

APPROXIMATION OF A LINEAR DYNAMIC PROCESS MODEL USING THE FREQUENCY APPROACH AND A NON-QUADRATIC MEASURE OF THE MODEL ERROR

KRZYSZTOF B. JANISZOWSKI

Institute of Automatic Control and Robotics
Warsaw University of Technology, ul. A. Boboli 8, 02-525 Warsaw, Poland
e-mail: kjanisz@mchtr.pw.edu.pl

The paper presents a novel approach to approximation of a linear transfer function model, based on dynamic properties represented by a frequency response, e.g., determined as a result of discrete-time identification. The approximation is derived for minimization of a non-quadratic performance index. This index can be determined as an exponent or absolute norm of an error. Two algorithms for determination of the approximation coefficients are considered, a batch processing one and a recursive scheme, based on the well-known on-line identification algorithm. The proposed approach is not sensitive to local outliers present in the original frequency response. Application of the approach and its features are presented on examples of two simple dynamic systems.

Keywords: approximation method, frequency domain, non-quadratic criterion, recursive algorithm.

1. Introduction

Investigation of process dynamics may result in different representations. Commonly used identification schemes (Ljung and Söderström, 1987) result in determination of linear parametric models, which have compact form of some transfer function, sometimes of a quite high order. Before application of this model in controller design, it is helpful to get a simplified, lower-order transfer function in a continuous time or a discrete time domain. In the case of determination of a fuzzy parameterized TSK model, a simplified representation is very advantageous, too. In the case of, e.g., determination of a model composed of a static actuator part and an inertial representation of a dynamic part, the simplified model is valuable, too. A common feature for all the mentioned problems is the possibility of determination of a frequency characteristic for the investigated model. Then the underlying computational problem is presented as determination of a linear transfer function model, derived from the known frequency characteristic.

In electrical-engineering applications, the fitting of an approximation in frequency domain is often considered during the modelling of parts of energy transfer components (Deschrijver *et al.*, 2007; Gustavsen and Mo, 2007; Gustavsen, 2006; 2004; Lima *et al.*, 2005; Mohan

et al., 2004; Unbehauen and Rao, 1997), hence this problem is deeply investigated in terms of proper soft modelling (Deschrijver *et al.*, 2007; Grivet-Talocia *et al.*, 2005; Gustavsen, 2006; 2004) and even building some hardware models (Mohan *et al.*, 2004). In the area of control techniques, various ways are used (Kowalczyk and Kozłowski, 2011; Pintelon and Schoukens, 2004; Sreeram and Agatokhlis, 1991; Unbehauen and Rao, 1997; Varricchio *et al.*, 2004; Wahlberg and Mäkilä, 1996), but usually they are focused on weighted approximation with the least-squares measure of approximation errors.

Various approaches are used to determine these approximations (Ljung and Söderström, 1987). The problem of fitting a reasonable model for some frequency characteristic was formulated long ago (Levy, 1959) and investigated by many authors in classic transfer function form (Ljung and Söderström, 1987; Pintelon and Schoukens, 2004) or other representations (Wahlberg and Mäkilä, 1996). Various techniques are suggested, and among them most interesting are methods based on the basic approach of vector fitting and its ortho normal version (Deschrijver *et al.*, 2010; 2011; 2007; Grivet-Talocia *et al.*, 2005; Gustavsen and Mo, 2007; Gustavsen, 2006; 2004; Gustavsen and Semlyen, 1999). These approaches are usually focused on determination

of poles of the final transfer function and can be very precise. Based on measured and recorded frequency characteristics, the resulting models can fit to the data with negligible errors; however, some problems with outliers points have been observed and the l_1 norm was used in the fitting algorithm (Deschrijver *et al.*, 2010; 2011). Similar problems are observed during the fitting of continuous time models (Kowalczuk and Kozłowski, 2011), too.

The approach presented in the above-mentioned works is usually based on the assumption of the linear behaviour of processes considered and a proper choice of poles for model fitting to the estimated frequency plots. In identification of processes with dynamics reflecting some non-linear phenomena, e.g., saturation or hysteresis, the process behaviour depends on working conditions and sometimes is modelled by neural nets. The corresponding frequency plots, developed for these processes, can represent other transients than the usually investigated development of the model with the vector fitting approach. The vector fitting algorithms discussed above are focused on proper selection of transfer function poles and can be less efficient in the mentioned cases. The problem considered in the paper is focused on fitting a frequency response to the data that can have some outliers, far from any rational frequency response, and therefore the fitting algorithm can be defined in way different from usual least-squares algorithms.

An approximation with a non-quadratic performance index has been considered in various contributions (Fiodorov, 1994; Janiszowski, 1998; Kozłowski, 2003; Kowalczuk and Kozłowski, 2011), and some positive aspects of this approach have been pointed out. However, numerical problems with the convergence of the resulting algorithms have been mentioned, too. Recent applications (Deschrijver *et al.*, 2010; 2011) confirm an important feature of the non-quadratic performance index that the resulting models reflect low sensitivity to local distortions of the original frequency characteristic.

The problem of fitting a model to a given frequency characteristic, defined by a set of values $FR = \{c_1, \dots, c_N\}$, $c_k = FR(\omega_k) \in \mathbb{C}$, is usually defined on some interval $\Omega \supset \{\omega_1, \omega_2, \dots, \omega_N\}$. The problem can be presented as the minimization of the performance index

$$I_\Omega = \sum_{k=1}^N \|c_k - F_a(\omega_k)\|_{\alpha_k}, \quad (1)$$

where $F_a(\omega_k)$ is a frequency characteristic of the approximation. The introduced weights $\alpha_k \geq 0$ can induce a filter effect of different values c_k within interval Ω . The employed measure $\|\cdot\|$ can be L^2 or another norm. In the case of a norm resulting in least-squares fitting, the minimization problem of I_Ω is easy to solve and the corresponding relations for the optimal solution can be easily derived. However, in the case of the matching

problem, the application of the least-squares approach is not the best choice. On the other hand, the least sum of absolute errors (LSA) can be considered, but it induces poor convergence properties (Fiodorov, 1994; Janiszowski, 1998; Kozłowski, 2003). The nature of this effect can be explained by the following example.

