Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Non-Iterative Identification of IIR Wiener Systems Using Orthogonal Polynomial

Ibrahim Aljamaan * **David T. Westwick** * **Michael Foley** **

* *Department of Electrical and Computer Engineering, Schulich School of Engineering, 2500 University Drive NW, Calgary AB, T2N1N4, Canada (e-mail: dwestwic@ucalgary.ca and iaaljaman@ucalgary.ca).*
** *Department of Chemical and Petroleum Engineering, Schulich School of Engineering, 2500 University Drive NW, Calgary AB, T2N1N4, Canada (e-mail: mfoley@ucalgary.ca).*

Abstract: In this work, a non-iterative identification approach is presented for estimating a single-input single-output Wiener model, comprising an infinite impulse response discrete transfer function followed by static non-linearity. Global orthogonal basis functions and orthogonal Hermite polynomials are used as expansion bases for the linear subsystem and the non-linearity, respectively. A multi-index based method is used to transform the non-convex optimization over the parameter values into an over-parametrized linear regression. A singular value decomposition based method is then used to project the result of the over-parametrized linear regression onto the class of Wiener models, each comprising a linear element followed by a memoryless non-linearity. The advantages obtained by using orthogonal polynomials are illustrated using a series of simulation examples.

1. INTRODUCTION

The process of applying numerical tools and methods to construct an appropriate mathematical model of a dynamical system from measurements of its inputs and outputs is called system identification [1]. Many practical systems are non-linear, in that they don't obey the superposition and scaling principles. Saturation in power amplifiers, backlash in gears, and hysteresis in magnetic materials are examples of some of the most familiar non-linearities [2, 3]. While models of non-linear systems can be very complex, a wide variety of non-linear systems can be represented by non-linear block oriented models [7]. These relatively simple models consist of interconnected linear dynamic subsystems and memoryless non-linear elements.

The Hammerstein model consists of a memoryless non-linearity (N) followed by a linear time invariant filter (L), and is often denoted NL. The Hammerstein (NL), Wiener (LN), Wiener-Hammerstein (LNL) and Hammerstein-Wiener (NLN) models are the most common non-linear block structures [4, 5].

The simplest of these is the Hammerstein model [5, 6, 7], which has been extensively studied in the literature [5, 6, 7, 8, 9].

The reverse arrangement of the Hammerstein model is called the Wiener or LN model, where the linear dynamic subsystem comes before the non-linearity. This model is widely used in the literature as in [5, 7, 10, 11, 12, 13, 14, 15] and in practical applications such as modelling distillation columns [12, 13, 14] and radio frequency (RF) power amplifiers [15].

Different approaches to Wiener model identification have been presented in the literature. Many identification algorithms are based on Prediction Error methods [6, 16], but a non-iterative over-parametrization approach [10] has also been proposed. Lacy and Bernstein [10] considered the identification of Wiener models with Finite Impulse Response (FIR) linear elements and (possibly) non-invertible polynomial non-linearities. A multi-index was used to generate an over-parametrized model which was linear in the variables. We extended this method in [11] to deal with infinite impulse response (IIR) Wiener filters, represented by a finite sum of global orthogonal basis functions (GOBFs) (such as Laguerre filters) [9, 17].

The main contribution of the present work is to extend this non-iterative, overparameterization-based method to use orthogonal polynomials, such as the Hermite or Tchebyshev series, instead of the simple polynomials used in previous studies [10, 11]. The main reason for replacing the polynomials is to avoid the poor numerical properties inherent in polynomial regressions, and hence to improve the accuracy of the identification of both the linear and non-linear elements.

This paper is organized as follows: first, the proposed model is described and formulated in §**2**, after that, the identification algorithms for both FIR and IIR Wiener models proposed in [10] and [11] are summarized in §**3**. Then, Hermite orthogonal polynomials are reviewed and incorporated into the identification algorithm. In the 5^{th} part, simulation examples, qualitative and quantitative comparisons and discussion of the results are provided. Finally, Section **6** contains concluding remarks.

⋆ This work was supported by NSERC (Canada) and Ministry of Higher Education (Saudi Arabia).

