Fast Model Order Selection of the Nonlinearity in Hammerstein Systems

Grzegorz Mzyk
Institute of Computer Engineering, Control and Robotics, Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, e-mail: grzegorz.mzyk@pwr.wroc.pl

***Abstract*—In the paper we show a new three-stage algorithm identifying the Hammerstein system nonlinearity. The algorithm is designed to work when a poor a priori knowledge is available and when the measurement data set is small. In the first stage, a deconvolution routine is applied to output signal in order to diminish a (usually) harmful influence of the linear dynamic component. Such filtered output is then used in a standard nonparametric estimate (be it kernel or orthogonal series one) to recover the unknown characteristics. Finally, the results of nonparametric estimation are used in the algorithm selecting the best parametric model. The entire proposed approach is illustrated by the simulation example.**

***Keywords*—Nonparametric identification, structure detection, Hammerstein system, kernel regression, orthogonal series expansion, inverse filtering.**

I. Introduction

In the paper we identify the nonlinear static characteristic in Hammerstein system under lack of any prior knowledge and for moderate number of measurements. The situation is often met in practise and the selection of appropriate model is of fundamental meaning ([8], [19], [2], [12]). The classical approach (e.g. least squares) is connected with the risk of bad parametrization, which leads to the systematic approximation error. The parameters of both subsystems are aggregated (see e.g. [1]) and jointly estimated. It usually leads to complicated and badly conditioned numerical procedures. On the other hand, recent ideas of nonparametric regression can be applied without prior assumptions, but the convergence is relatively slower, because the signal produced by the dynamic component of Hammerstein system is treated as 'system noise'.

The paper is based on the results concerning combined parametric-nonparametric approach to block-oriented system identification, presented widely in [14], [15], [16], [13], [21], [22], and [24]. The main goal of the paper is to present the ready-for-use version of the identification procedure, which works with poor prior knowledge of the system, and is based on the collected input-output measurements only. In the proposed scheme, cooperation between parametric and nonparametric method allows to solve typical problems mentioned above. It consists of three stages:

- inverse filtering of the output y_k, with the parameters of the inverse filter obtained by any well-known parametric method (e.g. cummulant optimization [4], [7], [25], Box-Jenkins approach [3]),
- nonparametric (kernel or orthogonal) estimation of the regression function in the grid of points ([12], [17]),
- computing the distance between the obtained pattern and each candidating classes ([6], [5]), and selection of the best one (nearest neighbour approach [16]).

In Section II the problem is formulated in detail and in Sections III-V the following stages of the proposed procedure are described. Finally, in the simulation example we show how to select the optimal order of polynomial model of the nonlinearity.

This research was supported by Polish National Science Centre, Grant No. N N514 700640

II. Statement of the problem

A. Hammerstein system

We consider a Hammerstein system shown in Fig. 1, and assume the following.

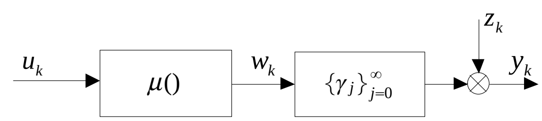

Figure 1. Hammerstein system

(A1) The nonlinear characteristic $\mu(u)$ is a Lipschitz function, i.e., it exists a positive constant $l < \infty$, such that for each $u, x \in R$ it holds $|\mu(u) - \mu(x)| \leqslant l\,|u - x|$.
(A2) The linear dynamics with the unknown impulse response $\{\gamma_i\}_{i=0}^{\infty}$ is stable and inverible (see the Appendix C).
(A3) The input $\{u_k\}$ and the noise $\{z_k\}$ are mutually independent i.i.d. random processes, $\sigma_u^2 = \text{var}u_k < \infty$, $\sigma_z^2 = \text{var}z_k < \infty$ and $Ez_k = 0$. It exists the input probability density $f(u)$.

The objective is to build the parametric model $\mu(u, c)$ of the characteristic $\mu(u)$ on the basis of the learning sequence $\{(u_k, y_k)\}_{k=1}^{N}$.