Let us consider the problem of matching a position of a section of a known length, equal to 10, by 6 measurements of the initial point (*IP*) position

$$\begin{aligned} X_1 &= \{x_{11}, x_{12}, \dots, x_{16}\} \\ &= \{10.1, 9.8, 10.2, 9.7, -10, 10.3\} \end{aligned}$$

and 6 measurements of the end point (*EP*) position

$$\begin{aligned} X_2 &= \{x_{21}, x_{22}, \dots, x_{26}\} \\ &= \{19.9, 20.1, 19.8, 20.2, 19.7, 20.3\}. \end{aligned}$$

Let the performance index, used for the matching of the proper position of *IP* and *EP*, be defined in the form of

$$\begin{aligned} I &= \sum_{k=1}^6 \|IP - x_{1k}\|^\gamma + \sum_{k=1}^6 \|EP - x_{2k}\|^\gamma \\ &= \sum_{k=1}^6 \|IP - x_{1k}\|^\gamma + \sum_{k=1}^6 \|IP + 10 - x_{2k}\|^\gamma \\ &= I(IP), \quad \gamma \geq 1. \end{aligned} \quad (2)$$

Close inspection of the set X_1 shows one point $x_{15} = -10$ quite distant from the other x_{1k} . A good matching placement of the investigated section is the points $IP \approx 10$ and $EP = 10 + IP \approx 20$, with the point x_{15} neglected. This single outlier x_{15} can be a result of some distortion, but an algorithm of matching should process this value, too.

Now let us consider the problem of matching the pair (*IP*, *EP*) with different performance indices used. The Least-Squares (LS) matching corresponds to the value of $\gamma = 2$ in (2). The most intuitive least sum of absolute errors is determined for $\gamma = 1$. As the last option, will be investigated the performance index *LSE* for a small exponent of $\gamma = 1.1$.

In the case of the determination of LS matching *IP* and *EP* to the index (2), the solution will attain a minimal value of I at $IP = 8.749$ and $EP = 18.749$. In the case of LSA value matching, the best fitting will yielded by the values $IP = 9.9$ and $EP = 19.9$ and will correspond to the above-mentioned intuitive solution. For *LSE*, the best matching will be produced by the solution $IP = 9.872$ and $EP = 19.872$, quite close to the *LSA* solution, but not exactly the same. Now let us observe variations of $I(x)$ as functions of $x = IP$, ($EP = x + 10$) for different values of the parameter ω , Fig. 1.

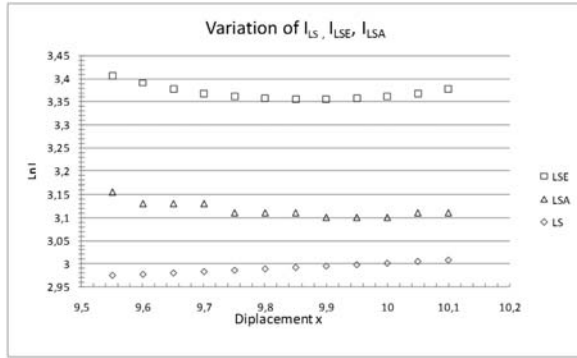


Fig. 1. Variations of performance indices I_{LS} , I_{LSA} and I_{LSE} in a logarithmic scale.

For the common presentation of trends of the compared indices, the values of $0.5 \cdot \ln(I_{LS})$, $\ln(I_{LSA})$ and $\ln(I_{LSE})$ in Fig. 1. The plots reflect some basic problems of matching with application of different methods, LS , LSA and LSE . LS matching is shifted quite apart from the intuitive solution ($IP = 10, EP = 20$) and the index I_{LS} slowly goes down and attains the minimal value for $IP = 8.749$, but the corresponding algorithm for I_{LS} minimization is very easy to implement. The most intuitive matching is attained by the LSA method, but it has a very serious drawback—the index value I_{LSA} is constant within intervals of argument x , as shown in Fig. 1. This feature, reflecting the fact that the LSA algorithm can be performed only in an iterative way, represents problems that arise in numerical calculations (Fedorov, 1994). The minimization of the LSA index yields intervals vanishing the numerical derivative. The plots, presented in Fig. 1, yield a conclusion that the best result for the LSE algorithm is quite close to the LSA matching, but the variation of I_{LSE} is a continuous function of the argument x with a unique minimum and therefore is more flexible in numerical processing.

The next section presents the formal statement of the approximation problem of fitting a transfer function to the known frequency characteristic. Derivation of the numerical approximation algorithm is based on reformulation of some nonlinear problem, resulting from matching complex values of the frequency characteristic for the investigated approximation F_a to the pattern characteristic FR . The approximation problem is transformed to minimization of a measure of errors between patterns $|FR - F_a|$ defined by different performance indices— LS , LSA or LSE . In Section 3 a simple theorem for the development of an effective way of calculation of solutions in the LSA and LSE sense is shown, followed by propositions of numerical algorithms for the approximation. The numerical calculations can be performed in a batch-processing scheme or can be based on the known, very old, recursive scheme, formerly developed by Young (1966), and next used in efficient

on-line recursive algorithms, employed for parameter identification (e.g., Ljung and Söderström, 1987). The recursive version seems to be more flexible and easy launching numerical calculations. As shown in Section 4, the introduced approximation approach based on the pseudo-inverse scheme can be used for solving different tasks. For the verification and the efficiency test of the proposed approach, we present problems where the pattern is well known and the expected results are more or less intuitive: reduction of the model order for a discrete-time SISO process and approximation of the model in a continuous-time domain. The results of application of LSE performance in the case of data with outliers compared with LS results close the set of examples. The conclusions contain a summary of investigations and are focused on new applications of the proposed approach.

2. Problem statement

The set $FR = \{c_k, k = 1 \dots, N\}$ of the estimated discrete-time frequency response, defined on the set $\Omega = \{\omega_k, k = 1, \dots, N\}$, represents dynamics of some SISO process. We are looking for a linear transfer function $G_a(s)$, of order n , that is defined by the relation

$$G_a(s) = \frac{B(s)}{A'(s)} = \frac{b_0 + b_1 s + \dots + b_{n-1} s^{n-1}}{1 + a_1 s + \dots + a_n s^n}, \quad (3)$$

where the coefficients b_i , a_i are not known. The frequency characteristic for the approximation $G_a(s)$ will be determined as

$$F_a(\omega_k) = G_a(s)|_{s=l\omega_k}, \quad \omega_k \in \Omega, \quad (4)$$

or, in the case of investigating a process controlled by a zero-order sampling unit with the transfer function $I(s)$, by

$$F_a(\omega_k) = \{G_a(s)I(s)\}|_{s=j\omega_k}, \quad \omega_k \in \Omega. \quad (5)$$

The measure of the approximation error is defined by the relation (1) with application of an index adjusted to the complex character of the frequency characteristic,

$$\begin{aligned} I_\alpha^\gamma &= \sum_{k=1}^N \alpha_k \{ [\text{Re} c_k - \text{Re} F_a(\omega_k)]^\omega \\ &\quad + [\text{Im} c_k - \text{Im} F_a(\omega_k)]^\gamma \} \\ &= \sum_{k=1}^N \|c_k - F_a(\omega_k)\|_\alpha^\omega, \\ \omega_k &\in \Omega, \quad \alpha_k \geq 0, \quad \gamma \in [1, 2]. \end{aligned} \quad (6)$$

The approximation aim is to determine a set of coefficients b_i , a_i (3), which attain a minimum of the performance index I_α^γ .

