Identification of Wiener Systems: the Least Amount of a priori Information

Er-Wei Bai and John Reyland Jr.
Dept. of Electrical and Computer Engineering
University of Iowa, Iowa City, Iowa 52242
er-wei-bai@uiowa.edu

Abstract—In this paper, we investigate what constitutes the least amount of a priori information on the nonlinearity so that the linear part is identifiable in the non-Gaussian case. Three types of a priori information are considered including quadrant information, point information and locally monotone information. In all three cases, identifiability has been established and corresponding identification algorithms are developed with their convergence proofs.

Index Terms—system identification, nonlinear systems, Hammerstein systems

I. INTRODUCTION

The Wiener nonlinear system is a cascade of a linear dynamic system followed by a static nonlinearity as shown in Figure 1, and is effective in representing a class of nonlinear systems. For this reason, identification of Wiener systems has been an active research area for many years and there exist several identification algorithms in the literature [1], [3], [4], [5], [7], [8], [9], [10], [11].

Fig. 1. Wiener system.

In Wiener system identification, only the input \(u(k)'s\) and output \(y(k)'s\) are assumed to be available and the internal variable \(x(k)'s\) are not available. Obviously, with no assumptions on the system and the input, the input-output data observations do not provide any information on the unknown internal variable \(x(k)'s\). In other words, input-output data sets are not very useful and do not reveal any information on the relationship between \(u(k)\) and \(x(k)\) as well as the relationship between \(x(k)\) and \(y(k)\) unless certain conditions are made on the system and/or the input. Most of works in the literature assume one of the following three assumptions, Gaussian random input [3], [4], [7], invertibility of the unknown nonlinearity [1], [7] and a known parameterization of the unknown nonlinearity with some unknown coefficients [1], [8], [9]. If the input is Gaussian, thanks to the Bussgang theorem, identification is possible without knowing the nonlinearity [3], [4], [7]. Without the Gaussian assumption, identification becomes nontrivial. It is usually assumed that the nonlinearity is monotone and invertible [1], [7], [12]. Further, either \(y = f(x)\) or \(x = f^{-1}(y)\) is expressed by a known basis [1], [8]. Another approach is to assume that the nonlinearity is a piecewise linear function [9] with unknown slopes and nodes that are estimated. This essentially reduces the Wiener identification problem into a parameter estimation problem that is much easier.

A fundamental question in Wiener system identification that has eluded researchers for so long is what is the least amount of a priori information on the unknown system so that identification is possible in the non-Gaussian input case? This is a very hard problem and can be addressed in many different ways. One direction is to ask what constitutes the least amount of a priori information on the nonlinearity so that the linear part of the system can be uniquely identified without the Gaussian assumption. Even this well formulated question is not easy because it is not possible to quantify what "the least amount of a priori information" is. Therefore, efforts have been devoted to address this question in an indirect way, i.e., to develop identification algorithms using as little a priori information as possible. To this end, some work has been reported in the literature [12] recently that a monotone assumption on the unknown nonlinearity ensures a solution for an FIR linear system, though not necessarily unique. This was the first time an identification algorithm was developed in a non-Gaussian case by only assuming monotonicity of a known possibly non-parametric nonlinearity.

The work reported in this paper follows along the same line as [12] by asking whether the monotonicness assumption is necessary. If unnecessary, what constitutes a less restricted condition so that the linear part of the system can be uniquely identified, at least in theory. Further, if identifiable, construct a numerical identification algorithm that calculates an estimates of the linear part. The paper provides answers to above two questions. In particular, we show in the paper that very little a priori information is needed on the nonlinearity so that the FIR linear part can be identified in the non-Gaussian case. The corresponding identification algorithms are also developed.

