Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 54-7 (2021) 755–760

Bayes Control of Hammerstein Systems

Mina Ferizbegovic * **Per Mattsson** ** **Thomas B. Schön** ** **Håkan Hjalmarsson** *

* *Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden (e-mail: {minafe, hjalmars}@ kth.se)*
** *Department of Information Technology, Uppsala University, Uppsala, Sweden (e-mail: {per.mattsson, thomas.schon}@it.uu.se)*

Abstract: In this paper, we consider data driven control of Hammerstein systems. For such systems a common control structure is a transfer function followed by a static output nonlinearity that tries to cancel the input nonlinearity of the system, which is modeled as a polynomial or piece-wise linear function. The linear part of the controller is used to achieve desired disturbance rejection and tracking properties. To design a linear part of the controller, we propose a weighted average risk criterion with the risk being the average of the squared L_2 tracking error. Here the average is with respect to the observations used in the controller and the weighting is with respect to how important it is to have good control for different impulse responses. This criterion corresponds to the average risk criterion leading to the Bayes estimator and we therefore call this approach Bayes control. By parametrizing the weighting function and estimating the corresponding hyperparameters we tune the weighting function to the information regarding the true impulse response contained in the data set available to the user for the control design. The numerical results show that the proposed methods result in stable controllers with performance comparable to the optimal controller, designed using the true input nonlinearity and true plant.

Keywords: Bayesian methods, Model reference control, Hammerstein system.

1. INTRODUCTION

Hammerstein systems are nonlinear systems that consist of a static input nonlinearity followed by a linear dynamic system. Existing methods for the identification of Hammerstein systems are instrumental variable methods (Stoica and Söderström, 1982), overparametrization methods (Bai, 2004), subspace methods (Goethals et al., 2005), kernel-based methods (Risuleo et al., 2015, 2017) etc.

Control of Hammerstein systems consists of two steps: identification of the Hammerstein system and the control design. It is usually performed by canceling the estimated input nonlinearity via inversion and using robust control theory to synthesize a controller for the remaining linear dynamics of the system (Bloemen et al., 2000; Ingram et al., 2005). This paper aims to provide the control method for Hammerstein systems that cancels the estimated input nonlinearity and provides a data-driven robust control design for the linear dynamics in a deterministic setting, which takes into account errors made by canceling the nonlinearity. In particular, for the linear control design, we consider the model reference control problem for the Hammerstein system, where the control objective is to design a fixed order controller such that the closed-loop system response resembles a given reference model. We take the model uncertainty into account by taking the average of this control objective with respect to a weighting function, which defines how important it is to have a good control performance for a given impulse response.

To solve the model reference control problem, both model-based and data-driven methods have been proposed. The model-based approach to this problem is to identify a model of the plant, then use the model to find the controller so that the closed-loop system consisting of the model and the controller corresponds to the given reference model, and subsequently perform a model-order reduction step if needed (Anderson and Liu, 1989). The main limitation of this approach is modeling errors. To account for this limitation, several data-driven methods have been proposed. Direct data-driven methods use only input/output data to derive the fixed order controller without explicitly forming a model. Both one-shot methods, as well as iterative methods where the controller is refined based on new data, have been developed. Well-know iterative methods include Iterative feedback-tunning (Hjalmarsson, 2002; Hjalmarsson et al., 1998), and Iterative Correlation-based Tuning (Karimi et al., 2004). The two most common non-iterative data-driven methods are Virtual Reference Feedback Tuning (Campi et al., 2002; Guardabassi and Savaresi, 2000), and Correlation-based Tuning (Karimi et al., 2007).

In Scampicchio et al. (2019), the authors propose a control design method for the model reference control problem for linear systems which accounts for model uncertainty. Given a data set to be used for the control design, and modeling the impulse response as a random vector, the design criterion is to minimize the posterior mean of the

⋆ This work was supported by the research environment NewLEADS—New Directions in Learning Dynamical Systems supported by the Swedish Research Council under contract 2016-06079; and Wallenberg AI, Autonomous Systems and Software Program (WASP), funded by Knut and Alice Wallenberg Foundation.

10.1016/j.ifacol.2021.08.452

model reference criterion which is taken as the squared L_2 error between the reference model and the achieved complementary sensitivity function. This method can thus be seen as a robust control design where the uncertainty in the estimate is used in the design criterion. Another feature of this method is that the hyperparameters of the prior distribution of the model parameters are tuned using marginalized maximum likelihood, i.e. Empirical Bayes. This means that the approach is not a *bona fide* Bayesian approach so that what is used as posterior is not actually a posterior. Nevertheless the results in Scampicchio et al. (2019) are very impressive.

