Direct Bayesian Identification of Inverse Linear Systems

Suzuki, Rikuto; Oomen, Tom; González, Rodrigo A.

DOI
10.1109/LCSYS.2025.3578480

Publication date
2025

Document Version
Final published version

Published in
IEEE Control Systems Letters

Citation (APA)
Suzuki, R., Oomen, T., & González, R. A. (2025). Direct Bayesian Identification of Inverse Linear Systems. *IEEE Control Systems Letters*, *9*, 1478-1483. https://doi.org/10.1109/LCSYS.2025.3578480

Direct Bayesian Identification of Inverse Linear Systems

Rikuto Suzuki, Tom Oomen, *Senior Member, IEEE*, and Rodrigo A. González

***Abstract*—The kernel-based inverse system identification framework enables accurate identification of systems with non-minimum phase dynamics, greatly expanding the potential of non-causal system identification approaches. The existing kernel-based inverse system identification method performs the estimation assuming noisy input data, while in practice noise is typically present only in the output measurements. To address this impracticality, we propose a Bayesian identification method that employs the Expectation-Maximization algorithm and the Markov chain Monte Carlo method to enable direct identification of the inverse system using the available data. Through numerical simulations, we found that the proposed method allows for an accurate estimation of inverse models, and outperforms an indirect approach in both model fit and variance. The proposed method can be used to develop enhanced data-driven feedforward control methods that allow for flexible design while incorporating design specifications.**

***Index Terms*—Identification, inverse systems, identification for control.**

I. Introduction

FEEDFORWARD control is essential in motion systems to effectively compensate for known disturbances and enhance reference tracking [1]. Since the optimal feedforward controller is given by the inverse of the true plant, substantial effort has been dedicated to deriving the inverse system dynamics through first principle modeling, grey box model-based approaches, or data-driven learning of the control law.

Data-driven feedforward control has been explored through iterative approaches that learn the user-specified model parameters, e.g., see [2] and [3]. Since a poor choice of model order can lead to large model errors, these approaches have been further extended towards non-parametric modeling to cope with more complex dynamics including unknown nonlinearities [4], parameter-varying systems [5], and state variable dependent disturbances [6]. In these data-driven approaches, system identification methods that estimate a model that captures the system dynamics are incorporated in the form of inverse system identification [7].

From a system identification perspective, recent studies on inverse model estimation suggest that estimating an inverse model directly from data may pose advantages over first estimating a forward model and subsequently inverting it [2], [8]. This is because the accuracy of the latter two-step approach is highly dependent on both the quality of the forward model and the accuracy of the model inversion [9]. To address these challenges, direct inverse system identification methods have been explored, based on kernel-based techniques [10], [11]. These methods effectively incorporate various system properties and balance the bias-variance tradeoff, especially in small-sample scenarios. From a Bayesian perspective, prior distributions allow for the effective integration of prior knowledge into the identification process [12], [13]. For instance, in the inverse system identification method developed in [14], [15], unstable inverse systems due to the non-minimum phase dynamics of the forward systems are viewed as non-causal and bounded operators. This approach has demonstrated that integrating such type of prior knowledge of the inverse system further enhances the accuracy of the inverse model. Additionally, kernel-based methods have been implemented in the form of Gaussian process-based feedforward control [4], [5], [6]. However, these methods require the optimal feedforward signal beforehand to generate a mapping from the reference to the corresponding optimal feedforward signal, and the optimality from the perspective of system identification such as consistency, e.g., see [16], is generally overlooked. Furthermore, the kernel-based inverse system identification method presented in [14], which leads to analytical solutions, may lack practicality in its formulation, as it is assumed that the input of the inverse system, i.e., the output signal, is perfectly known. Typically the opposite situation occurs, where the data is collected from the forward system and its output is known up to noisy measurements.

In contrast to the identification problem formulation in [14], inverse system identification using noisy output data of the forward system is computationally challenging. The main difficulty is due to the fact that the system inversion entails a transformation of random variables between forward and inverse system parameters, making it impossible to express the probabilistic relationship between the inverse system and observed data in a closed-form analytical expression. Addressing this intractability problem and enabling the direct identification of the inverse system in a practical manner is expected to further enhance the accuracy of the inverse model. This, in turn, introduces a new perspective on inverse

Received 17 March 2025; revised 14 May 2025; accepted 3 June 2025. Date of publication 10 June 2025; date of current version 11 July 2025. Recommended by Senior Editor A. P. Aguiar. *(Corresponding author: Rikuto Suzuki.)*

Rikuto Suzuki is with the Department of Electrical Engineering and Information Systems, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan (e-mail: r.suzuki@ctl.t.u-tokyo.ac.jp).

Tom Oomen is with the Department of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, and also with the Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, Delft, 2628 CN, The Netherlands.

Rodrigo A. González is with the Department of Mechanical Engineering, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

Digital Object Identifier 10.1109/LCSYS.2025.3578480

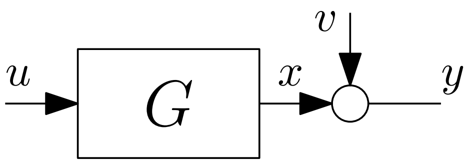

Fig. 1. Open-loop system with output measurement noise.

