A blind approach to identification of Hammerstein-Wiener systems corrupted by nonlinear-process noise

Jiandong Wang, Tongwen Chen, Long Wang

Abstract—This paper proposes a new blind approach to identification of Hammerstein-Wiener models, where a linear dynamics is embedded between two static nonlinearities. The blind approach directly aims at estimating immeasurable inner input and output, with noise effects in consideration. By exploiting input’s piece-wise constant property, the parameters of the inverse output nonlinearity and the denominator of the linear dynamics are consistently identified via an iterative instrumental-variable-based method from the output measurements only; next, a subspace direct equalization method estimates the immeasurable inner input. The blind approach does not require an explicit parametrization of the input nonlinearity; moreover, the input nonlinearities are more general than static nonlinearities and may include finite-memory nonlinearities such as hysteresis-relay and hysteresis backlash. The proposed blind approach is validated and compared with the blind approach proposed by Bai in 2002 through numerical simulations.

I. INTRODUCTION

Hammerstein-Wiener models form a class of block-oriented nonlinear models, where a linear dynamic system is embedded between two static nonlinearities. The existing identification methods for Hammerstein-Wiener models may be classified as iterative methods [9], [23], [5], [11], [7], and over-parameterized methods [1], [15], [17], [16]. Besides the above methods, there are also two other methods based on specially-designed input signals, namely, piece-wise constant input signals [2], and binary input signals followed with a multistep one [13], [18]. The key step in [13] and [18] used the binary input signals to eliminate the effect of input nonlinearity, while the piece-wise constant property of input signals is exploited in [2] to remove the requirements for the measurable inner signals.

The blind approach in [2], as well as the one to be proposed in this paper, has two important merits. First, identification of Hammerstein-Wiener systems is possible without an explicit parametrization of the input nonlinearity. This merit is valuable when the input nonlinearity has many possible structures or is hard to be represented by parametric models. Second, the input nonlinearity does not have to be static, but could be certain nonlinearities with finite memories such as the hysteresis-backlash and relay-backlash relay studied in [10]. Hence, Hammerstein-Wiener models in this context do not limit to the traditional class where only the static input nonlinearity is involved. By contrast, most of aforementioned methods do require the explicit parametrization of the input nonlinearity, and cannot be applied to the nonlinearities with finite memories. As one of industrial application, the blind approach can be applied to capture the nonlinearities of faulty control valves in feedback control systems, owing to the above two merits. These nonlinearities of control valves could be static such as deadband and saturation, or have finite memories such as backlash, or be a combination of them [6].

The main contribution of this paper is to propose a new blind approach for the identification of Hammerstein-Wiener models, which has two major differences with that in [2]: (i) we consider a particular noise-corrupted case in all the involved steps instead of the noise-free one in [2]; the properties of the noise will be specified in Section II. The proposed approach is based on the instrumental-variable (IV) mechanism, and provides asymptotically unbiased and consistent parameter estimates, while that in [2] adopts the ordinary least-squares (OLS) method and the estimated parameters are biased. (ii) The proposed approach uses the subspace direct equalization to estimate the immeasurable inner input of the linear dynamics. By contrast, the blind approach in [2] has to estimate the numerator of the linear dynamics on the basis of estimated inner output, whose estimation errors are perhaps magnified in the intermediate step of estimating the numerator. Hence, the proposed approach bypasses this intermediate step and avoids the possible error magnification. Also owing to the subspace direct equalization, the proposed blind approach has no discrimination of minimum or non-minimum phase linear dynamics, whereas [2] estimated the immeasurable inner input by taking a direct inverse of the identified linear dynamics if it is minimum-phase, or by exploiting Bezout identity for non-minimum phase linear dynamics.

This paper is an extension of our recent work on Hammerstein models [22]. The main difficulty in the extension arises from a fact that noise effects become much harder to eliminate due to the output nonlinearity. In particular, the first step in [22] is to estimate the denominator parameters of linear dynamics via the bias-compensated least-squares (BCLS) method; this is possible because only one term in the regressor is correlated with the noise (Equation (8) therein). In this paper, the first step is to simultaneously estimate the denominator parameters and the parameters of the inverse
output nonlinearity, where multiple terms in the regressor are correlated with the noise (to be clarified in Section III-A later). For this situation, the BCLS method is no longer applicable.