2. PROBLEM STATEMENT

The proposed model is a single-input single-output Wiener cascade with an additive disturbance after the non-linear element, as shown in Fig. 1. The measured input and output to/from the system are denoted $u(t)$ and $z(t)$, respectively, while $e(t)$ is an additive independent disturbance. This disturbance, as well as the outputs of the linear dynamic element and static non-linearity, denoted $x(t)$ and $y(t)$, respectively, is assumed to be inaccessible to the investigator.

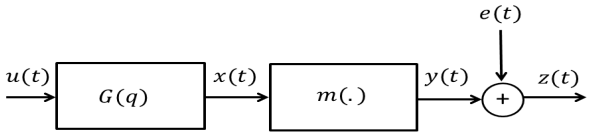

Fig. 1. Block diagram of a Wiener system with additive measurement noise, $e(t)$

The dynamics of the linear filter are expanded onto a basis of orthogonal filters, such as the Laguerre or Kautz filters (or any other GOBFs [20]). Thus, the unmeasured output of the linear block, $x(t)$, is expressed as a weighted sum of the outputs of the basis filters, as illustrated in Fig. 2 and formulated in (1).

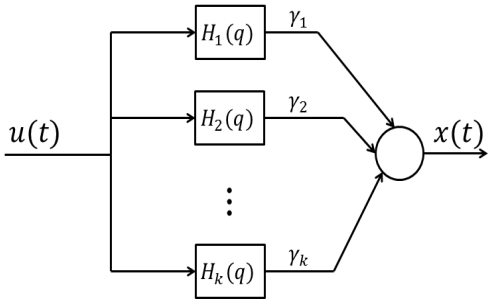

Fig. 2. The linear part of the system $x(t)$

$$\begin{aligned} x(t) &= \gamma_0 H_o(q)u(t) + \gamma_1 H_1(q)u(t) + \cdots + \gamma_n H_n(q)u(t) \\ &= \sum_{k=0}^{n} \gamma_k H_k(q)u(t) \end{aligned} \tag{1}$$

The basis filters, $H_k(q)$, are often just simple delays [10], in which case $H_k(q) = q^{-1}$, but Laguerre filters or any other GOBFs could also be used [11]. The expansion coefficients, γ_k for $k = 0 \dots n$, are the parameters that describe the linear element.

The signal resulting from the non-linearity is denoted $y(t)$ and can be expressed as:

$$y(t) = m\big(x(t)\big) \ = \sum_{i=0}^{p} a_i M_i\big(x(t)\big) \tag{2}$$

where $x(t)$ is the output of the linear element, as computed in (1), and the functions $M_i(x)$ for $i = 0 \dots p$ are polynomial basis functions. These may be orthogonal polynomials such as the Hermite, Legendre or Tchebyshev polynomials. Most often, they are simple polynomials, [10], in which case:

$$M_i(x) = x^i \tag{3}$$

Finally, the a_i are the expansion coefficients used to parametrize the non-linearity. It is well known that using orthogonal polynomials generally leads to better conditioned estimation problems [5]. In this study, the estimation algorithm proposed in Lacy and Bernstein in [10], and extended to use IIR linear filters in [11], will be reformulated in terms of orthogonal polynomials, in order to take advantage of their improved numerical properties.

The measured output of the system is obtained by adding a zero-mean Independent and identically distributed (IID) sequence, $e(t)$, to the output of the Wiener model $y(t)$:

$$z(t) = y(t) + e(t) \tag{4}$$

after substituting the (1) and (2), this:

$$y(t) = \sum_{i=0}^{p} a_i M_i\bigg(\sum_{k=0}^{n} \gamma_k H_k(q)u(t) \bigg) \tag{5}$$

3. IDENTIFICATION OF IIR WIENER MODELS

This section contains a summary of the multi-index based identification algorithm for Wiener [10, 11] systems.