In this paper, we present a control method for Hammerstein systems; see Section 2 for a detailed problem formulation. A discussion of the used nonlinearities is in Section 3. Our specific contributions are the following. First, we re-derive the method in Scampicchio et al. (2019), but without the Bayesian assumption on the impulse response in Section 4. For this, we will use an average risk criterion as a starting point. We call this type of control design Bayes control. Second, we will also show in Section 5 that the use of the Empirical Bayes estimator can be argued from a deterministic perspective as well. Third, we extend the approach from the linear setting detailed in Scampicchio et al. (2019) to Hammerstein systems in Section 6. A comparison of the proposed method with the optimal controller is presented in Section 7. The paper ends with some concluding remarks.

Notation. $T(g)$ is a lower Toeplitz matrix with first column equal to the vector g. g_X denotes the vector of the first n impulse response coefficients of the transfer function X. For a transfer function F, $\|F\|_\Phi^2 := \frac{1}{2\pi}\int_{-\pi}^{\pi}|F(e^{i\omega})|^2\Phi(\omega)d\omega$. $\mathcal{N}(m, R)$ denotes the normal distribution with mean m and covariance matrix R.

2. PROBLEM STATEMENT

We consider the Hammerstein system, which is shown in Figure 1. The nonlinearity is a function of a known basis $(\gamma_1, \gamma_2, \cdots, \gamma_p)$ as

$$\bar{u}(t) = f(u(t)) = \sum_{k=1}^{p} c_k \gamma_k(u(t)),$$

where $u(t)$ is the system input, $\bar{u}(t)$ is the unknown output of the nonlinear block, and the coefficients $c = [c_1\ c_2 \ldots c_p]^\top$ are to be estimated. The linear dynamic block is

$$y(t) = G(q, g_G)\bar{u}(t) + w(t) = \sum_{k=0}^{n} g_k \bar{u}(t-k) + w(t), \quad (1)$$

where q is the time-shift operator, $y(t)$ is the output of the system which is corrupted by the white noise $w(t)$, and $g_G = [g_1\ g_2 \ldots g_n]^\top$ are impulse response coefficients to be estimated. We introduce the vector-based notation:

$$u_{\text{obs}} = [u(0)\cdots u(N-1)]^\top, \quad y_{\text{obs}} = [y(1)\cdots y(N)]^\top$$
$$\bar{u}_{\text{obs}} = [\bar{u}(0)\cdots \bar{u}(N-1)]^\top, \quad w_{\text{obs}} = [w(1)\cdots w(N)]^\top.$$

Then, $y_{\text{obs}} = \Phi_{\text{obs}} g_G + w_{\text{obs}}$, where $\Phi_{\text{obs}} = T(\bar{u}_{\text{obs}})$. Notice that Φ_{obs} depends on parameters c, as $\bar{u}_{\text{obs}} = F(u_{\text{obs}})c$.

The objective of model reference control is to design a controller that achieves a desired closed-loop performance

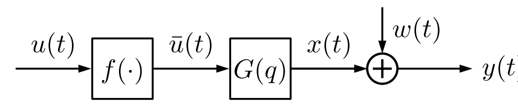

Fig. 1. Block scheme of the Hammerstein system.

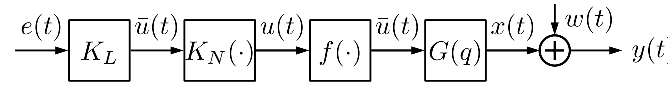

Fig. 2. Control of Hammerstein systems.

specified by a reference model $M(q)$. In this paper, we assume that the reference model is given, but discussion on how to choose the reference model can be found in, e.g., (Åström and Murray, 2010; Dehghani et al., 2006).

For a linear system we can use a feedback $u(t) = K_L(q;\rho)(r(t) - y(t)) = K_L(q;\rho)e(t)$, where $r(t)$ is the reference signal and K_L is parametrized by ρ. In this paper we will use the a linear parametrization

$$K_L(q;\rho) = \alpha(q)^\top\rho, \rho \in \mathbb{R}^{n_\rho}$$

where $\alpha(q)$ is a transfer function vector consisting of some pre-specified basis functions $\{\alpha_i(q)\}_{i=1}^{n_\rho}$.

In this paper, we consider a nonlinear Hammerstein system, so the controller will also contain a nonlinear part, see Figure 2. The goal of the linear control block K_L is to mitigate the effect of the input nonlinearity, as will be described in Section 3.

To determine how well the closed-loop system matches with the reference model $M(q)$, a reasonable loss function, used by Scampicchio et al. (2019), is

$$J(\rho) = \left\| M - \frac{K_L(q;\rho)G}{1 + K_L(q;\rho)G} \right\|_\Phi^2, \quad (2)$$

which corresponds to the average error between the desired output $y_M(t) := M(q)r(t)$ and the closed loop output achieved with the parameter ρ in the controller, when the reference signal is stationary with spectrum Φ.

In this paper we instead use the exact finite-time loss for a closed-loop experiment of length N with a given reference signal sequence $\{r(t)\}_{t=1}^N$

$$J_f(g_G, \rho, r) := \frac{1}{N}\sum_{t=1}^{N}(y_M(t) - y(t;\rho))^2. \quad (3)$$

We consider the finite-time loss function J_f rather than the infinite horizon loss J to avoid technical complications that arise regarding stability.

