Approximation of Confidence Sets for Output Error Systems Using Interval Analysis ⋆

Sándor Kolumbán * **István Vajk** * **Johan Schoukens** **

* *Department of Automation and Applied Informatics, Budapest University of Technology and Economics, Budapest, Hungary, (e-mail: {kolumban,vajk}@aut.bme.hu)*
** *Department of Fundamental Electricity and Instrumentation, Vrije Universiteit Brussel, Brussels, Belgium, (e-mail: Johan.Schoukens@vub.ac.be)*

Abstract: Standard identification techniques usually result in a single point estimate of the system parameters. This is justified in cases when the number of observations is large compared to the number of system parameters. However in case of small sample count it is more reasonable to identify a set of possible parameters which contain the nominal parameters with a given probability. These confidence sets cannot be calculated directly. The paper proposes interval analytic techniques to approximate confidence sets of model parameters up to arbitrary precision. The origins of interval analysis lie in the field of reliable computing, giving certified results for every computation. It has been used to solve global optimization problems numerically providing theoretical certificates on the optimality of the results. This method of global optimization is modified in a suitable way to generate the needed confidence sets. Introduction to interval analytic techniques is given and the methodology of global optimization via these is also presented. The modifications of this algorithm needed to construct the confidence sets are discussed and the method is illustrated on a simple example. The presented algorithm is focused on the output error model structure but the methodology can be extended to more general cases as well.

Keywords: confidence set, small sample identification, interval analysis, output error

1. INTRODUCTION

The goal of this paper is to present an algorithm capable of generating confidence sets for the parameter estimates of a system from measurement data. This type of identification output is favoured compared to single point estimates in cases where the number of measurement points is not high enough compared to the number of parameters to be estimated. This is the case in some biological and medical applications where the sampling possibilities are really restricted (Godfrey et al., 2011).

Interval analytic methods are used in the presented algorithm and these were the major points of inspiration as well. The use of interval analysis is not new to system identification. The methodology of unknown but bounded errors was already used some forty years ago (Schweppe, 1968) and it is inherently an interval analytic method (Jaulin and Walter, 1993).

⋆ This work has been supported partly by the fund of the Hungarian Academy of Sciences for control research, by the project "Talent care and cultivation in the scientific workshops of BME" financed by the grant TÁMOP - 4.2.2.B-10/1–2010-0009. Support was also provided by the Fund for Scientific Research (FWO-Vlaanderen), in part by the Flemish Government (Methusalem) and in part by the Belgian Government through the Interuniversity Poles of Attraction (IUAP VI/4) Program.

The start of interval analysis is said to be the publication of Moore (1966). Back then, the focus of interval analysis was to analyse the propagation of errors caused by numerical algorithms towards the final results. These are usually caused by rounding errors. In time these methods were used not just to bound the errors on the results but also to solve systems of equations. Later on a full optimization framework was developed on the bases of interval analytic concepts. The book of Hansen (1992) contains an extensive introduction to interval analysis and its application for solving linear and non-linear systems of equations and global optimization problems.

The use of interval analytic methods in system identification is partly concentrated around the framework of unknown but bounded errors. In that case the errors are assumed to be bounded meaning that for each measurement an interval can be specified in which the true value lies. Given these intervals one is interested in the set of model parameters which would imply that any particular model from the set would result in noiseless measurements in the given intervals (Belforte and Milanese, 1978). In other occurrences of interval analytic methods it is used as a global optimization procedure to solve the identification as an optimization problem (Kampen et al., 2011) or to get certified numerical estimates (Braems et al., 2003).

The rest of the paper is built up as follows. Section 2 contains the necessary introductions to interval analysis

and how global optimization problems are solved using these techniques. The identification problem of output error models and some notation is defined in Section 3. Following these, Section 4 defines confidence sets for parameter estimation with a given confidence level. The algorithm which approximates these sets up to any user defined precision is presented in Section 5, while Section 6 demonstrates the applicability of the method on a simple first order system. Concluding remarks are given in Section 7.

2. INTERVAL ANALYSIS, OVERVIEW

This section briefly presents the concepts of interval arithmetic and its application for global optimization. The notation follows that of (Hansen, 1992). In interval analysis, as the name shows, intervals are used instead of numbers. A closed bounded interval X is defined as

$$X = [a, b] = \{x \in \mathbb{R} | a \leq x \leq b\} \tag{1}$$

Simple numbers are represented as degenerate intervals. For example $2 = [2, 2]$. In case of numbers that are not representable by machine numbers, such as π or 0.1, small intervals are defined which contain the actual number. These intervals have the nearest smaller and larger machine numbers as bounds.

