Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Nonparametric models for Hammerstein-Wiener and Wiener-Hammerstein system identification

Riccardo S. Risuleo [*,**] **Håkan Hjalmarsson** [*]

[*] *ACCESS Linnaeus Centre, School of Electrical Engineering, KTH – Royal Institute of Technology, Stockholm, Sweden (e-mail: {risuleo; hjalmars}@kth.se)*
[**] *Department of Information Technology, Uppsala University, Uppsala, Sweden (e-mail: riccardo.risuleo@it.uu.se)*

Abstract: We propose a framework for modeling structured nonlinear systems using nonparametric Gaussian processes. In particular, we introduce a two-layer stochastic model of latent interconnected Gaussian processes suitable for modeling Hammerstein-Wiener and Wiener-Hammerstein cascades. The posterior distribution of the latent processes is intractable because of the nonlinear interactions in the model; hence, we propose a Markov Chain Monte Carlo method consisting of a Gibbs sampler where each step is implemented using elliptical-slice sampling. We present the results on two example nonlinear systems showing that they can effectively be modeled and identified using the proposed nonparametric modeling approach.

Keywords: Nonlinear system identification, Bayesian methods, Nonparametric methods

1. INTRODUCTION

In the identification of nonlinear dynamical systems, block-oriented approaches are often used for their ability to capture complex nonlinear relationships without sacrificing mathematical tractability (Giri and Bai, 2010).

Among the block-oriented models, three-block cascades offer great modelig flexibility and have been used with success in many applications.

For instance, the Wiener-Hammerstein cascade consists of a static nonlinear function sandwiched between two linear dynamical systems that has been used in the modeling of skeletal muscles (Bai et al., 2009; Dewhirst et al., 2010), power electronics (Oliver et al., 2009), heat exchangers and superheaters (Haryanto and Hong, 2013), and in model-predictive control applications (Ławryńczuk, 2016), among others.

The methods for Wiener-Hammerstein identification available in the literature follow four main directions: iterative nonlinear optimization methods (Marconato et al., 2012; Paduart et al., 2012; Tan et al., 2012), stochastic methods (Bershad et al., 2001; Pillonetto and Chiuso, 2009), frequency-domain methods (Westwick and Schoukens, 2012; Schoukens and Tiels, 2017), and two-stage methods (Vanbeylen, 2014; Schoukens et al., 2014; Giordano et al., 2018)

The Hammerstein-Wiener cascade, on the contrary, is a three-block cascade consisiting of a linear dynamical system sandwiched between two static nonlinear functions. These structures have been used to model radio frequency transmitters and power amplifiers (Taringou et al., 2010), the magnetosphere and ionosphere (Palanthandalam-Madapusi et al., 2005), the decomposition of stored crops (Nadimi et al., 2012), turbofan engines (Wang et al., 2017), and geothermal borefields (Atam et al., 2018). To estimate Hammerstein-Wiener models, various techniques have been proposed (see, for instance, Zhu, 2002; Hasiewicz and Mzyk, 2004; Ni et al., 2013).

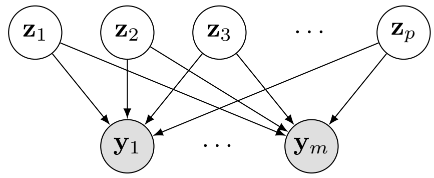

Fig. 1. Bayesian network of the two-layer stochastic model. Empty nodes indicate latent variables; shaded nodes indicate observed variables; edges indicate conditional dependencies.

Differently from the previously mentioned works, in this paper we formulate a unified nonparametric framework for modeling of three-block cascades. Modeling the impulse responses of the linear blocks and the characteristics of the nonlinear blocks using appropriate Gaussian-process models, we show that the Hammerstein-Wiener and Wiener-Hammerstein cascades can both be formulated as two-layer stochastic models such as the one presented in Fig. 1. The model consists of a set of a-priori independent latent variables obeserved through some nonlinear likelihood functions. We propose a sampling method to approximate the posterior distribution of the latent variables using a Markov-chain Monte Carlo algorithm. The method consists of a Gibbs sampler where each step is implemented using an elliptical-slice sampler targeting the full conditional distribution of one of the latent variables.

⋆ This work was supported by the Swedish Research Council under the project *NewLEADS – New directions in learning dynamical systems* (contract 2016-06079). Corresponding author R. S. Risuleo.

The rest of the paper is organized as follows: in Sec. 2, we propose the general two-layer stochastic model; in Sec. 3, we propose the sampling approach targeting the posterior distribution of the latent variables of the model; in Sec. 4, we show applications to Hammerstein-Wiener and Wiener-Hammerstein cascades with general noninvertible nonlinearities.

