Journal of Universal Computer Science, vol. 13, no. 2 (2007), 199-208
submitted: 31/7/06, accepted: 15/1/07, appeared: 28/2/07

Parameter Estimation of Systems Described by the Relation with Noisy Observations

Jerzy Świątek
(Institute of Information Science and Engineering
Wroclaw University of Technology, Poland
Jerzy.Swiatek@pwr.wroc.pl)

Abstract: In this paper the problem of parameter estimation of an input – output system is discussed. It is assumed that the system is described by the relation known with accuracy to some parameters. The possible noisy observations of system are described. The estimation algorithm based on maximum likelihood method is proposed. The presented approach is illustrated by analytical examples.

Keywords: System identification, parameter estimation, input – output system, relational systems, knowledge representation
Categories: H.2.1, I.2.4, I.2.11

1 Introduction

Investigation of computer systems for decision processes based on knowledge representation requires new methods of system modeling. The traditional mathematical model given by system of equations was very convenient for analytical investigations. The application of computer aided methods for processing observations, more generally the knowledge about investigated plant allows to consider a wide class of models. Particularly, the input – output system may be described by the set of facts given by logical statements about the input and output of this system. Sometimes such a description is given by an expert. The problem is to generalize the expert observation and propose the system description in form of relation defined on the set of input and output. In this case we can formulate the identification problem for the system described by the relation, similar to the identification problems of systems described by the equations [Bubnicki, 80a].

The problem of modeling and identification of systems described by the relation has been presented in previous works. Particularly in [Bubnicki, 80b] the general problem identification of relational system is presented. Application of relational system to knowledge representation is given in [Bubnicki, 90] and to control and identification in [Bubnicki, 88]. In [Swiatek, 89] the problem of optimal model choice is discussed. Some estimation problem is presented in [Swiatek, 90]. The application of maximum likelihood method is proposed in [Swiatek, 06]. Now for the noisy observations the estimation algorithm is proposed.

2 System Descriptions

Let us consider the input – output static system with input x and output y. Input and output are S and L – dimensional vectors, respectively. Input and output are elements of sets X and Y, which are subsets of R^S and R^L spaces, respectively, i.e.:

$$x \in X \subseteq R^S, \quad y \in Y \subseteq R^L .$$

The system is described by the set of facts concerning inputs and outputs. More precisely, the set of true logical statements about x and y is given. Consequently the logical function

$$F(x, y, a) \tag{1}$$

defined on the set of inputs and outputs is proposed, where F is a complex logical function and a is a K – dimensional vector of parameters from set of parameters A (i.e.: $a \in A \subseteq R^K$). In the system description only such (x, y) from $X \times Y$ may appear, for which the statement (1) is true. In this way the description of the system is given by the relation defined on $X \times Y$ i.e.:

$$\Re_a = \{(x, y) \in X \times Y : \quad F(x, y, a)\}. \tag{2}$$

On the relation $\Re_a$ the probability density function

$$g(x, y, a) \tag{3}$$

is defined. For example, let sets of inputs and outputs are real numbers and the facts concerning inputs and outputs are the following: input and output are positive numbers and the sum of input and output is not grater than a. Furthermore the probability density is monotonous. For the above system (2) and (3) have the forms:

$$\Re_a = \{(x, y) \in R^2 : \quad x \geq 0 \wedge y \geq 0 \wedge x + y \leq a\}, \tag{4}$$

$$g(x, y, a) = \begin{cases} 2a^{-2} & if \quad x \geq 0 \wedge y \geq 0 \wedge x + y \leq a \\ 0 & \qquad otherwise \end{cases}. \tag{5}$$

3 System Observations

Now it is assumed that the description of the system is known with accuracy to parameters, i.e. the functions F and g in (1)-(3) are known but the vector of parameters a must be estimated. To determine vector a, the noised observations are collected. More precisely, observed input x and output y are noised by z_x and z_y, $(z_x \in Z_x, z_y \in Z_y)$, respectively. The influences of random noise on the observed values are described by the following equations:

$$u = h_x(x, z_x), \tag{6}$$

$$v = h_y(y, z_y), \tag{7}$$

where: u and v are effects of input x and output y noisy observations, respectively $(u \in U, v \in V)$; h_x, h_y are known functions $h_x : X \times Z_x \to U$, $h_y : Y \times Z_y \to V$. Practically it means that collected experimental data are elements of the relation $\overline{\Re}_a$, obtained from relation $\Re_a$ (2) transformed by functions (6) and (7), i.e.:

$$\overline{\Re}_a = \left\{ \begin{array}{r} (u, v) \in U \times V : u = h_x(x, z_x), v = h_y(y, z_y), \\ \forall (x, y) \in \Re_a,\ z_x \in Z_x,\ z_y \in Z_y \end{array} \right\}. \tag{8}$$

