

$$P(K) = \int_{\mathrm{R}} x^2 K(x) \,\mathrm{d}x \tag{11}$$

$$\psi_p(g) = \frac{1}{m^2 g^{p+1}} \sum_{i=1}^{m} \sum_{j=1}^{m} L^{(p)} \left(\frac{x_i - x_j}{g} \right) \quad \text{for} \quad p = 1, 2, \ldots \quad . \tag{12}$$

The kernel *K*, applied in estimator (3), is used only in the last step. In all other steps, the different kernel *L*, more convenient for the plug-in method, may be used.

The value of the smoothing parameter *h* introduced in definition (3) is the same for all kernels, mapped to particular elements of the random sample. In “dense” areas of such elements, the above value should be lessened (which allows for better showing of specific features of the distribution), as opposed to areas where such elements are “sparse” and it should be increased (which causes additional smoothing of “tails”). The parameter modification procedure achieves this goal in compliance with the following algorithm:

(A) the kernel estimator $\hat{f}$ is specified according to basic formula (3);

(B) the modifying parameters $s_i > 0$ of the form

$$s_i = \left(\frac{\hat{f}(x_i)}{\bar{s}} \right)^{-c} \quad \text{for } i = 1, 2, \ldots, m \quad , \tag{13}$$

are calculated, while the nonnegative parameter c shows the intensity of the modification procedure, whereas $\bar{s}$ is the geometric mean of the numbers $\hat{f}(x_1)$, $\hat{f}(x_2)$, ..., $\hat{f}(x_m)$;

(C) the kernel estimator with the modification of the smoothing parameter, is ultimately defined as

$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^{m} \frac{1}{s_i} K \left(\frac{x - x_i}{hs_i} \right) \quad . \tag{14}$$

Note that taking $c = 0$ results in $s_i \equiv 1$ and consequently basic form (3).

Details of the above-presented methodology of statistical kernel estimators can be found in the books [8, 20, 22].

3. The Algorithm

3.1. Linear Case

As an example illustrating the investigations presented in this section, an optimal control [1] problem will be considered. The control performance index, which exists here, can also refer to quality of identification allowing the creation of an optimal procedure for estimation of object parameter values, thereby notably lowering excess sensitivity of such systems to the inaccuracy of modeling.

Thus, consider the following dynamic system:

$$\begin{bmatrix} \dot{X}_1(t) \\ \dot{X}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \dfrac{1}{M} \end{bmatrix} U(t) \quad , \tag{15}$$

where the positive parameter *M* represents a mass submitted to a force according to Newton’s second law of dynamics [1 – Section 7.2]. Then X_1, X_2 and U denote position and velocity of the mass, and the force regarded here as a control, respectively. Consider the time-optimal control task, the basic form of which consists of bringing the system’s state to the origin, in minimal and finite time, assuming the control values are bounded. Thus, the performance index $J_{\text{t-o}}$ is given here as

$$J_{\text{t-o}}(U) = T_U \quad , \tag{16}$$

where T_U denotes the time to reach the origin when the control *U* is used, assumed as infinity if the origin is not reached at all with this control. For details see the classic textbook [1 – Chapter 7]. Fundamental meaning for phenomena existing in the control system lies in the proper identification of value of the parameter *M*. The control is defined in relation to the value of the estimator $\hat{M}$, actually different from the value of the parameter M in the object. A detailed analysis can be found in the paper [16].

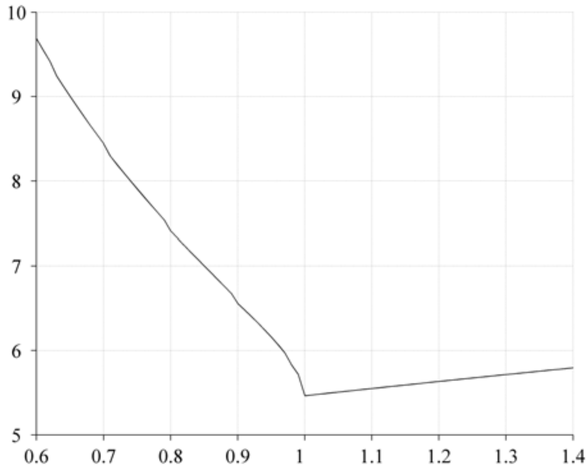

Figure 1. Values of performance index (16) for values of the estimator $\hat{M}$ (with $M = 1$)

In the purely hypothetical case of $\hat{M} = M$, i.e. when the value of the estimator of this parameter is equal to its true value, the process is regular in character. The system’s state reaches the origin in minimal and finite time. However, in the event of underestimation (i.e. $\hat{M} < M$), overregulation occur in the system

– its state oscillates around the origin and reaches it in a finite time, albeit larger than the minimal. Next, in the case of overestimation (i.e. when $\hat{M} > M$), the system's state moves along a sliding trajectory and finally reaches the origin in a finite time, again larger than the minimal. Figure 1 shows the graph of the performance index for values of the estimator $\hat{M}$. One can note that an increase in this index is roughly proportional to the estimation error $|\hat{M} - M|$, although with different coefficients for positive and negative errors. The resulting losses can so be described in the form of an asymmetrical linear loss function, i.e. given by formula (1) with $k = 1$.

