Conditional Parameter Identification with Asymmetrical Losses of Estimation Errors

Piotr Kulczycki[1,2] and Malgorzata Charytanowicz[1,3]

[1] Polish Academy of Sciences, Systems Research Institute,
Centre of Information Technology for Data Analysis Methods, Warsaw, Poland
[2] Cracow University of Technology,
Department of Automatic Control and Information Technology, Cracow, Poland
[3] Catholic University of Lublin, Institute of Mathematics and Computer Science,
Lublin, Poland
{Piotr.Kulczycki,Malgorzata.Charytanowicz}@ibspan.waw.pl

Abstract. In many scientific and practical tasks, the classical concepts for parameter identification are satisfactory and generally applied with success, although many specialized problems necessitate the use of methods created with specifically defined assumptions and conditions. This paper investigates the method of parameter identification for the case where losses resulting from estimation errors can be described in polynomial form with additional asymmetry representing different results of under- and overestimation. Most importantly, the method presented here considers the conditionality of this parameter, which in practice means its significant dependence on other quantities whose values can be obtained metrologically. To solve a problem in this form the Bayes approach was used, allowing a minimum expected value of losses to be achieved. The methodology was based on the nonparametric technique of statistical kernel estimators, which freed the worked out procedure from forms of probability distributions characterizing both the parameter under investigation and conditioning quantities. As a result, a ready to direct use algorithm has been presented here.

Keywords: parameter identification, Bayes approach, asymmetrical losses of estimation errors, conditional factors, nonparametric estimation, statistical kernel estimators, numerical algorithm.

1 Introduction

Parametric identification, i.e. assigning a concrete value to a parameter present in a model – despite its very traditional nature – has still great significance in modern scientific and applicational problems, which continuously increases together with the dominance of model-based methods and the growing, often specific, demands made on models used in practice. Fortunately, the development of modern advanced methods of parameter identification is facilitated by the dynamic expansion of contemporary computer technology, supported on the theoretical side by the procedures of information technology dedicated to them.

N.-T. Nguyen et al. (Eds.): ICCCI 2012, Part I, LNAI 7653, pp. 553–562, 2012.

The subject of this paper is an algorithm for parametric identification based on four premises:

1. minimization of expected value of losses arising from estimation errors, unavoidable in practice;
2. asymmetry of those losses, i.e. allowing for situations where losses occurring through underestimation are substantially different from losses resulting from overestimation;
3. arbitrariness of probability distributions appearing in the problem; and finally, worth particularly highlighting
4. conditionality of an identified parameter, that is its significant dependence on a factor (or factors), with values that can be in practice obtained metrologically.

The realization of the first will be through application of the Bayes approach [1].

The second by assuming the loss function resulting from estimation errors, in the asymmetrical form

$$l(\hat{y}, y) = \begin{cases} (-1)^k a_l (\hat{y} - y)^k & \text{for} \quad \hat{y} - y \leq 0 \\ a_r (\hat{y} - y)^k & \text{for} \quad \hat{y} - y \geq 0 \end{cases} , \tag{1}$$

with the given degree $k \in \mathbb{N} \setminus \{0\}$, where the coefficients a_l and a_r are positive, while y and $\hat{y}$ denote the values of the parameter under consideration and its estimator, respectively. The fact that the coefficients a_l and a_r may differ causes an asymmetry of the above function and enables the inclusion of different losses implied by over- and underestimation of the examined parameter. Limiting the form of function (1) to a polynomial seems not to decrease the generality of considerations in practical applications, offering an effective compromise between precision and complexity of results obtained. Moreover the possibility of change of the polynomial degree k allows a differing scale of protection against large estimation errors.

The third aspect is realized by applying nonparametric methodology of statistical kernel estimators [3, 13, 14] for defining probability characteristics.

Lastly – and worth highlighting once more – this paper is aimed at the conditional approach, i.e. where the value of the estimated parameter is strongly dependent on a conditional factor, for example in engineering practice it is often temperature. If the value of such a factor is metrologically available, then its inclusion can make the used model significantly more precise.

The preliminary version of this paper was presented as the publication [7]. More details is available in the paper [8], which will appear soon.

