Nonparametric Estimation for Control Engineering

PIOTR KULCZYCKI *
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6, PL-01-447 Warsaw
POLAND

Abstract: - The subject of this paper is the application of nonparametric estimation methods – in particular statistical kernel estimators – for control engineering. Such methods allow the useful characterization of probability distributions without arbitrary assumptions regarding their membership to a fixed class. A detailed description of the Bayes parameter estimation with asymmetrical polynomial loss function will be given, as will one for fault detection in dynamical systems as objects of automatic control, in the scope of detection, diagnosis and prognosis of malfunctions. To this aim the basics of data analysis and exploration tasks – identification of outliers, clustering, and classification – solved using uniform mathematical apparatus based on the kernel estimators methodology will also be considered. In every case the final result will be an algorithm ensuring that its practical implementation does not demand of the user detailed knowledge of the theoretical aspects, or laborious research and calculations.

Key-Words: - soft computing, nonparametric estimation, kernel estimators, information technology, data analysis and exploration, parametric identification, fault detection, optimal control, robust control

1 Introduction

The specification, based on experimental data, of functions which characterize an object under investigation, constitutes one of the main tasks in modern science and technological problems. A typical example here is the estimation of density function of random variable distribution from any given sample. The classical procedures rely here on arbitrary assumption of the form of this function, and then in specification of its parameters. These are called parametric methods. A valuable advantage is their theoretical and calculational simplicity, as well as their being commonly known and present in subject literature. Nowadays – along with the dynamic development of computer systems – nonparametric methods, whose main feature constitutes a lack of arbitrary assumptions of the form of a density function, are used more and more often. In a probabilistic approach, kernel estimators are becoming the principal method in this subject. Although their concept is relatively simple and their interpretation transparent, the applications are impossible without a high class of computer which, even until recently, significantly hindered theoretical, and especially practical research.

In this paper, first – in Section 2 – the basics of kernel estimators methodology are briefly presented. Thanks to availability and the possibilities of contemporary computer systems as well as the automation of metrological and data gathering processes, the universal character of kernel estimators allows for their broad application in various problems of modern science and technology. In Sections 3 and 4 the possibilities of applications of the kernel estimators methodology are shown using as examples the following subjects from control engineering:

- parametric identification illustrated in automatic control applications;
- data analysis and exploration – recognition of atypical elements (outliers), clustering, and classification – applied to the detection and diagnosis of devices working in real-time.

The following text also contains results of research in the field of kernel estimators carried out together with M. Charytanowicz, K. Daniel, P.A. Kowalski, S. Lukasik, A. Mazgaj, C. Prochot, and J. Waglowski. It also contains material presented in the works [9, 10]. An extended version of this paper will be published as [12].

2 Kernel Estimators

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

* Also: Cracow University of Technology, Head of Department of Automatic Control, ul. Warszawska 24, PL-31-155 Cracow, Poland, e-mail: kulczycki@pk.edu.pl, http://www.control.pk.edu.pl/kulczycki .

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

where the measurable, symmetrical with respect to zero and having a weak global maximum in this point, function $K : \mathbb{R}^n \to [0, \infty)$ fulfils the condition $\int_{\mathbb{R}^n} K(x)\,\mathrm{d}x = 1$ and is called a kernel, whereas the positive coefficient *h* is referred to as a smoothing parameter (for interpretation see Fig. 1). It is worth noting that a kernel estimator allows the identification of density for practically every distribution, without any assumptions concerning its membership to a fixed class. Atypical, complex distributions, also multimodal, are regarded here as textbook unimodal. In the multidimensional case this enables, among others, the discovery of total dependences between particular coordinates of the random variable under investigation.

