Identification of Hammerstein-Wiener Systems using Subspace Method and Separable Least-Squares

Hajime Ase* and Tohru Katayama†
* Mathematical Technology Institute, Chofu, Tokyo 182-0014, Japan
† Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan
E-mail: tohru_katayama(at)nifty.com

Abstract

We deal with identification of Hammerstein-Wiener systems, or NLN systems, in which a linear subsystem is sandwiched by two nonlinearities. First we identify an approximate linear state space model of the NLN system by using the ORT (orthogonal projection) subspace method or PO-MOESP method. Then, initialized by the estimated state space model, we optimize the output error (OE) model, which is derived based on the basis functions expansion of nonlinearities, by using a gradient-based optimization method. Numerical results are included to show the applicability of the present approach.

1 Introduction

Identification of block-oriented nonlinear systems, which consist of dynamic linear (L) and static nonlinear (N) blocks, has received much interest in control community. In particular, quite popular are Wiener systems, Hammerstein systems, Wiener-Hammerstein (LNL) systems and Hammerstein-Wiener (NLN) systems [16, 10].

In this paper, we deal with the identification of Hammerstein-Wiener systems, for which many papers have been published in the literature under different conditions and by various methods; for example, an overparametrization method [3], a blind approach [4], iterative methods [5, 8, 26, 31], subspace methods [11, 22], an EM-based algorithm [28], an instrumental variable method [21]; moreover in many papers it is assumed that the output nonlinearity is invertible [4, 8, 11, 24, 29, 31] to get an estimate of the input of the output nonlinearity. Also, the NLN models are used in many applications, including prediction of magnetic-field [22] and process controls [30, 31], etc.

To identify the Hammerstein-Wiener system, we use an output error (OE) model derived by the basis functions expansion of nonlinearities; therefore the invertibility of the nonlinearities is not required. Adapting the method that uses the linear approximations to initialize the subsequent optimization algorithm [8, 1, 18], we first identify an approximate linear model of the Hammerstein-Wiener system by applying the ORT (orthogonal projection) [17] or PO-MOESP [25] method to input-output data as if they are generated from a linear system. Then, initialized by the estimated linear model, we apply the separable least-squares [12, 13] to optimize the mean square error of the OE model, where a version of the DDLC-based gradient search is employed [27]. It is shown by numerical simulations that the performance of the present method is quite good for the identification of Hammerstein-Wiener systems, though the computational load is bit expensive.

The organization of the paper is as follows. The problem is formulated in Section 2. Section 3 briefly reviews the linear approximation of NLN and LNL systems by defining the best linear model and the equivalent linear model, and Section 4 outlines the identification procedure. In Section 5, the OE model is derived by using the basis function expansion of nonlinearities, and the identification method based on the separable least-squares is described. The overall algorithm is summarized in Section 6. Moreover, some numerical results comparing the use of ORT and PO-MOESP methods for initializations are included in Section 7. Section 8 concludes the paper.

2 Model and problem formulation

\[ \begin{align*}
  u & \rightarrow h(\cdot) \rightarrow G(z) \rightarrow w \rightarrow f(\cdot) \rightarrow y^0 + y \\
  & \uparrow \\
  & \nu 
\end{align*} \]

Fig. 1: NLN system.

We consider the identification of a single-input single-output (SISO) Hammerstein-Wiener system in Fig. 1, where the input nonlinear block \( h(\cdot) \) and the output nonlinear block \( f(\cdot) \) are the actuator and process nonlinearities, respectively, in control applications [31].

Suppose that a realization of the linear system is given by \( G(z) = [A, B, C, D] := C(zI - A)^{-1}B + D \). Since \( v(t) = h(u(t)) \), the nonlinear system of Fig. 1 is expressed as

\[ \begin{align*}
  x(t+1) &= Ax(t) + Bh(u(t)) \\
  w(t) &= Cx(t) + Dh(u(t)) \\
  y(t) &= f(w(t)) + \nu(t)
\end{align*} \]

where \( x(t) \in \mathbb{R}^n \) is the state vector, and \( A \in \mathbb{R}^{n \times n} \), \( B \in \mathbb{R}^n \), \( C \in \mathbb{R}^{1 \times n} \), \( D \in \mathbb{R} \) are system parameters.
Assumption 1 (i) The input \( u(t) \) is a stationary Gaussian process with mean zero and finite covariance. (ii) The output noise \( v(t) \) is a stationary Gaussian white noise with mean zero and is uncorrelated with the inputs \( u(\tau), \tau = 0, \pm 1, \cdots \). (iii) The linear system \( G(z) \) is stable, and the realization \((A, B, C)\) is minimal. (iv) The two static nonlinearities are measurable functions, and the variances of outputs of the nonlinearities are bounded, i.e.

\[
E[|f(w)|^2] < \infty, \quad E[|h(u)|^2] < \infty
\]

(v) The nonlinearities are well approximated by basis function expansions [19], but their invertibility is not required (since the OE model is used). □

In this paper, we develop a method of identifying the Hammerstein-Wiener system of Fig. 1 by using given input-output data \( \{u(t), y(t)\mid t = 1, 2, \cdots, N\} \), where the initial linear models identified by the subspace methods [17, 25] together with nonlinear models are optimized by the separable least-squares [12].

