Kernel-based identification using Lebesgue-sampled data

González, Rodrigo A.; Tiels, Koen; Oomen, Tom

DOI
10.1016/j.automatica.2024.111648

Publication date
2024

Document Version
Final published version

Published in
Automatica

Automatica 164 (2024) 111648

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Kernel-based identification using Lebesgue-sampled data☆

Rodrigo A. González [a,*], Koen Tiels [a], Tom Oomen [a,b]

[a] Control Systems Technology Section, Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[b] Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Article history:
Received 10 March 2023
Received in revised form 9 January 2024
Accepted 23 February 2024
Available online 1 April 2024

Keywords:
System identification
Event-based sampling
Kernel-based methods
Regularization
Impulse response estimation

ABSTRACT

Sampling in control applications is increasingly done non-equidistantly in time. This includes applications in motion control, networked control, resource-aware control, and event-based control. Some of these applications, like the ones where displacement is tracked using incremental encoders, are driven by signals that are only measured when their values cross fixed thresholds in the amplitude domain. This paper introduces a non-parametric estimator of the impulse response and transfer function of continuous-time systems based on such amplitude-equidistant sampling strategy, known as Lebesgue sampling. To this end, kernel methods are developed to formulate an algorithm that adequately takes into account the bounded output uncertainty between the event timestamps, which ultimately leads to more accurate models and more efficient output sampling compared to the equidistantly-sampled kernel-based approach. The efficacy of our proposed method is demonstrated through a mass–spring damper example with encoder measurements and extensive Monte Carlo simulation studies on system benchmarks.

1. Introduction

In system identification and control design, it is common to assume that the signals are sampled equidistantly in time. However, it is now well known that event-based sampling schemes can lead to improvements in control performance, as well as in resource efficiency (Åström & Bernhardsson, 2003). In particular, one of the most popular event-based sampling methods is Lebesgue sampling. The event associated with this sampling scheme is the crossing of fixed thresholds in the amplitude domain of the continuous-time signal of interest. Such type of sampling can be found in incremental encoders (Merry, van de Molengraft, & Steinbuch, 2013), and also in networked control systems, where the goal is to reduce resource utilization without affecting network throughput (Liu, Wang, He, & Zhou, 2014).

The Lebesgue sampling paradigm provides knowledge on what amplitude band the signals are located in at each instant of time. In this sense, this type of sampling is related to quantization, since a measurement (or lack of) at any instant in time that does not correspond to an event can be viewed as a quantized measurement. There has been extensive work on how to identify systems based on quantized measurements. The maximum likelihood estimator based on the Expectation–Maximization algorithm (EM) has been derived for finite impulse response (FIR) systems in Godoy, Goodwin, Agüero, Marelli, and Wigren (2011), while Chen, Zhao, and Ljung (2012) develop a regularized FIR estimator for binary measurements. An approximate maximum likelihood approach is studied in Risuleo, Bottegal, and Hjalmarsson (2019), and Bottegal, Hjalmarsson, and Pillonetto (2017) propose a kernel-based method for estimating FIR models. Other approaches have been pursued for the identification of IIR systems (Piga, Mejari, & Forgione, 2021; Pouliquen, Pigeon, Gehan, & Goudjil, 2019), ARX systems (Agüero, González, & Carvajal, 2017), and event-based sampling of FIR models with binary observations (Diao, Guo, & Sun, 2018).

The problem that is addressed in this paper is the estimation of non-parametric continuous-time models from Lebesgue-sampled output data. To this end, we seek estimators that can (1) provide a *continuous-time* impulse or transfer function estimate from possibly noisy and short data records, and (2) exploit the entirety of the output information contained in the irregular sampling instants and the bounded intersample behavior. Our interest in continuous-time models stems from the fact that they can provide physical interpretability, which is relevant when dealing with applications such as the identification of positioning systems with incremental encoder sensing (Strijbosch & Oomen, 2022). Furthermore, direct identification of continuous-time systems can deal with non-uniformly sampled data, which can be the case

☆ The material in this paper was partially presented at the 22nd IFAC World Congress (IFAC 2021), July 9–14, 2023, Yokohama, Japan. This paper was recommended for publication in revised form by Associate Editor Tianshi Chen under the direction of Editor Alessandro Chiuso.

* Corresponding author.
E-mail addresses: r.a.gonzalez@tue.nl (R.A. González), k.tiels@tue.nl (K. Tiels), t.a.e.oomen@tue.nl (T. Oomen).

https://doi.org/10.1016/j.automatica.2024.111648

for event-based sampling schemes, and they can incorporate the full continuous-time input information in the construction of the estimators (González, Rojas, Pan, & Welsh, 2021), which solves the bias problems encountered in discrete-time when the intersample behavior of the input is misspecified (Schoukens, Pintelon, & Van Hamme, 1994).

Although there has been recent work on non-parametric identification for continuous-time systems using kernel methods that use non-equidistantly sampled data (Pillonetto & De Nicolao, 2010; Scandella, Mazzoleni, Formentin, & Previdi, 2022), these works do not incorporate the intersample behavior information provided by a Lebesgue sampling framework, i.e., the lower and upper bounds on the unsampled output in between the time-stamps are not exploited. In Kawaguchi, Hikono, Maruta, and Adachi (2016) and Pouliquen, Goudjil, Gehan, and Pigeon (2016), continuous-time systems with Lebesgue-sampled and binary outputs are considered, although such results are only valid for parametric models with fixed model structures. On the other hand, the approaches in Bottegal et al. (2017), Chen, Zhao, and Ljung (2012) and Risuleo et al. (2019) for identification with quantized data might be used for obtaining a non-parametric discrete-time representation that can later be converted into continuous-time. However, this conversion is in many cases ill-defined or ill-conditioned, which drives the need for directly estimating a continuous-time system from the input–output data (Garnier & Young, 2014).

In summary, the main contributions of this paper are:

(C1) We introduce a loss function (in terms of the continuous-time impulse response to be estimated) that incorporates the intersample information we obtain through Lebesgue sampling. This loss function, after regularization, has an optimum that can be characterized by the generalized representer theorem (Schölkopf, Herbrich, & Smola, 2001; Wahba, 1990), and is related to a maximum *a posteriori* (MAP) optimization problem for Lebesgue-sampled data.

(C2) Once the kernel-regularized estimator is written as a finite linear combination of representers, we propose an iterative procedure that delivers the associated weights based on the MAP Expectation–Maximization (MAP-EM) method. We also contrast this procedure with a midpoint approach for identification with quantized data (Risuleo et al., 2019).

(C3) The hyperparameters that describe the kernel and noise variance are computed from an Empirical Bayes (EB) approach. We make the high-dimensional integral optimization problem tractable by

  (C3.1) Providing closed-form expressions for the kernel matrix in terms of the input samples and the kernel hyperparameters, which is made explicit for the stable-spline kernels, and

  (C3.2) Proposing an EM algorithm that iteratively computes the optimal hyperparameter vector. Such algorithm is presented in a matrix-inversion-free form by leveraging Cholesky and QR factorizations. While the noise variance estimate has a closed-form expression for its iterations, the other two hyperparameters are computed via a simple non-convex optimization step.

(C4) We obtain a closed-form expression for the estimated continuous-time transfer function in terms of the representer weight vector, the input samples, and an integrated version of the kernel in the frequency domain.

(C5) The proposed method is tested via Monte Carlo simulations.

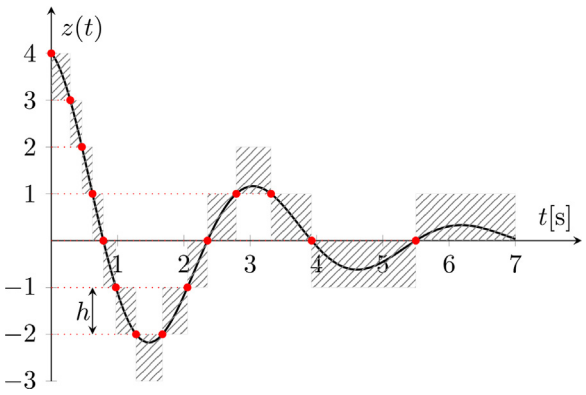

Fig. 1. Lebesgue sampling of a signal $z(t)$ with threshold amplitude $h = 1$. The red dots indicate the sampling instants and thresholds being crossed, and the dashed gray rectangles show the regions where $z(t)$ is known to be located.

The remainder of the paper is organized as follows. In Section 2, the problem of interest is stated, and practical aspects of Lebesgue-sampled system identification are covered. Section 3 introduces the ideas and notation behind non-parametric continuous-time system identification using kernel methods. Section 4 contains the main contribution of this paper, namely, the derivation of a kernel-based estimator for continuous-time, linear and time-invariant (LTI), Lebesgue-sampled systems. Numerical studies are presented in Section 5, while Section 6 provides concluding remarks.

Preliminary results related to the current manuscript are presented in González, Tiels, and Oomen (2023). The present paper substantially extends these results by (1) providing a MAP interpretation to the novel cost function being minimized for identification, (2) introducing an initialization for the MAP-EM approach, (3) proposing more computationally efficient optimization problems for the hyperparameters and (4) deriving the estimated transfer function description in closed form. Additional simulation setups are tested and presented in this paper, and all proofs can be found in the Appendix.

2. Setup and problem formulation

2.1. System and setup

Consider the following LTI, asymptotically stable, strictly causal, continuous-time system

$$x(t) = \int_0^\infty g(\tau)u(t-\tau)\mathrm{d}\tau, \tag{1}$$

where u is the input, which is assumed to be a causal function in t (i.e., $u(t) = 0$ for $t < 0$) that is deterministic and exogenous, and g is the impulse response of the LTI system. The transfer function of the LTI system, defined as the Laplace transform of the impulse response g, is denoted as $G(s)$, where s denotes the Laplace complex variable. The frequency response function associated with g is given by evaluating $G(s)$ at $s = \mathrm{i}\omega$.

The input $u(t)$ is assumed to be perfectly known, i.e., there is no noise in its measurement. The output $x(t)$ is corrupted by additive noise $v(t)$, which results in a continuous-time signal $z(t) = x(t) + v(t)$. Assume that we have access to N_L data points of the Lebesgue-sampled version of $z(t)$, as in Fig. 1. That is, given the threshold amplitude $h > 0$ and the continuous-time signal $z(t)$, we have at disposal the sampled sequence $\{y_\mathrm{L}(t_l)\}_{l=1}^{N_\mathrm{L}}$ that satisfies $y_\mathrm{L}(t_l) = z(t_l)$. The sampling times (or time-stamps) t_l, $l = 1, 2, \ldots, N_\mathrm{L}$, are the instants in time at which $z(t)$ crosses

a fixed threshold hm_l, with $m_l \in \mathbb{Z}$. Formally, we characterize the time-stamps by

$$t_l = \min\{\tau \in (t_{l-1}, \infty) : z(\tau) = mh \text{ for some } m \in \mathbb{Z}\},$$
$$m_l = z(t_l)/h.$$

Without loss of generality and for simplicity only, we assume that $t_1 = 0$. The goal is to obtain an estimate of the continuous-time system g using the continuous-time input $\{u(t)\}_{t\in[0,t_{N_L}]}$ and the Lebesgue-sampled output data $\{y_L(t_l)\}_{l=1}^{N_L}$.

2.2. Practical framework for Lebesgue-sampled system identification

Incremental encoders operate on this kind of sampling principle (Merry et al., 2013). In practice, a light source emits a beam directed towards a slotted disk or strip, and the output of two light detectors are recorded. These two signals allow the encoder to detect the direction of the rotation. These signals are evaluated at a high sampling rate compared to that of the input sequence, typically generated by a zero-order-hold (ZOH) device (Strijbosch & Oomen, 2019). The quantity h represents the uncertainty in the measurements of the incremental encoder, which is inversely proportional to its resolution. In low-resolution incremental encoders, the quantization effect produced by h, in conjunction with the non-equidistant nature of the sampling mechanism, can impact the performance and design of iterative learning control (Strijbosch & Oomen, 2022) or repetitive control (Kon, Strijbosch, Koekebakker, & Oomen, 2021).