2 Preliminaries: Statistical Kernel Estimators

Let the *n*-dimensional random variable X be given, with a distribution characterized by the density f. Its kernel estimator $\hat{f} : \mathbb{R}^n \to [0, \infty)$, calculated using experimentally obtained values for the *m*-element random sample

$$x_1, \; x_2, \ldots, x_m \quad , \tag{2}$$

in its basic form is defined as

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

where $m \in \mathbb{N} \setminus \{0\}$, the coefficient $h > 0$ is called a smoothing parameter, while the measurable function $K : \mathbb{R}^n \to [0, \infty)$ of unit integral $\int_{\mathbb{R}^n} K(x) \mathrm{d}x = 1$, symmetrical with respect to zero and having a weak global maximum in this place, takes the name of a kernel. The choice of form of the kernel K and the calculation of the smoothing parameter h is made most often with the criterion of the mean integrated square error.

Thus, the choice of the kernel form has – from a statistical point of view – no practical meaning and thanks to this, it becomes possible to take into account primarily properties of the estimator obtained or calculational aspects, advantageous from the point of view of the applicational problem under investigation; for broader discussion see the books [3 – Section 3.1.3; 14 – Sections 2.7 and 4.5]. In practice, for the one-dimensional case (i.e. when $n = 1$), the function K is assumed most often to be the density of a common probability distribution. 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

$$K(x) = K\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) = \mathcal{K}_1(x_1)\, \mathcal{K}_2(x_2) \ldots \mathcal{K}_n(x_n) \quad , \tag{4}$$

where $\mathcal{K}_i$ for $i = 1, 2, \ldots, n$ denotes the previously-mentioned one-dimensional kernels, while the expression h^n appearing in the basic formula (3) should be replaced by $h_1 \cdot h_2 \cdot \ldots \cdot h_n$, the product of the smoothing parameters for particular coordinates.

The fixing of the smoothing parameter h has significant meaning for quality of estimation. Fortunately – from the applicational point of view – many suitable procedures for calculating the value of the parameter h on the basis of random sample (2) have been worked out. For broader discussion of the above tasks see the monographs [3, 13, 14]. In particular, for the one-dimensional case, the simple and effective plug-in method [3 – Section 3.1.5; 14 – Section 3.6.1] is especially recommended. Of course this method can also be applied in the *n*-dimensional case when product kernel (4) is used, sequentially n times for each coordinate.

Practical applications may also use additional procedures generally improving the quality of estimator (3). For the method presented in this paper, the modification of the smoothing parameter [3 – Section 3.1.6; 13 – Section 5.3.1] is strongly recommended.

The above concept will now be generalized for the conditional case. Here, besides the basic (sometimes termed the describing) n_Y -dimensional random variable Y, let

also be given the n_W-dimensional random variable W, called hereinafter the conditioning random variable. Their composition $X = \begin{bmatrix} Y \\ W \end{bmatrix}$ is a random variable of dimension $n_Y + n_W$. Assume that distributions of the variables X and, in consequence, W have densities, denoted below as $f_X : \mathbb{R}^{n_Y+n_W} \to [0,\infty)$ and $f_W : \mathbb{R}^{n_W} \to [0,\infty)$, respectively. Let also be given the so-called conditioning value, that is the fixed value of conditioning random variable $w^* \in \mathbb{R}^{n_W}$, such that

$$f_W(w^*) > 0 \quad . \tag{5}$$

Then the function $f_{Y|W=w^*} : \mathbb{R}^{n_Y} \to [0,\infty)$ given by

$$f_{Y|W=w^*}(y) = \frac{f_X(y,w^*)}{f_W(w^*)} \qquad \text{for every} \quad y \in \mathbb{R}^{n_Y} \tag{6}$$

constitutes a conditional density of probability distribution of the random variable Y for the conditioning value w^*. The conditional density $f_{Y|W=w^*}$ can so be treated as a "classic" density, whose form has been made more accurate in practical applications with w^* – a concrete value taken by the conditioning variable W in a given situation.

Let therefore the random sample

$$\begin{bmatrix} y_1 \\ w_1 \end{bmatrix}, \begin{bmatrix} y_2 \\ w_2 \end{bmatrix}, \ldots, \begin{bmatrix} y_m \\ w_m \end{bmatrix}, \tag{7}$$

obtained from the variable $X = \begin{bmatrix} Y \\ W \end{bmatrix}$, be given. The particular elements of this sample are interpreted as the values y_i taken in measurements from the random variable Y, when the conditioning variable W assumes the respective values w_i. Using the methodology presented in the first part of the section below, on the basis of sample (7) one can calculate $\hat{f}_X$, i.e. the kernel estimator of density of the random variable X probability distribution, while the sample

$$w_1, w_2, \ldots, w_m \tag{8}$$

gives $\hat{f}_W$ – the kernel density estimator for the conditioning variable W. The kernel estimator of conditional density of the random variable Y probability distribution for the conditioning value w^*, is defined then – a natural consequence of formula (6) – as the function $\hat{f}_{Y|W=w^*} : \mathbb{R}^{n_Y} \to [0,\infty)$ given by

$$\hat{f}_{Y|W=w^*}(y) = \frac{\hat{f}_X(y,w^*)}{\hat{f}_W(w^*)} \quad . \tag{9}$$