It may be noted that we cannot identify the three block elements independently in the NLN structure of Fig. 1 from the input-output data, since for any constants \( a, b, c \) with \( abc = 1 \), the realizations \((f, G, h)\) and \((af, bG, ch)\) cannot be distinguished. Thus, we should normalize the gains of three elements to get a unique parametrization [4]. In this paper, we use internal variables to adjust these gains [2], and restrict the domain of the nonlinearities to some interval \([-\delta, \delta]\), \( \delta > 0 \).

3 Linear approximation of NLN and LNL systems

![LNL system](image)

Fig. 2: LNL system.

In this section, we define the equivalent linear models and consider the best linear approximations of NLN and LNL systems.

Definition 1 The equivalent linear model of the NLN system of Fig. 1 is defined by

\[
G_e(z) := F^e G(z) H^e
\]

where \( F^e \) and \( H^e \) are equivalent gains of \( f(\cdot) \) and \( h(\cdot) \), respectively; they are defined by [23]

\[
H^e = \frac{E(h(u(t))u(t))}{E(u^2(t))}, \quad F^e = \frac{E(f(w(t))w(t))}{E(w^2(t))}
\]

Similarly, the equivalent linear model of the LNL system of Fig. 2 is defined by

\[
G_e(z) := G_1(z) F^e G_2(z)
\]

where \( F^e \) is the equivalent gain

\[
F^e = \frac{E(f(v(t))v(t))}{E(v^2(t))}
\]

For the equivalent gain, see [23]. □

Definition 2 For the NLN and LNL systems, the best linear mean square approximation problem of \( y(t) \) in terms of the past inputs \( U^e_t := \{u(t), u(t-1), \cdots\} \) is to find \( G_d(z) \) that minimizes the mean square error

\[
E\{y^2(t) - G_d(z)u(t)^2\}
\]

The optimal \( G_d(z) \) is called the best linear approximation, or the best linear model [9]. □

It is shown [1] that for the LNL system of Fig. 2, if the input is a zero mean Gaussian process, the best linear approximation of the output based on the past inputs is given by the orthogonal projection of the output onto the past inputs, i.e.

\[
y^d(t) := G_d(z)u(t) = \hat{E}\{y(t) \mid U^e_t\}
\]

where \( \hat{E}\{\cdot \mid \cdot\} \) denotes the orthogonal projection.

Moreover, using the result of Bussgang [7], it has been shown that under the Gaussian assumption, the best linear model \( G_d(z) \) is equal to the equivalent linear model, i.e. \( G_d(z) = G_e(z) \) holds [9, 2], where \( G_e(z) \) is given by (3) and that it is consistently identified by using the ORT subspace method [18]. The following example is motivated by [9].

Example 1 Consider the best linear approximation problem for a Hammerstein system and a Wiener system with a cubic nonlinearity shown in Fig. 3, where \( h(u) = u^3, f(w) = w^3 \) with \( b \neq 0 \).

![Hammerstein and Wiener systems](image)

Fig. 3: (a) Hammerstein system. (b) Wiener system.

It follows from (5) that for both nonlinear systems of Fig. 3, the best linear approximation of the output \( y(t) \) is given by the orthogonal projection

\[
y^d(t) = \hat{E}\{y(t) \mid u(t), u(t-1), \cdots\}
\]

where \( K_0 = R_{yu}(0)/R_{uu}(0), K_1 = R_{yu}(1)/R_{uu}(0) \).
As mentioned above, if the input is Gaussian, the best linear models of Figs. 3(a) and 3(b) are given by the equivalent linear models \( G_d(z) = (1 + b z^{-1}) H^e \) and \( G_d(z) = F^e (1 + b z^{-1}) \), respectively.

Suppose that the input \( u(t) \) is a white noise with a uniform distribution \( U(-a, a) \), \( a > 0 \). For the Hammerstein system of Fig. 3(a), it follows that \( R_{uw}(0) = a^2/3 \), \( R_{uw}(0) = a^4/5 \), \( R_{uw}(1) = a^6/5 \), so that \( H^e = 3 a^2/5 \) and \( K_0 = H^e \), \( K_1 = b H^e \). Hence, the best linear model is given by \( G_d(z) = (1 + b z^{-1}) H^e \), which is exactly the equivalent linear model.