The rest of the paper is organized as follows. Section II describes the problem to be solved and the necessary assumptions. Because the blind approach directly aims at estimating immeasurable inner signals, Sections III and IV devote to estimating these inner signals. Section V compares the performance of the proposed blind approach with that in [2] via numerical simulations. Some concluding remarks are given in Section VI.

Some standard notation is used throughout the paper. Symbol $\mathbb{Z}_+$ stands for the set of nonnegative integers, $\|\cdot\|_2$ denotes the Euclidean norm, the superscript $(T)$ represents the transpose, and $q$ is a forward shift operator, e.g., $q x(t) = x(t + 1)$.

II. Problem Description

Consider a discrete-time Hammerstein-Wiener model depicted in Fig. 1. We make the following assumptions throughout the paper:

A1. The linear system and noise dynamics can be described by an AutoRegressive with eXternal input (ARX) model,

$$x(t) = \frac{B(q)}{A(q)} v(t) + \frac{1}{A(q)} e(t),$$  \hspace{1cm} (1)

where

$$A(q) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n},$$

$$B(q) = b_1 q^{-1} + b_2 q^{-2} + \cdots + b_n q^{-n}. $$

Here the order $n$ is known a priori and $e(t)$ is white noise with zero-mean and variance $\sigma_e^2$.

A2. The input $u(t)$ is piece-wise constant for $p$ consecutive samples, where $p$ is a positive integer no less than $(n + 1)$, i.e., $p \geq n + 1$.

A3. The input $u(t)$ is persistently exciting (PE) with multiple-level amplitudes.

A4. The input nonlinearity $f_1(\cdot)$ preserves the piece-wise constant property and the PE property of $u(t)$, so that the immeasurable inner input $v(t)$ is piece-wise constant and PE. An explicit parametrization of $f_1(\cdot)$ is not assumed to be known.

A5. The numerator $B(q)$ of $G(q)$ does not have a zero at 1, i.e., $\sum_{j=1}^{n} b_j \neq 0$.

A6. The output nonlinearity $f_O(\cdot)$ is static and one-to-one so that its inverse $f_O^{-1}(\cdot)$ exists with a linear-parameters representation,

$$x(t) = \sum_{i=1}^{m} r_i \varphi_i(y(t)), $$  \hspace{1cm} (2)

where $r_1 \equiv 1$ and $\varphi_i(\cdot)$’s are the user-selected basis functions.

In Fig. 1, the noise is placed in front of $f_O(\cdot)$ and is additive to the output of the linear dynamics. It stands for the nonlinear-process noise and seems more realistic from a process operation point of view [23]. Assumption 1 also assumes that the noise dynamics shares the same poles with the linear dynamics, or equivalently, the equation error of (1) is white. Owing to the space limitation, this paper is limited to the case of white equation errors only.

Assumptions A2 and A4 are fundamental characters of the blind approach. First, the piece-wise constant property of $u(t)$ in Assumption A2 is the main requirement for the blind approach. It is satisfied in several scenarios. A common one arises from the so-called under-sampling approach [2], that is, $p$ consecutive samples of $u(t)$ are kept the same by user’s design, or from the fast-sampling output approach — see Example 1 in Section V. Second, Assumption A4 is not only obviously satisfied by all static nonlinearities including both continuous and discontinuous static nonlinearities, but also by certain nonlinearities with finite memories, e.g., the hysteresis-backlash and hysteresis-relay nonlinearities studied in [10]. Hence, Hammerstein-Wiener models in this context do not limit to the traditional class where only the static input nonlinearity is involved.

Assumption A3 is a standard identifiability condition for nonlinear systems, see, e.g. Chapter 2 in [12]. Assumption A5, that is, the steady-state gain of $G(q)$ is non-zero, is a mild assumption satisfied by many systems. For Hammerstein-Wiener models, “identifiability” is understood with the gain ambiguities among $f_1(\cdot)$, $G(q)$ and $f_O(\cdot)$. In Assumption A6, $r_1 = 1$ is one way of normalization to remove the gain ambiguity between $G(q)$ and $f_O(\cdot)$, some constraints are added later to remove the gain ambiguity between $f_1(\cdot)$ and $G(q)$ (see the last paragraph in Section IV). The basis functions $\varphi_i(\cdot)$ are typically polynomials, cubic spines, or radial basis functions.

Under these assumptions, the objective is to identify $f_1(\cdot)$, $G(q)$ and $f_O(\cdot)$ from the measurements of $u(t)$ and $y(t)$. To reach the objective, the blind approach directly aims at estimating the immeasurable inner signals $v(t)$ and $x(t)$, which is discussed in the next two sections.

