Control and Cybernetics

vol. **34** (2005) No. 4

An algorithm for Bayes parameter identification with quadratic asymmetrical loss function

by

Piotr Kulczycki[1] **and Aleksander Mazgaj**[2]

[1]Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warszawa, Poland
e-mail: kulczycki@ibspan.waw.pl

[2] Department of Automatic Control
Faculty of Electrical and Computer Engineering
Cracow University of Technology
Warszawska 24, 31-155 Cracow, Poland
e-mail: amazgaj@pk.edu.pl

Abstract: The paper deals with the estimation problem of model parameter values, in tasks where overestimation implies results other than underestimation, and where losses arising from this can be described by a quadratic function with different coefficients characterizing positive and negative errors. In the approach presented, the Bayes decision rule was used, allowing for minimizing potential losses. Calculation algorithms were based on the theory of statistical kernel estimators, which frees the method from distribution type. The result constitutes a complete numerical procedure enabling effective calculation of the value of an identified parameter or – in the multidimensional case – the vector of parameters. The method is aimed at both of the main contemporary approaches to uncertainty modeling: probabilistic and fuzzy logic. It is universal in nature and can be applied in a wide range of tasks of engineering, economy, sociology, biomedicine and other related fields.

Keywords: parameter identification, Bayes decision, quadratic asymmetrical loss function, kernel estimators, optimal control.

1. Introduction

A basic role in many problems of modern science and practical applications is played by the proper identification of parameters used in a model describing reality under consideration, e.g. an object or process. Due to obvious metrological reasons, a precise – in the strictest sense of the word – measure of their values is never possible. In real applications, the underestimation of a parameter value

very often implies different results from overestimation, both in quality and quantity. A task where losses resulting from this can be described by a linear function with different coefficients characterizing positive and negative errors was analysed in Kulczycki (2000, 2001), Kulczycki and Wisniewski (2002). In this paper, the problem has been transposed onto the more complex case of an quadratic asymmetrical function.

As an example to illustrate the reason for the above task, an optimal control problem will be considered. Such systems have shown themselves in practice to be very sensitive to the inaccuracy of modeling, which was – in fact – the main limit of their applications. However, the performance index defined here primarily for the purposes of control, can also refer to quality of identification allowing for creation of an optimal procedure for estimation of model parameter values, notably lowering this sensitivity. Thus, consider the following task of optimal control for a quadratic performance index with infinite end time, unit matrix/parameter for the integrand function of the performance index (Athans and Falb, 1966; Section 9.5). The object is the dynamic system

$$\dot{x}(t) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \lambda \end{bmatrix} u(t) \,, \tag{1}$$

where $\lambda \in \mathbb{R}\backslash\{0\}$. Moreover, let $\Lambda \in \mathbb{R}\backslash\{0\}$ represent an estimator of the parameter λ. An optimal feedback controller is defined on the basis of the value Λ, not necessarily equal to the value of the parameter λ existing in the object. The values of the performance index obtained for the particular Λ, are shown in Fig. 1. One can see that the resulting graph can be described with

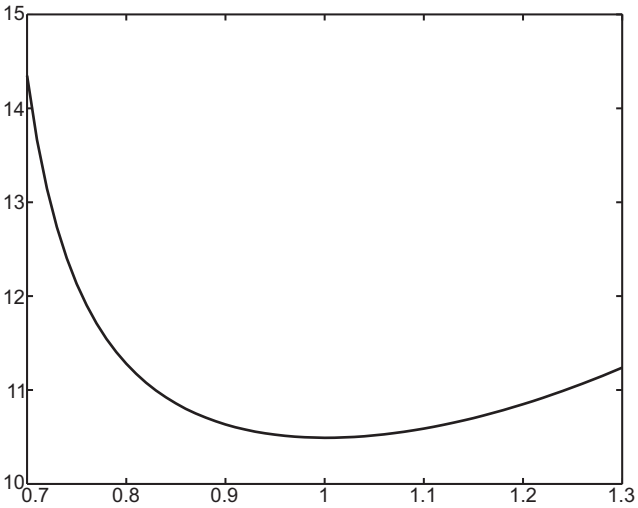

Figure 1. Performance index value as a function of the parameter Λ; ($\lambda = 1$)).

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 different effects on the performance index value. Similar outcomes can also be obtained for time-optimal control, therefore in the second main type of classical optimal control. Over- and underestimation of object parameters produce here original phenomena – sliding trajectories and over-regulations, with a significantly different influence on the performance index.

Similar conditioning can also be shown for many problems outside the area of optimal 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 irrational dread of change. Therefore, if one describes the psychological preferences of an ordinary person applying mathematical means similar to the performance index used in optimal control, then the shape of those preferences takes the quadratic and asymmetrical form displayed in Fig. 1. Here an inverse relationship to losses and gains is represented by asymmetry, and fear of extremes by the quadratic form.

The uncertainty of the examined parameters will be considered in this paper for a probabilistic approach. The methodology of statistical kernel estimators is applied to estimate the distribution of probability measure, which makes the result independent of arbitrary assumptions concerning the type of this distribution. An algorithm based on the Bayes decision rule is proposed, which allows for obtaining minimal expectation value for potential losses. The basic form of the procedure investigated here can be easily generalized for different aspects common in applications – as an example a multidimensional case, when the vector of parameters is submitted to identification, will be considered in detail. Furthermore, the investigated method can be used for other uncertainty approaches apart from that of probability, e.g. fuzzy logic.

It must be stressed that the main goal of this paper is the presentation of an algorithm in its complete form, ensuring – in particular – that its practical implementation does not demand of the user detailed knowledge of the theoretical aspects or laborious research and calculations.

This paper is based on dissertation (Mazgaj, 2005), where additional details of the problems concerned may be found.

2. Mathematical preliminaries

2.1. Elements of decision theory

The main aim of decision theory (Berger, 1980) is the selection of a concrete decision, based on a representation of measure characterizing the imprecision of states of nature. Thus, assume the space $\mathbb{R}^n$ as the set of all states of nature,

and the density of distribution of a measure on $\mathbb{R}^n$ for such imprecision (e.g. the density of distribution of a probability measure for a probabilistic approach or a membership function for a fuzzy logic approach) denoted by M. Let there be given also the nonempty set of possible decisions D, as well as the loss function $l : D \times \mathbb{R}^n \to \mathbb{R}$ while its values $l(d, z)$ can be interpreted as losses occurring in a hypothetical case when the state of nature (deterministic or crisp in probabilistic or fuzzy approaches, respectively) is z and the decision d is taken. If for every $d \in D$ the integral $\int_{\mathbb{R}^n} l(d, z)M(z)\mathrm{d}z$ exists, then the Bayes loss function $l_B : D \to \mathbb{R} \cup \{\pm\infty\}$ can be defined as

$$l_B(d) = \int_{\mathbb{R}^n} l(d, z)M(z)\mathrm{d}z \,. \tag{2}$$

Every element $d_B \in D$ such that

$$l_B(d_B) = \min_{d \in D} l_B(d) \tag{3}$$

is called a Bayes decision, and the procedure of its discovery – a Bayes decision rule. The Bayes decision is chosen in such a way, therefore, as to minimize the mean value (with respect to the measure represented by the function M) of losses following the decision d. In tasks investigated further in this paper, the Bayes decision will constitute a Bayes estimator or a Bayes defuzzyficator for the probabilistic or fuzzy approaches, respectively.