2. TWO-LAYER STOCHASTIC MODEL

We consider a stochastic model with p independent latent variables $\mathbf{z}_1, \ldots, \mathbf{z}_p$. Each latent variable $\mathbf{z}_i$ represents an unknown quantity in the model and is the realization of a zero-mean Gaussian process observed at some known input locations $x_1^i, \ldots, x_n^i$ (which may represent external excitation signals, reference values, or time, among others). Hence, from the Gaussian-process model, each latent variable has a multivariate normal distribution,

$$\mathbf{z}_i \sim \mathcal{N}(0, K_{\mathbf{z}_i}), \tag{1}$$

where the covariance matrix $K_{\mathbf{z}_i}$ is determined by the covariance function (the *kernel*) of the Gaussian process,

$$\big[K_{\mathbf{z}_i}\big]_{j,k} = K_i(x_j^i,\, x_k^i;\, \zeta_i),$$

where ζ_i is a set of prior hyperparameters that determine the shape of the kernel. In this work, we assume that the hyperparameters are given, or determined with some external procedure (e.g., cross validation or marginal likelihood maximization). Approaches to efficiently estimate these hyperparameters from data are still object of research.

We suppose that we have available m independent vectors of observations $\mathbf{y}_1, \ldots, \mathbf{y}_m$, that depend on the latent variables according to some joint likelihood function

$$\mathrm{p}(\mathcal{Y} \mid \mathcal{Z}; \lambda) = \mathrm{p}(\mathbf{y}_i \mid \mathcal{Z}; \lambda_i) \cdots \mathrm{p}(\mathbf{y}_m \mid \mathcal{Z}; \lambda_m), \tag{2}$$

where $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$, $\mathcal{Z} = \big\{\mathbf{z}_1, \ldots, \mathbf{z}_p\big\}$, and where $\lambda = \{\lambda_1, \ldots, \lambda_m\}$ are hyperparameters describing the observation models. The Bayesian network of the two-layer stochastic model is presented in Fig. 1.

In Section 4, we show how two classical structures from block-oriented system identification, the Wiener-Hammerstein and the Hammerstein-Wiener cascades, can be modeled using such two-layer stocastic models.

The general estimation problem we consider is then to estimate the posterior distribution,

$$\mathrm{p}(\mathcal{Z} \mid \mathcal{Y}), \tag{3}$$

of the latent Gaussian processes given the observations $\mathcal{Y}$. Note that, in the following, we will drop explicit dependencies on the kernel and the likelihood hyperparameters λ and ζ for notational convenience.

3. MONTE CARLO APPROXIMATE INFERENCE

In the general case, the Gaussian priors of the latent variables (1) and the likelihood function (2) are not conjugate, so the posterior distribution (3) is intractable. Therefore, we make a Monte Carlo approximation of the posterior according to

$$\mathrm{p}(\mathcal{Z} \mid \mathcal{Y}) \approx \frac{1}{M} \sum_{m=1}^{M} \delta\big(\mathbf{z}_1 - \bar{\mathbf{z}}_1^{(m)}\big) \cdots \delta\big(\mathbf{z}_p - \bar{\mathbf{z}}_p^{(m)}\big)\,, \tag{4}$$

where the particles $\big\{\bar{\mathbf{z}}_i^{(m)}\big\}_{m=1}^{M}$ are drawn using a Markov chain sampler targeting the posterior distribution (3).

To create the Markov chain, we use a Gibbs sampling procedure: we start from an initialization $\mathbf{z}_1^{(0)}, \ldots, \mathbf{z}_p^{(0)}$, and we iteratively sample each latent variable conditioned on the data and all the remaining latent variables,

$$\begin{aligned}
\mathbf{z}_1^{(k+1)} &\sim \mathrm{p}\big(\mathbf{z}_1 \mid \mathcal{Y}, \mathbf{z}_2^{(k)}, \ldots, \mathbf{z}_p^{(k)}\big)\,,\\
\mathbf{z}_2^{(k+1)} &\sim \mathrm{p}\big(\mathbf{z}_2 \mid \mathcal{Y}, \mathbf{z}_1^{(k+1)}, \mathbf{z}_3^{(k)}, \ldots, \mathbf{z}_p^{(k)}\big)\,,\\
&\;\vdots\\
\mathbf{z}_{p-1}^{(k+1)} &\sim \mathrm{p}\big(\mathbf{z}_{p-1} \mid \mathcal{Y}, \mathbf{z}_1^{(k+1)}, \ldots, \mathbf{z}_{p-2}^{(k+1)}, \mathbf{z}_p^{(k)}\big)\\
\mathbf{z}_p^{(k+1)} &\sim \mathrm{p}\big(\mathbf{z}_p \mid \mathcal{Y}, \mathbf{z}_1^{(k+1)}, \ldots, \mathbf{z}_{p-1}^{(k+1)}\big)
\end{aligned} \tag{5}$$

Note that, from Fig. 1, there may be conditional independence properties that reduce the number of conditioning variables in (5). In particular, $\mathbf{z}_i$ is conditionally independent of all the measurements $\mathbf{y}_j$ whose likelihood function (2) does not contain $\mathbf{z}_i$.

Iteratively running (5) for a large number of iterations and discarding the initial samples, we obtain a sequence of values that can be used in (4) to approximate the posterior distribution (3) and compute point estimates of interest and (Bayesian) credible intervals.