III. Estimation of $x(t)$

With the noise effects in consideration, the parameters are consistently identified via the over-parametrized method and the singular-value decomposition (SVD) technique. Next, an iterative method is proposed in order to overcome a dimension problem caused by over-parametrization.
A. Over-parameterized method

Rewrite (1) in a linear regression form as
\[ x(t) + \sum_{j=1}^{n} a_j x(t-j) = \sum_{j=1}^{n} b_j v(t-j) + e(t). \] (3)

With \( r_1 = 1 \) in Assumption A6, we replace the immeasurable signal \( x(t) \) with that in (2) to rewrite (3) as
\[ \varphi_1(y(t)) = \phi^T(t) \theta + \sum_{j=1}^{n} b_j v(t-j) + e(t), \]
where \( \phi(t) \) and \( \theta \) are defined respectively as
\[ \phi^T(t) := \begin{bmatrix} \varphi_2(y(t)) & \cdots & \varphi_m(y(t)) \\ \varphi_1(y(t-1)) & \cdots & \varphi_m(y(t-1)) \\ \vdots & \cdots & \vdots \\ \varphi_1(y(t-n)) & \cdots & \varphi_m(y(t-n)) \end{bmatrix}, \]
\[ \theta^T := \begin{bmatrix} r_2 & \cdots & r_m & a_1 & \cdots & a_1 r_m & \cdots & a_n & \cdots & a_n r_m \end{bmatrix}. \] (4)

Let \( t = kp + l \) for \( k \in \mathbb{Z}_+ \) and \( l \in [0, p) \), where \( p \) is the number of consecutive constant samples of \( u(t) \) defined in Assumption A2. Define
\[ \Delta_{\varphi_1}(k;l) := \phi(kp + l) - \phi(kp + l - 1), \] (5)
and \( \Delta_{\varphi_2}(k;l), \Delta_{\varphi_m}(k;l), \) and \( \Delta_{\varphi}(k;l) \) analogously. Subtracting two consecutive samples of \( \varphi_1(y(t)) \) gives
\[ \Delta_{\varphi_1}(k;l) = \Delta_{\varphi_1}^T(k;l) \theta + \sum_{j=1}^{n} b_j \Delta_{\varphi_m}(k;l-j) + \Delta_{\varphi}(k;l). \] (6)

Assumption A4 says that if \( u(t) \) is piece-wise constant, then \( v(t) \) inherits the same piece-wise constant property from \( u(t) \), i.e.,
\[ v(t) - v(t-1) = 0, \text{ for } (kp + 1) \leq t \leq (kp + p - 1), \] (7)
for all \( t, k \in \mathbb{Z}_+ \). The property in (7) says that
\[ \Delta_{\varphi}(k;l) = 0, \text{ for } l = 1, 2, \cdots, p - 1. \] (8)

Assumption A2, namely, \( p \geq n + 1 \), says that if \( l \in [n+1, p] \), then the set \( \{l-1, l-2, \cdots, l-n\} \) is a subset of \( \{1, 2, \cdots, p-1\} \), and all the immeasurable terms of \( \Delta_{\varphi}(k;l-j) \) in (6) become zeros. Using the property (8), (6) reduces to
\[ \Delta_{\varphi_1}(k;l) = \Delta_{\varphi_1}^T(k;l) \theta + \Delta_{\varphi}(k;l), \forall l \in [n+1, p], \] (9)
for all \( k \in \mathbb{Z}_+ \).

Let us take a careful look at (9). First, under Assumption A1, the equation error \( e(t) \) is white, so is \( \Delta_{\varphi}(k;l) \). However, it is straightforward to see that \( \Delta_{\varphi}(k;l) \) is correlated with the first \( 2m - 1 \) terms of \( \Delta_{\varphi_1}(k;l) \) in (5), namely, \( \Delta_{\varphi_2}(k;l), \cdots, \Delta_{\varphi_m}(k;l) \) and \( \Delta_{\varphi_1}(k;l-1), \cdots, \Delta_{\varphi_m}(k;l-1) \); moreover, the correlation involves the higher-order moments of \( e(t) \). If the OLS method is applied to (9) as in [2] to yield \( \hat{\theta} \), an estimate of \( \theta \), \( \hat{\theta} \) would be biased due to the above correlation. Second, (9) is a static linear regression, because it is linear in the parameter \( \theta \), and the time index is \( k \) instead of \( t \) for all the involved signals in (9).