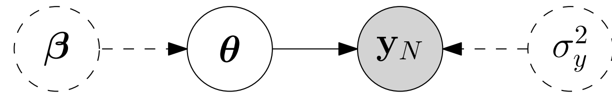

Fig. 2. Bayesian network considered in the proposed inverse system identification method. Each solid node represents a random variable, and solid links express probabilistic relationships between these variables. The dotted nodes represent hyperparameters specified for the group of random variables connected by the dotted links. Node $\mathbf{y}_N$ is shaded to indicate that such node corresponds to the observed data. For background on the Bayesian network, see [20].

model estimation methods and contributes to improving the marginal performance of data-driven feedforward control.

In this letter, a direct Bayesian identification method for inverse systems in the forward open-loop setting is developed, where the intractability of the hyperparameter optimization problem is addressed via Markov chain Monte Carlo (MCMC) techniques [17] and the Expectation Maximization (EM) algorithm [18]. The main contributions of this letter are as follows.

- C1 A kernel-based identification approach that allows for the direct identification of the inverse system from the data collected from a forward system is introduced. To this end, a procedure to optimize the hyperparameters associated with the inverse model via the Expectation-Maximization (EM) algorithm in conjunction with MCMC techniques is formalized to solve the computational intractability of the E-step.
- C2 The performance of this estimator is compared against 1) the kernel-based approach in [14], and 2) the traditional two-step approach, in which the forward system is identified first and subsequently its inverse is computed.

This letter is organized as follows. In Section II the system setup and problem statement are described. In Section III we derive the proposed direct identification method for inverse linear systems. In Section IV, simulation results are presented, and we provide concluding remarks in Section V.

Notation: The shift operator q satisfies $qx(t) = x(t+1)$. All vectors and matrices are written in bold, and matrices are denoted with capital letters. $[\mathbf{A}]_{i,j}$ denotes the i, j element of a matrix $\mathbf{A}$, and single subscripts are used for vectors, e.g., $[\mathbf{x}]_i$. The probability density function (pdf) of a random variable $\mathbf{x}$ is denoted as $p_{\mathbf{x}}(\cdot;\cdot)$. Here, the known or deterministic terms are placed after the semicolon. or notational simplicity, we omit the explicit dependency on the input vector $\mathbf{u}_N$. For instance, the likelihood can be expressed as $p_{\mathbf{y}_N}(\mathbf{y}_N;\boldsymbol{\rho})$ instead of the more detailed form $p(\mathbf{y}_N;\mathbf{u}_N,\boldsymbol{\rho})$.

II. Preliminaries and Problem Formulation

A. System Description

Consider a single-input single-output (SISO), discrete-time, linear time-invariant (LTI), causal and asymptotically stable system $G(q)$ in Figure 1, with input signal $u(t)$, output signal $x(t)$, and impulse response $\xi(t)$. The measured output signal $y(t)$ is contaminated by i.i.d. zero mean Gaussian-distributed noise sequence $v(t)$ with variance σ_y^2, i.e., $v(t) \sim \mathcal{N}(0, \sigma_y^2)$, uncorrelated with the input signal $u(t)$. That is,

$$y(t) = \sum_{\tau=0}^{\infty} \xi(\tau)u(t-\tau) + v(t). \tag{1}$$

B. Non-Causal Impulse Response Interpretation

Within the context of feedforward control, our interest is centered in the inverse relationship between the input and output sequences in (1). An inverse system is non-causal if the corresponding forward system is strictly proper, i.e., if the relative degree is greater than zero, or if it has a non-minimum phase zero that becomes an unstable pole when inverted. In the standard (causal) interpretation, the presence of poles outside the unit disk on the complex plane indicates that the associated impulse response of the system is unbounded, often calling the system unstable. On the other hand, a non-causal interpretation of the system is also possible by selecting a different region of convergence for the z-transform, which is used to calculate the impulse response [14, Th. 1].

Theorem 1 (Non-Causal Exact Inversion [14, Th. 1]): Let a system $G(z)$ be given such that $G^{-1}(z)$ belongs to the set of real, rational, discrete-time systems without poles on the unit circle $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$. Then, there exists a non-causal sequence $\boldsymbol{\theta} \in \ell_1(\mathbb{Z})$ such that, for any signal $x(t) \in \ell_2(\mathbb{Z})$, the signal

$$u(t) = \sum_{\tau=-\infty}^{\infty} \theta(\tau)x(t-\tau) \in \ell_2(\mathbb{Z}) \tag{2}$$

leads to exact inversion $x(t) = G(q)u(t)$.

This theorem is particularly useful for identifying unstable systems [19] and facilitates inverse model estimation when the inverse system is unstable due to the non-minimum phase dynamics of the forward system. By interpreting the unstable inverse system as a non-causal but bounded operator, a kernel-based approach can incorporate this property through an appropriate choice of kernel [14].

C. Problem Formulation

Based on the forward open-loop setting in Figure 1 and the available data $\{u(t), y(t)\}_{t=1}^N$, we aim to develop a direct Bayesian identification method for the inverse system $G^{-1}(q)$, incorporating prior information directly into the identification problem.

III. Direct Bayesian Identification of Inverse Linear Systems

In this section, a direct Bayesian identification method for inverse linear systems is developed. The forward impulse response $\boldsymbol{\xi}$ is viewed as a function of the impulse response of the inverse system, $\boldsymbol{\theta}$, i.e., $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\theta})$, which yields the probabilistic relationship shown in Figure 2 to perform Bayesian estimation.