The layout of the paper is as follows. The system and problems are introduced in Section 2. Section 3 is devoted to a priori information in terms of quadrant knowledge of the unknown nonlinearity. It is shown that with quadrant information, the linear part can be uniquely identified and further an efficient global numerical algorithm is presented.
Section 4 discusses the case that the unknown nonlinearity passes through the origin. With this little a priori information, it is shown that the linear part can be again uniquely identified and the corresponding numerical algorithm is developed along with its convergence proof. Then in Section 5, the results are extended to a priori information in terms of local monotonousness. Similar results are established. Finally, some concluding remarks are provided in Section 6. Because of the page limit, all proofs are not included and interested readers can email the authors for the pre-prints.

II. THE PROBLEM STATEMENT

The Wiener system we consider in this paper is shown in Figure 1 and is described by

\[ x(k) = (u(k), u(k - 1), ..., u(k - n + 1))h \]
\[ y(k) = f(x(k)), \quad \phi_T(k) \quad k = 1, 2, ..., N \]

where \( u(k), x(k) \) and \( y(k) \) are the input, the internal and the output variables respectively. \( h \in \mathbb{R}^n \) is the parameter vector to be identified with \( \|h\| = 1 \) and the first non-zero element being positive. This condition is necessary and standard in Wiener system identification because of scaling ambiguity in the product term. The nonlinearity \( f(\cdot) \) and the internal variable \( x(k) \)'s are unknown. Throughout of the paper, we assume that the variables \( x(k) \)'s and \( y(k) \)'s are bounded,

\[ |x(k)|, |y(k)| \leq M < \infty \]  

(II.2)

for all \( k \) and the input \( u(k) \)'s are iid (not necessarily Gaussian).

We say that the linear system or the parameter vector \( h \) is identifiable from the input-output data set \( \{\phi(k), y(k)\}_{k=1}^{N} \) if \( h \) can be uniquely determined from the system model (II.1) and the data set \( \{\phi(k), y(k)\}_{k=1}^{N} \) independent of the unknown nonlinearity \( f(\cdot) \), or equivalently, there does not exist a different pair \( (h', f') \) that would produce the identical input-output data set \( \{\phi(k), y(k)\}_{k=1}^{N} \). The question we want to address is what is a priori information on the unknown \( f(\cdot) \) so that the parameter vector \( h \) is identifiable and if identifiable, how to calculate it? Further, keep in mind, our goal is to find the least amount of a priori information on \( f(\cdot) \) so that \( h \) is identifiable.

Clearly, all we have is

\[
\begin{pmatrix}
x(1) \\
x(2) \\
\vdots \\
x(N)
\end{pmatrix} =
\begin{pmatrix}
\phi^T(1) \\
\phi^T(2) \\
\vdots \\
\phi^T(N)
\end{pmatrix} h \quad \text{and} \quad
\begin{pmatrix}
y(1) \\
y(2) \\
\vdots \\
y(N)
\end{pmatrix} =
\begin{pmatrix}
f(x(1)) \\
f(x(2)) \\
\vdots \\
f(x(N))
\end{pmatrix}
\]

with unknown \( x(k) \)'s and \( f \).

III. QUADRANT INFORMATION

In this section, we discuss a case in which a priori information on \( f(\cdot) \) is the quadrant information.

Assumption III.1: The unknown nonlinearity is strictly in the first and third quadrants. In other works,

\[ \text{sign}(x(k)) = \text{sign}(y(k)), \quad k = 1, ..., N \]

The case that the nonlinearity lies in the second and fourth quadrants can be similarly discussed. With Assumption (III.1), the nonlinearity can be non-smooth, non-monotonomous and no information on the values of \( x(k) \) is available. However, it is guaranteed that \( \text{sign}(y(k)) = \text{sign}(x(k)) = \text{sign}(\phi_T(k)h) \). Obviously identification of \( h \) has to rely on the following sign equations

\[
\begin{align*}
sign(y(1)) &= sign(x(1)) = sign(\phi^T(1)h) \\
&\vdots \\
sign(y(N)) &= sign(x(N)) = sign(\phi^T(N)h).
\end{align*}
\]

The first question is obviously the identifiability, i.e., if the above equations based only on the signs are sufficient to uniquely determine the unknown \( h \). The following result shows a positive answer.

