ALTERNATING MINIMIZATION APPROACH FOR IDENTIFICATION OF PIECEWISE CONTINUOUS HAMMERSTEIN SYSTEMS

Hiroki Kuroda, Masao Yamagishi, and Isao Yamada

Dept. of Information and Communications Engineering, Tokyo Institute of Technology
Email:{kuroda,myamagi,isao}@sp.ce.titech.ac.jp

ABSTRACT

Identification of piecewise continuous Hammerstein systems, which consist of the cascade of memoryless piecewise continuous systems followed by linear systems, is an important problem in engineering. In the identification process, major existing approaches approximate exact minimization of the non-convex cost function for the piecewise continuous system, which degrades the identification accuracy in some occasions. In this paper, we propose an alternating minimization approach for identification of the piecewise continuous Hammerstein system by alleviating the difficulty in minimization of the cost function for the piecewise continuous system. We first decompose this minimization into quadratic subproblems by sorting the magnitudes of input signals. Then, based on this decomposition, the proposed method exactly minimizes the cost function by finite comparison of the solutions of the subproblems. Numerical examples show the effectiveness of the proposed method.

Index Terms— System identification, Hammerstein system, discontinuous nonlinearity.

1. INTRODUCTION

Identification of Hammerstein systems which exhibit discontinuous nonlinearity is an important problem in, e.g., automatic control [1–3] and acoustic echo cancellation [4, 5]. More precisely, we consider identification of the Hammerstein system which consists of the cascade of a memoryless nonlinear system followed by a finite impulse response (FIR) system, and the nonlinear part is (possibly) discontinuous w.r.t. the input signal. In many situations, such nonlinearity can be well approximated by a piecewise continuous Hammerstein system, i.e., the nonlinear part is approximated by a piecewise continuous system which possesses unknown discontinuous points [4–11].

Meanwhile, for the globally continuous Hammerstein system, a common identification method is to alternatingly minimize the cost function w.r.t. linear and nonlinear parts [12–14]. However, for the piecewise continuous Hammerstein system, since the cost function for the nonlinear part is nonconvex due to the discontinuous points, major existing approaches approximate the exact minimization by replacing unknown discontinuous points to tentative estimates [6–11]. In some occasions, this approximation causes serious degradation of the identification accuracy. Another approaches [4, 5] use gradient descent to suppress the cost function (note: the use of gradient descent can be found for similar models, e.g., [15, 16]). However, gradient descent would not be suitable for this scenario because the cost function has plateaus w.r.t. the discontinuous points.

In this paper, by alleviating the difficulty in the minimization of the cost function for the nonlinear part, we propose an alternating minimization approach for identification of the piecewise continuous Hammerstein system.

To minimize the cost function for the nonlinear part, we exploit the fact that the cost function is quadratic over regions which can be specified by sorting the input signals of the system. Based on this fact, we reduce minimization of the cost function for the nonlinear part to quadratic subproblems. More precisely, the proposed method first clarifies regions where the cost function is quadratic by sorting input signals, and obtain the quadratic form in each region. Then, since the minimizer of the quadratic function in each region can be computed in the closed form, the global minimizer can be obtained by finite comparison of the solutions.

The proposed method is particularly useful for the situations where the number of discontinuity is few enough because the number of subproblems is roughly $(N+M)^L$ where L is the number of discontinuous points, N is the length of the FIR system, and M is the number of observations used for the identification. Such examples include, e.g., preload and deadzone nonlinearity [6–8] where the number of discontinuity is at-most two.

To show the effectiveness of the proposed method, we present numerical examples on identification of the Hammerstein system with preload and dead-zone nonlinearity. We compare the proposed method with the algorithm presented in [6]. The results show the superior performance of the proposed method.

This work is an extension of our previous paper [17] where the clipping nonlinearity is considered.