The output, $z(t)$, as is given in (5), but using simple polynomials as in (3), becomes:

$$z(t) = \sum_{i=0}^{p} a_i \big(\sum_{k=0}^{n} \gamma_k H_k(q)u(t) \big)^i + e(t) \tag{6}$$

To represent the proposed model in a multi-index form, we define an $n+1$ element multi-index vector α by:

$$\alpha = [\alpha_0 \ \ \alpha_1 \ \ \cdots \ \ \alpha_n]^T \in \mathbb{N}_0^{n+1} \tag{7}$$

and the vector μ which contains the outputs of the basis filters due to the input $u(t)$:

$$\mu(t) = [H_0(q)u(t) \ \ H_1(q)u(t) \ .... \ H_n(q)u(t)] \tag{8}$$

The vector γ contains the expansion coefficients of the linear filter:

$$\gamma = [\gamma_0 \ \ \gamma_1 \ \ \cdots \ \ \gamma_n]^T \in \mathbb{R}^{n+1} \tag{9}$$

The operation of raising a vector such as $\mu(t)$ or γ to the multi-index α is defined in (10) and (11).

$$\mu(t)^\alpha = \mu_0(t)^{\alpha_0} \mu_1(t)^{\alpha_1} \cdots \mu_n(t)^{\alpha_n} \ = \prod_{i=0}^{n} \mu_i(t)^{\alpha_i} \tag{10}$$

$$\gamma^{\alpha} = \gamma_0^{\alpha_0}\gamma_1^{\alpha_1}\cdots\gamma_n^{\alpha_n} = \prod_{i=0}^{n}\gamma_i^{\alpha_i} \tag{11}$$

As a final piece of multi-index notation, define the norm and factorial of a the multi-index, α, as follows:

$$|\alpha| = \alpha_0 + \alpha_1 + \cdots + \alpha_n = \sum_{i=0}^{n}\alpha_i \tag{12}$$

$$\alpha! = (\alpha_0!)(\alpha_1!)\cdots(\alpha_n!) = \prod_{i=0}^{n}\alpha_i! \tag{13}$$

The equivalent formulation for the final model output $z(t)$, (6), using a multi-index representation , as derived in (10), is,

$$z(t) = \sum_{i=0}^{p} a_i \sum_{|\alpha|=i}\frac{|\alpha|!}{\alpha!}\gamma^{\alpha}\mu(t)^{\alpha} + e(t) \tag{14}$$

After combining the two sums this becomes:

$$z(t) = \sum_{|\alpha|=0}^{p}\frac{|\alpha|!}{\alpha!}a_{|\alpha|}\mu(t)^{\alpha}\gamma^{\alpha} + e(t) \tag{15}$$

Let $\phi(t)$ contain the known elements in (15). Thus

$$\phi(t) = \left[\frac{|\alpha|!}{\alpha!}\mu(t)^{\alpha}\right]_{|\alpha|\leq p} \mathbb{R}^{D_p^{n+1}} \tag{16}$$

and let θ contain all of the unknown elements from (15),

$$\boldsymbol{\theta} = [a_{|\alpha|}\gamma^{\alpha}]_{|\alpha|\leq p}]^T \in \mathbb{R}^{D_p^{n+1}} \tag{17}$$

Rewriting (15) in terms of $\phi(t)$ and θ

$$z(t) = \phi(t)^T\boldsymbol{\theta} + e(t) \tag{18}$$

reveals a linear regression, which may be solved using least squares. Note, however, that the number of entries in θ is larger than the number of unknowns in the vectors a and γ.

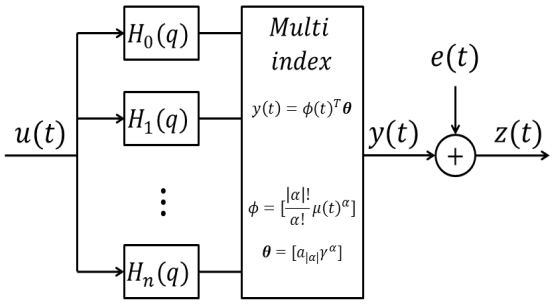

Fig. 3. Multi-index block diagram illustrating Wiener model output, $y(t)$, computation. Where the linear dynamics have been expanded onto a general basis filters ($H_k(q)$)

By rearranging the problem to be able to use a least squares method as follows:

$$\boldsymbol{z} = \boldsymbol{\Phi}^{\boldsymbol{T}}\boldsymbol{\theta} + e(t) \tag{19}$$

where $\boldsymbol{z}$ and $\boldsymbol{\Phi}$ are known and formulated, respectively, in (20) and (21):

$$\boldsymbol{z} = [z(N)\cdots z(\ell)]^T\mathbb{R}^{\ell-N} \tag{20}$$

and

$$\boldsymbol{\Phi} = [\phi(N)\cdots\phi(\ell)]^T\mathbb{R}^{D_p^{n+1}\times\ell-N} \tag{21}$$

Note that the notation N is the total number of data and the range is assumed to be between $t = \ell\cdots N$.