If for the estimator $\hat{f}_W$ one uses a kernel with positive values, then the inequality $\hat{f}_W(w^*) > 0$ implied by condition (5) is fulfilled for any $w^* \in \mathbb{R}^{n_W}$.

In the case when for the estimators $\hat{f}_X$ and $\hat{f}_W$ the product kernel (4) is used, applying in pairs the same positive kernels to the estimator $\hat{f}_X$ for coordinates which correspond to the vector W and to the estimator $\hat{f}_W$, then the expression for the kernel estimator of conditional density becomes particularly helpful for practical applications. Formula (9) can then be specified to the form

$$\hat{f}_{Y|W=w^*}(y) = \hat{f}_{Y|W=w^*}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_Y} \end{bmatrix}\right) = \tag{10}$$

$$= \frac{\frac{1}{h_1 h_2 \ldots h_{n_Y}} \sum_{i=1}^{m} \mathcal{K}_1\left(\frac{y_1 - y_{i,1}}{h_1}\right) \mathcal{K}_2\left(\frac{y_2 - y_{i,2}}{h_2}\right) \ldots \mathcal{K}_{n_Y}\left(\frac{y_{n_Y} - y_{i,n_Y}}{h_{n_Y}}\right) \mathcal{K}_{n_Y+1}\left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}}\right) \mathcal{K}_{n_Y+2}\left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}}\right) \ldots \mathcal{K}_{n_Y+n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}}\right)}{\sum_{i=1}^{m} \mathcal{K}_{n_Y+1}\left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}}\right) \mathcal{K}_{n_Y+2}\left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}}\right) \ldots \mathcal{K}_{n_Y+n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}}\right)},$$

where h_1, $h_2, \ldots, h_{n_Y+n_W}$ represent – respectively – smoothing parameters mapped to particular coordinates of the random variable X, while the coordinates of the vectors w^*, x_i and w_i are denoted as

$$w^* = \begin{bmatrix} w_1^* \\ w_2^* \\ \vdots \\ w_{n_W}^* \end{bmatrix} \quad \text{and} \quad y_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,n_Y} \end{bmatrix}, \quad w_i = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,n_W} \end{bmatrix} \quad \text{for} \quad i = 1, 2, \ldots, m \quad . \tag{11}$$

Define the so-called conditioning parameters d_i for $i = 1, 2, \ldots, m$ by the formula

$$d_i = \mathcal{K}_{n_Y+1}\left(\frac{w_1^* - w_{i,1}}{h_{n_Y+1}}\right) \mathcal{K}_{n_Y+2}\left(\frac{w_2^* - w_{i,2}}{h_{n_Y+2}}\right) \ldots \mathcal{K}_{n_Y+n_W}\left(\frac{w_{n_W}^* - w_{i,n_W}}{h_{n_Y+n_W}}\right) \quad . \tag{12}$$

Thanks to the assumption of positive values for the kernels $\mathcal{K}_{n_Y+1}$, $\mathcal{K}_{n_Y+2}, \ldots, \mathcal{K}_{n_Y+n_W}$, these parameters are also positive. So the kernel estimator of conditional density (10) can be presented in the form

$$\hat{f}_{Y|W=w^*}(y) = \hat{f}_{Y|W=w^*}\left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_Y} \end{bmatrix}\right) = \frac{1}{h_1 h_2 \ldots h_{n_Y} \sum_{i=1}^{m} d_i} \sum_{i=1}^{m} d_i \mathcal{K}_1\left(\frac{y_1 - y_{i,1}}{h_1}\right) \mathcal{K}_2\left(\frac{y_2 - y_{i,2}}{h_2}\right) \ldots \mathcal{K}_{n_Y}\left(\frac{y_{n_Y} - y_{i,n_Y}}{h_{n_Y}}\right). \tag{13}$$

The value of the parameter d_i characterizes the "distance" of the given conditioning value w^* from w_i – that of the conditioning variable for which the *i*-th element of

the random sample was obtained. Then estimator (13) can be interpreted as the linear combination of kernels mapped to particular elements of a random sample obtained for the variable Y, when the coefficients of this combination characterize how representative these elements are for the given value w^*.