Results of operations having interval operands are defined in a way that the resulting interval contains all possible results for any possible choices of operands. More formally in case of operations with two operands if (2) is satisfied

$$\forall x \in X, y \in Y : \underline{f} \leq f(x, y), \overline{f} \geq f(x, y) \tag{2}$$

then the result is correctly defined as

$$f(X, Y) = [\underline{f}, \overline{f}] \tag{3}$$

The basic concept is that every operation defined for numbers is also defined for intervals in a way that the possible results for numbers are contained in the result set for interval operands. It is not required that the defined result interval should be tight around the possible results but it is beneficial. The following examples illustrate the definition of summation, multiplication and exponentiation.

$$[-1, 1] + [1, 2] = [0, 3] \tag{4}$$

$$[-1, 1] \cdot [1, 2] = [-2, 2] \tag{5}$$

$$[-1, 1]^2 = [0, 1] \tag{6}$$

$$[-1, 1] \cdot [-1, 1] = [-1, 1] \tag{7}$$

One of the fundamental drawbacks of carrying out calculations with interval operands is illustrated by the examples (6) and (7). When working with interval variables multiplication $X \cdot X$ and squaring X^2 does not give the same result. This is called decoupling and it is a consequence of the definition of interval operations (2). Another typical example of decoupling is that if $X = [\underline{x}, \overline{x}]$ then $X - X = [\underline{x} - \overline{x}, \overline{x} - \underline{x}] \neq [0, 0]$.

Decoupling has an important role in every interval arithmetic based algorithm. The implementation and sequence of operations is really important. As a special case special care should be taken to calculate the simulated output of linear systems for a given input signal.

For more details on interval arithmetic, such as using the extended real numbers and handling division by zero, the reader is directed to (Hansen, 1992).

Using interval analytic operations one can evaluate an objective function over a whole domain of operands. Since almost every computer architecture allows the configuration of the rounding behaviour, it can be guaranteed that the boundaries of the result intervals are calculated with outward rounding. This fact helps to discard whole portions of the variable space in case of global optimization.

The rest of this section presents the most basic type of interval analytic global optimization algorithm. This will be modified in Section 5 to approximate confidence sets. The presented version is a non-derivative version of the state-of-the-art interval analytic optimization algorithms but it will be enough for the purposes of this paper. Only "unconstrained" optimization is considered where the optimization is given as

$$\min \ : \ f(x) \qquad x \in X^{(0)} \subset \mathbb{R}^n \tag{8}$$

This problem formulation is called unconstrained because of the simple structure of $X^{(0)}$ as it is a box without real structural constraints.

$$X_i^{(0)} = \left[\underline{x_i}, \overline{x_i}\right] \quad i = 1, \ldots, n \tag{9}$$

Interval analytic optimization algorithms can approximate globally optimal solutions up to any user given precision ε. Along the course of the algorithm two lists of boxes $(L1, L2)$ are maintained and a known upper bound for the optimal value $(\overline{f})$. $L1$ contains boxes representing portions of the variable space which may contain the global solutions. $L2$ contains portions of the variable space which have diameter less than ε and at the current state of the algorithm these may contain the global solution.

Algorithm 1 shows the outline of a the non-derivative global optimization algorithm.

Algorithm 1 Global optimization

(1) $L1 \leftarrow X^{(0)}$, $L2 \leftarrow \emptyset$
(2) $\overline{f} = \infty$
(3) while $L1$ is not empty
(4) select an element $X^{(i)}$ from $L1$
(5) $\overline{f} = \min\left(\overline{f}, \sup\left(\underline{f}\left(C\left(X^{(i)}\right)\right)\right)\right)$
(6) if $\inf(f(X^{(i)})) > \overline{f}$ continue from (3)
(7) split $X^{(i)}$ into smaller boxes LX
(8) add elements of LX smaller in diameter than ε to $L2$
(9) add elements of LX larger in diameter than ε to $L1$
(10) endwhile
(11) foreach $X^{(i)}$ in $L2$
(12) if $\inf\left(\underline{f}\left(C\left(X^{(i)}\right)\right)\right) > \overline{f}$ discard $X^{(i)}$ from $L2$
(13) endforeach

