
ASYMMETRICAL CONDITIONAL BAYES PARAMETER IDENTIFICATION FOR CONTROL ENGINEERING

**PIOTR KULCZYCKI¹ and MALGORZATA
CHARYTANOWICZ²**

¹Cracow University of Technology, Department of
Automatic Control, Cracow, Poland

²Polish Academy of Sciences, Systems Research
Institute, Warsaw, Poland

One of the main problems in control engineering practice results from unavoidable errors in specifying parameters existing in the object model and the necessity to deal with the unwanted phenomena arising as a result. In this article, a Bayes methodology considering both asymmetrical and conditional aspects is applied for this purpose, with the application of kernel estimators methodology. Use of the Bayes rule enables minimum potential losses to be assumed, while the asymmetry of the occurring loss function also enables the inclusion of different results for under- and overestimation. A conditional approach allows researchers to obtain a more precise result thanks to using information entered as the fixed (e.g., current) values of conditioning factors of continuous and/or binary (also categorical) type. The nonparametric methodology of statistical kernel estimators frees the procedure from arbitrary assumptions concerning the forms of distributions characterizing both the parameter under investigation and conditioning factors. The generalizations introduced here also allow different relevance of particular random sample elements to be taken into account.

Address correspondence to Piotr Kulczycki, Cracow University of Technology, Department of Automatic Control, ul. Warszawska 24, PL-31-155 Cracow, Poland. E-mail: kulczycki@pk.edu.pl

INTRODUCTION

In contemporary control engineering, the model of a controlling object constitutes one of the two most important—after the controlling algorithm itself—elements in deciding the efficiency of a designed device (Morrison 1991; Nusse and Yorke 1997; Soderstrom and Stoica 1994). Despite it seeming that the development of innovative methods, based on knowledge engineering and data exploratory analysis, will slowly lessen the importance of the models, first, the above methods are actually hope for the future rather than the present, and second, procedures based on models will always remain at least valuable assistance in the preliminary analysis of problems.

Regardless of precision of the identification method used, it is of course impossible to obtain an exact model. This concerns its structure as well as the parameters occurring within. In the first case, it is mainly due to the enormous amount of phenomena influencing the considered object, the description of which would result in the model being too complex and consequently of no use for the needs of control algorithm design. In the second, it is not only because of natural metrological limitations but also the fact that one parameter often represents an entire spectrum of phenomena impossible to precisely represent as a single number.

In order to overcome the above limitations in the precision of models, two basic concepts are used in developing control algorithms. The first consists of designing control procedures with imprecision of models already assumed in their nature and consequently robust to resulting errors—in practice they are robust only to some of them in a bounded range, and moreover are sometimes so complicated that they are of theoretical significance only. The second concept dictates that the identification of a model is subject to a higher goal, which is the control algorithm itself, and consequently models are worked out minimizing errors relative to their malevolence regarding quality of the control algorithm. This article offers a contribution to the latter concept.

The subject of the investigations presented here is the specification of parameters existing in the models of a controlled object. The typical procedures for parameter identification generally available in literature do not take into account different—in sign as well as size—results of their under- and overestimation. The method worked out in this article, based on the Bayes decision rule, allows for the consideration of this aspect, while the result obtained minimizes potential losses. Furthermore, the

proposed algorithm enables the influence of various conditioning factors—of continuous and binary (also categorical) type—to be taken into account, for example, temperature or other external conditioning factors of an environmental, technical, and economic nature. The mathematical apparatus relies on the statistical kernel estimators, which frees the method from the types of distributions of imprecise information as well as conditioning factors, and furthermore the introduction of generalizations allows to take into account various relevance of particular elements of a random sample. Although the problem is presented with reference to control problems, the concept is universal in nature and can be applied in a wide range of tasks in contemporary engineering, economy and management, environmental and social issues, biomedicine, and other related fields.

PRELIMINARIES: STATISTICAL KERNEL ESTIMATORS

Presented below is the concept of statistical kernel estimators of the distribution density of the n -dimensional random variable Z , whose n_c coordinates are continuous, whereas the remaining n_b are binary. Variables of both these types will be considered first separately and then together in one approach.