3.1 Singular Value Decomposition

As a result of its simplicity and accuracy, a singular value decomposition (SVD) based method is used extract the linear parameters, $\hat{\gamma_n}$, and the coefficients of the non-linearity, $\hat{a_i}$ from the over-parametrized vector $\boldsymbol{\theta}$, where the first coefficient of the non-linearity, is simply equal to the first parameter of $\boldsymbol{\theta}$:

$$\hat{a}_0 = \hat{\boldsymbol{\theta}}(1) \tag{22}$$

The rest of the parameters in $\boldsymbol{\theta}$ are placed into the matrix $A(\boldsymbol{\theta})$ such that it has the following structure:

$$A(\hat{\boldsymbol{\theta}}) = \psi\gamma^T = \begin{bmatrix} a_0 \\ a_1\gamma_1^{\alpha_1} \\ \vdots \\ a_n\gamma_n^{\alpha_n} \end{bmatrix} \tag{23}$$

where:

$$\psi = [a_{|\alpha|+1}\gamma^{\alpha}]_{|\alpha|=p} \tag{24}$$

Note that the vector $\hat{\theta}$ contains estimates of the unique elements of the system's Volterra kernels of order 0 through p, expanded onto a basis constructed from the GOBF basis filters [11]. Construct estimates of the Volterra kernels, by placing each unique coefficient in all of its symmetric locations in the appropriate kernel. Then, the matrix $A(\hat{\theta})$ may be constructed as follows:

$$A(\hat{\theta}) = \begin{bmatrix} K_1^T \\ K_2 \\ K_3(:,:,1) \\ K_3(:,:,2) \\ \vdots \\ K_3(:,:,n+1) \\ \vdots \end{bmatrix}$$

K_1 is an $n+1$ element vector, K_2 is a $n+1$ by $n+1$ matrix, and that K_i is an i'th order tensor, with $n+1$ elements in each dimension. MATLAB notation has been used to indicate matrix and tensor partitions. Thus, for kernels of order 3 and higher, where K_i is the estimate of the order i Volterra kernel. For kernels of order 3 and higher, $A(\hat{\theta})$ contains stacks of 2-dimensional slices of the kernel.

Computing the SVD of $A(\hat{\theta})$,

$$A(\hat{\boldsymbol{\theta}}) = USV^T \tag{25}$$

it is evident from (22) that:

$$\hat{\gamma}_n = S(1,1)V(1,n) \tag{26}$$

which can be used to recover the expansion coefficients of the linear element. Finally, the non-linearity is estimated

by applying a simple least squares regression, after computing the output of the estimated linear element, $\hat{x}(t)$.

Note that there are approaches other than the SVD method discussed here that may be used to estimate the parameter values, a_i and γ_k, from elements of the vector $\hat{\boldsymbol{\theta}}$. These include the other three techniques presented in Lacy and Bernstein in [10]

4. ORTHOGONAL POLYNOMIALS

Two polynomials are called orthogonal if the expected value of the product of their outputs is zero. Thus, the polynomial terms will only be orthogonal if their input has a specific probability distribution. In this section, one of the most common orthogonal polynomials available in the literature is discussed, namely the Hermite polynomials.

4.1 Hermite Polynomials

The Hermite polynomials are orthogonal, provided the input is a sample function of a standard Gaussian process (i.e. zero mean, unit variance). They are formulated by [5]

$$\mathcal{M}^n(u) = n! \sum_{m=0}^{n/2} \frac{(-1)^m}{m! \times 2^m (n-2m)!} u^{(n-2m)} \quad (27)$$

and have the recurrence relation:

$$\mathcal{M}^{n+1}(u) = u\mathcal{M}^n(u) - k\mathcal{M}^{n-1}(u) \quad (28)$$

where the first four Hermite polynomials are shown in Fig. 4 and defined by:

$$\begin{aligned}
\mathcal{M}^0(u) &= 1 \\
\mathcal{M}^1(u) &= u \\
\mathcal{M}^2(u) &= u^2 - 1 \\
\mathcal{M}^3(u) &= u^3 - 3u \\
\mathcal{M}^4(u) &= u^4 - 6u^2 + 3
\end{aligned}$$