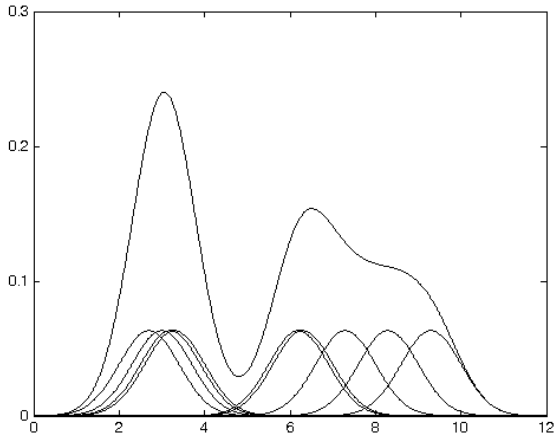

Fig. 1. Kernel density estimator (1) of one-dimensional random variable distribution.

Setting of the quantities introduced in definition (1), i.e. choice of the form of the kernel *K* as well as calculation of the value for the smoothing parameter *h*, is most often carried out according to the criterion of minimum of an integrated mean-square error. Broader discussion and practical algorithms are found in the books [8; 19, 20] [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, etc.) 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 – possibly fitting the model to an existing reality. For the first group one should recommend the modification of the smoothing parameter [8 – Section 3.1.6; 19 – Section 5.3.1] and a linear transformation [8 – Section 3.1.4; 19 – Section 4.2.1], while for the second, the boundaries of a support [8 – Section 3.1.8; 19 – Section 2.10]. It is worth mentioning also the possibility of applying data compensation and dimensionality reduction procedures – original and useful algorithms can be found in the book [18 – Sections 2.5 and 3.4].

Kernel estimators allow modeling of the distribution density – a basic functional characteristic of random variables. Consequently this is fundamental in obtaining other functional characteristics and parameters. For example, if in a one-dimensional case the kernel *K* is such chosen that its primitive $I(x) = \int_{-\infty}^{x} K(y)\,\mathrm{d}y$ may be analytically obtained, then the estimator of the distribution function

$$\hat{F}(x) = \frac{1}{m} \sum_{i=1}^{m} I\left(\frac{x - x_i}{h} \right) \qquad (2)$$

can be easy calculated. Next, if the kernel *K* has (strictly) positive values, the solution for the equation

$$\hat{F}(x) = r \qquad (3)$$

constitutes the kernel estimator of quantile of the order $r \in (0,1)$. For details and proofs of strong consistencies see [15].

Later in this paper, examples of applications of the kernel estimators methodology to control engineering tasks will be presented. First, Section 3 shows its use in calculate optimal – in the Bayes sense – values of parameters of automatic control objects, as an example of a subordinate factor with respect to the control algorithm. Next, Section 4 describes a fault detection system, after considerations regarding the basic procedures for data analysis and exploration, as an example of superior – with respect to such an algorithm – factor.

3 Parameter Identification

One of the main problems of control engineering is parameter identification – the specification of values of parameters existing in the model of an investigated object. In a typical practical task, *m* independent measurements of the parameter, although suffering from errors of different origin, are available. On this basis one should define the value which, from an overall point of view of the problem to be worked out, would best represent phenomena described by this parameter. Usual

[1] For calculating a smoothing parameter one can especially recommend the plug-in method in the one-dimensional case [8 – Section 3.1.5; 19 – Section 3.6.1], as well as the cross-validation method [8 – Section 3.1.5; 18 – Section 3.4.3] in the multidimensional. Comments for the choice of kernel may best be found in [8 – Section 3.1.3; 19 – Sections 2.7 and 4.5].

estimation procedures, based on minimum integrated mean-square error or maximum likelihood methods, are applied mostly because of their popularity and availability in subject literature, however they do not allow differing causes of estimation errors to be taken into account.

This problem will be illustrated for the example of optimal control. The performance index, fundamental for the above task, may be used for testing not only the quality of a control, but also the procedure of identifying model parameters. As an example let the system, whose dynamic is described by the following differential equation

$$\dot{x}(t) = \begin{bmatrix} v & 1 \\ 0 & v \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ v \end{bmatrix} u(t) \quad , \tag{4}$$

while $v \in \mathbb{R} \setminus \{0\}$, be given. If the optimal feedback controller with quadratic performance index has been constructed with the value $V \in \mathbb{R} \setminus \{0\}$ not necessarily equal to the parameter v existing in object (4), then the obtained graph for this index can be approximated with reasonable precision by a quadratic function where coefficients differ for negative and positive errors (Fig. 2). Treating obtained values of an examined parameter as realizations of a random variable with the distribution density f, one can calculate the value of the optimal – in the sense of minimum expectation value of performance index for the control – estimator using Bayes' decision rule [3]. If the distribution of the above random variable is obtained with the aid of kernel estimators, the algorithm worked out is suitable for calculational procedures and in consequence for practical use.