2.2. Statistical kernel estimators

Let the n-dimensional random variable X, with a distribution having density f, be given. Its kernel estimator $\hat{f} : \mathbb{R}^n \to [0, \infty)$ is calculated on the basis of the m-element 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

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right) , \tag{4}$$

where the function $K : \mathbb{R}^n \to [0, \infty)$, which is measurable, symmetrical relative to zero, and has a weak global maximum at this point, fulfils the condition $\int_{\mathbb{R}^n} K(x)\mathrm{d}x = 1$, and is called a kernel, whereas the positive coefficient h is known as a smoothing parameter. In the particular procedures concerning kernel estimators, additional requirements regarding the functions f and K are assumed, although these are not restrictive from an applicational point of view. It should be stated that the kernel estimators allow for identification of density of practically any distribution, without an assumption concerning its type.

Fixing values introduced in definition (4), i.e. choosing the form of the kernel K and calculating the value of the smoothing parameter h, is most often carried out using the mean squares criterion.

Thus, from the statistical point of view, the form of the kernel seems not to have an essential meaning, thanks to which it becomes possible for the choice of the function K to be arbitrary, taking into account above all the 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.

In the one-dimensional case, for the function K, the classical forms of densities of probability distributions (e.g. normal, Cauchy, triangular and others) or their linear combinations, are used. The most effective is the so-called Epanechnikov kernel

$$K_E(x) = \begin{cases} \frac{3}{4}(1-x^2) & \text{for } x \in [-1,1] \\ 0 & \text{for } x \in (-\infty,-1) \cup (1,\infty) \ . \end{cases} \tag{5}$$

In the multidimensional case, two natural generalizations of the above concept are used: radial and product kernels. However, the former is somewhat more effective, although from an applicational point of view, the difference is immaterial and the product kernel – significantly more convenient in analysis – is often favored in practical problems. The n-dimensional product kernel K can be expressed as

$$K(x) = K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \mathcal{K}(x_1) \cdot \mathcal{K}(x_2) \cdot \ldots \cdot \mathcal{K}(x_n) \ , \tag{6}$$

where $\mathcal{K}$ denotes the one-dimensional kernel. Among product kernels, the most effective is based on the one-dimensional Epanechnikov kernel, i.e. when $\mathcal{K}$ is given by formula (5).

As opposed to the form of the kernel, the value of the smoothing parameter has significant influence on the quality of the estimator obtained. In any case, convenient algorithms have been developed to secure calculation of the value of h close to optimal, on the basis of a random sample. As the product kernel will be used in this paper, a presentation of the one-dimensional case is sufficient. Thus, the most convenient algorithm is the so-called plug-in method. Its concept consists in the calculation of this parameter using an approximate method, and after r steps of 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 investigated in this paper, $r = 3$ was assumed. In this case the plug-in method consists in the application of the following formulas:

$$\tilde{\psi}_{10} = \frac{-945}{64\pi^{1/2}\hat{\sigma}^{11}} \ , \tag{7}$$

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

$$\hat{\sigma} = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m} x_i^2 - \frac{1}{m(m-1)}\left(\sum_{i=1}^{m} x_i\right)^2} \tag{8}$$

and

$$g_1 = \left(\frac{-2L^{(8)}(0)}{mP(L)\tilde{\psi}_{10}}\right)^{1/11} \tag{9}$$

$$g_2 = \left(\frac{-2L^{(6)}(0)}{mP(L)\psi_8(g_1)}\right)^{1/9} \tag{10}$$

$$g_3 = \left(\frac{-2L^4(0)}{mP(L)\psi_6(g_2)}\right)^{1/7}, \tag{11}$$

finally

$$h = \left(\frac{R(K)}{mP(K)^2\psi_4(g_3)}\right)^{1/5}, \tag{12}$$

where

$$R(K) = \int_{\mathbb{R}} K(x)^2 \,\mathrm{d}x \tag{13}$$

$$P(K) = \int_{\mathbb{R}} x^2 K(x) \,\mathrm{d}x \tag{14}$$

$$\psi_r(g) = \frac{1}{m^2 g^{r+1}} \sum_{i=1}^{m}\sum_{j=1}^{m} L^{(r)}\left(\frac{x_i - x_j}{g}\right). \tag{15}$$

The kernel K, applied in the kernel estimator (4), is used only in the last step. In all the other steps, a different kernel L, more convenient for the plug-in method, may be used.

The value of the smoothing parameter h introduced in definition (4) 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 showing better the 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 the basic formula (4);

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

$$s_i = \left(\frac{\hat{f}(x_i)}{\bar{s}}\right)^{-1/2} \quad \text{for} \quad i = 1, 2, \ldots, m \tag{16}$$

are calculated, where $\bar{s}$ is the geometric mean of the numbers $\hat{f}(x_1), \hat{f}(x_2), \ldots, \hat{f}(x_m)$, given by the logarithmic equation

$$\log(\bar{s}) = \frac{1}{m}\sum_{i=1}^{m}\log(\hat{f}(x_i)\,; \tag{17}$$

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

$$\hat{f}(x) = \frac{1}{mh^n}\sum_{i=1}^{m}\frac{1}{s_i^n}K\left(\frac{x-x_i}{hs_i}\right)\,. \tag{18}$$

Details of the above-presented methodology of statistical kernel estimators are found in Kulczycki (2005), Silverman (1986), Wand and Jones (1995).