As a result, we face a static linear regression with its error term correlated with regressors. This problem has been well studied in the field of econometrics, and can be solved by the instrumental variable (IV) method, the two-stage least-squares (2SLS) method, or the limited-information maximum likelihood method [20]. By observing a fact that \( \Delta_{\varphi_1}(k;l-2) \) does not contain \( \varphi_1(k;l) \) and \( \varphi_1(k;l-1) \) for \( i = 1, \cdots, m \), and thus are uncorrelated with the white noises \( e(kp + l) \) and \( e(kp + l-1) \) in \( \Delta_{\varphi}(k;l) \), we choose \( \Delta_{\varphi_1}(k;l-2) \) as the instrumental variable. In addition, since \( \Delta_{\varphi}(k;l) \) is white under Assumption A1, it is valid to use the 2SLS method, also known as the generalized IV method, to estimate the parameter \( \theta \). That is, \( \hat{\theta} \) is estimated by minimizing the following quadratic equation,
\[ \hat{\theta} = \arg \min_{\theta} \left\| \frac{1}{K} \sum_{k=1}^{K} \left( \Delta_{\varphi_1}(k;l) - \Delta_{\varphi_1}^T(k;l) \theta \right) \Delta_{\varphi}(k;l-2) \right\|^2_{W} \] (10)
where \( K \) is the length of the available samples, and \( W \) is a symmetric positive definite weighting matrix. The solution of (10) is readily found to be
\[ \hat{\theta} = (R^T W R)^{-1} (R^T W Y), \]
where
\[ R := \frac{1}{K} \sum_{k=1}^{K} \Delta_{\varphi_1}(k;l) \Delta_{\varphi}(k;l-2), \]
\[ Y := \frac{1}{K} \sum_{k=1}^{K} \Delta_{\varphi_1}(k;l) \Delta_{\varphi_1}(k;l-2). \]

The 2SLS method finds the optimal weighting matrix \( W \), minimizing the asymptotic covariance matrix of \( \hat{\theta} \) [20],
\[ W_{opt} = \left[ \frac{1}{K} \sum_{k=1}^{K} \Delta_{\varphi}(k;l-2) \Delta_{\varphi}^T(k;l-2) \right]^{-1}. \]

Next, we construct a matrix \( \Theta \in \mathbb{R}^{n \times m} \) composed of the entries of \( \hat{\theta} \),
\[ \Theta := \begin{bmatrix} a_1 & a_1 r_2 & \cdots & a_1 r_m \\ \vdots & \vdots & \ddots & \vdots \\ a_n & a_n r_2 & \cdots & a_n r_m \end{bmatrix}. \]

Lemma A.1 in [1] gives the optimal estimates of \( \hat{\Theta} := [r_2 \cdots r_m]^T \) and \( \hat{\Theta}^T := [a_1 \cdots a_n]^T \) as
\[ (\hat{\Theta}, \hat{\Theta}) = \arg \min_{(r,a)} \| \Theta - a \begin{bmatrix} 1 & r^T \end{bmatrix} \|^F, \] (11)
where \( F \) stands for the matrix Frobenious norm. Let the SVD of \( \Theta \) be \( \Theta = L_{\Theta} S_{\Theta} R_{\Theta}^T \). Then, the optimal estimates \( (\hat{r}, \hat{a}) \) in (11) are, in Matlab notation,
\[ \hat{r} = R_{\Theta} (2 : end,1) / R_{\Theta}(1,1), \]
\[ \hat{a} = L_{\Theta}(:,1) \cdot S_{\Theta}(1,1) \cdot L_{\Theta}(1,1). \] (12)
The consistency of \( \hat{r} \) and \( \hat{a} \) depends on that of \( \hat{\theta} \) in (10), because \( \hat{r} \) and \( \hat{a} \) in (12) are the globally optimal solutions of (11). If the matrix \( E \{ \sum_{k=1}^{K} \Delta_{y} (k; l-2) \Delta_{x}^{t} (k; l) \} \) has full rank, \( \hat{\theta} \) in (10) is consistent under the condition that the instrumental variable \( \Delta_{y} (k; l-2) \) is uncorrelated with the noise term \( \Delta_{x} (k; l) \). This condition is satisfied here because \( \Delta_{y} (k; l-2) \) does not contain \( \varphi_{i} (k; l) \) and \( \varphi_{i} (k; l-1) \) for \( i = 1, \ldots, m \), and thus are uncorrelated with the white noises \( e (kp + l) \) and \( e (kp + l - 1) \) in \( \Delta_{x} (k; l) \).