For the Wiener system of Fig. 3(b), we have

\[
R_{uw}(0) = \frac{a^2}{3} (1 + b^2), \quad R_{uw}(0) = \frac{a^4}{5} \left( 1 + \frac{10}{3} b^2 + b^4 \right)
\]

Thus the equivalent gain is given by

\[
F^e = \frac{a^2 (3 + b^2) (1 + 3b^2)}{5 (1 + b^2)}
\]

On the other hand, it follows that

\[
R_{wy}(0) = \frac{a^4 (3 + 5b^2)}{15}, \quad R_{wy}(1) = \frac{a^4 b (5 + 3b^2)}{15}
\]

so that

\[
K_0 = \frac{a^2 (3 + 5b^2)}{5}, \quad K_1 = \frac{a^2 b (5 + 3b^2)}{5}
\]

Hence, the best linear model becomes

\[
G_d(z) = \frac{a^2 (3 + 5b^2)}{5} \left[ 1 + \frac{b (5 + 3b^2)}{3 + 5b^2} z^{-1} \right]
\]

which is not equal to \( F^e G(z) \) unless \( b^2 = 1 \). This implies that in general, the best linear model is not the equivalent linear model for the Wiener system of Fig. 3(b) if the input is non-Gaussian.

For the Hammerstein-Wiener system in Fig. 1 with two nonlinear elements, we see that even if the input \( u(t) \) to the nonlinearity \( h(\cdot) \) is Gaussian, the output \( v(t) \) is non-Gaussian, so that the input \( w(t) \) to the nonlinearity \( f(\cdot) \) is non-Gaussian. Thus, we see that the best linear model of an NLN system is different from the equivalent linear model, i.e. \( G_d(z) \neq F^e G(z) H^e \).

4 Identification procedure of NLN systems

Identification procedure of the Hammerstein-Wiener system of Fig. 1 is divided into two steps:

**Step 1.** We apply the ORT or PO-MOESP subspace identification method to the input-output data to obtain an approximate linear model of the plant \( G_d(z) = (A^{(0)}, B^{(0)}, C^{(0)}, D^{(0)}) \), which, we assume, is close to a realization of the equivalent linear model \( G_e(z) = (A, B H^e, F^e C, F^e D H^e) \).

**Step 2.** Initialized by a realization of the linear model \( (A^{(0)}, B^{(0)}, C^{(0)}, D^{(0)}) \), we minimize the mean square OE by using a gradient-based search [27], where the linear part is expressed in terms of the DDLC-based parametrization [20].

For the computational procedures of the ORT and PO-MOESP subspace methods, see [17] and [25], respectively. Note that the standard conditions for subspace identification methods such as the PE (persistently exciting) condition of the input must be satisfied.

5 Identification of NLN system by separable least-squares

5.1 Output error model

We assume that the nonlinearities \( f(\cdot) \) and \( h(\cdot) \) are respectively modeled linear in parameters in terms of the basis function expansions

\[
y^0(t) = f(w(t)) \approx \sum_{l=1}^{L} \beta_l \zeta_l(w(t))
\]

and

\[
v(t) = h(u(t)) \approx \sum_{l=1}^{L} \alpha_l \zeta_l(u(t))
\]

where \( \zeta_l(\cdot), l = 1, \cdots, L \) are basis functions. Let \( \zeta_l(u(t)) := [\zeta_l(u(t)) \cdots \zeta_l(u(t))]^T \in \mathbb{R}^L \), and \( \alpha := [\alpha_1 \cdots \alpha_L]^T \in \mathbb{R}^L \). Then, it follows from (7) that the estimate of the output of \( h \) is given by

\[
\hat{v}(t) := [\alpha_1 \cdots \alpha_L] \zeta(u(t)) = \alpha^T \zeta(u(t))
\]

Also, let the overparametrized matrices be (33)

\[
B := B_0^T = \begin{bmatrix} B_1 \alpha_1 & \cdots & B_1 \alpha_L \\ \vdots & \ddots & \vdots \\ B_L \alpha_1 & \cdots & B_L \alpha_L \end{bmatrix} \in \mathbb{R}^{n \times L}
\]

\[
D := D_0^T = [D_0 \alpha_1 \cdots D_0 \alpha_L] \in \mathbb{R}^{1 \times L}
\]