The notation $C(X)$ denotes the center of the given box X. The algorithm starts with the whole variable space as a single box. When selecting a new box $X^{(i)}$ for processing it is sure that evaluating the objective function in the center of $X^{(i)}$ will give an upper bound on the globally optimal value. Also, if the interval evaluation of the objective function over $X^{(i)}$ results in a set which is above the

currently known upper bound, it is sure that the global optimum cannot be in $X^{(i)}$. If this cannot be decided then there might be a point in $X^{(i)}$ at which the objective function results in a better upper bound. In this case the set $X^{(i)}$ is split into a number of smaller boxes. If a box is smaller than the user specifications then it is considered at the end of the algorithm. If it is still large enough than it is added to the list of boxes needed to be processed, $L1$. Evaluating the objective function over smaller boxes reduces the effects of decoupling and results in sharper bounds on the result sets. At the end there is a list of ε small intervals. Evaluating the objective function on these intervals may result in sets which are completely above the known upper bound $\overline{f}$, these are discarded. The remaining sets will contain all globally optimal points. The algorithm can be restarted with a different precision $\varepsilon_2 < \varepsilon$ and $L1_2 = L2$, thus refining the results even more.

There is a number of different possibilities to enhance the performance of this variant of the optimization algorithm. The choice of $X^{(i)}$ from $L1$ and the choice of dimensions used during the splitting can be made using different heuristics but these are not considered here. For detailed discussion on these topics the reader is directed to (Hansen, 1992).

Algorithm 1 will be modified in Section 5 to approximate confidence sets with arbitrary precision. The algorithm implementation is based on the interval analytic toolbox Rump (1999).

3. IDENTIFICATION PROBLEM

The generalized Box-Jenkins model structure, as defined in (Ljung, 2003) is given by (10)

$$A(q)y[k] = \frac{B(q)}{F(q)}u[k] + \frac{C(q)}{D(q)}e[k] \tag{10}$$

where $A(q)$, $B(q)$, $C(q)$, $D(q)$ and $F(q)$ are polynomials in q^{-1} and q is the forward shift operator, meaning that $(qx)[k] = x[k+1]$ for any time series x. The noise source $e[k]$ is assumed to be white, zero mean and Gaussian. The polynomials $A(q)$, $C(q)$, $D(q)$ and $F(q)$ are monic, meaning that the coefficient of q^{-0} is one in each of these polynomials.

The identification problem is determined by the model structure chosen to be identified, the available measurement data and the measure of fitness used to rank specific models.

The rest of the paper concentrates on the output error (OE) model structure. This means that $A(q) = C(q) = D(q) = 1$. This is a simple model structure but complex enough to require the presented techniques for approximation of confidence sets. The presented procedure can be generalized to other model structures as well but it would take the focus away from the main points, the concept of confidence sets and the algorithm to approximate these.

Models are ranked according to the ℓ^2 norm of the noise sequence required by the model to generate the measured data. This means that the objective function value corresponding to a particular model is defined as

$$J(\theta) = \frac{1}{N}\sum_{k=1}^{N} e^2[k] \tag{11}$$

where N denotes the number of samples. It is important to note that if the values $e[k]$ are assumed to be zero mean independent identically distributed random variables then the objective function (11) is an unbiased and efficient estimate for the variance of their distribution noise.

Classical system identification aims at finding a single model in the possible set of models which minimizes the objective function (11). This approach is justified when the number of samples is large compared to the number of system parameters. However in case of relatively small sample count the best fitting model might be over-fitted, also it is not very reliable. Because of these issues it is reasonable to search for a set of parameters. The algorithm presented bellow searches for confidence sets which are centred around the globally optimal parameter values but also include parameters which are not significantly different from those. The next section provides a way to decide whether two models are considered to be significantly different or not.

4. CONFIDENCE SETS

The concept of confidence intervals is a basic element in elementary statistics, in the multivariate case these are usually called confidence sets or regions. In parametric statistical analysis a basic problem is that there is a number of independent and identically distributed samples from an unknown distribution and one wants to estimate the unknown parameters of this distribution. For instance in case of Gaussian samples $X_i \sim \mathcal{N}(\mu, \sigma^2)$ this means estimating the mean and the variance. The average of the samples is an unbiased and efficient estimate of the mean.

$$\overline{X} = \frac{1}{N}\sum_{k=1}^{N} X_i \tag{12}$$

However, since the Gaussian distribution is an absolutely continuous one, the probability $\mathbb{P}(\overline{X} = \mu) = 0$. To overcome the fact that every estimate has zero probability of being right, an interval is given around $\overline{X}$. If this interval is constructed in a way that it contains the actual expectation μ with probability p then it is called a p-confidence interval. For rigorous definition of confidence sets see (Anderson, 2003).