### B. Iterative method

The parameter \( \theta \) in (4) is actually the over-parameterized vector; thus, the dimension of \( \theta \), i.e., \( \dim (\theta) = nm + m - 1 \), increases quickly as the orders \( n = \dim (a) \) and \( m = \dim (r) \) increase. To overcome this problem, we are inspired by the normalized iterative method [3], [4], and propose an iterative method that takes \( \hat{r} \) or \( \hat{a} \) in (12) as the initial estimate as follows.

First, given the estimate

\[
\hat{r}^{(i-1)} = \begin{bmatrix} \hat{r}_{1}^{(i-1)} & \ldots & \hat{r}_{m}^{(i-1)} \end{bmatrix}^{T}
\]

in the \((i-1)\)-th iteration, we can rewrite (9) by isolating the parameter \( a \) as

\[
\Delta_{\hat{r}} (k; l) = \phi_{x}^{T} (k; l) a + \Delta_{x} (k; l),
\]

for \( l \in [n+1, p] \) and \( k \in \mathbb{Z}_{+} \). Here \( \Delta_{\hat{r}} (k; l) \) := \( \hat{r} (kp + l) - \hat{r} (kp + l - 1) \) with \( \hat{r} (t) = \sum_{j=1}^{m} \hat{r}_{j}^{(i-1)} \varphi_{j} (y (t)) \) (note that \( \hat{r}_{1}^{(i-1)} = 1 \)), and

\[
\phi_{x}^{T} (k; l) := - \begin{bmatrix} \Delta_{\hat{r}} (k; l-1) & \ldots & \Delta_{\hat{r}} (k; l-n) \end{bmatrix}.
\]

The BCLS method proposed by [19] can be used to estimate the parameter \( a \) based on (13); see also Section 3 of [22] for details.

Second, with the updated estimate \( \hat{a}^{(i)} \) based on (13), we rewrite (9) by isolating the parameter \( r \) as

\[
\psi (k; l) = \Psi^{T} (k; l) r + \Delta_{x} (k; l),
\]

for \( l \in [n+1, p] \) and \( k \in \mathbb{Z}_{+} \). Here \( \psi (k; l) \) := \( \sum_{j=1}^{m} \hat{a}_{j}^{(i)} \Delta_{\varphi_{j}} (k; l-j) \), and

\[
\Psi^{T} (k; l) = - \begin{bmatrix} \sum_{j=1}^{n} \hat{a}_{j}^{(i)} \Delta_{\varphi_{2}} (k; l-j) & \ldots & \sum_{j=1}^{n} \hat{a}_{j}^{(i)} \Delta_{\varphi_{m}} (k; l-j) \end{bmatrix}.
\]

Similar to (9), (14) is a static linear regression with its error term \( \Delta_{x} (k; l) \) correlated with the regressor \( \Psi (k; l) \). Hence, we adopt the same instrumental variable \( \Delta_{y} (k; l-2) \) as (10) and apply the 2SLS method to (14) to update the estimate of \( r \), which is similar to that of (10) and thus is omitted here.

In summary, with an initial estimate \( \hat{r}^{(0)} \) in (12), \( \hat{a}^{(i)} \) is first updated by the BCLS method based on (13), after which \( \hat{a}^{(i)} \) is obtained by the 2SLS method based on (14); the process is repeated by replacing \( i \) with \( i+1 \) till the estimated parameters \( \hat{a}^{(i)} \) and \( \hat{r}^{(i)} \) converge. We currently have no proof of the convergence for the iteration; however, the numerical example in Section V does demonstrate the convergent behaviors of \( \hat{a}^{(i)} \) and \( \hat{r}^{(i)} \). If the estimates converge, Theorem 1 in [19] proves that \( \hat{a}^{(i)} \) is asymptotically unbiased and consistent; so is \( \hat{r}^{(i)} \), by the arguments analogous to the results of the 2SLS method in Section III-A of this paper.

Once the parameter \( r \) is identified as \( \hat{r} = [\hat{r}_{2}, \ldots, \hat{r}_{m}] \), the immeasurable inner output \( x (t) \) is ready to be estimated as (with \( \hat{r}_{1} = 1 \)),

\[
\hat{x} (t) = \sum_{i=1}^{m} \hat{r}_{i} \varphi_{i} (y (t)).
\]

### IV. Estimation of \( v(t) \)

The section aims to estimating the immeasurable inner input \( v(t) \), after the inner output \( x(t) \) and the parameter \( a \) are estimated in Section III. The proposed subspace direct equalization method is inherited from our previous work [22] and thus only the necessary steps are given here for the completeness of presentation.