Theorem III.1: Consider the system (II.1) under Assumption (III.1). Assume that \( u(\cdot) \) is iid and the distribution is continuous containing an interval \([-a, a]\) for some \( 0 < a \). Then, for any unit vector \( h \neq h \), with probability one as \( N \to \infty \), there exists some \( 1 \leq k \leq N \) so that

\[ \text{sign}(y(k)) = sign(x(k)) = sign(\phi^T(k)h) \neq sign(\phi^T(k)\hat{h}) \]

where \( \hat{h} \) is a local minimum at \( h \) for the minimization problem in (II.1).

Identification algorithm under Assumption (III.1):

Consider the system (II.1) under Assumption (III.1).

Step 1: Collect data \( \{y(k)\}'s \) and \( \{y(k)\}'s \).

Step 2: Solve the following minimization problem to find the estimate \( \hat{h} \),

\[ \hat{h} = \arg \min_{\|h\|=1} \frac{1}{N} \sum_{k=1}^{N} (sign(y(k)) - sign(\phi^T(k)\hat{h}))^2 \]

\[ = \arg \min_{\|h\|=1} \frac{1}{N} \sum_{k=1}^{N} (sign(\phi^T(k)h) - sign(\phi^T(k)\hat{h}))^2 \]  

(III.1)

It is clear that there is only one global minimum at \( \hat{h} = h \) for the minimization problem (III.1) if \( N \) is large from the above theorem. It is not clear however if there is any local minimum that would make numerical calculations difficult. To this end, we make two observations. First, suppose the input \( u(\cdot) \) is such distributed that the probability distribution of the direction of \( \phi(k) \) is uniform. Then, the ratio of the number of the regressors \( \phi(k) \)'s that lie in the sector spanned by the angle \( \theta \) as shown in Figure 3 and the total number of the regressor \( N \) converges to, in probability as \( N \to \infty \),

\[ \frac{\phi}{2\pi} \cdot \frac{\text{area of the unit sphere}}{\text{area of the unit sphere}} = \frac{\theta}{2\pi}. \]

Secondly, let \( \hat{h} \neq h \) be any unit vector and \( \theta = \angle(h, \hat{h}), \) \(-\pi \leq \theta < \pi\), be the angle between \( h \) and \( \hat{h} \). Let \( S\theta(h, \hat{h}) = \{ah + \beta\hat{h}, \alpha, \beta \in \mathbb{R} \} \) be the two dimensional plane spanned by \( h \) and \( \hat{h} \). Let \( r(\eta), -\pi \leq \eta < \pi \), denote the unit circle
on the plane $Sp(h, \hat{h})$ such that $r(0) = h$ and $r(\theta) = \hat{h}$. Along the trajectory $r(\eta)$, $-\pi \leq \eta < \pi$, the unit sphere is divided into 4 sectors as illustrated in Figure 4. Since $\phi^T(k) h = \|\phi(k)\| \|h\| \cos(\angle(\phi(k), h))$, it follows that

$\phi(k) \in \text{sector } 1 \implies \text{sign}(\phi^T(k) h) = \text{sign}(\phi^T(k) \hat{h}) = 0$
$\phi(k) \in \text{sector } 2 \implies \text{sign}(\phi^T(k) h) = -\text{sign}(\phi^T(k) \hat{h}) = -2$
$\phi(k) \in \text{sector } 3 \implies \text{sign}(\phi^T(k) h) = \text{sign}(\phi^T(k) \hat{h}) = 0$
$\phi(k) \in \text{sector } 4 \implies \text{sign}(\phi^T(k) h) = \text{sign}(\phi^T(k) \hat{h}) = 2$

From the first observation, the ratio of $\phi(k)$’s in sectors 2 and 4 and the total number of $\phi(k)$’s converges to $2 \cdot \theta/2\pi = |\theta/2\pi|$ if the direction of $\phi(k)$ is uniformly distributed. The idea can be easily extended to a general distribution.

