Kernel-based identification with frequency domain side-information

Khosravi, Mohammad; Smith, Roy S.

DOI
10.1016/j.automatica.2022.110813

Publication date
2023

Document Version
Final published version

Published in
Automatica

Automatica 150 (2023) 110813

Contents lists available at ScienceDirect

Automatica

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

Kernel-based identification with frequency domain side-information[☆]

Mohammad Khosravi[a,*], Roy S. Smith[b]

[a] Delft Center for Systems and Control, Delft University of Technology, Netherlands
[b] Automatic Control Laboratory, ETH Zürich, Switzerland

ARTICLE INFO

Article history:
Received 2 November 2021
Received in revised form 24 April 2022
Accepted 23 November 2022
Available online xxxx

Keywords:
System identification
Side-information
Kernel-based methods
Frequency domain properties
Optimization

ABSTRACT

This paper discusses the problem of system identification when frequency domain side-information is available. We mainly consider the case where the side-information is provided as the $\mathcal{H}_\infty$-norm of the system being bounded by a given scalar. This framework allows considering different forms of frequency domain side-information, such as the dissipativity of the system. We propose a nonparametric identification approach for estimating the impulse response of the system under the given side-information. The estimation problem is formulated as a constrained optimization in a stable reproducing kernel Hilbert space, where suitable constraints are considered for incorporating the desired frequency domain features. The resulting optimization has an infinite-dimensional feasible set with an infinite number of constraints. We show that this problem is a well-defined convex program with a unique solution. We propose a heuristic that tightly approximates this unique solution. The proposed approach is equivalent to solving a finite-dimensional convex quadratically constrained quadratic program. The efficiency of the discussed method is verified by several numerical examples.

1. Introduction

System identification problem, initially introduced in Zadeh (1956), deals with the theory and techniques for estimating suitable mathematical models describing dynamical systems using measurement data. The topic has received substantial attention due to its broad applicability in numerous phenomena in science and technology (Ljung, 1999, 2010; Schoukens & Ljung, 2019). In many situations, identifying a dynamical system goes beyond fitting a mathematical model to the input–output measurement data. We may additionally need to integrate a specific attribute or a known feature of the system into the model. This side-information is possibly provided from our general understanding of the behavior of the system based on its inherent physical nature, or acquired from qualitative characteristics and phenomena observed from historical or experimental data. For example, in the identification of nonlinear dynamics various forms of side-information such as stability, region of attraction, dissipativity and many others are incorporated (Ahmadi & El Khadir, 2020; Hara, Inoue, & Sebe, 2019; Khosravi & Smith, 2021a, 2021e; Umenberger & Manchester, 2018).

The incorporation of side-information has been considered in the identification of linear dynamics, e.g., the low complexity of the model is imposed by considering sparsity promoting regularizations (Khosravi, Iannelli, et al., 2020; Khosravi, Yin, et al., 2020; Pillonetto, Chen, Chiuso, Nicolao, & Ljung, 2016; Shah, Bhaskar, Tang, & Recht, 2012; Smith, 2014). For penalizing the order of systems, the rank and the nuclear norm of the corresponding Hankel matrix are utilized in Fazel, Pong, Sun, and Tseng (2013), Mohan and Fazel (2010), Pillonetto et al. (2016) and Smith (2014). For the same purpose, the notion of atomic transfer functions and regularization based on the atomic norm are employed (Khosravi, Yin, Iannelli, Parsi & Smith, 2020; Shah et al., 2012). Identification with the side-information involving positivity features of the system such as compartmental structure and the internal or external positivity are also discussed in the literature (Benvenuti, De Santis, & Farina, 2002; De Santis & Farina, 2002; Grussler, Umenberger, & Manchester, 2017; Umenberger & Manchester, 2016). Other forms of side-information are studied in Abe, Inoue, and Adachi (2016), Goethals, Van Gestel, Suykens, Van Dooren, and De Moor (2003), Hoagg, Lacy, Erwin, and Bernstein (2004), Inoue (2019), Miller and De Callafon (2013) and Okada and Sugie (1996), e.g., including information on the location of the eigenvalues (Miller & De Callafon, 2013; Okada & Sugie, 1996), the moments of transfer function (Inoue, 2019), and the positive-realness (Goethals et al., 2003; Hoagg et al., 2004). The subspace identification method is employed to include information on the steady-state behavior (Alenany, Shang, Soliman, & Ziedan, 2011; Lacy & Bernstein, 2003; Yoshimura, Matsubayashi, & Inoue, 2019) such as the stability of the system (Lacy & Bernstein, 2003).

[☆] The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Gianluigi Pillonetto under the direction of Editor Alessandro Chiuso.

[*] Corresponding author.
E-mail addresses: mohammad.khosravi@tudelft.nl (M. Khosravi), rsmith@control.ee.ethz.ch (R.S. Smith).

https://doi.org/10.1016/j.automatica.2022.110813
0005-1098/

Starting from the seminal work Pillonetto and De Nicolao (Pillonetto & De Nicolao, 2010), a paradigm shift known as the *kernel-based* approach has emerged in system identification which allows integrating side-information (Chiuso & Pillonetto, 2019; Khosravi & Smith, 2021f, 2023; Ljung, Chen, & Mu, 2020; Pillonetto, Dinuzzo, Chen, De Nicolao, & Ljung, 2014). In this framework, the identification problem is formulated as a regularized regression in a reproducing kernel Hilbert space (RKHS) (Berlinet & Thomas-Agnan, 2011) where the regularization term, based on the norm of RKHS, penalizes solutions not compatible with the side-information. By a suitable choice of the kernel function, or by imposing appropriate constraints in the regression problem, one can incorporate a variety of side-information such as stability, resonant frequencies, smoothness of the impulse response, steady-state gain, and, internal or external positivity of the system (Chen, 2018; Khosravi & Smith, 2019, 2021d, 2021g; Marconato, Schoukens, & Schoukens, 2016; Zheng & Ohta, 2021). Moreover, employing a Tikhonov-like regularization in this framework leads to an improvement in the bias–variance trade-off (Pillonetto et al., 2014).

Input–output behavioral properties such as the $\mathcal{H}_\infty$-norm of the plant, or more generally frequency domain properties like the dissipativity of the system (Willems, 1972), can be considerably useful in feedback controller design (Brogliato, Lozano, Maschke, & Egeland, 2007; Zames, 1966). These features are a priori known for many systems due to their inherent physical nature, e.g., the electrical circuits where the energy is dissipated by the resistors (Willems, 1972). In addition, they can be verified using recently developed data-driven methods (Koch, Berberich, & Allgöwer, 2020; Koch, Montenbruck, & Allgöwer, 2019; Müller, Valenzuela, Proutiere, & Rojas, 2017; Romer, Berberich, Köhler, & Allgöwer, 2019; Romer, Montenbruck, & Allgöwer, 2017). Information about these attributes is potentially available for later use as side-information. However, existing research on LTI system identification with side-information (Abe et al., 2016) does not adequately address the inclusion of this frequency domain side-information. For example, in the standard implementation of subspace and kernel-based identification methods, the information on the bound of system's $\mathcal{H}_\infty$-norm is not encoded in the identified model (see the example in Section 3).