Filter \( \hat{x} (t) \) in (15) by the denominator estimate \( \hat{A} (q) \), i.e.,

\[
w (t) := \hat{A} (q) \hat{x} (t),
\]

and construct a blocked Hankel data matrix with the size of \( Lp \times (N-L) \),

\[
W (L) = \begin{bmatrix}
W (2) & W (3) & \cdots & W (K-L+1) \\
W (3) & W (4) & \cdots & W (K-L+2) \\
\vdots & \vdots & \ddots & \vdots \\
W (L+1) & W (L+2) & \cdots & W (K)
\end{bmatrix},
\]

Here \( W (k) := [w (pk+1), w (pk+2), \ldots, w (pk+p)]^{T} \), \( L \geq 1 \) is called the equalizer length, and \( K \) is the length of the data to be estimated. Selection of \( L \) affects a tradeoff between the higher computational cost and the increased estimation accuracy for a larger \( L \). It can be shown that because \( v (t) \) is PE under Assumption A3, the null space of \( W (L) \) has the dimension \((K-2L-1)\), i.e., in Matlab notation,

\[
W_N := \text{null} \left(W (L) \right) = R_W(:, (L+2):(K-L)),
\]

where \( R_W \) is from the SVD of \( W (L) = L_W S_W R^T_W \). Next, we construct a matrix \( W_N \) with the size \((L+1)(K-2L-1) \times K \),

\[
W_N = \begin{bmatrix}
W_n^T & 0 & \cdots & 0 \\
0 & W_n^T & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & W_n^T
\end{bmatrix}.
\]

A slow-rate counterpart of \( v(t) \), \( V (k) := v (pk) \), can be estimated by a least-squares method as

\[
\hat{V} = \arg \min_{V} \| W_N \cdot V \|_2^2,
\]

where \( V := [V (1) \cdots V (K)]^{T} \). If the SVD of \( W_N \) is \( W_N = L_X S_X R^T_X \), then \( \hat{V} \) is the last vector of \( R_X \), i.e.,

\[
\hat{V} = R_X(:, K).
\]
Here we impose the constraint \( \|V\|_2 = 1 \) and an assumption that the linear system \( G(q) \) has a positive gain, in order to remove the gain ambiguity between the input nonlinearity \( f_I(\cdot) \) and \( G(q) \). Thus, if necessary, the sign of \( V \) in (17) may be reversed. Owing to the property in (7), the inner input estimate \( \hat{v}(t) \) is readily obtained from \( V(k) \) by piecewise constant interpolation. The uniqueness/consistency of \( V \) was established in [22].

V. EXAMPLE

This section validates its performance via a numerical example in the comparison with the blind approach in [2].

![A sampled-data Hammerstein-Wiener model with output fast-sampling.](image)

**Example 1:** This is based on Example 2 in [2]. In Fig. 2, \( G_c(s) \) is a second-order continuous-time LTI system,

\[
\begin{align*}
\hat{z} &= \begin{bmatrix} 0 & 1 \\ -0.4 & -0.3 \end{bmatrix} \eta + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v, \\
x &= \begin{bmatrix} 0.6556 \\ 0.6556 \end{bmatrix} \eta,
\end{align*}
\]

and both the input and output nonlinearities are polynomials,

\[
\begin{align*}
v &= f_I(u_c) = 1.5u_c - 1.2u_c^2 + 0.8u_c^3 + u_c^4, \\
y_c &= f_O(x) = 0.9962x + 0.0996x^2.
\end{align*}
\]

The input nonlinearity \( f_I(\cdot) \) is changed to the hysteresis-relay and the hysteresis-backlash later. The updating period of the ZOH is \( T = 1.2 \) sec, and the input \( U(k) = u_c(kT) \), \( k = 1, 2, \ldots, 500 \), is a uniform i.i.d. random variable in \([-0.5, 0.5]\). The noise dynamics shares the same poles with \( G_c(s) \),

\[
N_c(s) = \frac{1}{s^2 + 0.3s + 0.4}.
\]