**Theorem III.2:** Consider the system (II.1) under Assumption (III.1). Assume that $u(\cdot)$ is iid and the distribution is continuous containing an interval $[-a, a]$ for some $0 < a$. Then, in probability, for any unit vector $\hat{h}$ with $\theta = \angle(h, \hat{h})$, $-\pi \leq \theta < \pi$,

$$
\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} (\text{sign}(\phi^T(k) h) - \text{sign}(\phi^T(k) \hat{h}))^2 = 4F(\theta)
$$

(III.2)

where $F(\theta) \geq 0$ is the probability that $\phi(k)$ lies in the sectors 2 and 4. Further

$$
F(\theta) = 0 \iff \theta = 0, \ F(\pi) = F(-\pi) = 1
$$

and $F(\theta)$ is strictly decreasing for $-\pi \leq \theta \leq 0$ and strictly increasing for $0 \leq \theta \leq \pi$.

The impact of the theorem is that for a large enough $N$, there is only one (local and global) minimum for the optimization (III.1) achieved at $\theta = 0$. Further, the estimation error is monotonous in terms of the distance “$\theta$” between the true but unknown $h$ and the estimate $\hat{h}$. Simply put, many existing minimization algorithms that calculate gradients numerically apply here.

We now provide a numerical simulation. Let the unknown system be

$$
x(k) = (0.8901, -0.3368, 0.3071) \phi(k), \quad k = 1, \ldots, 3000
$$

$$
y(k) = \sin(0.5 \times x(k)) + 0.3 \times x(k)^3 + v(k);
$$

(III.3)

The input $u(k)$ is iid uniformly in $[-1, 1]$, the noise $v(k)$ is Gaussian and the initial estimate is randomly generated. Table 1 shows the estimates and corresponding errors for various noise levels. The algorithm used is fminsearch, a simplex based method built in the MATLAB. The error is defined as $\|\hat{h} - h\|/\|h\| = \|\hat{h} - h\|$. The results shown in Table 1 are the averages of 100 Monte Carlo simulations. This demonstrates the effectiveness of the identification algorithm.

It was also simulated for the cases when $N = 3000, 10000, 30000$ and $100000$ respectively with the noise level at $20dB$. The results are in Table 2. Because of the sign function that is robust in terms of rejecting small noises, the estimation errors obviously have less to do with the noise level but more to do the number of data points. Recall that the convergence results are derived under the assumption $N \to \infty$. A large $N$ should lead to a small estimation error as supported by numerical simulations.

**IV. Point prior information**

In this section, we consider identification of Wiener systems with a point prior information $f(x_0) = y_0$. For simplicity, the point is assumed to be at the origin.

**Assumption IV.1:** In the set $|x(k)|, |y(k)| \leq M$,

$$
f(x) = 0 \iff x = 0
$$

<table>
<thead>
<tr>
<th>SNR</th>
<th>5dB</th>
<th>10dB</th>
<th>20dB</th>
<th>40dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>error</td>
<td>.0497</td>
<td>.0362</td>
<td>.0514</td>
<td>.0450</td>
</tr>
</tbody>
</table>

TABLE I

**Estimation error vs noise level, quadrant information.**

<table>
<thead>
<tr>
<th>$N$</th>
<th>3,000</th>
<th>10,000</th>
<th>30,000</th>
<th>100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>error</td>
<td>.0514</td>
<td>.0406</td>
<td>.0052</td>
<td>.0020</td>
</tr>
</tbody>
</table>

TABLE II

**Estimation error vs the data points $N$.**
and $f(\cdot)$ is continuous in the neighborhood of the origin. $x(k)$'s are unknown. Only $x(k) = 0$ is known if $y(k) = 0$. Hence, non-zero $y(k)$'s do not reveal any information on the unknown $x(k)$'s and are not very useful in terms of identification. Identification has to rely on $y(k) = 0$ or $y(k)$'s that are close to zero.