Moreover, define the parameter vector

\[
\theta := \begin{bmatrix} \text{vec}(A) \\ \text{vec}(B) \\ \text{vec}(C) \\ \text{vec}(D) \end{bmatrix} \in \mathbb{R}^{nq}
\]

where \( n_q := n^2 + (n + 1)L + n \), and vec(·) denotes a vector that stacks all the column vectors of a matrix on top of each other. Substituting (8) into (1a) and (1b) yields

\[
\hat{x}(t + 1, \theta) = A \hat{x}(t, \theta) + \hat{B} \zeta(u(t))
\]

\[
\hat{w}(t, \theta) = C \hat{x}(t, \theta) + \hat{D} \zeta(u(t))
\]

where we write \( \hat{x}(t, \theta) \) and \( \hat{w}(t, \theta) \) to show their dependence on the parameter vector \( \theta \). It follows from (6) that the noise-free output is approximated as
and of Wiener-Hammerstein systems \([1, 18]\). Fig. 4: Identification of NLN system by the OE model.

\[
y^0(t) = \sum_{i=1}^{L} \beta_i \zeta_i(\hat{w}(t, \theta)) = \zeta^T(\hat{w}(t, \theta))\beta
\]  

(11)

where \( \beta := [\beta_1, \ldots, \beta_L]^T \in \mathbb{R}^L \). It should be noted that \( \hat{y}(t) = y^0(t) \) holds, since \( \nu(t) \) is a white noise with mean zero. Thus (10) and (11) form an OE model for the NLN system, and the output error becomes

\[
\epsilon(t) = y(t) - \hat{y}^0(t), \quad t = 1, \ldots, N
\]  

(12)

The whole picture of the identification procedure by the OE model is depicted in Fig. 4.

5.2 Separable least-squares
Define the augmented \( N \)-dimensional vectors

\[
\mathcal{Y} := \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \quad \hat{\mathcal{Y}} := \begin{bmatrix} \hat{y}^0(1) \\ \vdots \\ \hat{y}^0(N) \end{bmatrix}, \quad \mathcal{E} := \begin{bmatrix} \epsilon(1) \\ \vdots \\ \epsilon(N) \end{bmatrix}
\]

and the regression matrix

\[
\Phi(\theta) := \begin{bmatrix} \zeta_1(\hat{w}(1, \theta)) & \cdots & \zeta_L(\hat{w}(1, \theta)) \\ \vdots & \ddots & \vdots \\ \zeta_1(\hat{w}(N, \theta)) & \cdots & \zeta_L(\hat{w}(N, \theta)) \end{bmatrix} \in \mathbb{R}^{N \times L}
\]

5.3 Gradient method
It follows from \([12]\) that the gradient of the error vector \( \mathcal{E}(\theta) \) with respect to \( \theta \) is given by

\[
J(\theta) := \frac{\partial \mathcal{E}}{\partial \theta} = -[\Phi(\theta)]^T \mathcal{Y} \in \mathbb{R}^{N \times n_\theta}
\]  

(15)

where \( \Phi(\theta) \) is a three dimensional tensor with \( [\Phi(\theta)]_{ilj} = \frac{\partial \Phi_i(\theta)}{\partial \theta_j} \). Since \( \Phi_i(\theta) = \zeta_i(\hat{w}(t, \theta)) \), we get

\[
[\Phi(\theta)]_{ilj} = \frac{\partial \zeta_i(\hat{w}(t, \theta))}{\partial \theta_j} = \zeta_i'(\hat{w}(t, \theta)) \frac{\partial \hat{w}(t, \theta)}{\partial \theta_j}
\]  

(16)

where \( t = 1, \ldots, N; l = 1, \ldots, L; j = 1, \ldots, n_\theta \).

Differentiating (10) with respect to \( \theta_j \), we see that the partial derivatives \( \frac{\partial \hat{w}(t, \theta)}{\partial \theta_j} \) satisfy

\[
\frac{\partial \hat{w}(t, 1 + \theta)}{\partial \theta_j} = A \frac{\partial \hat{x}(t, \theta)}{\partial \theta_j} + \frac{\partial A}{\partial \theta_j} \hat{x}(t, \theta) + \frac{\partial B}{\partial \theta_j} \zeta(u(t))
\]  

(17a)

\[
\frac{\partial \hat{w}(t, \theta)}{\partial \theta_j} = C \frac{\partial \hat{x}(t, \theta)}{\partial \theta_j} + \frac{\partial C}{\partial \theta_j} \hat{x}(t, \theta) + \frac{\partial D}{\partial \theta_j} \zeta(u(t))
\]  

(17b)