With this context in mind, we define $\Delta > 0$ as the (equidistant) sampling period of the amplitude detection mechanism. The following assumption is set in place:

Assumption 1. For every time instant $t = i\Delta$, the lower and upper threshold levels associated with the unsampled output $z(t)$ are known. The lower bound at each time instant $t = i\Delta$ is denoted as η_i, and it can be deduced unambiguously from $\{y_L(t_l)\}_{l=1}^{N_L}$.

Thus, a set-valued signal $y(i\Delta)$ can be defined as

$$y(i\Delta) = \mathcal{Q}_h\{z(i\Delta)\} := [\eta_i, \eta_i + h) \tag{2}$$

for $i = 0, 1, \ldots, N$, with $N := \lfloor t_{N_L}/\Delta \rfloor + 1$. To simplify our notation, we denote $\{z(i\Delta)\}_{i=0}^{N}$ as the vector $\mathbf{z}_{0:N}$, and we define the set describing the output measurements as

$$\mathcal{Y}_{1:N} = \left\{[z_1, \ldots, z_N]^\top \in \mathbb{R}^N : z_i \in y(i\Delta), i = 1, \ldots, N\right\}. \tag{3}$$

Assumption 1 eradicates possible inconsistencies that could occur if $z(t)$ is tangential to one of the threshold levels. Note that we do not assume that each time-stamp t_l is a multiple of Δ. Although such assumption is commonly used in intermittent sampling setups (Kon et al., 2021), and is well justified if the sampling period Δ is sufficiently small, we do not require it for the proposed method.

Assumption 2. The sampled disturbance $v(i\Delta)$ affecting the output $z(i\Delta)$ is an additive discrete-time independent and identically distributed (i.i.d.) Gaussian noise of zero mean and variance σ^2 (see Fig. 2).

The noise variance is not known beforehand, and the user may decide on estimating it from the data or selecting a value according to expert knowledge. For the former approach, it is possible to estimate the noise variance from other tools from identification with quantized measurements (Godoy et al., 2011), or to include it as an extra hyperparameter to be estimated in the proposed kernel approach.

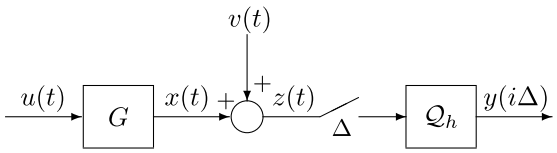

Fig. 2. Block diagram of the Lebesgue sampling scheme. Note that $\mathcal{Q}_h$ delivers a set-valued signal y, which is used for identification.

2.3. Problem formulation

Taking into consideration Assumptions 1 and 2, the problem we are interested in is as follows: Assume that the causal continuous-time input $\{u(t)\}_{t\in[0,t_{N_L}]}$ is perfectly known, i.e., there is no noise in its measurement, and that we have access to $\{y(i\Delta)\}_{i=0}^{N}$, i.e., the upper and lower threshold bounds of z. The goal is to estimate the underlying continuous-time impulse response g (or its transfer function $G(s)$) from the input and output data.

Remark 1. In many cases, the input in an identification experiment is generated by a zero-order-hold device of sampling period Δ_u, with $\Delta_u \gg \Delta$. When Δ_u is a multiple of Δ, we may consider the sampled input signal $\{u(i\Delta)\}_{i=0}^{N}$, instead of a fully continuous-time description for u. Clearly both viewpoints describe the same input and are thus equivalent if the intersample behavior of the sampled input is known and correctly incorporated in the construction of the algorithms.

Remark 2. We will only consider the output data that are produced by the input starting from $t = 0$. Since the system to be identified is assumed strictly causal, we discard the first output measurement $y(0)$.

3. Kernel-based continuous-time system identification: Preliminaries

This section provides the essential tools behind non-parametric continuous-time system identification using kernel methods, as detailed in, e.g., Dinuzzo (2015), Pillonetto, Dinuzzo, Chen, De Nicolao, and Ljung (2014). In particular, we introduce the notation that is subsequently employed in formulating the proposed non-parametric estimator for Lebesgue-sampled systems.

For unquantized data, estimating the continuous-time impulse response g in a kernel-based framework equates to solving a minimization problem of the form

$$\min_{g\in\mathcal{G}} \left(\sum_{i=1}^{N} L\big(z(i\Delta), (g * u)(i\Delta)\big) + \gamma \|g\|_{\mathcal{G}}^2 \right), \tag{4}$$

where $\mathcal{G}$ is a Hilbert space of functions, $L(\cdot)$ is a loss function of choice (not necessarily convex (Schölkopf et al., 2001)), and γ is a positive scalar regularization parameter. If the input signal and the space $\mathcal{G}$ are such that all the pointwise evaluated convolutions are bounded linear functionals, then there exist unique representers $\hat{g}_i$ such that $(g * u)(i\Delta) = \langle g, \hat{g}_i\rangle_{\mathcal{G}}$. With this in mind, the representer theorem (Dinuzzo & Schölkopf, 2012; Schölkopf et al., 2001) indicates that any optimal solution of (4) can be expressed as a *finite* linear combination of the form

$$\hat{g}(t) = \sum_{i=1}^{N} c_i \hat{g}_i(t), \tag{5}$$

where the optimal vector of coefficients $\hat{\mathbf{c}} := [c_1, c_2, \dots, c_N]^\top$ is obtained by

$$\hat{\mathbf{c}} = \arg\min_{\mathbf{c}\in\mathbb{R}^N} \left(\sum_{i=1}^{N} L\big(z(i\Delta), \mathbf{K}_i^\top \mathbf{c}\big) + \gamma \mathbf{c}^\top \mathbf{K} \mathbf{c} \right), \tag{6}$$

and $\mathbf{K}_i$ denotes the $(i+1)$th column of the kernel matrix $\mathbf{K}$. This matrix is assumed to be non-singular. More explicitly, the representers can be described in terms of the kernel function $k\colon \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$, which fully characterizes the Reproducing Kernel Hilbert Space (RKHS) $\mathcal{G}$. Indeed,

$$\hat{g}_i(t) = \int_0^\infty u(i\Delta - \tau) k(t, \tau) \mathrm{d}\tau,$$

and the entries of the kernel matrix are given by

$$\mathbf{K}_{ij} = \int_0^\infty \int_0^\infty u(i\Delta - \xi) u(j\Delta - \tau) k(\xi, \tau) \mathrm{d}\tau \mathrm{d}\xi. \tag{7}$$

One degree of freedom in this framework is the selection of the RKHS space $\mathcal{G}$, which is equivalent to choosing a suitable kernel k with hyperparameters $\boldsymbol{\beta}$. There are several kernels for continuous-time impulse response estimation (Pillonetto et al., 2014). For example, the stable-spline one of order $q \in \mathbb{N}$ is defined as

$$k(t, \tau) = s_q(e^{-\beta t}, e^{-\beta \tau}),$$

where the hyperparameter β is a strictly positive scalar, and s_q is the regular spline kernel of order q, given by Scandella et al. (2022, Prop. 2.1)

$$s_q(e^{-\beta t}, e^{-\beta \tau}) = \sum_{r=0}^{q-1} \gamma_{q,r} \begin{cases} e^{-\beta(2q-r-1)t} e^{-r\beta\tau} & \text{if } t \geq \tau, \\ e^{-\beta(2q-r-1)\tau} e^{-r\beta t} & \text{if } t < \tau, \end{cases} \tag{8}$$

where

$$\gamma_{q,r} = \frac{(-1)^{q+r-1}}{r!(2q-r-1)!}.$$

In practice, the hyperparameters $\boldsymbol{\beta}$, the positive gain γ in (4), and in some cases the noise variance σ^2, are tuned according to some fitting criteria such as cross validation, the SURE approach (Pillonetto & Chiuso, 2015) or Empirical Bayes (Pillonetto, Chen, Chiuso, De Nicolao, & Ljung, 2022).

Remark 3. The regularization problem in (4) admits a probabilistic interpretation in terms of MAP estimation. Under such perspective, the impulse response g is modeled as a Gaussian process with covariance being described by the kernel k. This interpretation is extended in Section 4.1 of this work to the context of Lebesgue sampling. For more details on the probabilistic interpretation for LTI systems, see Pillonetto et al. (2022, Chap. 7).

4. Non-parametric estimation using Lebesgue-sampled data

In this section, the non-parametric estimator for systems with Lebesgue-sampled data is developed. We divide this section in six parts, which are enumerated next:

(1) The Representer theorem for Lebesgue-sampled systems and its MAP interpretation;
(2) A method for initializing the computation of the weights related to each representer;
(3) The computation of the optimal weights using the MAP-EM algorithm;
(4) The kernel-hyperparameter optimization;
(5) A transfer function description for the impulse response estimate; and
(6) The full algorithm written in pseudocode.

4.1. Representer theorem for Lebesgue-sampled systems

The first goal, which constitutes Contribution C1 of this paper, is to derive a loss function L for estimating the impulse response via (4) which incorporates the set knowledge of the output, and to show how it relates with a MAP estimation problem. With that in mind, a Bayesian interpretation of kernel-based methods (Kimeldorf & Wahba, 1970; Pillonetto et al., 2022) involves computing the MAP estimate

$$\hat{g}_{\text{MAP}}(t) = \arg\max_g \big(\ell(g) + \log \mathrm{p}(g)\big), \tag{9}$$

where $\mathrm{p}(g)$ is the prior distribution of g, which is assumed to be a zero-mean Gaussian process with covariance $\mathbb{E}\{g(t)g(s)\} = k(t,s)/\gamma$, and $\ell(\cdot)$ denotes the log-likelihood function

$$\ell(g) = \log \mathrm{p}(\mathcal{Y}_{1:N}|g).$$

Intuitively, the MAP estimator (9) is related to the optimization problem in (4) by letting the *a priori* probability density of g be proportional to $\exp(-\gamma\|g\|_{\mathcal{G}}^2)$, and letting L in (4) be the negative log-likelihood of the measured output data. The main issue that is addressed in this paper is that this argument does not directly hold for g in our case, since the probability density of g is not well defined as it belongs to an infinite-dimensional function space (Bogachev, 1998). To this end, a key idea taken here is that it is possible to formalize this intuition by considering the MAP estimator of any *finite* set of samples $(g * u)(t_i)$ that contains the (noiseless) observation set $\{(g * u)(i\Delta)\}_{i=1}^N$. The following lemma uses this insight to provide a formal justification to the choice of L needed for estimating Lebesgue-sampled continuous-time systems.