The parameter under investigation, whose value is to be estimated, will be denoted by x hereinafter. In order to adhere to the principles of decision theory presented in Section 2.1, it will be treated here as the value of a random variable. According to point estimation methodology, it is assumed that the metrologically achieved measurements of the above parameter, i.e. x_1, $x_2, \ldots, x_m$, are the sum of its "true" (although unknown) value and random disturbances of various origin. The goal of this research is the calculation of the estimator of this parameter (hereinafter denoted by $\hat{x}$), which would approximate the "true" value – the best from the point of view of a practical problem investigated. In order to solve this task, the Bayes decision rule will be used, ensuring a minimum of expectation value of losses. According to the conditions formulated above, the loss function is assumed in asymmetrical linear form:

$$l(\hat{x},x) = \begin{cases} -a(\hat{x} - x) & \text{for} \quad \hat{x} - x \le 0 \\ b(\hat{x} - x) & \text{for} \quad \hat{x} - x \ge 0 \end{cases} , \tag{17}$$

while the coefficients a and b are positive and not necessarily equal to each other. Thus, the Bayes loss function (2) is given by the formula

$$l_B(\hat{x}) = b\int_{-\infty}^{\hat{x}} (\hat{x} - x) f(x)\,\mathrm{d}x - a\int_{\hat{x}}^{\infty} (\hat{x} - x) f(x)\,\mathrm{d}x , \tag{18}$$

where $f : \mathrm{R} \to [0,\infty)$ denotes the density of distribution of a random variable representing the uncertainty of states of nature, i.e. the parameter in question. It is readily shown that the function l_B attains its minimum at the value being a solution of the following equation with the argument $\hat{x}$:

$$\int_{-\infty}^{\hat{x}} f(x)\,\mathrm{d}x - \frac{a}{a+b} = 0 . \tag{19}$$

Since $0 < a/(a+b) < 1$, a solution for the above equation exists, and if the function f has connected support, e.g. it is positive, this solution is unique. Moreover, thanks to equality

$$\frac{a}{a+b} = \frac{\frac{a}{b}}{\frac{a}{b} + 1} , \tag{20}$$

it is not necessary to identify the parameters a and b separately, rather only their ratio.

The identification of the density f present in condition (19) will be carried out using statistical kernel estimators, presented in Section 2.2. Then one should choose a continuous kernel of positive values and also so that the function $I : \mathrm{R} \to \mathrm{R}$ such that $I(x) = \int_{-\infty}^{x} K(y)\,\mathrm{d}y$ can be expressed by relatively simple analytical formula. In consequence, this result in a similar property regarding the function $U_i : \mathrm{R} \to \mathrm{R}$ for any fixed $i = 1, 2, \ldots, m$ defined as

$$U_i(x) = \frac{1}{h}\int_{-\infty}^{x} K\left(\frac{y - x_i}{h}\right)\mathrm{d}y . \tag{21}$$

Then criterion (19) can be expressed equivalently in a form of

$$\frac{h}{m}\sum_{i=1}^{m} U_i(\hat{x}) - \frac{a}{(a+b)} = 0 . \tag{22}$$

If the left-hand side of the above formula is denoted by $L(\hat{x})$, its derivative is simply

$$L'(\hat{x}) = \hat{f}(\hat{x}) , \tag{23}$$

where $\hat{f}$ was given by definition (3). In this situation, the solution of criterion (19) can be calculated numerically on the basis of Newton's algorithm [6] as the limit of the sequence $\{\hat{x}_j\}_{j=0}^{\infty}$ defined by

$$\hat{x}_0 = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{24}$$

$$\hat{x}_{j+1} = \hat{x}_j - \frac{L(\hat{x}_j)}{L'(\hat{x}_j)} \quad \text{for} \quad j = 0, 1, \ldots , \tag{25}$$

with the functions L and L' being given by formulas (22)-(23), whereas a stop criterion takes on the form

$$|\hat{x}_j - \hat{x}_{j-1}| \le 0.01\,\hat{\sigma} , \tag{26}$$

where $\hat{\sigma}$ denotes the estimator of the standard deviation (5).