The noise source \( e \) is white noise with zero mean and variance \( \sigma_e^2 \). The output \( y_c \) is sampled with faster-sampling period \( h = T/5 = 0.24 \) sec, i.e., \( y(t) := y_c(th) \). Owing to the ZOH, the fast-rate input \( u(t) := u_c(th) \) is piece-wise constant for \( p = 5 \) consecutive samples. With the sampling period \( h = 0.24 \) sec, the true but unknown discrete-time LTI model \( G(q) \) is

\[
G(q) = \frac{0.1696q^{-1} - 0.1333q^{-2}}{1 - 1.908q^{-1} + 0.9305q^{-2}}.
\]

Under the input signal \( u(t) \) and Assumption A7, the inverse output nonlinearity \( f_O^{-1}(\cdot) \) can be approximated well by a fourth-order polynomial,

\[
x = f_O^{-1}(y) = 0.9947y - 0.0991y^2 + 0.0274y^3 - 0.0069y^4.
\]

That is, the basis function \( \varphi_i(\cdot) \) in (2) is selected as \( \varphi_i(y(t)) = y_i(t) \). The objective is to estimate the immeasurable inner signals \( v(t) \) and \( x(t) \), and identify \( f_I(\cdot), f_O(\cdot) \) and \( G(q) \) on the basis of the collected input and output data \( \{u(t), y(t)\}_{t=1}^N \).

<table>
<thead>
<tr>
<th>Noise Level</th>
<th>Proposed approach</th>
<th>Blind approach in [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma^2 = 0.001 )</td>
<td>( -1.9056 \pm 0.009 )</td>
<td>( -1.9056 \pm 0.0094 )</td>
</tr>
<tr>
<td>( \sigma^2 = 0.005 )</td>
<td>( 0.9277 \pm 0.0076 )</td>
<td>( 0.9277 \pm 0.0076 )</td>
</tr>
<tr>
<td>( \sigma^2 = 0.01 )</td>
<td>( 0.9009 \pm 0.0099 )</td>
<td>( 0.9009 \pm 0.0099 )</td>
</tr>
</tbody>
</table>

**TABLE I**

\( \hat{a} \) AND \( \hat{b} \) FROM THE PROPOSED BLIND APPROACH AND THAT IN [2]

The qualities of \( \hat{a} \) and \( \hat{b} \) play the most important role in the whole identification of Hammerstein-Wiener systems, because they are the basis for all the subsequent steps. We fix the input \( u(t) \) to be deterministic and run 100 Monte Carlo simulations. Each simulation runs for 500 sec so that the data length of \( u(t) \) and \( y(t) \) is around \( N = 2000 \). In the simulation, the noise level is \( \sigma^2 = 0.05 \). Table I lists the averaged values of \( \hat{a} \) and \( \hat{b} \) and their standard deviations obtained by two blind approaches, which clearly shows that the estimates from [2] are biased. Table II presents the results obtained by the proposed blind approach for the other three noise levels. Because the estimates in the second column of Table I and those in Table II are getting more accurate as \( \sigma^2 \) decreases, it is concluded that the estimates from the proposed blind approach are consistent.

<table>
<thead>
<tr>
<th>Noise Level</th>
<th>Proposed approach</th>
<th>Blind approach in [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma^2 = 0.001 )</td>
<td>( -1.9056 \pm 0.009 )</td>
<td>( -1.9056 \pm 0.0094 )</td>
</tr>
<tr>
<td>( \sigma^2 = 0.005 )</td>
<td>( 0.9277 \pm 0.0076 )</td>
<td>( 0.9277 \pm 0.0076 )</td>
</tr>
<tr>
<td>( \sigma^2 = 0.01 )</td>
<td>( 0.9009 \pm 0.0099 )</td>
<td>( 0.9009 \pm 0.0099 )</td>
</tr>
</tbody>
</table>

**TABLE II**

\( \hat{a} \) (2ND ROW) AND \( \hat{b} \) (3RD ROW) FROM THE PROPOSED BLIND APPROACH FOR DIFFERENT \( \sigma^2 \)'S

As discussed in Section II, the estimation of two inner signals \( x(t) \) and \( v(t) \) is the actual objective of the blind approach. A fitness in % measures the error between \( x(t) \) (or \( v(t) \)) and its estimate \( \hat{x}(t) \) (or \( \hat{v}(t) \)), e.g.,

\[
F_x = 100 \left( 1 - \frac{\|\hat{x}(t) - x(t)\|_2}{\|x(t) - \frac{1}{N} \sum_{i=1}^{N} x(t)\|_2} \right).
\]