This section generalizes the concept of confidence sets to system parameters. For the time being it is assumed that the globally optimal parameter vector θ^* is known. The question is that when can a different parameter vector θ be considered as significantly different with a given confidence p.

Let $e_{\theta^*}[\cdot]$ and $e_\theta[\cdot]$ denote the noise samples needed to generate the measurement data for the models θ^* and θ. Mind that the objective function (11) is actually an estimate of the noise variance. Thus the question is whether the two variance estimates $J(\theta^*)$ and $J(\theta)$ do differ significantly or not. Comparison of variances is done using the one sided F-test, which is based upon the fact that if both $e_{\theta^*}[\cdot]$ and $e_\theta[\cdot]$ are zero mean Gaussian with different variances then the ratio (13) follows F distribution with both degrees of freedom being N.

$$F = \frac{J(\theta)}{J(\theta^*)} \sim F(N,N) \tag{13}$$

If the F value is greater than the p percentile of the $F(N,N)$ distribution then the two noise sequences $e_{\theta^*}[\cdot]$ and $e_\theta[\cdot]$ are said to have significantly different variance. Thus the corresponding model parameters significantly differ. If there are parameters already estimated from the samples (such as the mean) then the degrees of freedom should be decreased with the number of estimated parameters.

Confidence sets with confidence level p around θ^* can be characterized with a positive multiplier $\alpha_{p,N-n} \geq 1$, where n denotes the number of unknown system parameters. This multiplier is chosen in a way that if $J(\theta) > \alpha_{p,N-n} J(\theta^*)$ then the two parameter vectors are considered to be significantly different. It is important to note that this multiplier does not depend on the actual value of θ^*. It is simply the p-percentile of the $F(N-n, N-n)$ distribution.

To summarize the contents of this section, the p level confidence set of model parameters is defined as

$$\Omega_p = \{\theta \ : \ J(\theta) \leq \alpha_{p,N-n} J(\theta^*)\} \tag{14}$$

Figure 1 shows the values of $\alpha_{p,N-n}$ for 90% confidence as a function of degrees of freedom. It is worth mentioning that as the sample count grows the multiplier α in the definition of the confidence set (14) converges to one. This means that even a small deviation from the globally optimal objective function value is considered significant. Since the objective function is continuous in the system parameters this fact transfers to the system parameters as well. Even small deviation in the parameters is considered significant, meaning that a point estimate is sufficiently reliable. On the other hand for small number of degrees of freedom the multiplier $\alpha_{p,N-n}$ is rather large. This means that a large range of objective function values is acceptable. This range can be generated by a large set of system parameters. Thus rendering a single point estimate basically useless.

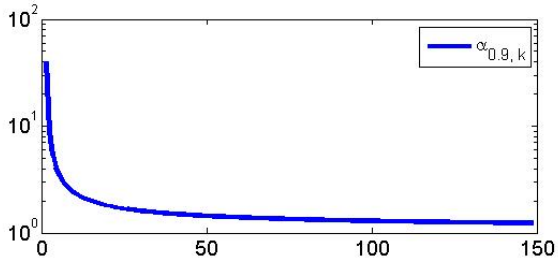

Fig. 1. The values of $\alpha_{p,N-n}$ with $p = 0.9$ and $N - n$ ranging from 1 to 150

5. APPROXIMATION OF CONFIDENCE SETS

The scope of this section is to present an algorithm which will approximate the confidence sets (14) to an arbitrary precision. As input the algorithm will have the desired confidence level p, the desired resolution ϵ, the measurement data $u[k]$ and $y[k]$ for $k = 1, \ldots, N$ and the order of the system needed to be identified.

The first step of the algorithm is to determine the multiplier α which is used to distinguish between the different parameter values. This is done as presented in the previous section.

As the second step the set of parameters $X^{(0)}$ needs to be defined. Since output error models (15) are considered, the decision variables θ are now composed of the coefficients of the polynomials $B(q)$ and $F(q)$. It assumed We stated in Section 3 that the polynomial $F(q)$ is monic. This is not a necessary restriction, it is one of many possible ways to eliminate having multiple solution to the problem. Multiple solutions are due to the fact that $J(\theta) = J(c\theta)$ for every $c \neq 0$.