In the linear case worked out above, the Cauchy kernel

$$K(x) = \frac{2}{\pi}\frac{1}{(1+x^2)^2} \tag{27}$$

is proposed. Then, for the generalized form of the kernel estimator (14):

$$U_i(x) = \frac{1}{\pi}\operatorname{arctg}\left(\frac{x-x_i}{hs_i}\right) + \frac{\dfrac{x-x_i}{hs_i}}{\pi\left[1+\left(\dfrac{x-x_i}{hs_i}\right)^2\right]} + \frac{1}{2} \quad . \quad (28)$$

(for the basic form (3) one should put $s_i \equiv 1$), and also

$$P(K) = 1 \quad (29)$$

$$R(K) = \frac{5}{4\pi} \quad . \quad (30)$$

Further if the kernel L present in the plug-in method is taken as the most commonly used here normal kernel

$$L(x)) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad , \quad (31)$$

then

$$P(L) = 1 \quad (32)$$

$$L^{(4)}(x) = \frac{1}{\sqrt{2\pi}}(x^4 - 6x^2 + 3)\exp(-\frac{1}{2}x^2) \quad (33)$$

$$L^{(6)}(x) = \frac{1}{\sqrt{2\pi}}(x^6 - 15x^4 + 45x^2 - 15)\exp(-\frac{1}{2}x^2) \quad (34)$$

$$L^{(8)}(x) = \frac{1}{\sqrt{2\pi}}(x^8 - 28x^6 + 210x^4 - 420x^2 + 105)\exp(-\frac{1}{2}x^2) \quad (35)$$

which completes all quantities necessary for implementing the algorithm worked out.

Primary investigations in the linear case were published in the paper [9]. The conditional version constitutes the subject of the article [13].

3.2. Quadratic Case

As an example to illustrate the reason for the case investigated below, consider the problem concerning the classical task of optimal control for the quadratic performance index with infinite end time

$$J_q(\boldsymbol{U}) = \int_0^\infty \boldsymbol{X}(t)^{\mathrm{T}} \boldsymbol{Q}\, \boldsymbol{X}(t) + \boldsymbol{U}(t)^{\mathrm{T}} \boldsymbol{R}\, \boldsymbol{U}(t)\, \mathrm{d}t \quad , \quad (36)$$

while $\boldsymbol{X}$ and $\boldsymbol{U}$ denote state and control vectors, whereas $\boldsymbol{Q}$ and $\boldsymbol{R}$ mean loss matrices defined nonnegative and positive, respectively. For details see [1 – Section 9]. The object is the dynamic system

$$\begin{bmatrix} \dot{X}_1(t) \\ \dot{X}_2(t) \end{bmatrix} = \begin{bmatrix} \Lambda & 1 \\ 0 & \Lambda \end{bmatrix} \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \Lambda \end{bmatrix} U(t) \quad , \quad (37)$$

while $\Lambda \in \mathrm{R}\backslash\{0\}$. Moreover, let $\hat{\Lambda} \in R \setminus \{0\}$ represent an estimator of the parameter Λ. An optimal feedback controller is defined on the basis of the value $\hat{\Lambda}$, not necessarily equal to the value of the parameter Λ existing in the object. For the sake of simplicity, assume the unit matrix $\boldsymbol{Q}$ as well as the matrix (here parameter) $\boldsymbol{R}$. The values of the performance index obtained for a particular $\hat{\Lambda}$, are shown in Figure 2. One can see that the resulting graph can be described with great precision by a quadratic function with different coefficients for positive and negative errors, which in fact proves that over- and underestimation of the parameter Λ have other results on the performance index value.

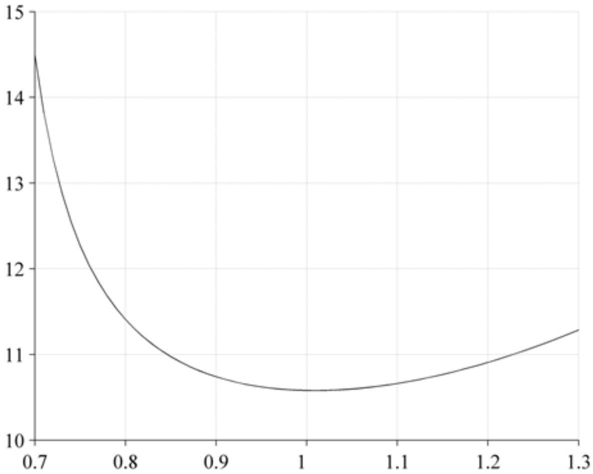

Figure 2. Values of performance index (36) for values of the estimator $\hat{\Lambda}$, with $\Lambda = 1$.

To use an analogous methodology to that of the linear case considered in the previous section, the loss function is assumed in quadratic and asymmetrical form defined as

$$l(\hat{x}, x) = \begin{cases} a\,(\hat{x} - x)^2 & \text{for} \quad \hat{x} - x \le 0 \\ b\,(\hat{x} - x)^2 & \text{for} \quad \hat{x} - x \ge 0 \end{cases} \quad , \quad (38)$$

while the coefficients a and b are positive and not necessarily equal to each other. Thus, the Bayes loss function (2) is given by the formula

$$l_B(\hat{x}) = a\int_{\hat{x}}^{\infty} (\hat{x} - x)^2 f(x)\,\mathrm{d}x + b\int_{-\infty}^{\hat{x}} (\hat{x} - x)^2 f(x)\,\mathrm{d}x \quad . \quad (39)$$

One can show that the function l_B attains its minimum at the value $\hat{x}$ being a solution of the equation

$$(a-b)\int_{-\infty}^{\hat{x}} (\hat{x} - x) f(x)\,\mathrm{d}x - a\int_{-\infty}^{\infty} (\hat{x} - x) f(x)\,\mathrm{d}x = 0 \quad . \quad (40)$$

This solution exists and is unique. As in the linear case, dividing the above equation by b, note that it is necessary to identify only the ratio of the parameters a and b.