When r is stationary with spectrum Φ this loss function converges to (2) as $N \to \infty$. We can write (3) on vector form by introducing $r := [r(1)\ \ldots\ r(N)]^\top \in \mathbb{R}^N$,

$$y_M := \begin{bmatrix} y_M(1) \\ \vdots \\ y_M(N) \end{bmatrix} = T(g_M)r,$$
$$y(g_G, \rho) := P(g_G, \rho)r,$$

where $P(g_G, \rho)$ is the lower Toeplitz matrix $P(g_G, \rho) := (I + T(g_G)T(g_{K(\rho)}))^{-1}T(g_G)T(g_{K(\rho)})$. This gives

$$J_f(g_G, \rho, r) = \frac{1}{N}\|y_M - y(g_G, \rho)\|^2.$$

We will assume r to be white, in which case for large N, we can use

$$J_f(g_G, \rho) = \frac{1}{N}\mathrm{Tr}\{\Delta^\top(g_G, \rho)\Delta(g_G, \rho)\}, \quad (4)$$

where $\Delta(g_G, \rho) = T(g_M) - P(g_G, \rho)$.

The problem considered in this paper is to select the controller parameters ρ such that the loss function is small for the true system, and coefficients c such that we mitigate the nonlinearity in the input. The available information about the system is an input-output data set $\{u_{\text{obs}}, y_{\text{obs}}\}$ collected from (1).

3. INPUT NONLINEARITY

We will consider two types of basis functions γ based on polynomial and piecewise-linear functions. We will explain how to construct a nonlinear controller K_N for both basis functions.

3.1 Polynomial input nonlinearity

For the polynomial input nonlinearity, we have that $\gamma_k(u(t)) = u(t)^{k-1}$. If we then choose the nonlinear controller block K_N as a root of the equation $f(u(t)) = \bar{u}(t)$, we will cancel the input nonlinearity. In this paper we assume that the nonlinearity is of odd order, i.e. p is even, as this guarantess that a real root always exists. In this way, we mitigate the effect of polynomial input nonlinearity.

3.2 Piecewise input nonlinearity

In the case of the piecewise input nonlinearity, we have to choose a set of grid points as $u^{\text{grid}} = \left\{u_1^{\text{grid}}, u_2^{\text{grid}}, \ldots, u_m^{\text{grid}}\right\}$. The basis function is $\gamma_1(u(t)) = 1$, and for $i \geq 2$

$$\gamma_i(u(t)) = \begin{cases} u(t) - u_{i-1}^{\text{grid}}, & \text{if } u(t) \in \left[u_{i-1}^{\text{grid}}, u_i^{\text{grid}}\right] \\ u_i^{\text{grid}} - u_{i-1}^{\text{grid}}, & \text{if } u(t) > u_i^{\text{grid}} \\ 0, & \text{otherwise.} \end{cases}$$

With these basis functions, the parameters in c corresponds to the slope between between the grid points.

In order to find a nonlinear control block K_N, we compute $x^{\text{grid}} = \left\{x_1^{\text{grid}}, x_2^{\text{grid}}, \ldots, x_p^{\text{grid}}\right\}$ using previously estimated coefficients c, where $x_i^{\text{grid}} = F(u_i^{\text{grid}})c$. It may not be possible to cancel the input nonlinearity for all values $\bar{u}(t)$. Therefore, we find a compensator for any value of $\bar{u}(t)$, by introducing the saturation function, such that

$$\hat{u}(t) = \begin{cases} \max(x^{\text{grid}}), & \text{if } \bar{u}(t) > \max(x^{\text{grid}}) \\ \min(x^{\text{grid}}), & \text{if } \bar{u}(t) < \min(x^{\text{grid}}) \\ \bar{u}(t), & \text{otherwise.} \end{cases}$$

Then, the output of the nonlinear controller block is

$$K_N(\bar{u}(t)) = u_{\text{ind-1}}^{\text{grid}} + \frac{\hat{u}_t - F_{1:\text{ind}-1}(u_{\text{ind-1}}^{\text{grid}})c_{1:\text{ind}-1}}{c_{\text{ind}}},$$

where

$$\text{ind}(\bar{u}(t)) = \left\{i, \text{ if } \hat{u}(t) \in \left[\min(x_{i-1}^{\text{grid}}, x_i^{\text{grid}}), \max(x_{i-1}^{\text{grid}}, x_i^{\text{grid}})\right]\right\}.$$

After constructing the nonlinear controller part K_N to mitigate uncertainty, we design a linear controller K_L.