Lemma 1. *Suppose that Assumptions 1 and 2 hold, and that g is a zero-mean Gaussian process that is independent of $\{v(i\Delta)\}_{i=0}^N$ and has covariance $\mathbb{E}\{g(t)g(s)\} = k(t,s)/\gamma$. Let $\{t_i\}_{i=1}^{N+M}$ be a finite set of real values such that $t_i = i\Delta$ for $i = 1, 2, \dots, N$, and where $\{t_i\}_{i=N+1}^{N+M}$ are arbitrary. Define the vector of noiseless output values*

$$\mathbf{x} = [(g * u)(t_1), (g * u)(t_2), \dots, (g * u)(t_{N+M})]^\top.$$

Furthermore, define $\breve{g}$ as the solution of the optimization problem

$$\min_{g\in\mathcal{G}} \left(-2 \sum_{i=1}^{N} \log \left[\int_{\eta_i}^{\eta_i + h} e^{\frac{-1}{2\sigma^2}\left[z_i - (g*u)(i\Delta)\right]^2} \mathrm{d}z_i \right] + \gamma \|g\|_{\mathcal{G}}^2 \right), \tag{10}$$

where $\|\cdot\|_{\mathcal{G}}$ is the RKHS norm induced by the kernel k. Then, the MAP estimate of $\mathbf{x}$ given $\mathcal{Y}_{1:N}$ is

$$\hat{\mathbf{x}} = [(\breve{g} * u)(t_1), (\breve{g} * u)(t_2), \dots, (\breve{g} * u)(t_{N+M})]^\top.$$

Proof. For the following analysis, define the first N elements in $\mathbf{x}$ as $\mathbf{x}_1$. The analysis with the first N elements is the relevant and non-standard step, since the MAP estimator of the last $M - N$ elements in $\mathbf{x}$ can be derived with a similar methodology to that in Proposition 5 of Aravkin, Bell, Burke, and Pillonetto (2014), and is therefore omitted. We first must compute the likelihood function $\log \mathrm{p}(\mathcal{Y}_{1:N}|\mathbf{x}_1)$. To this end, the probability density function of the output prior to sampling $\mathbf{z}_{1:N} = [z(\Delta), \dots, z(N\Delta)]^\top$ (conditioned on $\mathbf{x}_1$) is given by

$$\mathrm{p}\left(\mathbf{z}_{1:N}|\mathbf{x}_1\right) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \prod_{i=1}^{N} e^{-\frac{1}{2\sigma^2}\left[z(i\Delta) - (g*u)(i\Delta)\right]^2},$$

where we have used the fact that the additive noise is Gaussian and i.i.d. by Assumption 2. Therefore, the probability mass function of $\mathcal{Y}_{1:N}$ is

$$\mathrm{p}\left(\mathcal{Y}_{1:N}|\mathbf{x}_1\right)$$

$$= \mathbb{P}\big(z(\Delta) \in [\eta_1, \eta_1 + h), \ldots, z(N\Delta) \in [\eta_N, \eta_N + h) | \mathbf{x}_1\big)$$
$$= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \prod_{i=1}^{N} \int_{\eta_i}^{\eta_i+h} e^{-\frac{1}{2\sigma^2}\left[z_i - (g*u)(i\Delta)\right]^2} \mathrm{d}z_i. \tag{11}$$

From (11), the log-likelihood function $\ell(\mathbf{x}_1)$ can be written as

$$\ell(\mathbf{x}_1) = \sum_{i=1}^{N} \log\left[\int_{\eta_i}^{\eta_i+h} e^{-\frac{1}{2\sigma^2}\left[z_i - (g*u)(i\Delta)\right]^2} \mathrm{d}z_i\right] + C,$$

where C is a known constant. On the other hand, $\mathbf{x}_1$ is zero-mean and normally distributed with covariance that has entries given by

$$\mathbb{E}\{(g*u)(i\Delta)(g*u)(j\Delta)\}$$
$$= \mathbb{E}\left\{\int_0^\infty u(i\Delta - \tau)g(\tau)\mathrm{d}\tau \int_0^\infty u(j\Delta - \xi)g(\xi)\mathrm{d}\xi\right\}$$
$$= \int_0^\infty \int_0^\infty u(i\Delta - \tau)u(j\Delta - \xi)\mathbb{E}\{g(\tau)g(\xi)\}\mathrm{d}\tau\mathrm{d}\xi$$
$$= \mathbf{K}_{ij}/\gamma. \tag{12}$$

This leads to the following MAP estimator for $\mathbf{x}_1$:

$$\hat{\mathbf{x}}_1 = \arg\max_{\mathbf{x}_1}\big(\ell(\mathbf{x}_1) + \log \mathrm{p}(\mathbf{x}_1)\big)$$
$$= \arg\max_{\mathbf{x}_1}\left(\sum_{i=1}^{N} \log\left[\int_{\eta_i}^{\eta_i+h} e^{\frac{-1}{2\sigma^2}\left[z_i - (g*u)(i\Delta)\right]^2} \mathrm{d}z_i\right] - \frac{\gamma \mathbf{x}_1^\top \mathbf{K}^{-1}\mathbf{x}_1}{2}\right),$$

Under the representation $(g*u)(i\Delta) = \mathbf{K}_i^\top \mathbf{c}$ for $i = 1, 2, \ldots, N$, we obtain that $\hat{\mathbf{x}}_1 = \mathbf{K}\hat{\mathbf{c}}$, where

$$\hat{\mathbf{c}} = \arg\max_{\mathbf{c}\in\mathbb{R}^N}\left(\sum_{i=1}^{N} \log\left[\int_{\eta_i}^{\eta_i+h} e^{\frac{-1}{2\sigma^2}\left(z_i - \mathbf{K}_i^\top \mathbf{c}\right)^2} \mathrm{d}z_i\right] - \frac{\gamma \mathbf{c}^\top \mathbf{K}\mathbf{c}}{2}\right). \tag{13}$$

This is precisely the optimal weighting of the representers that describe the solution of (10) via the representer theorem. This completes the proof. □

Lemma 1 provides a relation between the impulse response minimization problem in (4) and the MAP estimator in a Lebesgue-sampling framework. More precisely, we have shown that the following choice of loss function for when the output is Lebesgue-sampled

$$L\big(y(i\Delta), (g*u)(i\Delta)\big)$$
$$= 2\sum_{i=1}^{N} \log\left[\int_{\eta_i}^{\eta_i+h} e^{\frac{-1}{2\sigma^2}\left[z_i - (g*u)(i\Delta)\right]^2} \mathrm{d}z_i\right]$$

leads to a MAP estimator of the noiseless output of the system prior to Lebesgue sampling. Since the integer M in Lemma 1 is arbitrary, $(\breve{g} * u)(t)$ represents a MAP estimator of the noiseless output for any time instant $t \in [\Delta, \Delta N]$.

The following subsections are focused on how to compute the minimizer of (10), and how to choose a specific kernel according to the Lebesgue-sampled data. The optimization problem in (10) does not have an explicit form as the Riemann sampling counterpart, i.e., the point-valued output case, Pillonetto et al. (2022). However, the representer theorem indicates that any optimal solution of (10) can anyway be expressed as a *finite* linear combination of the representers $\hat{g}_i$ of the form (5) with $\hat{\mathbf{c}}$ being given by (13). Next, we cover how to compute $\hat{\mathbf{c}}$, the optimal weighting of the representers $\hat{g}_i$, for a fixed kernel k and hyperparameters γ and σ^2.

4.2. Optimal weights with MAP-EM

In this subsection, we present a MAP-EM algorithm to obtain an iterative procedure that computes (13). The derivation of this iterative procedure, which ensures the computation of a local maximum of the cost in (13) under general conditions as a generalization of the standard EM approach (McLachlan & Krishnan, 2007; Wu, 1983), constitutes Contribution C2 of this paper. The approach consists of relating (13) to the MAP of a specific FIR model in discrete-time, to later apply the EM algorithm (Dempster, Laird, & Rubin, 1977) tailored for MAP estimation. This relation is made evident in the following lemma.

Lemma 2. *Consider the following model*

$$z(i\Delta) = \mathbf{K}_i^\top \mathbf{c} + e(i\Delta), \tag{14a}$$
$$y(i\Delta) = \mathcal{Q}_h\{z(i\Delta)\}, \tag{14b}$$

where $e(\Delta), \ldots, e(N\Delta)$ are i.i.d. Gaussian with variance σ^2, and $\mathbf{K}_i$, $i = 1, 2, \ldots, N$, is assumed known. Assume that $\mathbf{c}$ in (14a) has a Gaussian prior distribution, with zero mean and covariance $(\gamma\mathbf{K})^{-1}$. Then, the MAP estimator for $\mathbf{c}$ is given by $\hat{\mathbf{c}}$ in (13).

Proof. See Appendix A.1. □

By Lemma 2 we can view the computation of the weights $\hat{\mathbf{c}}$ in a MAP-EM framework if we set the unquantized data $\mathbf{z}_{1:N}$ as our hidden variable. In other words, we can optimize the *a posteriori* density for $\mathbf{c}$, which is exactly the objective function in (13), by iteratively (1) computing the conditional expectation of the log complete-data posterior density given the set measurements $\mathcal{Y}_{1:N}$ and the current estimate of $\hat{\mathbf{c}}$ (i.e., the E-step), and later (2) performing a maximization step (M-step). These two steps are outlined in Algorithm 1. Note that this method departs from the standard EM method in the objective function of the maximization step, which here includes the log prior density. The E-step is computed using a result from quantized FIR maximum likelihood estimation, while the M-step including the log prior density is presented in Theorem 1.

Algorithm 1 MAP-EM algorithm for the computation of $\hat{\mathbf{c}}$ in (13)

1: Select an initial estimate $\hat{\mathbf{c}}^{(1)}$, a maximum number of iterations M_{iter}, and a tolerance factor ϵ
2: $j \leftarrow 1$, flag $\leftarrow 1$
3: **while** $j \leq M_{\text{iter}}$ and flag $= 1$ **do**
4: **E-step**: Compute the expectation

$$Q(\mathbf{c}, \hat{\mathbf{c}}^{(j)}) = \mathbb{E}\left\{\log \mathrm{p}(\mathbf{z}_{1:N}, \mathcal{Y}_{1:N}|\mathbf{c}) | \mathcal{Y}_{1:N}, \hat{\mathbf{c}}^{(j)}\right\}. \tag{15}$$

5: **M-step**: Solve the optimization problem

$$\hat{\mathbf{c}}^{(j+1)} = \arg\max_{\mathbf{c}\in\mathbb{R}^N}\left(Q(\mathbf{c}, \hat{\mathbf{c}}^{(j)}) - \frac{\gamma \mathbf{c}^\top \mathbf{K}\mathbf{c}}{2}\right). \tag{16}$$

6: **if** $\dfrac{\|\hat{\mathbf{c}}^{(j+1)} - \hat{\mathbf{c}}^{(j)}\|_2}{\|\hat{\mathbf{c}}^{(j)}\|_2} < \epsilon$ **then**
7: flag $\leftarrow 0$
8: **end if**
9: $j \leftarrow j + 1$
10: **end while**

Lemma 3 (*Godoy et al., 2011, Lemma 5*)**.** *Consider the discrete-time model (14). The Q function in (15) satisfies*

$$Q(\mathbf{c}, \hat{\mathbf{c}}^{(j)}) = \frac{-1}{2\sigma^2}\sum_{i=1}^{N} \int_{\eta_i}^{\eta_i+h} (z_i - \mathbf{K}_i^\top \mathbf{c})^2 \mathrm{p}(z_i | y(i\Delta), \hat{\mathbf{c}}^{(j)})\mathrm{d}z_i + C,$$

where C is a constant.

Proof. See Godoy et al. (2011). □

Theorem 1. *The M-step in (16) is equivalent to*

$$\hat{\mathbf{c}}^{(j+1)} = (\mathbf{K} + \tilde{\gamma}\mathbf{I})^{-1}\tilde{\mathbf{z}}^{(j)}, \tag{17}$$

where $\tilde{\gamma} = \gamma\sigma^2$, and with the ith entry of $\tilde{\mathbf{z}}^{(j)}$ being given by

$$\tilde{z}_i^{(j)} = \mathbf{K}_i^\top \hat{\mathbf{c}}^{(j)} + \frac{\sqrt{\frac{2}{\pi}}\sigma\left(\exp\{-(b_i^{(j)})^2\} - \exp\{-(b_i^{(j)} + \frac{h}{\sqrt{2}\sigma})^2\}\right)}{\operatorname{erf}[b_i^{(j)} + \frac{h}{\sqrt{2}\sigma}] - \operatorname{erf}[b_i^{(j)}]}, \tag{18}$$

where $b_i^{(j)} := (\eta_i - \mathbf{K}_i^\top \hat{\mathbf{c}}^{(j)})/(\sqrt{2}\sigma)$, and the error function erf[x] *is defined by*

$$\operatorname{erf}[x] = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t.$$

Proof. See Appendix A.2. □

Theorem 1 reveals that the optimal weights $\hat{\mathbf{c}}$ can be computed from successive regularized least squares expressions. These have the same form as the standard solution for the optimal weights for unquantized data (Pillonetto et al., 2022, Theorem 7.3), but with an iteration-varying output vector $\tilde{\mathbf{z}}^{(j)}$. Interestingly, $\tilde{z}_i^{(j)}$ can be interpreted as the conditional mean of $z(i\Delta)$ given the available quantized data and the current weight vector $\hat{\mathbf{c}}^{(j)}$; see Eq. (38) of Appendix A.2 for this interpretation.