Solution of equation (40) for a general case is not an easy task. However, if estimation of the density f is reached using statistical kernel estimators, then – thanks to a proper choice of the kernel form – one can design an effective numerical algorithm to this end. Let, therefore, a continuous kernel of positive values, fulfilling the condition

$$\int_{-\infty}^{\infty} xK(x)\,\mathrm{d}x < \infty \tag{41}$$

be given. Besides the functions U_i introduced in Section 3.1, let for any fixed $i = 1, 2, \ldots, m$ the functions $V_i\colon \mathrm{R} \to \mathrm{R}$ be defined as

$$V_i(x) = \frac{1}{h}\int_{-\infty}^{x} yK\left(\frac{y - x_i}{h}\right)\mathrm{d}y \quad . \tag{42}$$

The kernel K should be chosen so the function $J\colon \mathrm{R} \to \mathrm{R}$ such that $J(x) = \int_{-\infty}^{x} yK(y)\,\mathrm{d}y$ be expressed by a convenient analytical formula.

If an expected value is estimated by the arithmetical mean value of a sample, then criterion (40) can be described equivalently as

$$\sum_{i=1}^{m}[(a-b)(\hat{x}U_i(\hat{x}) - V_i(\hat{x})) + ax_i] - a\hat{x}m = 0 \ . \tag{43}$$

If the left-hand side of the above formula is denoted by $L(\hat{x})$, then – using the equality $V_i'(\hat{x}) = \hat{x}U_i'(\hat{x})$ directly resulting from dependencies (21) and (42) – one can express the value of its derivative as

$$L'(\hat{x}) = \sum_{i=1}^{m}[(a-b)U_i(\hat{x})] - am \quad . \tag{44}$$

In this situation, the solution of criterion (40) can be calculated numerically on the basis of Newton's algorithm (24)-(25).

In the quadratic case also Cauchy kernel (27) is proposed; then formula (28) remains true and additionally for the general form of the kernel estimator (14):

$$V_i(x) = x_i\left(\frac{1}{\pi}\mathrm{arctg}\left(\frac{x - x_i}{hs_i}\right) + \frac{\frac{x - x_i}{hs_i}}{\pi\left[1 + \left(\frac{x - x_i}{hs_i}\right)^2\right]} + \frac{1}{2}\right) - \frac{hs_i}{\pi\left[1 + \left(\frac{x - x_i}{hs_i}\right)^2\right]} \ . \tag{45}$$

Also dependencies (29)-(35) remain unchanged.

Primary investigations concerning the quadratic case, including also the multidimensional, were published in the article [14].

3.3. Higher Degree Polynomial Case

In this section, detailed investigations presented earlier will be supplemented with the polynomial case, that is where the loss function is an asymmetrical monomial of the order $k \geq 2$ and is therefore given by the following formula:

$$l(\hat{x}, x) = \begin{cases} (-1)^k a(\hat{x} - x)^k & \text{for } \ \hat{x} - x \leq 0 \\ b(\hat{x} - x)^k & \text{for } \ \hat{x} - x \geq 0 \end{cases} , \tag{46}$$

while the coefficients a and b are positive, and may differ. Criterion for the optimal estimator $\hat{x}$ is given here in the form

$$\begin{aligned} &(-1)^k ak\int_{\hat{x}}^{\infty}(\hat{x} - x)^{k-1}f(x)\,\mathrm{d}x \\ &+ bk\int_{-\infty}^{\hat{x}}(\hat{x} - x)^{k-1}f(x)\,\mathrm{d}x = 0 \end{aligned} \ . \tag{47}$$

The solution of the above equation exists and is unique.

When the statistical kernel estimators are used with respect to the density f, it is possible again to create an efficient numerical algorithm enabling equation (47) to be solved. Let the kernel K be continuous, of positive values and fulfilling the following condition:

$$\int_{-\infty}^{\infty} x^{k-1}K(x)\,\mathrm{d}x < \infty \quad . \tag{48}$$

For clarity of presentation, the case $k = 3$ is presented below. Thus, equation (47) takes on the equivalent form

$$\begin{aligned} &(a+b)\left(\hat{x}^2\int_{-\infty}^{\hat{x}}f(x)\,\mathrm{d}x - 2\hat{x}\int_{-\infty}^{\hat{x}}xf(x)\,\mathrm{d}x + \int_{-\infty}^{\hat{x}}x^2 f(x)\,\mathrm{d}x\right) \\ &-a\left(\hat{x}^2 - 2\hat{x}\int_{-\infty}^{\infty}xf(x)\,\mathrm{d}x + \int_{-\infty}^{\infty}x^2f(x)\,\mathrm{d}x\right) = 0 \end{aligned} \ . \tag{49}$$