The main goals of this paper are to develop identification methods utilizing frequency domain side-information in a numerically tractable fashion and to provide suitable theoretical guarantees. First, we consider the case where the side-information is the $\mathcal{H}_\infty$-norm of the system being bounded by a given scalar. The framework considers various forms of frequency domain side-information, such as the dissipativity of the system. As kernel-based approaches provide a powerful framework, the identification problem is formulated as a constrained regularized regression problem in a stable RKHS (Chen & Pillonetto, 2018; Pillonetto et al., 2014) with suitable constraints to impose the frequency domain side-information on the estimation. The optimization problem is infinite-dimensional with an infinite number of constraints. We show that this problem is a convex program with a unique solution. Towards deriving a tractable solution scheme, we consider a suitable finite set of frequencies, and subsequently, an approximate estimation problem is formulated as an optimization problem with the same objective function but with constraints imposed according to these frequencies. The new problem attains a unique solution with a specific parametric form. Subsequently, an equivalent finite-dimensional convex optimization is derived as a convex quadratically constrained quadratic program (QCQP). By solving this optimization problem, the solution of the approximate problem is obtained. We derive proper bounds on the tightness of this approximation and provide theoretical guarantees on the convergence of the approximate solution to the solution of the original problem. Several numerical examples verify the efficiency of the proposed method.

2. Notation and preliminaries

The set of natural numbers, the set of non-negative integers, the set of real numbers, the set of non-negative real numbers, the set of complex numbers, the n-dimensional Euclidean space, the space of n by m real matrices, and the space of n by n real symmetric matrices are denoted by $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{C}$, $\mathbb{R}^n$, $\mathbb{R}^{n\times m}$, and $\mathbb{S}^n$, respectively. For any $z \in \mathbb{C}$, the real and imaginary part of z are denoted by $\mathrm{real}(z)$ and $\mathrm{imag}(z)$, respectively. The inner product and norm of Hilbert space $\mathscr{H}$ is denoted by $\langle\cdot,\cdot\rangle_{\mathscr{H}}$ and $\|\cdot\|_{\mathscr{H}}$, respectively, and when it is clear from the context, we drop the subscript. To handle discrete and continuous time in the same formulation, $\mathbb{T}$ denotes either $\mathbb{Z}_+$ or $\mathbb{R}_+$, and $\mathbb{T}_\pm$ is the set of scalars t where either $t \in \mathbb{T}$ or $-t \in \mathbb{T}$. Given measure space $\mathscr{X}$, the space of measurable functions $g : \mathscr{X} \to \mathbb{R}$ is denoted by $\mathbb{R}^{\mathscr{X}}$. The element $\mathrm{u} \in \mathbb{R}^{\mathscr{X}}$ is shown entry-wise as $\mathrm{u} = (u_x)_{x\in\mathscr{X}}$, or equivalently as $\mathrm{u} = \big(u(x)\big)_{x\in\mathscr{X}}$. Depending on the context of discussion, $\mathscr{L}^\infty$ refers either to $\ell^\infty(\mathbb{Z})$ or $L^\infty(\mathbb{R})$. Similarly, $\mathscr{L}^1$ is either $\ell^1(\mathbb{Z}_+)$ or $L^1(\mathbb{R}_+)$. For $p \in \{1,\infty\}$, the norm in $\mathscr{L}^p$ is denoted by $\|\cdot\|_p$. Let $(\mathbb{X}, \|\cdot\|_{\mathbb{X}})$ and $(\mathbb{Y}, \|\cdot\|_{\mathbb{Y}})$ be two normed vector spaces. The set of linear bounded (continuous) operators $\mathrm{A} : \mathbb{X} \to \mathbb{Y}$, denoted by $\mathcal{L}(\mathbb{X},\mathbb{Y})$, is a normed vector space with the norm defined as $\|\mathrm{A}\|_{\mathcal{L}(\mathbb{X},\mathbb{Y})} := \sup_{\mathrm{x}\in\mathbb{X},\|\mathrm{x}\|_{\mathbb{X}}\le 1} \|\mathrm{Ax}\|_{\mathbb{Y}}$. The identity matrix/operator and the zero vector are denoted by $\mathbb{I}$ and $\mathbf{0}$ respectively. Given $\mathscr{V} \subseteq \mathbb{X}$, the linear span of $\mathscr{V}$, denoted by $\mathrm{span}\mathscr{V}$, is a linear subspace of $\mathbb{X}$ containing linear combination of the elements of $\mathscr{V}$. Let $\mathcal{Y}$ be a set and $\mathcal{C} \subseteq \mathcal{Y}$. We define the function $\delta_{\mathcal{C}}$ as $\delta_{\mathcal{C}}(y) = 0$, if $y \in \mathcal{C}$ and $\delta_{\mathcal{C}}(y) = \infty$, otherwise. Similarly, function $\mathbf{1}_{\mathcal{C}}$ is defined as $\mathbf{1}_{\mathcal{C}}(y) = 1$, if $y \in \mathcal{C}$ and $\mathbf{1}_{\mathcal{C}}(y) = 0$, otherwise.

3. System identification with frequency domain side-information

Consider a stable single-input–single-output (SISO) LTI system $\mathcal{S}$ described with impulse response $\mathrm{g}^{(\mathcal{S})} := (g_t^{(\mathcal{S})})_{t\in\mathbb{T}} \in \mathscr{L}^1$ and transfer function $G^{(\mathcal{S})}$. For the case of discrete-time and the case of continuous-time, we have here respectively $\mathbb{T} := \mathbb{Z}_+$ and $\mathbb{T} := \mathbb{R}_+$. Let the system be actuated with a bounded input signal denoted by $\mathrm{u} = (u_t)_{t\in\mathbb{T}} \in \mathscr{L}^\infty$. Accordingly, for any $t \in \mathbb{T}_\pm$, one can define linear map $\mathrm{L}_t^{\mathrm{u}}$ over the space of stable causal impulse responses, for the discrete-time case, as

$$\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) := \sum_{s\in\mathbb{Z}_+} g_s u_{t-s}, \tag{1}$$

and similarly, for the case of continuous-time, as

$$\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) := \int_{\mathbb{R}_+} g_s u_{t-s}\,\mathrm{d}s. \tag{2}$$

Let the output of the system be measured at time instants $\mathscr{T} := \{t_i \mid i = 0,\ldots,n_{\mathscr{D}}-1\}$, for a given $n_{\mathscr{D}} \in \mathbb{N}$. More precisely, define y_t as

$$y_t := \mathrm{L}_t^{\mathrm{u}}(\mathrm{g}^{(\mathcal{S})}) + w_t, \qquad t \in \mathscr{T}, \tag{3}$$

where, for any $t \in \mathscr{T}$, w_t denotes the measurement uncertainty. Subsequently, let $\mathscr{D}$ be the set of input–output pairs, i.e., $\mathscr{D}$ is defined as $\mathscr{D} = \{(u_t, y_t) \mid t \in \mathscr{T}\}$.

Let assume we know that the $\mathcal{H}_\infty$-norm of system $\mathcal{S}$ is bounded by a given scalar $\rho \in \mathbb{R}_+$. The question is whether this side-information is naturally encoded in the identification problem. The following example elaborates this issue by demonstrating that the information on the bound of the system's $\mathcal{H}_\infty$-norm is not included in the models identified by the standard identification approaches such as the subspace and the kernel-based methods.

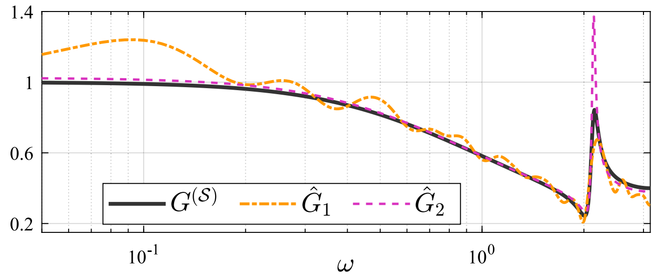

Fig. 1. The estimated models $\hat{G}_1$ and $\hat{G}_2$ for the system (4) are compared with the true transfer function.