Let therefore (Ω, Σ, P) be a probability space. Let also the n_c -dimensional continuous random variable $X : \Omega \rightarrow \mathbb{R}^{n_c}$ be given, with a distribution characterized by the density f_X . The corresponding kernel estimator $\hat{f}_X : \mathbb{R}^{n_c} \rightarrow [0, \infty)$, calculated using experimentally obtained values for the m -element random sample x_1, x_2, \dots, x_m , in its basic form is defined as

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

where $m \in \mathbb{N} \setminus \{0\}$, the coefficient $h > 0$ is called a smoothing parameter, while the measurable function $K_c : \mathbb{R}^{n_c} \rightarrow [0, \infty)$ of unit integral $\int_{\mathbb{R}^{n_c}} \hat{f}_X(x) dx = 1$, symmetrical with respect to zero and having a weak global maximum in this place, takes the name of a kernel. Thanks to such a formulated concept, kernel estimators enable the density of practically any distribution to be found, without prior knowledge concerning its membership to a fixed class, as occurs in classical parametrical methods.

Let then the n_b -dimensional binary random variable $Y : \Omega \rightarrow \mathbb{B}^{n_b}$, when $\mathbb{B} = \{0, 1\}$, be given. Its distribution density kernel estimator $\hat{f}_Y : \mathbb{B}^{n_b} \rightarrow [0, 1]$, calculated on the basis of experimentally obtained values of the random sample y_1, y_2, \dots, y_m , takes the form

$$\hat{f}_Y(y) = \frac{1}{m} \sum_{i=1}^m K_b(y, y_i), \quad (2)$$

where $m \in \mathbb{N} \setminus \{0\}$, and the kernel $K_b : \mathbb{B}^{n_b} \rightarrow [0, 1]$ is defined by

$$K_b(y, y_i) = \lambda^{n_b - d(y, y_i)} (1 - \lambda)^{d(y, y_i)}, \quad (3)$$

while $\lambda \in [0, 5; 1]$ fulfills the role of a smoothing parameter for the binary component, whereas the function $d : \mathbb{B}^{n_b} \times \mathbb{B}^{n_b} \rightarrow \mathbb{N}$, expressed as $d(y_1, y_2) = (y_1 - y_2)^T (y_1 - y_2)$, refers to a number of coordinates of the vectors y_1 and y_2 , which are different.

Taking the above all together, consider the $(n_c + n_b)$ -dimensional random variable $z \equiv \begin{bmatrix} X \\ Y \end{bmatrix}$, being a composition of the n_c -dimensional random variable X and n_b -dimensional binary variable Y . The kernel K used for the estimation of distribution density of the random variable Z takes the form

$$K(z, z_i) = K_c\left(\frac{x - x_i}{h}\right) K_b(y, y_i), \quad (4)$$

where $z \equiv \begin{bmatrix} x \\ y \end{bmatrix}$ and $z_i \equiv \begin{bmatrix} x_i \\ y_i \end{bmatrix}$ for $i = 1, 2, \dots, m$. Finally, the kernel estimator \hat{f}_Z of the density of distribution of the random variable Z , calculated on the basis of values of the m -element random sample z_1, z_2, \dots, z_m , can be defined as

$$\hat{f}_Z(z) = \frac{1}{mh^{n_c}} \sum_{i=1}^m K(z, z_i), \quad (5)$$

where the kernel K is given by Eq. (4).

Dependencies (1), (2), and (5) constitute a fundamental form of the kernel estimator for the random variables: continuous X , binary Y , and their composition Z , respectively. The tasks concerning the choice of form of the kernel K_c , as well as the smoothing parameters h and λ , are found in the books (Kulczycki 2005, Silverman 1986; Wand

and Jones 1995).¹ In particular, the choice of the kernel form has no practical meaning, and thanks to this it is possible to take into account primarily properties of the estimator obtained (e.g., its class of regularity, boundary of a support) or aspects of calculations, advantageous from the point of view of the applicational problem under consideration.

Practical applications may also use additional procedures, some generally improving the quality of the estimator, and others—optional—fitting the model to an existing reality. In light of the subject considered here, one can—in fact should—recommend for the first group the modification of the smoothing parameter (Kulczycki 2005, section 3.1.6; Silverman 1986, section 5.3.1) and a linear transformation (Kulczycki, section 3.1.4; Silverman, section 4.2.1), whereas for the second, the boundaries of a support of continuous variable X (Kulczycki, section 3.1.8; Silverman, section 2.10).