4. BAYES CONTROL

In this section, we will derive a robust control design method that uses a finite set of observations y_{obs} from the system [1]. The latter means that the controller parameter is a function of the data: $\rho = \rho(y_{\text{obs}})$. Hence, $J_f(g_G, \rho(y_{\text{obs}}))$ in (4) is the loss function for a given experimental set-up that generated the data. However, $J_f(g_G, \rho(y_{\text{obs}}))$ will be dependent on the fluctuations in the noise and in order to have a criterion that does not depend on this we introduce the averaged loss function

$$\begin{aligned} R(g_G, \rho(\cdot)) &= \mathbb{E}[J_f(g_G, \rho(y_{\text{obs}}))] \\ &= \int J_f(g_G, \rho(y_{\text{obs}}))p(y_{\text{obs}}|g_G)dy_{\text{obs}}, \end{aligned}$$

where $p(y_{\text{obs}}|g_G)$ denotes the distribution of the data given that the system impulse response is g_G. $R(g_G, \rho(\cdot))$ thus measures the long term performance of using the policy $\rho(\cdot)$ when the true system has impulse response g_G.

Our control design criterion is now defined as

$$r_g(\rho(\cdot)) = \int R(g_G, \rho(\cdot))p_g(g_G)dg_G,$$

where p_g is a non-negative weighting function, for convenience normalized so that $\int p_g(g_G)dg_G = 1$. The weighting should be chosen to be large for impulse responses for which it is important to have a good controller, and small otherwise. In particular it should be large for the true system and we will return to this issue in Section 5.

We call the controller minimizing $r_g(\rho(\cdot))$, i.e.

$$\rho^*(\cdot) := \arg\min_{\rho(\cdot)} r_g(\rho(\cdot)) \quad (5)$$

the Bayes controller. At first sight it may seem like a formidable task to solve this functional minimization problem. Now, the minimizer [2] of $R(g_G, \rho)$ is a function of g_G, let us denote this function by $\bar{\rho}(g_G)$. We can thus interpret our problem as that of finding a data-driven estimator $\rho(y_{\text{obs}})$ which accurately can estimate the function value of $\bar{\rho}(\cdot)$ at the true system parameter g_G, $\bar{\rho}(g_G)$. From this inference perspective of our control problem, $R(g_G, \rho(\cdot))$ can be interpreted as a *risk function*. It measures the long term error of our "estimator" $\rho(\cdot)$. We notice that this function is non-negative. Further, the function r_g can be interpreted as the average risk over the possible impulse responses g_G, using p_g as weighting (or density). This means that $\rho^*(\cdot)$ can be interpreted as the Bayes estimator of $\bar{\rho}(\cdot)$.

We can directly derive the solution by reversing the order of integration in the definition of r_g

$$r_g(\rho(\cdot)) = \int \int J_f(g_G, \rho(y_{\text{obs}}))p(y_{\text{obs}}|g_G)p_g(g_G)dg_G \, dy_{\text{obs}}$$

and noticing that since the integrand is non-negative we can obtain $\rho^*(y_{\text{obs}})$ by minimizing the inner integral with respect to $\rho(y_{\text{obs}})$. We thus have the following result.

Theorem 1. The Bayes controller (5) is given by

$$\rho^*(y_{\text{obs}}) = \arg\min_{\rho} \tilde{V}(\rho, y_{\text{obs}}), \quad (6)$$

[1] We consider that the data have been collected in open loop and therefore the corresponding input u_{obs} is seen as deterministic and not included in our observation vector.

[2] Here we consider ρ to be a fix parameter.

where $\tilde{V}(\rho, y_{\text{obs}}) = \int J_f(g_G, \rho) p(y_{\text{obs}}|g_G) p_g(g_G) dg_G$.

Remark: Other criteria than model reference control can be used. This only entails changing the loss function J_f.

Introducing

$$p(y_{\text{obs}}) = \int p(y_{\text{obs}}|g_G) p_g(g_G) dg_G,$$

$$p(g_G|y_{\text{obs}}) = \frac{p(y_{\text{obs}}|g_G) p_g(g_G)}{p(y_{\text{obs}})},$$

which can be interpreted as the marginalized likelihood of the observations and the posterior distribution of g_G given that g_G has distribution p_G, we can write $\tilde{V}(\rho, y_{\text{obs}})$ as

$$\tilde{V}(\rho, y_{\text{obs}}) = \int J_f(g_G, \rho) p(g_G|y_{\text{obs}}) dg_G \; p(y_{\text{obs}}),$$

where we can interpret

$$V(\rho, y_{\text{obs}}) = \int J_f(g_G, \rho) p(g_G|y_{\text{obs}}) dg_G = \mathbb{E}[J_f(g_G, \rho)|y_{\text{obs}}] \quad (7)$$

as the conditional mean of the loss function given the observations y_{obs}. This is the criterion introduced in Scampicchio et al. (2019). Above, we have shown that the same design criterion can be obtained starting from an average risk criterion *without* any stochastic assumptions on g_G, as minimization of V and $\tilde{V}$ are equivalent for given y_{obs}. This result parallels that the Bayes estimator can be derived either using stochastic assumptions on the quantity to be estimated or an average risk criterion using a deterministic assumption on the quantity of interest.