Remark 4. The iterations provided by the M-step in Theorem 1 require an initial estimate $\hat{\mathbf{c}}^{(1)}$. To this end, by noting that $\tilde{z}_i^{(j)} \in [\eta_i, \eta_i + h)$ for all $i = 1, 2, \ldots, N$, we may follow a best worst-case approach and set

$$\hat{\mathbf{c}}^{(1)} = (\mathbf{K} + \tilde{\gamma}\mathbf{I})^{-1}\tilde{\mathbf{z}}^{(0)}, \tag{19}$$

with the ith entry of $\tilde{\mathbf{z}}^{(0)}$ being the midpoints of each quantization level, i.e., $\tilde{z}_i^{(0)} = \eta_i + h/2$. This initialization coincides with the approach suggested in Risuleo et al. (2019) for constructing an approximate maximum likelihood estimator under quantized data.

4.3. Kernel hyper-parameter optimization

Here we consider the marginal likelihood method for computing an appropriate hyperparameter vector, also known as the Empirical Bayes approach. This approach, which has been proven useful in other contributions on kernel system identification (Bottegal et al., 2017; Pillonetto & De Nicolao, 2010; Pillonetto et al., 2014; Scandella et al., 2022), proposes to estimate the hyperparameter vector $\boldsymbol{\rho} = [\boldsymbol{\beta}^\top, \gamma, \sigma^2]^\top$ by solving the maximum likelihood problem

$$\hat{\boldsymbol{\rho}}_{\mathrm{EB}} = \arg\max_{\boldsymbol{\rho}\in\boldsymbol{\Gamma}} \mathrm{p}(\mathcal{Y}_{1:N}|\boldsymbol{\rho}), \tag{20}$$

where $\boldsymbol{\Gamma}$ denotes the admissible space of hyperparameters, which must consider $\gamma, \sigma^2 > 0$. To describe such optimization problem more explicitly, we first compute the probability density function of the output prior to Lebesgue sampling. This expression can be obtained directly by exploiting the fact that the additive noise is Gaussian and independent of g (which is also assumed Gaussian, and satisfies (12)), thus leading to

$$\mathbf{z}_{1:N}|\boldsymbol{\rho} \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_{\boldsymbol{\beta}}/\gamma + \sigma^2\mathbf{I}), \tag{21}$$

where we have made explicit the dependence of the kernel matrix $\mathbf{K}$ on the kernel hyperparameter vector $\boldsymbol{\beta}$. Therefore, the Empirical Bayes estimator for $\boldsymbol{\rho}$ is given by

$$\hat{\boldsymbol{\rho}}_{\mathrm{EB}} = \arg\max_{\boldsymbol{\rho}\in\boldsymbol{\Gamma}} \frac{1}{\sqrt{\det(2\pi[\mathbf{K}_{\boldsymbol{\beta}}/\gamma + \sigma^2\mathbf{I}])}} \times \int_{\mathbf{z}\in\mathcal{Y}_{1:N}} \exp\left\{-\frac{1}{2}\mathbf{z}^\top(\mathbf{K}_{\boldsymbol{\beta}}/\gamma + \sigma^2\mathbf{I})^{-1}\mathbf{z}\right\}\mathrm{d}\mathbf{z}, \tag{22}$$

where $\mathcal{Y}_{1:N}$ is defined in (3). This non-convex optimization problem involves an N-dimensional integral, which is hard to compute in general (see, e.g., Bottegal et al., 2017; Chen, Zhao, & Ljung, 2012). The intractability is here solved by optimizing (22) with EM along similar lines as in the previous subsection. For brevity, we derive the EM iterations jointly (both E and M steps) in Theorem 2.

Theorem 2. *The following iterative procedure is guaranteed to converge with probability 1 to a (local or global) maximum for the cost in (22):*

$$\hat{\boldsymbol{\rho}}^{(j+1)} = \arg\min_{\boldsymbol{\rho}\in\boldsymbol{\Gamma}} \left(\log\det(\mathbf{S}_{\boldsymbol{\rho}}) + \operatorname{tr}\{\mathbf{S}_{\boldsymbol{\rho}}^{-1}\bar{\mathbf{Q}}^{(j)}\}\right), \tag{23}$$

where $\mathbf{S}_{\boldsymbol{\rho}} := \mathbf{K}_{\boldsymbol{\beta}}/\gamma + \sigma^2\mathbf{I}$, and $\bar{\mathbf{Q}}^{(j)}$ is the second moment of $\mathbf{z}_{1:N}$ given the data and the jth iteration of $\hat{\boldsymbol{\rho}}$, i.e.,

$$\bar{\mathbf{Q}}^{(j)} = \mathbb{E}\{\mathbf{z}_{1:N}\mathbf{z}_{1:N}^\top|\mathcal{Y}_{1:N}, \hat{\boldsymbol{\rho}}^{(j)}\}. \tag{24}$$

Proof. See Appendix A.3. □

Remark 5. The $\bar{\mathbf{Q}}^{(j)}$ matrix in (24) cannot be computed in closed-form in general. In this paper, we extract samples of a multivariate truncated Gaussian distribution using the minimax tilting algorithm in Botev (2017) and we approximate the expectation in (24) via Monte Carlo integration.

The iterations in (23) to solve (22) can possibly be ill-conditioned and computationally costly to compute. In particular, the kernel matrix $\mathbf{K}$, with elements described in (7), is known to be difficult to compute for continuous-time system identification due to the presence of integrals instead of sums in the discrete-time case (Dinuzzo, 2015; Scandella et al., 2022). Here we provide the necessary details to explicitly write the elements of this matrix for any kernel k in terms of samples of an input with zero-order hold intersample behavior (recall Remark 1), which is later used in Theorem 3 for constructing more computationally efficient iterations for solving (22). The following lemma and its corollary (Corollary 1) constitute Contribution C3.1 of the paper.

Lemma 4. *Consider the kernel matrix $\mathbf{K}$ with entries described in (7). If $u(t)$ is constant between the time instants $t = 0, \Delta, 2\Delta, \ldots, N\Delta$, then $\mathbf{K}$ admits the decomposition*

$$\mathbf{K}_{\boldsymbol{\beta}} = \boldsymbol{\Phi}\mathcal{O}_{\boldsymbol{\beta}}\boldsymbol{\Phi}^\top, \tag{25}$$

where $\boldsymbol{\Phi}$ is given by

$$\boldsymbol{\Phi} = \begin{bmatrix} u(0) & & & 0 \\ u(\Delta) & u(0) & & \\ \vdots & & \ddots & \\ u([N-1]\Delta) & u([N-2]\Delta) & \cdots & u(0) \end{bmatrix}, \tag{26}$$

and the matrix $\mathcal{O}_{\boldsymbol{\beta}} \in \mathbb{R}^{N\times N}$ has entries

$$\mathcal{O}_{\boldsymbol{\beta},ij} = \int_{\Delta[i-1]}^{\Delta i}\int_{\Delta[j-1]}^{\Delta j} k(\xi, \tau)\mathrm{d}\tau\,\mathrm{d}\xi. \tag{27}$$

Proof. See Appendix A.4. □

Corollary 1. *Consider the kernel matrix $\mathbf{K}$ with entries described in (7), with k being the stable-spline kernel of order q in (8). If $u(t)$ is constant between the time instants $t = 0, \Delta, 2\Delta, \ldots, N\Delta$, then $\mathbf{K}$*

admits the decomposition $\mathbf{K}_\beta = \mathbf{\Phi}\mathcal{O}_\beta\mathbf{\Phi}^\top$, *where* $\mathbf{\Phi}$ *is given by (26) and the matrix* $\mathcal{O}_\beta \in \mathbb{R}^{N\times N}$ *has entries*

$$\mathcal{O}_{\beta,ij} = \sum_{r=0}^{q-1} \frac{\gamma_{q,r} e^{-\beta\Delta(2q-1)\max\{i,j\}}}{\beta^2 r(2q-r-1)} \begin{cases} a(\beta) & \text{if } i=j, \\ b_{i-j}(\beta) & \text{if } i\neq j, \end{cases}$$

where

$$a(\beta) = \frac{2[(2q-r-1) + re^{\beta\Delta(2q-1)} - (2q-1)e^{\beta\Delta r}]}{(2q-1)},$$
$$b_{i-j}(\beta) = e^{-\beta\Delta r(1-|i-j|)}(e^{\beta r\Delta}-1)(e^{\beta\Delta(2q-1)} - e^{\beta\Delta r}).$$

Proof. Direct from replacing $k(\xi,\tau)$ in (27) for (8) and solving the integrals. □

Remark 6. The continuous-time setting provides substantial freedom compared to discrete-time approaches for incorporating the intersample behavior of the input signal. Although Lemma 4 and Corollary 1 are exact only for zero-order hold inputs, these results can be extended in exact form (at the expense of more computations but avoiding numerical integration techniques), to any input with a specified intersample behavior (e.g., first-order hold, or B-splines used in a generalized hold framework Arriagada & Yuz, 2008). Throughout this paper, only ZOH is considered; extensions to other interpolation schemes are conceptually straightforward.

The description for $\mathbf{K}$ in Lemma 4 is now used to rewrite the iterations in (23) by considering an adequate QR factorization of the data at hand. For the following, we consider the change of variable $\tilde{\gamma} = \gamma\sigma^2$ and compute the Cholesky factorizations $\mathcal{O}_{\boldsymbol{\beta}}/\tilde{\gamma} = \mathbf{L}_\rho\mathbf{L}_\rho^\top$ and $\bar{\mathbf{Q}}^{(j)} = \mathbf{C}^{(j)}\mathbf{C}^{(j)\top}$, where $\mathbf{L}_\rho$ and $\mathbf{C}^{(j)}$ are upper triangular matrices with non-negative diagonal entries. We introduce the QR factorization

$$\begin{bmatrix} \mathbf{\Phi}\mathbf{L}_\rho & \mathbf{C}^{(j)} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} = \mathbf{Q}_\rho \begin{bmatrix} \mathbf{R}_{1,\rho} & \mathbf{R}_{2,\rho} \\ \mathbf{0} & \mathbf{R}_{3,\rho} \end{bmatrix}, \tag{28}$$

where $\mathbf{Q}_\rho$ is an orthogonal matrix (not to be confused with $\bar{\mathbf{Q}}^{(j)}$ in (24)), and $\mathbf{R}_{1,\rho}$, $\mathbf{R}_{3,\rho}$ are upper triangular matrices of dimension $N\times N$. Without loss of generality, we assume that they have positive diagonal entries. Note that the following identities are satisfied:

$$\mathbf{R}_{1,\rho}^\top\mathbf{R}_{1,\rho} = \mathbf{L}_\rho^\top\mathbf{\Phi}^\top\mathbf{\Phi}\mathbf{L}_\rho + \mathbf{I}, \tag{29a}$$
$$\mathbf{R}_{1,\rho}^\top\mathbf{R}_{2,\rho} = \mathbf{L}_\rho^\top\mathbf{\Phi}^\top\mathbf{C}^{(j)}, \tag{29b}$$
$$\mathbf{R}_{2,\rho}^\top\mathbf{R}_{2,\rho} + \mathbf{R}_{3,\rho}^\top\mathbf{R}_{3,\rho} = \mathbf{C}^{(j)\top}\mathbf{C}^{(j)}. \tag{29c}$$

Theorem 3 provides a straightforward implementation for computing the EM iterations of Theorem 2, which constitutes Contribution C3.2 of this paper.

