Identification of Nonlinear Systems using Orthonormal Bases

Juan C. Gómez

Laboratory for System Dynamics and Signal Processing
Universidad Nacional de Rosario
Argentina
jcgomez@fceia.unr.edu.ar

Outline

- Introduction
- Nonlinear Models → **Block-oriented Models**
- **Hammerstein model**
  1. Formulation of the Identification Problem
  2. Nonlinear Identification Algorithm
- **Wiener model**
  1. Formulation of the Identification Problem
  2. Nonlinear Identification Algorithm
- **Feedback Block-oriented model**
  1. Formulation of the Identification Problem
  2. Nonlinear Identification Algorithm
- Simulation Examples
- Conclusions

Introduction

- Non iterative algorithms for the identification of **Multivariable Block-oriented Nonlinear** models are presented.
- The algorithms are **numerically robust**, since they are based only on Least Squares Estimation (**LSE**) and Singular Value Decomposition (**SVD**). No nonlinear numerical optimization procedures are required.
- For the **Hammerstein** model **consistency** of the estimates is guaranteed under very weak assumptions on the persistency of excitation of the inputs, and **even in the presence of coloured noise**. For the **Wiener** model and the **Feedback** model ⟶ problems.
- Key in the derivation of the results is the representation of the linear part of the models using **orthonormal bases functions**.

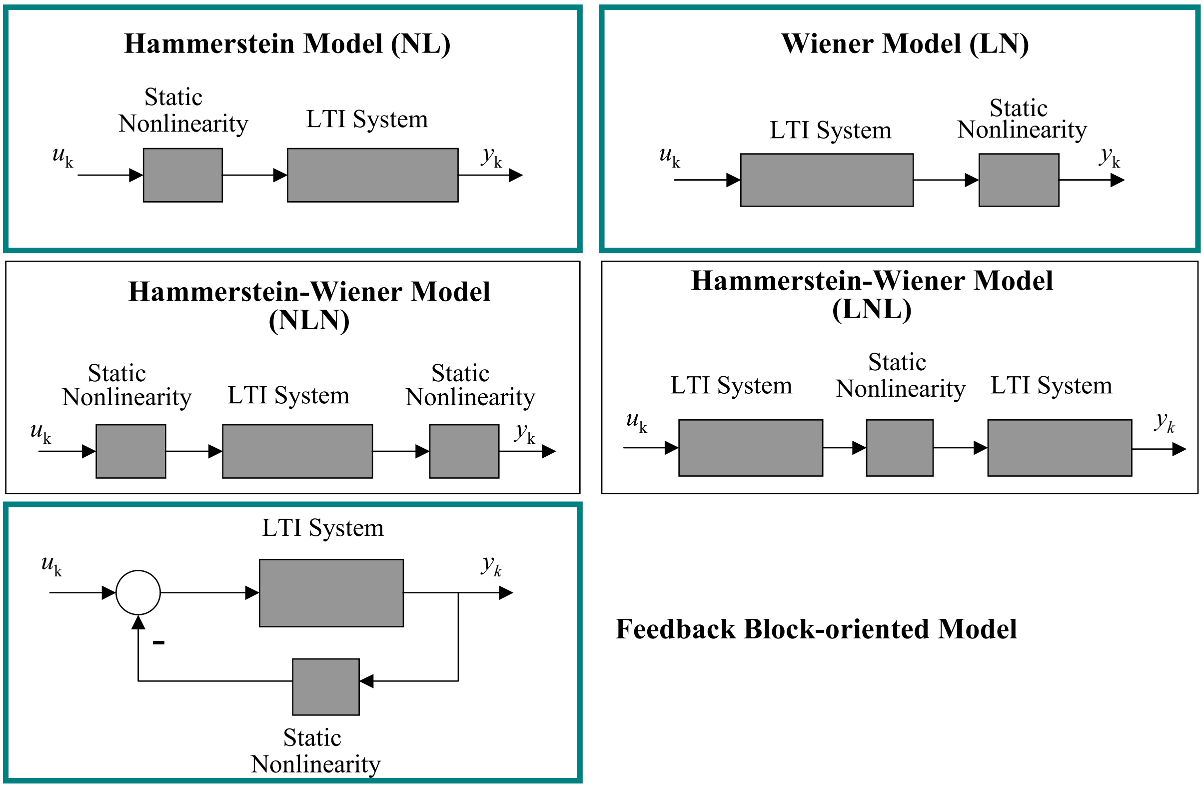
Motivation for Nonlinear Identification

- Most physical processes have a nonlinear behaviour, except in a limited range where they can be considered linear.
- The performance of controllers designed from a linear approximation is strongly influenced by a change in the operating point of the system.
- Nonlinear models are able to describe more accurately the global behaviour of the system, independently of the operating point.