It is assumed that functions h_x and h_y are one-to-one mapping with respect to z_x and z_y. It means that for given x and y there exist inverse functions with respect to z_x and z_y, i.e.:

$$z_x = h_x^{-1}(x, u), \tag{9}$$

$$z_y = h_y^{-1}(y, v). \tag{10}$$

In the sequence of observations it is assumed that noise z_x and z_y are independent realizations of random variables $\underline{z}_x$ and $\underline{z}_y$ with known probability density function

$$g_z(z_x, z_y). \tag{11}$$

It is possible to collect two different types of observations (Fig.1). The first (A) is the sequence of measured values of input and output. The second one (B) is the sequence of true statements on input and output.

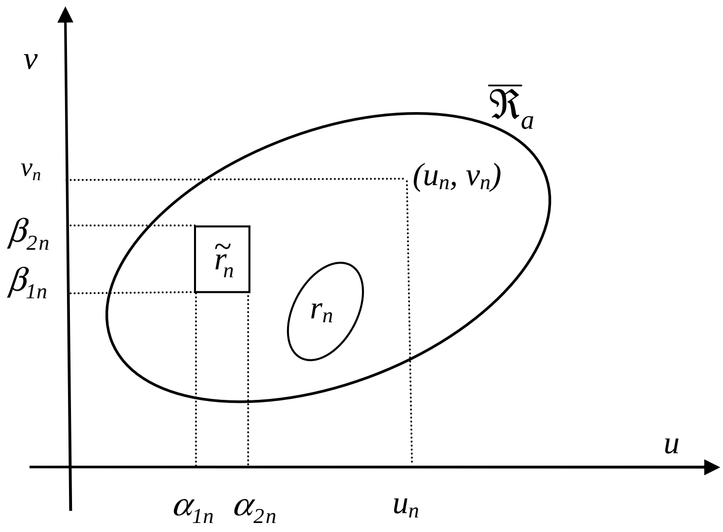

Figure 1: Relational system observations

- A – The sequence of input and output noisy measurements are collected, i.e.:

$$(u_n, v_n), \quad n = 1, 2, \ldots, N \,, \tag{12}$$

where: u_n, v_n are n-th measurements of input and output, respectively, N is the number of measurements.

- B – The sequence of true logical statements about input and output is given, i.e.:

$$r_n = \{(u,v) \in U \times V : \quad f_n(u,v)\}, \quad n = 1, 2, \ldots, N$$

(13)

where: f_n is n-th logical statement about input and output. Such a fact may be given by an expert. For example from the expert we know that for input x from the interval $u \in [\alpha_{1n}, \alpha_{2n}]$ the output y is in the interval $v \in [\beta_{1n}, \beta_{2n}]$ what will be denoted as observation:

$$\tilde{r}_n = \left\{(u,v) \in R^2 : \quad \alpha_{1n} \le u \le \alpha_{2n} \wedge \beta_{1n} \le v \le \beta_{2n}\right\}.$$

(14)

4 Parameter Estimation by Maximum Likelihood Method

Let us assume that the observations (*u*,*v*) from the relation $\overline{\Re}_a$ (8), corresponds to the (*x*,*y*), for which the statement *F* (1) is true, i.e.: $(x,y)\in\Re_a$ (2). For the given (*x*,*y*), because of the random measurement noises (z_x,z_y) the observed (*u*,*v*) is realization of conditional random variable $(\underline{u},\underline{v}/x,y)$. Notice that for the given (*x*, *y*) conditional random variable $(\underline{u},\underline{v}/x,y)$ can be obtained as the transformation of random variable ($\underline{z_x}$, $\underline{z_y}$) by (6) and (7). Consequently, conditional probability density function of random variable $(\underline{u},\underline{v}/x,y)$ is defined by probability density function g_z (11) and transformations h_x and h_y ((6) and (7)), i.e.:

$$g_1(u,v/x,y)=g_z\left(h_x^{-1}(x,u),h_y^{-1}(y,v)\right)\times\left|J_x\right|\times\left|J_y\right| \tag{15}$$

where J_x and J_y are Jacob's matrix of transformations (9) and (10). Practically we do not know which pair (*x*,*y*) is just observed. It can be any pair from the relation $\Re_a$ i.e.: such (*x*,*y*) for which statement (1) is true. So, the probability density function defined on the relation $\overline{\Re}_a$ is given by the formula:

$$\overline{g}(u,v,a)=\int_{\Re_a} g_1(u,v/x,y)\times g(x,y)dxdy \tag{16}$$

Now, let us come back to the collected data of the form (12). They are independent realization of random variables $(\underline{x},\underline{y})$ with probability density function (16). Consequently the likelihood function has the form:

$$W_A(a,U_N,V_N)=\prod_{n=1}^{N}\overline{g}(u_n,v_n,a), \tag{17}$$

where: $U_N=[u_1\quad u_2\quad\cdots\quad u_N]$ and $V_N=[v_1\quad v_2\quad\cdots\quad v_N]$. The estimate a_{AN} of vector *a* is obtained by maximization (17) with respect to *a*, i.e.:

$$a_{AN}=\Psi_A(U_N,V_N)\quad\rightarrow\quad W_A(a_{AN},U_N,V_N)=\min_{a\in A}W_A(a,U_N,V_N) \tag{18}$$

For the observations of the form B the true sentence of the form (13) is given. The probability that it is possible to obtain true observation r_n is determined by the following formula:

$$P_n(r_n,a)=\begin{cases}\int\limits_{r_n}\bar{g}(u,v,a)dxdy & if \quad r_n \subseteq \overline{\Re}_a \\ 0 & otherwise\end{cases}. \tag{19}$$

For further consideration it is assumed that

$$\forall n,m\in\{1,2,\dots,N\} \quad r_n\cap r_m=\varnothing \quad \vee \quad r_n\cap r_m=r_n=r_m,$$

consequently the likelihood function is:

$$W_B(a,\Gamma_N)=\begin{cases}\prod\limits_{n=1}^{N}\int\limits_{r_n}\bar{g}(u,v,a)dxdy & if \quad \forall n\in\{1,2,\dots,N\} \quad r_n\subseteq\overline{\Re}_a \\ 0 & otherwise\end{cases} \tag{20}$$

where $\Gamma_N=[r_1 \quad r_2 \quad \cdots \quad r_N]$. The estimates a_{BN} of vector *a*, for the measurements of type B, is obtained by maximization of the likelihood function (20) with respect to *a*, i.e.:

$$a_{BN}=\Psi_B(\Gamma_N) \quad \to \quad W_B(a_{BN},\Gamma_N)=\min_{a\in A}W_B(a,\Gamma_N). \tag{21}$$

Example: Let *x* and *y* are real numbers ($L = S = 1$). The system is described by the relation:

$$\Re_a=\{(x,y)\in R^2 : a^{(1)}-0.5\le x\le a^{(1)}+0.5\wedge a^{(2)}-0.5\le y\le a^{(2)}+0.5\} \tag{22}$$

and the probability density distribution (3) has the form:

$$g(x,y,a)=\begin{cases}1 & if \quad (x,y)\in\Re_a \\ 0 & otherwise\end{cases}. \tag{23}$$

In the system description (22) the vector of parameters $a^T=\lfloor a^{(1)} \; a^{(2)}\rfloor$ is unknown. To determine the vector of unknown parameters the sequence of observations (12) were collected. The noise is assumed to be additive, i.e. functions (7) and (8) have the form:

$$u=x+z_x, \quad v=y+z_y. \tag{24}$$

and noise (z_x, z_y) is the realization of random variable $(\underline{z_x}, \underline{z_y})$ and probability density function (11) has the form:

$$g_z(z_x, z_y) = \begin{cases} 1 & if \quad -0.5 \le z_x \le 0.5 \wedge -0.5 \le z_y \le 0.5 \\ 0 & otherwise \end{cases}. \tag{25}$$