Theorem 3. *The iterative procedure in (23) for computing* $\hat{\boldsymbol{\rho}}_{\text{EB}}$ *in (22) is equivalent to*

$$\begin{bmatrix} \hat{\tilde{\gamma}}^{(j+1)} \\ \hat{\boldsymbol{\beta}}^{(j+1)} \end{bmatrix} = \arg\min_{\tilde{\gamma},\boldsymbol{\beta}} \left(N\log\left(\|\mathbf{C}^{(j)}\|_{\text{F}}^2 - \|\mathbf{R}_{2,\rho}\|_{\text{F}}^2\right) + 2\log\det(\mathbf{R}_{1,\rho})\right), \tag{30}$$

$$\hat{\sigma}^{2(j+1)} = \frac{1}{N}\left(\|\mathbf{C}^{(j)}\|_{\text{F}}^2 - \|\mathbf{R}_{2,\hat{\rho}^{(j+1)}}\|_{\text{F}}^2\right), \tag{31}$$

where $\|\cdot\|_{\text{F}}$ *is the Frobenius norm,* $\mathbf{R}_{1,\rho}$ *and* $\mathbf{R}_{2,\rho}$ *are computed from (28), and* $\mathbf{C}^{(j)}$ *is the Cholesky factor of* $\bar{\mathbf{Q}}^{(j)}$ *in (24).*

Proof. See Appendix A.5. □

Remark 7. The expressions derived in Theorem 3 are related to the Empirical Bayes hyperparameter estimator computations for regularized least-squares in Chen and Ljung (2013) and González, Rojas, and Hjalmarsson (2021). In fact, in the absence of Lebesgue sampling, we would have $\bar{\mathbf{Q}}^{(j)} = \mathbf{z}_{1:N}\mathbf{z}_{1:N}^\top$, $\mathbf{C}^{(j)} = \mathbf{z}_{1:N}$, and the QR factorization in (28) is now a thin QR factorization (Horn & Johnson, 2012, Thm 2.1.14) that provides alternative closed-form expressions for computing the hyperparameter estimator in one iteration using similar formulas to (30) and (31). Contrary to the Riemann-sampling case, this work requires the EM algorithm to make the Empirical Bayes optimization tractable.

4.4. Transfer function description

The final theoretical contribution of this paper (Contribution C4) is the derivation of a more explicit expression for the estimated transfer function. Explicit expressions for general stable-spline kernels have been reported in Scandella et al. (2022) for unquantized output data with fully continuous-time inputs:

Proposition 1. *The transfer function associated to the minimizer of (10) can be written as*

$$\hat{G}(s) = \sum_{l=1}^{N} \hat{c}_l\hat{G}_l(s), \tag{32}$$

where $\{\hat{c}_l\}_{l=0}^N$ *is computed from (13), and*

$$\hat{G}_l(s) = \int_0^\infty K(s;\tau)u(l\Delta-\tau)\mathrm{d}\tau, \tag{33}$$

with $K(s;\tau)$ *being the Laplace transform of the kernel function* $k(t,\tau)$.

Proof. See Scandella et al. (2022). □

A similar expression to (32) also holds for this framework, as the only difference can be observed in the computation of the weights and the hyperparameters of the kernel (but not of the structure of the kernel itself). However, under the zero-order hold assumption on the input signal, we can provide an alternative representation of (32) for which the software implementation is easier and that does not rely on approximations of the intersample behavior of the input. This representation is stated in Lemma 5.

Lemma 5. *Consider the optimization problem in (10), where* $\mathcal{G}$ *is the RKHS induced by a kernel* k. *The transfer function associated to the minimizer of (10) can be written as*

$$\hat{G}(s) = \hat{\mathbf{c}}^\top\mathbf{\Phi}\mathcal{K}(s),$$

where $\hat{\mathbf{c}}$ *is computed from (13),* $\mathbf{\Phi}$ *is defined in (26), and* $\mathcal{K}(s)$ *is a vector of size* N *with entries* $\mathcal{K}_l(s)$ *given by the Laplace transform of the integrated kernel, i.e.,*

$$\mathcal{K}_l(s) = \int_0^\infty \left(\int_{\Delta[l-1]}^{\Delta l} k(t,\tau)\mathrm{d}\tau\right) e^{-st}\mathrm{d}t. \tag{34}$$

Proof. See Appendix A.6. □

Corollary 2. *If the RKHS* $\mathcal{G}$ *is induced by the stable-spline kernel of order* q, *then the transfer function associated to the minimizer of (10) can be written as* $\hat{G}(s) = \hat{\mathbf{c}}^\top\mathbf{\Phi}\mathcal{K}(s)$, *where* $\hat{\mathbf{c}}$ *is computed from (13),* $\mathbf{\Phi}$ *is defined in (26), and* $\mathcal{K}(s)$ *is a vector of size* N *with entries*

$\mathcal{K}_l(s)$ given by

$$\mathcal{K}_l(s) = \sum_{r=0}^{q-1} \frac{\gamma_{q,r} e^{-l\beta\Delta(2q-r-1)}(e^{\beta\Delta(2q-r-1)} - 1)}{\beta(s + r\beta)(2q - r - 1)} + \frac{(-1)^q \beta^{2q-1} e^{-l\Delta(s+\beta[2q-1])}(e^{\Delta(s+\beta[2q-1])} - 1)}{(s + \beta[2q - 1]) \prod_{k=0}^{2q-1}(s + k\beta)}.$$

Proof. Direct from replacing $k(t, \tau)$ in (34) for (8) and solving the integrals. □

In summary, the estimated transfer function of the Lebesgue-sampled continuous-time system of interest can be computed in a straightforward manner after the hyperparameter vector $\boldsymbol{\rho}$ and representer weighting vector $\hat{\mathbf{c}}$ are obtained. Both of these quantities have been proven to be computable from separate EM iterations in Theorems 3 and 1, respectively.

4.5. Algorithm

To conclude this section, the full algorithm for non-parametric identification of Lebesgue-sampled continuous-time systems is described in Algorithm 2. For simplicity we replace the hyperparameter γ for $\tilde{\gamma}$ in the description of the hyperparameter vector $\boldsymbol{\rho}$.

Algorithm 2 Kernel-based non-parametric identification for Lebesgue-sampled continuous-time systems

1: Input: $\mathbf{u}_{0:N-1}, \mathcal{Y}_{1:N}$, initial hyperparameter estimate $\hat{\boldsymbol{\rho}}^{(1)} = [\hat{\tilde{\gamma}}^{(1)}, \hat{\boldsymbol{\beta}}^{(1)\top}, \hat{\sigma}^{2(1)}]^\top$, maximum number of MAP-EM iterations M_{iter}
2: Form $\boldsymbol{\Phi}$ as in (26)
3: **for** $j = 1, 2, \ldots, M_{\text{iter}}$ **do**
4: Compute $\bar{\mathbf{Q}}^{(j)}$ from (24) using the minimax tilting algorithm in Botev (2017)
5: Factor $\bar{\mathbf{Q}}^{(j)} = \mathbf{C}^{(j)}\mathbf{C}^{(j)\top}$ and $\mathcal{O}_{\hat{\boldsymbol{\beta}}^{(j)}}/\hat{\tilde{\gamma}}^{(j)} = \mathbf{L}_{\hat{\boldsymbol{\rho}}^{(j)}}\mathbf{L}_{\hat{\boldsymbol{\rho}}^{(j)}}^\top$
6: Perform the QR factorization in (28)
7: Obtain $\hat{\boldsymbol{\rho}}^{(j+1)} = [\hat{\tilde{\gamma}}^{(j+1)}, \hat{\boldsymbol{\beta}}^{(j+1)}, \hat{\sigma}^{2(j+1)}]^\top$ from (30) and (31)
8: **end for**
9: Compute initial estimate $\hat{\mathbf{c}}^{(1)}$ from the midpoint approximation in (19)
10: **for** $j = 1, 2, \ldots, M_{\text{iter}}$ **do**
11: Obtain $\hat{\mathbf{c}}^{(j+1)}$ from (17) with $\mathbf{K}$, $\tilde{\gamma}$ and $\tilde{\mathbf{z}}^{(j)}$ computing using $\hat{\boldsymbol{\rho}}^{(M_{\text{iter}}+1)}$
12: **end for**
13: Output: estimated transfer function $\hat{G}(s) = \hat{\mathbf{c}}^{(M_{\text{iter}}+1)\top}\boldsymbol{\Phi}\mathcal{K}(s)$, with $\mathcal{K}(s)$ computed from (34) using $\hat{\boldsymbol{\beta}}^{(M_{\text{iter}}+1)}$.

Remark 8. Similarly as in lines 2 to 10 of Algorithm 1, instead of performing a fixed number of iterations, the iterations could be stopped after a stopping criterion is satisfied (line 6 in Algorithm 1). In case of the loop in lines 3 to 8 in Algorithm 2, this stopping criterion is defined as $\|\hat{\boldsymbol{\rho}}^{(j+1)} - \hat{\boldsymbol{\rho}}^{(j)}\|_2/\|\hat{\boldsymbol{\rho}}^{(j)}\|_2 < \epsilon$, while in case of the loop in lines 10 to 12 in Algorithm 2, it is defined as $\|\hat{\mathbf{c}}^{(j+1)} - \hat{\mathbf{c}}^{(j)}\|_2/\|\hat{\mathbf{c}}^{(j)}\|_2 < \epsilon$, where the values of ϵ could be different in each stopping criterion.

5. Simulations

The performance of the novel non-parametric estimator is tested on a series of extensive Monte Carlo simulations.

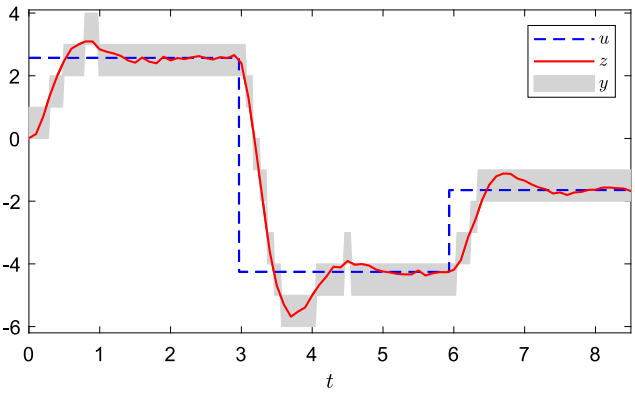

Fig. 3. Input and output signals of the system (35) corresponding to 8 [s] of one Monte Carlo run.

5.1. Practically relevant example

We consider a mass–spring–damper system with transfer function given by

$$G(s) = \frac{1}{\mathrm{m}s^2 + \mathrm{d}s + \mathrm{k}}, \tag{35}$$

with mass m = 0.05 [kg], damping coefficient d = 0.2 [Ns/m], and spring constant k = 1 [N/m]. The output is sensed with period $\Delta = 0.1$ [s], and $h = 1$ [m]. The input is a Gaussian white noise sequence of standard deviation 5 [N] passed through a zero-order hold device with period $\Delta_u = 3$ [s]. The output prior to the Lebesgue sampling is computed using the `lsim` command in MATLAB with sampling time 0.1[s], which delivers exact noiseless output values since the input is a zero-order hold signal. One hundred Monte Carlo runs are performed with a varying input and an additive Gaussian white noise prior to the Lebesgue sampling with standard deviation 0.05[m]. Each run has a total time duration of 30 [s] (i.e., 300 data points are sensed prior to Lebesgue sampling), and on average $N_{\text{L}} = 69$ output samples are obtained after sampling per run.

Three estimators are tested: the kernel-based continuous-time non-parametric estimator with equidistantly-sampled data (Pillonetto & De Nicolao, 2010; Scandella et al., 2022) using the stable-spline kernel of order 1 and the midpoint estimate $z(i\Delta) \approx \eta_i + h/2$ as output data ($\hat{g}_{\text{rie}}$), this same estimator but using the noisy output $z(i\Delta)$ prior to Lebesgue sampling as output data ($\hat{g}_{\text{or}}$), and the proposed approach (Algorithm 2 of this paper, $\hat{g}_{\text{leb}}$). Note that the oracle estimator $\hat{g}_{\text{or}}$ cannot be implemented in practice, since we do not have direct knowledge of the system output before the event-sampler. This estimator is different from the commonly-denominated oracle estimator that uses the unattainable kernel $k(\tau, \xi) = g(\tau)g(\xi)$ (Chen, Ohlsson, & Ljung, 2012). We measure the performance of each estimator with the fit metric

$$\text{fit} = 100\left(1 - \frac{\|\hat{\mathbf{x}}^j - \mathbf{x}\|_2}{\|\mathbf{x} - \bar{x}\mathbf{1}\|_2}\right),$$

where $\mathbf{x}$ is the noiseless output sequence (prior to Lebesgue sampling), $\hat{\mathbf{x}}^j$ is the simulated output sequence using the jth impulse response estimate, and $\bar{x}$ is the mean value of $\mathbf{x}$. The proposed estimator uses the stable-spline kernel of order 1 with a maximum number of EM iterations $M_{\text{iter}} = 40$, and 1000 samples of a multivariate truncated Gaussian distribution are obtained to compute $\bar{\mathbf{Q}}^{(j)}$ in (24).