Nonlinear Models

- Since the identification is carried out from observed input-output data, it is more natural to try to identify discrete-time models, rather than continuous-time ones.
- Many dynamical systems can be represented by the interconnection of static nonlinearities and LTI systems. These models are called **block-oriented** nonlinear models.
- **Hammerstein models** (cascade connection of a static nonlinearity followed by a LTI system) , **Wiener models** (where the order of the blocks is reversed), and **Feedback models** (static nonlinearity in the feedback loop around a LTI system), have been successfully used in a number of practical applications in the areas of chemical processes, biological processes, signal processing, communications, controls, etc.

Block-oriented Nonlinear Models



Feedback Block-oriented Model

Nonlinear Identification Algorithms for Hammertein-Wiener Models

• Iterative algorithms for nonlinear optimization (Narendra *et al.,* 1966) : convergence problems, existence of local minima, initialization problems, computationally intensive.

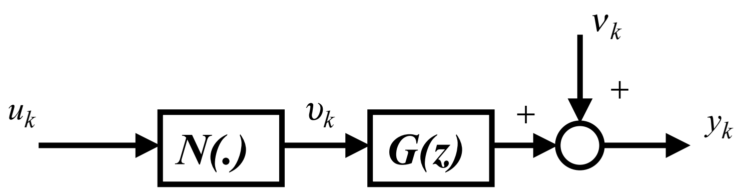
• Correlation techniques (Billings *et al.,* 1982) : rather restrictive requirement on the input being white noise.

• Recent approaches based on Least Squares techniques and Singular Value Decomposition (SVD) (Bai, 1998),(Gómez *et al.,* 2000): global convergence is guaranteed, numerically robust, not computationally intensive.

• Present work is a collaboration with Dr. Enrique Baeyens, Universidad de Valladolid, Spain.

Hammerstein Model

1.Problem Formulation



Let the **Hammerstein model** be described by:

$$y_k = G(q)N(u_k) + v_k \qquad (1)$$

where $G(q)$ is the transfer matrix of the LTI subsystem, and $N(\bullet)$ is the (static) input-output characteristic of the nonlinear subsystem, and where $y_k \in \Re^m$, $u_k \in \Re^n$, and $v_k \in \Re^m$ are the system output, input, and measurement noise vectors at time *k,* respectively.

It will be assumed that the **nonlinear subsystem** can be described as

$$N(u_k) = \sum_{i=1}^{r} a_i g_i(u_k) \tag{2}$$

where $g_i(\bullet): \Re^n \rightarrow \Re^n, (i = 1, \cdots, r)$ are known vector fields, and $a_i \in \Re^{n \times n}, (i = 1, \cdots, r)$ are unknown matrix parameters.

On the other hand, the **LTI subsystem** will be represented using rational orthonormal bases on $H_2(\mathbf{T})$ as

$$G(q) = \sum_{\ell=0}^{p-1} b_\ell \mathbf{B}_\ell(q) \tag{3}$$

where $b_\ell \in \Re^{m \times n}$ are unknown matrix parameters, and $\{\mathbf{B}_\ell(q)\}_{\ell=0}^{\infty}$ are rational orthonormal bases on $H_2(\mathbf{T})$.

<u>Identification problem:</u> to estimate the unknown parameter matrices $a_i \in \Re^{n \times n}, (i = 1, \cdots, r)$, and $b_\ell \in \Re^{m \times n}, (\ell = 0, \cdots, p-1)$ characterizing the nonlinear and the linear parts, respectively, from an N-point data set $\{u_k, y_k\}_{k=1}^{N}$ of observed input-output measurements.

2. Nonlinear Identification Algorithm

Considering (2) and (3), the input-output equation (1) can be written as

$$y_k = \sum_{\ell=0}^{p-1}\sum_{i=1}^{r} b_\ell a_i \mathbf{B}_\ell(q) g_i(u_k) + v_k \qquad (4)$$

Identifiability problem

Note: It is clear from (4) that the parameterization (1)-(3) is **not unique**, since any parameter matrices $b_\ell \alpha$, and $\alpha^{-1} a_i$, for some nonsingular matrix $\alpha \in \Re^{n\times n}$, provide the same input-output equation (1). To obtain a one-to-one parameterization, *i.e.,* for the system to be **identifiable**, additional constraints must be imposed on the parameter matrices. A standard technique is to normalize the parameter matrices, assuming for instance $\|a_i\|_2 = 1$ $(\text{or } \|b_\ell\|_2 = 1)$.