It is easy to see that in this example the relation $\overline{\Re}_a$ (8) is the form:

$$\overline{\Re}_a = \left\{(u,v) \in R^2 : a^{(1)} \le u \le a^{(1)} + 1.0 \wedge a^{(2)} \le v \le a^{(2)} + 1.0\right\} \tag{26}$$

and probability density function (16) observed random variables $(\underline{u}, \underline{v})$ defined on $\overline{\Re}_a$ is:

$$\overline{g}(u,v,a) = \begin{cases} 0.25\left(u - a^{(1)} + 1\right)\left(v - a^{(2)} + 1\right) & if \quad a^{(1)} - 1 \le u \le a^{(1)} \wedge a^{(2)} - 1 \le v \le a^{(2)} \\ 0.25\left(a^{(1)} - u + 1\right)\left(v - a^{(2)} + 1\right) & if \quad a^{(1)} \le u \le a^{(1)} + 1 \wedge a^{(2)} - 1 \le v \le a^{(2)} \\ 0.25\left(u - a^{(1)} + 1\right)\left(a^{(2)} - v + 1\right) & if \quad a^{(1)} - 1 \le u \le a^{(1)} \wedge a^{(2)} \le v \le a^{(2)} + 1 \\ 0.25\left(a^{(1)} - u + 1\right)\left(a^{(2)} - v + 1\right) & if \quad a^{(1)} \le u \le a^{(1)} + 1 \wedge a^{(2)} \le v \le a^{(2)} + 1 \end{cases} \tag{27}$$

The respective likelihood function (17) takes the form:

$$W_A(a, U_N, V_N) = \\ = \begin{cases} \overline{W}_A(a, U_N, V_N) \, if \ \forall n = 1,2,\dots,N \ a^{(1)} - 1 \le u_n \le a^{(1)} + 1 \wedge a^{(1)} - 1 \le u_n \le a^{(1)} + 1 \\ 0 \qquad otherwise \end{cases} \tag{28}$$

where:

$$\overline{W}_A(a, U_N, V_N) = \prod_{n \in N_1}\left(u_n - a^{(1)} + 1\right)\left(v_n - a^{(2)} + 1\right) \times \prod_{n \in N_2}\left(a^{(1)} - u_n + 1\right)\left(v_n - a^{(2)} + 1\right) \times \\ \times \prod_{n \in N_1}\left(u_n - a^{(1)} + 1\right)\left(a^{(2)} - v_n + 1\right) \times \prod_{n \in N_2}\left(a^{(1)} - u_n + 1\right)\left(a^{(2)} - v_n + 1\right), \tag{29}$$

and N_1, N_2, N_3, N_4 are sets of indexes defined as follows:

$$N_1 = \left\{n \in \overline{1,N} : a^{(1)} - 1 \le u_n \le a^{(1)} \wedge a^{(2)} - 1 \le v_n \le a^{(2)}\right\},$$

$$N_2 = \left\{n \in \overline{1,N} : a^{(1)} \le u_n \le a^{(1)} + 1 \wedge a^{(2)} - 1 \le v_n \le a^{(2)}\right\},$$

$$N_3 = \left\{n \in \overline{1,N} : a^{(1)} - 1 \le u_n \le a^{(1)} \wedge a^{(2)} \le v_n \le a^{(2)} + 1\right\},$$

$$N_4 = \left\{n \in \overline{1,N} : a^{(1)} \le u_n \le a^{(1)} + 1 \wedge a^{(2)} \le v_n \le a^{(2)} + 1\right\}.$$

The estimate a_{AN} of vector a is obtained by maximization (29) with respect to $a^{(1)}$ and $a^{(2)}$.To obtain the solution the numerical optimization method was used. For the described example the simulation study was performed. During simulation the vector of parameters $a^T = [4.0 \;\; 5.0]$. The sequence (x_n,y_n), $n=1,2,\ldots,N$ was generated with probability density given by (23) for N equal *10, 20, 30* and *40*. Then the sequences were noised (24) by the sequence generated with probability density given by (25) and this gives the sequence (12) for the investigated example. The obtained sequences (u_n,v_n), $n=1,2,\ldots,N$ were used to determine the estimates of unknown vector a. The simulations were repeated several times for the same N. The results are given in Table 1. In this table the estimates intervals are given for corresponding numerical experiment.