B. Regression function

Nonparametric, regression-based approach is based on the following property

$$R(u) \triangleq E\{y_k | u_k = u\} = \gamma_0 \mu(u) + \delta, \qquad (1)$$

where $\delta = E\mu(u_k) \cdot \sum_{i=1}^{\infty} \gamma_i$. By (1), the regression function $R(u)$ is the scaled and shifted version of the true static characteristic $\mu(u)$. Since the signal $\{w_k\}$ cannot be measured, the constant γ_0 is not identifiable. Hence, for clarity of exposition and without any loss of generality we can further assume that $\gamma_0 = b_0 = 1$, i.e., $R(u) = \mu(u) + \delta$. The additive constant δ can be determined only under additional knowledge, e.g., when the parametric model of the linear block is given, or when the static characteristic is known at least in one point. We also assume that $\mu(u)=0$ and hence $\mu(u)=R(u)-R(0)$ (for discussion see [14]).

III. Deconvolution of the output process

Under Assumption **(A2)** and the fact that $v_k = y_k - z_k$ we obtain, with arbitrarily small accuracy, that

$$\begin{aligned} y_k &= a_1 y_{k-1} + a_2 y_{k-2} + ... + a_p y_{k-p} + w_k + \widetilde{z}_k \quad (2)\\ &= a_1 y_{k-1} + ... + a_p y_{k-p} + Ew_k + \widetilde{w}_k + \widetilde{z}_k, \end{aligned}$$

where $\widetilde{w}_k \triangleq w_k - Ew_k$, and $\widetilde{z}_k \triangleq z_k - a_1 z_{k-1} - a_2 z_{k-2} - ... - a_p z_{k-p}$ are zero-mean stationary random processes. Equation (2) may be rewritten in the form

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + ... + a_p y_{k-p} + c + d_k, \quad (3)$$

where $c = Ew_k$ is some (not informative) constant, and

$$d_k = \widetilde{z}_k + \widetilde{w}_k \quad (4)$$

may be interpreted as a zero-mean correlated random disturbance. Introducing the regressor $\phi_k = (y_{k-1}, y_{k-2}, ..., y_{k-p}, 1)^T \in R^{(p+1)\times 1}$, and the vector of unknown parameters

$$\theta = (a_1, a_2, ..., a_p, c)^T \in R^{(p+1)\times 1}, \quad (5)$$

we get $y_k = \phi_k^T \theta + d_k$, $k = 1, 2, ..., N$, or in the compact, matrix-vector version $Y_N = \Phi_N \theta + D_N$, where $Y_N = (y_1, y_2, ..., y_N)^T \in R^{N\times 1}$, $\Phi_N = (\phi_1, \phi_2, ..., \phi_N)^T \in R^{N\times(p+1)}$, and $D_N = (d_1, d_2, ..., d_N)^T \in R^{N\times 1}$.

The parameter vector of the inverse filter (5) can be identified by the instrumental variables method

$$\begin{aligned} \widehat{\theta}_N &= (\widehat{a}_{1,N}, \widehat{a}_{2,N}, ..., \widehat{a}_{p,N}, \widehat{c}_N)^T \quad (6)\\ &= (\Psi_N^T \Phi_N)^{-1} \Psi_N^T Y_N, \\ \text{where } \Psi_N &= (\psi_1, \psi_2, ..., \psi_N)^T \in R^{N\times(p+1)}, \\ \psi_k &= (\psi_{k,1}, \psi_{k,2}, ..., \psi_{k,p}, \psi_{k,p+1})^T \in R^{(p+1)\times 1} \end{aligned}$$

is additional matrix including instruments ψ_k, which fulfill the following two standard conditions [15]

$$\det E\psi_k \phi_k^T \neq 0 \text{ and } E\psi_k d_k = 0, \quad (7)$$

and perform the following FIR output filtering

$$y_k^f = y_k - \widehat{a}_{1,N} y_{k-1} - \widehat{a}_{2,N} y_{k-2} - ... - \widehat{a}_{p,N} y_{k-p}. \quad (8)$$

In this point we also refer the reader to [4], [7], and [25], were the other deconvolution methods, which can be used alternatively to (8), are described.