To get the partial derivatives \( \frac{\partial \mathcal{E}}{\partial \theta_j} \), we compute (17) together with (10) for \( t = 1, \ldots, N \) with some initial conditions. By using (16), the \( j \)-th column of (15) is given by

\[
\begin{bmatrix} \frac{\partial \epsilon(1)}{\partial \theta_j} \\ \vdots \\ \frac{\partial \epsilon(N)}{\partial \theta_j} \end{bmatrix} = -\begin{bmatrix} \frac{\partial \zeta_1(\hat{w}(1, \theta))}{\partial \theta_j} & \cdots & \frac{\partial \zeta_L(\hat{w}(1, \theta))}{\partial \theta_j} \\ \vdots & \ddots & \vdots \\ \frac{\partial \zeta_1(\hat{w}(N, \theta))}{\partial \theta_j} & \cdots & \frac{\partial \zeta_L(\hat{w}(N, \theta))}{\partial \theta_j} \end{bmatrix} \Phi^T \mathcal{Y}
\]

Note that since \( \Phi^T(\theta) \in \mathbb{R}^{L \times N} \) and \( \mathcal{Y} \in \mathbb{R}^N \), the right-hand side of the above expression is \( N \)-dimensional. Thus, we get the Jacobian of (15).

We briefly explain the DDLC-based gradient method \([27]\). Let \( \hat{\theta}(r) \) be the \( r \)-th estimate of \( \theta \) in the gradient-based search, i.e.

\[
\hat{\theta}(r) := \begin{bmatrix} \text{vec}(A^{(r)}) \\ \text{vec}(B^{(r)}) \\ \text{vec}(C^{(r)}) \\ \text{vec}(D^{(r)}) \end{bmatrix} \in \mathbb{R}^{n_s}
\]  

(18)

To use the DDLC parametrization \([20]\), we define

\[
Q^{(r)} = \begin{bmatrix} (A^{(r)})^T \otimes I_n - I_n \otimes (C^{(r)}) \\ (B^{(r)})^T \otimes I_n \\ -I_n \otimes (C^{(r)}) \\ 0_{1 \times n^2} \end{bmatrix} \in \mathbb{R}^{n_s \times n^2}
\]  

(19)

where \( \otimes \) denotes the Kronecker product \([14]\). For the DDLC parametrization, the search directions are to be restricted to the orthogonal complement of \( \text{Im}(Q^{(r)}) \) \([20, 25, 27]\). Let the SVD of \( Q^{(r)} \) be

\[
Q^{(r)} = [U_1, U_2] \begin{bmatrix} S_1 & 0 \\ 0 & V^T \end{bmatrix} \quad S_1 \in \mathbb{R}^{n_s \times n^2}
\]
where $U_1 \in \mathbb{R}^{n_u \times n^2}$ and $U_2 \in \mathbb{R}^{n_u \times (n_u - n^2)}$ [14]. Since $P := U_2$ is orthogonal to $Q^{(r)}$, we define $R^{(r)} := J(\theta^{(r)})P \in \mathbb{R}^{N \times (n_u - n^2)}$. Then, the basic iteration algorithm is given by

$$d^{(r)} = -[(R^{(r)})^T R^{(r)} + \Lambda]^{-1}(R^{(r)})^T \mathcal{E}(\theta^{(r)})$$

where $\lambda \geq 0$ is a stabilizing parameter, and the step size $\sigma > 0$ is determined by a simple line search.

6 Overall algorithm of identification

The algorithm for identifying the Hammerstein-Wiener systems is summarized.

Algorithm for identifying NLN systems

Step 1. Based on input-output data, identify a linear model $G_d(z) = [A(0), B(0), C(0), D(0)]$ of the NLN system by the ORT or PO-MOESP method, and define $B(0) = B(0) \alpha(0)$ and $D(0) = D(0) \alpha(0)$. Set $r = 0$ and the initial vector $\theta^{(0)}$ as in (18). Then, go to Step 2.

Remark 1 Suppose that $\eta := (A(0), B(0), C(0), D(0))$ are given, and that $f(w) = w$ temporarily. Then, since $\tilde{w}(t, \eta) = \hat{g}(t)$ is linear with respect to $\alpha$, the initial estimate $\alpha(0)$ in Step 1 is obtained by solving a linear least-squares problem.

Step 2. (Jacobians) Let initial conditions be

$$x(1, \theta^{(r)}) = 0, \quad \frac{\partial x(1, \theta^{(r)})}{\partial \theta_j} = 0, \quad j = 1, \cdots, n_\theta$$

Compute (10), (17a) and (17b) for $t = 1, \cdots, N$. Then, compute the Jacobian $J(\theta^{(r)})$.

Step 3. Compute the SVD of (19) to get $R^{(r)}$. Then, update the current vector $\theta^{(r)}$ according to (20) and (21).