Returning to the subject of this article, described in the Introduction, in the case of estimation of a single parameter, the random variable Y is one-dimensional ($n_Y = 1$). This will be investigated further in the presented paper. However when one estimates a number of conditionally correlated parameters, then n_Y becomes equal to their number – this case will be commented upon at the end of Section 4.

More details concerning kernel estimators can be found in the books [3, 13, 14]. Exemplary applications are presented in the publications [4-6, 9, 10, 12].

3 Main Results

3.1 Linear Case

Let the parameter under investigation, whose value is to be estimated, denoted by $y \in \mathbb{R}$, be treated as the value of the random variable Y. Let also the n_W-dimensional conditional random variable W be given. The availability is assumed of the metrologically achieved measurements of the parameter y, i.e. y_1, $y_2, \ldots, y_m$, obtained for the values w_1, $w_2, \ldots, w_m$ of the conditional variable, respectively. Finally, let $w^* \in \mathbb{R}^{n_W}$ denote any fixed conditioning value. The goal is to calculate the estimator of this parameter, denoted by $\hat{y}_{w^*}$, optimal in the sense of minimum expected value of losses arising from errors of estimation, for conditioning value w^*. The case considered in this subsection is such that loss function (1) can be specified to the following asymmetrical linear form:

$$l(\hat{y}_{w^*}, y) = \begin{cases} -a_l\,(\hat{y}_{w^*} - y) & \text{for} \quad \hat{y}_{w^*} - y \le 0 \\ a_{\mathrm{r}}\,(\hat{y}_{w^*} - y) & \text{for} \quad \hat{y}_{w^*} - y \ge 0 \end{cases}, \tag{14}$$

while the coefficients a_l and a_r are positive and not necessarily equal to each other.

In order to solve such a task, the Bayes decision rule will be used [1]. The minimum expected value of losses arising from estimation errors occurs when the value is a solution of the following equation with the argument $\hat{y}_{w^*}$:

$$\int_{-\infty}^{\hat{y}_{w^*}} f_{Y|W=w^*}(y)\,\mathrm{d}\,y - \frac{a_l}{a_l + a_r} = 0 \quad , \tag{15}$$

where $f_{Y|W=w^*}$ denotes the density of distribution of the random variable Y representing the uncertainty of the parameter in question, for conditioning value w^*. Since $0 < a_l/(a_l + a_r) < 1$, a solution for the above equation exists, and if the function $f_{Y|W=w^*}$ has connected support, this solution is unique. Moreover, thanks to

equality $\frac{a_l}{a_l + a_r} = \frac{a_l / a_r}{a_l / a_r + 1}$, it is not necessary to identify the parameters a_l and a_r separately, rather only their ratio.

The identification of the density $f_{Y|W=w^*}$ will be carried out using statistical kernel estimators, presented in Section 2, with the – convenient here – form (13). Then as $\mathscr{K}_1$ (note that $n_Y = 1$) one should choose a continuous kernel of positive values, and also so that the function $\mathscr{J} : \mathbb{R} \to \mathbb{R}$ such that $\mathscr{J}(x) = \int_{-\infty}^{x} \mathscr{K}_1(y)\,\mathrm{d}\,y$ can be expressed by a relatively simple analytical formula. In consequence, this results in a similar property regarding the function $\mathscr{U}_i : \mathbb{R} \to \mathbb{R}$ for any fixed $i = 1, 2, \ldots, m$ defined as

$$\mathscr{U}_i(\hat{y}_{w^*}) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_{w^*}} \mathscr{K}_1\left(\frac{y - x_i}{h_1}\right) \mathrm{d}\,y \quad . \tag{16}$$

Criterion (15) can be expressed then equivalently in the form of

$$\sum_{i=1}^{m} d_i \mathscr{U}_i(\hat{y}_{w^*}) - \frac{a_l}{(a_l + a_r)} \sum_{i=1}^{m} d_i = 0 \quad . \tag{17}$$

If the left side of the above equation is denoted by $L(\hat{y}_{w^*})$, then $\lim_{\hat{y}_{w^*} \to -\infty} L(\hat{y}_{w^*}) < 0$, $\lim_{\hat{y}_{w^*} \to \infty} L(\hat{y}_{w^*}) > 0$, the function L is (strictly) increasing and its derivative is simply:

$$L'(\hat{y}_{w^*}) = \sum_{i=1}^{m} d_i \mathscr{K}_1\left(\frac{\hat{y}_{w^*} - x_i}{h_1}\right) \quad . \tag{18}$$