A. Bayesian Approach to Inverse System Identification

The Bayesian approach uses the posterior distribution as a measure for quantifying all the information that can be extracted from the observed data combined with the prior information [17]. Using Bayes' rule, the posterior distribution of the impulse response of the inverse model given the available data can be written as

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{y}_N;\boldsymbol{\rho}) = \frac{p_{\mathbf{y}_N}\left(\mathbf{y}_N|\boldsymbol{\theta};\sigma_y^2\right)p_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\beta})}{\int p(\mathbf{y}_N,\boldsymbol{\theta};\boldsymbol{\rho})\mathrm{d}\boldsymbol{\theta}}, \tag{3}$$

where $p_{\mathbf{y}_N}(\mathbf{y}_N|\boldsymbol{\theta};\sigma_y^2)$ is a likelihood function of the measured output vector $\mathbf{y}_N = [y(1),\dots,y(N)]^\top$ given the inverse impulse response $\boldsymbol{\theta}$, $p_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\rho})$ denotes the prior distribution of $\boldsymbol{\theta}$, and the denominator term is the normalizing factor of the pdf. These distributions may depend on parameters $\boldsymbol{\rho} = [\sigma_y^2, \boldsymbol{\beta}^\top]$, known as hyperparameters, which define the probabilistic properties of the random variables within the associated user-defined probabilistic structure. The maximum a posteriori estimate of the impulse response can be obtained via maximization of the log of the posterior distribution, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log p_{\mathbf{y}_N}\left(\mathbf{y}_N|\boldsymbol{\theta};\sigma_y^2\right) + \log p_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\beta}). \tag{4}$$

Let us consider a finite impulse response (FIR) model estimate $\hat{\boldsymbol{\theta}} \in \mathbb{R}^{n_\theta}$ and a prior distribution described by

$$\boldsymbol{\theta} \sim \mathcal{N}_{\boldsymbol{\theta}}(\mathbf{0}, \mathbf{K}_{\boldsymbol{\theta}}(\boldsymbol{\beta})), \tag{5}$$

with the covariance matrix $\mathbf{K}_{\boldsymbol{\theta}}(\boldsymbol{\beta}) \in \mathbb{R}^{n_\theta \times n_\theta}$. Then, under the assumption that the noise is Gaussian, the likelihood function $p_{\mathbf{y}_N}(\mathbf{y}_N|\boldsymbol{\theta};\sigma_y^2)$ admits the following Gaussian description

$$\mathbf{y}_N|\boldsymbol{\theta} \sim \mathcal{N}_{\mathbf{y}_N}\left(\boldsymbol{\Phi}_u^\top \boldsymbol{\xi}(\boldsymbol{\theta}), \sigma_y^2\mathbf{I}_N\right), \tag{6}$$

where the lower triangular Toeplitz matrix $\boldsymbol{\Phi}_u \in \mathbb{R}^{N\times n_\xi}$ that considers the order n_ξ of the forward FIR $\boldsymbol{\xi}(\boldsymbol{\theta})$ has as entries

$$[\boldsymbol{\Phi}_u]_{i,j} = \begin{cases} u(i-j+1) & 0 < i-j+1 \le n_\xi \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Thus, the optimization problem in (4) can be written as

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{N}\left(y(t) - \sum_{\tau=0}^{N}\xi(\theta,\tau)u(t-\tau)\right)^2 + \sigma_y^2\theta^\top \mathbf{K}_\theta^{-1}(\beta)\theta. \tag{8}$$

As in standard kernel-based methods for forward system identification, desired model properties such as smoothness and stability can be enforced by structuring the covariance matrix $\mathbf{K}_{\boldsymbol{\theta}}(\boldsymbol{\beta})$ through carefully chosen kernels. These kernels constrain the estimate to an induced reproducing kernel Hilbert space. To impose smoothness and exponential decay in the non-causal impulse response, the non-causal stable spline kernel [14] can be used, given by

$$k\left(t,t'\right) = \alpha\ \frac{\min\left(b(t),b(t')\right)^2}{6} \times \left(3\max(b(t),b(t')) - \min(b(t),b(t'))\right), \tag{9}$$

where

$$b(t) = \begin{cases} \lambda_{\mathrm{nc}}^{-t} & \text{if } t<0 \\ \lambda_{\mathrm{c}}^{t} & \text{if } t \ge 0, \end{cases} \tag{10}$$

with the hyperparameters $\boldsymbol{\beta} = [\alpha, \lambda_{\mathrm{c}}, \lambda_{\mathrm{nc}}]^\top$ for the kernel, in which $\alpha > 0$ is a scaling factor and $\lambda_{\mathrm{c}}, \lambda_{\mathrm{nc}} \in (0,1)$ denote decay rates for $t \ge 0$ and $t < 0$ respectively. We use this probabilistic interpretation of kernels to specify the prior distribution and to impose prior information into the identification problem. For more information on kernel-based methods and their relation to Bayesian estimation, we refer to [10], [13], [21], [22].