3. Algorithm

3.1. Basic, one-dimensional case

The parameter under investigation, whose value is to be estimated, will be denoted by x. The basic one-dimensional case, i.e. where $x \in \mathbb{R}$, is presented below. In order to adhere to principles of decision theory presented in Section 2.1, the parameter x will be treated here as the value of a one-dimensional 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 study 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 the minimum of expected value of losses. According to the conditions formulated in the Introduction, the loss function is assumed in quadratic and asymmetrical form:

$$l(\hat{x}, x) = \begin{cases} a(\hat{x}-x)^2 & \text{for } \hat{x}-x \le 0 \\ b(\hat{x}-x)^2 & \text{for } \hat{x}-x \ge 0 \end{cases}\,, \tag{19}$$

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

$$l_B(\hat{x}) = a\int_{\hat{x}}^{\infty}(\hat{x}-x)^2 f(x)\mathrm{d}x + b\int_{-\infty}^{\hat{x}}(\hat{x}-x)^2 f(x)\mathrm{d}x\,, \tag{20}$$

where $f : \mathbb{R} \to [0, \infty)$ 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 fulfils its minimum for the value being a solution of the following equation with the argument $\hat{x}$:

$$(a-b) \int_{-\infty}^{\hat{x}} (\hat{x} - x) f(x) \mathrm{d}x - a \int_{-\infty}^{\infty} (\hat{x} - x) f(x) \mathrm{d}x = 0 \,. \tag{21}$$

This solution exists and is unique. Upon dividing the above equation by b, one obtains its equivalent form

$$\left(\frac{a}{b} - 1\right) \int_{-\infty}^{\hat{x}} (\hat{x} - x) f(x) \mathrm{d}x = \frac{a}{b} \int_{-\infty}^{\infty} (\hat{x} - x) f(x) \mathrm{d}x \,. \tag{22}$$

This implies that it is not necessary to identify the parameters a and b separately, but only their ratio.

Solution of equation (21) for a general case is not an easy task. However, if estimation of the density f is obtained 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, and also fulfilling the condition

$$\int_{-\infty}^{\infty} xK(x) \mathrm{d}x < \infty \tag{23}$$

be used. For any fixed $i = 1, 2, \ldots, m$ the functions $U_i : \mathbb{R} \to \mathbb{R}$ and $V_i : \mathbb{R} \to \mathbb{R}$ given by

$$U_i(x) = \frac{1}{hs_i} \int_{-\infty}^{x} K\left(\frac{y - x_i}{hs_i}\right) \mathrm{d}y \tag{24}$$

$$V_i(x) = \frac{1}{hs_i} \int_{-\infty}^{x} yK\left(\frac{y - x_i}{hs_i}\right) \mathrm{d}y \tag{25}$$

can then be defined. The choice of form of the kernel K should be made so that the functions $I : \mathbb{R} \to \mathbb{R}$ and $J : \mathbb{R} \to \mathbb{R}$ such that

$$I(x) = \int_{-\infty}^{x} K(y) \mathrm{d}y \tag{26}$$

$$J(x) = \int_{-\infty}^{x} yK(y) \mathrm{d}y \tag{27}$$

can be expressed by relatively simple analytical formulas, which results in a similar property regarding the functions U_i and V_i.

If an expected value is estimated by the arithmetical mean value of a sample, then criterion (21) can be described equivalently as

$$\sum_{i=1}^{m}[(a-b)(\hat{x}U_i(\hat{x}) - V_i(\hat{x}) + ax_i] - a\hat{x}m = 0\ . \tag{28}$$

Denoting the left side of the above formula by $L(\hat{x})$, one can – thanks to the equality $V_i'(\hat{x}) = \hat{x}U_i'(\hat{x})$ directly resulting from dependencies (24)-(25) – express the value of its derivative as

$$L'(\hat{x}) = \sum_{i=1}^{m}[(a-b)U_i(\hat{x})] - am\ . \tag{29}$$

In this situation, the solution of equation (21) can be calculated numerically on the basis of Newton's algorithm as the limit of the sequence $\{\hat{x}_j\}_{j=0}^{\infty}$ defined by

$$\hat{x}_0 = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{30}$$

$$\hat{x}_{j+1} = \hat{x}_j - \frac{L(\hat{x}_j)}{L'(\hat{x}_j)} \quad \text{for} \quad j = 0, 1, \dots\ , \tag{31}$$

with the functions L and L' being given by formulas (28)-(29), whereas the stop criterion takes on the form

$$|\hat{x}_j - \hat{x}_{j-1}| \leq 0.01\hat{\sigma}\ , \tag{32}$$

where $\hat{\sigma}$ denotes the estimator of the standard deviation (8). Depending on the conditions of application under investigation, the above form of Newton's algorithm can be modified and enhanced by additional aspects, available in the literature on the subject (e.g. Dalquist and Bjork, 1983; Stoer and Bulirsch, 1983).

As mentioned, an important positive feature consists in the possibility to choose the kernel form with regard to the requirements of the practical task investigated. In the above considerations, the following four requirements for this subject have been formulated:

(A) continuity and positivity of the function K;

(B) the finite first moment of the above function, required by condition (23);

(C) the function I defined by dependence (26) should be expressed by a relatively simple analytical formula;

(D) similarly with respect to the function J, given by formula (27).

The most effective Epanechnikov kernel (5) does not fulfil the first of the above conditions. Because of this, the Cauchy kernel

$$K_C(x) = \frac{2}{\pi}\frac{1}{(1+x^2)^2} \tag{33}$$

can be proposed, or more general – a linear combination of the Epanechnikov (5) and Cauchy (33) kernels:

$$K(x) = \alpha K_E(x) + (1-\alpha)K_C(x) \,, \tag{34}$$

with

$$\alpha \in [0,1) \,. \tag{35}$$

If the constant α takes on a value close to 1, then kernel (34) has effectiveness near to optimal, however, due to the Cauchy kernel having positive values, the combination (34) will also have this feature. In practice it is recommended that

$$\alpha = 0.9 \,, \tag{36}$$

which ensures effectiveness close to the one for the optimal Epanechnikov kernel, although it is not so close to 1, to avoid computational difficulties connected with approximations of numerical methods. Note, however, that in the case of $\alpha = 0$, only the Cauchy kernel is used, leading to a drop of about half in calculation time.

For any fixed $i = 1, 2, \ldots, m$, the function U_i defined by formula (24) is therefore a linear combination of appropriate functions for the Epanechnikov and Cauchy kernels:

$$U_i(x) = \alpha U_{i,E}(x) + (1-\alpha)U_{i,C}(x) \,, \tag{37}$$

while

$$U_{i,E}(x) = \begin{cases} 0 & \text{for } x < x_i - hs_i \\ \dfrac{-x^3 + 3x_i x^2 + 3(h^2 s_i^2 - x_i^2)x + x_i^3 + 2h^3 s_i^3 - 3h^2 s_i^2 x_i}{4(hs_i)^3} & \\ & \text{for } x_i - hs_i \le x \le x_i + hs_i \\ 1 & \text{for } x > x_i + hs_i \end{cases} \tag{38}$$

$$U_{i,C}(x) = \frac{1}{\pi}\operatorname{arctg}\left(\frac{x-x_i}{hs_i}\right) + \frac{\dfrac{x-x_i}{hs_i}}{\pi\left[1 + \left(\dfrac{x-x_i}{hs_i}\right)^2\right]} + \frac{1}{2} \,. \tag{39}$$