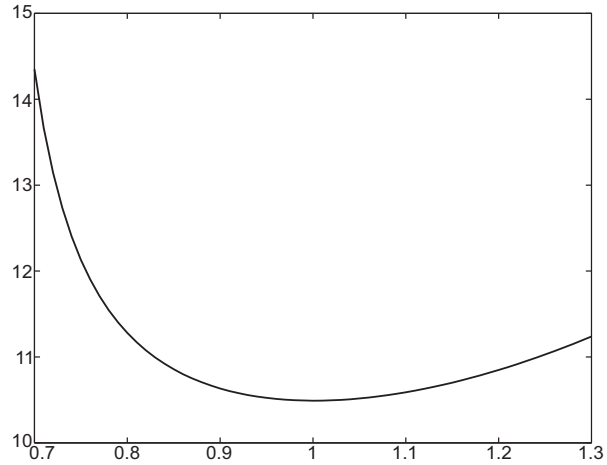

Fig. 2. Performance index value as a function of the parameter V; ($v = 1$).

First, a basic case will be investigated, considering the single parameter *v*. As sets of possible decisions as well as states of nature, the set of real numbers $\mathbb{R}$ can be assumed, while the loss function is given in the asymmetrical quadratic form:

$$l(\hat{v}, v) = \begin{cases} p(\hat{v} - v)^2 & \text{for} \quad \hat{v} - v \le 0 \\ q(\hat{v} - v)^2 & \text{for} \quad \hat{v} - v \ge 0 \end{cases} , \tag{5}$$

where $p, q > 0$ (note that these coefficients can be different) and $\hat{v}$ denotes the desired value of the Bayes decision which here fulfills the role of optimal estimator. This value is then given as the solution of the following equation with the argument $\hat{v}$:

$$(p - q) \int_{-\infty}^{\hat{v}} (\hat{v} - v) f(v)\, \mathrm{d}v = p \int_{-\infty}^{\infty} (\hat{v} - v) f(v)\, \mathrm{d}v \ . \tag{6}$$

Solving the above criterion is generally no easy task. If, however, the kernel estimators methodology is used in specifying the density f, then, thanks to the proper choice of the kernel form, the effective numerical algorithm can be obtained. Let therefore *m* measurements for the examined parameter be given, treated as random sample $x_1, x_2, \ldots, x_m$. For the chosen kernel K one may define the following real mappings $I(x) = \int_{-\infty}^{x} K(y)\, \mathrm{d}y$, $J(x) = \int_{-\infty}^{x} yK(y)$, and finally:

$$\bar{I}(x) = \frac{1}{mh} \sum_{i=1}^{m} I\left(\frac{x - x_i}{h} \right) \tag{7}$$

$$\bar{J}(x) = \frac{1}{mh} \sum_{i=1}^{m} J\left(\frac{x - x_i}{h} \right) \ . \tag{8}$$

Then, criterion (6) takes the form of the equation

$$(p - q)[\hat{v}\bar{I}(\hat{v}) - \bar{J}(\hat{v})] - p\hat{v} = -p \frac{1}{m} \sum_{i=1}^{m} x_i \ . \tag{9}$$