**Lemma IV.1:** Consider the system (II.1) under Assumption (IV.1). Assume that the probability density function of $u(\cdot)$ has a positive support containing a non-empty interval $[-a, a]$ for some $0 < a < 1$. Then, $h$ is identifiable for a given input-output data set \{$(\phi(k), y(k))$\} if and only if there exist some $1 \leq p_1 < p_2 < \ldots < p_k \leq N$ so that $x(p_1) = x(p_2) = \ldots = x(p_k) = 0$ (or equivalently $y(p_1) = \ldots = y(p_k) = 0$) and the corresponding matrix $\Phi(p_1, p_2, \ldots, p_k)$ satisfies

$$\text{rank } \begin{pmatrix} \phi^T(p_1) \\ \phi^T(p_2) \\ \vdots \\ \phi^T(p_k) \end{pmatrix} = n - 1.$$

From the above results, it is obvious that to identify $h$, the data has to be collected when $y(k) = 0$. Exact values of $y(k) = 0$ are unlikely in practice. The hope is that the values $y(k) \approx 0$ will be good enough if a small identification error is allowed. In the following, we will show this is indeed the case. To this end, we need some preliminary works.

Let $e_1$, $e_2$, ..., $e_{n-1}$ and $e_n = h$ be an orthonormal basis of $\mathbb{R}^n$. Recall $\|h\| = 1$. Given any small $\epsilon_1 > 0$, construct a truncated cone $C_i$ around $e_i$, $i = 1, 2, \ldots, n$. For $i = 1, 2, \ldots, n - 1$,

$$\phi \in C_i(\epsilon_1) \iff \cos(\angle(\phi, e_i)) \geq 1 - \epsilon_1 \text{ and } a \leq \|\phi\| \leq 1$$

where $\angle(\phi, e_i)$ is the angle between $\phi$ and $e_i$ and $a$ is defined in (IV.1). For $i = n$,

$$\phi \in C_n(\epsilon_1) \iff \cos(\angle(\phi, e_n)) \geq 1 - \epsilon_1 \text{ and } 0 < \|\phi\| \leq \epsilon.$$

![Fig. 4. Truncated cones.](image)

See Figure 2 for a three dimensional case. The following lemma is straightforward.

**Lemma IV.2:** Under conditions of Lemma (IV.1), for any $\epsilon > 0$ and $0 < a < 1$, there exists a small $\epsilon(\epsilon, a) > 0$ so that for all $\epsilon_1 \leq \epsilon$, we have

- $|\cos(\angle(\phi, e))| \leq \epsilon$, $\phi \in C_1$, $\psi \in C_j$ and $|\phi^T\psi| = \|\phi\| \cdot \|\psi\| |\cos(\angle(\phi, \psi))| \leq \epsilon.$
- $\text{rank } \begin{pmatrix} \phi^T(i_1) \\ \phi^T(i_2) \\ \vdots \\ \phi^T(i_n) \end{pmatrix} = n$ if $\phi(i_1) \in C_1$, $\phi(i_2) \in C_2$, ..., $\phi(i_n) \in C_n$.
- Let $\phi(i) \in C_i$, $i = 1, 2, \ldots, n - 1$ and write

$$\phi(i) = \beta_{1,i}e_1 + \beta_{2,i}e_2 + \ldots + \beta_{n-1,i}e_{n-1} + \beta_{n,i}h$$

in terms of the basis vectors $e_1, \ldots, e_{n-1}, e_n = h$. Then, $|\beta_{l,i}| > \sum_{j=1, j \neq i}^{n-1} |\beta_{l,j}|$, $i = 1, 2, \ldots, n - 1$ for sufficient small $\epsilon$.