Exemplary applications of these procedures are described in the publications (Kulczycki 2002a, b; 2007; 2008).

The unconditional probabilistic approach of the concept presented here was shown in the article (Kulczycki 2001a). Many aspects of applications to the sharpening of imprecise information in medical problems, are also found in the publication (Kulczycki and Charytanowicz 2005). The preliminary version of this article was presented as the paper (Kulczycki and Charytanowicz 2007).

FORMULATION OF A PROBLEM UNDER INVESTIGATION

As a simple illustrated example, which serves as an introduction to the problem worked out in this article, consider the following dynamic system (see e.g. Athans and Falb 1966, section 7.2):

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} u(t) \quad x(0) = x_0 = \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix} \quad (6)$$

describing among others the mass $m > 0$ (here often called the “load” because of its practical interpretation) submitted to the force u . The first state variable, x_1 , denotes the position of the mass and the second, x_2 , its

¹For calculating a smoothing parameter, one can especially recommend the plug-in method in the one-dimensional case (Kulczycki 2005, section 3.1.5; Wand and Jones 1995, section 3.6.1), as well as the cross-validation method (Kulczycki section 3.1.5; Silverman 1986, section 3.4.3) in the multidimensional. For a binary variable, see (Kulczycki section 3.1.8; Silverman section 6.1.4). Comments for choice of the kernel may best be found in Kulczycki (section 3.1.3) and Wand and Jones (sections 2.7 and 4.5).

velocity; the parameters x_{01} and x_{02} represent their initial values, respectively. Such a system provides the basis for most research in a wide range of robotics tasks. Let a time-optimal control with the target $x_T = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ be considered; therefore, its goal is to reach the origin of coordinates from the initial value x_0 , in a minimal and finite time, using a control with bounded values.

Detailed analysis of the sensitivity of such a created system provides the following conclusions. Let the value of the parameter m occurring in the object be still denoted as m ; however the value used in feedback controller equations be marked hereinafter by M . If $M = m$, the control is regular, typical for classical time-optimal control theory (Athans and Falb 1966, section 7.2), that is, the state of the system is brought to the switching curve, and being permanently included in this curve hereafter, it reaches the target in a minimal and finite time. In the case $M < m$, as a result of the oscillations around the target, overregulations occur in the system; the target is reached in a finite time. And finally, when $M > m$, after the switching curve is crossed, sliding trajectories appear in the system. Here, too, the target is reached in a finite time. In both of the last two cases (i.e., with $M \neq m$), the time to reach the target state increases from the optimal more or less proportionally to the difference between the values M and m but with significantly differing coefficients—the ratio of which being about 5, while the greater refers to the case $M < m$, when less effective oscillations occur. Similar results also appear for many other control systems, although the above example—basic for robotics—provides satisfactory illustration with this commonly known task.

More details can be found in the publication (Kulczycki 2001b). Results of further research were published in the article (Kulczycki and Wisniewski 2002).

The value of many parameters existing in the models of real objects is influenced by external quantities such as temperature, humidity, expenditure of material gradually used in the technological process, and other factors, mostly of a technical nature, but also economic and environmental. These may be quantities of a continuous or binary character,² multidimensional in both cases. Implementation in an online controlling

²It is worth noting here that binary variables may refer in practice not only to variables with two possible states in their character but also to variables of a different nature but of less importance to a considered problem, which can be reduced to a binary form to simplify the model. Also, so-called categorical variables (i.e., multi-states with undefined order between states, e.g., representing colors or particular manufacturers of a given product) can be denoted with the use of multidimensional binary variables.

algorithm currently measuring their values in many cases allows the model to be made more precise by the information included in them.

In the practice of parametrical identification, one often has the random sample v_1, v_2, \dots, v_m containing experimentally obtained measurements of the parameter under investigation, achieved when conditional factors have assumed the values z_1, z_2, \dots, z_m , respectively. Moreover, the information included in particular measurements may not be of equal relevance (i.e., in many problems, the more current a particular measurement, the greater its relevance). In order to take the above into account, the nonnegative coefficients w_1, w_2, \dots, w_m , such that $\sum_{i=1}^m w_i = m$, are introduced, with values reflecting the relevance of particular elements of a random sample (if $w_i \equiv 1$, then the relevance of these elements remains the same).