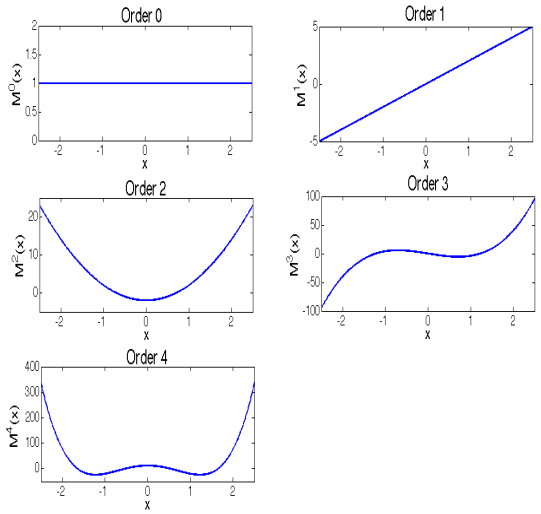

Fig. 4. Hermite polynomials of orders 0 through 4.

4.2 Multi-index algorithm using Hermite Polynomials

To expand the multiple-input polynomial (described by Eq. (14) and illustrated in Fig. 3) onto a basis of Hermite polynomials, the basis elements $\mu(t)^\alpha$ are replaced with:

$$\begin{aligned}
\mathcal{M}^{\alpha_i}(\mu_i(t)) &= \mathcal{M}^{\alpha_0}(\mu_0(t))\mathcal{M}^{\alpha_1}(\mu_1(t)) \cdots \mathcal{M}^{\alpha_n}(\mu_n(t)) \\
&= \prod_{i=0}^{n} \mathcal{M}^{\alpha_i}(\mu_i(t)) \quad (29)
\end{aligned}$$

thus, instead of raising each element in $\mu(t)$ to a power, it is transformed by the corresponding Hermite polynomial.

When simple polynomials are used, the multi-index generates all of the unique terms in the Volterra series up to degree p (whether a basis of simple delays or GOBFs is used for the dynamics, as in [10] and [11], respectively). The polynomial non-linearity has been replaced with Hermite orthogonal polynomials to reduce the numerical conditioning problems inherent in polynomial estimation. Expanding the Volterra kernels using a basis of Hermite polynomials results in the Wiener series [5]. Thus, the over-parametrized linear regression resulting from the multi-index generates estimates of all of the unique entries in the system's Wiener kernels. Since the Volterra and Wiener kernels of Wiener-Hammerstein systems are proportional to one another [21], the same SVD based approach used in the previous section may be used to recover the coefficients of the linear element.

4.3 Scaling Orthogonal Hermite Polynomials

Since the Hermite polynomials are only orthogonal for a standard normal input, the basis filters $H_k(q)$ must all be scaled so that their outputs have unit variance. This added scaling must be taken into account when constructing the linear element from its estimated coefficients.

5. SIMULATION

The simulations used the structure of Fig. 1, with a third-order Butterworth filter with a normalized cut-off frequency of $W_n = 0.2455$ as the linear element followed by a third order polynomial for the non-linearity.

$$y(t) = 1 + x(t) + x^2(t)/2 + x^3(t)/3$$

A sum of 7 Laguerre filters was used to approximate the third order Butterworth filter shown in Fig. 5. In these simulations, the Laguerre pole was chosen manually based in the dynamics of the simulated linear system, resulting in the value $\overline{\alpha} = 0.22151$. In practical applications, $\overline{\alpha}$ would be chosen using an iterative optimization [22].

Four different experiments were conducted. In two of the experiments, the input $u(t)$ was a zero-mean IID sequence of Gaussian random variables with variance ≈ 4. In the other experiments, the input was uniformly distributed between -1 and 1. The system was identified for both input signals using a mulit-index based expansion onto either simple polynomials, or Hermite polynomials.

Each test involved a 50 trial Monte-Carlo simulation. The data length was $N = 2^{14} = 16384$ with a Signal To Noise Ratio (SNR) of 10 **dB**.

The effect of the scaling, used to produce unit variance inputs to ensure the orthogonality of the Hermite polynomials was also tested. Two complete sets of simulations were performed, one with and one without this scaling.