We are now in a position to state the main results of this section.

**Theorem IV.1:** Consider the system (II.1) under Assumption (IV.1). Assume that the probability density function of $u(\cdot)$ has a positive support containing a non-empty interval $[-a, a]$ for some $0 < a < 1$. Then, 1. For any small $\epsilon > 0$, with probability one as $N \to \infty$, there exists a sequence $\phi(i_l)$, $l = 1, 2, \ldots, n$ so that $|x(i_l)| = |d_u| \leq \epsilon$ and $\text{rank } \Phi(i_1, i_2, \ldots, i_n) = n$.
2. The matrix $\Phi(i_1, i_2, \ldots, i_n)$ can be written as $\Phi(i_1, i_2, \ldots, i_n) = Q + E(\epsilon)$ for some $Q$ and $E(\epsilon)$, where $\text{rank } Q = n - 1$ independent of $\epsilon$ and $E(\epsilon) \to 0$ as $\epsilon \to 0$. Further, let

$$\Phi(i_1, i_2, \ldots, i_n) = U(\epsilon)\Sigma(\epsilon)(V_1(\epsilon), V_2(\epsilon), \ldots, V_n(\epsilon))^T$$

be the SVD decomposition of $\Phi(i_1, i_2, \ldots, i_n)$. Then, modulus of $\pm$ signs,

$$\|V_n(\epsilon) - h\| \to 0, \text{ as } \epsilon \to 0.$$

Based on the theorem, we introduce the identification algorithm.

**Identification algorithm under Assumption (IV.1):**

Consider the system (II.1) under Assumption (IV.1). For any given small $\epsilon$,

Step 1: Collect data $\phi(k)$'s and $y(k)$'s.
Step 2: Construct a submatrix $\Phi(i_1, i_2, \ldots, i_l)$ of $\Phi(1, 2, \ldots, N)$ by deleting $k$'s row if $|y(k)| > \epsilon$.
Step 3: Calculate SVD decomposition $\Phi(i_1, \ldots, i_l) = U(\epsilon)\Sigma(\epsilon)(V_1, \ldots, V_n)^T$.
Step 4: Define $\hat{h} = \pm V_n$ so that the first non-zero element of $\hat{h}$ is positive.

From the theorem, $\hat{h} \approx h$ if $\epsilon$ is small and further $\hat{h} \to h$ as $\epsilon \to 0$.

We now test the algorithm on the same example as in the previous section. Consider the example (III.3) under identical simulation conditions with the assumption $f(0) = 0$ and $\epsilon = 0.01$. The results shown in Table 3 are the averages of 100 Monte Carlo simulations. This demonstrates the effectiveness of the identification algorithm.
TABLE III
Estimation error vs noise level, \(f(0) = 0\).

<table>
<thead>
<tr>
<th>SNR</th>
<th>5dB</th>
<th>10dB</th>
<th>20dB</th>
<th>40dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>error</td>
<td>.0248</td>
<td>.0083</td>
<td>.0015</td>
<td>.0012</td>
</tr>
</tbody>
</table>

TABLE IV
Estimation error vs noise level, locally monotonous.

<table>
<thead>
<tr>
<th>SNR</th>
<th>5dB</th>
<th>10dB</th>
<th>20dB</th>
<th>40dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>error</td>
<td>.0043</td>
<td>.0019</td>
<td>.00002</td>
<td>.00000</td>
</tr>
</tbody>
</table>

V. LOCALLY MONOTONOUS NONLINEARITIES

In this section, we investigate a class of nonlinearities that are assumed to be monotonous locally. Consider the system (II.1) and assume

Assumption V.1: There exists an interval \(f < \tilde{f} \leq M\) and within the interval \(f(x) \in [\tilde{f}, \tilde{f}]\), \(f(\cdot)\) is continuous and

\[
f(x_1) = f(x_2) \iff x_1 = x_2
\]

Clearly, \(f(\cdot)\) is monotonous if \(y = f(x) \in [\tilde{f}, \tilde{f}]\).