Finally, let the parameter under investigation be treated as a random variable denoted hereinafter as $V : \Omega \rightarrow \mathbb{R}$. Let also the n -dimensional random variable $Z : \Omega \rightarrow \mathbb{R}^n$ representing conditioning factors be given and $U = \mathbb{R}^{n+1}$ mean a $(n + 1)$ -dimensional random variable, whose first coordinate represents the investigated parameter, and the remaining n denote further coordinates of the random variable, continuous and binary, that is, $U = \begin{bmatrix} V \\ Z \end{bmatrix}$. Assume also that the distribution of its probability measure has the density $f : \mathbb{R}^{n+1} \rightarrow [0, \infty)$. When this density is specified on the basis of concrete values of the investigated parameter obtained successively for different values of the conditioning factors, one can use, for this purpose, the kernel estimators methodology, described in the section "Preliminaries: Statistical Kernel Estimators." The random sample required there constitutes $(n + 1)$ -dimensional vectors of the

form $\begin{bmatrix} v_i, \{w_i\} \\ z_{i,1} \\ z_{i,2} \\ \vdots \\ z_{i,n} \end{bmatrix}$ for $i = 1, 2, \dots, m$. The first coordinate $v_i, \{w_i\}$ represents

the value of the investigated parameter (the above notation $v_i, \{w_i\}$ means that the relevance of the element v_i is described by the coefficient w_i), whereas the further n coordinates $z_{i,1}, z_{i,2}, \dots, z_{i,n}$ denote the next components of realizations of the conditioning variable Z , for which the above value was obtained.

In concrete practical situations, during control in real time, the realization $Z(\omega)$ of the conditioning random variable becomes fixed. It

is worth noting that in the set of obtained conditioning values $\begin{bmatrix} z_{i,1} \\ z_{i,2} \\ \vdots \\ z_{i,n} \end{bmatrix}$, there may not be a value equal to the realization $Z(\omega)$, or their low

number may not be enough for responsible inference. On the other hand, the kernel estimator allows for proper averaging of data collected.

In practice, when—according to the conditions assumed earlier—underestimation has a significantly different influence than overestimation, the applied procedure should take asymmetry of errors into account. The procedure worked out in later parts of this article fulfills the above demands, providing such an algorithm that the use of its final form does not require the user to have a deeper knowledge of the theoretical aspects.

BAYES ESTIMATION ALGORITHM

The formula for the optimal value of the parameter under investigation will be worked out below based on the Bayes decision rule (Berger 1980). Let the following be given: the sets of states of nature $S = \mathbb{R}$ and allowable decisions $D = \mathbb{R}$, as well as the loss function

$$l(\hat{v}, v) = \begin{cases} -p(\hat{v} - v) & \text{for } \hat{v} - v \leq 0 \\ q(\hat{v} - v) & \text{for } \hat{v} - v \geq 0 \end{cases}, \quad (7)$$

where $v \in S$ and $\hat{v} \in D$, while the coefficients p and q are positive and—it should be stressed—not necessarily equal. Assume that the value of the random variable $Z(\omega)$ is fixed; then, for loss function (7), the Bayes loss function $l_b : D \rightarrow \mathbb{R}$ takes the form

$$l_b(\hat{v}) = \int_{-\infty}^{\hat{v}} q(\hat{v} - v) f_{|Z(\omega)}(v) dv - \int_{\hat{v}}^{\infty} p(\hat{v} - v) f_{|Z(\omega)}(v) dv, \quad (8)$$

where $f_{|Z(\omega)}$ denotes the distribution density of states of nature with the fixed random factor $Z(\omega)$. It is readily shown that the function l_b assumes a minimum for the value of the argument \hat{v} fulfilling the following criterion:

$$\int_{-\infty}^{\hat{v}} f_{|Z(\omega)}(v) dv = \frac{p}{p+q} \int_{-\infty}^{\infty} f_{|Z(\omega)}(v) dv. \quad (9)$$

Because the coefficients p and q are positive, $0 < p/(p+q) < 1$. Therefore, it is easy to see that if the function $f_{|Z(\omega)}$ has a connected support, then the solution of Eq. (9) exists and is unique. Finally, the value of the solution of Eq. (9) constitutes the optimal—in the Bayes sense, i.e.,

assuming a minimal expectation value of losses—value of the parameter under investigation with the fixed random factor $Z(\omega)$.