IV. Nonparametric estimation of the regression function

As was mentioned above, the standard nonparametric methods ([9]-[12]) for nonlinearity recovering work completely independently of the shape of impulse response of the linear dynamics. The cost paid for simplicity, robustness and universality is a high variance of estimates. In the standard approach the Hammerstein system is treated in fact as a nonlinear static element corrupted by a correlated noise. Namely, one can specify three components of the output, i.e.,

$$y_k = \mu(u_k) + \sum_{i=1}^{\infty} \gamma_i \mu(u_{k-i}) + z_k. \quad (9)$$

The most part of the signal y_k is in a sense wasted, because the "system noise" $\xi_k \triangleq \sum_{i=1}^{\infty} \gamma_i \mu(u_{k-i})$ produced by linear dynamics is treated as an additional disturbance. In the Sections IV-A and IV-B the most important properties of the kernel regression estimates and orthogonal series expansion estimation methods are reminded.

A. Kernel method

The kernel estimate based on the filtered data has the form [18][12]

$$\widehat{\mu}_N(u) = \widehat{R}_N(u) - \widehat{R}_N(0), \quad \widehat{R}_N(u) = \frac{\sum_{k=1}^N y_k^f K\left(\frac{u_k - u}{h_N}\right)}{\sum_{k=1}^N K\left(\frac{u_k - u}{h_N}\right)}, \quad (10)$$

where h_N is a bandwidth parameter, such that

$$h_N \to 0 \text{ and } Nh_N \to \infty, \text{ as } N \to \infty, \quad (11)$$

and $K()$ is a kernel function, which fulfils the following conditions

$$K(x) \geqslant 0, \ \sup K(x) < \infty \text{ and } \int K(x)dx < \infty. \quad (12)$$

B. Orthogonal series expansion method

Similarly, the filtered output can be involved in the orthogonal expansion procedure. Denoting $R(u) = g(u)/f(u)$, where $g(u) \triangleq R(u)f(u)$, let $\{\varphi_i(u)\}_{i=0}^{\infty}$ be the complete set of orthonormal functions in the input domain. For square-integrable $g()$ and $f()$ one can write $g(u) = \sum_{i=0}^{\infty} \alpha_i \varphi_i(u)$, $f(u) = \sum_{i=0}^{\infty} \beta_i \varphi_i(u)$, where $\alpha_i = Ey_k \varphi_i(u_k)$ and $\beta_i = E\varphi_i(u_k)$, are orthogonal series representations of $g(u)$ and $f(u)$ in the basis $\{\varphi_i(u)\}_{i=0}^{\infty}$. The standard estimates of the coefficients α_i's and β_i's are $\widehat{\alpha}_{i,N} = \frac{1}{N}\sum_{k=1}^N y_k^f \varphi_i(u_k)$, and $\widehat{\beta}_{i,N} = \frac{1}{N}\sum_{k=1}^N \varphi_i(u_k)$, which leads to the following ratio estimate of $\mu(u)$

$$\widehat{\mu}_N(u) = \widehat{R}_N(u) - \widehat{R}_N(0), \quad \widehat{R}_N(u) = \frac{\sum_{i=0}^{q(N)} \widehat{\alpha}_{i,N} \varphi_i(u)}{\sum_{i=0}^{q(N)} \widehat{\beta}_{i,N} \varphi_i(u)}, \quad (13)$$

where $q(N)$ is some cut-off level [9].

V. Selection of the best model

Assume that we have M candidates for the parametric class of model

$$\left\{ \overline{R}^{(1)}(u, \theta_1) \right\}, \ \left\{ \overline{R}^{(2)}(u, \theta_2) \right\}, ..., \ \left\{ \overline{R}^{(M)}(u, \theta_M) \right\} \quad (14)$$

and the weights/costs $\varkappa_1, \varkappa_2, ..., \varkappa_M$, representing priorities of these models (e.g. dependent on the computational complexity). For the grid of N_0 fixed/selected points $\overline{u}_1, \overline{u}_2, ..., \overline{u}_{N_0}$, (e.g. uniformly distributed on the input domain) we compute nonparametric estimates $\widehat{R}_N(\overline{u}_1), \widehat{R}_N(\overline{u}_2), ..., \widehat{R}_N(\overline{u}_{N_0})$.