The OE model is given as

$$y[k] = \frac{b_1 q^{-1} + \cdots + b_n q^{-n}}{a_0 + a_1 q^{-1} + \cdots + a_n q^{-n}} u[k] + e[k] \tag{15}$$

In order to be able to use interval analytic calculations efficiently the model is searched for as the sum of first and second order systems and it is assumed that the system has no poles with multiplicity more than one. In this case the system can be decomposed to a number of first order components and a number of second order components with complex conjugate poles.

$$\begin{aligned} y[k] = &\sum_{f=1}^{F} \frac{b_{f,1} q^{-1}}{a_{f,0} + a_{f,1} q^{-1}} u[k] + \\ &+ \sum_{s=1}^{S} \frac{b_{s,1} q^{-1} + b_{s,2} q^{-2}}{a_{s,0} + a_{s,1} q^{-1} + a_{s,2} q^{-2}} u[k] + \\ &+ e[k] \end{aligned} \tag{16}$$

In case of the first order components the parameter vector is composed a $\theta_f = [a_{f,0} \ a_{f,1} \ b_{f,1}]^T$. Multiple solutions are excluded by restricting the optimization to norm one vectors $\|\theta_f\| = 1$ with $a_{f,0} > 0$. The efficient way to obey these constraints is to translate the problem into polar coordinates ϕ_f where $\phi_f(1) \in [0, \pi/2 - \epsilon]$ and $\phi_f(2) \in [0, 2\pi]$. $\phi_f(1)$ is separated from $\pi/2$ to avoid situations with $a_{f,0} = 0$. The coordinate transformation is given as

$$\begin{aligned} a_{f,0} &= \cos(\phi_f(1)) \\ a_{f,1} &= \sin(\phi_f(1)) \cos(\phi_f(2)) \\ b_{f,1} &= \sin(\phi_f(1)) \sin(\phi_f(2)) \end{aligned} \tag{17}$$

Calculating the output of a first order component with zero initial conditions can be realized using the state space representation $A = -\frac{a_{f,1}}{a_{f,0}}$ $B = \frac{b_{f,1}}{a_{f,0}}$ $C = 1$ $D = 0$. Using the Toeplitz matrix H generated by the column vector with entries CA^kB, the vector of output values Y_f is given by $Y_f = HU$ where U is the column vector input values.

$$Y_f = \begin{bmatrix} 0 & & & \\ CB & 0 & & \\ CAB & CB & 0 & \\ \vdots & & & \ddots \\ CA^{N-2}B & \ldots & CAB & CB \ 0 \end{bmatrix} U \tag{18}$$

Second order systems are parametrized in a different manner. Every second order component $H(q)$ can be decomposed as the sum of two first order components which are complex conjugates of each other.

$$H(q) = H_1(q) + \overline{H_1}(q) = \\ = \frac{B_1 e^{i\rho_1} q^{-1}}{A_0 + A_1 e^{i\rho_2} q^{-1}} + \frac{B_1 e^{-i\rho_1} q^{-1}}{A_0 + A_1 e^{-i\rho_2} q^{-1}} = \\ = \frac{2B_1 A_0 \cos(\rho_1) q^{-1} + 2B_1 A_1 \cos(\rho_2 - \rho_1) q^{-2}}{A_0^2 + 2A_0 A_1 \cos(\rho_2) q^{-1} + A_1^2 q^{-2}} \quad (19)$$

The parameter vector for second order systems is $\theta_s = [A_{s,0}\ A_{s,1}\ B_{s,1}\ \rho_1\ \rho_2]^T$. It is immediate that the values of A_0, A_1 and B_1 can be multiplied with any non-zero number without changing the system. Thus, the representation of these parameters is converted to polar coordinates again. The last two unknowns ρ_1 and ρ_2 are already angular values. The first two elements of the vector ϕ_s correspond to the polar representation of A_0, A_1 and B_1, while the last two elements are exactly ρ_1 and ρ_2. The sign of A_0 is fixed to be positive so the first two coordinates are just as in the first order case, $\phi_s(1) \in [0, \pi/2 - \epsilon]$ and $\phi_s(2) \in [0, 2\pi]$. If ρ_2 can have values from $[0\ \pi/2]$ and ρ_1 from $[0\ 2\pi]$ then all possible second order systems can described as

$$\begin{aligned} A_{s,0} &= \cos(\phi_s(1)) \\ A_{s,1} &= \sin(\phi_s(1)) \cos(\phi_s(2)) \\ B_{s,1} &= \sin(\phi_s(1)) \sin(\phi_s(2)) \\ \rho_{s,1} &= \phi_s(3) \\ \rho_{s,2} &= \phi_s(4) \end{aligned} \quad (20)$$