It is also worth underlining that, with respect to the equation

$$\frac{p}{p+q} = \frac{\frac{p}{q}}{\frac{p}{q} + 1}, \quad (10)$$

there is no need to define the value of the coefficients p and q separately, rather only the ratio p/q .

In the algorithm proposed here, the density f will be specified using statistical kernel estimators. There is applied below the main kernel estimators formula (5) assuming therefore the following form:

$$\hat{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \frac{1}{mh^{n_c+1}} \sum_{i=1}^m K_c\left(\frac{x-x_i}{h}\right) K_b(y, y_i). \quad (11)$$

What is more, for the needs of further considerations, this dependence is generalized to

$$\hat{f}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \frac{1}{mh^{n_c+1}} \sum_{i=1}^m w_i K_c\left(\frac{x-x_i}{h}\right) K_b(y, y_i), \quad (12)$$

while the coefficients w_i for $i = 1, 2, \dots, m$ are nonnegative and such that $\sum_{i=1}^m w_i = m$. For the simplified forms of kernel estimators (1) and (2), the above formulas are subject to obvious reductions.

Let therefore the $(n+1)$ -dimensional random variable $U = \begin{bmatrix} V \\ Z \end{bmatrix}$ be given, whose first coordinate V represents the parameter under investigation, while the remaining n coordinates constitute further components of the conditioning random variable $Z = \begin{bmatrix} X \\ Y \end{bmatrix}$, where X is an n_c -dimensional continuous variable and Y – an n_b -dimensional binary variable. Introducing the natural notations $u = \begin{bmatrix} v \\ z \end{bmatrix}$ as well as $z = \begin{bmatrix} x \\ y \end{bmatrix}$, one can write

$$u = \begin{bmatrix} v \\ z \end{bmatrix} = \begin{bmatrix} v \\ x \\ y \end{bmatrix}. \quad (13)$$

The density f is characterized using the kernel estimator calculated on the basis of m elements

$$u_i = \begin{bmatrix} v_i, \{w_i\} \\ z_i \end{bmatrix} = \begin{bmatrix} v_i, \{w_i\} \\ x_i \\ y_i \end{bmatrix} \quad \text{for } i = 1, 2, \dots, m, \quad (14)$$

on which are composed the particular values of the investigated parameter v_i (together with their relevancies w_i), obtained experimentally for different values of the conditioning variable $z_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$. According to Eq. (12), the function $\hat{f} : \mathbb{R}^{n_c+1} \times \mathbb{B}^{n_b} \rightarrow [0, \infty)$ given as

$$\hat{f}\left(\begin{bmatrix} v \\ x \\ y \end{bmatrix}\right) = \hat{f}\left(\begin{bmatrix} v \\ x \\ y \end{bmatrix}\right) = \frac{1}{mh^{n_c+1}} \sum_{i=1}^m w_i K_c \left(\frac{\begin{bmatrix} v \\ x \end{bmatrix} - \begin{bmatrix} v_i \\ x_i \end{bmatrix}}{h} \right) K_b(y, y_i) \quad (15)$$

is the kernel estimator of the density f . The measurable function $K_c : \mathbb{R}^{n_c+1} \rightarrow [0, \infty)$ fulfills the condition $\int_{\mathbb{R}^{n_c+1}} K_c\left(\begin{bmatrix} s \\ x \end{bmatrix}\right) ds dx = 1$, which for any fixed $x \in \mathbb{R}^{n_c}$ guarantees the existence of its primitive $I_c : \mathbb{R} \rightarrow [0, \infty)$ with respect to the first coordinate, that is,

$$I_c\left(\begin{bmatrix} v \\ x \end{bmatrix}\right) = \int_{-\infty}^v K_c\left(\begin{bmatrix} s \\ x \end{bmatrix}\right) ds. \quad (16)$$