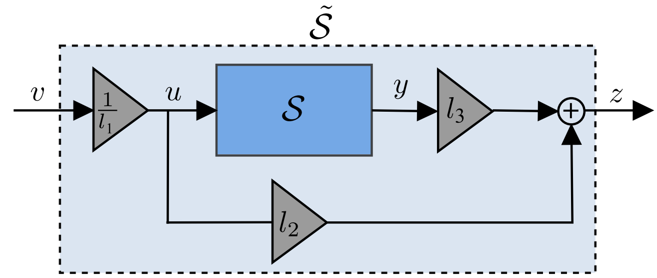

Fig. 2. The equivalent system $\tilde{\mathcal{S}}$ with $\mathcal{H}_\infty$-norm less than or equal to 1.

Example. Let $\mathcal{S}$ be a discrete-time system described by the following transfer function

$$G^{(\mathcal{S})}(z) = \frac{1}{2z-1} + \frac{3}{100}\,\frac{z-1}{z^2+z+0.9}. \tag{4}$$

To obtain set of data $\mathscr{D}$, we actuate the system with a random white Gaussian signal of length $n_{\mathscr{D}} = 150$, and then, the output of system is measured where SNR $= 14.5$ dB (see Fig. 1). Additionally, assume we are given the side-information $\|G^{(\mathcal{S})}\|_{\mathcal{H}_\infty} \leq 1$. We employ MATLAB's SYSTEM IDENTIFICATION TOOLBOX (Ljung & Singh, 2012) to estimate models $\hat{G}_1$ and $\hat{G}_2$ for the system using `impulseest` and `n4sid`, respectively. The results are shown in Fig. 1 where we have $\|\hat{G}_1\|_{\mathcal{H}_\infty} = 1.24$ and $\|\hat{G}_2\|_{\mathcal{H}_\infty} = 1.38$. One can see that the models estimated by the mentioned standard identification methods do not comply the side-information. △

Motivated by this example, we introduce the main problem discussed in this paper as system identification with side-information on the $\mathcal{H}_\infty$-norm of the system.

Problem 1. Given the set of input–output data $\mathscr{D}$, estimate the impulse response of system $\mathcal{S}$ satisfying the side-information $\|G^{(\mathcal{S})}\|_{\mathcal{H}_\infty} \leq \rho$, where ρ is a given non-negative real scalar.

This problem can be extended to the identification problem with more general characterization of dissipativity side-information. More precisely, let $\mathrm{Q} \in \mathbb{S}^2$ be an indefinite matrix and assume that in addition to the given set of data $\mathscr{D}$, we know that the system $\mathcal{S}$ is *dissipative* with respect to quadratic supply rate function $s_{\mathrm{Q}}(u, y)$ defined as

$$\begin{aligned} s_{\mathrm{Q}}(u,y) &:= \begin{bmatrix} u & y \end{bmatrix} \mathrm{Q} \begin{bmatrix} u \\ y \end{bmatrix} \\ &= \begin{bmatrix} u & y \end{bmatrix} \begin{bmatrix} q_u & q_{uy} \\ q_{uy} & q_y \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix} \\ &= q_u u^2 + 2q_{uy}uy + q_y y^2, \end{aligned} \tag{5}$$

where $q_y < 0$[1] (Antoulas, 2005; Haddad & Chellaboina, 2011). For the case of $s_{\mathrm{Q}}(u,y) = \rho^2u^2 - y^2$ where $\rho \in \mathbb{R}_+$, this dissipativity prior knowledge is equivalent to being $\mathscr{L}_2$-gain or $\mathcal{H}_\infty$-norm of the system not larger than ρ. Given the above side-information on the dissipativity of system $\mathcal{S}$, a desired identification procedure for estimating the impulse response of the system should suitably utilize this information and also guarantee that the identified model satisfies the given feature. The more general problem is the following.

Problem 2. Given the set of data $\mathscr{D}$ and considering the side-information on the dissipativity of system with respect to the supply rate $s_{\mathrm{Q}}(u, y)$, estimate the impulse response of system $\mathcal{S}$ satisfying the dissipativity side-information.

Note that we have

$$\mathrm{Q} = \begin{bmatrix} q_u & q_{uy} \\ q_{uy} & q_y \end{bmatrix} = \begin{bmatrix} l_1 & l_2 \\ 0 & l_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} l_1 & 0 \\ l_2 & l_3 \end{bmatrix} \tag{6}$$

where $l_1 = (\det \mathrm{Q}/q_y)^{\frac{1}{2}}$, $l_2 = -q_{uy}/(-q_y)^{\frac{1}{2}}$, and $l_3 = (-q_y)^{\frac{1}{2}}$, and subsequently, one can define the system $\tilde{\mathcal{S}}$ with input $v = l_1u$ and output $z = l_2u + l_3y$ as shown in Fig. 2. Then, $\tilde{\mathcal{S}}$ is a dissipative system with respect to supply rate function $s(v, z) = u^2 - y^2$, or equivalently, $\|G^{(\tilde{\mathcal{S}})}\|_{\mathcal{H}_\infty} \leq 1$ where $G^{(\tilde{\mathcal{S}})}$ is the transfer function of system $\tilde{\mathcal{S}}$. Meanwhile, using the introduced change of variables and based on $\mathscr{D}$, we can define set of data $\mathscr{D}_{\tilde{\mathcal{S}}} := \{(v_t, z_t) \mid t \in \mathscr{T}\}$, where $v_t := l_1u_t$ and $z_t := l_2u_t + l_3y_t$, for $t \in \mathscr{T}$. Accordingly, the problem is equivalent to identifying system $\tilde{\mathcal{S}}$ given the set of data $\mathscr{D}_{\tilde{\mathcal{S}}}$ and the side-information $\|G^{(\tilde{\mathcal{S}})}\|_{\mathcal{H}_\infty} \leq 1$. Therefore, in order to address Problem 2 it suffices to find a solution approach for Problem 1.

Problem 1 can be further extended to the case where system $\mathcal{S}$ is approximately known. More precisely, let $\bar{\mathcal{S}}$ be a known system with transfer function $G^{(\bar{\mathcal{S}})}$ and we are given that $\|G^{(\mathcal{S})} - G^{(\bar{\mathcal{S}})}\|_{\mathcal{H}_\infty} \leq \rho$, for a known $\rho \in \mathbb{R}_+$. This can be the case where the system might have been previously modeled or identified as $\bar{\mathcal{S}}$, and we would now like to identify an improved version of the model. The following identification problem addresses this case.

Problem 3. Given the set of input–output data $\mathscr{D}$, estimate the impulse response of system $\mathcal{S}$ satisfying the side-information $\|G^{(\mathcal{S})} - G^{(\bar{\mathcal{S}})}\|_{\mathcal{H}_\infty} \leq \rho$, where $\bar{\mathcal{S}}$ is a known system with transfer function $G^{(\bar{\mathcal{S}})}$ and $\rho \in \mathbb{R}_+$ is a given scalar.