The output of second order systems in the presented parametrization can be calculated as follows. Consider the state space representation

$$A = -\frac{A_{s,1}}{A_{s,0}} e^{i\rho_{s,2}}\ B = \frac{B_{s,1}}{A_{s,0}} e^{i\rho_{s,1}}\ C = 1\ D = 0 \quad (21)$$

With this description the output of a second order system assuming zero initial conditions is given as

$$Y_s = 2\ \mathrm{Re}(Y_s^*) \qquad Y_s^* = HU \quad (22)$$

where H is defined as for the first order systems but with the current state space representation.

The decomposition to first and second order systems is necessary to reduce the effect of decoupling during the evaluation of the objective function. In case of the output error model this is done by simply simulating the system with the measured input signals and subtracting the results from the measured output values. These differences are the estimated output errors. The effect of decoupling between interval parameters is minimized by the calculation of the Toeplitz matrix H. In each entry of the matrix the exponentiation can be carried out without decoupling. When calculating the simulated output, different rows and entries within the same row are decoupled but that cannot be eliminated from the process.

If the order of the system to be estimated is n and every pole has multiplicity one than there are at most $\lfloor n/2 \rfloor + 1$ different decompositions of it depending on the number of second order components. Independent from the number of second order systems in the given realization the length of the parameter vector is always $2n$. In case of k first order components the first $2k$ elements are understood as representations of first order systems put one after the other while the rest is interpreted as parameters of second order systems. With this representation all possible system parameters can be described with $\lfloor n/2 \rfloor + 1$ different boxes. Each box corresponds to a given number of first order components in the system decomposition. This list of boxes is used to start the algorithm that generates the confidence sets of the parameter estimates.

Algorithm 2 presents the outline of the modified algorithm used to approximate the confidence set.

Algorithm 2 Confidence set approximation

(1) $L1 \leftarrow StartList$, $L2 \leftarrow \emptyset$
(2) $\overline{f} = \infty$
(3) while $L1$ is not empty
(4) select an element $X^{(i)}$ from $L1$
(5) $\overline{f} = \min\left(\overline{f}, f\left(C\left(X^{(i)}\right)\right)\right)$
(6) if $\inf(f(X^{(i)})) > \alpha\overline{f}$ continue from (3)
(7) split $X^{(i)}$ into smaller boxes LX
(8) add elements of LX smaller in diameter than ε to $L2$
(9) add elements of LX larger in diameter than ε to $L1$
(10) endwhile
(11) foreach $X^{(i)}$ in $L2$
(12) if $\inf\left(f\left(C\left(X^{(i)}\right)\right)\right) > \alpha\overline{f}$ discard $X^{(i)}$ from $L2$
(13) endforeach

As the starting step, the list of possible boxes is set. As opposed to Algorithm 1 in this case already a list of possible boxes is given. These boxes are defined in a special way. Their first coordinate is an integer number, showing the number of first order terms in the decomposition. After this, there are dimensions corresponding to the polar coordinates of the individual first and second order systems. Splitting is never carried out on the integer dimension.

After setting the initial list of boxes a loop evaluates items in this list. Every selected item is evaluated at its center point. Sets are split in the same manner as in Algorithm 1. Major differences are present in two steps. One of them is that sets are not discarded when their lower bound is larger than the known upper bound but only if it is greater than $\alpha\overline{f}$ (step 6). This ensures that no models are discarded unless they are significantly different from the yet unknown globally optimal solution. During the execution of the algorithm the known upper bound of the globally optimal objective function value decreases to the true value. Parallel with this domains which were not discarded in previous iterations may become discarded and only those boxes are kept which are not significantly different from the global optima.

The second difference is in the way how sets are added to the list $L1$. In order to avoid multiple solutions in the variable space, some constraints should be obeyed by the algorithm. The first and second order terms in (16) should be ordered for example in increasing order of static gain within themselves.

At the end of the algorithm the list $L2$ contains sets for which the objective function value is under $\alpha J(\theta^*)$. If this is not the case then the distance of the worst parameter in the set from one that is resulting in a small enough objective function value is less then ε.

It should be noted that this formulation of the algorithm assumes that the system was started from zero state. Otherwise the starting state should also be part of the optimization variables.