Now, with any fixed $i = 1, 2, \ldots, m$, let the functions U_i and V_i defined by dependencies (21) and (42) be given, and furthermore $W_i\colon \mathrm{R} \to \mathrm{R}$ be introduced as

$$W_i(x) = \frac{1}{h}\int_{-\infty}^{x} y^2K\left(\frac{y - x_i}{h}\right)\mathrm{d}y \quad . \tag{50}$$

Making use of the above notations, condition (49) can be expressed in the following form

$$\begin{aligned} &\sum_{i=1}^{m}\Big[(a+b)\left(x^2U_i(x) - 2xV_i(x) + W_i(x)\right) + 2ax_i x \\ &-\lim_{x\to\infty}W_i(x)\Big] - amx^2 = 0 \end{aligned} \ . \tag{51}$$

If the left-hand side of the above formula is denoted as $L(x)$, then – also taking into account the equalities $V_i'(x) = xU_i'(x)$ and $W_i'(x) = xV_i'(x)$ resulting from dependencies (21), (42) and (50) – the derivative of the function L is

$$\begin{aligned} L'(x) = &\sum_{i=1}^{m}\Big[2(a+b)\left(xU_i(x) - V_i(x)\right) + 2ax_i\Big] \\ &-2amx \end{aligned} \ . \tag{52}$$

Finally, the desired estimator can be calculated numerically through Newton's algorithm (24)-(25), while the functions L and L' are given by formulas (51)-(52).

The Cauchy kernel (27) must by modified here to the form

$$K(x) = \frac{8}{3\pi} \frac{1}{(1+x^2)^3} \quad . \tag{53}$$

An increase of the power in the denominator has been implied with the necessity of ensuring the fulfillment of condition (48). Here for the general form of the kernel estimator (14):

$$U_i(x) = \frac{\left(\frac{x-x_i}{hs_i}\right)^3}{\pi\left(1+\left(\frac{x-x_i}{hs_i}\right)^2\right)^2} + \frac{5\frac{x-x_i}{hs_i}}{3\pi\left(1+\left(\frac{x-x_i}{hs_i}\right)^2\right)^2} + \frac{1}{\pi}\operatorname{arctg}\left(\frac{x-x_i}{hs_i}\right) + \frac{1}{2} \tag{54}$$

$$V_i(x) = -\frac{2}{3\pi}\frac{hs_i}{\left[1+\left(\frac{x-x_i}{hs_i}\right)^2\right]^2} + x_i\left(\frac{\left(\frac{x-x_i}{hs_i}\right)^3}{\pi\left[1+\left(\frac{x-x_i}{hs_i}\right)^2\right]^2} + \frac{5\frac{x-x_i}{hs_i}}{3\pi\left[1+\left(\frac{x-x_i}{hs_i}\right)^2\right]^2} + \frac{1}{\pi}\operatorname{arctg}\left(\frac{x-x_i}{hs_i}\right) + \frac{1}{2}\right) \tag{55}$$

$$W_i(x) = -\frac{4hs_i x_i}{3\pi\left[1+\left(\frac{x-x_i}{hs_i}\right)^2\right]^2} + x_i^2\left(\frac{\left(\frac{x-x_i}{hs_i}\right)^3}{\pi\left[1+\left(\frac{x-x_i}{hs_i}\right)^2\right]^2} + \frac{5\frac{x-x_i}{hs_i}}{3\pi\left[1+\left(\frac{x-x_i}{hs_i}\right)^2\right]^2} + \frac{1}{\pi}\operatorname{arctg}\left(\frac{x-x_i}{hs_i}\right) + \frac{1}{2}\right) \tag{56}$$

$$+ h^2 s_i^2\left(\frac{\left(\frac{x-x_i}{hs_i}\right)^3}{3\pi\left[1+\left(\frac{x-x_i}{hs_i}\right)^2\right]^2} - \frac{\frac{x-x_i}{hs_i}}{3\pi\left[1+\left(\frac{x-x_i}{hs_i}\right)^2\right]^2} + \frac{1}{3\pi}\operatorname{arctg}\left(\frac{x-x_i}{hs_i}\right) + \frac{1}{6}\right) .$$

The constants used within the plug-in method are:

$$P(K) = \frac{1}{3} \tag{57}$$

$$R(K) = \frac{7}{4\pi} \quad ; \tag{58}$$

dependencies (31)-(35) remain unchanged.

The above investigations can be similarly transposed to a higher order of asymmetrical polynomial loss function (1), although on account of their extreme nature, they seem to be useful mainly for atypical applicational tasks.

4. Numerical Simulations Results

The correctness of the algorithm designed here has been checked in detail using a numerical simulation. The results are shown below for five values of the ratio $a/b = 1/10$, $1/3$, 1, 3, 10. Investigations were carried out for m = 10, 20, 50, 100, 200, 500, 1000. In every case 1000 samples were obtained, and the tables below display mean and standard deviation values of results calculated on the basis of these samples, described using the standard notation „mean value ± standard deviation".