Let $\Delta\mathcal{S}$ be the system with the transfer function $G^{(\Delta\mathcal{S})} := G^{(\mathcal{S})} - G^{(\bar{\mathcal{S}})}$ and the impulse response $\mathrm{g}^{(\Delta\mathcal{S})} := \mathrm{g}^{(\mathcal{S})} - \mathrm{g}^{(\bar{\mathcal{S}})}$, where $\mathrm{g}^{(\bar{\mathcal{S}})}$ is the impulse response of system $\bar{\mathcal{S}}$. Since $\mathrm{g}^{(\bar{\mathcal{S}})}$ is known, one can obtain $\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}^{(\bar{\mathcal{S}})})$, and subsequently, define d_t as $d_t := y_t - \mathrm{L}_t^{\mathrm{u}}(\mathrm{g}^{(\bar{\mathcal{S}})})$, for any $t \in \mathscr{T}$. Due to (3) and the linearity of $\mathrm{L}_t^{\mathrm{u}}$, we have

$$d_t = \mathrm{L}_t^{\mathrm{u}}(\mathrm{g}^{(\mathcal{S})}) + w_t - \mathrm{L}_t^{\mathrm{u}}(\mathrm{g}^{(\bar{\mathcal{S}})}) = \mathrm{L}_t^{\mathrm{u}}(\mathrm{g}^{(\Delta\mathcal{S})}) + w_t, \quad t \in \mathscr{T}.$$

Accordingly, we can define $\mathscr{D}_{\Delta\mathcal{S}} := \{(u_t, d_t) \mid t \in \mathscr{T}\}$ as the input–output pairs of data for the system $\Delta\mathcal{S}$. Therefore, Problem 3 is equivalent to identifying system $\Delta\mathcal{S}$ given the set of data $\mathscr{D}_{\Delta\mathcal{S}}$ and the side-information $\|G^{(\Delta\mathcal{S})}\|_{\mathcal{H}_\infty} \leq \rho$, which is in form of Problem 1.

Let $\mathbb{T} = \mathbb{Z}_+$, $\mathscr{T} := \{i = 0, \ldots, n_{\mathscr{D}} - 1\}$, and consider the following identification problem:

Problem 4. Given the set of input–output data $\mathscr{D}$, estimate the impulse response of system $\mathcal{S}$ satisfying the frequency domain side-information $\|WG^{(\mathcal{S})}\|_{\mathcal{H}_\infty} \leq \rho$, where $\rho \in \mathbb{R}_+$ is a given scalar, and weight W is a known stable transfer function with stable casual inverse.

[1] In other words, given that the system is initially at rest, for any input–output pairs $(u_t, y_y)_{t\in\mathbb{T}}$ and any $\tau \in \mathbb{T}$, we have $\int_0^\tau s_{\mathrm{Q}}(u_t, y_t)\mathrm{d}t \geq 0$, if $\mathbb{T} = \mathbb{R}_+$, and $\sum_{t=0}^{\tau} s_{\mathrm{Q}}(u_t, y_t) \geq 0$, if $\mathbb{T} = \mathbb{Z}_+$.

Let the output signal $\mathrm{y} = (y_t)_{t=0}^{n_{\mathscr{D}}-1}$ be filtered by system W and $\mathrm{p} := (p_t)_{t=0}^{n_{\mathscr{D}}-1}$ be the resulting filtered signal. One can see that $\mathscr{D}_H := \{(u_t, p_t) | t \in \mathscr{T}\}$ is a set of input–output data for the system with transfer function $H := WG^{(s)}$, where we know that $\|H\|_{\mathcal{H}_\infty} \le \rho$. Let $\hat{H}$ be the solution of Problem 1 for this setting. Then, $\hat{G} := W^{-1}\hat{H}$ is a solution to Problem 4.

Considering the above discussion, a solution approach for Problem 1 also addresses Problems 2–4. In the remainder of the paper, we discuss solving Problem 1.

4. The estimation problem: Existence and uniqueness of the solution

We investigate the existence and uniqueness properties for the solution of Problem 1.

Let $\mathscr{F} \subseteq \mathbb{R}^{\mathbb{T}}$ be a suitable functional space of stable impulse responses taken as the hypothesis set for the estimation Problem 1. Given bounded signal $\mathrm{u} \in \mathbb{R}^{\mathbb{T}}$, with respect to each $t \in \mathbb{T}$, we have the linear map $\mathrm{L}_t^{\mathrm{u}} : \mathscr{F} \to \mathbb{R}$ as defined in (1) and (2). Based on this definition and the set of data $\mathscr{D}$, one can define the *empirical loss* function or the fitting error function, $\mathcal{E}_{\mathscr{D}} : \mathscr{F} \to \mathbb{R}$, as the sum of squared error. In other words, for a given candidate impulse response $\mathrm{g} \in \mathscr{F}$, we have that

$$\mathcal{E}_{\mathscr{D}}(\mathrm{g}) := \sum_{t \in \mathscr{T}} (\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) - y_t)^2. \tag{7}$$

Define $\mathscr{G} \subseteq \mathscr{L}^1$ as the set of impulse responses corresponding to the systems with $\mathcal{H}_\infty$-norm less than or equal to ρ. More precisely, we have

$$\mathscr{G} := \left\{ \mathrm{g} = (\mathrm{g}_t)_{t \in \mathbb{Z}_+} \in \mathscr{L}^1 \,\middle|\, \sup_{\omega \in [0,\pi]} \left| \sum_{t \in \mathbb{Z}_+} \mathrm{g}_t \mathrm{e}^{-\mathrm{j}\omega t} \right| \le \rho \right\}, \tag{8}$$

and

$$\mathscr{G} := \left\{ \mathrm{g} = (\mathrm{g}_t)_{t \in \mathbb{R}_+} \in \mathscr{L}^1 \,\middle|\, \sup_{\omega \in \mathbb{R}_+} \left| \int_{\mathbb{R}_+} \mathrm{g}_t \mathrm{e}^{-\mathrm{j}\omega t} \mathrm{d}t \right| \le \rho \right\}, \tag{9}$$

respectively for $\mathbb{T} = \mathbb{Z}_+$ and $\mathbb{T} = \mathbb{R}_+$. Note that since each element of $\mathscr{G}$ belongs to $\mathscr{L}^1$ the summation in (8) and the integration in (9) are well-defined. Accordingly, in order to address Problem 1, we solve the following optimization problem

$$\begin{aligned} \min_{\mathrm{g} \in \mathscr{F}} \quad & \mathcal{E}_{\mathscr{D}}(\mathrm{g}) + \lambda \mathcal{R}(\mathrm{g}) \\ \text{s.t.} \quad & \mathrm{g} \in \mathscr{G}, \end{aligned} \tag{10}$$

where $\mathcal{R} : \mathscr{F} \to \mathbb{R}_+$ is a suitable regularization function and $\lambda > 0$ is the regularization weight. Due to the definition of $\mathscr{G}$, the optimization problem (10) is an infinite-dimensional program with an uncountably infinite number of inequality constraints, which is not tractable in the current form. Accordingly, we need to address the following questions:

(1) *What is a suitable candidate for hypothesis set $\mathscr{F}$?*
(2) *Does optimization problem (10) admit a solution? Is this solution unique?*
(3) *How can we obtain the solution of (10) or a tight approximation for it?*

The first two questions are addressed throughout the remainder of this section. The last question is postponed to Section 5.

4.1. Stable reproducing kernel Hilbert spaces

The hypothesis space taken for estimating the unknown impulse response is a type of Hilbert spaces known as *reproducing kernel Hilbert spaces* (RKHS) which are introduced briefly below (see Aronszajn, 1950; Berlinet & Thomas-Agnan, 2011, for more details). The structure of RKHS provides a suitable framework for investigating the problem and obtaining a tractable scheme for solving (10).