Now, define

\[
\psi(i, j) = \phi(i) - \phi(j), \quad z(i, j) = y(i) - y(j).
\]

If we identify \(\psi\) as \(\phi\) and \(z\) as \(y\), identification of the parameter vector \(h\) under the locally monotonous prior information can be cast into identification under either quadrant or point prior information case. If \(z(i, j) = y(i) - y(j) = f(x(i)) - f(x(j)) = 0\) and \(f(x(i)), f(x(j)) \in [\tilde{f}, \tilde{f}]\), we have \(x(i) - x(j) = 0\). Recall that this is exactly the point information case. Therefore, the following theorem is a straightforward extension of the results derived in the previous section.

Theorem V.1: Consider the system (II.1) under Assumption (V.1). Assume that the input \(u(k)\)'s are iid (not necessarily Gaussian) and the probability density function of \(y = f(x)\) is positive in the interval \([\tilde{f}, \tilde{f}]\). Then, \(h\) is identifiable if and only if there exist two sequences \(1 \leq i_1 < i_2 \ldots < i_k \leq N\), \(1 \leq j_1 < j_2 \ldots < j_k \leq N\) so that \(y(i_1), y(j_1) \in [\tilde{f}, \tilde{f}]\) and \(z(i_1, j_1) = y(i_1) - y(j_1) = 0\), \(l = 1, \ldots, k\) (or equivalently \(x(i_1) - x(j_1) = 0\), ..., \(x(i_k) - x(j_k) = 0\)) and the corresponding matrix \(\Psi(1, 2, \ldots, k)\) satisfies

\[
\text{rank } \begin{pmatrix}
\phi^T(i_1) - \phi^T(j_1) \\
\phi^T(i_2) - \phi^T(j_2) \\
\vdots \\
\phi^T(i_k) - \phi^T(j_k)
\end{pmatrix} = n - 1.
\]

2. For any small \(\varepsilon > 0\), with probability one as \(N \to \infty\), there exist two sequences \(\phi(i_l)\) and \(\phi(j_l)\), \(l = 1, 2, \ldots, n\) so that \(y(i_l), y(j_l) \in [\tilde{f}, \tilde{f}]\), \(|y(i_l) - y(j_l)| \leq \varepsilon\) and

\[
\text{rank } \begin{pmatrix}
\phi^T(i_1) - \phi^T(j_1) \\
\phi^T(i_2) - \phi^T(j_2) \\
\vdots \\
\phi^T(i_k) - \phi^T(j_k)
\end{pmatrix} = n.
\]

3. The matrix \(\Psi(1, 2, \ldots, n)\) can be written as

\[
\Psi(1, 2, \ldots, n) = Q + E(\varepsilon)
\]

for some \(Q\) and \(E(\varepsilon)\), where rank \(Q\) = \(n - 1\) independent of \(\varepsilon\) and \(E(\varepsilon) \to 0\) as \(\varepsilon \to 0\). Further, let

\[
\Psi(1, 2, \ldots, n) = U(\varepsilon)\Sigma(\varepsilon)(V_1(\varepsilon), V_2(\varepsilon), \ldots, V_n(\varepsilon))^T
\]

be the SVD decomposition of \(\Psi(1, 2, \ldots, n)\). Then, modulus of \(\pm\) signs,

\[
||V_n(\varepsilon) - h|| \to 0, \quad \varepsilon \to 0.
\]

Based on the theorem, we can define the identification algorithm when the unknown linearity is locally monotonous.

Identification algorithm 1 under Assumption (V.1):

Consider the system (II.1) under Assumption (V.1). For any given small \(\epsilon\),

Step 1: Collect data \(\phi(k)\)'s and \(y(k)\)'s for those \(y(k) \in [\tilde{f}, \tilde{f}]\).