6. DEMONSTRATION OF THE METHOD

This section illustrates the algorithm with a sample run on a simple first order system. First order systems assuming zero starting state contain only two decision variables this makes them suitable for illustration because the state of the lists $L1$ and $L2$ can be visualized.

The sample system is given with the impulse transfer function

$$G(q) = \frac{0.2658q^{-1}}{0.8282 - 0.5460q^{-1}} \tag{23}$$

The input of the identification is given in Figure 2. The input signal is the black one, the true output is the blue signal whereas the red signal is the noisy measurement.

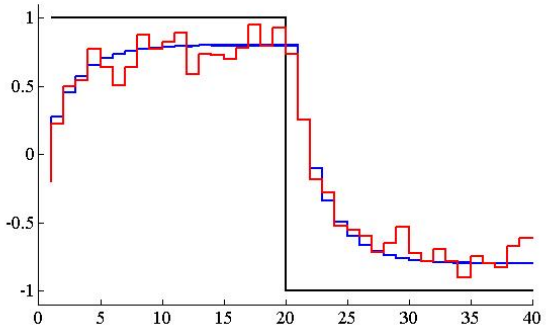

Fig. 2. The input of the identification problem.

Using the oe routine of the Matlab Control System Toolbox, the obtained globally optimal model is

$$G(q) = \frac{0.2325q^{-1}}{0.8172 - 0.5256q^{-1}} \quad \theta^* = \begin{bmatrix} 0.8172 \\ -0.5256 \\ 0.2325 \end{bmatrix} \tag{24}$$

Since the coefficient of a_0 in the denominator is constrained to be positive and the norm of the vector θ is constrained to be one the solution set is on the half sphere positive in its first coordinate. The state of Algorithm 2 can be visualized by showing the projection of the half sphere onto the plane (a_1, b_1).

The algorithm is started with confidence level $p = 0.9$. Figure 3 shows the state of the algorithm after 25 iterations. There are 89 boxes in the list $L1$. Each such box is presented in the figure with a different colour. It can be seen that large portions of the variable space are already processed and discarded as they cannot contain the optimal solution.

Figure 4 shows the state of the algorithm after 225 iterations. The sharp edges around large blocks indicate that these portions of the decision space are discarded in blocks. It is also visible that the decision space is coved with fine resolution boxes around the global optimum. The reason for this is that approximating the confidence set

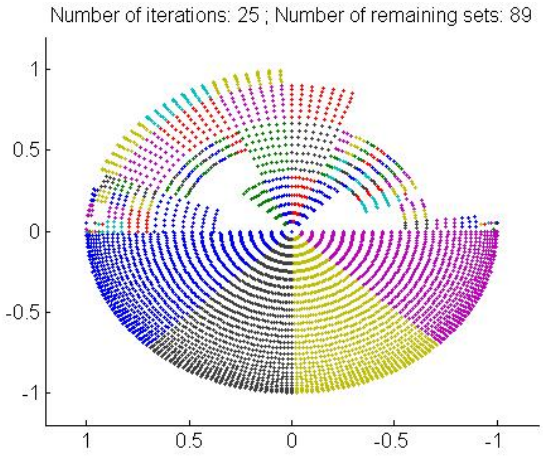

Fig. 3. The state of the algorithm after 25 iteration.

requires fine resolution covering of the neighbourhood of the optimum point.

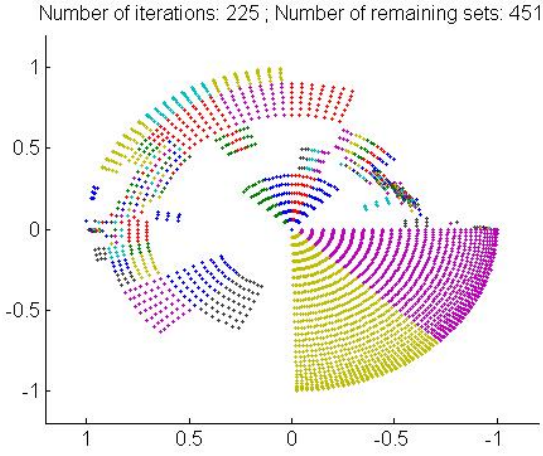

Fig. 4. The state of the algorithm after 225 iteration.

Letting the algorithm finish, the list $L2$ contains boxes covering the confidence set around the optimal value. Figure 5 shows the center of each box inf the resulting list. The red point is the globally optimal solution to the problem found by both Matlab and Algorithm 2.