Defining

$$\theta = [b_0 a_1, \ldots, b_0 a_r, \ldots, b_{p-1} a_1, \ldots, b_{p-1} a_r]^T$$

$$\phi_k = [\mathbf{B}_0(q) g_1(u_k)^T, \ldots, \mathbf{B}_0(q) g_r(u_k)^T, \cdots, \mathbf{B}_{p-1}(q) g_1(u_k)^T, \ldots, \mathbf{B}_{p-1}(q) g_r(u_k)^T]^T$$

the input/output equation (4) can be written as a **linear regressor**

$$y_k = \theta^T \phi_k + v_k \qquad (4)$$

Considering an N -point data set, equation (4) can be written in matrix form as

$$Y_N = \Phi_N^T \theta + V_N \qquad (5)$$

where

$$Y_N = [y_1^T, \ldots, y_N^T]^T, V_N = [v_1^T, \ldots, v_N^T]^T, \mathbf{\Phi}_N = [\phi_1, \ldots, \phi_N]$$

The Least Squares Estimate is given by

$$\hat{\theta} = (\Phi_N \Phi_N^T)^{-1} \Phi_N Y_N \tag{6}$$

The problem is now how to estimate the parameter matrices a_i $(i = 1, \cdots, r)$ and b_ℓ $(\ell = 0, \cdots, p\text{-}1)$ from the estimate $\hat{\theta}$ in (6).

Defining the matrices

$$\Theta_{ab} = \begin{pmatrix} a_1^T b_0^T & a_1^T b_1^T & \ldots & a_1^T b_{p-1}^T \\ a_2^T b_0^T & a_2^T b_1^T & \ldots & a_2^T b_{p-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ a_r^T b_0^T & a_r^T b_1^T & \ldots & a_r^T b_{p-1}^T \end{pmatrix} = ab^T, \tag{7}$$

$$a = [a_1, a_2, \cdots, a_r]^T,$$

$$b = [b_0^T, b_1^T, \cdots, b_{p-1}^T]^T,$$

it is easy to see that

$$\theta = \text{blockvec}(\Theta_{ab})$$

so that an estimate $\hat{\Theta}_{ab}$ can be obtained from the estimate $\hat{\theta}$ in (6).

The closest, in the 2-norm sense, estimates $\hat{a}$ and $\hat{b}$ are such they minimize the norm

$$\left\| \hat{\Theta}_{ab} - \hat{a}\hat{b}^T \right\|_2^2$$

That is

$$(\hat{a}, \hat{b}) = \underset{a,b}{\text{argmin}} \left\| \hat{\Theta}_{ab} - ab^T \right\|_2^2. \tag{8}$$

The solution to this optimization problem is provided by the SVD of $\hat{\Theta}_{ab}$.

Main Result: Let $\hat{\Theta}_{ab} \in \Re^{nr \times mp}$ have rank *k>n,* and let its economy size SVD be partioned as

$$\hat{\Theta}_{ab} = U\Sigma V^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{9}$$

with $U_1 \in \Re^{nr \times n}$, $V_1 \in \Re^{mp \times n}$, and $\Sigma_1 = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$.

Then

$$(\hat{a}, \hat{b}) = \underset{a,b}{\mathrm{argmin}} \left\| \hat{\Theta}_{ab} - ab^T \right\|_2^2 = (U_1, V_1 \Sigma_1), \tag{10}$$

and the approximation error is given by

$$\left\| \hat{\Theta}_{ab} - \hat{a}\hat{b}^T \right\|_2^2 = \sigma_{n+1}^2. \tag{11}$$

Identification Algorithm

The identification algorithm can be summarized as follows.

Step 1: Compute the LSE $\hat{\theta}$ in (6), and the matrix $\hat{\Theta}_{ab}$ such that

$$\hat{\theta} = \mathrm{blockvec}(\hat{\Theta}_{ab}).$$

Step 2: Compute the *economy size* SVD of $\hat{\Theta}_{ab}$, and the partition of this decomposition as in (9).

Step 3: Compute the estimates of the parameter matrices *a* and *b* as

$$\hat{a} = U_1 ,$$

$$\hat{b} = V_1 \Sigma_1 ,$$

respectively.

Consistency Analysis

Result: Let $\hat{a}$ and $\hat{b}$ be computed using the proposed identification algorithm. Then, assuming that the uniqueness condition $\|a_i\|_2 = 1$ holds, and that the regressors ϕ_k are persistently exciting (PE),

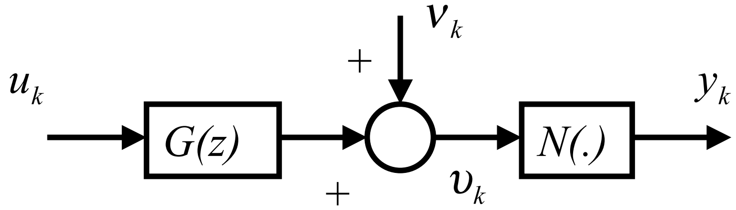
$$\hat{a} \xrightarrow{a.s.} a\,,$$
$$\hat{b} \xrightarrow{a.s.} b\,,$$

as $N \to \infty$. The result holds even in the presence of **coloured noise**.