Definition 1 (Berlinet & Thomas-Agnan, 2011)**.** Let $\mathscr{H} \subseteq \mathbb{R}^{\mathbb{T}}$ be a Hilbert space endowed with inner product $\langle \cdot, \cdot \rangle_{\mathscr{H}}$ and induced norm $\| \cdot \|_{\mathscr{H}}$. Then, $\mathscr{H}$ is a *reproducing kernel Hilbert space* (RKHS) if for any $t \in \mathbb{T}$, we have $\sup\big\{ |\mathrm{g}_t| \,\big|\, \mathrm{g} := (\mathrm{g}_t)_{t \in \mathbb{T}} \in \mathscr{H}, \|\mathrm{g}\|_{\mathscr{H}} \le 1 \big\} < \infty$.

Along with the RKHS, the notion of kernel is introduced in the next definition.

Definition 2 (Berlinet & Thomas-Agnan, 2011)**.** The function $\Bbbk : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ is a *kernel*, or more precisely, a *positive definite kernel*, when for any $m \in \mathbb{N}$, $t, s, t_1, \ldots, t_m \in \mathbb{T}$ and $a_1, \ldots, a_m \in \mathbb{R}$, we have $\Bbbk(t, s) = \Bbbk(s, t)$ and $\sum_{1 \le i,j \le m} a_i a_j \Bbbk(t_i, t_j) \ge 0$. For each $t \in \mathbb{T}$, the *section* of kernel $\Bbbk$ at t is defined as the function $\Bbbk(t, \cdot) : \mathbb{T} \to \mathbb{R}$ and denoted by $\Bbbk_t$.

The next theorem shows the connection between RKHSs and the kernels.

Theorem 1 (Berlinet & Thomas-Agnan, 2011)**.** *Given a kernel $\Bbbk : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$, there exists a RKHS $\mathcal{H}_{\Bbbk} \subseteq \mathrm{g} \in \mathbb{R}^{\mathbb{T}}$, endowed with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_{\Bbbk}}$ and norm $\| \cdot \|_{\mathcal{H}_{\Bbbk}}$, such that for any $t \in \mathbb{T}$, we have $\Bbbk_t \in \mathcal{H}_{\Bbbk}$, and $\langle \mathrm{g}, \Bbbk_t \rangle_{\mathcal{H}_{\Bbbk}} = \mathrm{g}_t$, for all $\mathrm{g} = (\mathrm{g}_t)_{t \in \mathbb{T}} \in \mathcal{H}_{\Bbbk}$. The second feature is called the reproducing property.*

Based on Theorem 1, we know that a RKHS is uniquely characterized by a kernel. Therefore, in the current context, the kernel $\Bbbk$ is chosen suitably such that each impulse response $\mathrm{g} \in \mathcal{H}_{\Bbbk}$ represents a stable system in the bounded-input–bounded-output (BIBO) sense. More precisely, we should have $\mathcal{H}_{\Bbbk} \subseteq \mathscr{L}^1$. When kernel $\Bbbk$ satisfies this feature, it is called a *stable kernel* (Chen & Pillonetto, 2018). The following theorem provides a necessary and sufficient condition for the stability of a given kernel.

Theorem 2 (Carmeli, De Vito, & Toigo, 2006; Chen & Pillonetto, 2018)**.** *Let $\Bbbk : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ be a kernel. Then, $\Bbbk$ is stable if and only if, when $\mathbb{T} = \mathbb{Z}_+$, we have $\sum_{t \in \mathbb{Z}_+} \big| \sum_{s \in \mathbb{Z}_+} u_s \Bbbk(t, s) \big| < \infty$, and, when $\mathbb{T} = \mathbb{R}_+$, we have $\int_{\mathbb{R}_+} \big| \int_{\mathbb{R}_+} u_s \Bbbk(t, s) \mathrm{d}s \big| \mathrm{d}t < \infty$, for any $\mathrm{u} = (u_t)_{t \in \mathbb{T}} \in \mathscr{L}^\infty$.*

The most common stable kernel in the literature (Pillonetto et al., 2014) is the *tuned/correlated* (TC) kernel defined as follows

$$\Bbbk(s, t) = \begin{cases} \alpha^{\max(s,t)} & \text{if } \mathbb{T} = \mathbb{Z}_+, \\ \mathrm{e}^{-\beta \max(s,t)} & \text{if } \mathbb{T} = \mathbb{R}_+, \end{cases} \tag{11}$$

where $\alpha \in [0, 1)$ and $\beta > 0$. By setting α as $\mathrm{e}^{-\beta}$, one can obtain same definition for both cases in (11). This is the default kernel employed in `impulseest` function of MATLAB's SYSTEM IDENTIFICATION TOOLBOX. One can easily see that if kernel $\mathbb{h}$ is *dominated* by TC kernel $\Bbbk$, i.e., there exists $\gamma \in \mathbb{R}_+$ such that for any $s, t \in \mathbb{T}$ we have $|\mathbb{h}(s, t)| \le \gamma \Bbbk(s, t)$, then $\mathbb{h}$ is a stable kernel. This highlights the importance of TC kernels.

4.2. The estimation problem in stable reproducing kernel Hilbert spaces

Let $\Bbbk$ be a stable kernel and $\mathcal{H}_{\Bbbk}$ be the corresponding RKHS. Motivated by Ljung et al. (2020), Pillonetto and De Nicolao (2010), Pillonetto et al. (2014) and the above discussion, we set $\mathcal{H}_{\Bbbk}$ as the hypothesis space for the estimation problem, i.e., $\mathscr{F} = \mathcal{H}_{\Bbbk}$. Subsequently, we can introduce a suitable kernel-based regularization. More precisely, let the regularization function $\mathcal{R} : \mathcal{H}_{\Bbbk} \to \mathbb{R}_+$

be defined as $\mathcal{R}(\mathrm{g}) := \|\mathrm{g}\|^2_{\mathcal{H}_\Bbbk}$. Therefore, in correspondence with the estimation Problem 1, we have the following optimization problem:

$$\begin{aligned} \min_{\mathrm{g}\in\mathcal{H}_\Bbbk} \quad & \sum_{t\in\mathscr{T}} (\mathrm{L}^{\mathrm{u}}_t(\mathrm{g}) - y_t)^2 + \lambda \,\|\mathrm{g}\|^2_{\mathcal{H}_\Bbbk}, \\ \text{s.t.} \quad & |G(\mathrm{j}\omega)| \le \rho, \quad \forall \omega \in \Omega_{\mathbb{T}}, \end{aligned} \tag{12}$$

where G denotes the transfer function corresponding to the impulse response g, $\Omega_{\mathbb{T}} := [0,\pi]$ when $\mathbb{T} = \mathbb{Z}_+$, and $\Omega_{\mathbb{T}} := \mathbb{R}_+$ when $\mathbb{T} = \mathbb{R}_+$. Note that by abuse of notation, for both cases of discrete-time and continuous-time, we employ same expression $G(\mathrm{j}\omega)$ for the transfer function of the system.

The constraints introduced in (12) are the Fourier transform of the impulse response g at different frequencies. Since g belongs to RKHS $\mathcal{H}_\Bbbk$, one needs to study the frequency response and its properties in the domain of RKHS $\mathcal{H}_\Bbbk$.