Similarly one can denote for the function V_i given by (25):

$$V_i(x) = \alpha V_{i,E}(x) + (1-\alpha)V_{i,C}(x) \,, \tag{40}$$

where

$$V_{i,E}(x) = \begin{cases} 0 & \text{for } x < x_i - hs_i \\ \dfrac{-3x^4 + 8x_i x^3 + 6(h^2 s_i^2 - x_i^2)x^2 + x_i^4 - 6h^2 s_i^2 x_i^2 + 8h^3 s_i^3 x_i - 3h^4 s_i^4}{16(hs_i)^3} & \text{for } x_i - hs_i \le x \le x_i + hs_i \\ x_i & \text{for } x > x_i + hs_i \end{cases} \tag{41}$$

$$V_{i,C}(x) = x_i \left(\frac{1}{\pi} \operatorname{arctg}\left(\frac{x - x_i}{hs_i}\right) + \frac{\dfrac{x - x_i}{hs_i}}{\pi\left[1 + \left(\dfrac{x - x_i}{hs_i}\right)^2\right]} + \frac{1}{2} \right) - \frac{hs_i}{\pi\left[1 + \left(\dfrac{x - x_i}{hs_i}\right)^2\right]} . \tag{42}$$

When using linear combination (34) with $\alpha \neq 0$, the smoothing parameter h should be calculated separately for each component kernel (5) and (33), which enables to obtain proper form separately for each of them and, consequently, a suitable one – from the practical point of view – for their linear combination. In Section 2.2 the third order plug-in method was proposed for the above purpose. For Epanechnikov (5) and Cauchy (33) kernels, the quantities occurring in formulas (9)-(12) are, respectively, given as

$$P(K_E) = \frac{3}{15} \tag{43}$$

$$R(K_E) = \frac{9}{15} \tag{44}$$

and

$$P(K_C) = 1 \tag{45}$$

$$R(K_C) = \frac{5}{4\pi} . \tag{46}$$

As the kernel L, introduced in the plug-in method, Cauchy kernel can be proposed. The quantities occurring in formulas (9)-(11) are expressed, respectively, as:

$$P(L) = P(K_C) = 1 \tag{47}$$

$$L^{(4)}(x) = K_C^{(4)}(x) = \frac{48}{\pi} \frac{35x^4 - 42x^2 + 3}{(1 + x^2)^6} \tag{48}$$

$$L^{(6)}(x) = K_C^{(6)}(x) = \frac{5760}{\pi} \frac{21x^6 - 63x^4 + 27x^2 - 1}{(1+x^2)^8} \tag{49}$$

$$L^{(8)}(x) = K_C^{(8)}(x) = \frac{80640}{\pi} \frac{165x^8 - 924x^6 + 990x^4 - 220x^2 + 5}{(1+x^2)^{10}} \,. \tag{50}$$

The modifying parameters $s_1, s_2, \ldots, s_m$ should be directly calculated based on the algorithm presented in Section 2.2.

Finally, in this way, the complete algorithm has been investigated, allowing for calculation of the value of the Bayes estimator for the quadratic and asymmetrical loss function (19), in the basic one-dimensional case. All quantities necessary for its numerical implementation have been directly given, as a result of which, after the introduction of the random sample value $x_1, x_2, \ldots, x_m$ and the assumed value of the quotient $\frac{a}{b}$, one obtains the desired estimator value. Thus, assuming the kernel K in the form (34), with the application of (5), (33), and possibly (36), on the basis of the random sample value $x_1, x_2, \ldots, x_m$ one can calculate – using the procedures described by formulas (7)-(12) and (15) with derivations (43)-(50) – the values of the smoothing parameters h separately for Epanechnikov and Cauchy kernels, and then according to the procedure presented in Section 2.2 – the modifying parameters $s_1, s_2, \ldots, s_m$. For the assumed quotient $\frac{a}{b}$, this allows for defining the functions U_i as well as V_i based on dependencies (37)-(42), and also – thanks to equalities (28)-(29) – L as well as L'. The above completes the quantities required to use Newton's algorithm (30)-(32), and, consequently, to obtain the desired estimator value.

3.2. Multidimensional case

The proposed method can be easily generalized for different cases commonly appearing in applications: multidimensional (the vector of parameters), conditional (a dependence on external factors) and polynomial (a loss function of a degree higher than quadratic); for a commentary see Kulczycki and Mazgaj (2003). As a particularly important example for practice, a multidimensional case, when a number of parameters (i.e. their vector) is submitted for estimation, will be presented in details below. For the sake of clarity of presentation, a two-dimensional case will be worked out here. The idea itself may be transposed for larger dimensions, although at a natural – in such a situation – cost of increasing complexity.

In order to illustrate the conditioning of such a problem, the example from optimal control examined in the Introduction will be continued now. Consider the dynamical system resulting from inclusion of the inertia of an actuator in model (1), therefore replacement of the control u with the additional coordinate x_3, and adding the third differential equation in the form

$$\dot{x}_3(t) = -\tau x_3(t) + u(t) \,, \tag{51}$$

with the inertia constant $\tau > 0$. Joining thus formulas (1) and (51), one obtains the following differential equation:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & \lambda \\ 0 & 0 & -\tau \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u(t) \,. \tag{52}$$

Submitted for estimation, therefore, are the parameters derived from the different subsystems: λ directly from the object and τ characterizing the actuator, and so are independent. The feedback controller is defined based on the values of the estimators Λ and T, which differ from the values of the parameters λ and τ existing in the object. The results of errors of estimation of both of these parameters are, however, correlated – in effect they can cumulate or be partially eliminated. A spatial graph for the performance index for particular Λ and T, i.e. the two-dimensional counterpart of one shown in Fig. 1, can be approximated with great accuracy using a two-dimensional, quadratic and asymmetrical loss function, of the form which will be defined below.

Let therefore the estimated parameters $x_1, x_2 \in \mathbb{R} \times \mathbb{R}$ be treated as the two-dimensional vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and their estimators: $\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$. Similarly to the one-dimensional case, the vector of parameters will be taken to be the value of a two-dimensional random variable. Ultimately, the two-dimensional quadratic and asymmetrical loss function is given by the following dependence:

$$l\left(\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{cases} a_l(\hat{x}_1 - x_1)^2 + a_{ld}(\hat{x}_1 - x_1)(\hat{x}_2 - x_2) + a_d(\hat{x}_2 - x_2)^2 \\ \qquad \text{if } \hat{x}_1 - x_1 \leq 0 \text{ and } \hat{x}_2 - x_2 \leq 0 \\ a_p(x_1 - x_1)^2 + a_{pd}(\hat{x}_1 - x_1)(\hat{x}_2 - x_2) + a_d(\hat{x}_2 - x_2)^2 \\ \qquad \text{if } \hat{x}_1 - x_1 \geq \text{ and } \hat{x}_2 - x_2 \leq 0 \\ a_l(\hat{x}_1 - x_1)^2 + a_{lg}(\hat{x}_1 - x_1)(\hat{x}_2 - x_2) + a_g(\hat{x}_2 - x_2)^2 \\ \qquad \text{if } \hat{x}_1 - x_1 \leq 0 \text{ and } x_2 - x_2 \geq 0 \\ a_p(\hat{x}_1 - x_1)^2 + a_{pg}(\hat{x}_1 - x_1)(\hat{x}_2 - x_2) + a_g(\hat{x}_2 - x_2)^2 \\ \qquad \text{if } \hat{x}_1 - x_1 \geq 0 \text{ and } \hat{x}_2 - x_2 \geq 0 \end{cases} \tag{53}$$