Let us denote by ζ_k the value of the frequency response of G_a for ω_k

$$G_a(j\omega_k) = \frac{b_0 + b_1 s_k + \dots + b_{n-1} s_k^{n-1}}{1 + a_1 s_k + \dots + a_n s_k^n} \Big|_{s_k = j\omega_k} = \zeta_k, \quad k = 1, \dots, N. \quad (7)$$

The complex value ζ_k depends on polynomials A and B defined for complex frequency s_k ,

$$\begin{aligned} \zeta_k(1 + A(s_k)) &= B(s_k) \\ \Rightarrow \zeta_k &= B(s_k) - \zeta_k A(s_k), \quad k = 1, \dots, N. \end{aligned} \quad (8)$$

The error $e_k = c_k - \zeta_k \in \mathbb{C}$ of the approximation is

$$e_k = c_k - \zeta_k = c_k - B(s_k) + \zeta_k A(s_k). \quad (9)$$

Now let us introduce some notation useful in the next steps of the algorithm:

$$\begin{aligned} B(s_k) &= \sum_{i=0}^{n-1} b_i s_k^i = B S_k, \\ B &= [b_0, b_1, \dots, b_{n-1}], \\ S_k &= \begin{bmatrix} 1 \\ s_k \\ \vdots \\ s_k^{n-1} \end{bmatrix}, \\ A(s_k) &= \sum_{i=1}^n a_i s_k^i = A S_k S_k, \\ A &= [a_1, a_2, \dots, a_n] \\ A, B &\in \mathbb{R}^n, \quad S_k \in \mathbb{C}^n. \end{aligned} \quad (10)$$

The unknown coefficients vectors A, B are real, but the terms of the measure I_α^γ in (6) are in the complex domain. Let us distinguish the real and imaginary parts of c_k and vectors representing the values of $B(s_k)$ and $\gamma_k A(s_k)$ in the relation (9) as

$$\begin{aligned} r_k &= \operatorname{Re} c_k, \quad i_k = \operatorname{Im} c_k \\ RB_k &= \begin{bmatrix} 1 \\ \operatorname{Re} s_k \\ \vdots \\ \operatorname{Re} s_k^{n-1} \end{bmatrix}, \quad IB_k = \begin{bmatrix} 0 \\ \operatorname{Im} s_k \\ \vdots \\ \operatorname{Im} s_k^{n-1} \end{bmatrix}, \\ RA_k &= \begin{bmatrix} \operatorname{Re}(s_k \zeta_k) \\ \operatorname{Re}(s_k^2 \zeta_k) \\ \vdots \\ \operatorname{Re}(s_k^n \zeta_k) \end{bmatrix}, \quad IA_k = \begin{bmatrix} \operatorname{Im}(s_k \zeta_k) \\ \operatorname{Im}(s_k^2 \zeta_k) \\ \vdots \\ \operatorname{Im}(s_k^n \zeta_k) \end{bmatrix}. \end{aligned} \quad (11)$$

The perfect fitting of the frequency characteristic of $G_a(s)$ to the pattern values c_k will yield two conditions: $r_k \cong BRB_k - ARA_k, i_k \cong BIB_k - AIA_k$. The measure

of the fitting error of the investigated approximation can be represented by the index

$$\begin{aligned} I_\alpha^\gamma(\Theta) &= \sum_{k=1}^N \alpha_k \{ |r_k - BRB_k + ARA_k|^\gamma \\ &\quad + |i_k - BIB_k + AIA_k|^\gamma \} \\ &= \sum_{k=1}^N \alpha_k \{ |r_k - \Theta R_k|^\gamma + |i_k - \Theta I_k|^\gamma \}, \\ \Theta &= [B, A] \in \mathbb{R}^{2n}, \\ R_k &= \begin{bmatrix} RB_k \\ -RA_k \end{bmatrix} \in \mathbb{R}^{2n}, \\ I_k &= \begin{bmatrix} IB_k \\ -IA_k \end{bmatrix} \in \mathbb{R}^{2n}. \end{aligned} \quad (12)$$

The minimization of the index $I_\alpha^\gamma(\Theta)$ for N values of frequency ω involves the problem of minimization in $2N$ points, due to the complex character of values c_k and the form of the index (12). The algorithm for minimization depends on unknown values of ζ_k (9), which have to be estimated after determination of the model coefficients vector Θ . Hence, the minimization algorithm has to be arranged in an iterative way.

The minimization of

$$I_\alpha^\gamma(\Theta) = \|V\Theta - Y\|_\alpha^\gamma \rightarrow \min_{\Theta},$$

$$V = \begin{bmatrix} R_1 \\ I_1 \\ \vdots \\ R_N \\ I_N \end{bmatrix}, \quad Y = \begin{bmatrix} r_1 \\ i_1 \\ \vdots \\ r_N \\ i_N \end{bmatrix},$$

$$\alpha = \operatorname{diag}[\alpha_1, \alpha_2, \dots, \alpha_N], \quad R_i, I_i \in \mathbb{R}^{2n}, \quad (13)$$

can be arranged in different ways. In the paper it is assumed that the resulting transfer function coefficient will yield a frequency response close to the known values of c_k , and therefore in the first step the values ζ_k were replaced by c_k in (11). The next runs of the algorithm have to be iterative, with determination of ζ_k calculated as in (7). A batch-processing algorithm is proposed as Algorithm 1.

The start of a minimization algorithm for $I_\alpha^2(\Theta_0)$ is simple. The LS approximation is direct and after basic calculations can be solved as shown in (14), where $j = 0$ is the index of the iteration

$$\begin{aligned} \Theta_j^{LS} &= [V^{jT} \alpha V^j]^{-1} V^{jT} \alpha Y, \quad \alpha \in \mathbb{R}^{N \times N}, \\ V^j &= \begin{bmatrix} R_1^j \\ I_1^j \\ \vdots \\ R_N^j \\ I_N^j \end{bmatrix}, \end{aligned} \quad (14)$$

Algorithm 1. Batch-processing.