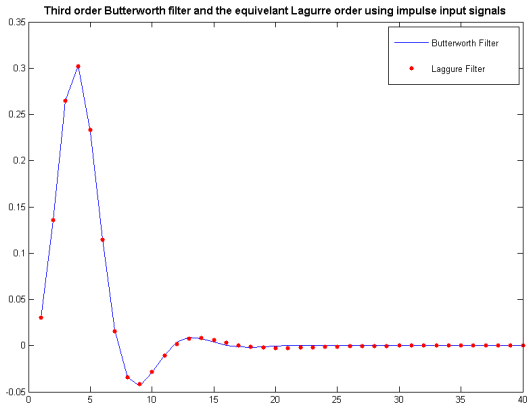

Fig. 5. Third order Butterworth filter and the fitted Laguerre filter using impulse input

The results of all experiments are presented in Table 1 which shows that the case with a Gaussian input, scaled filters and a Hermite polynomial expansion of the non-linearity generated the most accurate and consistent results. Note as well that reductions in the condition numbers of the regressions occurred when using a Hermite expansion, as compared to a simple polynomial expansion (rows listed as xxx.H and xxx.P, respectively). This continued to hold when the input was uniformly distributed, even though the polynomials were no-longer orthogonal.

Table 1. The sample average and standard deviation of the MSE the impulse response estimate of the linear subsystem as well as the sample average and standard deviation of the condition numbers of the overparametrized linear regressions for all experiments.

Exp.	AveMSE	stdMSE	AveCo	stdCo
Scaled				
Gauss.H	$0.330e^{-8}$	$0.143e^{-8}$	7.17	0.265
Gauss.P	$4.879e^{-8}$	$2.233e^{-8}$	25.86	0.656
Unif.H	$0.497e^{-8}$	$0.316e^{-8}$	13.33	0.618
Unif.P	$7.788e^{-8}$	$4.022e^{-8}$	33.17	0.824
Unscaled				
Gauss.H	$2.667e^{-8}$	$1.388e^{-8}$	117.50	4.713
Gauss.P	$6.345e^{-8}$	$3.187e^{-8}$	178.05	5.484
Unif.H	$6.735e^{-8}$	$4.099e^{-8}$	231.26	10.181
Unif.P	$15.61e^{-8}$	$9.556e^{-8}$	324.15	11.605

The simulated non-linearity is plotted in Fig. 6, superimposed on the 50 estimates produced in the Monte-Carlo simulation which used a Gaussian input, Hermite polynomial expansion and scaling. Figure 7 illustrates that the estimated impulse responses of the estimated filters in 50 runs fits properly the exact one. Tables 2 and 3 contains the standard errors of the estimated linear filter coefficients based on the 50 run Monte-Carlo simulations. Table 2 shows the results for Gaussian inputs, whereas the results from the Uniform input trials are shown in Table 3. From these 2 tables, it clear that the smallest errors occurs when the Hermite is scaled and the input is Gaussian and the worst happens in the case of unscaled Polynomials with non-Gaussian input. Fig. 8 shows the average condition numbers of the 50 runs against the average Mean Squares Error (MSE) of all the experiments.

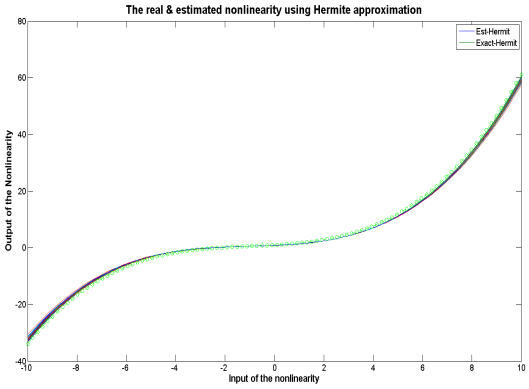

Fig. 6. The exact and estimated 50 runs non-linearity by fitting Hermite Polynomials for the Hermite approximated over-parametrized coefficients