B. Hyperparameter Estimation Using the EM Algorithm

The hyperparameters $\boldsymbol{\rho}$ used to specify the prior distribution and noise variance need to be estimated to solve the optimization problem in (8). We consider the maximization of the marginal likelihood function of the measured output $\mathbf{y}_N$, i.e.,

$$\hat{\boldsymbol{\rho}} = \arg\max_{\boldsymbol{\rho}} \log p_{\mathbf{y}_N}(\mathbf{y}_N;\boldsymbol{\rho}), \tag{11}$$

which is known as the Empirical Bayes method [23]. This marginal likelihood function does not admit a closed-form expression because the function depends on the impulse response $\boldsymbol{\theta}$, which we cannot directly observe from the data. To deal with this issue, we incorporate these latent variables by marginalizing over $\boldsymbol{\theta}$, i.e.,

$$\hat{\boldsymbol{\rho}} = \arg\max_{\boldsymbol{\rho}} \log \int p(\mathbf{y}_N,\boldsymbol{\theta};\boldsymbol{\rho})\mathrm{d}\boldsymbol{\theta}. \tag{12}$$

The maximum likelihood problem above can be solved using the EM algorithm [18], which iteratively refines parameter estimates by alternating between an expectation step (E-step) and a maximization step (M-step). This iterative procedure guarantees monotonic convergence to a stationary point of the marginal likelihood function in (12), which often corresponds to the global optimum if the initialization is appropriate [24]. A straightforward and common approach for initialization, used in our simulation, is a grid search over several initialization candidates and retain the one that yields the maximum value of the cost function [25]. The E and M steps are described as follows.

E-step: Compute the conditional expectation of the log-likelihood of the joint distribution with respect to the posterior joint distribution given estimates $\hat{\boldsymbol{\rho}}^{(k)}$

$$Q\left(\boldsymbol{\rho};\hat{\boldsymbol{\rho}}^{(k)}\right) = \int \log p(\mathbf{y}_N,\boldsymbol{\theta};\boldsymbol{\rho})p_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}|\mathbf{y}_N;\hat{\boldsymbol{\rho}}^{(k)}\right)\mathrm{d}\boldsymbol{\theta}. \tag{13}$$

M-step: Update the estimates $\hat{\boldsymbol{\rho}}^{(k+1)}$

$$\hat{\boldsymbol{\rho}}^{(k+1)} = \arg\max_{\boldsymbol{\rho}} Q\left(\boldsymbol{\rho};\hat{\boldsymbol{\rho}}^{(k)}\right). \tag{14}$$

$$\begin{aligned} Q\left(\boldsymbol{\rho};\hat{\boldsymbol{\rho}}^{(k)}\right) &= \int \log p_{\mathbf{y}_N}\left(\mathbf{y}_N|\boldsymbol{\theta};\sigma_y^2\right)p_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}|\mathbf{y}_N;\hat{\boldsymbol{\rho}}^{(k)}\right)\mathrm{d}\boldsymbol{\theta} + \int \log p_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\beta})p_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}|\mathbf{y}_N;\hat{\boldsymbol{\rho}}^{(k)}\right)\mathrm{d}\boldsymbol{\theta} \\ &= -\frac{1}{2\sigma_y^2}\mathbb{E}_{\boldsymbol{\theta}}\{||\mathbf{y}_N - \boldsymbol{\Phi}_u^\top\boldsymbol{\xi}(\boldsymbol{\theta})||^2 \mid \mathbf{y}_N;\hat{\boldsymbol{\rho}}^{(k)}\} - \frac{1}{2}\mathrm{tr}\{\mathbf{K}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\beta})\mathbb{E}_{\boldsymbol{\theta}}\{\boldsymbol{\theta}\boldsymbol{\theta}^\top \mid \mathbf{y}_N;\hat{\boldsymbol{\rho}}^{(k)}\}\} - \frac{N}{2}\log\sigma_y^2 \\ &\quad -\frac{1}{2}\log\det\mathbf{K}_{\boldsymbol{\theta}}(\boldsymbol{\beta}) - N\log 2\pi \end{aligned} \tag{15}$$

The conditional expectation $Q(\boldsymbol{\rho}; \hat{\boldsymbol{\rho}}^{(k)})$ in (13) is taken with respect to the posterior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{y}_N; \hat{\boldsymbol{\rho}}^{(k)})$, which can be rewritten as (15), shown at the bottom of the next page. Thus, in the M-step, the hyperparameter estimate $\hat{\boldsymbol{\rho}}^{k+1}$ can be updated by maximizing the conditional expectation $Q(\boldsymbol{\rho}; \hat{\boldsymbol{\rho}}^{(k)})$ with respect to each hyperparameter. This can be done by solving the optimization problem given by

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \arg\min_{\boldsymbol{\beta}} \frac{1}{2}\mathrm{tr}\left\{\mathbf{K}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\beta})\mathbb{E}_{\boldsymbol{\theta}}\{\boldsymbol{\theta}\boldsymbol{\theta}^\top \mid \mathbf{y}_N; \hat{\boldsymbol{\rho}}^{(k)}\}\right\} + \log\det \mathbf{K}_{\boldsymbol{\theta}}(\boldsymbol{\beta}), \quad (16)$$

and the update of the noise variance $\sigma_y^{2^{k+1}}$

$$\hat{\sigma}_y^{2^{(k+1)}} = \frac{\mathbb{E}_{\boldsymbol{\theta}}\{||\mathbf{y}_N - \boldsymbol{\Phi}_u^\top \boldsymbol{\xi}(\boldsymbol{\theta})||^2 \mid \mathbf{y}_N; \hat{\boldsymbol{\rho}}^{(k)}\}}{N}. \quad (17)$$

In order to compute the conditional expectations in (16) and (17), the prior on the inverse FIR $\boldsymbol{\theta}$ must be transformed to the prior on the forward FIR $\boldsymbol{\xi}$. However, since this transformation of random variables involves a nonlinear transformation and numerical computation of poles of the corresponding forward model, the posterior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{y}_N; \hat{\boldsymbol{\rho}}^{(k)})$ and hence the conditional expectation $Q(\boldsymbol{\rho}; \hat{\boldsymbol{\rho}}^{(k)})$ does not have an analytic expression and must be approximated. In the next section, an approach used to approximate the posterior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{y}_N; \hat{\boldsymbol{\rho}}^{(k)})$ is explained, which permits the computation of (16) and (17).