Its solution exists and is unique. Denoting the left and right sides of the above equation as follows

$$L(\hat{v}) = (p - q)[\hat{v}\bar{J}(\hat{v}) - \bar{I}(\hat{v})] - p\hat{v} \tag{10}$$

$$P = -p \frac{1}{m} \sum_{i=1}^{m} x_i \tag{11}$$

and calculating the derivative of function (10):

$$L'(\hat{v}) = (p - q)\bar{J}(\hat{v}) - p \ , \tag{12}$$

one can then – using Newton's iterative method – effectively obtain a solution for criterion (5) as a limit of the sequence $\{v^{(k)}\}_{k=0}^{\infty}$ given by the formulas

$$v^{(0)} = \frac{1}{m} \sum_{i=1}^{m} x_i \tag{13}$$

$$v^{(k+1)} = v^{(k)} + \frac{P - L(v^{(k)})}{L'(v^{(k)})} \qquad \text{for } k = 0, 1, \ldots \ . \tag{14}$$

The properties of kernel estimators allow generalizations of the above concept to be made for the multidimensional (parameters' vector), polynomial (loss function to a power greater than two), and conditional (dependence on conditional factors) cases.

Thus, in the multidimensional case, i.e. the task of a parameters' vector estimation, it is possible to carry out the above-proposed procedure with respect to a multidimensional random variable. As an example, if for the two-dimensional parameters' vector $v=[v_1, v_2]^{\mathrm{T}}$ one assumes the loss function in the following asymmetrical quadratic form:

$$l\left(\begin{bmatrix}\hat{v}_1\\ \hat{v}_2\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix}\right)= \tag{15}$$

$$\begin{cases} c_l(\hat{v}_1-v_1)^2 + c_{ld}(\hat{v}_1-v_1)(\hat{v}_2-v_2) + c_d(\hat{v}_2-v_2)^2 \\ \qquad\qquad \text{when } \hat{v}_1-v_1\le 0 \text{ i } \hat{v}_2-v_2\le 0 \\ c_p(\hat{v}_1-v_1)^2 + c_{pd}(\hat{v}_1-v_1)(\hat{v}_2-v_2) + c_d(\hat{v}_2-v_2)^2 \\ \qquad\qquad \text{when } \hat{v}_1-v_1\ge 0 \text{ i } \hat{v}_2-v_2\le 0 \\ c_l(\hat{v}_1-v_1)^2 + c_{lg}(\hat{v}_1-v_1)(\hat{v}_2-v_2) + c_g(\hat{v}_2-v_2)^2 \\ \qquad\qquad \text{when } \hat{v}_1-v_1\le 0 \text{ i } \hat{v}_2-v_2\ge 0 \\ c_p(\hat{v}_1-v_1)^2 + c_{pg}(\hat{v}_1-v_1)(\hat{v}_2-v_2) + c_g(\hat{v}_2-v_2)^2 \\ \qquad\qquad \text{when } \hat{v}_1-v_1\ge 0 \text{ i } \hat{v}_2-v_2\ge 0 \end{cases},$$

where $c_l, c_p, c_g, c_d > 0$, $c_{ld}, c_{pg} \ge 0$ and $c_{pd}, c_{lg} \le 0$, then here criterion (6) takes on the form of two equations, defining Bayes' decisions $\hat{v}_1$ and $\hat{v}_2$ of similar, albeit slightly more complex form.

The form of the asymmetrical quadratic loss function (5) may be generalized to a polynomial case:

$$l(\hat{v},v)=\begin{cases}(-1)^k\, p(\hat{v}-v)^k & \text{for} \quad \hat{v}-v\le 0 \\ q(\hat{v}-v)^k & \text{for} \quad \hat{v}-v\ge 0\end{cases}, \tag{16}$$

while $p,q>0$ and $k=2,3,\ldots$. For instance when $k=3$, a criterion analogical to equation (6) becomes

$$(p+q)\int_{-\infty}^{\hat{v}}(\hat{v}-v)^2 f(v)\,\mathrm{d}v = \mathrm{q}\int_{-\infty}^{\infty}(\hat{v}-v)^2 f(v)\,\mathrm{d}v \ . \tag{17}$$

Similar investigations concerning an asymmetrical linear loss function, i.e. in the case when formula (5) is replaced with the following dependence:

$$l(\hat{v},v)=\begin{cases}-p(\hat{v}-v) & \text{for} \quad \hat{v}-v\le 0 \\ q(\hat{v}-v) & \text{for} \quad \hat{v}-v\ge 0\end{cases}, \tag{18}$$

with $p,q>0$, were presented in the paper [7].