Notations: Let $\mathbb{N}^*$ and $\mathbb{R}$ denote the sets of all positive integers and all real numbers, respectively. For matrices or

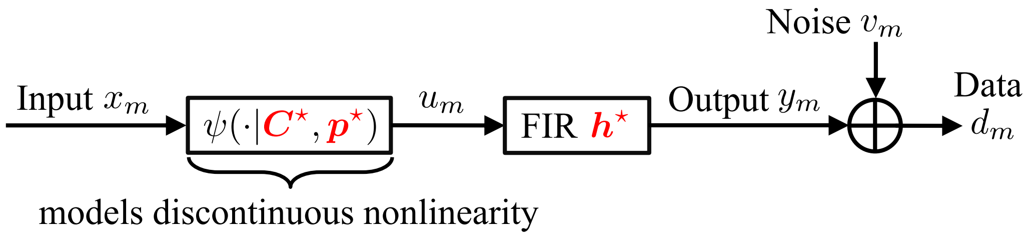

Fig. 1. An illustration of a Hammerstein system with discontinuous nonlinearity where the nonlinear part is approximated by the piecewise continuous system $\psi(\cdot|\boldsymbol{C}^\star,\boldsymbol{p}^\star)$.

vectors, we denote the transpose by $(\cdot)^\top$. For $\boldsymbol{x}\in\mathbb{R}^N$ and $\boldsymbol{X}\in\mathbb{R}^{N\times M}$, $[\boldsymbol{x}]_n$ and $[\boldsymbol{X}]_{n,m}$ respectively denote the n-th component of $\boldsymbol{x}$ and the (n,m)-entry of $\boldsymbol{X}$.

2. PRELIMINARIES

2.1. Problem Formulation

As shown in Fig. 1, we consider identification of the Hammerstein system which consists of a memoryless piecewise continuous system followed by an FIR system. Let $x_m\in\mathbb{R}$ and $y_m\in\mathbb{R}$ respectively denote the input and output signal of the system at time instant $m\in\mathbb{N}^*$. Then, available observation (x_m,d_m) for system identification is defined by

$$\begin{aligned} u_m &:= \psi(x_m|\boldsymbol{C}^\star,\boldsymbol{p}^\star),\\ y_m &:= [u_m,\ldots,u_{m-N+1}]\boldsymbol{h}^\star,\\ d_m &:= y_m+v_m, \end{aligned}$$

where $\boldsymbol{h}^\star\in\mathbb{R}^N$ is the impulse response vector, and $v_m\in\mathbb{R}$ is the observation noise. The piecewise continuous system $\psi(\cdot|\boldsymbol{C}^\star,\boldsymbol{p}^\star)$ is defined by

$$\begin{aligned}&\psi(x|\boldsymbol{C}^\star,\boldsymbol{p}^\star)\\ &:=\begin{cases}\sum_{k=1}^K[\boldsymbol{C}^\star]_{k,1}\varphi_k(x), & (x\in(-\infty,[\boldsymbol{p}^\star]_1)),\\ \quad\vdots & \\ \sum_{k=1}^K[\boldsymbol{C}^\star]_{k,\ell}\varphi_k(x), & (x\in[[\boldsymbol{p}^\star]_{\ell-1},[\boldsymbol{p}^\star]_\ell)),\\ \quad\vdots & \\ \sum_{k=1}^K[\boldsymbol{C}^\star]_{k,L+1}\varphi_k(x), & (x\in[[\boldsymbol{p}^\star]_L,\infty)),\end{cases}\end{aligned} \tag{1}$$

where $\varphi_k\colon\mathbb{R}\to\mathbb{R}$ $(k=1,\ldots,K)$ are known continuous functions, $\boldsymbol{C}^\star\in\mathbb{R}^{K\times(L+1)}$ consists of unknown coefficients in each segment, $\boldsymbol{p}^\star\in\mathbb{R}^L$ consists of unknown (possibly) discontinuous points $[\boldsymbol{p}^\star]_1<\cdots<[\boldsymbol{p}^\star]_L$. A popular model is the piecewise affine function, i.e., $\varphi_k(x)=x^{k-1}$ $(k=1,2)$ [4,6–11].