Key in the proof of this result is the fact that the regressors are deterministic, since depend only on past inputs **(orthonormal basis model structure).**

Wiener model

1. Problem Formulation



We assume that *N(.)* is invertible, and that its inverse can be represented as

$$N^{-1}(y_k) = \sum_{i=1}^{r} a_i g_i(y_k) \qquad (12)$$

where $g_i(\bullet): \Re^m \to \Re^m, (i = 1, \cdots, r)$ are known vector fields, and $a_i \in \Re^{m \times m}, (i = 1, \cdots, r)$ are unknown matrix parameters.

Without loss of generality it will be assumed that $a_1 = I_m$

On the other hand, the **LTI subsystem** will be represented using rational orthonormal bases on $H_2(\mathbf{T})$ as

$$G(q) = \sum_{\ell=0}^{p-1} b_\ell \mathbf{B}_\ell(q) \tag{13}$$

where $b_\ell \in \Re^{m\times n}$ are unknown matrix parameters, and $\{\mathbf{B}_\ell(q)\}_{\ell=0}^{\infty}$ are rational orthonormal bases on $H_2(\mathbf{T})$.

Identification problem: to estimate the unknown parameter matrices $a_i \in \Re^{m\times m}, (i = 2,\cdots,r)$, and $b_\ell \in \Re^{m\times n}, (\ell = 0,\cdots,p-1)$ characterizing the nonlinear and the linear parts, respectively, from an N-point data set $\{u_k, y_k\}_{k=1}^{N}$ of observed input-output measurements.

2. Nonlinear Identification Algorithm

The intermediate variable υ_k can be written as

$$\upsilon_k = G(q)u_k + \nu_k$$

and also as

$$\upsilon_k = N^{-1}(y_k)$$

Equating the right-hand sides of both equations and considering the parame-terization of the linear and nonlinear blocks

$$g_1(y_k) = -\sum_{i=2}^{r} a_i g_i(y_k) + \sum_{\ell=0}^{p-1} b_\ell \mathbf{B}_\ell(q)u_k + \nu_k \tag{14}$$

which is a linear regression. Defining

$$\theta = [a_2, a_3, \cdots, a_r, b_0, b_1, \cdots, b_{p-1}]^T$$

$$\phi_k = [-g_2^T(y_k), -g_3^T(y_k), \cdots, -g_r^T(y_k), \mathbf{B}_0(q)u_k^T, \cdots, \mathbf{B}_{p-1}(q)u_k^T]^T$$

we can write

$$g_1(y_k) = \theta^T \phi_k + v_k$$

Now, an estimate of the parameter matrix θ can be computed by minimizing a quadratic criterion on the prediction errors

$$\varepsilon_k = g_1(y_k) - \theta^T \phi_k$$

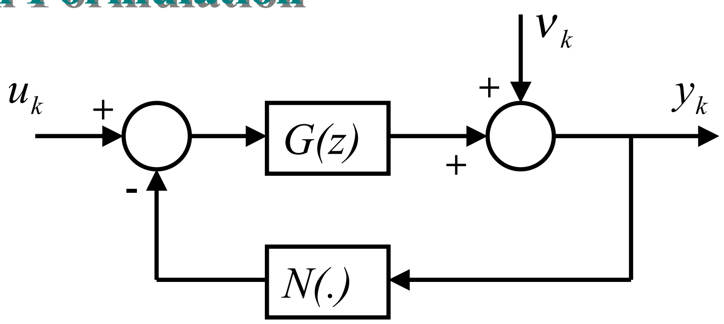
(*i.e.,* the least squares estimate). The solution is given by

$$\hat{\theta} = (\Phi_N \Phi_N^T)^{-1} \Phi_N Y_N$$

Consistency ⟶ **problems (noise free-case)**

Feedback block-oriented model

1. Problem Formulation



$$N(y_k) = \sum_{i=1}^{r} a_i g_i(y_k)$$ **nonlinear subsystem**

$$G(z) = \sum_{\ell=0}^{p-1} b_\ell \mathbf{B}_\ell(z)$$ **LTI subsystem**

$$y_k = \sum_{\ell=0}^{p-1} b_\ell \mathbf{B}_\ell(q) u_k - \sum_{\ell=0}^{p-1} \sum_{i=1}^{r} b_\ell a_i \mathbf{B}_\ell(q) g_i(y_k) + v_k \qquad (15)$$