where $a_l, a_p, a_g, a_d > 0$, $a_{ld}, a_{pg} \geq 0$ and $a_{pd}, a_{lg} \leq 0$. The coefficients a_{ld}, a_{pd}, a_{lg}, a_{pg} represent the complementary correlation of estimation errors for both parameters. It is also worth noting that in the case when the parameters a_{ld}, a_{pd}, a_{lg}, a_{pg} equal zero, the problem is reduced to two separate tasks in the basic one-dimensional form, investigated in the previous section.

Assume independence of the estimated parameters. This for example can mean that they originate from different subsystems of an object under research, which was illustrated by the example from the area of optimal control at the

beginning of this section. Then the density f representing their uncertainty may be shown as the product of the one-dimensional densities $f_1 : \mathbb{R} \to [0, \infty)$ and $f_2 : \mathbb{R} \to [0, \infty)$ corresponding to particular composites, i.e.

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) \, . \tag{54}$$

The Bayes loss function (2) assumes its minimum for the value, which is the solution for the following two equations with arguments $\hat{x}_1$ and $\hat{x}_2$:

$$\begin{aligned}
&\int_{-\infty}^{\hat{x}_1} f_1(x_1)\mathrm{d}x_1 \left[(a_{pg} - a_{pd} - a_{lg} + a_{ld}) \left(\hat{x}_2 \int_{-\infty}^{\hat{x}_2} f_2(x_2)\mathrm{d}x_2 - \int_{-\infty}^{\hat{x}_2} x_2 f_2(x_2)\mathrm{d}x_2 \right) + \right. \\
&\quad \left. + (a_{pd} - a_{ld}) \left(\hat{x}_2 - \int_{-\infty}^{\infty} x_2 f_2(x_2)\mathrm{d}x_2 \right) \right] + 2a_l \left(\hat{x}_1 - \int_{-\infty}^{\infty} x_1 f_1(x_1)\mathrm{d}x_1 \right) + \\
&\quad + 2(a_p - a_l) \left(\hat{x}_1 \int_{-\infty}^{\hat{x}_1} f_1(x_1)\mathrm{d}x_1 - \int_{-\infty}^{\hat{x}_1} x_1 f_1(x_1)\mathrm{d}x_1 \right) + \\
&\quad + a_{ld} \left(\hat{x}_2 - \int_{-\infty}^{\infty} x_2 f_2(x_2)\mathrm{d}x_2 \right) + \\
&\quad + (a_{lg} - a_{ld}) \left(\hat{x}_2 \int_{-\infty}^{\hat{x}_2} f_2(x_2)\mathrm{d}x_2 - \int_{-\infty}^{\hat{x}_2} x_2 f_2(x_2)\mathrm{d}x_2 \right) = 0
\end{aligned} \tag{55}$$

and

$$\begin{aligned}
&\int_{-\infty}^{\hat{x}_2} f_2(x_2)\mathrm{d}x_2 \left[(a_{pg} - a_{pd} - a_{lg} + a_{ld}) \left(\hat{x}_1 \int_{-\infty}^{\hat{x}_1} f_1(x_1)\mathrm{d}x_1 - \int_{-\infty}^{\hat{x}_1} x_1 f_1(x_1)\mathrm{d}x_1 \right) + \right. \\
&\quad \left. + (a_{lg} - a_{ld}) \left(\hat{x}_1 - \int_{-\infty}^{\infty} x_1 f_1(x_1)\mathrm{d}x_1 \right) \right] + 2a_d \left(\hat{x}_2 - \int_{-\infty}^{\infty} x_2 f_2(x_2)\mathrm{d}x_2 \right) + \\
&\quad + 2(a_g - a_d) \left(\hat{x}_2 \int_{-\infty}^{\hat{x}_2} f_2(x_2)\mathrm{d}x_2 - \int_{-\infty}^{\hat{x}_2} x_2 f_2(x_2)\mathrm{d}x_2 \right) + \\
&\quad + a_{ld} \left(\hat{x}_1 - \int_{-\infty}^{\infty} x_1 f_1(x_1)\mathrm{d}x_1 \right) + \\
&\quad + (a_{pd} - a_{ld}) \left(\hat{x}_1 \int_{-\infty}^{\hat{x}_1} f_1(x_1)\mathrm{d}x_1 - \int_{-\infty}^{\hat{x}_1} x_1 f_1(x_1)\mathrm{d}x_1 \right) = 0 \, .
\end{aligned} \tag{56}$$

This solution exists and it is unique. The above equations (55)-(56) constitute a two-dimensional equivalent to criterion (21).

The densities f_1 and f_2 were estimated by the statistical kernel estimators $\hat{f}_1$ and $\hat{f}_2$, respectively, using the kernel K assumed to fulfill conditions (A)-(D) formulated in Section 3.1. Note that the modifying parameters should be calculated based on the algorithm from Section 2.2 applying product kernel (6) with the above form of the function K (although denoted there as $\mathcal{K}$) – consequently, the obtained values $s_1, s_2, \ldots, s_m$ will be the same in further parts of the procedure for both coordinates.

The solution of equations (55)-(56) may be calculated through efficient numerical methods. To this end with any fixed $i = 1, 2, \ldots, m$, one can define the functions $U_{i,1} : \mathbb{R} \to \mathbb{R}$, $U_{i,2} : \mathbb{R} \to \mathbb{R}$, $V_{i,1} : \mathbb{R} \to \mathbb{R}$ and $V_{i,2} : \mathbb{R} \to \mathbb{R}$, given by formulas (24)-(25), while x, x_i, h should be replaced with quantities relating to the first and second coordinates, i.e. x_1, $x_{i,1}$, h_1 in the case of the functions $U_{i,1}$ and $V_{i,1}$, as well as x_2, $x_{i,2}$, h_2 for $U_{i,2}$ and $V_{i,2}$, respectively. After introducing the above notations, criteria (55)-(56) can be described in the following form:

$$\begin{aligned}\sum_{i=1}^{m} \Big\{ & U_{i,1}(x_1) \Big[(a_{pg} - a_{pd} - a_{lg} + a_{ld}) \left(x_2 U_{i,2}(x_2) - V_{i,2}(x_2) \right) + \\ & + (a_{pd} - a_{ld})(x_2 - x_{i,2}) \Big] + 2a_l(x_1 - x_{i,1}) + \\ & + 2(a_p - a_l) \left(x_1 U_{i,1}(x_1) - V_{i,1}(x_1) \right) + a_{ld}(x_2 - x_{i,2}) + \\ & + (a_{lg} - a_{ld}) \left(x_2 U_{i,2}(x_2) - V_{i,2}(x_2) \right) \Big\} = 0 \end{aligned} \tag{57}$$

$$\begin{aligned}\sum_{i=1}^{m} \Big\{ & U_{i,2}(x_2) \Big[(a_{pg} - a_{pd} - a_{lg} + a_{ld}) \left(x_1 U_{i,1}(x_1) - V_{i,1}(x_1) \right) + \\ & + (a_{lg} - a_{ld})(x_1 - x_{i,1}) \Big] + 2a_d(x_2 - x_{i,2}) + \\ & + 2(a_g - a_d) \left(x_2 U_{i,2}(x_2) - V_{i,2}(x_2) \right) + a_{ld}(x_1 - x_{i,1}) + \\ & + (a_{pd} - a_{ld}) \left(x_1 U_{i,1}(x_1) - V_{i,1}(x_1) \right) \Big\} = 0 \,. \end{aligned} \tag{58}$$

If one denotes the left hand sides of the above equations as $L_1(x_1, x_2)$ and $L_2(x_1, x_2)$, the values of the partial derivatives of the functions L_1 and L_2 with respect to x_1 and x_2 are given as

$$\begin{aligned}\frac{\partial L_1(x_1, x_2)}{\partial x_1} = \sum_{i=1}^{m} \Bigg\{ & \frac{1}{h_1 s_i} K\left(\frac{x_1 - x_{i,1}}{h_1 s_i} \right) \Big[(a_{pg} - a_{pd} - a_{lg} + a_{ld}) \times \\ & \left(x_2 U_{i,2}(x_2) - V_{i,2}(x_2) \right) + (a_{pd} - a_{ld})(x_2 - x_{i,2}) \Big] + \\ & + 2(a_p - a_l) U_{i,1}(x_1) + 2a_l \Bigg\} \end{aligned} \tag{59}$$

$$\frac{\partial L_2(x_1,x_2)}{\partial x_2} = \sum_{i=1}^{m} \left\{ \frac{1}{h_2 s_i} K\left(\frac{x_2 - x_{i,2}}{h_2 s_i}\right) \Big[(a_{pg} - a_{pd} - a_{lg} + a_{ld}) \times \right.$$
$$(x_1 U_{i,1}(x_1) - V_{i,1}(x_1)) + (a_{lg} - a_{ld})(x_1 - x_{i,1}) \Big] +$$
$$\left. +2(a_g - a_d) U_{i,2}(x_2) + 2a_d \right\} \quad . \tag{60}$$

Then the solution of equations (55)-(56) can be calculated through Newton's algorithm, as the limit of the two-dimensional sequence $\left\{ \begin{array}{c} \hat{x}_{j,1} \\ \hat{x}_{j,2} \end{array} \right\}_{j=0}^{\infty}$ defined by

$$\hat{x}_{0,1} = \frac{1}{m} \sum_{i=1}^{m} x_{i,1} \tag{61}$$

$$\hat{x}_{0,2} = \frac{1}{m} \sum_{i=1}^{m} x_{i,2} \tag{62}$$

$$\hat{x}_{j+1,1} = \hat{x}_{j,1} - \frac{L_1(\hat{x}_{j,1}, \hat{x}_{j,2})}{\dfrac{\partial L_1(\hat{x}_{j,1}, \hat{x}_{j,2})}{\partial x_1}} \quad \text{for} \quad j = 0, 1, \ldots \tag{63}$$

$$\hat{x}_{j+1,2} = \hat{x}_{j,2} - \frac{L_2(\hat{x}_{j,1}, \hat{x}_{j,2})}{\dfrac{\partial L_2(\hat{x}_{j,1}, \hat{x}_{j,2})}{\partial x_2}} \quad \text{for} \quad j = 0, 1, \ldots \ , \tag{64}$$

while the quantities in the above dependencies are given by formulas (57)-(60), whereas the stop condition takes the form of the conjunction of the following inequalities:

$$|\hat{x}_{j,1} - \hat{x}_{j-1,1}| \ \leqslant 0.01 \ \hat{\sigma}_1 \tag{65}$$

$$|\hat{x}_{j,2} - \hat{x}_{j-1,2}| \leq 0.01 \hat{\sigma}_2 \ , \tag{66}$$

where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ denote the estimators of standard deviations for particular coordinates, whose values can be obtained on the basis of formula (8).

For the kernel K it is recommended to once again take the linear combination of the Epanechnikov and Cauchy kernels (34). Separate calculations of the smoothing parameters are desired not only for both coordinates of the parameters' vector, but when $\alpha \neq 0$, also for the component kernels K_E and K_C. Proper formulas for the values of the functions $U_{i,1}$, $U_{i,2}$ and $V_{i,1}$, $V_{i,2}$ are given by dependencies (37)-(42), which must be used respectively for the first and second coordinates.

4. Other approaches to uncertainty modeling

The method investigated in this paper has been presented until now for a probabilistic approach, currently the most commonly used way of describing uncertainty. However, the formula itself has a universal character and can be applied

to other concepts of this description. The above task is considered in this section with fuzzy logic as an example.

Thus, the membership function of a fuzzy number is often estimated on the basis of the verbal opinions of a group of experts, expressed intuitively with an inevitable imprecision. Let therefore m values $x_1, x_2, \ldots, x_m$ belonging to the membership function domain be the subject of the experts' opinions. Each of them receives an expert's score by the mapping of the nonnegative number $w_1, w_2, \ldots, w_m$ representing the value of this score (not all equal to zero), and also positive parameters $s_1, s_2, \ldots, s_m$ characterizing imprecision in the sense that, if an expert's score is less precise, the respective parameter is bigger. Then – with regard to the kernel estimator for a probabilistic approach (18) – one can define the membership function of a fuzzy number by the following formula:

$$\mu(x) = c \sum_{i=1}^{m} \frac{w_i}{s_i} K\left(\frac{x - x_i}{h s_i}\right), \tag{67}$$

whereby the meanings of the function K and the parameter h remain unchanged with respect to the basic definition (4) for $n = 1$, and the positive constant c is chosen, so that

$$\max_{x \in \mathbb{R}} \mu(x) = 1 . \tag{68}$$

For the kernel types used in practice such a constant exists.