For simplicity, this paper considers batch estimation of $(\boldsymbol{h}^\star,\boldsymbol{C}^\star,\boldsymbol{p}^\star)$ from known observations $d_M,\ldots,d_1$ and $x_M,\ldots,x_{1-N+1}$, and define the cost function by

$$\begin{aligned}&\Theta(\boldsymbol{h},\boldsymbol{C},\boldsymbol{p})\\ &:=\sum_{m=1}^M(d_m-[\psi(x_m|\boldsymbol{C},\boldsymbol{p}),\ldots,\psi(x_{m-N+1}|\boldsymbol{C},\boldsymbol{p})]\boldsymbol{h})^2.\end{aligned} \tag{2}$$

3. PROPOSED METHOD

We propose to estimate $(\boldsymbol{h}^\star,\boldsymbol{C}^\star,\boldsymbol{p}^\star)$ by the estimation sequence $(\hat{\boldsymbol{h}}_j,\hat{\boldsymbol{C}}_j,\hat{\boldsymbol{p}}_j)_{j\in\mathbb{N}^*}$ generated through alternating minimization of the cost function:

$$\hat{\boldsymbol{h}}_j\in\arg\min_{\boldsymbol{h}\in\mathbb{R}^N}\Theta(\boldsymbol{h},\hat{\boldsymbol{C}}_{j-1},\hat{\boldsymbol{p}}_{j-1}), \tag{3}$$

$$(\hat{\boldsymbol{C}}_j,\hat{\boldsymbol{p}}_j)\in\arg\min_{(\boldsymbol{C},\boldsymbol{p})\in\mathcal{V}}\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p}), \tag{4}$$

from initial guess $(\hat{\boldsymbol{C}}_0,\hat{\boldsymbol{p}}_0)\in\mathcal{V}$ where $\mathcal{V}:=\{(\boldsymbol{C},\boldsymbol{p})\in\mathbb{R}^{K\times(L+1)}\times\mathbb{R}^L|[\boldsymbol{p}]_1<\cdots<[\boldsymbol{p}]_L\}$ is a constraint set to keep the order of the entries of $\boldsymbol{p}$. The first problem (3) is reduced to a system of linear equations because $\Theta(\boldsymbol{h},\hat{\boldsymbol{C}}_{j-1},\hat{\boldsymbol{p}}_{j-1})$ is quadratic w.r.t. $\boldsymbol{h}$. In the next section, we present an algorithm to exactly minimize non-convex $\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})$ in (4) by decomposing this minimization into quadratic subproblems.

3.1. Identification of Piecewise Continuous Systems

We begin by rewriting $\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})$ in (4) from the expression in (2) to

$$\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})=\|\boldsymbol{d}-\hat{\boldsymbol{H}}_j\boldsymbol{\psi}(\boldsymbol{x}|\boldsymbol{C},\boldsymbol{p})\|^2, \tag{5}$$

where $\boldsymbol{d}:=[d_M,\ldots,d_1]\in\mathbb{R}^M$, $\hat{\boldsymbol{H}}_j\in\mathbb{R}^{M\times\bar{N}}$ represents the convolution of $\hat{\boldsymbol{h}}_j$, $\boldsymbol{x}:=[x_M,\ldots,x_{1-N+1}]\in\mathbb{R}^{\bar{N}}$, $\boldsymbol{\psi}(\boldsymbol{x}|\boldsymbol{C},\boldsymbol{p}):=[\psi(x_M|\boldsymbol{C},\boldsymbol{p}),\ldots,\ \psi(x_{1-N+1}|\boldsymbol{C},\boldsymbol{p})]^\top\in\mathbb{R}^{\bar{N}}$, $\bar{N}=N+M-1$, and $\|\cdot\|$ denotes the Euclidian (ℓ_2) norm. More precisely, $\hat{\boldsymbol{H}}_j$ is defined by

$$\hat{\boldsymbol{H}}_j:=\begin{pmatrix}[\hat{\boldsymbol{h}}_j]_1 & \cdots & [\hat{\boldsymbol{h}}_j]_N & 0 & \cdots & 0\\ 0 & [\hat{\boldsymbol{h}}_j]_1 & \cdots & [\hat{\boldsymbol{h}}_j]_N & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & [\hat{\boldsymbol{h}}_j]_1 & \cdots & \cdots & [\hat{\boldsymbol{h}}_j]_N\end{pmatrix}.$$

In the next proposition, we first clarify the regions where $\boldsymbol{\psi}(\boldsymbol{x}|\boldsymbol{C},\boldsymbol{p})$ is affine w.r.t. $(\boldsymbol{C},\boldsymbol{p})$, which implies that $\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})$ is quadratic w.r.t. $(\boldsymbol{C},\boldsymbol{p})$ in the regions. Subsequently, we reduce the minimization of $\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})$ into quadratic subproblems.