In the case where the examined parameter depends heavily on a conditioning factor, the proposed procedures can be generalized with the purpose of including its influence. Let therefore the conditioning factor be given in the form of the random variable Z, whose distribution has a density. Statistical kernel estimators can be used to estimate the density of total distribution of the random variable $[X,Z]^{\mathrm{T}}$ being the comparison of the variables X and Z. During application of the procedure, after fixing the concrete value of the conditional factor $Z(\omega)$, calculation can be applied to a "cross-section" defined by this value. When significant conditioning for factors represented by variable Z occurs to an object under consideration, such an approach can considerably improve the quality of received results. For the linear case see the paper [11]; the results for the quadratic and polynomial cases will be published soon.

The presented method was verified experimentally for various optimal control systems. In particular in the case of the task presented at the beginning of this section, the values of the performance index was even 43% less than obtained with traditional use of the mean value of a random sample as the estimator $\hat{v}$.

The above text also contains material from research carried out together with M. Charytanowicz and A. Mazgaj, published in the common papers [11, 16].

4 Fault Detection

The fault detection and diagnosis problem has lately become one of the most important challenges in modern control engineering. Early discovery of anomalies appearing in the operation of a controlled system, from an industrial robot to a nuclear reactor, most often allows serious incidents and even catastrophes to be avoided, which could save material damage, or loss of human life. Secondly, confirmation of kind and location of these anomalies is of fundamental meaning, especially when supervising large systems like complex chemical installations, as well as modern ships and airplanes. The importance of the above actions is multiplied by a psychological factor expressed by an increased feeling of safety, as well as – for the producer – prestige and commercial reputation. Finally, economic reasons often translate into a significant decrease in running costs, above all by ensuring the proper technological conditions as well as rationalizing overhauls and reducing repairs. Among the many different procedures used with this aim, the most universal are statistical methods. This paper presents the concept of a fault detection system, based on the kernel estimators methodology, covering:

- detection, so discovery of the existence of potential anomalies in the technical state of a device under supervision;
- diagnosis, that is identification of these anomalies;

- prognosis, i.e. warning of the threat of their occurrence in the near future, together with anticipated classification.

First, Section 4.1 presents possible applications of kernel estimators to the fundamental problems of data analysis and exploration: recognition of atypical elements (outliers), clustering and classification. It is worth noting that use of a single methodology for all investigated tasks significantly simplifies the process of synthesis of a fault detection system being worked upon. Consequently, Section 4.2, where the fault detection system designed here is described, will consist mainly of references to earlier material, and integrate them into one coherent idea.

4.1 Data Analysis and Exploration

The application of kernel estimators for recognition of atypical elements, clustering, and classification will be subsequently investigated below. In all three cases the *n*-dimensional random variable $X:\Omega\to\mathbb{R}^n$ is considered.

First, in many problems of data analysis the task of recognizing atypical elements (outliers) – those which differ greatly from the general population – arises. This enables the elimination of such elements from the available set of data, which increases its homogeneity (uniformity), and facilitates analysis, especially in complex and unusual cases. In practice, the recognition process for outliers is most often carried out using procedures of statistical hypotheses testing [2]. The significance test based on the kernel estimators methodology will now be described.

Let therefore the random sample $x_1,\ x_2,\ldots,x_m$ treated as representative, therefore including a set of elements as typical as possible, be given. Furthermore, let $r\in(0,1)$ denote an assumed significance level. The hypothesis that $\tilde{x}\in\mathbb{R}^n$ is a typical element will be tested against the hypothesis that it is not, and therefore should be treated as an outlier. The statistic $S:\mathbb{R}^n\to[0,\infty)$, used here, can be defined by

$$S(\tilde{x})=\hat{f}(\tilde{x}) \quad , \tag{19}$$

where $\hat{f}$ denotes a kernel estimator of density, obtained for the random sample $x_1,\ x_2,\ldots,x_m$ mentioned above, while the critical set takes the left-sided form $A=(-\infty,a]$, when a constitutes the kernel estimator of quantile of the order r (see the end of Section 2), calculated for the sample $\hat{f}(x_1),\ \hat{f}(x_2),\ldots,\hat{f}(x_m)$, with the assumption that random variable support is bounded to nonnegative numbers.