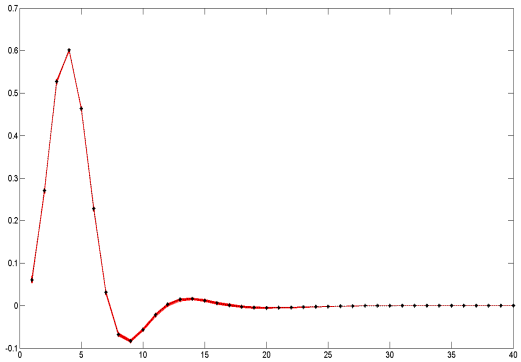

Fig. 7. The estimated impulse response for 50 runs of the linear subsystem and the exact one for the case Hermit Gaussian scaled

6. CONCLUSION

In this paper, we extended the multi-index based over-parametrization method for Wiener system identification proposed in [10], replacing the simple polynomial expansion with an orthogonal expansion based on Hermite polynomials. Simulations demonstrated the increased accuracy that result from the better numerical properties of the orthogonal polynomials.

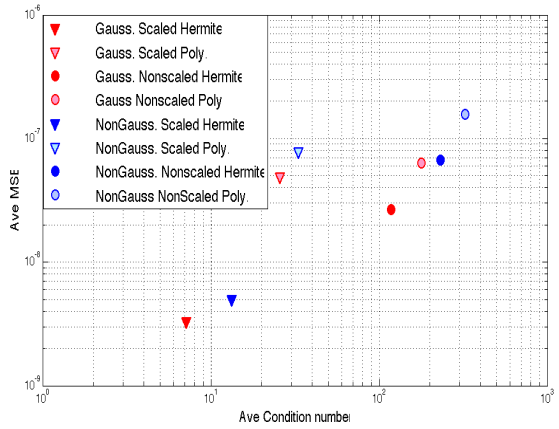

Fig. 8. The condition number against the MSE values

Table 2. Standard errors in the estimated linear subsystem parameters for the four experiments, when the input is Gaussian

Par.	S.Hermite	Uns.Hermite	S.Poly	Uns.Poly
$\mathbf{h_0}$	$0.1861e^{-4}$	$0.5928e^{-4}$	$0.6991e^{-4}$	$0.897e^{-4}$
$\mathbf{h_1}$	$0.1375e^{-4}$	$0.4031e^{-4}$	$0.5438e^{-4}$	$0.601e^{-4}$
$\mathbf{h_2}$	$0.2414e^{-4}$	$0.7001e^{-4}$	$0.8210e^{-4}$	$1.087e^{-4}$
$\mathbf{h_3}$	$0.2082e^{-4}$	$0.4644e^{-4}$	$0.8245e^{-4}$	$0.720e^{-4}$
$\mathbf{h_4}$	$0.2301e^{-4}$	$0.5403e^{-4}$	$0.7733e^{-4}$	$0.823e^{-4}$
$\mathbf{h_5}$	$0.2156e^{-4}$	$0.6192e^{-4}$	$0.9331e^{-4}$	$0.960e^{-4}$
$\mathbf{h_6}$	$0.1804e^{-4}$	$0.5888e^{-4}$	$0.8070e^{-4}$	$0.915e^{-4}$
$\mathbf{h_7}$	$0.2167e^{-4}$	$0.6751e^{-4}$	$0.7887e^{-4}$	$1.053e^{-4}$

Table 3. Standard errors in the estimated linear subsystem parameters for the four experiments, when the input is uniform

Par.	S.Hermite	Uns.Hermite	S.Poly	Uns.Poly
$\mathbf{h_0}$	$0.0332e^{-3}$	$0.0967e^{-3}$	$0.1081e^{-3}$	$0.1517e^{-3}$
$\mathbf{h_1}$	$0.0188e^{-3}$	$0.0504e^{-3}$	$0.0627e^{-3}$	$0.0777e^{-3}$
$\mathbf{h_2}$	$0.0254e^{-3}$	$0.0953e^{-3}$	$0.1106e^{-3}$	$0.1510e^{-3}$
$\mathbf{h_3}$	$0.0206e^{-3}$	$0.1049e^{-3}$	$0.0972e^{-3}$	$0.1573e^{-3}$
$\mathbf{h_4}$	$0.0203e^{-3}$	$0.1021e^{-3}$	$0.0933e^{-3}$	$0.1547e^{-3}$
$\mathbf{h_5}$	$0.0298e^{-3}$	$0.0975e^{-3}$	$0.0977e^{-3}$	$0.1463e^{-3}$
$\mathbf{h_6}$	$0.0252e^{-3}$	$0.0963e^{-3}$	$0.1035e^{-3}$	$0.1432e^{-3}$
$\mathbf{h_7}$	$0.0192e^{-3}$	$0.0793e^{-3}$	$0.1038e^{-3}$	$0.1176e^{-3}$

REFERENCES

[1] R. Pintelon and J. Schoukens, "System Identification A Frequency Domain Approach", *Wiley-IEEE Press*, 2012.