Using the Gibbs sampler, we can sample the joint posterior distribution of $\mathcal{Z}$ by sampling each variable $\mathbf{z}_i$ in turns. To sample $\mathbf{z}_i$, we notice that we are drawing from a conditional density that is proportional to the product of a nonlinear likelihood function and a Gaussian prior distribution,

$$\mathrm{p}\big(\mathbf{z}_i \mid \mathcal{Y}, \mathcal{Z}_{\setminus i}\big) \propto \mathrm{p}\big(\mathcal{Y} \mid \mathcal{Z}_{\setminus i} \cup \{\mathbf{z}_i\}\big)\, \mathrm{p}(\mathbf{z}_i)$$

where $\mathcal{Z}_{\setminus i} = \big\{\mathbf{z}_j,\ j \neq i\big\}$. hence, to sample each step, we can use elliptical slice sampling (ESS).

As the name suggests, ESS is a modification of the standard slice sampler (Neal, 2003) specialized for drawing samples from a target distribution that is proportional to the product of a Gaussian distribution and a nonlinear likelihood function (Murray et al., 2010). To sample the latent variable, ESS starts from a sample $\mathbf{z}_i^{(k)}$ and draws a point ν randomly from the prior $\mathrm{p}(\mathbf{z}_i)$. Then, samples $\mathbf{z}_i'$ are proposed along an ellipse passing through $\mathbf{z}_i^{(k)}$ and ν with an adaptively decreasing step size until the proposal has a likelihood that exceeds a threshold L determined by the starting point $\mathbf{z}_i^{(k)}$, in which case it is accepted as the next sample in the chain. The ESS procedure is presented in detail in Alg. 1 (adapted from Murray et al., 2010).

The whole Gibbs sampling algorithm targeting the joint posterior (3) is presented in Alg. 2.

4. APPLICATION TO THREE-BLOCK CASCADES

In this section, we present two examples of block-oriented nonlinear systems that can be modeled as two-layer stochastic models. To this end, we first introduce nonparametric Gaussian-process models on the unknowns; then, we manipulate the models to put them in the structure described by Fig. 1, in order to run Alg. 2 and approximate the posterior distribution of the latent variables.

Algorithm 1 Elliptical slice sampling updating the ith latent variable of a two-layer stochastic model.

procedure ESS(i, $\mathcal{Z}$)
  $\nu \sim \mathrm{p}(\mathbf{z}_i), \quad u \sim \mathcal{U}[0,1]$ ▷ $\mathcal{U}$ is uniform
  $L \leftarrow \log \mathrm{p}(\mathcal{Y} \mid \mathcal{Z}) + \log u$ ▷ Likelihood threshold
  $\tau \sim \mathcal{U}[0, 2\pi]$ ▷ Step-size parameter
  $\tau_{\min} \leftarrow \tau - 2\pi, \quad \tau_{\max} \leftarrow \tau$
  $\mathbf{z}'_i \leftarrow \mathbf{z}_i \cos\tau + \nu \sin\tau$ ▷ Proposal on the ellipse
  if $\log \mathrm{p}(\mathcal{Y} \mid \mathcal{Z}_{\setminus i} \cup \{\mathbf{z}'_i\}) < L$ **then**
    if $\tau < 0$ **then** $\tau_{\min} \leftarrow \tau$ **else** $\tau_{\max} \leftarrow \tau$
    $\tau \sim \mathcal{U}[\tau_{\min}, \tau_{\max}]$
    goto 6 ▷ Update proposal
  end if
  return $\mathcal{Z} = \mathcal{Z}_{\setminus i} \cup \{\mathbf{z}'_i\}$ ▷ Accept proposal
end procedure

Algorithm 2 Gibbs sampling algorithm targeting the joint posterior of the two-layer model returning M samples after a burn-in of B samples.

procedure 2LayerGibbs(M, B)
  initialize $\mathcal{Z}$ ▷ Arbitrary initialization
  for $m = -B, \dots, M$ **do**
    for $i = 0, \dots, p$ **do** ▷ For each latent variable
      $\mathcal{Z} \leftarrow \mathrm{ESS}(i, \mathcal{Z})$ ▷ update variable
    end for
    if $m > 0$ **then** ▷ Burn-in finished
      $\mathcal{Z}^{(m)} \leftarrow \mathcal{Z}$ ▷ Save samples
    end if
  end for
  return $\left\{\mathcal{Z}^{(m)}\right\}_{m=1}^{M}$
end procedure

4.1 Wiener-Hammerstein Cascades

Consider the Wiener-Hammerstein cascade in Fig. 2; it consists of a static nonlinear function $f(\cdot)$ sandwiched between two linear dynamical systems (represented by the impulse responses g_1 and g_2). The output is measured with additive Gaussian white noise.

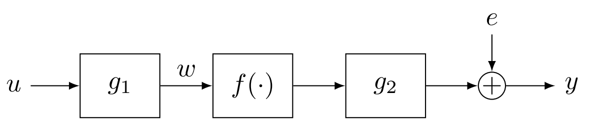

Fig. 2. The Wiener-Hammerstein cascade.