Step 4. (Convergence) Compute the value of the output error of (14) for $\theta^{(r+1)}$. If the convergence condition is satisfied, retrieve the systems parameters

$$\theta^{(r+1)} \rightarrow (A, \hat{B}, C, \hat{D})$$

Then, go to Step 5. Otherwise, set $r := r + 1$ and go to Step 2.

Step 5. (Estimation of nonlinearities) The coefficients of $f(\cdot)$ are given by $\beta = \Phi(\theta^{(r+1)}) \gamma$. Then, compute the rank-one approximation of the overparametrized input matrices, i.e. $\left[ \begin{array}{c} \hat{B} \\ \hat{D} \end{array} \right] \simeq \hat{U}_1 \hat{\sigma}_1 \hat{V}_1^T$ [3, 15], where $\hat{\sigma}_1$ is the largest singular value, and where $\hat{U}_1 \in \mathbb{R}^{n+1}$ and $\hat{V}_1 \in \mathbb{R}^L$ are respectively the left and right singular vectors corresponding to $\hat{\sigma}_1$. Put $\gamma := \hat{V}_1(1)$. Then, we get the coefficients $\alpha$ for $h(\cdot)$ with $\alpha(1) = 1$ by $\alpha := \hat{V}_1(\gamma)$, and define $\left[ \begin{array}{c} \hat{B} \\ \hat{D} \end{array} \right] := \hat{U}_1 \hat{\sigma}_1 \gamma$. Thus, the estimates of $(A, B, C, D)$ and $\alpha, \beta$ are obtained.

7 Numerical examples

Numerical examples are included to show the applicability of the identifiability method derived in this paper.

Case 1: Consider the Hammerstein-Wiener system of Fig. 1, where the linear block is a 5th-order system $G(z) = 2.8(0.0275z + 0.0551) \overline{z^2 - 2.3443z^4 + 3.081z^3 - 2.527z^2 + 1.2415z - 0.3686}$ and the nonlinearities are 3rd-order polynomials of the form $f(w) = (4/9)w^3$, $h(u) = 1.5u - (2/9)u^3$. It should be noted that since the Legendre polynomial expansion with $L = 3$ is used, the true nonlinearities are in the model set formed by basis functions.

We assume that the input $u(t)$ and the noise $\nu(t)$ are uncorrelated Gaussian white noises with means zero; the simulation conditions are shown in Table 1, where the standard deviations $\sigma_u, \sigma_\nu$ and the output $S/N$ ratio are sample values, and where $k$ is the number of block rows in subspace identification methods.

| \hline $\sigma_u$ & $\sigma_\nu$ & $\sigma_\nu$ & $S/N$ & $N$ & $k$ \\
|---|---|---|---|---|---|
| 0.5 & 0.6163 & 0.3758 & 0.02 & 25.4 & 30,000 & 15 \\
\hline

We have performed Monte Carlo simulations using $M = 50$ sets of input-output data with $N = 30,000$. A typical convergence of the OE norm, i.e. the square root $(V(\theta))^{1/2}$ of (14), is displayed in Fig. 5, where it takes about 37.3 sec per iteration.

| \hline Initial Optimized \\
|---|---|
| \hline PO-MOESP | Mean 0.23155 0.02596 |
| Std 0.00708 0.00071 |
| ORT | Mean 0.23097 0.02573 |
| Std 0.00696 0.00077 |
\hline

Table 2 displays the average and standard deviation of the OE norm initialized by the PO-MOESP and ORT methods. The initial performance is computed by using $\theta^{(0)}, \beta^{(0)} = \Phi(\theta^{(0)}) \gamma$ and the input-output data, see Remark 1.
We clearly see that the initial performance is substantially improved by the optimization, resulting in the good performance close to a lower bound. Note that a lower bound of $(E\{V_N(0)\})^{1/2}$ is given by the standard deviation of output noise $\sigma_v = 0.02$ in the prediction error method using the OE model.

Fig. 6: Initial estimates of poles of the linear system by ORT (left) and PO-MOESP (right) methods.

Fig. 7: Optimized estimates of poles of the linear system with initial estimates of Fig. 6 by ORT (left) and PO-MOESP (right) methods.

Fig. 8: Initial Bode magnitudes estimated by ORT (left) and PO-MOESP (right) methods.

We see from Fig. 6 that the estimates of the poles of $A$-matrix by the ORT method is much better than those by the PO-MOESP. Also, Fig. 7 displays the plot of the optimized poles with initial estimates shown in Fig. 6, where we see that two results initialized by the ORT and PO-MOESP methods are not distinguishable; this fact is also confirmed by comparing Figs. 8 and 9 that show the initial and optimized Bode plots corresponding to Figs. 6 and 7, respectively.