Step 0. At the start of iterations, the initial guess of ζ_k^0 is substituted by c_k , $k = 1, \dots, N$. For the known frequency values ω_k the vectors RB_k, IB_k, RA_k, IA_k in (11) can be calculated and the estimate Θ_0 can be determined by minimization of $I(\Theta)$ in (13).

Step 1. For the known initial component vector Θ_0 the values of ζ_k^1 , $k = 1, \dots, N$ can be recalculated, consistent with the determination (7), (11) of the vectors RB_k, IB_k, RA_k, IA_k . The next step of minimization (13) can be performed and yields the estimation of the vector of the model coefficients Θ_1 . After this step, based on (7), the estimate of the model error and the resulting performance index $I_\alpha^\gamma(\Theta_1)$ in (12) are calculated.

Step 2 and onward. For the known vector Θ_j the values of ζ_k^j , $k = 1, \dots, N$ can be calculated and the next iteration of the vectors RB_k, IB_k, RA_k, IA_k is performed. The next step of minimization of (13) $I_\alpha^\gamma(\Theta_j)$ is made and yields the next approximation of Θ_{j+1} . The minimization of (13) proceeds until the value of I_j is decreasing.

The only problem here is the convergence of the proposed algorithm. Step 0 always produces an estimate Θ_0 , which can be close to or quite distant from the pattern set FR . The process of minimization of I^2 has to be supervised. After the first 2 to 3 steady iterations of this algorithm, the procedure is convergent. However, sometimes, especially in the case of quite a small order n of the transfer function $G_a(s)$ in (7), the algorithm may converge to a false solution. In the presence of big outliers in the pattern plot of c_k , this effect was observed, too.

Now we will discuss algorithms for effective calculation with non-square performance indices, with exponent $\gamma < 2$. These algorithms are not based on direct rules as in (14) and need a recursive way of calculations.

3. Minimization of LSA and LSE error measures

Let us first consider the well-known problem of minimization in the sense of the weighted minimal sum of error squares of the linear regression model

$$\hat{Y} = V\Theta, \quad \hat{Y} \in \mathbb{R}^M, \quad V \in \mathbb{R}^{M \times m}, \quad \Theta \in \mathbb{R}^m \quad (15)$$

for the vector $Y \in \mathbb{R}^M$, where V is a matrix of stored data and Θ is a vector of unknown model coefficients. The least-squares estimation for the minimization of the model error in the weighted LS sense (Ljung and Söderström,

1987),

$$I_{WLS} = \{(Y - V\Theta)^T W (Y - V\Theta)\}, \\ W \in \mathbb{R}^{M \times M}, \\ \text{diag}[w_1, \dots, w_M], \quad w_i \geq 0, \quad (16)$$

where W is the matrix of weights, is defined by the well-known formula (Ljung and Söderström, 1987)

$$\Theta_{WLS} = [V^T W V]^{-1} V^T W Y. \quad (17)$$

Now let us consider a special form of the WLS estimator with dynamic weight coefficients w , dependent on the error volume produced by the resulting model.

Theorem 1. Consider a linear form

$$\hat{Z} = X\psi, \quad \hat{Z} \in \mathbb{R}^M, \quad X \in \mathbb{R}^{M \times m}, \quad \psi \in \mathbb{R}^m, \quad (18)$$

which is a model of a variable z , presented by the vector $Z = [z_1, z_2, \dots, z_M]^T$ and X is a matrix composed of row vectors $x_k \in \mathbb{R}^m$, $k = 1, \dots, M$. If $\psi'' \in \mathbb{R}^m$ denotes a vector of the model coefficients, which attains a minimum of the performance index

$$I = \sum_{k=1}^M \|z_k - x_k \psi''\|^\gamma, \quad x_k \in \mathbb{R}^m, \quad \gamma \geq 1, \quad (19)$$

and E is equal to

$$E = \text{diag}\{e_1, e_2, \dots, e_M\}, \\ e_k = \frac{1}{\|z_k - \hat{z}_k(\psi)\|^{2-\gamma}}, \quad k = 1, \dots, M, \\ \hat{z}_k = x_k \psi'', \quad (20)$$

then the vector ψ'' satisfies the relation

$$\psi'' = [X^T E X]^{-1} X^T E Z. \quad (21)$$

Proof. The sum $I(\psi'') = \sum_{k=1}^M |z_k - \hat{z}_k(\psi)|^\gamma$ can be presented as

$$I(\psi'') = \sum_{k=1}^M \left[(z_k - \hat{z}_k(\psi))^2 \right]^{\gamma/2}.$$

I is a convex function of ψ . A necessary condition for a minimum of $I(\psi)$ yields

$$\frac{\partial}{\partial \psi} I(\psi) \\ = \frac{\partial}{\partial \psi} \sum_{k=1}^M \left[(z_k - \hat{z}_k(\psi))^2 \right]^{\gamma/2} \\ = \sum_{k=1}^M \frac{\gamma}{2} \frac{1}{\left[(z_k - \hat{z}_k(\psi))^2 \right]^{(2-\gamma)/2}} \frac{\partial}{\partial \psi} (z_k - \hat{z}_k(\psi))^2 \\ \equiv 0.$$

This relation can be presented in the following form:

$$\frac{\partial}{\partial \psi} I = \sum_{k=1}^M e_k \frac{\partial}{\partial \psi} (z_k - \hat{z}_k(\psi))^2 \equiv 0,$$

$$e_k = \frac{1}{[(z_k - \hat{z}_k(\psi))^2]^{(2-\gamma)/2}}.$$

The above expression is equivalent to the necessary condition for a minimum of the weighted least square problem (e.g., Ljung and Söderström, 1987), where the weights are equal to $e_k = 1/\epsilon_k$, and then the coefficient vector ψ'' is determined as in (21). This completes the proof. ■

The theorem yields the condition for ψ'' , but first the vector of estimation errors $e = [e_1, e_2, \dots, e_n]$ has to be known. In case of off-line estimation, Algorithm 2 is proposed.

Algorithm 2. Off-line estimation in the LSE sense.