Figure 6 contains the simulated outputs of the models in the obtained confidence set. The black line corresponds to the globally optimal estimate while the blue lines correspond to some system parameters in the confidence set.

7. CONCLUSIONS

The execution time of the presented algorithm increases with the system order due to a number reasons. One of them is that covering a high dimensional set with boxes require larger number of boxes. Also this presented simple nonderivative version of the algorithm uses to many evaluations of the objective function. This issue can be handled by using more advanced versions of Algorithm 1.

The paper presented the interval analytic framework for global optimization. The notion of confidence sets was

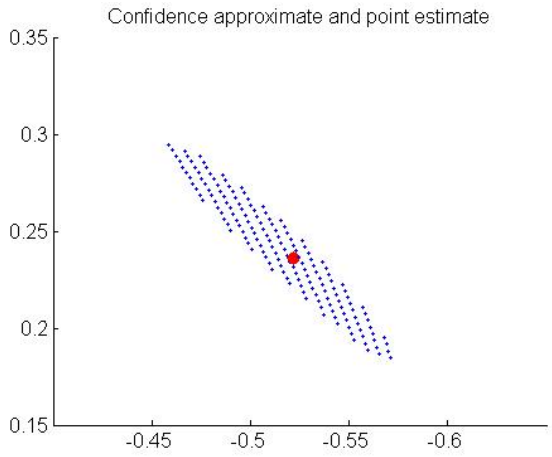

Fig. 5. The center of each remaining box in the confidence set and the global optimum.

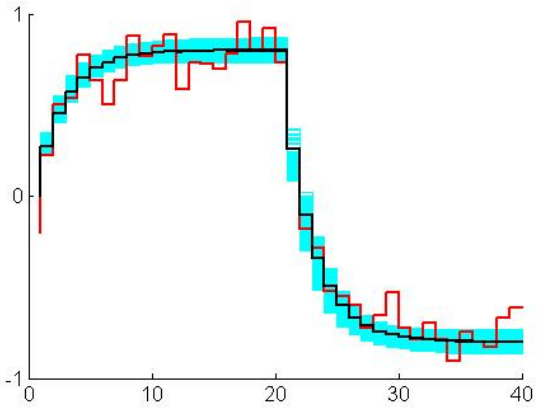

Fig. 6. Simulated outputs of the systems within the confidence set.

defined for system parameters in a way that it allowed to decide whether a given system parameter is significantly different from the possibly yet unknown globally optimal one or not. This allowed the modification of the previously presented global optimization algorithm in a way to construct the p-level confidence sets of the identification. The method was illustrated on a simple first order system while showing the inner progress of the algorithm as well.

REFERENCES

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley.

Belforte, G. and Milanese, M. (1978). Parameter estimation with unknown but bounded error of measure. In *Proc. International Conference on Cybernetics and Society (IEEE SMC)*, volume II, 1329–1332.

Braems, I., Jaulin, L., Kieffer, M., Ramdani, N., and E.Walter (2003). Reliable parameter estimation in presence of uncertain variables that are not estimated. *13th IFAC Symposium On System Identification*, 1856–1861.

Godfrey, K.R., Arundel, P.A., Dong, Z., and Bryant, R. (2011). Modelling the double peak phenomenon in pharmacokinetics. *Comput. Methods Prog. Biomed.*, 104(2), 62–69.

Hansen, E. (1992). *Global Optimization Using Interval Analysis*. Dekker.

Jaulin, L. and Walter, E. (1993). Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4), 1053 – 1064.

Kampen, E., Chu, Q.P., and Mulder, J.A. (2011). Interval analysis as a system identification tool. In F. Holzapfel and S. Theil (eds.), *Advances in Aerospace Guidance, Navigation and Control*, 333–343. Springer Berlin Heidelberg.

Ljung, L. (2003). Initialisation aspects for subspace and output-error identification methods. Technical Report LiTH-ISY-R-2565, Department of Electrical Engineering, Linkping University, SE-581 83 Linkping, Sweden.

Moore, R.E. (1966). *Interval Analysis*. Prentice-Hall.

Rump, S. (1999). Intlab - interval laboratory. In T. Csendes (ed.), *Developments in Reliable Computing*, 77–104. Kluwer Academic Publishers.

Schweppe, F. (1968). Recursive state estimation: Unknown but bounded errors and system inputs. *Automatic Control, IEEE Transactions on*, 13(1), 22 – 28.