First, it was assumed that the uncertainty of the estimated parameter has standard normal distribution:

$$N(0,1) \quad ; \tag{59}$$

in the hereafter-used natural notation $N(\mu,\sigma)$ denotes normal distribution with the expected value $\mu \in \mathrm{R}$ and the standard deviation $\sigma > 0$. The above classical example of distribution (59) is taken at the beginning only for its simplicity and ease of interpretation – in such a simple case use of the complex nonparametric estimation method is in practice redundant.

In Tables 1 and 2 results are compared for twinned quadratic and cubic cases, respectively. The theoretical value of the estimator is shown here below the values of the ratio a/b.

Thus, if $a/b = 1$, then – as mentioned – the estimator investigated in this paper reduces to the expected value, now amounting to zero for distribution (59). The condition $a/b = 1/3$ means that losses caused by overestimation are three times greater then those arising from underestimation – the estimator should therefore take a value less then the expectation value. Notably, it is transferred to –0,436 and –0,344 for quadratic and cubic cases, respectively. This effect is intensified when $a/b = 1/10$ – the estimators' values are then –0,901 and –0,716, respectively. Conversely, in the case $a/b = 3$ the losses relating to overestimation are less than those resulting from underestimation, and so the estimator should exceed the expectation value – in fact it equals 0,436 and 0,344 for quadratic and cubic cases, respectively (due to the symmetry of the considered distribution (59) these values are opposite with respect to those obtained for $a/b = 1/3$). When $a/b = 10$ this effect is again intensified – the values of the investigated estimator are 0,901 and 0,716, respectively.

Table 1. Results for distribution (59) for the quadratic case (k = 2)

m	$\frac{a}{b}=\frac{1}{10}$	$\frac{a}{b}=\frac{1}{3}$	$\frac{a}{b}=1$	$\frac{a}{b}=3$	$\frac{a}{b}=10$
	–0,9015	–0,4363	0,0000	0,4363	0,9015
10	–0,8978 ± 0,3986	–0,4337 ± 0,3326	0,0012 ± 0,3123	0,4376 ± 0,3299	0,9010 ± 0,3956
20	–0,9223 ± 0,2802	–0,4415 ± 0,2364	0,0027 ± 0,2240	0,4474 ± 0,2380	0,9282 ± 0,2871
50	–0,9335 ± 0,1743	–0,4485 ± 0,1466	–0,0024 ± 0,1402	0,4441 ± 0,1498	0,9288 ± 0,1787
100	–0,9314 ± 0,1217	–0,4463 ± 0,1019	0,0001 ± 0,0966	0,4468 ± 0,1029	0,9315 ± 0,1239
200	–0,9257 ± 0,0885	–0,4441 ± 0,0754	0,0007 ± 0,0716	0,4459 ± 0,0753	0,9282 ± 0,0892
500	–0,9207 ± 0,0558	–0,4427 ± 0,0466	0,0003 ± 0,0439	0,4435 ± 0,0469	0,9221 ± 0,0571
1000	–0,9201 ± 0,0400	–0,4429 ± 0,0330	0,0001 ± 0,0310	0,4433 ± 0,0329	0,9208 ± 0,0399

Table 2. Results for distribution (59) for the cubic case (k = 3)

m	$\frac{a}{b}=\frac{1}{10}$	$\frac{a}{b}=\frac{1}{3}$	$\frac{a}{b}=1$	$\frac{a}{b}=3$	$\frac{a}{b}=10$
	–0,7158	–0,3436	0,0000	0,3436	0,7158
10	–0,6812 ± 0,3926	–0,3293 ± 0,3485	0,0012 ± 0,3340	0,3310 ± 0,3469	0,6802 ± 0,3893
20	–0,7267 ± 0,2882	–0,3473 ± 0,2528	0,0029 ± 0,2434	0,3526 ± 0,2562	0,7305 ± 0,2949
50	–0,7550 ± 0,1860	–0,3614 ± 0,1591	–0,0023 ± 0,1517	0,3566 ± 0,1591	0,7495 ± 0,1846
100	–0,7609 ± 0,1303	–0,3623 ± 0,1093	–0,0004 ± 0,1034	0,3612 ± 0,1093	0,7587 ± 0,1302
200	–0,7574 ± 0,0944	–0,3597 ± 0,0801	0,0013 ± 0,0764	0,3623 ± 0,0808	0,7601 ± 0,0964
500	–0,7540 ± 0,0596	–0,3584 ± 0,0499	0,0010 ± 0,0476	0,3604 ± 0,0510	0,7567 ± 0,0623
1000	–0,7543 ± 0,0443	–0,3588 ± 0,0365	0,0004 ± 0,0344	0,3597 ± 0,0365	0,7554 ± 0,0442

Comparing further results for the quadratic and cubic cases one can note that in the former the values of the estimator are closer to zero. This effect is intuitively justified, as for large arguments, the values of the cubic function are greater than the quadratic, and consequently, the tendency to eliminate extreme results appears. It is also worth noticing that estimator values in the linear case for $a/b = 1/10$, 1/3, 1, 3, 10 equal –1,335, –0,675, 0, 0,675, 1,335, respectively, which is additionally confirmed in the above interpretation.