Step 1. Compute an estimate ψ^0 with the initial value of $e_j = 1, j = 1, \dots, M$,

$$\psi^0 = [X^T X]^{-1} X^T Z. \quad (22)$$

Step 2. Determine the absolute values of output errors ϵ_k^0 ,

$$\epsilon_k^0 = |z_k - \hat{z}_k(\psi)|^\gamma, \quad k = 1, \dots, M,$$

$$\gamma \in [1, 1.2], \quad (23)$$

and set the weighing coefficients

$$e_k^0 = \begin{cases} 1/\epsilon_k, & \epsilon_k > \epsilon_{\min}, \\ \epsilon_k = \max, & \epsilon_k \leq \epsilon_{\min}, \end{cases}$$

$$k = 1, \dots, M, \quad \max = 1/\epsilon_{\min}, \quad (24)$$

where ϵ_{\min} can be determined by quantization or measurement errors. For the initial step $j = 0$ the weight matrix is equal to

$$E^0 = \text{diag} [e_1^0, e_2^0, \dots, e_M^0]. \quad (25)$$

Step 3. Compute the estimate ψ^1 in (21),

$$\psi^1 = [X^T E^0 X]^{-1} X^T E^0 Z, \quad (26)$$

with weights e_k^1 defined as in (24).

Step 4. For the next iterations with indices $j = 2, 3, \dots$

Steps 2 and 3 are repeated with new determined model errors (23) and updated values in the weight matrix E^j (25) until the index I in (19) reaches a limit value, i.e., it is not decreasing within a determined number κ of the last runs.

Algorithm 2 is based on batch processing of recorded data. Due to the observed behaviour of the *LSA* performance index presented in the Introduction, the convergence of this iterative processing may yield some problems (Fiodorov, 1994; Janiszowski, 1998). *LSE* estimation yields better convergence properties. The only difference between *LSA* and *LSE* estimation is the value of the weights coefficients (20), and the *LSE* index can be expressed as

$$I = \sum_{k=1}^M |z_k - x_k \psi_{LSE}|^\gamma$$

$$= \sum_{k=1}^M e_k |z_k - x_k \psi_{LSE}|^2,$$

$$e_k = |z_k - x_k \psi_{LSE}|^\lambda, \quad \lambda = 2 - \gamma, \quad (27)$$

and in effect these weights e_k will be used in estimation (26) for *LSE* off-line estimation.

The batch processing algorithm yields one problem in calculations: the first estimates in *LSA* or *LSE* are received after the processing of all M recorded points.

The convergence rate can be improved when estimation for one run, i.e., M steps, is determined recursively, like in on-line schemes (Janiszowski, 1998; Kozłowski, 2003; Kowalczyk and Kozłowski, 2011). The efficiency of the recursive scheme was observed quite long ago (Young, 1966), and next its positive features were confirmed (Ljung and Söderström, 1987) and used in other applications (Janiszowski, 1998; Kozłowski, 2003; Kowalczyk and Kozłowski, 2011). A proposition of such an algorithm is shown below.

In comparison with the usually used on-line identification schemes, where the index k represents discrete time instants, in the above algorithm this index represents values of radial frequency $\omega_k \in \Omega$. The choice of these values is discussed in the next section. The proposed starting point (35) can be used or a primary fitting with the *LS* scheme for $\lambda = 2$ can be tested.

The application of the recursive calculation (28)–(35) to estimation of model coefficients has an advantage with respect to the batch processing scheme proposed by (22)–(26). From the very beginning the approximation Θ slowly but steadily converges to the expected *LSE* or *LSA* estimates, does not contain an initial phase of the ψ_{LSA}^0 estimation (22), which can push the initial result far away from the proper resolution. This advantage is accompanied by a convergence problem. The first iterations of $\psi_{LSE}(j)$ have to be well prepared, i.e., the values of $\omega_1, \dots, \omega_q, (q \ll M)$, close to 0, yield very small variations in c_k and the convergence of estimation $\psi_{LSE}(j)$ in this phase of data processing is slow and not very well conditioned. Therefore, a set of the values c_k has to be carefully selected to yield sufficient changes into

Algorithm 3. On-line estimation in the LSA sense.

Step 1. Estimate the model error for the step j of iteration,

$$\begin{aligned}\epsilon_k &= z_k - x_k \psi_{LSE}(k-1), \quad k = 1, \dots, M, \\ x_k &= [x_{1k}, x_{2k}, \dots, x_{mk}].\end{aligned}\quad (28)$$

Step 2. Determine the weight coefficient e_j ,

$$e_k = \begin{cases} 1/|\epsilon_k|^\lambda, & |\epsilon_k| \geq \epsilon_{\min}, \\ \max = 1/\epsilon_{\min}, & |\epsilon_k| < \epsilon_{\min}. \end{cases}\quad (29)$$

Step 3. Calculate the vector of the filtered model inputs and gain vector χ ,

$$v_k = e_k x_k, \quad v_k \in \mathbb{R}^m, \quad (30)$$

$$\mu_k = [\rho + v_k \mathbf{P}_{k-1} v_k^T]^{-1}, \quad \mu_k, \rho \in \mathbb{R}, \quad (31)$$

$$\chi = \mu_k \mathbf{P}_{k-1} v_k^T, \quad \chi \in \mathbb{R}^m. \quad (32)$$

Step 4. Update the model coefficient vector,

$$\begin{aligned}\psi_{LSE}(k) &= \psi_{LSE}(k-1) + \text{sign}(\epsilon_k) \chi |\epsilon_k|^{\gamma-1}, \\ \alpha_{LSE} &\in \mathbb{R}^m.\end{aligned}\quad (33)$$

Step 5. Update the estimate of the matrix \mathbf{P} ,

$$\mathbf{P}_k = [1 - \chi v_k] \mathbf{P}_{k-1} \rho^{-1}, \quad \mathbf{P}_k \in \mathbb{R}^{mm}, \quad (34)$$

where $\rho = [0.99, 1.0]$ is a forgetting factor.

Step 6. The initial conditions for the above algorithm are

$$\begin{aligned}\psi_{LSE}(0) &= [0, 0, \dots, 0], \quad \alpha_{LSE} \in \mathbb{R}^m, \\ \mathbf{P}_0 &= \text{diag}(\kappa),\end{aligned}\quad (35)$$

and $\kappa \in [10^8, 10^{20}]$ initialises the matrix (\mathbf{P}) .

estimation of c_k in the approximation process. This topic will be investigated and considered in further sections.

4. Application of the proposed algorithm for approximation and reduction of the model order for SISO systems

The problem of approximation can have many applications like, e.g., reduction of the order of discrete-time model dynamics for SISO system, approximation of the process model in a continuous time domain based on the discrete-time frequency response with reduction of the model order (or without), approximation of a model in a discrete or a continuous time domain with the filtering of the response in a

specified frequency interval, change in the sampling interval Δ in discrete-time representation and others. In all these applications the main tool is efficient approximation in the frequency domain, and the final result can be achieved by an adequate statement of the problem and proper pre-processing of the data used.