Next, for each class $l = 1, 2, ..., M$ minimize the loss function (see e.g. [20] and [15])

$$\widehat{Q}_l(\theta_l) = \sum_{i=1}^{N_0} \left(\widehat{R}_N(\overline{u}_i) - \overline{R}^{(l)}(\overline{u}_i, \theta_l) \right)^2, \qquad (15)$$

with respect to θ_l, providing

$$\widehat{\theta}_l^* = \arg\min_{\theta_l} \widehat{Q}_l(\theta_l). \qquad (16)$$

The natural choice is to select the 'nearest neighbour' model, i.e.

$$\left\{ \overline{R}^{(l_0)}(u, \theta_{l_0}) \right\}, \text{ such that } l_0 = \arg\min_l \varkappa_l \widehat{Q}_l(\widehat{\theta}_l^*). \qquad (17)$$

VI. Numerical example

To illustrate the behaviour of the method we simulated the following Hammerstein system:

$$\mu(u) = \text{sgn}(u) + |u|, \qquad v_k = 0.9v_{k-1} + w_k. \qquad (18)$$

The random input u_k and random noise z_k were uniformly distributed on $[-1, 1]$ and $[-0.1, 0.1]$, respectively. The instrumental variables were generated according to $\psi_k = (u_{k-1}, u_{k-2}, ..., u_{k-p}, 1)^T$. In the kernel-type estimation algorithm (10) we applied the window kernel $K(x) = \begin{cases} 1, \text{ as } |x| < 1 \\ 0, \text{ elsewhere} \end{cases}$, and set $h(N) \sim N^{-1/5}$. We implemented competition between polynomial models

$$\overline{R}^{(l)}(u, \theta_l) = p_0 + p_1 u + ... + p_l u^l, \qquad (19)$$

with various orders. Each model was assigned with the penalty $\varkappa_l = l$. The results ($l_0 = 3$) are presented in Fig. 2 and 3.

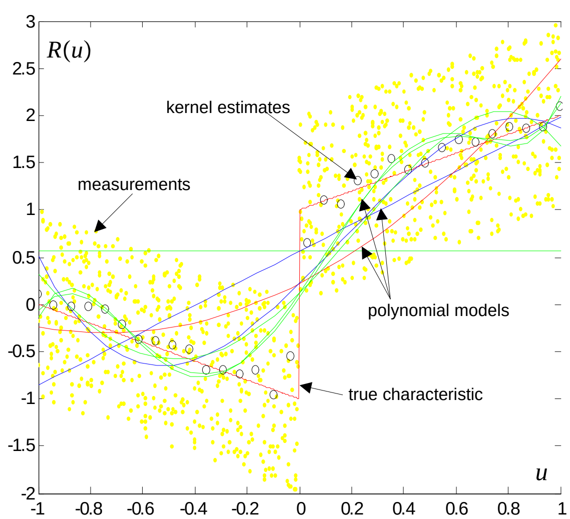

Figure 2. Results of identification for various models

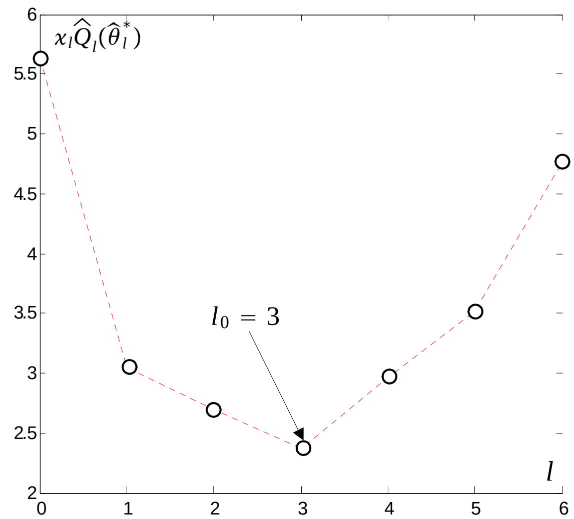

Figure 3. The cost function versus the order of polynomial model

VII. Conclusions

The combined parametric-nonparametric methods give promising results for a broad class of problems. The set of candidating classes can be selected arbitrarily. Moreover, since the decision is made on the basis of nonparametric estimates computed in relatively slow number of points, storing all the measurements in the memory is not necessary. Computation complexity of the nonlinear optimization task can be simplified by smart selection of the estimation points (i.e. the random learning points are replaced by the deterministic ones). This fact can have fundamental meaning if the identified nonlinear characteristic is not linear with respect to parameters.

As regards the open problems left for future consideration, the following generalizations seems to be of the highest importance:

- identification in the correlated input case, and
- identification of time-varying nonlinearities.