C. Approximating the Posterior Distribution Using the Metropolis-Hastings (MH) Algorithm

The maximum of $Q(\boldsymbol{\rho}; \hat{\boldsymbol{\rho}}^{(k)})$ in (13) depends on conditional expectations of functions of $\boldsymbol{\theta}$, which are not computable in closed form. To address this, we approximate these expectations by sampling from the posterior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{y}_N; \hat{\boldsymbol{\rho}}^{(k)})$. As seen in (3), this posterior is proportional to the product of two Gaussian distributions, making it suitable for sampling via the Metropolis-Hastings (MH) algorithm, a Markov Chain Monte Carlo (MCMC) method [17], [26]. The MH algorithm consists of sampling and evaluation steps; at each iteration j, the algorithm uses a user-defined Gaussian proposal distribution $q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}}^{(j)})$ centered at a previous sample $\bar{\boldsymbol{\theta}}^{(j)}$ to draw a next sample $\bar{\boldsymbol{\theta}}^{\star}$. The drawn sample is compared to the previous sample by computing the acceptance probability, which is used to decide whether the sample is accepted or rejected. This value is given by

$$\alpha = \min\left(1, \frac{\tilde{\gamma}(\bar{\boldsymbol{\theta}}^{\star})}{\tilde{\gamma}(\bar{\boldsymbol{\theta}}^{(j)})}\frac{q(\bar{\boldsymbol{\theta}}^{(j)}|\bar{\boldsymbol{\theta}}^{\star})}{q(\bar{\boldsymbol{\theta}}^{\star}|\bar{\boldsymbol{\theta}}^{(j)})}\right), \quad (18)$$

where $\tilde{\gamma}(\bar{\boldsymbol{\theta}}^{(j)})$ is the unnormalized posterior distribution, i.e.,

$$\tilde{\gamma}\left(\bar{\boldsymbol{\theta}}^{(j)}\right) = p_{\mathbf{y}_N}\left(\mathbf{y}_N|\bar{\boldsymbol{\theta}}^{(j)}; \hat{\boldsymbol{\rho}}^{(k)}\right)p_{\boldsymbol{\theta}}\left(\bar{\boldsymbol{\theta}}^{(j)}; \hat{\boldsymbol{\rho}}^{(k)}\right). \quad (19)$$

The MH algorithm requires several iterations before the obtained samples converge to represent the target posterior distribution. The period until convergence is called the burn-in period, and samples $\{\bar{\boldsymbol{\theta}}^{(j)}: j \in \{1, \ldots, M_{\text{burn-in}}\}\}$ obtained during this period $M_{\text{burn-in}}$ must be discarded. The length of the burn-in period depends on the initial conditions and the dimensions of the target posterior distribution, and it needs to be adjusted by trial and error.

Algorithm 1 Direct Bayesian Inverse System Identification

Require: Input $\mathbf{u}_N$, measured output $\mathbf{y}_N$, initial hyperparameter estimate $\hat{\boldsymbol{\rho}}^{(0)} = [\hat{\boldsymbol{\beta}}^{(0)^\top}, \hat{\sigma}_y^{2^{(0)}}]^\top$, initial sample $\bar{\boldsymbol{\theta}}^{(1)}$ of the impulse response, proposal distribution $q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}}^{(j)})$, unnormalized target distribution $\tilde{\gamma}(\bar{\boldsymbol{\theta}}^{(j)}) = p_{\mathbf{y}_N}(\mathbf{y}_N|\bar{\boldsymbol{\theta}}^{(j)}; \hat{\boldsymbol{\rho}}^{(k)})p_{\boldsymbol{\theta}}(\bar{\boldsymbol{\theta}}^{(j)}; \hat{\boldsymbol{\rho}}^{(k)})$, number of samples M_{MH} to draw and burn-in period $M_{\text{burn-in}}$ in the MH algorithm, maximum number of EM iteration M_{EM}

for $k = 1, 2, \ldots, M_{\text{EM}}$ **do**
  for $j = 1, 2, \ldots, M_{\text{MH}}$ **do**
    Draw a sample $\bar{\boldsymbol{\theta}}^{\star}$ from $q(\boldsymbol{\theta}|\bar{\boldsymbol{\theta}}^{(j)})$
    Compute α from (18)
    Sample b from a uniform distribution on [0, 1]
    Accept the sample $\bar{\boldsymbol{\theta}}^{\star}$, i.e., $\bar{\boldsymbol{\theta}}^{(j+1)} = \bar{\boldsymbol{\theta}}^{\star}$, if $b \leq \alpha$, otherwise reject, i.e., $\bar{\boldsymbol{\theta}}^{(j+1)} = \bar{\boldsymbol{\theta}}^{(j)}$
  end for
  Compute $\hat{\boldsymbol{\rho}}^{(k)}$ from (16), (17)
  Update $\bar{\boldsymbol{\theta}}^{(1)}$ using (20)
end for
Compute $\hat{\boldsymbol{\theta}}$ from (8).

Ensure: Estimated FIR $\hat{\boldsymbol{\theta}}$ of the inverse system.