Proposition 1 (Decomposition into quadratic subproblems)**.** Sort the entries of $\boldsymbol{x}$ into $\tilde{x}_1,\dots,\tilde{x}_{\bar{N}}$ in ascending order, i.e.,

$$\tilde{x}_i := [\boldsymbol{x}]_{\tau(i)} \ (i=1,\dots,\bar{N})$$

where the bijection $\tau\colon \{1,\dots,\bar{N}\}\to\{1,\dots,\bar{N}\}$ satisfies[1]

$$[\boldsymbol{x}]_{\tau(1)} < \cdots < [\boldsymbol{x}]_{\tau(\bar{N})}.$$

Then, $\boldsymbol{\psi}(\boldsymbol{x}|\boldsymbol{C},\boldsymbol{p})$ is affine for $[\boldsymbol{p}]_\ell \in (\tilde{x}_{n_\ell},\tilde{x}_{n_\ell+1}]$ $(\ell = 1,\dots,L)$ $(1\le n_1<\cdots<n_L\le\bar{N}-1)$ as[2]

$$\boldsymbol{\psi}(\boldsymbol{x}|\boldsymbol{C},\boldsymbol{p}) = \sum_{\ell=1}^{L+1}\sum_{i_\ell=n_{\ell-1}+1}^{n_\ell}\sum_{k=1}^{K}[\boldsymbol{C}]_{k,\ell}\varphi_k(\tilde{x}_{i_\ell})\boldsymbol{e}_{\tau(i_\ell)}, \quad (6)$$

where $\{\boldsymbol{e}_j\}_{j=1}^{\bar{N}}$ is the standard orthonormal basis of $\mathbb{R}^{\bar{N}}$, i.e., j-th entry of $\boldsymbol{e}_j$ is one and the others are zeros. By substituting (6) into (5), for $[\boldsymbol{p}]_\ell \in (\tilde{x}_{n_\ell},\tilde{x}_{n_\ell+1}]$ $(\ell=1,\dots,L)$, we have

$$\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p}) = \left\|\boldsymbol{d} - \sum_{\ell=1}^{L+1}\sum_{k=1}^{K}[\boldsymbol{C}]_{k,\ell}\boldsymbol{a}^{(k)}_{n_{\ell-1},n_\ell}\right\|^2,$$

where

$$\boldsymbol{a}^{(k)}_{n_{\ell-1},n_\ell} := \sum_{i=n_{\ell-1}+1}^{n_\ell}\varphi_k(\tilde{x}_i)[\hat{\boldsymbol{H}}_j]_{:,\tau(i)}, \quad (7)$$

and $[\hat{\boldsymbol{H}}_j]_{:,\tau(i)} \in \mathbb{R}^M$ denotes the $\tau(i)$-th column vector of $\hat{\boldsymbol{H}}_j$. Based on this expression, minimization of $\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})$ can be decomposed as[3]

$$\begin{aligned}&\min_{(\boldsymbol{C},\boldsymbol{p})\in\mathcal{V}}\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})\\ &= \min_{(n_1,\dots,n_L)\in\mathcal{Q}}\ \min_{\boldsymbol{C}\in\mathbb{R}^{K\times(L+1)}}\left\|\boldsymbol{d} - \sum_{\ell=1}^{L+1}\sum_{k=1}^{K}[\boldsymbol{C}]_{k,\ell}\boldsymbol{a}^{(k)}_{n_{\ell-1},n_\ell}\right\|^2,\end{aligned}$$

where $\mathcal{Q} := \{(n_1,\dots,n_L)\in\{1,\dots,\bar{N}-1\}^L \mid n_1<\cdots<n_L\}$.

Based on this decomposition, we can compute the global minimizer by comparing the solutions of the quadratic subproblems. Detailed steps are shown in the following algorithm.

[1] For simplicity, we here assume that entries of $\boldsymbol{x}$ are different. This assumption can be relaxed easily.