Defining

$$\theta = \left[b_0, b_1, \cdots, b_{p-1}, b_0 a_1, \cdots, b_0 a_r, \cdots, b_{p-1} a_1, \cdots, b_{p-1} a_r\right]^T$$

$$\phi_k = \left[\mathbf{B}_0(q) u_k^T, \cdots, \mathbf{B}_{p-1}(q) u_k^T, -\mathbf{B}_0(q) g_1^T(y_k), \cdots, -\mathbf{B}_0(q) g_r^T(y_k), \right.$$
$$\left. \cdots, -\mathbf{B}_{p-1}(q) g_1^T(y_k), \cdots, -\mathbf{B}_{p-1}(q) g_r^T(y_k)\right]^T$$

the input-output equation (15) can be written as

$$y_k = \theta^T \phi_k + v_k$$

which is a linear regression. As in the case of the Hammerstein and the Wiener models, the least squares estimate of θ is given by

$$\hat{\theta} = (\Phi_N \Phi_N^T)^{-1} \Phi_N Y_N$$

with similar definitions for Φ_N and Y_N.

The parameter matrix θ can be written as

$$\theta = \left[b_0, \cdots, b_{p-1}, \text{blockvec}(\Theta_{ab})^T\right]^T$$

So that estimates $\hat{b}$ and $\hat{\Theta}_{ab}$ can be obtained from the LSE $\hat{\theta}$.

An estimate of matrix a can be obtained by solving the 2-norm minimization problem

$$\hat{a} = \underset{a}{\text{argmin}} \left\{ \left\| \hat{\Theta}_{ab} - a\hat{b}^T \right\|_2^2 \right\}$$

which yields

$$\hat{a} = \hat{\Theta}_{ab} \hat{b} \left(\hat{b}^T \hat{b}\right)^{-1}$$

Consistency ⟶ **problems (white noise)**

Simulation Examples

1. Hammerstein model

❑ The True System

$$G(z) = \frac{z^2 + 0.7z - 1.5}{z^3 + 0.9z^2 + 0.15z + 0.002}$$

linear subsystem

$$N(u_k) = 0.8585\,u_k + 0.0149\,u_k^2 - 0.5113\,u_k^3 - 0.0263\,u_k^4$$

nonlinear subsystem

❑ The input and noise

$$u_k = \sin(0.0005\pi k) + 0.5\sin(0.0015\pi k) + \\ + 0.3\sin(0.0025\pi k) + 0.1\sin(0.0035\pi k)$$

(a bad) input

$$\Phi_v(\omega) = \frac{0.64 \times 10^{-8}}{1.2 - 0.4\cos(\omega)}$$

Spectrum of the zero mean coloured noise

❑ The Orthonormal Bases

$$\mathbf{B}_\ell(q) = \left(\frac{\sqrt{1-|\xi_\ell|^2}}{q-\xi_\ell}\right)\prod_{i=0}^{\ell-1}\left(\frac{1-\xi_i q}{q-\xi_i}\right)$$

Orthonormal Bases with Fixed Poles

Generalization of the standard FIR, Laguerre, and Kautz Bases.

❑ The chosen basis poles

$$\{-0.01, -0.2, -0.7\}$$

Basis poles (3rd order linear model)

True poles at $\{0.0124, -0.2399, -0.6725\}$

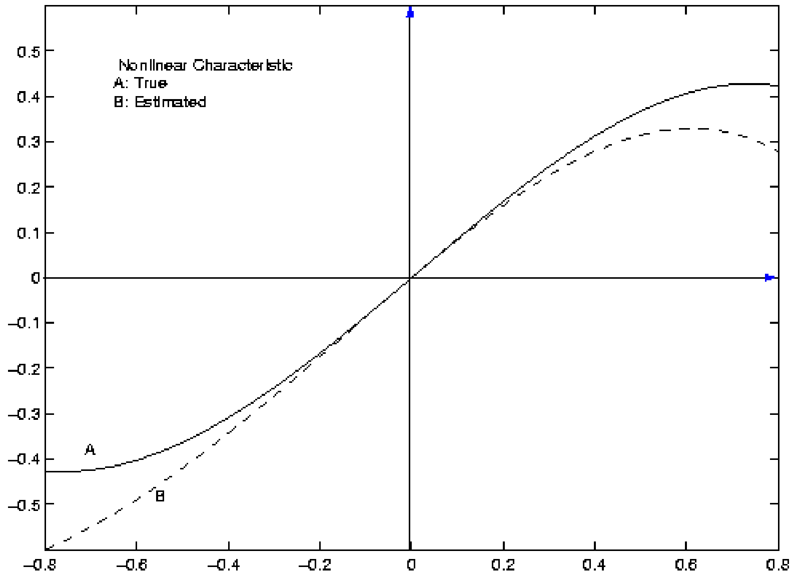
❑ The Estimated Transfer Function

$$\hat{G}(z) = \frac{1.0012z^2 + 0.6808z - 1.4832}{z^3 + 0.91z^2 + 0.149z + 0.0014}$$