5. TUNING THE WEIGHTING FUNCTION

Ideally we would like the weighting function $p_g(g_G)$ to be largest around the true value of g_G^o of g_G. To fix ideas suppose that p_G is the probability density function (pdf) of a normal distributed random variable $\mathcal{N}(0, P)$. Pretending that we know g_G^o this means that we should choose P such that the density $\mathcal{N}(0, P)$ is large at g_G, meaning that we should skew the distribution in the direction of g_G. Suppose that P is parametrized by $\beta \in \mathbb{R}^{n_\beta}$ so that $P = P(\beta)$, let us denote the corresponding normal density by $p_g(g_G; \beta)$. Then different optimization criteria can be formulated so that $p_g(g_G; \beta)$ is maximized around g_G^o. The most obvious would be to take

$$\beta = \arg\max_{\beta} p_g(g_G^o, \beta).$$

We can of course replace p_g by $\log p_g$ and for our normal distribution $\mathcal{N}(0, P(\beta))$ this gives

$$\beta = \arg\max_{\beta} (g_G^o)^\top P^{-1}(\beta) g_G^o + \log\det P(\beta)$$

after removing terms independent of P. Now, g_G^o is not available so we will have to use a surrogate in this expression. One possibility would be the maximum likelihood estimate which, assuming the noise to be $w \sim \mathcal{N}(0, \sigma^2 I)$, is given by

$$\hat{g}_{\text{ML}} = (\Phi_{\text{obs}}^\top \Phi_{\text{obs}})^{-1} \Phi_{\text{obs}}^\top y_{\text{obs}}.$$

We can also take the distribution of this estimate into account. Being linear in y_{obs}, $\hat{g}_{\text{ML}} \sim \mathcal{N}\left(g_G^o, \sigma^2 (\Phi_{\text{obs}}^\top \Phi_{\text{obs}})^{-1}\right)$. This means that

$$\mathbb{E}\left[\hat{g}_{\text{ML}} \hat{g}_{\text{ML}}^\top\right] = g_G^o (g_G^o)^\top + \sigma^2 (\Phi_{\text{obs}}^\top \Phi_{\text{obs}})^{-1}.$$

Pursuing the same idea as when g_G^o was assumed known leads to that one way to select β could be

$$\begin{aligned}\beta = \arg\max_{\beta} \; & \hat{g}_{\text{ML}}^\top \left(P(\beta) + \sigma^2 (\Phi_{\text{obs}}^\top \Phi_{\text{obs}})^{-1}\right)^{-1} \hat{g}_{\text{ML}} \\ & + \log\det\left(P(\beta) + \sigma^2 (\Phi_{\text{obs}}^\top \Phi_{\text{obs}})^{-1}\right).\end{aligned} \quad (8)$$

This criterion is nothing but the well known Empirical Bayes estimate, also known as the marginalized maximum likelihood estimate. Here we have shown that this estimate can be viewed as a way of tuning the weighting function p_G so that it is large where data tells us that the true system is without reference to that g_G itself is random and normal distributed. Thus combining Empirical Bayes estimation and Bayes control can be viewed as a data driven robust control design method in a deterministic setting with respect to the system itself.

6. IMPLEMENTATION

In this section, we present approximations and simplifications necessary for solving (6). For notational convenience we will use probabilistic notation, e.g. denote p_G as the distribution of g_G. Note, however that in our model it is only the measurement noise $w(t)$ that is considered stochastic. We will therefore use the probabilistic interpretation (7)

$$V(\rho, y_{\text{obs}}) = \mathbb{E}[J_f(g_G, \rho)|\, y_{\text{obs}}], \quad (9)$$

of the cost function, where the expectation is over g conditioned on y_{obs}. In order to be able to minimize this function, we first need to be able to compute it efficiently.

Firstly we need the conditional distribution $p(g_G|y_{\text{obs}})$ which in turn depends on p_g. Here we will make the simple assumption that the weighting function p_g corresponds to the pdf of a normal distribution with zero mean and covariance matrix P, i.e. we can think of g_G as being $g_G \sim \mathcal{N}(0, P)$. Furthermore, we assume that the noise $w \sim \mathcal{N}\left(0, \sigma^2 I\right)$, and this gives

$$p\left(\begin{bmatrix} g_G \\ y_{\text{obs}} \end{bmatrix}; \theta\right) \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P & P\Phi_{\text{obs}}^\top \\ \Phi_{\text{obs}} P & \Phi_{\text{obs}} P \Phi_{\text{obs}}^\top + \sigma^2 I \end{bmatrix}\right),$$

so that $g_G|y_{\text{obs}} \sim \mathcal{N}\left(\hat{g}_p, \hat{P}_p\right)$, where

$$\hat{g}_p = P\Phi_{\text{obs}}^\top (\Phi_{\text{obs}} P \Phi_{\text{obs}}^\top + \sigma^2 I)^{-1} y,$$

$$\hat{P}_p = P - P\Phi_{\text{obs}}^\top (\Phi_{\text{obs}} P \Phi_{\text{obs}}^\top + \sigma^2 I)^{-1} \Phi_{\text{obs}} P.$$

With, as in the previous section $P = P(\beta)$, the obtained controller is thus a function of the vector $\theta = \begin{bmatrix} c^\top & \sigma^2 & \beta^\top \end{bmatrix}^\top$, which we call the vector of hyperparameters. Following the argument in the preceeding section we will use the Empirical Bayes approach (Maritz and Lwin (1989)) to this end implemented using the Expectation-Maximization (EM) method. The following theorem, taken from Risuleo et al. (2015), is used to update θ.