We suppose that we have collected N measurements of the output in a vector of samples $\mathbf{y}$. Similarly, we have collected the values of the input in the vector $\mathbf{u}$. Then, the N samples of the internal signal w can be represented by the vector

$$\mathbf{w} = \mathbf{G}_1 \mathbf{u},$$

where $\mathbf{G}_1$ is the $N \times N$ lower-triangular Toeplitz matrix of the impulse response samples that represents the convolution operated by g_1:

$$[\mathbf{G}_1]_{i,j} = \begin{cases} g_1[i-j+1] & \text{if } 0 < i-j+1 \le N, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

The whole system can be represented, in vector form, by

$$\mathbf{y} = \mathbf{G}_2 f(\mathbf{G}_1 \mathbf{u}) + \mathbf{e},$$

where $\mathbf{G}_2$, analogously to (6), is the Toeplitz matrix that represents the convolutions operated by g_2 and where the function $f(\cdot)$ acts on vectors elementwise. We suppose that the vector of noise samples $\mathbf{e}$ contains Gaussian white noise with variance σ^2.

We model the impulse responses of the linear systems as independent zero-mean Gaussian processes. This means that the vectors of impulse responses are multivariate Gaussian vectors with

$$\mathbf{g}_1 \sim \mathcal{N}(0, \mathbf{K}_{\mathbf{g}_1}), \qquad \mathbf{g}_2 \sim \mathcal{N}(0, \mathbf{K}_{\mathbf{g}_2}), \tag{7}$$

where the covariance matrices are determined by an appropriately chosen kernel function—for instance the *first order stable spline kernel* (Pillonetto et al., 2014):

$$\left[\mathbf{K}_{\mathbf{g}_1}\right]_{i,j} = \alpha_1 \cdot \beta_1^{\max(i,j)}, \tag{8}$$

where $\alpha_1 > 0$ is a scaling parameter that determines the overall amplitude of g_1 and $\beta_1 \in (0,1)$ is a shaping parameter that determines the overall exponential decay of g_1. We use a similar model for $\mathbf{K}_{\mathbf{g}_2}$.

In addition, we model the static nonlinearity using a zero-mean Gaussian process

$$\varphi(\cdot) \sim \mathcal{GP}\left(0, H(\cdot,\cdot)\right), \tag{9}$$

with a covariance function $H(\cdot,\cdot)$ suitable for functional estimation—for instance, the squared exponential kernel

$$H(x_i, x_j) = \eta \exp\left\{-\frac{1}{\rho}\,(x_i - x_j)^2\right\}, \tag{10}$$

where the length-scale parameter ρ determines the overall smoothness of the functions and η is a scaling parameter. Note, however, that the method is general and in no way limited to the presented kernel functions.

From the Gaussian-process model (9), we have that the intermediate variable $\mathbf{f} = \varphi(\mathbf{G}_1\mathbf{u})$, is a multivariate Gaussian vector with marginal distribution given by

$$\mathbf{f} \mid \mathbf{g}_1 \sim \mathcal{N}(0, \mathbf{H}), \tag{11}$$

where the elements of the marginal covariance matrix $\mathbf{H}$ are given by the covariance function (10) evaluated in the entries of $\mathbf{G}_1\mathbf{u}$. Note that this covariance matrix is a function of the random variable $\mathbf{g}_1$; however, we keep this dependency implicit for notational convenience.

Conditioned on the intermediate variable $\mathbf{f}$ and on the impulse response of the second system in the cascade, the output samples have a joint multivariate Gaussian distribution given by the independent noise samples:

$$\mathbf{y} \mid \mathbf{f}, \mathbf{g}_2 \sim \mathcal{N}\left(\mathbf{G}_2\mathbf{f},\ \sigma^2\mathbf{I}\right).$$

The complete nonparametric model of the Wiener-Hammerstein cascade is presented in Fig. 3 (left). It is important to note that the Gaussian process model (9) for the static nonlinearity is independent of the random variables $\mathbf{g}_1$ and $\mathbf{g}_2$; however, in (11), we are computing the marginal distribution of $\varphi(\cdot)$ evaluated in $\mathbf{G}_1\mathbf{u}$, hence, the distribution of $\mathbf{f}$ depends on the impulse response $\mathbf{g}_1$ (see Fig. 3, middle).

To formulate the nonparametric Wiener-Hammerstein model as a two-layer model, we marginalize out $\mathbf{f}$ according to

$$\mathrm{p}(\mathbf{y} \mid \mathbf{g}_1, \mathbf{g}_2) = \int \mathrm{p}(\mathbf{y} \mid \mathbf{f}, \mathbf{g}_2)\, \mathrm{p}(\mathbf{f} \mid \mathbf{g}_1)\, \mathrm{d}\mathbf{f}.$$

This marginalization is given by:

$$\mathbf{y} \mid \mathbf{g}_1, \mathbf{g}_2 \sim \mathcal{N}(0, \mathbf{G}_2\mathbf{H}\mathbf{G}_2^T + \sigma^2\mathbf{I}). \tag{12}$$

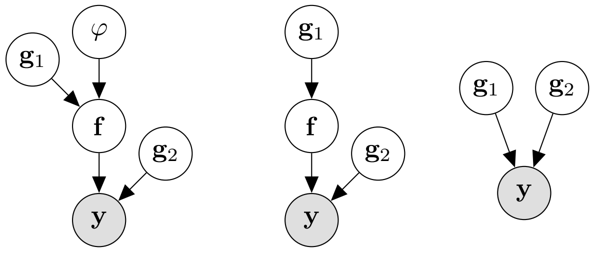

Fig. 3. Bayesian networks of the Wiener-Hammerstein model. *Left:* the complete nonparametric model. *Middle:* the model after integrating out φ (note that the vector $\mathbf{f}$ is conditionally dependent of $\mathbf{g}_1$). *Right:* Model after integrating out $\mathbf{f}$.