In this situation, the solution of criterion (15) can be effectively calculated on the basis of Newton's algorithm [2] as the limit of the sequence $\{\hat{y}_{w^*,j}\}_{j=0}^{\infty}$ defined by

$$\hat{y}_{w^*,0} = \frac{\sum_{i=1}^{m} d_i\, y_i}{\sum_{i=1}^{m} d_i} \tag{19}$$

$$\hat{y}_{w^*,j+1} = \hat{y}_{w^*,j} - \frac{L(\hat{y}_{w^*,j})}{L'(\hat{y}_{w^*,j})} \qquad \text{for} \quad j = 0, 1, \ldots \quad , \tag{20}$$

with the functions L and L' being given by dependencies (17)-(18), whereas a stop criterion takes on the form

$$| \hat{y}_{w^*,j} - \hat{y}_{w^*,j-1} | \le 0.01\, \hat{\sigma}_Y \quad , \tag{21}$$

while $\hat{\sigma}_Y$ denotes the estimator of the standard deviation of the random variable Y.

3.2 Quadratic Case

The conditionings of the problem investigated in this subsection are similar to the previous one, although asymmetric linear form of the loss function (14) is substituted by the asymmetric quadratic:

$$l(\hat{y}_{w^*}, y) = \begin{cases} a_l\,(\hat{y}_{w^*} - y)^2 & \text{for} \quad \hat{y}_{w^*} - y \le 0 \\ a_{\mathrm{r}}\,(\hat{y}_{w^*} - y)^2 & \text{for} \quad \hat{y}_{w^*} - y \ge 0 \end{cases} , \tag{22}$$

while the coefficients a_l and a_r are positive and not necessarily equal to each other. The minimum expected value of losses arising from estimation errors can in this case be calculated for the value $\hat{y}_{w^*}$ being a solution of the equation

$$(a_l - a_r) \int_{-\infty}^{\hat{y}_{w^*}} (\hat{y}_{w^*} - y) f_{Y|W=w^*}(y)\,\mathrm{d}\,y - a_l \int_{-\infty}^{\infty} (\hat{y}_{w^*} - y) f_{Y|W=w^*}(y)\,\mathrm{d}\,y = 0 \quad . \tag{23}$$

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

Using kernel estimators in form (13) to identify the density $f_{Y|W=w^*}$, one can design an effective numerical algorithm to this end. Let, therefore, a continuous kernel $\mathscr{K}_1$ of positive values, fulfilling the condition

$$\int_{-\infty}^{\infty} y \mathscr{K}_1(y)\,\mathrm{d}\,y < \infty \tag{24}$$

be given. Besides the functions $\mathscr{U}_i$ introduced by the dependence (16), let for any fixed $i = 1, 2, \ldots, m$ the functions $\mathscr{V}_i : \mathbb{R} \to \mathbb{R}$ be defined as

$$\mathscr{V}_i(\hat{y}_{w^*}) = \frac{1}{h_1} \int_{-\infty}^{\hat{y}_{w^*}} y \mathscr{K}_1\left(\frac{y - y_i}{h_1}\right) \mathrm{d}\,y \quad . \tag{25}$$

The kernel $\mathscr{K}_1$ should be chosen so that – apart from the requirements formulated above – the function $\mathscr{J} : \mathbb{R} \to \mathbb{R}$ such that $\mathscr{J}(x) = \int_{-\infty}^{x} y \mathscr{K}_1(y)\,\mathrm{d}\,y$ be expressed by a convenient analytical formula.

Criterion (23) can then be described equivalently as

$$\sum_{i=1}^{m} d_i [(a_l - a_r)(\hat{y}_{w^*}\,\mathscr{U}_i(\hat{y}_{w^*}) - \mathscr{V}_i(\hat{y}_{w^*})) + a_l\,y_i] - m a_l\,\hat{y}_{w^*} \sum_{i=1}^{m} d_i = 0 \quad . \tag{26}$$

If the left side of the above formula is denoted by $L(\hat{y}_{w^*})$, then one can express the value of its derivative as

$$L'(\hat{y}_{w^*}) = \sum_{i=1}^{m} d_i[(a_l - a_r)\mathcal{U}_i(\hat{y}_{w^*})] - ma_l \sum_{i=1}^{m} d_i \quad . \tag{27}$$

In this situation, the solution of criterion (23) can be calculated numerically on the basis of Newton's algorithm (19)-(21) with the functions L and L' defined by dependencies (26)-(27).