Theorem 2. Assume that the hyperparameters at k-th interation $\hat{\theta}^k$ is available. The updated estimate

$$\hat{\theta}^{k+1} = \left[\hat{c}^{k+1}, \hat{\sigma}^{2,k+1}, \hat{\beta}^{k+1}\right]$$

is obtained utilizing the following steps:

- The coefficients of the input nonlinearity are update by $\hat{c}^{k+1} = (\hat{A}^k)^{-1} \hat{b}^k$, where

$$\hat{A}^k = (RF(u_{\text{obs}}))^\top ((\hat{P}_p^k + \hat{g}_p^k \hat{g}_p^{k\top}) \otimes I_N) RF(u),$$

$$\hat{b}^k = F(u_{\text{obs}})^\top T_N(\hat{g}_p^k)^\top y,$$

  and $R \in \mathbb{R}^{Nn \times N}$ is a matrix such that we have $Ru_{\text{obs}} = \text{vec}(T_n(u_{\text{obs}}))$. The mean and covariance distribution of

posterior distribution at iteration k, i.e, calculated using Φ_{obs}^k and $\hat{\sigma}^{2,k}$, are $\hat{g}_p^k$ and $\hat{P}_p^k$, respectively.
- The noise variance is updated using
$$\hat{\sigma}^{2,k+1} = \frac{1}{N}\left(\left\|y - \Phi_{\text{obs}}^{k+1}\hat{g}_p^k\right\|_2^2 + \operatorname{Tr}\left\{\Phi_{\text{obs}}^{k+1}\hat{P}_p^k\Phi_{\text{obs}}^{k+1\top}\right\}\right),$$
where $\Phi_{\text{obs}}^{k+1} = T_n(F(u_{\text{obs}})\hat{c}^{k+1})$;
- The updated kernel parameters are updated by
$$\hat{\beta}^{k+1} = \arg\min_{\beta} \operatorname{Tr}\left\{P(\beta)^{-1}(\hat{P}_p^k + \hat{g}_p^k\hat{g}_p^{k\top})\right\} + \log\det P(\beta).$$

The previous expression used to select β is not the same as in the previous section in (8) because it is derived using the EM method to implement Empirical Bayes approach.

Next, we need to compute the cost function (9). While there are a number of approaches to approximately do this when the pdf is Gaussian, we will here use an importance sampling approach to reduce the asymptotic variance of the estimator. We use the proposal distribution $q(x) = \mathcal{N}(x|\hat{g}_p, 2\hat{P}_p)$ that we sample m times
$$V(\rho, y_{\text{obs}}) \approx \frac{1}{m}\sum_{k=1}^{m}\frac{p(g_{G,k})}{q(g_{G,k})}J_f(g_{G,k},\rho), \tag{10}$$
where $J_f(g_{G,k},\rho)$ is defined in (4), $g_{G,k}$, $k = 1,\dots,m$ are independent samples from the proposal distribution $\mathcal{N}\left(\hat{g}_p, 2\hat{P}_p\right)$, and $p(x) = \mathcal{N}(x|\hat{g}_p,\hat{P}_p)$. Finally, ρ is found by numerical minimization of the expression (10). Finding ρ is a non-convex problem, which is not surprising since solving many other problems using kernel-based methods is also non-convex, e.g., tuning the hyperparameters (8).

7. NUMERICAL EXPERIMENTS

In this section, we perform a simulation study [3] to illustrate the performance of the proposed method. Mainly, we observe how the simulations support that the proposed method performs comparably to the optimal controller. Also, we illustrate that the obtained results depend on how well the input nonlinearity can be modeled.

We compare the following methods:

- the optimal controller (`opt`) according to the model reference control objective (2), designed using true input nonlinearity and the true plant G;
- the proposed method using a polynomial input nonlinearity (`bc-p`);
- the proposed method using a piecewise linear input nonlinearity (`bc-pw`);
- the proposed method without a nonlinearity (`bc-lin`).