Similar conclusions can be drawn for the next distribution under research:

$$0{,}25\ N(-5,2) + 0{,}5\ N(0,1) + 0{,}25\ N(5,2) \quad . \qquad (60)$$

The obtained results are shown in Table 3. This distribution is trimodal. Its expectation value equals 0, and the standard deviation is $\sqrt{15} \cong 3{,}9$, so almost four times greater than that for distribution (59). On consideration of this fact one can infer that the results are comparative to those previously obtained for distribution (59), presented in Table 1. It should be concluded that the accuracy of estimation does not generally depend on the number of the modal values of the investigated distribution. Despite a significant change in the type of distribution examined, the procedure for calculating the estimator did not change in any way. This is due to the application of the nonparametric statistical kernel estimators methodology, the use of which is in practice independent of the distribution under research.

In every case represented by specific columns in Tables 1-3, together with the increase in the random sample size m, the average error of the estimation and its standard deviation decrease to zero. From an applicational point of view these are fundamental properties required of estimators used in practice. Above all this trait means that, as the size of samples increases, the obtained estimator values tend to the theoretical, while their “dispersion” decreases. This allows any assumed precision to be acquired, albeit after ensuring the proper sample size. This in practice implies a necessity to reach a compromise between these two quantities. A satisfactory degree of precision was obtained when the size of the sample was between 10 and 200, i.e. for $m \in [10, 200]$; in particular, the large values became necessary when the difference between parameters a and b increased.

Below are presented the results obtained for the optimal control tasks considered in Sections 3.1. and 3.2 as the motivations. In both cases, the uncertainty of the parameter in question was assumed to be of uniform distribution on the interval $[0{,}5\ ;\ 1{,}5]$.

Table 3. Results for distribution (60) for the quadratic case (k = 2)

m	$\frac{a}{b}=\frac{1}{10}$	$\frac{a}{b}=\frac{1}{3}$	$\frac{a}{b}=1$	$\frac{a}{b}=3$	$\frac{a}{b}=10$
	–3,6717	–1,6928	0,0000	1,6928	3,6717
10	–3,4493 ± 1,3887	–1,6209 ± 1,1688	0,0258 ± 1,1098	1,6882 ± 1,2891	3,5199 ± 1,5712
20	–3,5702 ± 0,9856	–1,6578 ± 0,8518	0,0141 ± 0,7900	1,6940 ± 0,9338	3,6146 ± 1,1102
50	–3,6540 ± 0,6066	–1,6832 ± 0,5639	0,0017 ± 0,5041	1,6906 ± 0,5971	3,6595 ± 0,6547
100	–3,6941 ± 0,4215	–1,7104 ± 0,3922	–0,0135 ± 0,3444	1,6829 ± 0,4220	3,6717 ± 0,4642
200	–3,6796 ± 0,2850	–1,6870 ± 0,2765	0,0095 ± 0,2483	1,7091 ± 0,3052	3,6964 ± 0,3233
500	–3,6902 ± 0,1850	–1,7031 ± 0,1821	–0,0057 ± 0,1584	1,6914 ± 0,1914	3,6796 ± 0,2004
1000	–3,6894 ± 0,1285	–1,7003 ± 0,1253	–0,0007 ± 0,1086	1,6998 ± 0,1319	3,6897 ± 0,1378

In the case of the time-optimal control problem considered in Section 3.1, approximating the loss function from Figure 1 by an asymmetrical linear function with $a/b = 15{,}5$, the feedback controller based on the procedure proposed here had a significant – even to about 40% – advantage over the classical controller based on the value of the mean.

For the optimal control task with the quadratic performance index from Section 3.2, the loss function shown in Figure 2 was assumed in the asymmetrical quadratic form with $a/b = 5{,}2$. For those realizations for which the system seemed to be stable, the controller based on the procedure proposed here also had a significant – even to about 50% – advantage with respect to the one using the mean value. Moreover, there was a greater distance from the instability area. In the case of the feedback controller obtained with the estimator worked out here, the system became unstable for the value of the object parameter $\Lambda \cong 1{,}7$, although in the case of the system designed using the mean value, this had already happened for $\Lambda \cong 1{,}5$.

5. Conclusions

This article has presented the method of estimating the values of model parameters, dedicated to those cases where the dependence of losses implied by estimation error can be approximated by asymmetrical and polynomial function. Asymmetry here represents the different influences of under- and overestimation of the parameter's estimator on the value of these losses, whereas the degree of the polynomial signifies how acceptable large errors are. The method worked out here is universal in character and can be applied in many areas of science and practice, also outside engineering.