Estimated Transfer Function

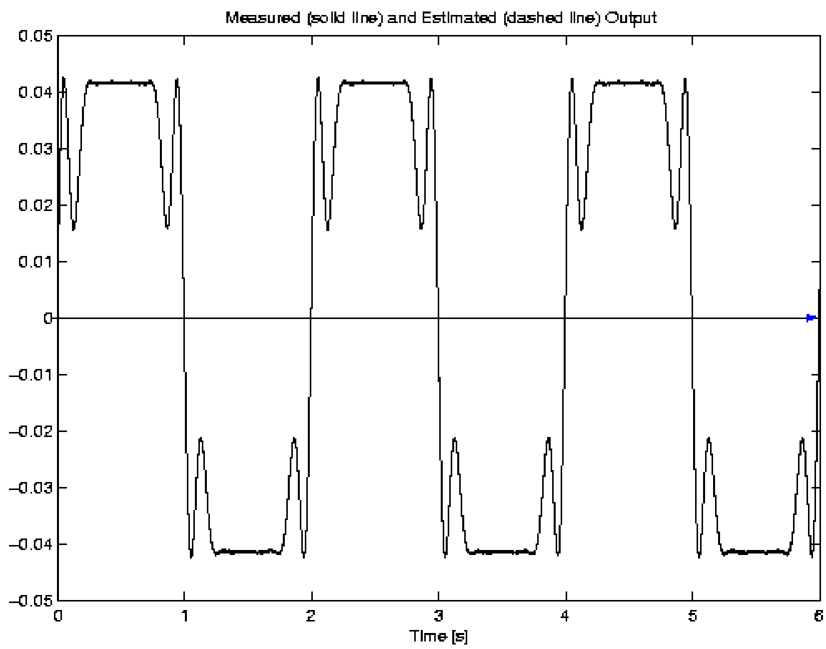
❑ The Estimated Nonlinear Model

$$\hat{N}(u_k) = 0.8829\,u_k - 0.0747\,u_k^2 - 0.4483\,u_k^3 - 0.1183\,u_k^4$$ **Estimated nonlinear model**



True (solid line) and Estimated (dashed line) nonlinear characteristic.

❑ True and Estimated Output

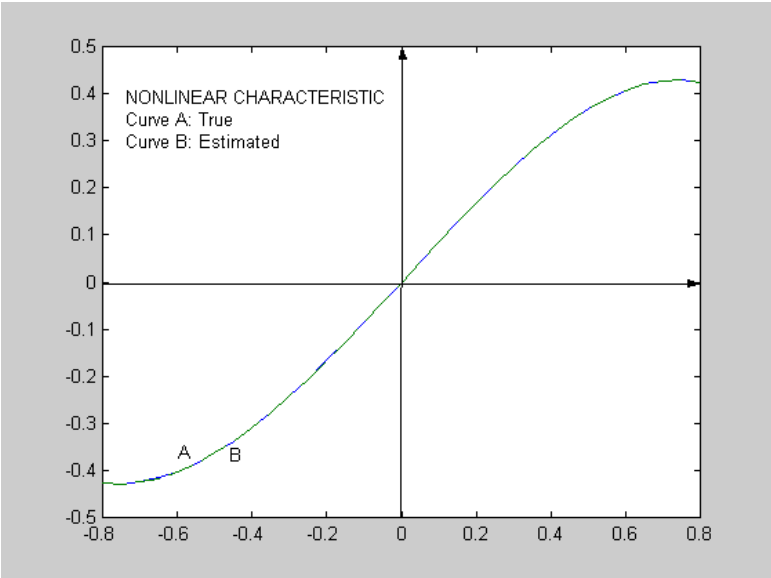


True (solid line) and Estimated (dashed line) Output.