	N	10	20	30	40
$a^{(1)} = 4.0$	$a_{AN}^{(1)}$	$[3.43, 4.55]$	$[3.62, 4.37]$	$[3.71, 4.26]$	$[3.82, 4.12]$
$a^{(2)} = 5.0$	$a_{AN}^{(2)}$	$[4.41, 5.58]$	$[4.63, 5.34]$	$[4.76, 5.22]$	$[4.86, 5.07]$

Table 1: Simulation results

5 Parameter Estimation by Bayes' Methods

Additionally, if we assume that vector of parameters a in the description (1), (2) and (3) is the realization of random variable $\underline{a}$ with known probability density function $g_a(a)$, then the Bayes' approach may be used. In this case, for the sequence (12) the estimate a_{CN} of vector a may be obtained as the solution of the following optimization problem:

$$a_{CN} = \Psi_C(U_N, V_N) \rightarrow r(a_{CN}, U_N, V_N) = \min_{\bar{a} \in A} r_C(\bar{a}, U_N, V_N) \tag{30}$$

where $r_C(\bar{a}, U_N, V_N)$ is the conditional risk defined:

$$r_C(\bar{a}, U_N, V_N) = \\ = \int_A L(a,\bar{a}) g_a(a) \prod_{n=1}^{N} \int_{\Re_a} g_z\left(h_x^{-1}(x,u), h_y^{-1}(y,v)\right) \times g(x,y) \times |J_x| \times |J_y| dxdyda \tag{31}$$

where $L(a,\bar{a})$ is the loss function, and $\bar{a}$ is possible decision.

Similarly for the sequence (13) the estimate a_{DN} of vector a may be obtained as the solution of the optimization problem:

$$a_{DN} = \Psi_D(\Gamma_N) \to r(a_{DN}, \Gamma_N) = \min_{\bar{a} \in A} r_D(\bar{a}, \Gamma_N), \tag{32}$$

where $r_D(\bar{a}, \Gamma_N)$ is the conditional risk defined:

$$r(\bar{a}, \Gamma_N) = \\ = \int_A L(a, \bar{a}) g_a(a) \prod_{n=1}^{N} \int_{r_n} \int_{\Re_a} g_z\left(h_x^{-1}(x,u), h_y^{-1}(y,v)\right) \times g(x,y) \times |J_x| \times |J_y| dxdydudvda \tag{33}$$

For different loss functions $L(a, \bar{a})$ the different Bayes' method may be obtained.

6 Final Remarks

The problem of modeling of system described by the relation has been discussed. The static system is described by a set of fact facts about input and output. The set of true facts gives the relation defined on the set of inputs and outputs. In this paper it was assumed that description is known with accuracy to parameters. To determine unknown model parameters the estimation algorithm was proposed. Two different kinds of observations were used. The first case corresponds to traditional measurements, i.e. for given input the output is measured. The other observations are true logical sentences about inputs and outputs. The both kinds of observations are assumed to be noised. For both cases the estimation algorithms based on maximum likelihood and Bayes' approaches have been proposed. The presented approach is illustrated by simple analytical examples and a simulation study.

References

[Bubnicki, 80a] Bubnicki Z., Identification of Control Plants, Elsevier, Amsterdam-Oxford-New York, 1980.

[Bubnicki, 90] Bubnicki Z., Introduction to the Expert Systems, PWN, Warsaw, 1990 (in Polish).

[Bubnicki, 80b] Bubnicki Z., Introduction to the relation systems theory, Proc. of 7th Polish-Russian Symposium on Optimization and Control of Cybernetics Systems, Warsaw, IBS PAN, 1980, pp. 9-15.

[Bubnicki, 88] Bubnicki Z., Problems of control and identification. Expert systems. Proc. of 3rd Polish-Russian Scientific-Technological Conference "Computer Automation in Industry", Technical University of Wroclaw, 1988, pp.27-36.

[Swiatek, 89] Swiatek J., Same problems of identification for relation systems, Systems Science vol. 15, no. 1, 1989, pp. 26-34.

[Swiatek, 90] Swiatek J., Parameter estimation of systems described by the relation. Proc of 7th International conference on Systems Engineering, Las Vegas, 1990, pp.274-281.

[Swiatek, 06] Swiatek J., Parameter Estimation by Maximum Likelihood Method of Systems Described by the Relation, Proceedings of the 7th International Conference on Artificial Intelligence and Soft Computing - ICAISC 2006, pp. 1217-1222.