Appendix

A. Nonparametric descriptions of linear dynamics

Generally, the linear subsystem cannot be described with use of finite number of parameters. Each discrete-time linear dynamic element has the following two equivalent descriptions [3]

$$v_k = w_k + \gamma_1 w_{k-1} + ... = w_k + \sum_{i=1}^{\infty} \gamma_i w_{k-i}, \qquad (20)$$

$$v_k = w_k + \kappa_1 v_{k-1} + ... = w_k + \sum_{i=1}^{\infty} \kappa_i v_{k-i}. \qquad (21)$$

TABLE I
CONDITIONS FOR STABILITY AND INVERTIBILITY.

	stable	invertible
MA(s)	always	zeros of $B(z^{-1})$ outside the circle
AR(p)	zeros of $A(z^{-1})$ outside the circle	always
ARMA(p,s)	zeros of $A(z^{-1})$ outside the circle	zeros of $B(z^{-1})$ outside the circle

B. Parametric models of linear dynamics

Definition 1: The system is called *MA(s)* with parameters $\{b_i\}_{i=1}^{s}$ if $\gamma_i = b_i$ for $i \leqslant s$, and $\gamma_i = 0$ for $i > s$, i.e.,

$$\begin{aligned} v_k &= w_k + \gamma_1 w_{k-1} + ... + \gamma_s w_{k-s}, \quad (22) \\ v_k &= w_k + b_1 w_{k-1} + ... + b_s w_{k-s} = B(z^{-1}) w_k, \\ B(z^{-1}) &= 1 + b_1 z^{-1} + ... + b_s z^{-s}. \end{aligned}$$

Definition 2: The system is called *AR(p)* with parameters $\{a_i\}_{i=1}^{p}$ if $\kappa_i = -a_i$ for $i \leqslant p$, $\kappa_i = 0$ for $i > p$, i.e.,

$$\begin{aligned} v_k &= w_k + ... + \kappa_p v_{k-p} \\ v_k + a_1 v_{k-1} + ... + a_p v_{k-p} &= A(z^{-1}) v_k = w_k \quad (23) \\ A(z^{-1}) &= 1 + a_1 z^{-1} + ... + a_p z^{-p} \end{aligned}$$

Definition 3: The system is called *ARMA(p,s)* if

$$A(z^{-1}) v_k = B(z^{-1}) w_k. \quad (24)$$

C. Stability and invertibility

Definition 4: The system is *stable* if and only if $\sum_{i=1}^{\infty} |\gamma_i| < \infty$.

Definition 5: The system is *invertible* if and only if $\sum_{i=1}^{\infty} |\kappa_i| < \infty$.

Remark 1: It holds that $\lim_{i\to\infty} \gamma_i = 0$ if the system is stable, and $\lim_{i\to\infty} \kappa_i = 0$ if the system is invertible.

Remark 2: If the linear system is stable, then it can be approximated with arbitrarily small approximation error by the *MA(s)* model.

Remark 3: If the linear system is invertible, then it can be approximated with arbitrarily small approximation error by the *AR(p)* model.

D. List of symbols

N – number of the input-output data $\{(u_k, y_k)\}_{k=1}^{N}$
N_0 – number of the estimation points in nonparametric stage
M – number of the candidating classes of models
$\mu()$ – identified nonlinear characteristic of the static block
γ – impulse response of the linear dynamic block
$R()$ – regression function in the Hammerstein system
a – vector of parameters of the inverse filter
$f()$ – input probability density function
$K()$ – kernel function
h – bandwidth parameter
θ_l^* – parameters of the l-th approximation model
$Q_l()$ – cost function of the l-th model

REFERENCES

[1] E. W. Bai, ”An optimal two-stage identification algorithm for Hammerstein-Wiener nonlinear systems”, *Automatica*, vol. 34, No. 3, pp. 333-338, 1998.

[2] E. W. Bai, M. Fu, ”A blind approach to Hammerstein model identification”, *IEEE Transactions on Signal Processing*, vol. 50, No. 7, pp. 1610-1619, 2002.

[3] G. Box, G. Jenkins, *Time Series Analysis – Forecasting and Control*, Holden-Day, San Francisco, 1976.