A typical data set is shown in Fig. 3. Note that the task of the proposed estimator is particularly challenging, since the overshoot of the output signal z is rarely captured in the y signal band

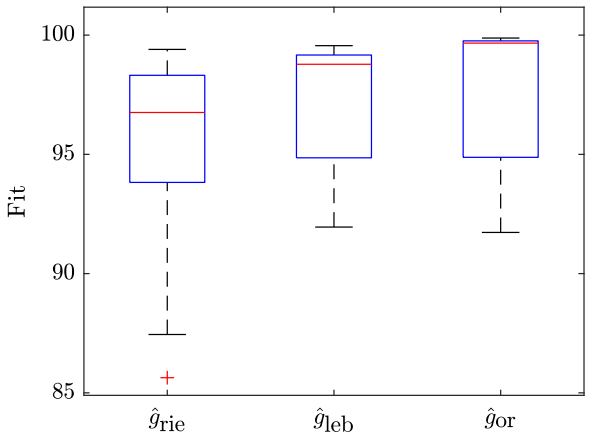

Fig. 4. Boxplots of the fit metric for the case study, Section 5.1. The Lebesgue-sampling-based estimator $\hat{g}_{\text{leb}}$ achieves a better performance than the Riemann approach $\hat{g}_{\text{rie}}$.

due to the coarse grid produced by the threshold level h. To show the statistical performance of each estimator, we present the boxplots of the fit metric for each estimator in Fig. 4. A graphical illustration of the proximity of the estimated frequency responses to the frequency response of the true system is presented in Fig. 5, which shows 20 Bode magnitude plots of the frequency response estimates (obtained via Corollary 2) of each method, obtained from 20 noise realizations. As expected, the proposed approach achieves on average a better fit than the estimator that only uses the midpoint values $\eta_i + h/2$ as output. The $\hat{g}_{\text{leb}}$ estimator is only slightly outperformed by the oracle estimator, despite having a low resolution for the output measurement mechanism and a 77% reduction in output data samples on average.

Remark 9. An additional test has been conducted to assess the necessity of EM iterations for computing the optimal weights $\hat{\mathbf{c}}$. Under the same experimental conditions as above, we have compared the fit of the Lebesgue approach employing the initial estimate (19) for the weight vector against the fit achieved with the estimator computed from the EM iterations outlined in Theorem 1. We have observed that incorporating EM iterations for the weight vector has led to a better fit in 96 out of 100 Monte Carlo runs. This suggests that performing EM iterations for computing the weight vector is crucial for achieving the best performance.

5.2. Effect of the threshold amplitude h

The threshold amplitude plays an important role in the accuracy of any system identification method, since it is directly related to the size of the set uncertainty of the output measurement. The system in (35) is identified under the same experimental conditions as Section 5.1, but now with $0.1[m]$ as standard deviation of the additive noise. Six different values of h are tested, and for each value, one hundred Monte Carlo runs are recorded.

The boxplots in Fig. 6 show that the performance of the standard (Riemann) non-parametric estimator severely deteriorates as the threshold amplitude h grows. In sharp contrast, the proposed estimator remains accurate even when h is large compared to the amplitude range of the unsampled output. In Table 1, we have registered the average number of effective samples that are obtained for each simulation study. These numbers confirm the advantage of Lebesgue sampling over equidistant sampling in terms of resource efficiency, since the correct utilization of the set-uncertainty in the Lebesgue sampling strategy can lead to a

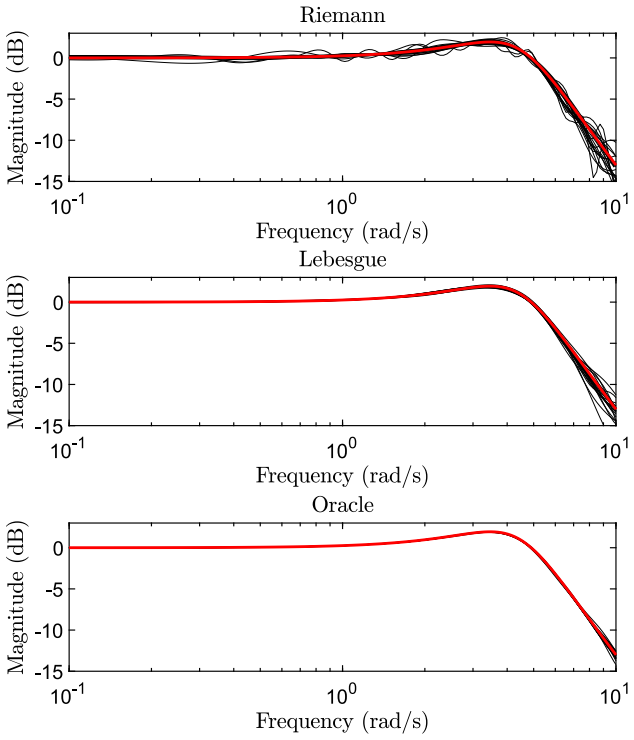

Fig. 5. Bode magnitude plots of 20 Monte Carlo runs (black), compared to the true frequency response (red). Upper plot: equidistantly-sampled approach (Pillonetto & De Nicolao, 2010); middle plot: proposed method; lower plot: oracle method (unattainable). The Bode plots of the Lebesgue-sampling approach, obtained via Lemma 5, show much less variability than the Riemann approach over the Monte Carlo runs, and are comparable to the estimates produced by the oracle method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Average number of output samples retrieved from the Monte Carlo experiments for each threshold distance h. For reference, the number of samples for the equidistantly-sampled estimator is 300.

h [m]	1	1.2	1.5	1.8	2	2.5
Samples	79.5	69.2	59.7	51.7	47.5	38.5

sevenfold reduction in output data used in the identification process (from 300 to 38.5) with only minor performance detriment compared to Riemann sampling with $h = 1$.

5.3. Other benchmark systems

To show that the proposed estimator also performs well under different system setups, the next tests consider three more systems:

$$G_{\text{A}}(s) = \frac{-6400s + 1600}{s^4 + 5s^3 + 408s^2 + 416s + 1600},$$

$$G_{\text{B}}(s) = \frac{27}{20}\frac{-2000s^3 - 3600s^2 - 2095s - 396}{1350s^4 + 7695s^3 + 12852s^2 + 7796s + 1520},$$

$$G_{\text{C}}(s) = \frac{-3.025s^3 - 15.676s^2 - 32.802s - 88.827}{s^4 + 16.52s^3 + 65.534s^2 + 235.01 + 292.948},$$

all of which have been used as benchmarks in other works on continuous-time system identification methods (Scandella et al., 2022). In particular, the Rao–Garnier system ($G_{\text{A}}(s)$ in this work) has been tested in numerous works (Garnier, 2015; Ljung, 2009; Rao & Garnier, 2002), and is particularly challenging to identify due to its damped step response and stiffness. All systems have

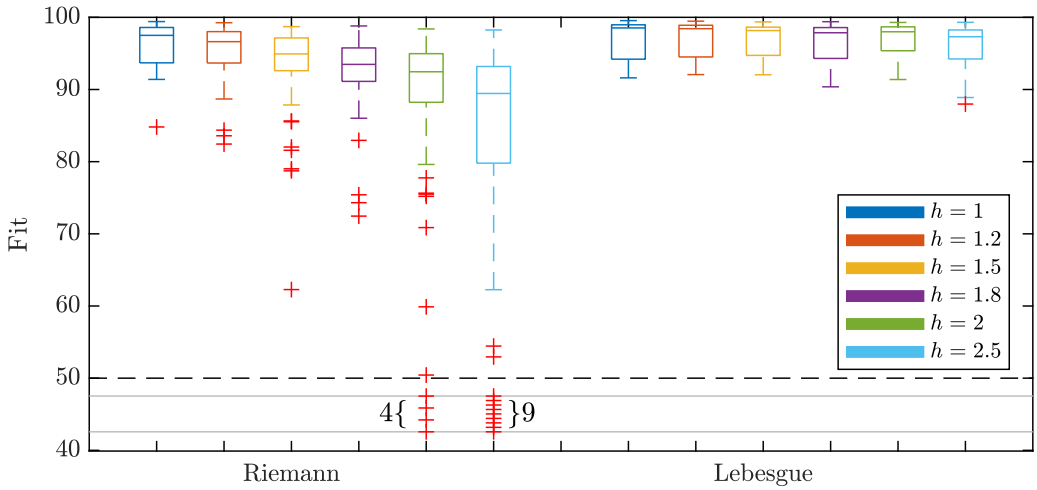

Fig. 6. Boxplots of the fit metric for different values of threshold amplitude h, Section 5.2. Riemann sampling (left), Lebesgue sampling (right). While the estimator using the Riemann-sampling approach severely deteriorates its performance for coarser threshold grids, the proposed method produces excellent results for all values of h in this study.

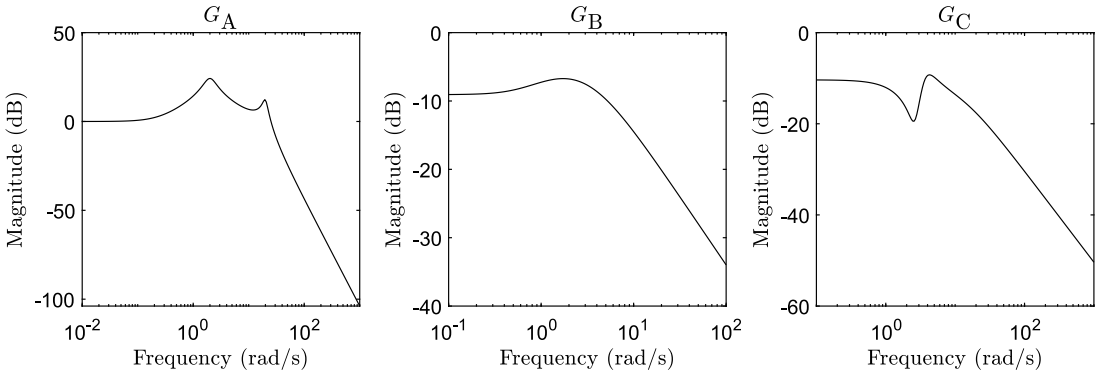

Fig. 7. Bode magnitude plots of the three systems in Section 5.3. From left to right: $G_A(s)$, $G_B(s)$ and $G_C(s)$.

Table 2
Experimental conditions for each system studied in Section 5.3.

	Δ	h	σ	SNR [dB]
$G_A(s)$	0.01	2.5	0.3	28.79
$G_B(s)$	0.03	0.2	0.03	22.82
$G_C(s)$	0.03	0.2	0.03	17.69

been excited by a Gaussian white noise of unit variance passed through a ZOH with period $\Delta_u = 3$[s]. The Bode plots of these systems are given in Fig. 7, and the experimental conditions that are tested can be found in Table 2, where we have also included the signal to noise ratio (SNR) between the output previous to Lebesgue sampling, z, and the additive noise, v. In Fig. 8, we compare the fit metric of the proposed estimator to the Riemann and oracle estimators described in Section 5.1 using 100 Monte Carlo runs. The results show that the Lebesgue sampling-based estimator outperforms the approach with equidistant sampling in all the systems considered in this study. Note that although the experimental conditions for $G_A(s)$ give a better SNR, the performance is affected by a large threshold amplitude h compared to the other cases.

6. Conclusions

The approach developed in this paper allows one to accurately identify Lebesgue-sampled systems based on input and output data. The main idea is to use all the available information for identification and control when dealing with Lebesgue-sampled signals. The proposed identification method, which is inspired by MAP estimation, kernel methods, and the EM algorithm, exploits the set uncertainty information in the output measurements to deliver more accurate models than the Riemann-sampling approach, while needing much fewer output samples. Thus, our method can enable systems with incremental encoders or with intermittent observations to be operated over less stringent sampling conditions (i.e., larger threshold amplitudes) without a severe loss in modeling accuracy. We have confirmed the advantages of the proposed algorithm in terms of statistical performance and resource efficiency in a series of extensive Monte Carlo simulations.

Acknowledgment

This work is part of the research program VIDI with project number 15698, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).