Given a sufficiently large number of samples $\{\bar{\boldsymbol{\theta}}^{(j)}: j \in \{M_{\text{burn-in}} + 1, \ldots, M_{\text{MH}}\}\}$ drawn from the approximated posterior distribution $p_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{y}_N; \hat{\boldsymbol{\rho}}^{(k)})$, the conditional expectation for the E-step in (13) of the EM algorithm can be executed by approximating the expectation terms:

$$\mathbb{E}_{\boldsymbol{\theta}}\{\mathbf{f}(\boldsymbol{\theta}) \mid \mathbf{y}_N; \hat{\boldsymbol{\rho}}^{(k)}\} \approx \frac{1}{M}\sum_{j=1}^{M}\mathbf{f}\left(\bar{\boldsymbol{\theta}}^{(j)}\right), \quad (20)$$

where $\mathbf{f}(\cdot)$ is a function of $\boldsymbol{\theta}$, given by, e.g., the arguments of the expectations in (16) and (17). With an appropriate choice of proposal distribution, the MH algorithm yields sample means of the form (20) that converge to the true expectations as $M \to \infty$ by the law of large numbers, under mild conditions (see, e.g., [17]). We summarize the proposed identification procedure in Algorithm 1.

IV. Simulation Example

In this section, the proposed direct Bayesian identification method for inverse systems is validated in simulation. The considered data-generating system is described by $y(t) = G(q)u(t) + v(t)$ as depicted in Figure 1, with

$$G(z) = \frac{0.004(z-0.75)(z-0.6)\left(z^2 - 30z + 225.9\right)}{\left(z^2 - 1.2z + 0.4\right)\left(z^2 - 1.5z + 0.5725\right)}. \quad (21)$$

The true system $G(z)$ is asymptotically stable but has a conjugate pair of non-minimum phase zeros at $z = 15 \pm i0.96$, thus its inverse system is unstable. The corresponding forward and inverse impulse responses are shown in Figure 3. The impulse response of the inverse system quickly decays to 0, while that of the forward system decays slower. This implies that fewer parameters are required to estimate the FIR model of the inverse system than that of the forward system. We generate 100 datasets $\{u(t), y(t)\}_{t=1}^{N}$, where in each dataset, the input $u(t)$ and output noise $v(t)$ are uncorrelated, i.i.d., zero-mean, normally distributed sequences of length $N = 60$. The output signal-to-noise ratio is 14dB.

A. Existing Approaches

In addition to the proposed method, we implement and compare two alternative feasible approaches. Firstly, the regularized least-squares approach [14] yields an inverse FIR model estimate $\hat{\boldsymbol{\theta}}$ via

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{N} \left(u(t) - \sum_{\tau=-m_{\mathrm{nc}}}^{m_{\mathrm{c}}-1} \theta(\tau) y(t-\tau) \right)^2 + \gamma \boldsymbol{\theta}^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{\beta}) \boldsymbol{\theta}, \quad (22)$$

where γ denotes a regularization constant, m_{nc} and m_{c} denote the number of non-causal and causal parameters, and $\mathbf{K}_{\boldsymbol{\theta}}(\boldsymbol{\beta})$ is a kernel or regularization matrix used to incorporate desired model properties into the optimization problem. In addition to the regularization-induced bias, the method in [14] introduces additional bias, as the identification framework cannot be tailored to the correct noise statistics when output noise is considered [27, Ch. 3].

An alternative to the aforementioned approach is the traditional two-step method, where the forward model is first estimated and then inverted to obtain the inverse model. Using the kernel-based approach, a consistent estimator can be proposed under mild conditions [28], leading to a forward FIR model estimate $\hat{\boldsymbol{\xi}}$ obtained via the optimization problem

$$\hat{\boldsymbol{\xi}} = \arg\min_{\boldsymbol{\xi}} \sum_{t=1}^{N} \left(y(t) - \sum_{\tau=0}^{m_{\mathrm{f}}-1} \xi(\tau) u(t-\tau) \right)^2 + \gamma \boldsymbol{\xi}^\top \mathbf{K}_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\beta}) \boldsymbol{\xi}. \quad (23)$$

However, since the inversion is performed outside the optimization process, the accuracy of the inverse model for finite samples is not guaranteed. Moreover, prior information about the inverse system cannot be directly incorporated into this identification framework, which could otherwise improve the accuracy of the inverse model [14]. All kernel based identification methods used in this study exhibit a bias variance tradeoff that depends on the number of estimated parameters: increasing the number of coefficients reduces bias but leads to higher variance in the resulting model [29]. Furthermore, the number of estimated parameters needs to be chosen in such a way that the resulting model is capable of capturing the dominant system dynamics.

B. Experimental Settings of Compared Identification Methods

The non-causal inverse system is identified using the FIR model structure given by

$$\widehat{G^{-1}}(z) = \sum_{\tau=-m_{\mathrm{nc}}}^{m_{\mathrm{c}}-1} \hat{\theta}(\tau) z^{-\tau}. \quad (24)$$

We estimate the model using 4 different methods, whose specifications are detailed below.

- Proposal: The proposed direct Bayesian inverse system identification method with numbers of causal and anti-causal parameters $m_{\mathrm{c}} = 2, m_{\mathrm{nc}} = 4$, respectively, for the inverse model with the non-causal stable spline kernel introduced in [14]. A total of 1×10^5 samples are drawn in the MCMC algorithm with 1×10^4 samples for the burn-in period. The magnitude of the covariance of the proposal distribution is tuned so that the acceptance rate is in between 25% and 35%.