[4] J. A. Cadzow, ”Blind deconvolution via cumulant extrema”, *IEEE Signal Processing Magazine*, vol. 13, pp. 24-42, May 1996.

[5] R. A. DeVore, "Nonlinear approximation", *Acta numerica*, Cambridge University Press, pp. 51-150, 1998.

[6] R. A. DeVore, G. G. Lorenz, *Constructive Approximation*, Springer Grundlehren, vol. 303, 1993.

[7] C. C. Feng, C. Y. Chi, ”Performance of cumulant based inverse filters for blind deconvolution”, *IEEE Transactions on Signal Processing*, vol. 47, No. 7, pp. 1922-1935, 2002.

[8] G. B. Giannakis, E. Serpedin, ”A bibliography on nonlinear system identification”, *Signal Processing*, vol. 81, pp. 533-580, 2001.

[9] W. Greblicki, ”Non-parametric orthogonal series identification of Hammerstein systems”, *International Journal of Systems Science*, vol. 20, No. 12, pp. 2355-2367, 1989.

[10] W. Greblicki, M. Pawlak, ”Identification of discrete Hammerstein systems using kernel regression estimates”, *IEEE Transactions on Automatic Control*, vol.31, pp. 74-77, 1986.

[11] W. Greblicki, M. Pawlak, ”Cascade non-linear system identification by a non-parametric method”, *International Journal of Systems Science*, vol. 25, No. 1, pp. 129-153, 1994.

[12] W. Greblicki, M. Pawlak, *Nonparametric System Identification*, Cambridge University Press, Cambridge, 2008.

[13] W. Greblicki, G. Mzyk, "Semiparametric approach to Hammerstein system identification", In: Proceedings of the 15th IFAC Symposium on System Identification, pp. 1680–1685, Saint-Malo, France, 2009.

[14] Z. Hasiewicz, G. Mzyk, ”Combined parametric-nonparametric identification of Hammerstein systems”, *IEEE Transactions on Automatic Control*, vol. 49, pp. 1370-1376, 2004.

[15] Z. Hasiewicz, G. Mzyk, ”Hammerstein system identification by nonparametric instrumental variables”, *International Journal of Control*, vol. 82, No. 3, pp. 440-455, 2009.

[16] Z. Hasiewicz, G. Mzyk, P. Śliwiński, Quasi-parametric recovery of Hammerstein system nonlinearity by smart model selection, *Lecture Notes in Computer Science. Lecture Notes in Artificial Intelligence*, vol. 6114, pp. 34-41, 2010.

[17] Z. Hasiewicz, M. Pawlak, P. Śliwiński, ”Nonparametric identification of nonlinearities in block-oriented systems by orthogonal wavelets with compact support”, *IEEE Transactions on Circuits and Systems. I Regular Papers*, vol. 52, No. 2, pp. 427-442, 2005.

[18] W. Härdle, *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, 1990.

[19] I. W. Hunter, M. J. Korenberg, ”The identification of nonlinear biological systems: Wiener and Hammerstein cascade models”, *Biological Cybernetics*, vol. 55, pp. 135-144, 1986.

[20] R. I. Jennrich, "Asymptotic Properties of Non-Linear Least Squares Estimators", *Ann. Math. Statist., vol.* 40, No. 2, pp. 633-643, 1969.

[21] G. Mzyk, ”Generalized kernel regression estimate for the identification of Hammerstein systems”, *International Journal of Applied Mathematics and Computer Science*, vol. 17, No. 2, pp. 101-109, 2007.

[22] G. Mzyk, "Nonlinearity recovering in Hammerstein system from short measurement sequence", *IEEE Signal Processing Letters*, vol. 16, No. 9, pp. 762-765, 2009.

[23] P. Stoica, T. Söderström, ”Instrumental variable methods for system identification”, *Circuits Systems Signal Processing*, vol. 21, No. 1, pp. 1-9, 2002.

[24] P. Śliwiński, J. Rozenblit, M. W. Marcellin, R. Klempous, "Wavelet amendment of polynomial models in nonlinear system identification", *IEEE Transactions on Automatic Control*, vol. 54, No. 4, pp. 820–825, 2009.

[25] J. K. Tugnait, ”Estimation of linear parametric models using inverse filter criteria and higher order statistics”, *IEEE Transactions on Signal Processing*, vol. 41, No. 11, pp. 3196-3199, 1993.