Appendix

A.1. Proof of Lemma 2

Proof. The MAP estimator for $\mathbf{c}$ is computed by

$$\hat{\mathbf{c}}_{\text{MAP}} = \arg\max_{\mathbf{c}\in\mathbb{R}^N} \mathrm{p}(\mathcal{Y}_{1:N}|\mathbf{c})\,\mathrm{p}(\mathbf{c})$$

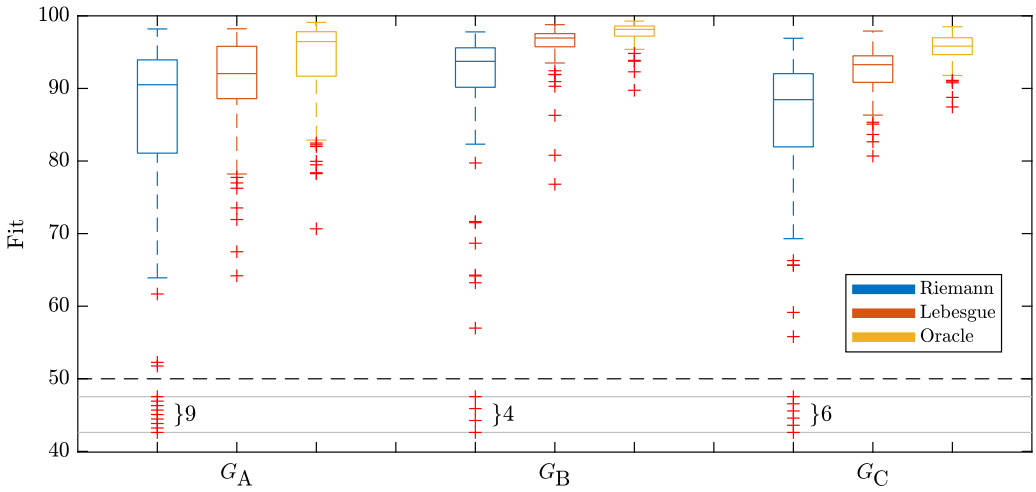

Fig. 8. Boxplots of the fit metric for the three systems in Section 5.3. The proposed Lebesgue-sampling approach leads to an important gain in model fit compared to the Riemann approach in all the benchmark systems of this study.

$$= \arg\max_{\mathbf{c}\in\mathbb{R}^N} \left(-\frac{N}{2}\log(2\pi\sigma^2) - \frac{\log\det(2\pi\mathbf{K}^{-1}/\gamma)}{2} + \sum_{i=1}^{N} \log\left[\int_{\eta_i}^{\eta_i+h} e^{\frac{-1}{2\sigma^2}\left(z_i-\mathbf{K}_i^\top\mathbf{c}\right)^2} \mathrm{d}z_i\right] - \frac{\gamma\mathbf{c}^\top\mathbf{K}\mathbf{c}}{2}\right), \tag{36}$$

where we have used the same derivation as for $\ell(\mathbf{x}_1)$ in Section 4.1 for computing the log-likelihood term. By comparing (36) to (13), we find that $\hat{\mathbf{c}}$ in (13) is simply the maximum*a posteriori* estimate of $\mathbf{c}$ within the model in (14). □

A.2. Proof of Theorem 1

Proof. Since the Q function provided by Lemma 3 is concave in $\mathbf{c}$, it is sufficient to obtain the point(s) which make the gradient of the objective function equal to zero. The gradient of $Q(\mathbf{c}, \hat{\mathbf{c}}^{(j)}) - \gamma\mathbf{c}^\top\mathbf{K}\mathbf{c}/2$ is given by

$$\frac{\partial}{\partial\mathbf{c}}\left(Q(\mathbf{c},\hat{\mathbf{c}}^{(j)}) - \frac{\gamma\mathbf{c}^\top\mathbf{K}\mathbf{c}}{2}\right) = \frac{-1}{\sigma^2}\sum_{i=1}^{N}\int_{\eta_i}^{\eta_i+h}\mathbf{K}_i(\mathbf{K}_i^\top\mathbf{c}-z_i)\mathrm{p}(z_i|y(i\Delta),\hat{\mathbf{c}}^{(j)})\mathrm{d}z_i - \gamma\mathbf{K}\mathbf{c}.$$

Setting the gradient to zero yields

$$\hat{\mathbf{c}}^{(j+1)} = \left(\sum_{i=1}^{N}\mathbf{K}_i\mathbf{K}_i^\top + \gamma\sigma^2\mathbf{K}\right)^{-1}\sum_{i=1}^{N}\mathbf{K}_i\tilde{z}_i^{(j)}, \tag{37}$$

where we have defined the conditional mean $\tilde{z}_i^{(j)}$ as

$$\tilde{z}_i^{(j)} = \int_{\eta_i}^{\eta_i+h} z_i\mathrm{p}(z_i|y(i\Delta),\hat{\mathbf{c}}^{(j)})\mathrm{d}z_i, \tag{38}$$

and where we have used the fact that, for all $i = 1, 2, \ldots, N$,

$$\int_{\eta_i}^{\eta_i+h}\mathrm{p}(z_i|y(i\Delta),\hat{\mathbf{c}}^{(j)})\mathrm{d}z_i = 1.$$

The iterations in (17) are obtained from (37) by rewriting the sum related to $\tilde{z}_i^{(j)}$ conveniently and using the fact that

$$\sum_{i=1}^{N}\mathbf{K}_i\mathbf{K}_i^\top = \mathbf{K}^2,$$

which holds since $\mathbf{K}$ is symmetric. Finally, the explicit expression for $\tilde{z}_i^{(j)}$ in (18) can be obtained directly from expanding the following alternative expression for (18) based on applying Bayes' theorem on the conditional expectation in (38):

$$\tilde{z}_i^{(j)} = \frac{\int_{\eta_i}^{\eta_i+h} z_i\exp\left(\frac{-1}{2\sigma^2}[z_i-\mathbf{K}_i^\top\hat{\mathbf{c}}^{(j)}]^2\right)\mathrm{d}z_i}{\int_{\eta_i}^{\eta_i+h}\exp\left(\frac{-1}{2\sigma^2}[z_i-\mathbf{K}_i^\top\hat{\mathbf{c}}^{(j)}]^2\right)\mathrm{d}z_i}. \quad \square$$

A.3. Proof of Theorem 2

Proof. We seek to derive the EM iterations for computing the maximum likelihood estimate in (20). By setting the latent variable as $\mathbf{z}_{1:N}$, we must compute the following Q function

$$Q(\boldsymbol{\rho},\hat{\boldsymbol{\rho}}^{(j)}) = \mathbb{E}\left\{\log\mathrm{p}(\mathbf{z}_{1:N},\mathcal{Y}_{1:N}|\boldsymbol{\rho})|\mathcal{Y}_{1:N},\hat{\boldsymbol{\rho}}^{(j)}\right\},$$

where it can be shown that (cf. Eq. (19) of Godoy et al. (2011))

$$\mathrm{p}(\mathbf{z}_{1:N},\mathcal{Y}_{1:N}|\boldsymbol{\rho}) = \begin{cases}\mathrm{p}(\mathbf{z}_{1:N}|\boldsymbol{\rho}) & \text{if } \mathbf{z}_{1:N}\in\mathcal{Y}_{1:N},\\ 0 & \text{otherwise,}\end{cases}$$

which, by exploiting (21), leads to

$$-2Q(\boldsymbol{\rho},\hat{\boldsymbol{\rho}}^{(j)}) = \log\det(2\pi\mathbf{S}_{\boldsymbol{\rho}}) + \mathbb{E}\{\mathbf{z}_{1:N}^\top\mathbf{S}_{\boldsymbol{\rho}}^{-1}\mathbf{z}_{1:N}|\mathcal{Y}_{1:N},\hat{\boldsymbol{\rho}}^{(j)}\}.$$

The iterations in (23) follow from applying the commutativity property of the trace function to the expectation above.

The minimization of $-2Q(\boldsymbol{\rho},\hat{\boldsymbol{\rho}}^{(j)})$ with respect to $\boldsymbol{\rho}$ provides the M-step of the EM iterations for computing a maximum of the likelihood of interest, which in turn is equivalent to solving (locally or globally) the optimization problem in (22). □

A.4. Proof of Lemma 4

Proof. Consider the zero-order hold representation (valid for $t\in[0,\Delta N)$),

$$u(t) = \sum_{k=0}^{N-1}u(k\Delta)\mathbb{1}(\Delta k\le t<\Delta[k+1]), \tag{39}$$

with $\mathbb{1}(\cdot)$ being the indicator function (i.e., 1 if $(\cdot)$ is satisfied, and 0 otherwise). Thus, we compute

$$u(i\Delta-\xi)u(j\Delta-\tau) = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1}u(k\Delta)u(l\Delta) \times\mathbb{1}(\Delta[i-k-1]<\tau\le\Delta[i-k]\wedge\Delta[j-l-1]<\xi\le\Delta[j-l]). \tag{40}$$

Since the integral of interest ranges from $0 < \tau, \xi < \infty$, the elements of (40) for $i - k \leq 0$ and $j - l \leq 0$ can be discarded. In other words, within the domain of integration, we can write $u(i\Delta - \xi)u(j\Delta - \tau)$ as (40) but with summation upper limits $i-1$ and $j-1$ instead of $N-1$, respectively. Thus, interchanging summation and integration yields

$$\mathbf{K}_{ij} = \sum_{k=0}^{i-1}\sum_{l=0}^{j-1} u(k\Delta)u(l\Delta)\int_{\Delta[i-k-1]}^{\Delta[i-k]}\int_{\Delta[j-l-1]}^{\Delta[j-l]} k(\xi,\tau)\mathrm{d}\tau\mathrm{d}\xi.$$

Alternatively, we can write this entry of the kernel matrix as $\mathbf{U}_j^\top \mathcal{O}_{\boldsymbol{\beta}}\mathbf{U}_i$, where $\mathbf{U}_j$ and $\mathbf{U}_i$ are the jth and ith columns of $\boldsymbol{\Phi}^\top$, respectively, and $\mathcal{O}_{\boldsymbol{\beta}}$ has entries that are given by (27). Since $\mathcal{O}_{\boldsymbol{\beta}}$ does not depend on i nor j, it is possible to describe the complete matrix $\mathbf{K}$ by stacking the column vectors $\mathbf{U}_j$ and $\mathbf{U}_i$, leading to (25). □

A.5. Proof of Theorem 3

Proof. Let us first rewrite the log det term in (23). Thanks to the Weinstein–Aronszajn identity (Horn & Johnson, 2012, 1.3.P28), we have

$$\begin{aligned}\log\det(\mathbf{S}_\rho) &= \log\det(\sigma^2\mathbf{I}) + \log\det(\boldsymbol{\Phi}\mathbf{L}_\rho\mathbf{L}_\rho^\top\boldsymbol{\Phi}^\top + \mathbf{I})\\ &= N\log\sigma^2 + \log\det(\mathbf{L}_\rho^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\mathbf{L}_\rho + \mathbf{I})\\ &= N\log\sigma^2 + 2\log\det(\mathbf{R}_{1,\rho}),\end{aligned} \tag{41}$$

where the identity in (29a) has been used in the last step. We now study the trace term in (23). Note that, by the matrix inversion lemma,

$$\mathbf{S}_\rho^{-1} = \sigma^{-2}\mathbf{I} - \sigma^{-2}\boldsymbol{\Phi}\mathbf{L}_\rho(\mathbf{L}_\rho^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\mathbf{L}_\rho + \mathbf{I})^{-1}\mathbf{L}_\rho^\top\boldsymbol{\Phi}^\top,$$

which leads to

$$\begin{aligned}\mathrm{tr}\{\mathbf{S}_\rho^{-1}\bar{\mathbf{Q}}^{(j)}\} &= \mathrm{tr}\{{\mathbf{C}^{(j)}}^\top\mathbf{S}_\rho^{-1}\mathbf{C}^{(j)}\}\\ &= \frac{\mathrm{tr}\{\bar{\mathbf{Q}}^{(j)}\}}{\sigma^2} - \frac{\mathrm{tr}\{{\mathbf{C}^{(j)}}^\top\boldsymbol{\Phi}\mathbf{L}_\rho(\mathbf{L}_\rho^\top\boldsymbol{\Phi}^\top\boldsymbol{\Phi}\mathbf{L}_\rho + \mathbf{I})^{-1}\mathbf{L}_\rho^\top\boldsymbol{\Phi}^\top\mathbf{C}^{(j)}\}}{\sigma^2}.\end{aligned}$$