To find the distribution of the uncertainty measure of an estimated parameter, statistical kernel estimators were used, which made the investigated procedure independent of distribution type. The solution was based on the Bayes decision rule, which allows a minimum – generally understood – average losses value to be obtained. As a result the complete algorithm was worked out, enabling the value of the estimator to be calculated on the basis of the measurements of the examined quantity, and the definition of degree – often natural in practical applications – of the polynomial and the ratio of the values of a losses function's coefficients. The herein-presented procedure is complete, and its practical implementation requires neither detailed knowledge of the theoretical aspects nor laborious research.

Although the uncertainty of the examined parameter was considered in the most common probabilistic approach, the worked out method can also be used for other types of uncertainty, for example that based on fuzzy logic [7]. In this case one is able to calculate the value of the optimal defuzzyficator or the preference function [5] – the procedure proposed here allows the Bayes preference function for fuzzy numbers to be obtained.

The propriety of the presented algorithms has been verified numerically. Thus, depending on the relation between the coefficients of the loss function, the estimator's value properly changed to one of smaller losses, while the size of this change was defined by the polynomial degree. As the random sample size increased, the average error of the estimation and its standard deviation tended to zero.

One may construe that the benefits arising from application of the method proposed in this paper are greater the more complex the control system is, and over- and under-estimation of model's parameters have a more differing influence on performance index, i.e. when asymmetry of the loss function is more distinct.

References

[1] **M. Athans, P.L. Falb.** Optimal Control. *McGraw-Hill, New York*, 1966.

[2] **J.O. Berger.** Statistical Decision Theory. *Springer-Verlag, New York*, 1980.

[3] **A. van den Bos.** Parameter Estimation for Scientists and Engineers. *Wiley, Hoboken*, 2007.

[4] **J. Caldwell, Y.M. Ram.** Mathematical Modelling. *Concepts and Case Studies, Springer, Berlin*, 1999.

[5] **J. Fodor, M. Roubens.** Fuzzy Preference Modelling and Multicriteria Decision Support. *Kluwer, Dordrecht*, 1994.

[6] **D. Kincaid, W. Cheney.** Numerical Analysis. *Brooks/ Cole, Pacific Grove*, 2002.

[7] **J. Kacprzyk.** Multistage Fuzzy Control: A Model-Based Approach to Control and Decision-Making. *Wiley, Chichester*, 1997.

[8] **P. Kulczycki.** Estymatory jadrowe w analizie systemowej. *WNT, Warsaw*, 2005.

[9] **P. Kulczycki.** An Algorithm for Bayes Parameter Identification. *Journal of Dynamic Systems, Measurement, and Control, Vol*. 123, 2001, 611-614.

[10] **P. Kulczycki.** Estymatory jadrowe w zagadnieniach badan systemowych. *P. Kulczycki, O. Hryniewicz, J. Kacprzyk (eds.), Techniki informacyjne w badaniach systemowych, WNT, Warsaw*, 2007, 79-105.

[11] **P. Kulczycki.** Kernel Estimators in Industrial Applications. *B. Prasad (ed.), Soft Computing Applications in Industry, Springer-Verlag, Berlin*, 2008, 69-91.

[12] **P. Kulczycki.** Applicational Possibilities of Nonparametric Estimation of Distribution Density for Control Engineering. *Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol*. 56, 2008, 347-359.

[13] **P. Kulczycki, M. Charytanowicz.** Asymmetrical Conditional Bayes Parameter Identification for Control Engineering. *Cybernetics and Systems, Vol*. 39, 2008, 229-243.

[14] **P. Kulczycki, A. Mazgaj.** An algorithm for Bayes parameter identification with quadratic asymmetrical loss function. *Control and Cybernetics, Vol*. 34, 2005, 1127-1148.

[15] **P. Kulczycki, A. Mazgaj.** Bayes Parameter Identification with Polynomial Asymmetrical Loss Function. *Proceedings of the* 17*th World Congress IFAC, Seoul*, 6-11 *July* 2008, 12395-12400, CD: 12395.

[16] **P. Kulczycki, R. Wisniewski.** Fuzzy Controller for a System with Uncertain Load. *Fuzzy Sets and Systems, Vol*. 131, 2002, 185-195.

[17] **L. Ljung.** System Identification: Theory for the User. *Prentice Hall, Englewood Cliffs*, 1998.

[18] **F. Morrison.** The Art of Modeling Dynamic Systems. *Multiscience Press, New York*, 1991.

[19] **H.E. Nusse, J.A. Yorke.** Dynamics: Numerical Explorations. *Springer-Verlag, New York*, 1997.

[20] **B.W. Silverman.** Density Estimation for Statistics and Data Analysis. *Chapman and Hall, New York*, 1986.

[21] **T.S. Soderstrom, P. Stoica.** System Identification. *Prentice Hall, Englewood Cliffs*, 1994.

[22] **M.P. Wand, M.C. Jones.** Kernel Smoothing. *Chapman and Hall, London*, 1995.

Received July 2008.