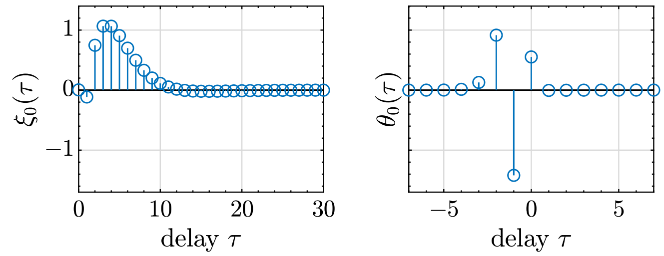

Fig. 3. The true impulse responses of the forward system $G(z)$ (left figure) and the inverse system $G^{-1}(z)$ (right figure) in (21).

- Naive: Implementation of the kernel-based approach in [14] with the cost function in (22). The number of causal and anti-causal parameters is given by $m_{\mathrm{c}} = 2$ and $m_{\mathrm{nc}} = 4$ respectively, and the non-causal stable spline kernel is considered.
- Two-step: The two-step approach in which the forward system is identified via (23) first with a number of parameters $m_{\mathrm{f}} = 25$ using the stable spline kernel [30] and subsequently the corresponding inverse system is computed using Theorem 1 with number of causal and anti-causal parameters $m_{\mathrm{c}} = 2, m_{\mathrm{nc}} = 4$ for the inverse model.
- Oracle: The kernel-based identification of inverse systems as in [14] using the artificial data $\{x(t), \tilde{u}(t)\}_{t=1}^N$ where $\tilde{u}(t)$ is the noise-corrupted input signal, i.e., $\tilde{u}(t) = u(t) + w(t)$ with white noise signal $w(t)$.

Each identification method is initialized using the hyperparameters $\hat{\boldsymbol{\beta}}^{(0)}$ of the kernel corresponding to the minimum of the cost function in (13) among the hyperparameter candidates, given the variance $\hat{\sigma}_y^2$ of the output noise and the impulse response $\hat{\boldsymbol{\xi}}^{(0)}$ of the forward system estimated by the standard regularized least-squares method. To evaluate the methods, the following measure for performance is used:

$$\mathrm{Fit} = 100 \left(1 - \left[\frac{\sum_{t=1}^{N} |u(t) - \hat{u}(t)|^2}{\sum_{t=1}^{N} |u(t) - \bar{u}(t)|^2} \right]^{1/2} \right), \quad (25)$$

with

$$\hat{u}(t) = \sum_{\tau=-m_{\mathrm{nc}}}^{m_{\mathrm{c}}-1} \hat{\theta}(\tau) x(t-\tau), \quad \bar{u}(t) = \frac{1}{N} \sum_{t=1}^{N} u(t). \quad (26)$$

C. Simulation Results

The resulting box plots for the 100 fits corresponding to the proposed, and conventional two-step and naive method for the system in (21) are depicted in Figure 4, with average fits shown in Table I. While the conventional two-step approach estimates the forward model reasonably well (left box plot in Fig. 4), small forward model inaccuracies can lead to larger errors in the inverse model. The two-step method also suffers from high variance in the inverse model, despite the true inverse FIR being representable with fewer parameters. In contrast, the proposed method achieves lower variance and higher accuracy. These results show that the proposed method outperforms the conventional two-step approach, offering greater accuracy and stability for inverse model estimation at the cost of higher computation. While the proposed method requires a greater computational cost, taking approximately one hour per model compared to a few seconds for the conventional approach, both are performed offline, making this cost acceptable in many applications.

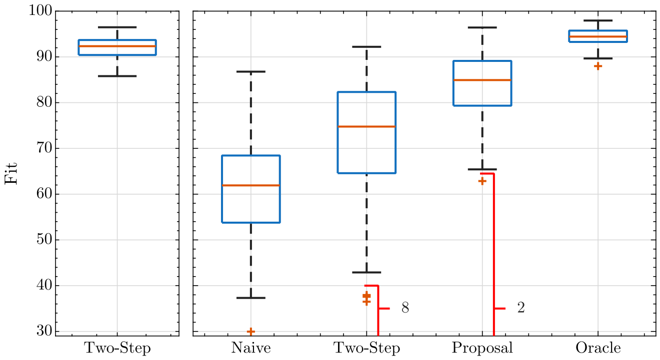

Fig. 4. Boxplots of the 100 fits for the two-step forward models (left figure), and inverse models (right figure) for the system in (21).

TABLE I
AVERAGE FIT (25) OF THE 100 FITS EXCLUDING THE OUTLIERS FOR THE INVERSE MODELS FOR THE SYSTEM IN (21)

Naive	Two-Step	Proposal	Oracle
60.93	74.02	84.42	94.31

V. CONCLUSION

The direct Bayesian identification method for inverse systems enables the estimation of the inverse model directly from data while incorporating prior information about the inverse system. To estimate the hyperparameters associated with this prior, an Expectation-Maximization algorithm is employed to handle the intractability of the marginal likelihood function, alongside a Metropolis-Hastings algorithm that draws samples from the posterior distribution to approximate the impulse response of the inverse model. These algorithms are applied iteratively until the hyperparameters converge, after which the inverse model is estimated. The proposed method is validated through numerical simulations, demonstrating improved accuracy compared to the conventional two-step approach.