Secondly, the aim of clustering is the division of a data set – for example given in the form of the random sample $x_1,\ x_2,\ldots,x_m$ – into subgroups (clusters), with every one including elements "similar" to each other, but with significant differences between particular subgroups [5]. In practice this often allows the decomposition of a large data set with differing characteristics of elements into subsets containing elements of similar properties, which considerably facilitates further analysis, or even makes it possible at all. The following clustering procedure based on kernel estimators, taking advantage of the gradient methods concept [4] will be presented now.

Here the natural assumption is made that clusters are associated to modes – local maximums of the density kernel estimator $\hat{f}$, calculated for the considered random sample $x_1,\ x_2,\ldots,x_m$. Within this procedure, particular elements are moved in a direction defined by a gradient, according to the following iterative algorithm:

$$x_j^0=x_j \quad \text{for} \quad j=1,2,\ldots,m \tag{20}$$

$$x_j^{k+1}=x_j^k+b\frac{\nabla\hat{f}(x_j^k)}{\hat{f}(x_j^k)} \quad \text{for} \quad j=1,2,\ldots,m \ \text{ and } \ k=0,1,\ldots \ , \tag{21}$$

where $b>0$ and ∇ denotes a gradient. Thanks to the proper choice of form of the kernel *K*, a suitable analytical formula for the gradient ∇ becomes possible. In practice the value $b=h^2/(n+2)$ may be used. As a result of the following iterative steps, the elements of the random sample move successively, focusing more and more clearly on a certain number of clusters. They can be defined after completing the k^*-th step, where k^* means the smallest number k such that $|D_k-D_{k-1}|\le c\,D_0$, where $c>0$ and $D_0=\sum_{i=1}^m\sum_{j=i+1}^m d(x_i,x_j)$, $D_{k-1}=\sum_{i=1}^m\sum_{j=i+1}^m d(x_i^{k-1},x_j^{k-1})$, $D_k=\sum_{i=1}^m\sum_{j=i+1}^m d(x_i^k,x_j^k)$, i.e. they are the sums of the distances between particular elements of the random sample under consideration before the beginning of algorithm (20)-(21) and having performed the (*k*–1)-th and *k*-th steps, respectively. For practical purposes $c=0.001$ may be used. Thus, after k^*-th step, one should calculate the kernel estimator for mutual distances of the elements $x_1^{k^*},\ x_2^{k^*},\ldots,x_m^{k^*}$ (under the assumption of nonnegative support of the random variable), and next, the value can be found where this estimator takes on the local minimum for the smallest value of its argument, omitting a possible minimum in zero. Finally, particular clusters are assigned those elements, whose distance to at least one of the others, is

not greater than the above value. Thanks to the possibility of change in the smoothing parameter value, it becomes possible to affect the range of a number of obtained clusters, albeit without arbitrary assumptions concerning the strict value of this number, which allows it to be suited to a true data structure. Moreover, possible changes in intensity of the smoothing parameter modification procedure enable influence on the proportion of clusters located in dense areas of random sample elements to the number of clusters on the "tails" of the distribution. The detailed description of the above procedure can be found in the paper [13].

Thirdly, the application of kernel estimators in a classification task is considered. Let the number $J \in \mathbb{N} \setminus \{0,1\}$ be given. Assume also, that the possessed random sample $x_1, x_2, \ldots, x_m$ has been divided into $J \in \mathbb{N} \setminus \{0,1\}$ nonempty and separate subsets $\{x_1, x_2, \ldots, x_{m_1}\}$, $\{x_1, x_2, \ldots, x_{m_2}\}$, ..., $\{x_1, x_2, \ldots, x_{m_J}\}$, while $\sum_{j=1}^{J} m_j = m$, representing classes with features as mutually different as possible. The classification task requires deciding into which of them the given element $\tilde{x} \in \mathbb{R}^n$ should be reckoned [5].