All methods are implemented in MATLAB 2018a, with `bc-p` and `bc-pw`, and `bc-lin` using $m = 10$ samples in the importance sampling to estimate the proposed cost function (9), and the length of the horizon is $N = 200$. To model impulse response, we have used the first-order stable spline kernel (also known as Tuned-Correlated (TC) kernel) (Chen et al., 2012; Pillonetto and De Nicolao, 2010) and order $n = 100$ for all cases. We will model input nonlinearity as a polynomial of 5th order for `bc-p`, and as a piecewise linear input nonlinearity with 10 intervals for `bc-pw`. As the coefficients of linear and nonlinear part of the Hammerstein system, can be only determined up to a scaling constant, to provide uniqueness we fix $\|g\|_2 = 1$. We update the hyperparameters θ using Theorem 2 until $\left\|\theta^{k+1} - \theta^k\right\|_2 \le 10^{-3}$. Since minimizing the cost function (10) is non-convex optimization problem, the obtained solution depends on an initial controller. Our initial controller is chosen by minimizing $\frac{1}{m}\sum_{k=1}^m J_{app}(\rho, G_k)$ where $J_{app}(\rho, G_k) = \|M - K_L(q;\rho)G_k(1-M)\|$ is an approximation of $J(g_{G,k},\rho)$.

7.1 Case study 1

In the first case study, we will consider the benchmark example of Landau et al. (1995) for assessing the performance of the proposed method. Here the discrete-time function of the plant is
$$G(q^{-1}) = \frac{0.28261q^{-3} + 0.50666q^{-4}}{A(q^{-1})},$$
where $A(q^{-1}) = 1 - 1.41833q^{-1} + 1.58939q^{-2} - 1.31608q^{-3} + 0.88642q^{-4}$. The controller has the following structure
$$K(\rho, q^{-1}) = \frac{\sum_{k=0}^{5}\rho_i q^{-k}}{1 - q^{-1}}$$
and the closed-loop reference model is given by
$$M(q^{-1}) = \frac{K^(\rho^*, q^{-1})G(q^{-1})}{(1 + K^(\rho^*, q^{-1})G(q^{-1})},$$
where $\rho^* = [0.2045, -0.2715, 0.2931, -0.2396, 0.1643, 0.0084]^\top$. We consider the true input nonlinearity to be
$$\bar{u}(t) = \begin{cases} u(t) + 1, & \text{if } u(t) < -1 \\ u(t) - 1, & \text{if } u(t) > 1 \\ 0, & \text{otherwise.} \end{cases}$$
The output of the nonlinear part K_N^* is
$$u(t) = \begin{cases} \bar{u}(t) - 1, & \text{if } \bar{u}(t) < 0 \\ \bar{u}(t) + 1, & \text{if } \bar{u}(t) > 0 \\ 0, & \text{otherwise.} \end{cases}$$
The input u_{obs} with length $N_{\text{obs}} = 500$ and the reference r are zero-mean Gaussian noise signals with standard deviation $\sigma_u = 1$ and $\sigma_r = 1$. The output is disturbed by zero-mean Gaussian noise with $\sigma = 0.5$ uncorrelated with the input signal. 50 Monte Carlo simulations are performed with different noise realizations. The results are presented on the left side of Figure 3. We observe that `bc-pw` and `bc-p` performe almost as good as the optimal controller `opt`.

7.2 Case study 2

The second example originates from Boutayeb et al. (1996). The plant is given by
$$G(q^{-1}) = \frac{q^{-1} + 0.6q^{-2}}{1 - q^{-1} + 0.8q^{-2}}.$$
The true input nonlinearity is the polynomial
$$\bar{u}(t) = f(u(t)) = 2.8u(t) - 4.8u(t)^2 + 5.7u(t)^3.$$
The control objective is expressed by the reference model
$$M(q^{-1}) = \frac{q^{-3}(1-\alpha)^2}{(1-\alpha q^{-1})^2},$$
where $\alpha = e^{-0.5}$. The controller structure is given by
$$K(\rho, q^{-1}) = \frac{\sum_{k=0}^{2}\rho_i q^{-k}}{1 - q^{-1}}.$$
The input u_{obs}, with length $N_{\text{obs}} = 500$, and the reference r are zero-mean Gaussian noise signals with $\sigma_u = 1$ and

[3] Matlab code for reproducing numerical examples is available online `https://github.com/minaf/hammerstein_control`.

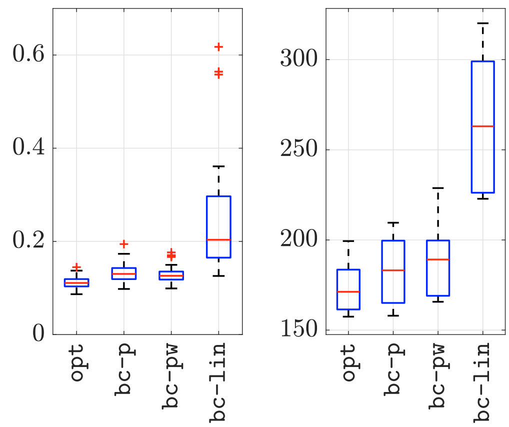

Fig. 3. Comparision of the performance using the cost function (3) for: i) the first case (left), ii) the second case (right).

$\sigma_r = 50$. The output is disturbed by zero-mean Gaussian noise with $\sigma = 10$ uncorrelated with the input signal. 50 Monte Carlo simulations are performed with different noise realizations. The results are presented on the right side of Figure 3. As in the first case, we observe that `bc-p` and `bc-pw` are comparable to the optimal controller.