Definition 3. With respect each $\omega \in \Omega_{\mathbb{T}}$, we define maps $\mathcal{F}^{(\mathrm{r})}_\omega : \mathcal{H}_\Bbbk \to \mathbb{R}$ and $\mathcal{F}^{(\mathrm{i})}_\omega : \mathcal{H}_\Bbbk \to \mathbb{R}$ such that for any $\mathrm{g} = (\mathrm{g}_t)_{t\in\mathbb{T}} \in \mathcal{H}_\Bbbk$, we have

$$\begin{aligned} \mathcal{F}^{(\mathrm{r})}_\omega(\mathrm{g}) &:= \sum_{t\in\mathbb{Z}_+} \mathrm{g}_t \cos(\omega t), \\ \mathcal{F}^{(\mathrm{i})}_\omega(\mathrm{g}) &:= -\sum_{t\in\mathbb{Z}_+} \mathrm{g}_t \sin(\omega t), \end{aligned} \tag{13}$$

when $\mathbb{T} = \mathbb{Z}_+$, and

$$\begin{aligned} \mathcal{F}^{(\mathrm{r})}_\omega(\mathrm{g}) &:= \int_{\mathbb{R}_+} \mathrm{g}_t \cos(\omega t)\mathrm{d}t, \\ \mathcal{F}^{(\mathrm{i})}_\omega(\mathrm{g}) &:= -\int_{\mathbb{R}_+} \mathrm{g}_t \sin(\omega t)\mathrm{d}t, \end{aligned} \tag{14}$$

when $\mathbb{T} = \mathbb{R}_+$. Also $\mathcal{F}_\omega : \mathcal{H}_\Bbbk \to \mathbb{C}$ is defined as $\mathcal{F}_\omega(\mathrm{g}) = \mathcal{F}^{(\mathrm{r})}_\omega(\mathrm{g}) + \mathrm{j}\mathcal{F}^{(\mathrm{i})}_\omega(\mathrm{g})$, for any $\mathrm{g} \in \mathcal{H}_\Bbbk$.

Based on the definition of $\mathcal{F}_\omega$, for any $\mathrm{g} \in \mathcal{H}_\Bbbk$ and $\omega \in \Omega_{\mathbb{T}}$, we have $G(\mathrm{j}\omega) = \mathcal{F}_\omega(\mathrm{g})$, where G is the transfer function corresponding to the impulse response g. Accordingly, the problem (12) can be re-written in the following form

$$\begin{aligned} \min_{\mathrm{g}\in\mathcal{H}_\Bbbk} \quad & \sum_{t\in\mathscr{T}} (\mathrm{L}^{\mathrm{u}}_t(\mathrm{g}) - y_t)^2 + \lambda \,\|\mathrm{g}\|^2_{\mathcal{H}_\Bbbk}, \\ \text{s.t.} \quad & |\mathcal{F}_\omega(\mathrm{g})| \le \rho, \quad \forall \omega \in \Omega_{\mathbb{T}}. \end{aligned} \tag{15}$$

Remark 1. One can see that $\mathcal{F}_\omega$ is the Fourier transform restricted to the RKHS $\mathcal{H}_\Bbbk$ evaluated for the frequency $\omega \in \Omega_{\mathbb{T}}$. However, we should note that it does not necessarily inherit the same properties of standard Fourier transform. More precisely, the structure of $\mathcal{H}_\Bbbk$ plays a key role.

Before proceeding further, we need to introduce a notion based on the kernel $\Bbbk$. For any $n \in \mathbb{Z}_+$, define $\mu_n \in [0,\infty]$ as

$$\mu_n := \begin{cases} \displaystyle\sum_{t\in\mathbb{Z}_+} t^n\, \Bbbk(t,t)^{\frac{1}{2}}, & \text{if } \mathbb{T} = \mathbb{Z}_+, \\ \displaystyle\int_{\mathbb{R}_+} t^n\, \Bbbk(t,t)^{\frac{1}{2}}\,\mathrm{d}t, & \text{if } \mathbb{T} = \mathbb{R}_+. \end{cases} \tag{16}$$

The next lemma introduces properties of $\mathcal{F}^{(\mathrm{r})}_\omega$, $\mathcal{F}^{(\mathrm{i})}_\omega$ and $\mathcal{F}_\omega$.

Lemma 3. *Assume that $\mu_0 < \infty$. Then, the following hold:*
(1) The maps $\mathcal{F}^{(r)}_\omega, \mathcal{F}^{(i)}_\omega : \mathcal{H}_\Bbbk \to \mathbb{R}$ and $\mathcal{F}_\omega : \mathcal{H}_\Bbbk \to \mathbb{C}$ are linear continuous with

$$\|\mathcal{F}^{(r)}_\omega\|_{\mathcal{L}(\mathcal{H}_\Bbbk,\mathbb{R})}, \|\mathcal{F}^{(i)}_\omega\|_{\mathcal{L}(\mathcal{H}_\Bbbk,\mathbb{R})} \le \mu_0. \tag{17}$$

(2) There exist unique elements $\varphi^{(r)}_\omega = (\varphi^{(r)}_{\omega,t})_{t\in\mathbb{T}}$ and $\varphi^{(i)}_\omega = (\varphi^{(i)}_{\omega,t})_{t\in\mathbb{T}}$ in $\mathcal{H}_\Bbbk$ such that, for any $\mathrm{g} \in \mathcal{H}_\Bbbk$, we have

$$\begin{aligned} \mathcal{F}^{(r)}_\omega(\mathrm{g}) &= \langle \varphi^{(r)}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk}, \\ \mathcal{F}^{(i)}_\omega(\mathrm{g}) &= \langle \varphi^{(i)}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk}. \end{aligned} \tag{18}$$

(3) For any $t \in \mathbb{T}$, when $\mathbb{T} = \mathbb{Z}_+$, we have

$$\begin{aligned} \varphi^{(r)}_{\omega,t} &= \sum_{s\in\mathbb{Z}_+} \Bbbk(t,s)\cos(\omega s), \\ \varphi^{(i)}_{\omega,t} &= -\sum_{s\in\mathbb{Z}_+} \Bbbk(t,s)\sin(\omega s), \end{aligned} \tag{19}$$

and, when $\mathbb{T} = \mathbb{R}_+$, we have

$$\begin{aligned} \varphi^{(r)}_{\omega,t} &= \int_{\mathbb{R}_+} \Bbbk(t,s)\cos(\omega s)\mathrm{d}s, \\ \varphi^{(i)}_{\omega,t} &= -\int_{\mathbb{R}_+} \Bbbk(t,s)\sin(\omega s)\mathrm{d}s. \end{aligned} \tag{20}$$

Proof. See Appendix A in Khosravi and Smith (2021c). □

Let $\Omega \subseteq \Omega_{\mathbb{T}}$ and $\eta \in \mathbb{R}_+$. Define set $\mathscr{G}_\Bbbk(\eta,\Omega)$ as

$$\mathscr{G}_\Bbbk(\eta,\Omega) := \left\{ \mathrm{g} \in \mathcal{H}_\Bbbk \;\middle|\; |\mathcal{F}_\omega(\mathrm{g})| \le \eta, \forall \omega \in \Omega \right\}. \tag{21}$$

One can see that the feasible set in (15) is a special case of this set. Based on Lemma 3, we study the main properties of $\mathscr{G}_\Bbbk(\eta,\Omega)$ in the next theorem.