The proposed approach will be first tested for reduction of the model order of a dynamic process. In this application the expected result is usually quite obvious and the efficiency of the approximation algorithm can be verified. The quality of approximation has to be verified, and therefore the following normalized performance index is proposed:

$$I_n = \left\{ \frac{1}{M} \sum_{k=1}^M \frac{|c_k - G(j\omega_k)|^2}{|c_k|^2} \right\}^{1/2}. \quad (36)$$

This form of the performance index does not impose any frequency *a priori* due to the introduced weighing normalizing factor $|c_k|^2$ in the denominator. Its form is different from the suggested *LSA* or *LSE* form (6) but corresponds to indices used in the literature (Ljung and Söderström, 1987; Mohan *et al.*, 2004; Sreeram and Agatokhlis, 1991) and therefore can be used for comparison.

4.1. Reduction of the model order for a discrete-time linear model of a SISO process. This problem usually is yielded as a result of process identification, performed with defined sampling interval Δ . The estimated discrete-time transfer function of the SISO process,

$$\begin{aligned}G^*(z^{-1}) &= \frac{B(z^{-1})z^{-d}}{1 + A(z^{-1})} \\ &= \frac{(b_0 + b_1 z^{-1} + \dots + b_p)z^{-d}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}}, \\ T_0 &= d\Delta,\end{aligned}\quad (37)$$

with the specified delay T_0 is next investigated in the form of the lower order $r \leq p$ model (37) with the same discrete-time delay d . At the beginning a set *FR* of given frequency characteristic values has to be created for some set Ω . The set of frequency values Ω has an upper bound ω_{up} , in the given case limited by the Shannon's frequency π/δ . The initial frequency $\omega_1 = \omega_{\text{low}}$ and the following values ω_k , its number and choice is a decision to be made by the user. In our case, the value of ω_1 is set to $10^{-5} \cdot \omega_{\text{up}}$. The number of generated frequency values ω_k should be sufficiently large and is defined, in the presented examples, as $M' \gg 300$. The interval $[\omega_{\text{low}}, \omega_{\text{up}}]$ is first divided in a logarithmic scale according to the rule

$$\Delta\omega = \exp \left\{ \frac{\ln(\omega_{\text{up}}/\omega_{\text{low}})}{M'} \right\}, \quad \omega_1 = \omega_{\text{low}}. \quad (38)$$

Next, the following selection of the frequency ω value is introduced. The value $\omega_k \in \Omega$ is multiplied by $\Delta\omega$. If the new value $\omega' = \omega_k \cdot \Delta\omega$ presents a sufficient variation of argument, i.e., $\arg G^*[j(\omega')] - \arg G^*[j(\omega_k)] > \Delta \arg$, then this value ω' is included into Ω as ω_{k+1} . Otherwise, ω' is increased again, $\omega' = \omega' \cdot \Delta\omega$ and the condition for argument variation is tested. The above rule generates a set Ω of M frequency values ω_k that preserve a sufficient variation of the investigated frequency response,

$$\begin{aligned} \Omega &= \{\omega_k, k = 1, \dots, M\}, \\ FR &= \{c_k = G^*(j\omega_k), k = 1, \dots, M\}, \\ |\arg G^*(j\omega_k) - \arg G^*(j\omega_{k-1})| &> \Delta \arg, \\ k &= 1, \dots, M. \end{aligned} \quad (39)$$

As an example, we use used a model of the eight order discrete-time transfer function (Sreeram and Agatkhli, 1991),

$$G_8^* = \frac{a_8(z-1)}{b_8(z-1)} \quad (40)$$

where

$$\begin{aligned} a_8(z-1) &= 0.2012z^{-1} + 0.1395z^{-2} - 0.0263z^{-3} \\ &\quad + 0.0190z^{-4} - 0.0645z^{-5} - 0.0328z^{-6} \\ &\quad + 0.0055z^{-7} - 0.0088z^{-8}, \\ b_8(z-1) &= 1 - 0.6307z^{-1} + 0.4185z^{-2} - 0.0788z^{-3} \\ &\quad + 0.0570z^{-4} + 0.1935z^{-5} + 0.0983z^{-6} \\ &\quad - 0.0165z^{-7} + 0.0023z^{-8}. \end{aligned}$$

In the work of Sreeram and Agatkhli (1991) the reduction into a second order model yields the following transfer function:

$$G_2^*(z^{-1}) = \frac{0.23159z^{-1} - 0.15233z^{-2}}{1 - 1.75851z^{-1} + 0.83862z^{-2}}, \quad (41)$$

which, after evaluation with I_n in (36), gives the value of $I_n = 0.043$. The application of the proposed *LSE* approach, with $\gamma = 1.2$, based on $M = 289$ and points selected as in (38), (39) gives the following results for reduction to the third and second orders:

$$\begin{aligned} G_3^{LSE}(z^{-1}) &= \frac{0.21309z^{-1} - 0.21427z^{-2} + 0.04096z^{-3}}{1 - 2.21016z^{-1} + 1.63034z^{-2} - 0.37974z^{-3}} \end{aligned}$$

and the performance index I_n equals $6.1 \cdot 10^{-5}$, which is a very good result. The result for second order reduction is quite good, too,

$$G_2^{LSE}(z^{-1}) = \frac{0.23089z^{-1} - 0.15083z^{-2}}{1 - 1.75752z^{-1} + 0.83751z^{-2}}, \quad (42)$$

and is qualified with $I_n = 0.0047$. Both results present quite good fitting and show that the *LSE* estimation scheme can be satisfactorily efficient.

4.2. Approximation in the continuous time domain, based on frequency characteristics derived from a discrete-time model of a SISO process.

This problem usually appears during process identification based on sampled data, when the determined model has to be expressed in continuous time form, which is more easy to handle and evaluate. It is clear that sampled data deliver information in a limited frequency domain $[0, \omega_S]$. Quite often data are gathered in an experiment where the investigated process is controlled by a microcontroller device, with constant sampling time Δ . This fact has to be included into determination of investigated dynamics. The observed dynamic properties in the form of frequency characteristic, denoted by c'_k , are products of process and sampling unit G_s dynamics

$$\begin{aligned} c'_k &= G^*(j\omega_k) \cdot G_s(j\omega_k), \\ G_s(j\omega_k) &= \frac{1 - \exp(-j\omega_k\Delta)}{j\omega_k}. \end{aligned} \quad (43)$$

The determined values of the frequency characteristic have to be redefined at the beginning of the approximation process,

$$c_k = c' [G_s(j\omega_k)]^{-1}, \quad k = 1, \dots, M. \quad (44)$$

These transformed values c_k are next used for approximation of the transfer function for investigated continuous time representation of a SISO process.