REFERENCES

[1] F. Boeren, T. Oomen, and M. Steinbuch, "Accuracy aspects in motion feedforward tuning," in *Proc. Amer. Control Conf.*, 2014, pp. 2178–2183.
[2] F. Boeren, T. Oomen, and M. Steinbuch, "Iterative motion Feedforward tuning: A data-driven approach based on instrumental variable identification," *Control Eng. Pract.*, vol. 37, pp. 11–19, Aug. 2015.
[3] F. Boeren, L. Blanken, D. Bruijnen, and T. Oomen, "Optimal estimation of rational feedforward control via instrumental variables: With application to a wafer stage," *Asian J. Control*, vol. 20, no. 3, pp. 975–992, 2018.
[4] M. Van Meer, M. Poot, J. Portegies, and T. Oomen, "Gaussian process based feedforward control for nonlinear systems with flexible tasks: With application to a printer with friction," *IFAC-PapersOnLine*, vol. 55, no. 37, pp. 241–246, 2022. [Online]. Available: www.sciencedirect.com
[5] M. van Haren, L. Blanken, and T. Oomen, "A kernel-based identification approach to LPV feedforward: With application to motion systems," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 6063–6068, Jul. 2023. [Online]. Available: https://research.tue.nl/en/publications/a-kernel-based-identification-approach-to-lpv-feedforward-with-ap
[6] R. Sasaki et al., "State variable dependent disturbances compensation using Gaussian process regression: With application to core-type linear motor," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics, AIM*, 2024, pp. 416–421.
[7] A. Uwadukunze, X. Bombois, M. Forgione, M. Gilson, and M. Albisser, "Identification for feedforward control of wiener systems," *IFAC-PapersOnLine*, vol. 58, no. 21, pp. 208–213, 2024.
[8] Y. Jung and M. Enqvist, "Estimating models of inverse systems," in *Proc. IEEE Conf. Decis. Control*, 2013, pp. 7143–7148.
[9] S. Devasia, "Should model-based inverse inputs be used as feedforward under plant uncertainty?" *IEEE Trans. Autom. Control*, vol. 47, no. 11, pp. 1865–1871, Nov. 2002.
[10] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.
[11] J. Lataire and T. Chen, "Transfer function and transient estimation by Gaussian process regression in the frequency domain," *Automatica*, vol. 72, pp. 217–229, Oct. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109816302357
[12] V. Peterka, "Bayesian system identification," *Automatica*, vol. 17, no. 1, pp. 41–53, 1981. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0005109881900832
[13] A. Chiuso, "Regularization and Bayesian learning in dynamical systems: Past, present and future," *Annu. Rev. Control*, vol. 41, pp. 24–38, Jun. 2016.
[14] L. Blanken and T. Oomen, "Kernel-based identification of non-causal systems with application to inverse model control," *Automatica*, vol. 114, Apr. 2020, Art. no. 108830.
[15] X. Fang and T. Chen, "On kernel design for regularized non-causal system identification," *Automatica*, vol. 159, Jan. 2024, Art. no. 111335. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0005109823005010
[16] L. Ljung, *System Identification: Theory for the User*. Hoboken, NJ, USA: Prentice-Hall, 1999. [Online]. Available: https://books.google.nl/books?id=nHFoQgAACAAJ
[17] B. Ninness and S. Henriksen, "Bayesian system identification via Markov chain monte carlo techniques," *Automatica*, vol. 46, no. 1, pp. 40–51, Jan. 2010.
[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. Ser. B, Methodological*, vol. 39, no. 1, pp. 1–22, 1977. [Online]. Available: http://dx.doi.org/10.1111/j.2517-6161.1977.tb01600.x
[19] Y. Fujimoto, "Kernel-based Regularization for unstable systems," in *Proc. IEEE 61st Conf. Decis. Control (CDC)*, 2022, pp. 209–214.
[20] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[21] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and Gaussian processes-revisited," *Automatica*, vol. 48, no. 8, pp. 1525–1535, 2012.
[22] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2005. [Online]. Available: https://doi.org/10.7551/mitpress/3206.001.0001
[23] J. S. Maritz and T. Lwin, *Empirical Bayes Methods With Applications*, 2nd ed. Boca Raton, FL, USA: Chapman and Hall/CRC, 1989. [Online]. Available: https://www.taylorfrancis.com/books/mono/10.1201/9781351071666/empirical-bayes-methods-applications-maritz
[24] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, 2nd ed. Hoboken, NJ, USA: Wiley-Intersci., 2008. [Online]. Available: http://www.123library.org/book_details/?id=21298
[25] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, 1983. [Online]. Available: http://www.jstor.org/stable/2240463
[26] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970. [Online]. Available: http://www.jstor.org/stable/2334940
[27] T. Söderström, *Errors-in-Variables Methods in System Identification*. Cham, Switzerland: Springer, 2018. [Online]. Available: http://link.springer.com/10.1007/978-3-319-75001-9
[28] Y. Ju, B. Mu, L. Ljung, and T. Chen, "Asymptotic theory for regularized system identification part I: Empirical Bayes hyperparameter estimator," *IEEE Trans. Autom. Control*, vol. 68, no. 12, pp. 7224–7239, Dec. 2023.
[29] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, *Regularized System Identification: Learning Dynamic Models from Data*. Cham, Switzerland: Springer Nat., 2022. [Online]. Available: https://library.oapen.org/bitstream/id/d12d395f-06ed-43c8-9234-3452fb8a611c/978-3-030-95860-2.pdf
[30] G. Pillonetto and G. De Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, pp. 81–93, Jan. 2010. [Online]. Available: www.elsevier.com/locate/automatica