Fig. 9: Optimized Bode magnitudes with initial estimates of Fig. 8 by ORT (left) and PO-MOESP (right) methods.

Moreover, as in Fig. 10, the optimized nonlinearities show a very good agreement with the true nonlinearities; thus we can safely say that the identification results for the case where the nonlinearities are in the model sets are quite good in the sense that the optimized performance is close to its lower bound.

Case 2: Suppose that the linear system is given by the 4th-order model

$$G(z) = \frac{0.6(0.1208 + 0.1812)}{z^4 - 1.992z^3 + 2.203z^2 - 1.841z + 0.8941}$$

and the nonlinearities are

$$f(w) = \frac{w}{\sqrt{0.1 + 0.9w^2}}$$

$$h(u) = \begin{cases} 
1.5(u - 0.5), & u \geq 0.5 \\
0, & -0.5 < u < 0.5 \\
1.5(u + 0.5), & u \leq -0.5 
\end{cases}$$

Note that $f(\cdot)$ and $h(\cdot)$ are not in the model set formed by the Legendre polynomial expansion, so that a higher order expansion is needed to approximate them; we choose $L = 13$ by trial and error. The simulation conditions in Case 2 are shown in Table 3, where the number of input-output data is $N = 10,000$.

| $\sigma_u$ | $\sigma_w$ | $\sigma_yo$ | $\sigma_v$ | $S/N$ | $N$ | $k$ |
|---|---|---|---|---|---|---|
| 0.6 | 0.5181 | 0.7296 | 0.03 | 27.7 | 10,000 | 15 |
8 Conclusions

We have developed a method of identifying SISO Hammerstein-Wiener systems using the OE model derived by the basis functions expansion of nonlinearities. First we identify a linear state space model of the NLN system by using a subspace method. Then, initialized by the estimated state space model, we apply the separable least-squares to optimize the OE model by using a gradient-based search.

Table 4: Performance of identification initialized by the estimates due to PO-MOESP and ORT methods, where the number of Monte Carlo simulations is $M = 50$.

| Method     | Initial Mean | Optimized Mean |
|------------|--------------|----------------|
| PO-MOESP   | 0.27203      | 0.06021        |
| ORT        | 0.27304      | 0.06020        |

Table 4 shows the performance of simulations initialized by the ORT and PO-MOESP methods. The average OE norm is 0.0620, which is twice as large as a lower bound $\sigma_v = 0.03$, i.e. the standard deviation of the output noise. This degradation is due to the fact that the nonlinearities are not in the model sets.

Fig. 11: Initial estimates of poles of the linear system by ORT (left) and PO-MOESP (right) methods.

Fig. 12: Optimized estimates of poles of the linear system with initial estimates of Fig. 11 due to ORT (left) and PO-MOESP (right) methods.

Fig. 13: Initial Bode magnitudes estimated by ORT (left) and PO-MOESP (right) methods.

Fig. 14: Optimized Bode magnitudes with initial estimates of Fig. 13 due to ORT (left) and PO-MOESP (right) methods.

Fig. 15: Optimized estimates of input nonlinearity $h(\cdot)$ (left) and output nonlinearity $f(\cdot)$ (right) initialized by the estimates due to ORT method.
Two numerical results are included to show the applicability of the present approach; the performance of identification is quite good for both initializations by the ORT and PO-MOESP methods. Thus we can use any subspace identification method for initializations. A computational cost to solve nonconvex optimization problems is rather expensive; thus a future problem is to reduce it by using a more efficient method.

References

[1] H. Ase and T. Katayama, A subspace-based identification of Wiener-Hammerstein benchmark model, *Control Engineering Practice*, Vol. 44, pp. 126–137, November 2015.

[2] H. Ase and T. Katayama, A subspace-based identification of two-channel Wiener systems, *Proc. 17th IFAC Symposium on System Identification (SYSID 2015)*, Beijing, October 2015, pp. 638–643.

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

[4] E. W. Bai, A blind approach to the Hammerstein-Wiener model identification, *Automatica*, Vol. 38, No. 6, pp. 967–979, 2002.

[5] D. Bauer and B. Ninness, Asymptotic properties of least-squares estimation of Hammerstein-Wiener models, *International Journal of Control*, Vol. 75, No. 1, pp. 35–51, 2002.

[6] J. Bruls, C. T. Chou, B. R. J. Haverkamp and M. Verhaegen, Linear and nonlinear system identification using separable least-squares, *European Journal of Control*, Vol. 5, No. 1, pp. 116–128, 1999.