Due to the gain ambiguity, \( \hat{x}(t) \) and \( \hat{v}(t) \) are perhaps scaled beforehand in order to be compatible with \( x(t) \) and \( v(t) \). Table III lists the averaged fitness of the two inner signals for various noise levels. For each noise level, 100 Monte Carlo simulations are implemented. In Table III, the top and bottom rows are the fitnesses \( F_x \) of \( \hat{x}(t) \) and the fitnesses \( F_v \) of \( \hat{v}(t) \), respectively. The averaged fitness \( F_x \) and \( F_v \) from the proposed blind approach are consistently higher than those from the blind approach in [2]; in addition, the standard deviations of \( F_x \) and \( F_v \) from the proposed blind approach are also smaller. This is explained by the facts that the proposed blind approach considers the noise effects and skips the estimation of the numerator \( B(q) \) to avoid estimation error magnification.
**TABLE III**

<table>
<thead>
<tr>
<th>$\sigma_e^2$</th>
<th>Proposed approach</th>
<th>Blind approach in [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>98.4569 ± 0.4822</td>
<td>97.7049 ± 0.6888</td>
</tr>
<tr>
<td></td>
<td>95.4553 ± 0.3631</td>
<td>91.2376 ± 2.8711</td>
</tr>
<tr>
<td>0.005</td>
<td>98.2811 ± 0.3723</td>
<td>97.6763 ± 1.8820</td>
</tr>
<tr>
<td></td>
<td>92.1719 ± 1.2153</td>
<td>85.5691 ± 4.3101</td>
</tr>
<tr>
<td>0.01</td>
<td>97.6991 ± 0.6472</td>
<td>91.7630 ± 2.0847</td>
</tr>
<tr>
<td></td>
<td>89.0233 ± 1.8271</td>
<td>74.9185 ± 6.9017</td>
</tr>
<tr>
<td>0.05</td>
<td>95.1076 ± 3.0999</td>
<td>80.0820 ± 2.2253</td>
</tr>
<tr>
<td></td>
<td>76.6602 ± 3.2369</td>
<td>43.3276 ± 12.2701</td>
</tr>
</tbody>
</table>

Fig. 3. A graph of $\hat{v}(t)$ and $u(t)$ in blue circles and the true input nonlinearity in red dots (Hysteresis-backlash (a), Hysteresis-relay (b)).

Finally, the input nonlinearity is modified to two nonlinearities with finite memory studied in [10]: the hysteresis-relay

$$v(t) = \begin{cases} 
0.5, & \text{if } (u(t) > 0.2) \\
\sigma(-0.2 \leq u(t) \leq 0.2 \text{ and } v(t-1) = 0.5) & \text{if } (u(t) < -0.2) \\
-0.5, & \text{if } (-0.2 \leq u(t) \leq 0.2 \text{ and } v(t-1) = -0.5) 
\end{cases}$$

and the hysteresis-backlash

$$u(t) = \begin{cases} 
u(t) + 0.2, & \text{if } u(t) \leq v(t-1) - 0.2 \\
u(t) - 0.2, & \text{if } u(t) \geq v(t-1) + 0.2 \\
v(t-1), & \text{if } v(t-1) - 0.2 < u(t) < v(t-1) + 0.2 
\end{cases}$$

Under the noise level $\sigma_e^2 = 0.001$, Fig. 3 depicts the estimated hysteresis-relay and hysteresis-backlash input nonlinearities, via the graphs of $\hat{v}(t)$ and $u(t)$ (in circle), where the true input nonlinearity (in dots) is shown for the purpose of the comparison. They demonstrate that the blind approach is applicable to certain input nonlinearities with finite memories satisfying Assumption A4. Many other identification approaches mentioned in Section I cannot be applied for such Hammerstein-Wiener models.

**VI. CONCLUDING REMARKS**

This paper proposed a new blind approach to identification of Hammerstein-Wiener models with nonlinear-process noise effects in consideration. There are also some problems to be solved. First, Example 1 demonstrates the convergent behaviors of $\hat{a}^{(t)}$ and $\hat{b}^{(t)}$; however, there is currently no proof of the convergence for the proposed iterative method. Second, as shown in Table I, the parameter estimates from the proposed blind approach have larger variances than those from Bai (2002). This implies that the selected instrumental variables are sub-optimal, but it is now unclear how to find better/optimal instrumental variables.

**REFERENCES**