Collecting (7) and (12), we see that the Wiener-Hammerstein structure here discussed can be modeled using a two-layer stochastic representation (see Fig. 3, right):

$$\begin{cases} \mathbf{g}_1 \sim \mathcal{N}(0, \mathbf{K}_{\mathbf{g}_1}), \\ \mathbf{g}_2 \sim \mathcal{N}(0, \mathbf{K}_{\mathbf{g}_1}), \\ \mathbf{y} \mid \mathbf{g}_1, \mathbf{g}_2 \sim \mathcal{N}(0, \mathbf{G}_2\mathbf{H}\mathbf{G}_2^T + \sigma^2\mathbf{I}). \end{cases}$$

To evaluate the proposed approach, we consider a Wiener-Hammerstein cascade with

$$G_1(q) = \kappa_1 \frac{1 - 0.5q^{-1} + 0.6q^{-2}}{1 - 0.3q^{-1} - 0.45q^{-2} + 0.175q^{-3}},$$
$$G_2(q) = \kappa_2 \frac{1}{1 - 0.8q^{-1}},$$

where κ_1 and κ_2 are such that the blocks have unit gain. The static nonlinearity is given by the smooth and noninvertible function

$$f(x) = \frac{\sin(4\pi x)}{4\pi x}.$$

We generated $N = 300$ samples of the output of the system in response to a white-noise input, uniform in the interval $[-1, 1]$, and we corrupted the output measurements with Gaussian white noise with variance equal to 10% of the variance of the noiseless output.

We considered a two-layer model, where the impulse responses are modeled with stable-spline kernels such as (8) and the static nonlinearity is modeled using a squared exponential kernel (10).

Using the procedure presented in Alg. 2, we estimated posterior distribution of $\mathbf{g}_1$ and $\mathbf{g}_2$ using $M = 1000$ samples, collected after burn-in of $B = 100$ samples. The approximate wallclock time needed to estimate one system using the method is 2 minutes.

The hyperparameters of the stable-spline kernels were set using a validation set of 150 samples (50% of initial the training set): α_i were choosen among 6 values logarithmically spaced between 10^{-3} and 10^2; β_i were choosen among 6 values uniformly spaced between 0.2 and 0.8. As the model is not identifiable, the hyperparameters of the static nonlinearity were fixed arbitrarily to $\eta = 10$ and $\rho = 5$. Different models were estimated using 50% of the data as training and then evaluated using the prediction error on the remaining 50% of the data. Finally, the hyperparameters that minimized prediction error were used to sample the posterior using the whole dataset.

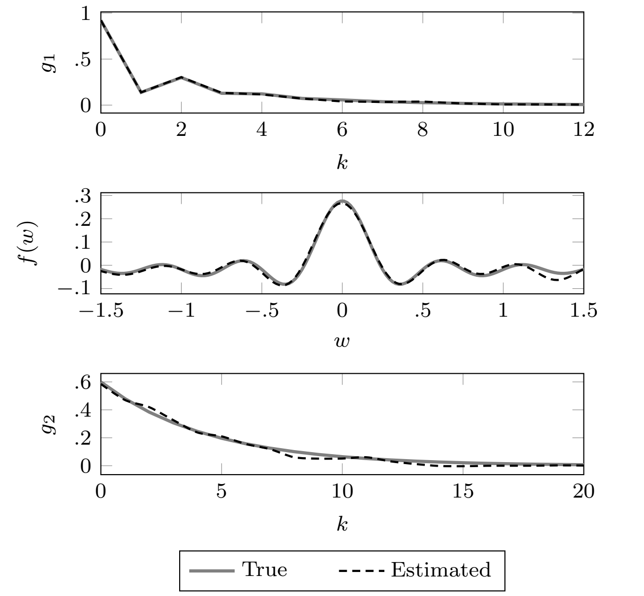

Fig. 4. Result of the identification of Wiener-Hammerstein cascades. Note that the blocks of the estimated system have been scaled to match the true system.

The result of this simulation is presented in Fig. 4. In the figure we see the characteristics of the blocks of the true system compared with the blocks estimated by the method. The marginalized $f(\cdot)$ was estimated using its posterior mean given the estimated g_1 and g_2. From this simulation, it appears that the proposed two-layer stochastic model is effective at modeling nonparametric Wiener-Hammerstein cascades.

4.2 Hammerstein-Wiener cascades

Consider the Hammerstein-Wiener cascade in Fig. 5; as we did for the Wiener-Hammerstein case, we represent the linear block with the vector of impulse response samples $\mathbf{g}$. The output of the cascade is subject to an additive noise $\mathbf{e}$.