[7] J. J. Bussgang, Crosscorrelation functions of amplitude-distorted Gaussian signals, Technical Report No. 216, Research Laboratory of Electronics, MIT, 1952.

[8] P. Crama and J. Schoukens, Hammerstein-Wiener system estimator initialization, *Automatica*, Vol. 40, No. 9, pp. 1543–1550, 2004.

[9] M. Enqvist and L. Ljung, Linear approximations of nonlinear FIR systems for separable input processes, *Automatica*, Vol. 41, No. 3, pp. 459–473, 2005.

[10] F. Giri and E. W.-B. Bai (eds.), *Block-oriented Nonlinear System Identification*, Springer, 2010.

[11] I. Goethals, K. Pelekmans, L. Hoegaerts, J. Suykens and B. De Moor, Subspace intersection identification of Hammerstein-Wiener systems, *Proc. 44th IEEE CDC & ECC 2005*, Seville, 2005, pp. 7108–7113.

[12] G. H. Golub and V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, *SIAM J. Numerical Analysis*, Vol. 10, No. 2, pp. 413–432, 1973.

[13] G. H. Golub and V. Pereyra, Separable nonlinear least squares: the variable projection method and its applications, *Inverse Problems*, Vol. 19, No. 2, pp. 1–26, 2003.

[14] G. H. Golub and C. F. Van Loan: *Matrix Computations* (4th ed.), The Johns Hopkins University Press, 2013.

[15] J. C. Gómez and E. Baeyens, Subspace identification of multivariable Hammerstein and Wiener models, *CD-ROM Preprints 15th IFAC World Congress*, Barcelona, 2002, pp. 55–60.

[16] R. Haber and H. Unbehauen, Structure identification of nonlinear dynamic systems - A survey on input/output approaches, *Automatica*, Vol. 26, No. 4, pp. 651–677, 1990.

[17] T. Katayama, *Subspace Methods for System Identification*, Springer, 2005.

[18] T. Katayama and H. Ase, Linear approximation and identification of MIMO Wiener-Hammerstein systems, *Automatica*, Vol. 71, pp. 118–124, September 2016.

[19] L. Ljung, *System Identification – Theory for the User* (2nd ed.), Prentice-Hall, 1999.

[20] T. McKelvey, A. Helmersson and T. Riba, Data driven local coordinates for multivariable linear systems and their applications to system identification, *Automatica*, Vol. 40, No. 9, pp. 1629–1635, 2004.

[21] B. Ni, M. Gilson and H. Garnier, Refined instrumental variable method for Hammerstein-Wiener continuous-time model identification, *IET Control Theory & Appl.*, Vol. 7, No. 9, pp. 1276–1286, 2013.

[22] H. J. Palanthandalam-Madapusi, D. S. Bernstein and A. J. Ridley, Identifying periodically switching block-structured models to predict magnetic-field fluctuations, *Control Systems Magazine*, Vol. 27, No. 5, pp. 109–123, 2007.

[23] J. B. Roberts and P. D. Spanos, *Random Vibration and Statistical Linearization*, Wiley, 1990 (Dover 2003).

[24] M. Schoukens, E. W. Bai, and Y. Rolain, Identification of Hammerstein-Wiener Systems, *Proc. 16th IFAC Symposium on System Identification*, Brussels, July 2012, pp. 274–279.

[25] M. Verhaegen and V. Verdult, *Filtering and System Identification*, Cambridge Univ. Press, 2007.

[26] J. Vörös, An iterative method for Hammerstein-Wiener systems parameter identification, *Journal of Electrical Engineering*, Vol. 55, Nos. 11–12, pp. 328–311, 2004.

[27] A. Wills and B. Ninness, On gradient-based search for multivariable system estimates, *IEEE Trans. Automatic Control*, Vol. 53, No. 1, pp. 298–306, 2008.

[28] A. Wills, T. B. Schön, L. Ljung and B. Ninness, Identification of Hammerstein-Wiener models, *Automatica*, Vol. 49, No. 1, pp. 70–81, 2013.

[29] F. Yu, Z. Mao and M. Jia, Recursive identification for Hammerstein-Wiener systems with dead-zone input nonlinearity, *Journal of Process Control*, Vol. 23, No. 9, pp. 1108–1115, 2013.

[30] D. Zambrano, S. Tayamon, B. Carlsson and T. Wigren, Identification of a discrete-time nonlinear Hammerstein-Wiener model for a selective catalytic reduction system, *2011 American Control Conference*, San Francisco, 2011, pp. 78–83.

[31] Y. Zhu, Estimation of an N-L-N Hammerstein-Wiener model, *Automatica*, Vol. 38, No. 9, pp. 1607–1614, 2002.