❑ A more persistently exciting input

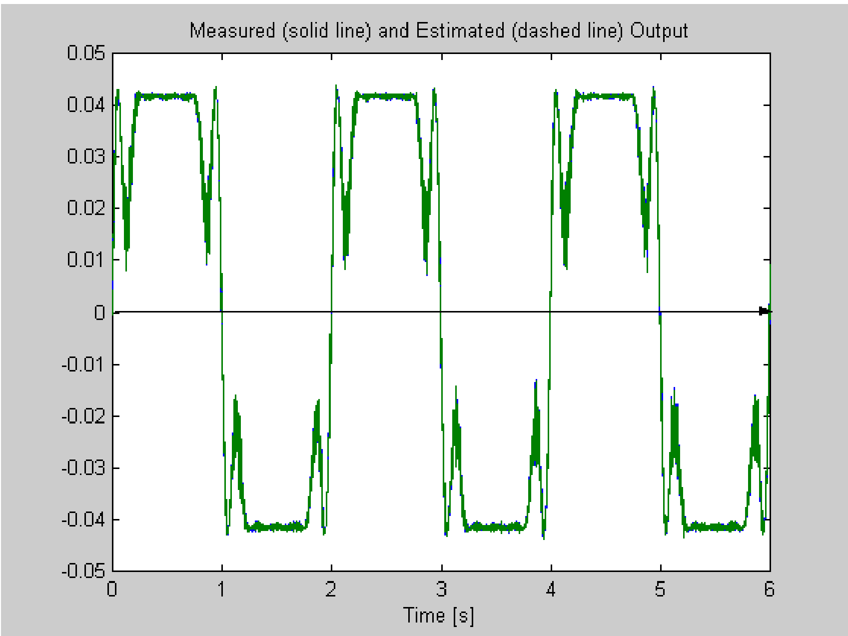
$$u_k = \sin(0.0005\pi k) + 0.5\sin(0.0015\pi k) + $$
$$+ 0.3\sin(0.0025\pi k) + 0.1\sin(0.0035\pi k) + \gamma_k$$

γ_k white noise with variance 10^{-6}



True (solid line) and Estimated (dashed line) nonlinear characteristic (indistinguishable one from the other)..

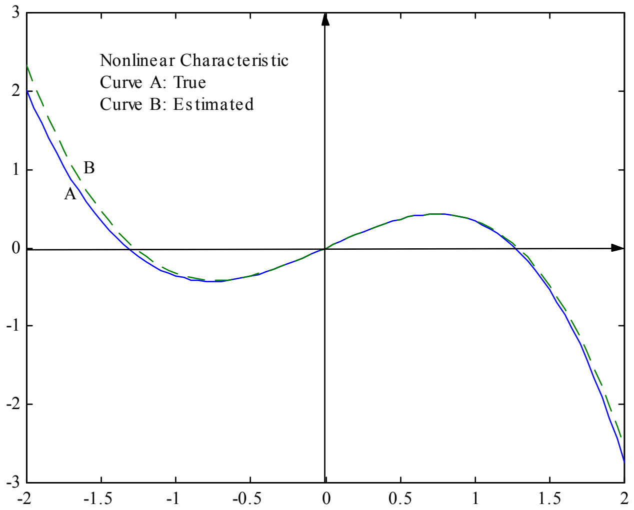
❑ True and Estimated Output



True (solid line) and Estimated (dashed line) Output.

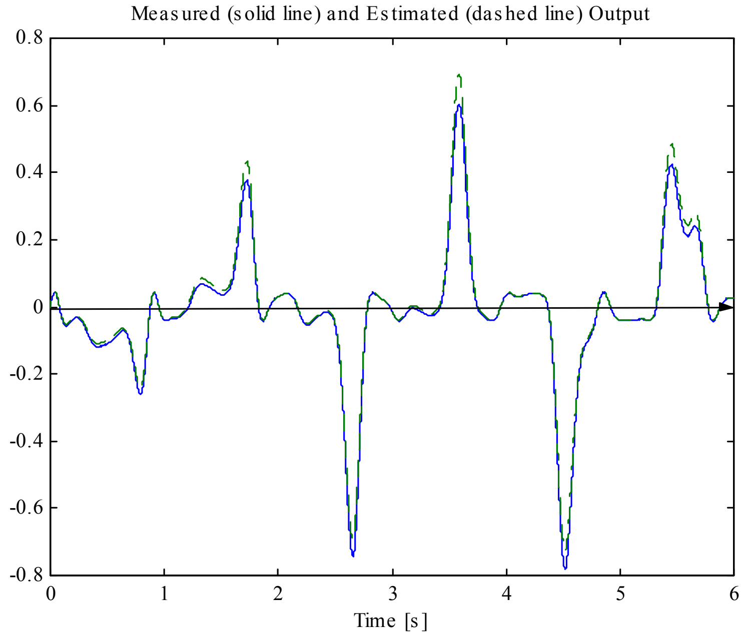
❑ An intermediate persistently exciting input

$$u_k = 2\sin(0.0005\pi k) + 0.5\sin(0.00157\pi k) + $$
$$+ 0.3\sin(0.002735\pi k) + 0.1\sin(0.003815\pi k)$$