[2] With a slight abuse of the notation, we let $n_0 := 0$ and $n_{L+1} = \bar{N}$.

[3] We restrict $[\boldsymbol{p}]_\ell \in (\tilde{x}_1,\tilde{x}_{\bar{N}}]$ $(\ell=1,\dots,L)$ since a minimizer of $\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})$ can be found in this region.

Algorithm 1 (Exact minimization via decomposition into quadratic subproblems)**.**

1. Solve quadratic subproblems

$$\boldsymbol{C}^*_{n_1,\dots,n_L} \in \arg\min_{\boldsymbol{C}\in\mathbb{R}^{K\times(L+1)}}\left\|\boldsymbol{d} - \sum_{\ell=1}^{L+1}\sum_{k=1}^{K}[\boldsymbol{C}]_{k,\ell}\boldsymbol{a}^{(k)}_{n_{\ell-1},n_\ell}\right\|^2$$

for every $(n_1,\dots,n_L)\in\mathcal{Q}$.

2. Compute $(\hat{\boldsymbol{C}}_j,\hat{\boldsymbol{p}}_j) \in \arg\min_{(\boldsymbol{C},\boldsymbol{p})\in\mathcal{V}}\Theta(\hat{\boldsymbol{h}}_j,\boldsymbol{C},\boldsymbol{p})$ by comparing the solutions of the quadratic problems:

$$\begin{aligned}&\hat{\boldsymbol{C}}_j = \boldsymbol{C}^*_{n_1^*,\dots,n_L^*},\\ &[\hat{\boldsymbol{p}}_j]_\ell \in (\tilde{x}_{n_\ell^*},\tilde{x}_{n_\ell^*+1}] \quad (\ell=1,\dots,L),\end{aligned}$$

where

$$(n_1^*,\dots,n_L^*) \in \arg\min_{(n_1,\dots,n_L)\in\mathcal{Q}}\left\|\boldsymbol{d} - \sum_{\ell=1}^{L+1}\sum_{k=1}^{K}[\boldsymbol{C}^*_{n_1,\dots,n_L}]_{k,\ell}\boldsymbol{a}^{(k)}_{n_{\ell-1},n_\ell}\right\|^2.$$

Remark 1 (Extension for symmetric nonlinear models)**.** In some typical situations, the nonlinear parts of Hammerstein systems are odd-symmetric w.r.t. the input signal, e.g., [4, 6, 8]. This type of nonlinearities can be modeled through (1), but more efficiently modeled as

$$\psi_{\mathrm{sym}}(x|\boldsymbol{A}^\star,\boldsymbol{q}^\star) := \begin{cases}\sum_{k=1}^{K}[\boldsymbol{A}^\star]_{k,1}\varphi_k(\mathrm{sgn}(x)), & (|x|\in[0,[\boldsymbol{q}^\star]_1)),\\ \quad\vdots & \\ \sum_{k=1}^{K}[\boldsymbol{A}^\star]_{k,\ell}\varphi_k(\mathrm{sgn}(x)), & (|x|\in[[\boldsymbol{q}^\star]_{\ell-1},[\boldsymbol{q}^\star]_\ell)),\\ \quad\vdots & \\ \sum_{k=1}^{K}[\boldsymbol{A}^\star]_{k,L+1}\varphi_k(\mathrm{sgn}(x)), & (|x|\in[[\boldsymbol{q}^\star]_L,\infty)),\end{cases}$$

where $\mathrm{sgn}(x)$ denotes the sign of x. For $\psi_{\mathrm{sym}}(\cdot|\boldsymbol{A}^\star,\boldsymbol{q}^\star)$, by sorting the absolute values of the entries of $\boldsymbol{x}$, we can develop a more efficient algorithm than Algorithm 1.