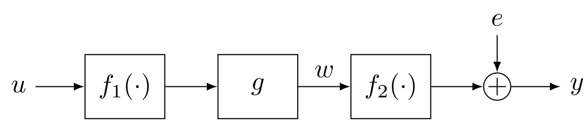

Fig. 5. The Hammerstein-Wiener cascade.

If we suppose that we have collected N measurements of the input in a vector of samples $\mathbf{u}$ then the samples of the internal signal w can be represented by the vector

$$\mathbf{w} = \mathbf{G}f_1(\mathbf{u}),$$

where $f_1(\cdot)$ is evaluated elementwise and where $\mathbf{G}$, analogously to (6), is the $N \times N$ lower-triangular Toeplitz matrix that represents the convolution operated by the linear system. Similarly, the N samples of the output can be represented, in vectorized form, as

$$\mathbf{y} = f_2(\mathbf{w}) + \mathbf{e},$$

where $\mathbf{e}$ is a vector of samples of Gaussian white mesurement noise with variance σ^2.

We model the impulse response of the linear system g with a zero-mean Gaussian process, obtaining the following multivariate Gaussian distribution for the vector $\mathbf{g}$:

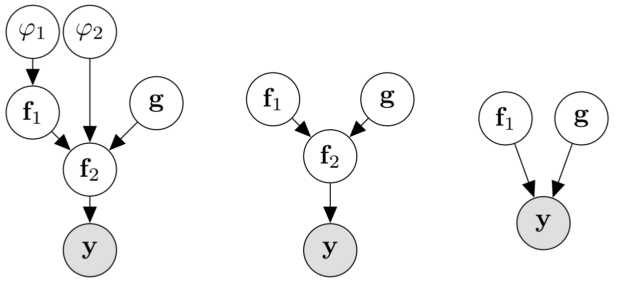

Fig. 6. Bayesian networks of the Hammerstein-Wiener model. *Left:* the complete nonparametric model. *Middle:* the model after integrating out $\varphi_1(\cdot)$ and $\varphi_2(\cdot)$ (note that the vector f_2 is conditionally dependent on g and f_1). *Right:* Model after integrating out the vector f_2.

$$\mathbf{g} \sim \mathcal{N}(0, K_{\mathbf{g}}), \tag{13}$$

where the covariance matrix is determined by the covariance function of the Gaussian process—for instance, the stable-spline kernel (8).

Similarly to what we did in the Wiener-Hammerstein case, we use zero-mean Gaussian-process models for the static nonlinearities,

$$\varphi_1(\cdot) \sim \mathcal{GP}\big(0, H_1(\cdot,\cdot)\big), \quad \varphi_2(\cdot) \sim \mathcal{GP}\big(0, H_2(\cdot,\cdot)\big),$$

for appropriate covariance functions $H_1(\cdot,\cdot)$ and $H_2(\cdot,\cdot)$—for example, the squared exponential kernel (10).

Considering the intermediate random variables $\mathbf{f}_1 = \varphi_1(\mathbf{u})$ and $\mathbf{f}_2 = \varphi_2(\mathbf{G}\mathbf{f}_1)$. We can write the marginal distributions

$$\mathbf{f}_1 \sim \mathcal{N}(0, \mathbf{H}_1), \quad \mathbf{f}_2 \mid \mathbf{f}_1, \mathbf{g} \sim \mathcal{N}(0, \mathbf{H}_2), \tag{14}$$

where the elements of $\mathbf{H}_1$ are given the covariance function evaluated in the entries of $\mathbf{u}$, and the elements of $\mathbf{H}_2$ are given by the covariance function evaluated in the entries of $\mathbf{G}\mathbf{f}_1$. Note that the covariance matrix $\mathbf{H}_2$ has a dependency on $\mathbf{f}_1$ and $\mathbf{g}$ which we keep implicit for notational convenience.

Conditioned on $\mathbf{f}_2$, the output has a multivariate normal distribution given by

$$\mathbf{y} \mid \mathbf{f}_2 \sim \mathcal{N}(\mathbf{f}_2, \sigma^2\mathbf{I}).$$

This complete nonparametric model of the Hammerstein-Wiener cascade is presented in Fig. 6 (left). Note again that the Gaussian process models are *a priori* independent and the dependency of $\mathbf{f}_2$ and $\mathbf{g}$ and $\mathbf{f}_1$ is introduced by marginalization (see Fig. 6, middle).