In many theoretical and application tasks the need arises to characterize a fuzzy number with a real number, in this case called defuzzyfier. The procedure investigated in this paper can be applied to obtain a Bayes defuzzyfier with quadratic and asymmetrical loss function. An equation which allows the calculation of such a value may be expressed in a form similar to formula (21):

$$(a - b) \int_{-\infty}^{\hat{x}} (\hat{x} - x)\mu(x)\mathrm{d}x - a \int_{-\infty}^{\infty} (\hat{x} - x)\mu(x)\mathrm{d}x = 0 \tag{69}$$

and proceeded according to Section 3.1. It is worth noting that – with regard to the form of dependence (69) – the value of the parameter c defined by condition (68) need not in fact be calculated, as it should be cancelled out anyway.

5. Experimental verification

Correct functioning of the procedure presented here has been verified using numerical simulation[1]. The first to be considered is the one-dimensional case,

[1]The calculations were carried out in ACK CYFRONET AGH on SGI 2800 (grant KBN/SGI2800/PK/019/2003) and Sun Fire 6800 (grant MNiI/Sun6800/PK/057/2004) computers.

investigated in Section 3.1. At the beginning assume that the uncertainty of an estimated parameter has the distribution

$$0.75N(0,1) + 0.25N(4,1) \tag{70}$$

and is therefore a linear combination of normal distributions with unique standard deviations, but with expectation values 0 and 4, respectively. The coefficients for this combination are not equal – the values 0.75 and 0.25 state that the first ,,share" is three times greater than the second. The distribution under investigation is thus bimodal and asymmetrical. Its expected value equals 1, while its standard deviation is 2. Table 1 shows the results obtained for five quotient values $\frac{a}{b} = \frac{1}{10}$, $\frac{1}{3}$, 1, 3, 10 and the random sample sizes m=10, 20, 50, 100, 200, 500, 1000. For every case 1000 samples were generated, and the mean value and standard deviation of the results calculated on the basis of them have been denoted using the usual "mean value ± standard deviation" notation. The theoretical values of the estimator are given below the quotient values $\frac{a}{b}$.

Table 1. Results for basic one-dimensional case with distribution (70)

m	$\frac{a}{b} = \frac{1}{10}$	$\frac{a}{b} = \frac{1}{3}$	$\frac{a}{b} = 1$	$\frac{a}{b} = 3$	$\frac{a}{b} = 10$
	-0.448	0.221	1.000	2.011	3.154
10	-0.515 ± 0.504	0.200 ± 0.544	0.997 ± 0.674	1.947 ± 0.827	3.016 ± 0.940
20	-0.511 ± 0.367	0.200 ± 0.389	1.003 ± 0.478	1.995 ± 0.587	3.130 ± 0.630
50	-0.500 ± 0.228	0.201 ± 0.241	1.000 ± 0.300	2.013 ± 0.374	3.172 ± 0.384
100	-0.489 ± 0.159	0.205 ± 0.170	0.998 ± 0.212	2.012 ± 0.268	3.173 ± 0.268
200	-0.476 ± 0.114	0.211 ± 0.123	1.002 ± 0.152	2.020 ± 0.190	3.181 ± 0.187
500	-0.467 ± 0.073	0.214 ± 0.078	1.000 ± 0.096	2.017 ± 0.119	3.173 ± 0.116
1000	-0.463 ± 0.052	0.216 ± 0.054	1.002 ± 0.067	2.018 ± 0.083	3.172 ± 0.080

Thus, if $\frac{a}{b} = 1$ then – resulting directly from the form of criterion (20) – the estimator investigated in this paper reduces to the expectation value, equal to 1 in the case examined here. The condition $\frac{a}{b} = \frac{1}{3}$ means that losses caused by overestimation are three times greater than those implied by underestimation – the estimator should therefore take a value less than the expectation value. Actually, it is shifted to 0.221. This effect is intensified when $\frac{a}{b} = \frac{1}{10}$; the estimator value is then –0.448. Contrary to this, in the case of $\frac{a}{b} = 3$ the losses connected with overestimation are less than those implied by underestimation, therefore the estimator should increase the expected value – it is now equal to 2.011. When $\frac{a}{b} = 10$ this effect is intensified and the value of the estimator investigated in this paper equals 3.154.

In each case represented by a particular column of the above table, as the size of a random sample m increases, the mean estimation error and its standard

deviation tend to zero. From an applicational point of view, these fundamental properties are demanded of estimators used in practice. This property, above all, states that as the sample size increases, the estimators' values achieved tend to a theoretical value, and their "dispersion" decreases. This allows for the obtaining of any desired precision, although the proper sample size must be guaranteed. In practice this implies a necessity for compromise between these two quantities.

Similar results were also obtained for the typical unimodal distribution

$$N(0,1)\ , \tag{71}$$

as well as trimodal

$$0.25N(-5,2) + 0.5N(0,1) + 0.25N(5,2)\ . \tag{72}$$

In particular, in relation to values of standard deviations of the particular distributions (70)-(72), the obtained mean values and variances of estimation errors were, respectively, similar. Moreover, despite significant changes in the type of distribution tested, the procedure of calculating the estimator did not undergo any modification. The above features result from the application of nonparametric methodology of statistical kernel estimators, independent of distribution occurring.

The results regarding the two-dimensional case, worked out in Section 3.2, will now be presented. For clarity of interpretation and comparative analysis, a typical two-dimensional standard normal distribution

$$N\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right]\right) \tag{73}$$

is considered. The obtained results are displayed in Table 2, with similar notations like previously. The first of its five columns is dedicated to the basic verification of correctness of the procedure worked out: thus, if the values of the coefficients a_{ld}, a_{pd}, a_{lg}, a_{pg} are taken equal zero, then, as mentioned before, the two-dimensional problem is reduced to two separate one-dimensional tasks – the results are therefore in accordance with these expectations. In the other four columns of Table 2, it is also shown that changes in the coefficients a_{ld}, a_{pd}, a_{lg}, a_{pg} imply corresponding shifts of the estimator in directions related to smaller losses. The last column also confirms that equality of respective coefficients causes mutual compensation of their influence – in this case the rise of the conditions $|a_{ld}| = |a_{pd}|$ and $|a_{lg}| = |a_{pg}|$, with $a_l = a_p$, results in the values of the estimator of the first coordinate being close to zero.

Remarks regarding complex multimodal distributions are identical to those presented earlier for the one-dimensional case: the results and also the algorithm did not undergo any change.