[2] E. Slotine and W. Li, "Applied Nonlinear Control", *Prentice-Hall, Inc.*, 1991.

[3] H. K. Khalil, "Nonlinear Systems", *Prentice-Hall, Inc.*, 2002.

[4] O. Nelles, "Nonlinear System Identification", *Springer-Verlag Berlin Heidelberg*, 2001.

[5] D.T. Westwick and R.E. Kearney, "Identification of Nonlinear Physiological System", *Wiley-IEEE Press*, 2003.

[6] L. Ljung, "System Identification Theory for the User", *Upper Saddle River, NJ Prentice-Hall*, 1999.

[7] F. Giri and E. Bai, "Block-oriented Nonlinear System Identification", *Springer, Berlin*, 2010.

[8] L.C. Evans, "Partial Differential Equations", *Graduate Study of Mathematics, American Mathematical Society, Providence, RI, USA*, vol. 19, 1998.

[9] A. Dankers and D. Westwick, "Nonlinear system identification using optimally selected Laguerre filter banks", *American Control Conference*, 2006, pp. 2369-2374.

[10] S.L. Lacy and D.S. Bernstein, "Identification of FIR Wiener systems with unknown, non-invertible, polynomial non-linearities", *International Journal of Control*, vol. 67, No. 15, 2003, pp. 1500-1507.

[11] I. A. Aljamaan, D.T. Westwick and M. Foley, " Non-Iterative Identification of IIR Wiener Systems with Non-Invertible Nonlinities", *IFAC International Workshop on Adaptation and Learning in Control and Signal Processing*, pp. 390-395, 2013.

[12] S.J Norquay, A Palazoglu, and J.A Romagnoli, "Application of wiener model predictive control (wmpc) to an industrial c2-splitter", *Journal of Process Control*, vol. 9(6), 1999, pp. 461473.

[13] M. Verhaegen, "Identification Of The Temperature-Product Quality Relationship In a Multi-Component Distillation Column", *Chemical Engineering Communications*, vol. 136, 1998, pp. 111132.

[14] S.J Norquay and A Palazoglu and J.A Romagnoli, "Application of Wiener model predictive control (WMPC) to an industrial C2-splitter", *Journal of Process Control*, vol. 9, 1999, pp. 461-473.

[15] C.J. Clark, G. Chrisikos, M.S. Muha, S. Michael, A.A. Moulthrop and and C.P Silva, "Time-domain envelope measurement technique with application to wideband power amplifier modeling", *Microwave Theory and Techniques, IEEE Transactions on*, vol. 46, 1998, pp. 2531-2540.

[16] A. Papoulis and S. Pillai, "Probability, Random Variables and Stochastic Processes", *McGraw-Hill Fourth Edition*, 2002.

[17] V.Z. Marmarelis, "Identification of nonlinear biological systems using Laguerre expansions of kernels", *Annals of Biomedical Engineering*, vol. 21, 1993, pp. 573-589.

[18] J.S. Bendat, A.G. Piersol, "Random Data: Analysis and Measurement Procedures", *Wiley Press, 4th Edition*, 2010.

[19] N. Wiener, "Nonlinear Problems in Random Theory ", *Technology Press Research Monograph. John Wiely Sons, New York*, 1958.

[20] C.F. Dunkl and Y. Xu , "Orthogonal Polynomials of Several Variables", *Encyclopedia of Mathematics and its Applications, (Cambridge, UK: Cambridge University Press)*, vol. 181, 2001.

[21] W.J. Rugh, "Nonlinear System Theory: The Volterra/Wiener Approach", *The Johns Hopkins University Press*, 1981.

[22] A. Dankers and D. Westwick, "On the Relationship between the Enforced Convergence Criterion and the Asymptotically Optimal Laguerre Pole", *IEEE Transactions on Automatic Control*, vol. 57, No.5, 2012, pp. 1102-1109.