The efficiency of the proposed approach can be presented on an example of a simple, second order linear dynamic process, described by a transfer function of the form

$$G_2(s) = \frac{20}{1 + 40s} - \frac{1}{1 + 4s} = \frac{19 + 40s}{1 + 44s + 160s^2}. \quad (45)$$

This transfer function was used for generation of a set of data with sampling interval $\Delta = 1$ s. In this example the sensitivity of approximations to the presence of outliers is tested. The process of gathering frequency characteristic data was performed

1. without any distortions (*FC*),
2. with outliers added to the initial set *FC* with more points of reduced values (*FC₋*),
3. with outliers added to the initial set *FC* with more points of increased values (*FC₊*).

All data sets *FC*, *FC₋*, *FC₊*, containing each 300 points, are generated with an algorithm satisfying conditions presented in relation (39). Next, these data

are used for approximation of the resulting transfer function. The outliers are introduced by changing the corresponding amplitudes of c_k and observing the reaction of approximation algorithm for such distortions.

In the case of undisturbed data FC and the second order of the investigated transfer function, the LS , LSE and LSA approximation results are practically the same and equal to the original transfer function $G_2(s)$ (45). But for the first order approximations for the LS , LSA and LSE schemes, the results are different:

$$\begin{aligned} G_{1LS}(s) &= \frac{19.38452}{1 + 48.30591s}, \\ G_{1LSA}(s) &= \frac{19.00002}{1 + 41.89483s}, \\ G_{1LSE}(s) &= \frac{19.00003}{1 + 41.89572s}. \end{aligned} \quad (46)$$

The LS approximation is visibly worse, which can be observed in Fig. 2, where the plots for the LS and LSA approximation, together with the original frequency characteristic FC , are shown. We can observe a difference at low frequency values ω_k : the gain of the LS approximation is too high. This effect is visible considering the form of the transfer function $G_{1LS}(s)$, too. The static gain for this transfer function is equal to 19.384. The frequency characteristic begins from 19 and quite the same value is present in approximations $G_{1LSE}(s)$ and $G_{1LSA}(s)$. The differences between these transfer functions were negligible and therefore only the frequency characteristic for $G_{1LSA}(s)$ is presented in Fig. 2. At the high frequencies ω_k , both approximations LS and LSA present the same plot, typical for the first order transfer function.

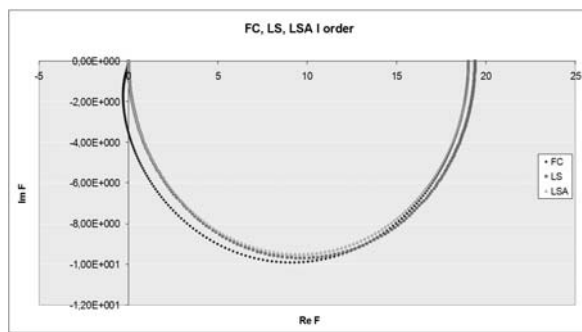


Fig. 2. Frequency characteristics for the first-order approximations determined with the LS and LSA schemes compared with the original pattern FC .

The observed difference is not the only one. In the case of outliers in the frequency pattern FC_- , can be observed, cf. some other behaviour of the approximations, Fig. 3. The set FC_- of frequency characteristics, with visible outliers, is used for the determination of the first order approximations with the LSE and LS algorithms.

In this case, the LSE approximation scheme produces quite the same effect as in the case of the original set FC . The outliers visibly destroy the variation of the LS_- approximation plot. The attraction of outliers pushed the resulting LS_- approximation quite apart from the set FC_- .

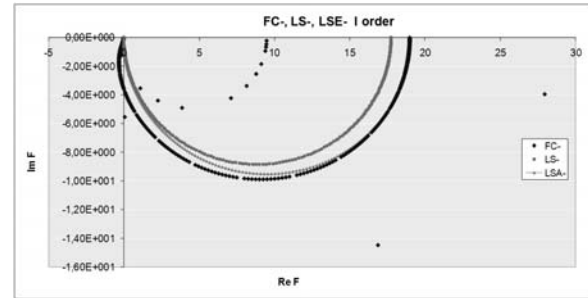


Fig. 3. Frequency characteristics for the first-order approximations LS_- , the LSA_- determined with LSA_- and LS schemes, compared with the pattern FC_- with outliers.

The case of positive gathered outliers is presented in Fig. 4. The approximations LS_+ and LSE_+ represent approximation results with application of the corresponding algorithms, but this time for the second order of the transfer function. The difference between LSE_+ and FC is hardly visible with distinctions of outliers. The LS_+ results are different—again the outliers pushed the LS_+ results in the direction of the transfer with an increased gain.

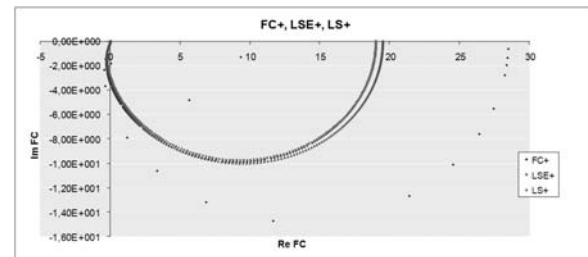


Fig. 4. Frequency characteristics for the second-order approximations LS_- , LSE_- determined with the LSE and LS schemes, compared with the pattern FC_- with outliers.

The transfer functions determined within approximation processing are equal to

$$\begin{aligned} G_{2LS+}(s) &= \frac{19.49801 + 29.6896s}{1 + 44.63748s + 118.786s^2}, \\ G_{2LSE+}(s) &= \frac{19.03859 + 28.5043s}{1 + 44.6002s + 114.020s^2}. \end{aligned} \quad (47)$$

Brief inspection of these forms confirms an increased static gain in the LS_+ approximation and practically the proper value in the LSE_+ approximation result. The index

(36) was not calculated because of the presence of the outliers. However, comparison of the presented figures shows that the LSE_+ approximation covers practically the original plot FC_+ excluding the outliers. LS_+ is close but reflects a small displacement in the direction of the outliers.