True (solid line) and Estimated (dashed line) nonlinear characteristic

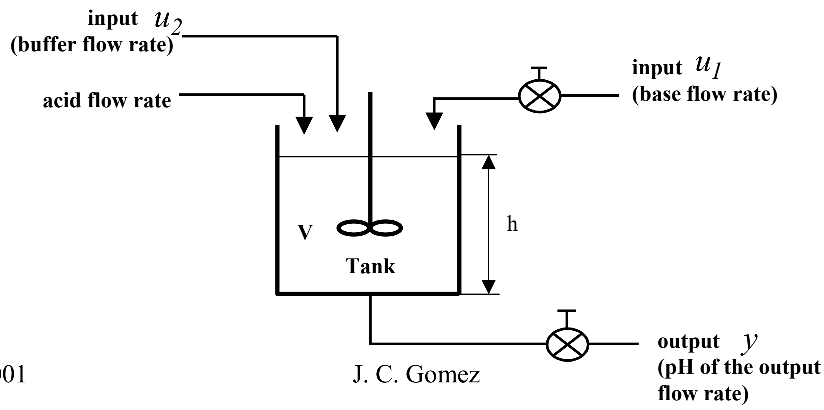
❑ True and Estimated Output



True (solid line) and Estimated (dashed line) Output.

2. Wiener model

- **The process:** pH neutralization process in a constant volume stirring tank considered in (Henson & Seborg, 1992). (Bench-scale plant at the University of California, Santa Barbara).
- The **model** was derived using the concept of reaction invariants (highly nonlinear model, with the output given in implicit form: **titration curve**).
- The **inputs** to the system are:
  - u_1 : the base flow rate
  - u_2 : the buffer flow rate
- The **output** is:
  - y : the pH of the solution in the tank.

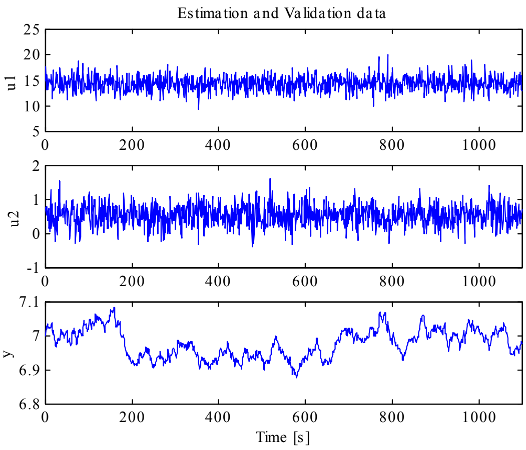


•Simulation:

- System excited with band-limited white noise around the nominal operating point.
- Linear Subsystem: Orthonormal Bases with fixed Poles at:

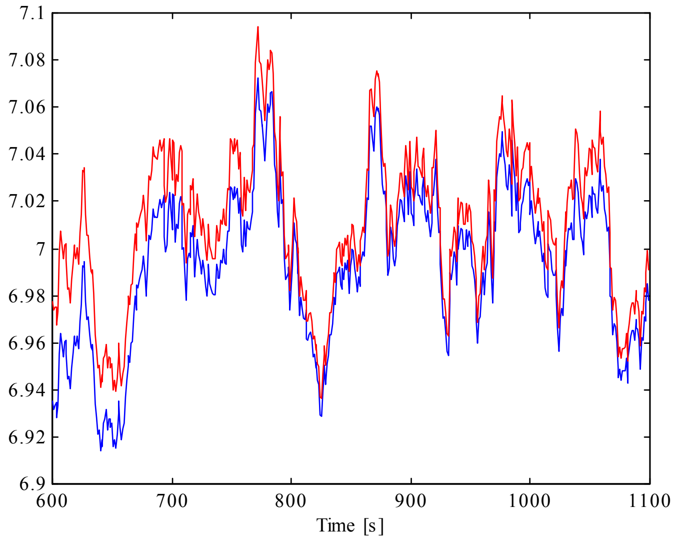
  {0.97 , 0.98 , 0.98 , 0.99, 0.99}

- Nonlinear Subsystem: 3rd. order polynomial.



Input/Output Data:

First 600 data used for **Estimation**, remaining 500 data used for **Validation**



True (blue) and Estimated (red) Output (Validation Data)

Conclusions

- Noniterative methods for the identification of **Multivariable Block-oriented Nonlinear Models** have been presented.
- The proposed methods are **numerically robust**, since they depend only on **Lest Squares Estimation** and **Singular Value Decomposition**. No nonlinear numerical optimization procedures are required.
- For the **Hammerstein** model, the method provides **consistent estimates** under weak assumptions on the persistency of excitation of the inputs, even in the presence of **coloured noise**. For the **Wiener** model, and the **Feedback** model, consistency can only be guaranteed in the noise-free case.
- The key issue is the representation of the LTI subsystem using **Orthonormal Basis Functions → deterministic regressors**.
- In addition, the use of orthonormal bases allows the incorporation of *a priori* information about system dynamics → **improvement in estimation accuracy** by choosing the poles of the bases close to the true poles.