Remark 2 (Efficient implementation through recursive computation)**.** In Algorithm 1, calculation of $\boldsymbol{a}^{(k)}_{n,m}$ for every $(n,m)\in\{(m',n')\in\{1,\dots,\bar{N}\}^2 \mid m'<n'\}$ by the definition (7) requires roughly $\mathcal{O}(M\bar{N}^3)$ multiplication. By exploiting the relation

$$\boldsymbol{a}^{(k)}_{n,m} = \boldsymbol{a}^{(k)}_{0,\bar{N}} - \boldsymbol{a}^{(k)}_{0,n-1} - \boldsymbol{a}^{(k)}_{m,\bar{N}},$$

$\boldsymbol{a}^{(k)}_{n,m}$ can be computed through simple addition/subtraction from $\left(\boldsymbol{a}^{(k)}_{0,n}\right)_{n=0}^{\bar{N}}$ and $\left(\boldsymbol{a}^{(k)}_{m,\bar{N}}\right)_{m=0}^{\bar{N}}$. Moreover, by computing

Table 1. Normalized mean square error of $(\boldsymbol{h}^\star, \boldsymbol{C}^\star, \boldsymbol{p}^\star)$ is shown in dB for different impulse response vectors $\boldsymbol{h}^\star$. For the existing method [6], we show the the results for different upper bounds $p_{\rm upb} = 3$ and $p_{\rm upb} = p^\star = 2$.

Situation	Algorithm	$\|\boldsymbol{h}^\star - \hat{\boldsymbol{h}}\|^2/\|\boldsymbol{h}^\star\|^2$	$\|\boldsymbol{C}^\star - \hat{\boldsymbol{C}}\|_{\rm fro}^2/\|\boldsymbol{C}^\star\|_{\rm fro}^2$	$\|\boldsymbol{p}^\star - \hat{\boldsymbol{p}}\|_2^2/\|\boldsymbol{p}^\star\|_2^2$
$(N, M) = (4, 200)$	Proposed (Algorithm 1)	-37.9	-18.7	-35.1
	Existing [6] with $p_{\rm upb} = 3$	-25.5	-2.4	omitted
	Existing [6] with $p_{\rm upb} = p^\star$	-37.9	-18.7	omitted
$(N, M) = (256, 1000)$	Proposed (Algorithm 1)	-24.7	-23.6	-51.5
	Existing [6] with $p_{\rm upb} = 3$	-11.4	-4.5	omitted
	Existing [6] with $p_{\rm upb} = p^\star$	-24.7	-23.6	omitted

$\big(\boldsymbol{a}_{0,n}^{(k)}\big)_{n=0}^{\bar{N}}$ and $\big(\boldsymbol{a}_{m,\bar{N}}^{(k)}\big)_{m=0}^{\bar{N}}$ through recursive calculation

$$\boldsymbol{a}_{0,n}^{(k)} = \boldsymbol{a}_{0,n-1}^{(k)} + \varphi_k(\tilde{x}_n)[\hat{\boldsymbol{H}}_j]_{:,\tau(n)},$$
$$\boldsymbol{a}_{m,\bar{N}}^{(k)} = \boldsymbol{a}_{m+1,\bar{N}}^{(k)} + \varphi_k(\tilde{x}_{m+1})[\hat{\boldsymbol{H}}_j]_{:,\tau(m+1)},$$

we can reduce the order of multiplication to $\mathcal{O}(M\bar{N})$.

Remark 3 (Overall computational costs of Algorithm 1)**.** The number of subproblems in step 1 of Algorithm 1 is roughly $\bar{N}^L$. Thus, for $M > K(L+1)$, computational cost for solving all subproblems is roughly $\mathcal{O}(\bar{N}^L(M^2 + (KL)^3))$, which suggests that Algorithm 1 is practical for small enough L (see Remark 2 for computational costs to obtain $\boldsymbol{a}_{n_{\ell-1},n_\ell}^{(k)}$).

4. NUMERICAL EXAMPLES

To show the effectiveness of the proposed approach, we conduct numerical experiments on identification of the Hammerstein system with symmetric preload and dead-zone nonlinearity defined as

$$\mathrm{pdz}(x|a^\star, b^\star, p^\star) := \begin{cases} 0, & \text{if } |x| \le p_\star, \\ a^\star x + b^\star \mathrm{sgn}(x), & \text{otherwise}, \end{cases} \quad (8)$$

which can be interpreted as $\psi(x|\boldsymbol{C}^\star, \boldsymbol{p}^\star)$ in (1) with $\varphi_k(x) = x^{k-1}$ $(k = 1, 2)$, $[\boldsymbol{p}^\star]_1 = -p^\star$, $[\boldsymbol{p}^\star]_2 = p^\star$, $[\boldsymbol{C}^\star]_{2,1} = [\boldsymbol{C}^\star]_{2,3} = a^\star$, $[\boldsymbol{C}^\star]_{1,1} = -b^\star$, $[\boldsymbol{C}^\star]_{1,3} = b^\star$, and $[\boldsymbol{C}^\star]_{1,2} = [\boldsymbol{C}^\star]_{2,2} = 0$.