Theorem 4. *Assume that $\mu_0 < \infty$. Then, $\mathscr{G}_\Bbbk(\eta,\Omega)$ is non-empty, closed and convex.*

Proof. From (21), one can see that

$$\begin{aligned} \mathscr{G}_\Bbbk(\eta,\Omega) &= \left\{ \mathrm{g} \in \mathcal{H}_\Bbbk \mid |\mathcal{F}_\omega(\mathrm{g})| \le \eta, \forall \omega \in \Omega \right\} \\ &= \bigcap_{\omega\in\Omega} \left\{ \mathrm{g} \in \mathcal{H}_\Bbbk \mid |\mathcal{F}_\omega(\mathrm{g})| \le \eta \right\} \\ &= \bigcap_{\omega\in\Omega} \mathscr{G}_\Bbbk\big(\eta, \{\omega\}\big). \end{aligned}$$

Therefore, it is enough to show that, for any $\omega \in \Omega_{\mathbb{T}}$, $\mathscr{G}_\Bbbk(\eta,\{\omega\})$ is a closed and convex subset of $\mathcal{H}_\Bbbk$. Since $|\mathcal{F}_\omega(\mathrm{g})|^2 = |\mathcal{F}^{(\mathrm{r})}_\omega(\mathrm{g})|^2 + |\mathcal{F}^{(\mathrm{i})}_\omega(\mathrm{g})|^2$, we have that

$$\mathscr{G}_\Bbbk(\eta,\{\omega\}) = \left\{ \mathrm{g} \in \mathcal{H}_\Bbbk \;\middle|\; |\mathcal{F}^{(\mathrm{r})}_\omega(\mathrm{g})|^2 + |\mathcal{F}^{(\mathrm{i})}_\omega(\mathrm{g})| \le \eta^2 \right\}.$$

Due to Lemma 3, we know that $\mathcal{F}^{(\mathrm{r})}_\omega(\mathrm{g}) = \langle \varphi^{(\mathrm{r})}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk}$ and $\mathcal{F}^{(\mathrm{i})}_\omega(\mathrm{g}) = \langle \varphi^{(\mathrm{i})}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk}$, for any $\mathrm{g} \in \mathcal{H}_\Bbbk$. With respect to each $\theta \in [0, \frac{\pi}{2}]$, define set $\mathscr{G}_\Bbbk(\eta,\omega,\theta)$ as

$$\begin{aligned} \mathscr{G}_\Bbbk(\eta,\omega,\theta) := \Big\{ \mathrm{g} \in \mathcal{H}_\Bbbk \Big| & |\langle \varphi^{(\mathrm{r})}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk}| \le \eta\cos\theta, \\ & |\langle \varphi^{(\mathrm{i})}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk}| \le \eta\sin\theta \Big\}. \end{aligned}$$

Note that $\mathscr{G}_\Bbbk(\eta,\omega,\theta)$ is the intersection of

$$\begin{aligned} &\left\{ \mathrm{g} \in \mathcal{H}_\Bbbk | \langle \varphi^{(\mathrm{r})}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk} \le \eta\cos\theta \right\}, \\ &\left\{ \mathrm{g} \in \mathcal{H}_\Bbbk | \langle \varphi^{(\mathrm{r})}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk} \ge -\eta\cos\theta \right\}, \\ &\left\{ \mathrm{g} \in \mathcal{H}_\Bbbk | \langle \varphi^{(\mathrm{i})}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk} \le \eta\sin\theta \right\}, \\ &\left\{ \mathrm{g} \in \mathcal{H}_\Bbbk | \langle \varphi^{(\mathrm{i})}_\omega, \mathrm{g}\rangle_{\mathcal{H}_\Bbbk} \ge -\eta\sin\theta \right\}, \end{aligned} \tag{22}$$

which are closed half-spaces in $\mathcal{H}_\Bbbk$. Therefore, $\mathscr{G}_\Bbbk(\eta,\omega,\theta)$ is a closed and convex set. Since we have $\mathscr{G}_\Bbbk(\eta,\{\omega\}) = \bigcap_{\theta\in[0,\frac{\pi}{2}]} \mathscr{G}_\Bbbk(\eta,\omega,\theta)$, the set $\mathscr{G}_\Bbbk(\eta,\{\omega\})$ is closed and convex as well. Note that for $\mathbf{0} \in \mathcal{H}_\Bbbk$ and any $\omega \in \Omega$, we have $\mathcal{F}_\omega(\mathbf{0}) = 0$. Therefore, $|\mathcal{F}_\omega(\mathbf{0})| = 0 \le \eta$ and subsequently, $\mathbf{0} \in \mathscr{G}(\eta,\Omega)$. This shows that $\mathscr{G}(\eta,\Omega)$ is a non-empty set and concludes the proof. □

Before proceeding to the main theorem of this section, we need to present an auxiliary lemma.

Lemma 5. *Assume that $\mu_0 < \infty$ and $\mathrm{u} \in \mathscr{L}^\infty$. Then, for any $t \in \mathbb{T}$, the map $\mathrm{L}_t^{\mathrm{u}} : \mathcal{H}_{\Bbbk} \to \mathbb{R}$, defined in (1) and (2), is linear and continuous with $\|\mathrm{L}_t^{\mathrm{u}}\|_{\mathcal{L}(\mathcal{H}_{\Bbbk},\mathbb{R})} \leq \mu_0 \|\mathrm{u}\|_\infty$. Also, there exists unique $\varphi_t^{(u)} := (\varphi_{t,s}^{(u)})_{s\in\mathbb{T}} \in \mathcal{H}_{\Bbbk}$ such that $\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) = \langle \varphi_t^{(u)}, \mathrm{g}\rangle_{\mathcal{H}_{\Bbbk}}$, for any $\mathrm{g} \in \mathcal{H}_{\Bbbk}$. Moreover, for any $s \in \mathbb{T}$, we have*

$$\varphi_{t,s}^{(u)} = \mathrm{L}_t^{\mathrm{u}}(\Bbbk_s) = \begin{cases} \displaystyle\sum_{\tau\in\mathbb{Z}_+} \Bbbk(s,\tau)u_{t-\tau}, & \text{if } \mathbb{T} = \mathbb{Z}_+, \\ \displaystyle\int_{\mathbb{R}_+} \Bbbk(s,\tau)u_{t-\tau}\mathrm{d}\tau, & \text{if } \mathbb{T} = \mathbb{R}_+. \end{cases} \tag{23}$$

Proof. See Appendix B in Khosravi and Smith (2021c). □

Theorem 6. *Assume that $\mu_0 > 0$ and consider the following program*

$$\min_{\mathrm{g}\in\mathscr{G}_{\Bbbk}(\eta,\Omega)} \sum_{t\in\mathscr{T}} (\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) - y_t)^2 + \lambda \|\mathrm{g}\|_{\mathcal{H}_{\Bbbk}}^2. \tag{24}$$

Then, (24) is a convex optimization problem with a unique *solution $\mathrm{g}_\Omega^\star$. Moreover, we have*

$$\|\mathrm{g}_\Omega^\star\|_{\mathcal{H}_{\Bbbk}} \leq \lambda^{-\frac{1}{2}} \|\mathrm{y}\|. \tag{25}$$

where vector y is defined as $\mathrm{y} = [y_t]_{t\in\mathscr{T}}$.