Following the application of the kernel estimators to find the density $f_{|Z(\omega)}$, Eq. (9) then takes the form

$$\begin{aligned} & \sum_{i=1}^m w_i K_b(Y(\omega), y_i) I_c \left(\frac{\begin{bmatrix} \hat{v} \\ X(\omega) \end{bmatrix} - \begin{bmatrix} v_i \\ x_i \end{bmatrix}}{h} \right) \\ &= \frac{p}{p+q} \frac{1}{h^{n_c+1}} \sum_{i=1}^m w_i K_b(Y(\omega), y_i) \int_{-\infty}^{\infty} K_c \left(\frac{\begin{bmatrix} v \\ X(\omega) \end{bmatrix} - \begin{bmatrix} v_i \\ x_i \end{bmatrix}}{h} \right) dv. \quad (17) \end{aligned}$$

The solution of the above equation constitutes at last the optimal value of the parameter under investigation with the fixed random factor $Z(\omega) = \begin{bmatrix} X(\omega) \\ Y(\omega) \end{bmatrix}$. If a kernel assuming positive values is used to construct the estimator, then the left side is a function of the argument \hat{v} , continuous and strictly increasing from zero to a positive value, which—thanks to the dependence $0 < p/(p+q) < 1$ —is higher than

the value of the right side. The solution of Eq. (17) therefore exists and is unique.

In practice, this solution can be calculated effectively based on Newton’s numerical method (Kincaid and Cheney 2002). The desired value is given then as a limit of the sequence $\{\hat{v}_k\}_{k=0}^\infty$ defined by formulas

$$\hat{v}_0 = \frac{1}{\sum_{i=1}^m w_i} \sum_{i=1}^m w_i v_i \tag{18}$$

$$\begin{aligned} \hat{v}_{k+1} = \hat{v}_k + & \frac{\frac{p}{p+q} \frac{1}{h^{n_c+1}} \sum_{i=1}^m w_i K_b(Y(\omega), y_i) \int_{-\infty}^{\infty} K_c \left(\frac{\begin{bmatrix} v \\ X(\omega) \end{bmatrix} - \begin{bmatrix} v_i \\ x_i \end{bmatrix}}{h} \right) dv}{\frac{1}{h^{n_c+1}} \sum_{i=1}^m w_i K_b(Y(\omega), y_i) K_c \left(\frac{\begin{bmatrix} \hat{v}_k \\ X(\omega) \end{bmatrix} - \begin{bmatrix} v_i \\ x_i \end{bmatrix}}{h} \right)} \\ & - \frac{\sum_{i=1}^m w_i K_b(Y(\omega), y_i) I_c \left(\frac{\begin{bmatrix} \hat{v}_k \\ X(\omega) \end{bmatrix} - \begin{bmatrix} v_i \\ x_i \end{bmatrix}}{h} \right)}{\frac{1}{h^{n_c+1}} \sum_{i=1}^m w_i K_b(Y(\omega), y_i) K_c \left(\frac{\begin{bmatrix} \hat{v}_k \\ X(\omega) \end{bmatrix} - \begin{bmatrix} v_i \\ x_i \end{bmatrix}}{h} \right)} \tag{19} \end{aligned}$$

for $k = 0, 1, \dots,$

with the stop condition $|\hat{v}_k - \hat{v}_{k-1}| \leq 0,01\hat{\sigma}_V$, where $\hat{\sigma}_V$ denotes an estimator of the standard deviation obtained for the sample v_1, v_2, \dots, v_m .

As mentioned in the section “Preliminaries: Statistical Kernel Estimators,” the assumed form of the kernel K_c has no practical meaning from the statistical point of view, and thanks to this, in applications a suitable choice for the needs of a concrete applicational task is possible. In the procedure worked out above, it was demanded that the kernel K_c assumes positive values, and its primitive with respect to the first coordinate is expressed by a convenient analytic formula. These properties

ensure the Cauchy kernel defined as

$$K_c\left(\begin{bmatrix} v \\ x \end{bmatrix}\right) = K_c\left(\begin{bmatrix} v \\ x_1 \\ x_2 \\ \vdots \\ x_{n_c} \end{bmatrix}\right) = \frac{1}{A} \frac{1}{(1 + v^2 + x_1^2 + x_2^2 + \dots + x_{n_c}^2)^a}, \quad (20)$$