According to [6], we set the system to be identified as $a^\star = 6$, $b^\star = 2$, $p^\star = 2$, and $\boldsymbol{h}^\star = [0.7947, 0.2649, -0.5298, -0.1325]^\top \in \mathbb{R}^4$. We also show a numerical example for the impulse response vector of larger size by generating $\boldsymbol{h}^\star \in \mathbb{R}^{256}$ from Gauss distribution $\mathcal{N}(0, 1)$. Note that both examples are normalized as $\|\boldsymbol{h}^\star\| = 1$ and $[\boldsymbol{h}^\star]_1 \ge 0$. We generate the input signal x_m by uniform distribution $[-5, 5]$, and set v_m as white Gaussian noise of 20dB SNR. We use $M = 200$ observations for $\boldsymbol{h}^\star \in \mathbb{R}^4$, and $M = 1000$ for $\boldsymbol{h}^\star \in \mathbb{R}^{256}$.

We compare the proposed approach (3) and (4) with the algorithm presented in [6] which approximately solves (4) by using an upper bound of $p^\star$. For both proposed and existing methods, $\hat{\boldsymbol{h}}_1$ is computed by (3) with identity initialization, i.e., $(\boldsymbol{C}_0, \boldsymbol{p}_0)$ is defined so that $\psi(x|\boldsymbol{C}_0, \boldsymbol{p}_0) = x$. To resolve the scalar ambiguity, we perform the normalization step after (3) so that $\|\hat{\boldsymbol{h}}_j\| = 1$ and $[\hat{\boldsymbol{h}}_j]_1 \ge 0$. Note that we utilize an efficient implementation by exploiting symmetry in $\boldsymbol{C}^\star$ and $\boldsymbol{p}^\star$ for the preload and dead-zone model (8) (note: this efficient implementation is essentially based on Remark 1).

The results are shown in Table 1 where we compare the algorithms in terms of normalized mean square error of $(\boldsymbol{h}^\star, \boldsymbol{C}^\star, \boldsymbol{p}^\star)$ averaged over 100 trials. Note that, as the estimate, we adopt 50 iterations of (3) and (4), i.e., $(\hat{\boldsymbol{h}}, \hat{\boldsymbol{C}}, \hat{\boldsymbol{p}}) = (\hat{\boldsymbol{h}}_{50}, \hat{\boldsymbol{C}}_{50}, \hat{\boldsymbol{p}}_{50})$. The existing method in [6] requires an upper bound $p_{\rm upb} \ge p^\star$, so we show the results for $p_{\rm upb} = 3$ and $p_{\rm upb} = p^\star = 2$. Note that we omit the result on $p^\star$ for the existing method because this method requires another identification process for the estimation of $p^\star$ by using the input-output correlation. It can be seen that the proposed method outperforms the existing method with $p_{\rm upb} = 3$, and achieve the performance same to that with the ideal upper bound $p_{\rm upb} = p^\star$. For impulse response vector of larger size $N = 256$, similar result can be seen with the moderate number of observations $M = 1000$.

5. CONCLUSION

For the identification of a piecewise continuous Hammerstein system, we presented an alternating minimization approach (3) and (4) by alleviating the difficulty in the minimization of $\Theta(\hat{\boldsymbol{h}}_j, \boldsymbol{C}, \boldsymbol{p})$ in (4). We first show that minimization of $\Theta(\hat{\boldsymbol{h}}_j, \boldsymbol{C}, \boldsymbol{p})$ can be reduced to minimization of quadratic subproblems by sorting the magnitudes of input signals. Then, the proposed method computes the global minimizer by finite comparison of the solution of the subproblems. Numerical examples show the effectiveness of the proposed approach. Future work includes development of a more efficient algorithm by exploiting the sparsity in $\boldsymbol{u} := \boldsymbol{\psi}(\boldsymbol{x}|\boldsymbol{C}, \boldsymbol{p})$ based on the techniques in [18].