The kernel estimators methodology provides a natural mathematical tool for solving the above problem in the optimal – in the sense of minimum for expectation of losses – Bayes approach. Let thus $\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_J$ denote kernel estimators of density calculated for subsets $\{x_1, x_2, \ldots, x_{m_1}\}$, $\{x_1, x_2, \ldots, x_{m_2}\}$, ..., $\{x_1, x_2, \ldots, x_{m_J}\}$, respectively, treated here as samples. If sizes $m_1, m_2, \ldots, m_J$ are proportional to the "frequency" of appearance of elements from particular classes, the considered element $\tilde{x}$ should be reckoned into the class for which the value

$$m_1 \hat{f}_1(\tilde{x}),\ m_2 \hat{f}_2(\tilde{x}), \ldots, m_J \hat{f}_J(\tilde{x}) \qquad (22)$$

is the greatest.

The above text also contains material from research carried out together with M. Charytanowicz, K. Daniel, and C. Prochot, published in the common papers [13, 14, 17].

4.2 Fault Detection System

The procedures presented in the previous section provide a complete and methodologically consistent mathematical tool to design an effective fault detection system for dynamical systems, covering detection, diagnosis, and also prognosis associated with them.

Assume that the technical state of a device under supervision may be characterized by a finite number of quantities measurable in real-time. These will be denoted in the form of the vector $x \in \mathbb{R}^n$, called a symptom vector. One can interpret this name noting that symptoms of any occurring anomalies should find the appropriate reflection in the features of a such-defined vector. More strictly, it is required that both correct functioning conditions and any type of diagnosed fault are connected with the most different sets of values and/or dissimilar relations between coordinates of the above vector as possible.

Assume also the availability of a fixed set of values of the symptom vector, representative for correct functioning conditions of a supervised device:

$$x_1, x_2, \ldots, x_{m_0} \ , \qquad (23)$$

as well as the set

$$x_1, x_2, \ldots, x_M \ , \qquad (24)$$

characteristic in the case of occurrence of anomalies. From the point of view of transparency of the designed fault detection system, in particular its function of diagnosis, it is worth dividing set (24) into $J \in \mathbb{N} \setminus \{0,1\}$ the most possibly different – in the sense of the values of particular coordinates of the symptom vector and/or relations between them – subsets assigned to the previously assumed types of diagnosed faults:

$$x_1, x_2, \ldots, x_{m_1} \qquad (25)$$

$$x_1, x_2, \ldots, x_{m_2} \qquad (26)$$

$$\vdots$$

$$x_1, x_2, \ldots, x_{m_J} \ , \qquad (27)$$

while $\sum_{j=1}^{J} m_j = M$. Where there is no such division, one can automatically divide set (24) into subsets (25)-(29) using the clustering algorithm presented in Section 4.1, although this then often requires laborious interpretation concerning each of them.

Fault detection will first be considered. With this aim the procedure for the recognition of atypical elements, described at the beginning of Section 4.1, can be applied. Assume therefore that the random sample considered there, including elements treated as typical, constitutes set (23) representing the correct functioning conditions for a supervised device, while $\tilde{x}$ denotes its current state. Applying the above mentioned procedure for the recognition of atypical elements, one can confirm if the present conditions should be regarded as typical or rather not, thus showing the appearance of anomalies.

For fault diagnosis, if one already is in possession of samples (25)-(27) characterizing particular types of faults being diagnosed, then after the above described detection of anomalies, one can – applying directly the

procedure for Bayes classification presented at the end of Section 4.1 – infer which of them is being dealt with. Note that the range of faults which can be discovered by detection may significantly exceed all types of faults assumed to be diagnosed.