Step 2. Sort out the collected data in a decreasing order \(y(k_1) \geq y(k_2) \geq \ldots \geq y(k_i)\) and define \(z(i, i + 1) = y(k_i) - y(k_{i+1})\), \(i \in \mathbb{N}\) and

\[
\Psi(1, \ldots, l - 1) = \begin{pmatrix}
\psi^T(1, 2) \\
\psi^T(2, 3) \\
\vdots \\
\psi^T(l - 1, l)
\end{pmatrix}
\]

Step 3: Construct a submatrix \(\Psi(1, 2, \ldots, p)\) of \(\Psi(1, 2, \ldots, l - 1)\) by deleting \(q\)'s row if \(|z(q, q + 1)| > \epsilon\).

Step 4: Calculate SVD decomposition \(\Psi(1, \ldots, p) = U \Sigma(V_1, \ldots, V_n)^T\).

Step 5: Define \(\hat{h} = \pm V_n\) so that the first non-zero element of \(\hat{h}\) is positive.

As before, from the theorem, \(\hat{h} \approx h\) if \(\epsilon\) is small and further \(\hat{h} \to h\) as \(\epsilon \to 0\).

We now test the algorithm on the same example (III.3) as in the previous sections under identical simulation conditions. The only difference is that the unknown nonlinearity is assumed to be monotonous for \(y = f(x) \in [-1.5, 1.5]\).

With \(\epsilon = 0.01\), Table 4 shows the simulation errors for various noise levels. The results are the averages of 100 Monte Carlo simulations. This demonstrates again the effectiveness of the identification algorithm.

This algorithm seems to perform better. One explanation is that this algorithm utilizes the data \(y \in [-1.5, 1.5]\) and the previous one only uses the data \(y\) close to zero. Simply put, more data is allowed for this algorithm than the previous one and therefore, the effect of noises is small.

Similarly, the algorithm developed for the quadrant prior information in Section 3 can be used, with some minor modifications, to find an estimate \(\hat{h}\) under Assumption (V.1).

Identification algorithm 2 under Assumption (V.1):

Consider the system (II.1) under Assumption (V.1).

Step 1: Collect data \(\phi(k_i)\)'s and \(y(k_i)\)'s, \(i = 1, \ldots, l + 1\), for those \(y(k_i) \in [\tilde{f}, \tilde{f}]\).

Step 2. Define \(z(i, i + 1) = y(k_i) - y(k_{i+1})\), \(i \in \mathbb{N}\) and

\[
\phi(k_i) - \phi(k_{i+1})
\]
Step 3: Solve the following minimization problem to find an estimate $\hat{h}$,

$$\hat{h} = \arg\min_{\|\hat{h}\|=1} \frac{1}{l} \sum_{i=1}^{l} (\text{sign}(z(i, i+1)) - \text{sign}(\psi^T(i, i+1)\hat{h}))^2.$$ 

Note that

$$\text{sign}(z(i, i + 1)) = \text{sign}(x(k_i) - x(k_{i+1}))$$

because of the monotonous assumption. Thus, all the convergence results presented in Section 3 can be similarly developed for this algorithm as well. We leave the details as an exercise for the readers.

VI. Concluding remarks

It has been shown that little a priori information is needed to identify the FIR linear part of a Wiener system. In all the cases, the input is not necessarily Gaussian. Also, the unknown nonlinearity is not necessarily invertible and not parametrized in any way.

In the paper, the focus is on the theoretical study of the identifiability towards the least amount a priori information needed to identify the linear part. Noise analysis was not emphasized, though numerical simulations appear to suggest the algorithms perform well in the presence of noises. Analytical study of noise effects will be very interesting. Also, to focus on the least amount of a priori information, only an FIR linear part is considered. Extensions of the results to an IIR linear part will be also interesting. Some extension are trivial and others do not seem trivial.

References