5. Conclusions

The approach presented in this paper is based on the approximation algorithm with the option of minimization of the non-quadratic error measure. The results shown confirm that both aspects can be combined in one algorithm and deliver good results. The primary iterative working batch processing algorithm (7)–(13) considered presents worse convergence than the recursive version (26)–(34). This effect can be a result of recursive scheme flexibility. The solution is in fact improved in each step of the recursive processing. Interesting is the fact that the recursive algorithm, invented initially for statistic estimation of dynamic models using measured time series, ca. 50 years ago (Young, 1966), can be now successfully used for calculation of some other problems, like recursive approximation. However, it should be stressed that positive results can be obtained only after special processing (38)–(39) of data. This choice of the frequency set Ω and the set of the corresponding complex values FR (39) can be completed by a stochastic draw of pairs $\{\omega_k, c_k\}$ for the recursive algorithm (28)–(35). This procedure does not present unstable solutions mentioned in Section 3. Future works on the proposed algorithm will consider application in the case of weakly nonlinear continuous time systems, e.g., a chemical reactor with dynamics dependent on operation point.

Recursive fitting, based on the LSE ($\gamma = 1.2$) performance index I_Ω^γ is quite effective and very close to LSA approximation. This observation, shown in Example 2, was confirmed for different investigated cases. In all cases the LSE recursive approximation procedure was stable and produced reasonable results. Introduction of a variable value of the exponent γ into the performance index created an additional degree of freedom into algorithms, based on minimization of performance indices as introduced in (1) or (12).

The proposed approach can create a new way for solving problems where non-quadratic criteria are reasonable and shall be used, especially where outliers create problems of LS approach convergence.

Acknowledgment

The paper was partially supported by a grant of the Dean of the Mechatronic Department of the Warsaw University of Technology for 2012.

References

- Deschrijver, D., Gustavsen, B. and Dhaene, T. (2007). Advancements in iterative methods for rational approximation in the frequency domain, *IEEE Transactions on Power Delivery* **22**(3): 1633–1642.
- Deschrijver, D., Knockaert, L. and Dhaene, T. (2010). Improving robustness of vector fitting to outliers in data, *IEEE Electronics Letters* **46**(17): 1200–1201.
- Deschrijver, D., Knockaert, L. and Dhaene, T. (2011). Robust macromodeling of frequency responses with outliers, *15th IEEE Workshop on Signal Propagation on Interconnects (SPI), Naples, Italy*, pp. 21–24.
- Fedorov, E. (1994). Least absolute values estimation: Computational aspects, *IEEE Transactions on Automatic Control* **39**(3): 626–630.
- Grivet-Talocia, S., Bandinu, M. and Canavero, F. (2005). An automatic algorithm for equivalent circuit extraction from noisy frequency responses, *2005 International Symposium on Electromagnetic Compatibility, EMC 2005, Zurich, Switzerland*, Vol. 1, pp. 163–168.
- Gustavsen, B. (2004). Wide band modeling of power transformers, *IEEE Transactions on Power Delivery* **19**(1): 414–422.
- Gustavsen, B. (2006). Relaxed vector fitting algorithm for rational approximation of frequency domain responses, *IEEE Workshop on Signal Propagation on Interconnects, Berlin, Germany*, pp. 97–100.
- Gustavsen, B. and Mo, O. (2007). Interfacing convolution based linear models to an electromagnetic transients program, *Conference on Power Systems Transients, Lyon, France*, pp. 4–7.
- Gustavsen, B. and Semlyen, A. (1999). Rational approximation of frequency domain responses by vector fitting, *IEEE Transactions on Power Delivery* **14**(3): 1052–1061.
- Janiszowski, K. (1998). Towards least sum of absolute errors estimation, *IFAC Symposium on Large Scale Systems LSS'98, Patras, Greece*, pp. 613–619.
- Kowalczyk, Z. and Kozłowski, J. (2011). Non-quadratic quality criteria in parameter estimation of continuous-time models, *IET Control Theory & Applications* **5**(13): 1494–1508.
- Kozłowski, J. (2003). Nonquadratic quality indices in estimation, approximation and control, *IEEE Conference MMAR'2003, Międzyzdroje, Poland*, pp. 277–282.
- Levy, E.C. (1959). Complex-curve fitting, *IRE Transactions on Automatic Control* **AC-4**(1): 37–43.
- Lima, A.C.S., Fernandes, A. and Carneiro, S., J. (2005). Rational approximation of frequency domain responses in the s and z planes, *IEEE Power Engineering Society General Meeting, San Francisco, CA, USA*, Vol. 1, pp. 126–131.
- Ljung, L. and Söderström, T. (1987). *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA.
- Mohan, R., Choi, M.J., Mick, S., Hart, F., Chandrasekar, K., Cangelaris, A., Franzon, P. and Steer, M. (2004).

- Causal reduced-order modeling of distributed structures in a transient circuit simulator, *IEEE Transactions on Microwave Theory and Techniques* **52**(9): 2207–2214.
- Pintelon, R. and Schoukens, J. (2004). *System Identification: A Frequency Domain Approach*, John Wiley & Sons, New York, NY.
- Sreeram, V. and Agatokhlis, P. (1991). Model reduction of linear discrete-time systems via impulse response Gramians, *International Journal on Control* **53**(1): 129–144.
- Unbehauen, H. and Rao, G. (1997). Identification of continuous-time systems: A tutorial, *11th IFAC Symposium on System Identification, Kitakyushu, Japan*, pp. 1023–1049.
- Varricchio, S., Gomes, S. and Martins, N. (2004). Modal analysis of industrial system harmonics using the s-domain approach, *IEEE Transactions on Power Delivery* **19**(3): 1232–1237.
- Wahlberg, B. and Mäkilä, P. (1996). On approximation of stable linear dynamical systems using Laguerre and Kautz functions, *Automatica* **32**(5): 693–708.
- Young, P.C. (1966). Process parameter estimation and self adaptive control, in P.H. Hammnood (Ed.), *Theory of Self Adaptive Control Systems*, Vol. 1, Plenum Press, New York, NY, p. 118.



Krzysztof B. Janiszowski, a leader of the Activators for Automation and Robotics Division at the Mechatronics Department of the Warsaw University of Technology, specializes in identification, modeling and control techniques. His application areas include identification of industrial plants, algorithms for control and fast prototyping in fluidic servo-mechanisms, models of human circulation systems, software packages for modeling and control of industrial processes.

Received: 7 December 2012

Revised: 5 July 2013

Re-revised: 12 October 2013