while $a = [(n_c + 3)/2]$ and $A = (n_c + 1)V_{n_c+1} \int_0^\infty r^{n_c} / (1 + r^2)^a dr$, where $[b]$ means the integer part of the number $b \in \mathbb{R}$, while V_n denotes the volume of an n -dimensional unit ball in the space \mathbb{R}^n and amounts to

$$V_n = \begin{cases} \frac{(2\pi)^{n/2}}{2 \cdot 4 \cdot \dots \cdot n} & \text{for } n \text{ even} \\ \frac{2(2\pi)^{(n-1)/2}}{1 \cdot 3 \cdot \dots \cdot n} & \text{for } n \text{ odd} \end{cases} \quad (21)$$

In turn, for the fixed dimension n_c , the expression $\int_0^\infty u^{n_c} / (1 + u^2)^a du$ can be easily calculated based on recurrence formulas, which are found in typical integral tables.

EXPERIMENTAL VERIFICATION

The correct functioning of the algorithm presented in this article has been at first confirmed with detailed numerical simulation. In any case under research, as the sample size increased, the obtained parameter value converged to the theoretical, and the standard deviation to zero. The above asymptotic features are of fundamental significance from an applicational point of view, as they prove that it is possible to obtain any precision wished, although this requires the assurance of a sufficient random sample size. In practice, therefore, the necessity of the right compromise between these quantities is called for.

The convergence speed increased as the value $p/(p + q)$ was closer to 0,5, although even for extreme cases (i.e., when it neared 0 or 1), satisfactory results were achieved. The convergence speed also increased as the value of the conditioning variable came within the range of its modal values; however, even for far-away areas, the results also were fair. It is worth noting that the cases where $r = 0,1$ or $r = 0,9$ and when the value of the conditioning variable appears in the neighborhood of its second standard deviation generally become very difficult problems, which are naturally associated with greater demands, and for some classical methods actually impossible to carry out in practice at all.

The above corollaries have been successfully proven in numerous obtained results of numerical simulation, including those for multimodal and asymmetrical distributions, taking into account binary conditioning factors, inference for lack of data from the neighborhood of a fixed value of a conditioning variable, as well as an additional procedure for boundaries of a support; for many details, see the work of Charytanowicz (2005).

To demonstrate the influence of the conditioning factor, assume—for simplicity and transparency of the example—that $n_c = 1$ and $n_b = 0$, while the distribution of the random variable $v = \begin{bmatrix} V \\ Z \end{bmatrix}$ is normal with the expectation value $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and the covariance matrix $\begin{bmatrix} 1 & 0,7 \\ 0,7 & 1 \end{bmatrix}$. In Table 1, the asymptotic (i.e., with $m \rightarrow \infty$) values of the parameter under investigation, obtained for the different ratios p/q and values of the conditioning random variable $Z(\omega)$, are shown. Comparing particular columns, one can note that the value of the conditioning variable significantly influences the achieved value of the investigated parameter, offering notable possibilities for making a model more accurate thanks to the influence of information provided in a properly precise (e.g., current) value of this variable, even if in practical problems the correlation coefficient is significantly less than 0,7 assumed in the case described above.

Returning to the example presented at the beginning of the section “Formulation of a Problem Under Investigation,” from the time-optimal control field, extensive simulation and experimental research was carried out and confirmed the concept described. In particular assuming the coefficients p and q such that $p/q = 0,2$, sliding trajectories clearly outnumbered overregulations in the system, giving a time of about 10% less than that for results obtained using the mean value, and about 5% shorter than in the case of the classical sliding mode control (Utkin 1992), where sliding trajectories had a notably larger number of switchings, which lowered their effectiveness. The algorithm worked out here was also applied in the identification of other parameters

Table 1. Asymptotic value of the parameter under investigation in relation to the conditioning factor value

$Z(\omega) \backslash p/q$	0,1	0,3	0,5	0,7	0,9
0	-0,92	-0,37	0,00	0,37	0,92
1	-0,22	0,33	0,70	1,07	1,62
2	0,48	1,03	1,40	1,77	2,32

existing in the following related systems: inertia coefficient of a drive and value of motion resistance (Kulczycki 2000). The application of the concept presented in this article simultaneously in different causes of uncertainty—e.g., mass (load), inertia of a drive, motion resistance—did not hinder the operation of the system; on the contrary, under- and overestimations of different parameters showed a tendency to mutual compensation. In this situation, the increase in the time to reach the target set was less than the simple sum of such increases resulting from the uncertainty of particular factors.