4 Final Remarks and Conclusions

The parameter identification for problems where losses arising from overestimation and underestimation are different and can be described by an asymmetrical and polynomial function and – what is worth particularly highlighting – in the presence conditioning quantities, was investigated in this paper. The linear and quadratic cases were described with details in Section 3. The similar results can be obtained for polynomials of higher degrees – they are presented in the paper [8], which will appear soon; in this publication the concrete formulas for the recommended kernels are also provided.

The functioning and features of the algorithm presented here were positively confirmed with detailed numerical and experimental verification, also for a multidimensional conditioning variable and multimodal, asymmetrical and complex distributions of the variables Y and W, as well as those including additional aspects, e.g. bounded supports, lack of data from the neighborhood of a given conditioning variable x^*, as well as the occurrence of discrete, binary and categorized coordinates of the conditioning variable W. For a broad description of the numerical verification results see the paper [8].

The concept presented in this publication was also verified experimentally by applying it to a task of identification of dynamic systems submitted to robust control, and also in medical applications, in establishing optimal dosages of anesthetic considering patients' body mass and general condition, as well as strategic sales in selecting policy for a mobile phone operator when negotiating with a business client characterized by many vastly different factors. Generally it is worth stressing that in every case investigated, 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 circumstantial changes in values for these factors, as well as structural object nonstationarity.

Finally, it is worth adding that the concept developed here can be generalized to a multidimensional case, i.e. where the vector of conditionally correlated parameters is identified. However, in this case, both the analytical criteria for optimal parameter values as well as their later numerical implementation, become too complicated for practical application given today's. Similarly it is possible to assume loss function (1) in an asymmetrical form of different degree of polynomial for negative and positive estimation errors. However such a case seems to have only theoretical significance, with no applicational connotations.

Acknowledgments . Our heartfelt thanks go to our colleague Dr. Aleksander Mazgaj, with whom we commenced the research presented here. With his consent, this text also contains results of joint research [11].

References

1. Berger, J.O.: Statistical Decision Theory. Springer, New York (1980)
2. Kincaid, D., Cheney, W.: Numerical Analysis. Brooks/Cole, Pacific Grove (2002)
3. Kulczycki, P.: Estymatory jadrowe w analizie systemowej. WNT, Warsaw (2005)
4. Kulczycki, P.: Estymatory jadrowe w badaniach systemowych. In: Kulczycki, P., Hryniewicz, O., Kacprzyk, J. (eds.) Techniki informacyjne w badaniach systemowych, pp. 79–105. WNT, Warsaw (2007)
5. Kulczycki, P.: Kernel Estimators in Industrial Applications. In: Prasad, B. (ed.) Soft Computing Applications in Industry, pp. 69–91. Springer, Berlin (2008)
6. Kulczycki, P., Charytanowicz, M., Kowalski, P.A., Łukasik, S.: The Com-plete Gradient Clustering Algorithm: Properties in Practical Applications. Journal of Applied Statistics 39, 1211–1224 (2012)
7. Kulczycki, P., Charytanowicz, M.: Warunkowa bayesowska identyfikacja parametryczna z niesymetrycznymi stratami. In: Proc. XVII Krajowa Konferencja Automatyki, Kielce, Poland, June 19-22, pp. 215–226 (2011); post-conference publication in Malinowski, K., Dindorf, R. (eds.) Postepy automatyki i robotyki, vol. 1, pp. 107–122. WPS, Kielce (2011)
8. Kulczycki, P., Charytanowicz, M.: Conditional Parameter Identification with Different Losses of Under- and Overestimation. Applied Mathematical Modelling (in press, 2012)
9. Kulczycki, P., Daniel, K.: Metoda wspomagania strategii marketingowej operatora telefonii komorkowej. Przeglad Statystyczny 56, 116–134 (2009)
10. Kulczycki, P., Kowalski, P.A.: Bayes classification of imprecise informa-tion of interval type. Control and Cybernetics 40, 101–123 (2011)
11. Kulczycki, P., Mazgaj, A.: Parameter Identification for Asymmetrical Polynomial Loss Function. Information Technology and Control 38, 51–60 (2009); Errata: vol. 38, pp. 167–168
12. Kulczycki, P., Waglowski, J.: On the application of statistical kernel estimators for the demand-based design of a wireless data transmission system. Control and Cybernetics 34, 1149–1167 (2005)
13. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)
14. Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman and Hall, London (1995)