Finally, results are presented, obtained using the estimator investigated here in the example from the field of optimal control, described in the Introduction

Table 2. Results for two-dimensional case with distribution (73)

m	a_l=10.0, a_p=1.0 a_{lg}=0.0, a_{pg}=0.0 a_{ld}=0.0, a_{pd}=0.0 a_g=1.0, a_d=10.0	a_l=1.0, a_p=1.0 a_{lg}=0.0, a_{pg}=0.0 a_{ld}=3.0, a_{pd}=0.0 a_g=1.0, a_d=1.0	a_l=1.0, a_p=1.0 a_{lg}=0.0, a_{pg}=0.0 a_{ld}=10.0, a_{pd}=0.0 a_g=1.0, a_d=1.0	a_l=1.0, a_p=1.0 a_{lg}=0.0, a_{pg}=0.0 a_{ld}=10.0, a_{pd}=-3.0 a_g=1.0, a_d=1.0	a_l=1.0, a_p=1.0 a_{lg}=0.0, a_{pg}=0.0 a_{ld}=10.0, a_{pd}=–10.0 a_g=1.0, a_d=1.0
	0.9015 0.9015	0.1956 0.1956	0.3985 0.3985	0.2020 0.6163	0.0000 0.8211
10	0.8933 ± 0.3721 0.9130 ± 0.3957	0.1859 ± 0.3322 0.1916 ± 0.3304	0.3635 ± 0.3900 0.3592 ± 0.3766	0.1805 ± 0.3976 0.5648 ± 0.3617	0.0016 ± 0.4258 0.7487 ± 0.3726
20	0.9164 ± 0.2660 0.9342 ± 0.2809	0.1933 ± 0.2307 0.1978 ± 0.2329	0.3855 ± 0.2748 0.3873 ± 0.2715	0.1952 ± 0.2786 0.5991 ± 0.2573	-0.0027 ± 0.3009 0.7943 ± 0.2651
50	0.9265 ± 0.1763 0.9366 ± 0.1790	0.1970 ± 0.1498 0.2000 ± 0.1456	0.3987 ± 0.1794 0.3990 ± 0.1722	0.2026 ± 0.1802 0.6166 ± 0.1636	-0.0002 ± 0.1939 0.8174 ± 0.1693
100	0.9270 ± 0.1235 0.9331 ± 0.1262	0.1989 ± 0.1060 0.1987 ± 0.1020	0.4035 ± 0.1282 0.3997 ± 0.1242	0.2074 ± 0.1277 0.6178 ± 0.1151	0.0035 ± 0.1371 0.8221 ± 0.1188
200	0.9260 ± 0.0883 0.9273 ± 0.0925	0.1981 ± 0.0746 0.1985 ± 0.0731	0.4018 ± 0.0900 0.4030 ± 0.0891	0.2052 ± 0.0906 0.6196 ± 0.0833	0.0003 ± 0.0985 0.8230 ± 0.0854
500	0.9228 ± 0.0592 0.9200 ± 0.0585	0.1987 ± 0.0488 0.1948 ± 0.0481	0.4038 ± 0.0591 0.3989 ± 0.0578	0.2059 ± 0.0591 0.6176 ± 0.6176	0.0004 ± 0.0640 0.8228 ± 0.0557
1000	0.9212 ± 0.0416 0.9210 ± 0.0398	0.1977 ± 0.0336 0.1974 ± 0.0328	0.4025 ± 0.0403 0.4031 ± 0.0403	0.2050 ± 0.0405 0.6210 ± 0.0376	0.0001 ± 0.0439 0.8259 ± 0.0391

as motivation. Thus, on the basis of Fig. 1 the quotient $\frac{a}{b}$ has been fixed at 5.2. It is assumed that the uncertainty of the estimated parameter is of uniform distribution in the interval $[0.5; 1.5]$. For the purpose of comparison, optimal feedback controllers obtained for the parameters Λ, assumed as the arithmetical mean value of elements of the random sample and then the estimator proposed in this paper, were synthesized. Taking into account only those results for which the system was stable, the value of the performance index turned out to be about 40% less in the second case. Moreover, the margin of stability was significantly increased. Generally, the benefits resulting from the application of the methods worked out here are greater for the more complex systems, and when overestimation and underestimation of the model parameters have very differing effects on the performance index, i.e. when the asymmetry of the loss function is sharper.

The presented method was also initially verified experimentally with the aid of an industrial robot, confirming the accuracy of the presented concept. Results – although more complicated because of nonlinearities and complexity of the model applied – yielded similar conclusions to those presented above, obtained using numerical simulation.

6. Summary

This paper has presented the method of parameter identification for those tasks, where the losses resulting from estimation errors can be described by the quadratic and asymmetrical function. The asymmetry represents here the differing influences of under- and overestimation. Besides the basic one-dimensional task, the multidimensional case – i.e. when a vector of parameters is submitted to identification – has been investigated as an example for possible generalizations. The method is universal in nature and can find applications in many areas of science and practice, also outside engineering. Although the uncertainty of the parameters was considered for the most common probabilistic approach, the procedure worked out can equally be used for descriptions of uncertainty other than probability, for example based on fuzzy logic – in this case one can obtain the value of an optimal defuzzyfier.

To estimate the distribution of the uncertainty measure of the tested parameters, the statistical kernel estimators were used, which made the investigated procedure independent of the distribution type. The solution of the problem was based upon the Bayes decision rule, thereby allowing for obtaining of minimum mean value of losses. Finally, a complete algorithm was designed enabling the calculation of the value of an estimator on the basis of measurements of the tested quantity, and also following the fixing – often natural in practical tasks – of the ratio of the loss function coefficient.

It should be stressed that the procedure elaborated is complete – all formulas necessary for its direct application have been given in this paper.

References

Athans, M. and Falb, P.L. (1966) *Optimal Control.* McGraw-Hill, New York.

Berger, J.O. (1980) *Statistical Decision Theory.* Springer-Verlag, New York.

Dahlquist, G. and Bjorck, A. (1983) *Numerical Methods.* Prentice-Hall, Englewood Clifs.

Kulczycki, P. (2000) Fuzzy Controller for Mechanical Systems. *IEEE Transactions on Fuzzy Systems*, **8**, 5, 645-652.

Kulczycki, P. (2001) An Algorithm for Bayes Parameter Identification. *Journal of Dynamic Systems, Measurement, and Control*, **123**, 4, 611-614, 2001.

Kulczycki, P. (2005) *Estymatory jądrowe w analizie systemowej.* WNT, Warsaw, in press.

Kulczycki, P. and Mazgaj A. (2003) Parameter Identification with Nonsymmetrical Quadratic Loss Function for Optimal Control. *9th IEEE International Conference on Methods and Models in Automation and Robotics*, Międzyzdroje (Poland), 25-28 August 2003, **1**, 689-694, CD: I1-5.

Kulczycki, P. and Wisniewski R. (2002) Fuzzy Controller for a System with Uncertain Load. *Fuzzy Sets and Systems*, **131**, 2, 185-195.

Mazgaj, A. (2005) *Modelowanie niepewności parametrów obiektu dla potrzeb sterowania optymalnego.* Ph.D.-thesis, AGH – Cracow University of Science and Technology, Cracow (Poland).

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

Stoer, J. and Bulirsch, R. (1983) *Introduction to Numerical Analysis.* Springer-Verlag, New York.

Wand, M.P. and Jones, M.C. (1995) *Kernel Smoothing.* Chapman and Hall, London, 1995.