Acknowledgment

This work was supported in part by JSPS Grants-in-Aid Grant Numbers 29-6921, 26730128 and B-15H02752.

6. REFERENCES

[1] G. Tao and C. A. C. d. Wit, "Special issue on adaptive systems with non-smooth non-linearities," *International Journal of Adaptive Control and Signal Processing*, vol. 11, no. 1, pp. 1–81, 1997.

[2] G. Tao and F. L. Lewis, *Adaptive Control of Nonsmooth Dynamic Systems*, Springer-Verlag, 2001.

[3] F. Giri and E.-W. Bai, *Block-Oriented Nonlinear System Identification*, vol. 1, Springer-Verlag, 2010.

[4] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Process.*, vol. 80, no. 9, pp. 1747–1760, 2000.

[5] M. I. Mossi, C. Yemdji, N. Evans, C. Beaugeant, F. M. Plante, and F. Marfouq, "Dual amplifier and loudspeaker compensation using fast convergent and cascaded approaches to nonlinear acoustic echo cancellation," in *Proc. IEEE ICASSP*, 2012, pp. 41–44.

[6] Y. Liu and E.-W. Bai, "Iterative identification of Hammerstein systems," *Automatica*, vol. 43, no. 2, pp. 346–354, 2007.

[7] J. Vörös, "Parameter identification of discontinuous Hammerstein systems," *Automatica*, vol. 33, no. 6, pp. 1141–1146, 1997.

[8] J. Vörös, "Recursive identification of Hammerstein systems with discontinuous nonlinearities containing dead-zones," *IEEE Trans. Automat. Ctrl.*, vol. 48, no. 12, pp. 2203–2206, Dec. 2003.

[9] D. Wang, F. Ding, and Y. Chu, "Data filtering based recursive least squares algorithm for Hammerstein systems using the key-term separation principle," *Information Sciences*, vol. 222, pp. 203–212, 2013.

[10] J. Chen, X. Wang, and R. Ding, "Gradient based estimation algorithm for Hammerstein systems with saturation and dead-zone nonlinearities," *Applied Mathematical Modelling*, vol. 36, no. 1, pp. 238–243, 2012.

[11] F. Wang, K. Xing, and X. Xu, "Parameter estimation of piecewise Hammerstein systems," *Transactions of the Institute of Measurement and Control*, vol. 36, no. 8, pp. 1024–1032, 2014.

[12] E.-W. Bai and D. Li, "Convergence of the iterative Hammerstein system identification algorithm," *IEEE Trans. Automat. Ctrl.*, vol. 49, no. 11, pp. 1929–1940, Nov. 2004.

[13] S. Rangan, G. Wolodkin, and K. Poolla, "New results for Hammerstein system identification," in *Proc. IEEE Conference on Decision and Control*, Dec. 1995, vol. 1, pp. 697–702.

[14] K. Narendra and P. Gallman, "An iterative method for the identification of nonlinear systems using a Hammerstein model," *IEEE Trans. Automat. Ctrl.*, vol. 11, no. 3, pp. 546–550, July 1966.

[15] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Nonlinear spline adaptive filtering," *Signal Processing*, vol. 93, no. 4, pp. 772–783, 2013.

[16] N. D. Vanli and S. S. Kozat, "A comprehensive approach to universal piecewise nonlinear regression based on trees," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5471–5486, Oct. 2014.

[17] H. Kuroda, M. Yamagishi, and I. Yamada, "Nonlinear acoustic echo cancellation by exact-online adaptive alternating minimization," *IEICE Trans. Fundamentals*, vol. E99-A, no. 10, pp. 2027–2036, Oct. 2016.

[18] H. Kuroda, M. Yamagishi, and I. Yamada, "Recovery of piecewise finite-dimensional continuous signals by exploiting sparsity," in *International Conference on Sampling Theory and Applications*, July 2017, pp. 609–613.