This expression, when written in terms of $\mathbf{R}_{1,\rho}$ and $\mathbf{R}_{2,\rho}$ via (29a) and (29b), is simply

$$\begin{aligned}\mathrm{tr}\{\mathbf{S}_\rho^{-1}\bar{\mathbf{Q}}^{(j)}\} &= \frac{\mathrm{tr}\{\bar{\mathbf{Q}}^{(j)} - \mathbf{R}_{2,\rho}^\top\mathbf{R}_{2,\rho}\}}{\sigma^2}\\ &= \frac{\|\mathbf{C}^{(j)}\|_{\mathrm{F}}^2 - \|\mathbf{R}_{2,\rho}\|_{\mathrm{F}}^2}{\sigma^2},\end{aligned} \tag{42}$$

where we have used the definition of the Frobenius norm in the last line. By combining the results in (41) and (42), we reach

$$\begin{aligned}\hat{\boldsymbol{\rho}}^{(j+1)} = \arg\min_{\boldsymbol{\rho}\in\boldsymbol{\Gamma}} \Bigg(& N\log\sigma^2\\ & + 2\log\det(\mathbf{R}_{1,\rho}) + \frac{\|\mathbf{C}^{(j)}\|_{\mathrm{F}}^2 - \|\mathbf{R}_{2,\rho}\|_{\mathrm{F}}^2}{\sigma^2}\Bigg).\end{aligned} \tag{43}$$

Since both $\mathbf{R}_{1,\rho}$ and $\mathbf{R}_{2,\rho}$ depend on $\mathbf{L}_\rho$, which in turn is already factored by a scalar variable $1/\gamma$, the dependence on σ in the $\mathbf{R}$ matrices is redundant for the optimization above. Therefore, we can concentrate the cost function by minimizing (43) over σ^2 first, which leads to (31). Replacing (31) in (43) and neglecting constant terms leads to (30), which concludes the proof. □

A.6. Proof of Lemma 5

Proof. Under the zero-order hold intersample behavior assumption, the input description in (39) permits rewriting the convolution in (33) as

$$\hat{G}_l(s) = \sum_{k=0}^{l-1} u(k\Delta)\int_{\Delta[l-k-1]}^{\Delta[l-k]} K(s;\tau)\mathrm{d}\tau.$$

Interchanging the integrals above leads to $\hat{G}_l(s)$ being equal to the lth row of $\boldsymbol{\Phi}$ multiplied by $\mathcal{K}$ defined in (34). This fact, together with the representer theorem description (32), leads to the desired result. □

References

Agüero, J. C., González, K., & Carvajal, R. (2017). EM-based identification of ARX systems having quantized output data. *IFAC-PapersOnLine*, *50*(1), 8367–8372.

Aravkin, A. Y., Bell, B. M., Burke, J. V., & Pillonetto, G. (2014). The connection between Bayesian estimation of a Gaussian random field and RKHS. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(7), 1518–1524.

Arriagada, I. A., & Yuz, J. I. (2008). On the relationship between splines, sampling zeros and numerical integration in sampled-data models for linear systems. In *2008 American Control Conference* (pp. 3665–3670). IEEE.

Åström, K. J., & Bernhardsson, B. (2003). Systems with Lebesgue sampling. In *Directions in mathematical systems theory and optimization* (pp. 1–13). Springer.

Bogachev, V. I. (1998). *Gaussian measures*, (no. 62), American Mathematical Society.

Botev, Z. I. (2017). The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *79*(1), 125–148.

Bottegal, G., Hjalmarsson, H., & Pillonetto, G. (2017). A new kernel-based approach to system identification with quantized output data. *Automatica*, *85*, 145–152.

Chen, T., & Ljung, L. (2013). Implementation of algorithms for tuning parameters in regularized least squares problems in system identification. *Automatica*, *49*(7), 2213–2220.

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

Chen, T., Zhao, Y., & Ljung, L. (2012). Impulse response estimation with binary measurements: A regularized FIR model approach. *IFAC Proceedings Volumes*, *45*(16), 113–118.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *39*(1), 1–22.

Diao, J.-D., Guo, J., & Sun, C.-Y. (2018). Event-triggered identification of FIR systems with binary-valued output observations. *Automatica*, *98*, 95–102.

Dinuzzo, F. (2015). Kernels for linear time invariant system identification. *SIAM Journal on Control and Optimization*, *53*(5), 3299–3317.

Dinuzzo, F., & Schölkopf, B. (2012). The representer theorem for Hilbert spaces: a necessary and sufficient condition. *Advances in Neural Information Processing Systems*, *25*.

Garnier, H. (2015). Direct continuous-time approaches to system identification. Overview and benefits for practical applications. *European Journal of Control*, *24*, 50–62.

Garnier, H., & Young, P. C. (2014). The advantages of directly identifying continuous-time transfer function models in practical applications. *International Journal of Control*, *87*(7), 1319–1338.

Godoy, B. I., Goodwin, G. C., Agüero, J. C., Marelli, D., & Wigren, T. (2011). On identification of FIR systems having quantized output data. *Automatica*, *47*(9), 1905–1915.

González, R. A., Rojas, C. R., & Hjalmarsson, H. (2021). Non-causal regularized least-squares for continuous-time system identification with band-limited input excitations. In *Proceedings of the 60th IEEE Conference on Decision and Control* (pp. 114–119).

González, R. A., Rojas, C. R., Pan, S., & Welsh, J. S. (2021). The SRIVC algorithm for continuous-time system identification with arbitrary input excitation in open and closed loop. In *Proceedings of the 60th IEEE Conference on Decision and Control* (pp. 3004–3009).

González, R. A., Tiels, K., & Oomen, T. (2023). Identifying Lebesgue-sampled continuous-time impulse response models: A kernel-based approach. In *IFAC World Congress on Automatic Control*.

Horn, R. A., & Johnson, C. R. (2012). *Matrix Analysis* (2nd edn). Cambridge University Press.

Kawaguchi, T., Hikono, S., Maruta, I., & Adachi, S. (2016). System identification under Lebesgue sampling and its asymptotic property. In *Proceedings of the 55th IEEE Conference on Decision and Control* (pp. 2079–2084).

Kimeldorf, G. S., & Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, *41*(2), 495–502.

Kon, J., Strijbosch, N., Koekebakker, S., & Oomen, T. (2021). Intermittent sampling in repetitive control: exploiting time-varying measurements. In *Proceedings of the 60th IEEE Conference on Decision and Control* (pp. 6566–6571).

Liu, Q., Wang, Z., He, X., & Zhou, D. (2014). A survey of event-based strategies on control and estimation. *Systems Science & Control Engineering: An Open Access Journal*, *2*(1), 90–97.

Ljung, L. (2009). Experiments with identification of continuous time models. In *15th IFAC Symposium on System Identification, vol. 42, no. 10* (10), (pp. 1175–1180). Elsevier.
McLachlan, G. J., & Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley & Sons.
Merry, R. J. E., van de Molengraft, M. J. G., & Steinbuch, M. (2013). Optimal higher-order encoder time-stamping. *Mechatronics*, *23*(5), 481–490.
Piga, D., Mejari, M., & Forgione, M. (2021). Learning dynamical systems from quantized observations: a Bayesian perspective. *IEEE Transactions on Automatic Control*.
Pillonetto, G., Chen, T., Chiuso, A., De Nicolao, G., & Ljung, L. (2022). *Regularized System Identification*. Springer.
Pillonetto, G., & Chiuso, A. (2015). Tuning complexity in regularized kernel-based regression and linear system identification: The robustness of the marginal likelihood estimator. *Automatica*, *58*, 106–117.
Pillonetto, G., & De Nicolao, G. (2010). A new kernel-based approach for linear system identification. *Automatica*, *46*(1), 81–93.
Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., & Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, *50*(3), 657–682.
Pouliquen, M., Goudjil, A., Gehan, O., & Pigeon, E. (2016). Continuous-time system identification using binary measurements. In *Proceedings of the 55th IEEE conference on decision and control* (pp. 3787–3792).
Pouliquen, M., Pigeon, E., Gehan, O., & Goudjil, A. (2019). Identification using binary measurements for IIR systems. *IEEE Transactions on Automatic Control*, *65*(2), 786–793.
Rao, G. P., & Garnier, H. (2002). Numerical illustrations of the relevance of direct continuous-time model identification. *35*, In *15th triennial IFAC world congress on automatic control, vol. 35, no. 1* (1), (pp. 133–138).
Risuleo, R. S., Bottegal, G., & Hjalmarsson, H. (2019). Identification of linear models from quantized data: a midpoint-projection approach. *IEEE Transactions on Automatic Control*, *65*(7), 2801–2813.
Scandella, M., Mazzoleni, M., Formentin, S., & Previdi, F. (2022). Kernel- based identification of asymptotically stable continuous-time linear dynamical systems. *International Journal of Control*, *95*(6), 1668–1681.
Schölkopf, B., Herbrich, R., & Smola, A. J. (2001). A generalized representer theorem. In *International conference on computational learning theory* (pp. 416–426).
Schoukens, J., Pintelon, R., & Van Hamme, H. (1994). Identification of linear dynamic systems using piecewise constant excitations: use, misuse and alternatives. *Automatica*, *30*(7), 1153–1169.
Strijbosch, N., & Oomen, T. (2019). Beyond quantization in iterative learning control: Exploiting time-varying time-stamps. In *IEEE American control conference* (pp. 2984–2989).
Strijbosch, N., & Oomen, T. (2022). Iterative learning control for intermittently sampled data: Monotonic convergence, design, and applications. *Automatica*, *139*, Article 110171.
Wahba, G. (1990). *Spline models for observational data*. SIAM.
Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, 95–103.

Rodrigo A. González was born in Viña del Mar, Chile, in 1992. He earned the professional title of Ingeniero Civil Electrónico and a Masters degree in Electronic Engineering from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2016, obtaining the Best Electronic Engineering Student Award. He received his Ph.D. in Electrical Engineering from KTH Royal Institute of Technology, Stockholm, Sweden, in 2022. Since June 2022, he has been affiliated with the Department of Mechanical Engineering at Eindhoven University of Technology, where he currently serves as an Assistant Professor in the Control Systems Technology Section, starting in March 2024. His research focuses on the design and analysis of data-driven modeling, estimation, and control methods, with applications in the high-tech precision industry.

Koen Tiels received the MSc. degree in Electromechanical Engineering in 2010 and the PhD. degree in Engineering in 2015, both from the Vrije Universiteit Brussel (VUB). After post-doc positions at the VUB and Uppsala University, he joined the Eindhoven University of Technology (TU/e) in 2020. Currently, he is a University Lecturer at the department of Mechanical Engineering at the TU/e. His main interests are in the field of nonlinear system identification and machine learning.

Tom Oomen is full professor with the Department of Mechanical Engineering at the Eindhoven University of Technology. He is also a part-time full professor with the Delft University of Technology. He received the M.Sc. degree (cum laude) and Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands. He held visiting positions at KTH, Stockholm, Sweden, and at The University of Newcastle, Australia. He is a recipient of the 7th Grand Nagamori Award, the Corus Young Talent Graduation Award, the IFAC 2019 TC 4.2 Mechatronics Young Research Award, the 2015 IEEE Transactions on Control Systems Technology Outstanding Paper Award, the 2017 IFAC Mechatronics Best Paper Award, the 2019 IEEJ Journal of Industry Applications Best Paper Award, and recipient of a Veni and Vidi personal grant. He is currently a Senior Editor of IEEE Control Systems Letters (L-CSS) and Co-Editor-in-Chief of IFAC Mechatronics, and he has served on the editorial board of IEEE Transactions on Control Systems Technology. He has also been vice-chair for IFAC TC 4.2 and a member of the Eindhoven Young Academy of Engineering. His research interests are in the field of data-driven modeling, learning, and control, with applications in precision mechatronics.