8. CONCLUSION

In this paper, we have presented a data driven robust control design method based on an average risk criterion which we call Bayes control. We have shown that it has very close ties to the Bayesian kernel-based method proposed in Scampicchio et al. (2019), the conceptual difference lies in the use of a deterministic respective stochastic setting for the system. We find it reassuring that this approach can be motivated by different arguments. Details in the implementations differ as well. In Scampicchio et al. (2019) a classic approximation is used so that a convex optimization problem is obtained whereas here we use importance sampling to approximately compute the cost function which is minimized using gradient based search.

We have further extended on the scope of this approach. In Scampicchio et al. (2019) the case of model reference control for linear systems was treated and here we take the step to model reference control for Hammerstein systems.

We believe this to be a promising data driven control design method having many avenues of future research in terms of different systems, models and design criteria. For Hammerstein systems the next step is to consider non-parametric modeling of the input nonlinearity using a Gaussian process model with a suitably chosen kernel.

REFERENCES

Anderson, B.D. and Liu, Y. (1989). Controller reduction: concepts and approaches. *IEEE Transactions on Automatic Control*, 34(8), 645–664.

Åström, K.J. and Murray, R.M. (2010). *Feedback systems*. Princeton University Press.

Bai, E.W. (2004). Decoupling the linear and nonlinear parts in Hammerstein model identification. *Automatica*, 40(4), 671–676.

Bloemen, H., Van den Boom, T., and Verbruggen, H. (2000). Model-based predictive control for Hammerstein systems. In *IEEE Conference on Decision and Control*, 4963–4968.

Boutayeb, M., Aubry, D., and Darouach, M. (1996). A robust and recursive identification method for MISO Hammerstein model. In *UKACC International Conference on Control*, 234–239.

Campi, M.C., Lecchini, A., and Savaresi, S.M. (2002). Virtual reference feedback tuning: a direct method for the design of feedback controllers. *Automatica*, 38(8), 1337–1346.

Chen, T., Ohlsson, H., and Ljung, L. (2012). On the estimation of transfer functions, regularizations and Gaussian processes—Revisited. *Automatica*, 48(8), 1525–1535.

Dehghani, A., Lanzon, A., and Anderson, B.D. (2006). $\mathcal{H}_\infty$ design to generalize internal model control. *Automatica*, 42(11), 1959–1968.

Goethals, I., Pelckmans, K., Suykens, J.A., and De Moor, B. (2005). Subspace identification of Hammerstein systems using least squares support vector machines. *IEEE Transactions on Automatic Control*, 50(10), 1509–1519.

Guardabassi, G.O. and Savaresi, S.M. (2000). Virtual reference direct design method: an off-line approach to data-based control system design. *IEEE Transactions on Automatic Control*, 45(5), 954–959.

Hjalmarsson, H. (2002). Iterative feedback tuning—an overview. *International Journal of Adaptive Control and Signal Processing*, 16(5), 373–395.

Hjalmarsson, H., Gevers, M., Gunnarsson, S., and Lequin, O. (1998). Iterative feedback tuning: theory and applications. *IEEE Control Systems Magazine*, 18(4), 26–41.

Ingram, G.A., Franchek, M.A., Balakrishnan, V., and Surnilla, G. (2005). Robust SISO $\mathcal{H}_\infty$ controller design for nonlinear systems. *Control Engineering Practice*, 13(11), 1413–1423.

Karimi, A., Mišković, L., and Bonvin, D. (2004). Iterative correlation-based controller tuning. *Int. Journal of Adaptive Control and Signal Processing*, 18(8), 645–664.

Karimi, A., Van Heusden, K., and Bonvin, D. (2007). Non-iterative data-driven controller tuning using the correlation approach. In *European Control Conference*, 5189–5195.

Landau, I., Rey, D., Karimi, A., Voda, A., and Franco, A. (1995). A flexible transmission system as a benchmark for robust digital control. *European Journal of Control*, 1(2), 77–96.

Maritz, J.S. and Lwin, T. (1989). *Empirical Bayes methods*. Chapman and Hall London.

Pillonetto, G. and De Nicolao, G. (2010). A new kernel-based approach for linear system identification. *Automatica*, 46(1), 81–93.

Risuleo, R.S., Bottegal, G., and Hjalmarsson, H. (2015). A kernel-based approach to Hammerstein system identification. *IFAC Symp. System Identification*, 48, 1011–1016.

Risuleo, R.S., Bottegal, G., and Hjalmarsson, H. (2017). A nonparametric kernel-based approach to Hammerstein system identification. *Automatica*, 85, 234–247.

Scampicchio, A., Chiuso, A., Formentin, S., and Pillonetto, G. (2019). Bayesian kernel-based linear control design. In *IEEE Conference on Decision and Control*, 822–827.

Stoica, P. and Söderström, T. (1982). Instrumental-variable methods for identification of Hammerstein systems. *International Journal of Control*, 35(3), 459–476.