Finally, if subsequent values of the symptom vector, obtained successively during the supervising process, are available, then it is possible to realize fault prognosis. It can be carried out by separate forecasts of values of the function $\hat{f}$ given by dependence (19) and $m_1\hat{f}_1$, $m_2\hat{f}_2, \ldots, m_J\hat{f}_J$ to be seen in formula (22), and inferences based on these forecasts for detection and diagnosis, according to guidelines presented in the previous two paragraphs. To calculate the values of forecasts of the functions $\hat{f}$, $\hat{f}_1$, $\hat{f}_1, \ldots, \hat{f}_J$ it is recommended to use the classical linear regression method separately, though in a version enabling easy updating of a model during successive collection of subsequent current values of the symptom vector. Appropriate formulas are found in the book [1 – Chapter 3 and additionally Chapter 4].

The proper operation of the fault detection system investigated in this section was verified experimentally for a robust control applied to the task from a field of robotics [6]. Thus, in cases where the symptoms appeared abruptly, the anomalies of the device were promptly discovered and correctly recognized within the scope of detection and diagnosis. If, on the other hand, the fault was accompanied by a slow progression of symptoms, it was forecast with a correct indication of the type of fault about to occur (scope of prognosis), and later it was also discovered and identified in detection and diagnosis. One should underline that fault prognosis, still rare in practical applications, proved to be highly effective in the case of slowly progressing symptoms, discovering and identifying anomalies before the object's characteristics transgressed the range for correct conditions for a system's functioning, thanks to the proper recognition of the change in the trend of values of the symptom vector, which indicates an unfortunate direction of its evolution.

References:

[1] Abraham B., Ledolter J., *Statistical Methods for Forecasting*, Wiley, 1983.

[2] Barnett V., Lewis T., *Outliers in Statistical Data*, Wiley, 1994.

[3] Berger J.O., *Statistical Decision Theory*, Springer-Verlag, 1980.

[4] Fukunaga K., Hostetler L.D., The estimation of the gradient of a density function, with applications in Pattern Recognition, *IEEE Transactions on Information Theory*, Vol. 21, 1975, pp. 32-40.

[5] Hand J.H., Mannila H., Smyth P., *Principles of Data Mining*, MIT Press, 2001.

[6] Kulczycki P., Fuzzy Controller for Mechanical Systems, *IEEE Transactions on Fuzzy Systems*, Vol. 8, 2000, pp. 645-652.

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

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

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

[10] Kulczycki P., Statistical Kernel Estimators for Design of a Fault Detection, Diagnosis and Prognosis System, *The Open Cybernetics and Systemics Journal*, Vol. 2, 2008, pp. 180-184, http://www.bentham.org/open/tocsj/openaccess2.htm.

[11] Kulczycki P., Charytanowicz P., Asymmetrical Conditional Bayes Parameter Identification for Control Engineering, *Cybernetics and Systems*, Vol. 38, 2008, pp. 229-243.

[12] Kulczycki P., Applicational Possibilities of Nonparametric Estimation of Distribution Density for Control Engineering, in press.

[13] Kulczycki P., Charytanowicz P., A Complete Gradient Clustering Algorithm Formed with Kernel Estimators, in press.

[14] Kulczycki P., Daniel K., Algorytm wspomagania strategii marketingowej operatora telefonii komorkowej, in: *Badania operacyjne i systemowe 2006: metody i techniki*, Kacprzyk J., Budzinski R. (eds.), EXIT, 2006, pp. 245-256.

[15] Kulczycki P., Dawidowicz A.L., Kernel estimator of quantile, *Universitatis Iagellonicae Acta Mathematica*, Vol. XXXVII, 1999, pp. 325-336.

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

[17] Kulczycki P., Prochot C., Wykrywanie elementow odosobnionych za pomoca metod estymacji nieparametrycznej, in: *Badania operacyjne i systemowe: podejmowanie decyzji – podstawy teoretyczne i zastosowania*, Kulikowski R., Kacprzyk J., Slowinski R. (eds.), EXIT, 2004, pp. 313-328.

[18] Pal S.K., Mitra P., *Pattern Recognition Algorithms for Data Mining*, Chapman and Hall, 2004.

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

[20] Wand M.P., Jones M.C., *Kernel Smoothing*. Chapman and Hall, 1994.