To formulate the two-layer model of the Hammerstein-Wiener cascade, we can marginalize out $\mathbf{f}_2$ according to

$$\mathrm{p}(\mathbf{y}|\mathbf{f}_1, \mathbf{g}) = \int \mathrm{p}(\mathbf{y}|\mathbf{f}_2)\, \mathrm{p}(\mathbf{f}_2|\mathbf{f}_1, \mathbf{g})\, \mathrm{d}\mathbf{f}_2.$$

Similarly to the Wiener-Hammerstein case, also this marginalization has a closed-form solution:

$$\mathbf{y} \mid \mathbf{f}_1, \mathbf{g} \sim \mathcal{N}(0, \mathbf{H}_2 + \sigma^2\mathbf{I}). \tag{15}$$

Collecting (13), (14), and (15), we see that the Hammerstein-Wiener structure here discussed can be modeled using a two-layer stochastic representation (see Fig. 6, right):

$$\begin{cases} \mathbf{g} \sim \mathcal{N}(0, \mathbf{K}_{\mathbf{g}}), \\ \mathbf{f}_1 \sim \mathcal{N}(0, \mathbf{H}_1), \\ \mathbf{y} \mid \mathbf{f}_1, \mathbf{g} \sim \mathcal{N}(0, \mathbf{H}_2 + \sigma^2\mathbf{I}). \end{cases}$$

To evaluate the approach, we consider a Hammerstein-Wiener cascade with

$$f_1(x) = x^3, \qquad f_2(x) = \frac{\sin(\pi x)}{\pi x},$$
$$G(q) = \frac{3 - 0.2q^{-1} - 0.32q^{-3}}{3.08 - 0.308q^{-1} - 0.9855q^{-2} + 0.1848q^{-3}}.$$

We generated $N = 300$ samples of the output of the system in response to a white-noise input, uniform in the interval $[-1, 1]$. We corrupted the output measurements with Gaussian white noise with variance equal to 10% of the variance of the noiseless output.

We considered a two-layer model where the impulse response is modeled with the stable-spline kernel (8). For the nonlinear blocks, we used squared exponential covariances such as (10).

Using the procedure presented in Alg. 2, we estimated the posterior distribution of $\mathbf{g}$ and $\mathbf{f}_1$ using $M = 1000$ samples, collected after a burn-in of $B = 100$ samples.

The result of this simulation is presented in Fig. 7. In the figure we see the characteristics of the blocks of the true system compared with the blocks estimated by the method. From this simulation, it appears that the proposed two-layer stochastic model is also effective at estimating Hammerstein-Wiener cascades.

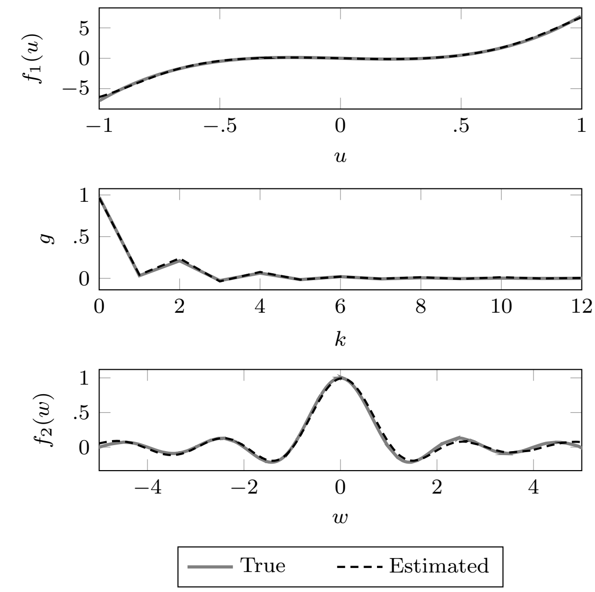

Fig. 7. Result of the identification of Hammerstein-Wiener cascades. Note that the blocks of the estimated system have been scaled to match the true system.

5. CONCLUSION

In this paper, we have proposed a two-layer stochastic model with an associated estimation algorithm. The estimation algorithm uses a Gibbs sampling procedure, paired with ESS, to target the posterior distribution of the latent variables. We have validated the approach on two example nonlinear systems showing that the proposed approach can be used to effectively identify nonparametric models of three-block cascades from data.

While the two layer stochastic approach seems very powerful and has many applications, there are limitations. In particular, the elliptical-slice sampling step involved in the algorithm can become expensive in large models or when there are many measurements: as shown in Algorithm 2, we need to sample every latent variable using a rejection scheme that involves the computation of the likelihood function. If the likelihood function is expensive to compute, this may lead to a large overall computational burden—for instance, in the examples considered here, the complexity of the likelihood function is cubic in the number of data. Accelerating the sampling procedure (possibly using variational approximations) is left as future work. In addition, there is no obvious way to tune the hyperparameters. In the simulation examples, we have used a validation set to select the hyperparameters. However, this is slow (for reference, the estimation is Sec. 4.1 took about 45 hours) and can possibly lead to overfitting. Extensions of the method attempting to estimate the hyperparameters from the marginal likelihood function are currently under study.

REFERENCES

Atam, E., Schulte, D.O., Arteconi, A., Sass, I., and Helsen, L. (2018). Control-oriented modeling of geothermal borefield thermal dynamics through Hammerstein-Wiener models. *Renewable Energy*, 120, 468–477.

Bai, E.W., Cai, Z., Dudley-Javorosk, S., and Shields, R.K. (2009). Identification of a modified Wiener–Hammerstein system and its application in electrically stimulated paralyzed skeletal muscle modeling. *Automatica*, 45(3), 736–743.