Finally, it is worth stressing that in every case researched, precision of the characteristics describing the parameter under investigation by providing the proper value for conditioning factors improved the result in proportion to the degree of differentiation of object features with respect to those factors. This occurred in the case of random changes in values for these factors, as well as regular object nonstationarity resulting for example from daily temperature fluctuations. In today's age of ever more available current measurement data and the possibility of instant inclusion in digital control algorithms used, this should be particularly underlined, providing real advantages from the procedure presented in this article.

REFERENCES

- Athans, M. and Falb, P. L. 1966. *Optimal control*. New York: McGraw-Hill.
- Berger, J. O. 1980. *Statistical decision theory*. New York: Springer-Verlag.
- Charytanowicz, M. 2005. *Bayesowska determinizacja informacji nieprecyzyjnej w zagadnieniach medycznych*. Warsaw: Systems Research Institute, Polish Academy of Sciences.
- Kincaid, D. and Cheney, W. 2002. *Numerical analysis*. Pacific Grove: Brooks/Cole.
- Kulczycki, P. 2000. Fuzzy controller for mechanical systems. *IEEE Transactions on Fuzzy Systems* 8: 645–652.
- Kulczycki, P. 2001a. An algorithm for Bayes parameter identification. *Journal of Dynamic Systems, Measurement, and Control* 123: 611–614.
- Kulczycki, P. 2001b. A Fuzzy Approach for Mechanical System with Uncertain Load. *Proceedings of the European Control Conference*, Porto, Portugal, 4–7 September 2001, pp. 2658–2663, CD: ecc4561.
- Kulczycki, P. 2002a. Statistical inference for fault detection: A complete algorithm based on kernel estimators. *Kybernetika* 38: 141–168.
- Kulczycki, P. 2002b. A test for comparing distribution functions with strongly unbalanced samples. *Statistica* LXII: 39–49.

- Kulczycki, P. 2005. *Estymatory jądrowe w analizie systemowej*. Warsaw: WNT.
- Kulczycki, P. 2007. Estymatory jądrowe w badaniach systemowych. In *Techniki informacyjne w badaniach systemowych*, edited by P. Kulczycki, O. Hryniewicz, and J. Kacprzyk. Warsaw: WNT, pp. 79–105.
- Kulczycki, P. 2008. Kernel estimators in industrial applications. In *Soft computing applications in industry*, edited by B. Prasad. Berlin: Springer-Verlag, pp. 69–91.
- Kulczycki, P. and Charytanowicz, M. 2007. Automatic Control Tasks Featuring Asymmetrical Conditional Identification of Parameters. *Proceedings of the 13th IEEE/IFAC International Conference on Methods and Models in Automation and Robotics*, Szczecin, Poland, 27–30 August 2007, CD: 0236.
- Kulczycki, P. and Charytanowicz, M. 2005. Bayes sharpening of imprecise information. *Applied Mathematics and Computer Science* 15: 393–404.
- Kulczycki, P. and Wisniewski, R. 2002. Fuzzy controller for a system with uncertain load. *Fuzzy Sets and Systems* 131: 185–195.
- Morrison, F. 1991. *The art of modeling dynamic systems*. New York: Multiscience Press.
- Nusse, H. E. and Yorke, J. A. 1997. *Dynamics: Numerical exploration*. New York: Springer-Verlag.
- Silverman, B. W. 1986. *Density estimation for statistics and data analysis*. London: Chapman and Hall.
- Soderstrom, T. S. and Stoica, P. 1994. *System identification*. Englewood Cliffs: Prentice Hall.
- Utkin, V. I. 1992. *Sliding modes in control optimization*. Berlin: Springer-Verlag.
- Wand, M. P. and Jones, M. C. 1995. *Kernel smoothing*. London: Chapman and Hall.