Bershad, N.J., Celka, P., and McLaughlin, S. (2001). Analysis of stochastic gradient identification of Wiener-Hammerstein systems for nonlinearities with Hermite polynomial expansions. *IEEE Trans. Signal Process.*, 49(5), 1060–1072.

Dewhirst, O.P., Simpson, D.M., Angarita, N., Allen, R., and Newland, P.L. (2010). Wiener-Hammerstein parameter estimation using differential evolution. In *Int. Conf. Bio-inspired Systems. Signal Process.*, 271–276.

Giordano, G., Gros, S., and Sjöberg, J. (2018). An improved method for Wiener–Hammerstein system identification based on the fractional approach. *Automatica*, 94, 349–360.

Giri, F. and Bai, E.W. (2010). *Block-oriented nonlinear system identification*. Springer.

Haryanto, A. and Hong, K.S. (2013). Maximum likelihood identification of Wiener–Hammerstein models. *Mechanical Systems and Signal Processing*, 41(1-2), 54–70.

Hasiewicz, Z. and Mzyk, G. (2004). Kernel instrumental variables for Hammerstein system identification. In *Proc. IEEE Int. Conf. Methods. Models. Autom. Robot.*

Ławryńczuk, M. (2016). Nonlinear predictive control of dynamic systems represented by Wiener–Hammerstein models. *Nonlinear Dynamics*, 86(2), 1193–1214.

Marconato, A., Sjöberg, J., and Schoukens, J. (2012). Initialization of nonlinear state-space models applied to the Wiener–Hammerstein benchmark. *Control Engineering Practice*, 20(11), 1126–1132.

Murray, I., Adams, R., and MacKay, D. (2010). Elliptical slice sampling. In *Proc. Int. Conf. Artif. Intell. Stat. (AISTAT)*, 541–548.

Nadimi, E.S., Green, O., Blanes-Vidal, V., Larsen, J.J., and Christensen, L.P. (2012). Hammerstein-Wiener model for the prediction of temperature variations inside silage stack-bales using wireless sensor networks. *Biosystems Eng.*, 112(3), 236–247.

Neal, R.M. (2003). Slice sampling. *Ann. Statist.*, 31(3), 705–767.

Ni, B., Gilson, M., and Garnier, H. (2013). Refined instrumental variable method for Hammerstein–Wiener continuous-time model identification. *IET Control Theory. Appl.*, 7(9), 1276–1286.

Oliver, J.A., Prieto, R., Cobos, J.A., Garcia, O., and Alou, P. (2009). Hybrid Wiener-Hammerstein structure for grey-box modeling of dc-dc converters. In *IEEE App. Pow. Elec. Conf. Expo. (APEC) 2009.*, 280–285. IEEE.

Paduart, J., Lauwers, L., Pintelon, R., and Schoukens, J. (2012). Identification of a Wiener–Hammerstein system using the polynomial nonlinear state space approach. *Control Eng. Prac.*, 20(11), 1133–1139.

Palanthandalam-Madapusi, H.J., Ridley, A.J., and Bernstein, D.S. (2005). Identification and prediction of ionospheric dynamics using a Hammerstein-Wiener model with radial basis functions. In *Proc. Amer. Control Conf., 2005*, 5052–5057. IEEE.

Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., and Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: a survey. *Automatica*, 50(3), 657–682.

Pillonetto, G. and Chiuso, A. (2009). Gaussian processes for Wiener–Hammerstein system identification. In *Proc. IFAC Symp. System Identification (SYSID)*, volume 15, 838–843.

Schoukens, M., Pintelon, R., and Rolain, Y. (2014). Identification of Wiener–Hammerstein systems by a nonparametric separation of the best linear approximation. *Automatica*, 50(2), 628–634.

Schoukens, M. and Tiels, K. (2017). Identification of block-oriented nonlinear systems starting from linear approximations: a survey. *Automatica*, 85, 272–292.

Tan, A.H., Wong, H.K., and Godfrey, K. (2012). Identification of a Wiener–Hammerstein system using an incremental nonlinear optimisation technique. *Control Eng. Prac.*, 20(11), 1140–1148.

Taringou, F., Hammi, O., Srinivasan, B., Malhame, R., and Ghannouchi, F.M. (2010). Behaviour modelling of wideband RF transmitters using Hammerstein-Wiener models. *IET Circuits Dev. Syst.*, 4(4), 282–290.

Vanbeylen, L. (2014). A fractional approach to identify Wiener–Hammerstein systems. *Automatica*, 50(3), 903–909.

Wang, J., Ye, Z., Hu, Z., Wu, X., Dimirovsky, G., and Yue, H. (2017). Model-based acceleration control of turbofan engines with a Hammerstein-Wiener representation. *Int. J. Turbo. Jet-Engines*, 34(2), 141–148.

Westwick, D.T. and Schoukens, J. (2012). Classification of the poles and zeros of the best linear approximations of Wiener-Hammerstein systems. *IFAC Proceedings Volumes*, 45(16), 470–475.

Zhu, Y. (2002). Estimation of an N-L-N Hammerstein-Wiener model. *Automatica*, 38(9), 1607–1614.