Proof. Define $\mathcal{J} : \mathcal{H}_{\Bbbk} \to \mathbb{R} \cup \{+\infty\}$ such that for any $\mathrm{g} \in \mathcal{H}_{\Bbbk}$ we have

$$\mathcal{J}(\mathrm{g}) = \sum_{t\in\mathscr{T}} (\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) - y_t)^2 + \lambda\|\mathrm{g}\|_{\mathcal{H}_{\Bbbk}}^2 + \delta_{\mathscr{G}_{\Bbbk}(\eta,\Omega)}(\mathrm{g}). \tag{26}$$

Since $\mathbf{0} \in \mathscr{G}_{\Bbbk}(\eta,\Omega)$, we have $\delta_{\mathscr{G}_{\Bbbk}(\eta,\Omega)}(\mathbf{0}) = 0$. From the definition of $\mathrm{L}_t^{\mathrm{u}}$ in (1) and (2), it is implied that $\mathrm{L}_t^{\mathrm{u}}(\mathbf{0}) = 0$, for each $t \in \mathscr{T}$, and subsequently, we have $\mathcal{J}(\mathbf{0}) = \sum_{t\in\mathscr{T}} y_t^2 = \|\mathrm{y}\|^2 < \infty$. From Theorem 4, we know that $\mathscr{G}_{\Bbbk}(\eta,\Omega)$ is a convex and closed set, and consequently, $\delta_{\mathscr{G}_{\Bbbk}(\eta,\Omega)}$ is a proper lower semi-continuous convex function (Peypouquet, 2015). Due to Lemma 5, we know that $\mathrm{L}_t^{\mathrm{u}} : \mathcal{H}_{\Bbbk} \to \mathbb{R}$ is a continuous linear map, for each $t \in \mathscr{T}$. Therefore, function $\mathcal{E}_{\mathscr{D}} : \mathcal{H}_{\Bbbk} \to \mathbb{R}$, defined in (7), is a convex and continuous function. Since $\lambda > 0$, we know that $\mathcal{J}$ is a proper and lower semi-continuous strongly convex function. Therefore, $\min_{\mathrm{g}\in\mathcal{H}_{\Bbbk}} \mathcal{J}(\mathrm{g})$ has a unique (finite) solution (Peypouquet, 2015), and subsequently, (24) is a convex program with a unique solution $\mathrm{g}_\Omega^\star$ with finite cost. Since $\mathbf{0} \in \mathscr{G}_{\Bbbk}(\eta,\Omega)$ and due to optimality of $\mathrm{g}_\Omega^\star$, we have $\mathcal{J}(\mathrm{g}_\Omega^\star) \leq \mathcal{J}(\mathbf{0})$. Subsequently, one can see

$$\lambda\|\mathrm{g}_\Omega^\star\|_{\mathcal{H}_{\Bbbk}}^2 \leq \sum_{t\in\mathscr{T}} (\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}_\Omega^\star) - y_t)^2 + \lambda\|\mathrm{g}_\Omega^\star\|_{\mathcal{H}_{\Bbbk}}^2 \leq \mathcal{J}(\mathbf{0}) = \sum_{t\in\mathscr{T}} y_t^2,$$

which implies (25). This concludes the proof. □

Corollary 7. *Assume that $\mu_0 < \infty$. Then, (15) is a convex optimization with a unique solution denoted by $\mathrm{g}^\star$. Moreover, $\mathrm{g}^\star$ satisfies inequality (25). A similar property holds for (12).*

Proof. In Theorem 6, set Ω and η respectively to $\Omega_{\mathbb{T}}$ and ρ. Then, the convexity of (15) as well as the existence and uniqueness of its solution immediately follows. Since (12) and (15) are equivalent, the same claim holds for (12). □

In the above discussion, we have assumed that $\mu_0 < \infty$. The next theorem shows that the boundedness of μ_0 is a valid assumption for most of the stable kernels introduced in the literature. In particular, it holds for TC kernel.

Theorem 8. *Consider kernel $\Bbbk$ and assume there exist $\beta, \gamma > 0$ such that $|\Bbbk(s,t)| \leq \gamma \mathrm{e}^{-\beta\max(s,t)}$, for any $s, t \in \mathbb{T}$. Then, for any $n \in \mathbb{Z}_+$, we have $\mu_n \leq \gamma^{\frac{1}{2}} (\frac{2}{\beta})^{n+1} n!$, where $n! := \prod_{k=1}^n k$, when $n \geq 1$, and, $0! := 1$.*

Proof. When $\mathbb{T} = \mathbb{Z}_+$, from the upper bound of kernel, we have

$$\mu_n \leq \sum_{t\in\mathbb{Z}_+} t^n\, \Bbbk(t,t)^{\frac{1}{2}} \leq \int_{\mathbb{R}_+} t^n (\gamma \mathrm{e}^{-\beta t})^{\frac{1}{2}} \mathrm{d}t,$$

where the second inequality is due to being $f(t) := \mathrm{e}^{-\frac{1}{2}\beta t}$ a non-increasing function. This inequality holds for the case of $\mathbb{T} = \mathbb{R}_+$ as well. Using change of variable $s = \frac{\beta}{2}t$, we have

$$\mu_n \leq \gamma^{\frac{1}{2}} \frac{2^{n+1}}{\beta^{n+1}} \int_{\mathbb{R}_+} s^n \mathrm{e}^{-s} \mathrm{d}s = \gamma^{\frac{1}{2}} \frac{2^{n+1}}{\beta^{n+1}} \Gamma(n+1),$$

where the equality is due to the definition of Gamma function. The claim follows from $\Gamma(k+1) = k!$, for $k \in \mathbb{Z}_+$. □

Based on Corollary 7, we know that the optimization problem (12) has a unique solution addressing Problem 1. Since this optimization problem is defined over an infinite-dimensional space with uncountably infinite number of constraints, obtaining this solution is not yet tractable in the current form. This is discussed in the next section.

5. Towards a tractable scheme

In this section, we present a tractable approach to derive the solution of the nonparametric estimation problem introduced in Section 4.

Consider optimization problem (15). For ease of discussion and without loss of generality, we assume $\rho = 1$. More precisely, one can use a change of variable and replace y_t with $\rho^{-1}y_t$ in $\mathscr{D}$ and then, set $\rho = 1$ in Problem 1 as well as in (15), or equivalently in (12). With respect to each $\omega \in \Omega_{\mathbb{T}}$, there exists a constraint in (15). Thus, we have uncountably infinite number of constraints which makes the problem intractable. To resolve this issue, we approximate the problem by considering a suitable finite subset of $\Omega_{\mathbb{T}}$. In order to investigate this possibility, we need the notion of *partition* introduced in the next definition.

Definition 4. We say $\mathscr{P}$ is a *partition* of interval $[a, b]$ if $\mathscr{P}$ is a finite subset of $[a, b]$ as $\mathscr{P} = \{\omega_i \mid i = 0, \ldots, n_{\mathscr{P}}\}$ where

$$a = \omega_0 < \omega_1 < \cdots < \omega_{n_{\mathscr{P}}} = b. \tag{27}$$

With respect to partition $\mathscr{P}$, the *mesh* of $\mathscr{P}$, denoted by $\mathrm{mesh}(\mathscr{P})$, is defined as

$$\mathrm{mesh}(\mathscr{P}) := \max\big\{|\omega_i - \omega_{i-1}| \mid i = 1, 2, \ldots, n_{\mathscr{P}}\big\}. \tag{28}$$

Now, let $\mathscr{P} = \{\omega_i \mid i = 0, \ldots, n_{\mathscr{P}}\}$ be a given partition. One can see that satisfying the constraints $|\mathcal{F}_\omega(\mathrm{g})| \leq 1$, for all $w \in \mathscr{P}$, does not necessarily imply that the desired feature $\sup_{\omega\in\Omega_{\mathbb{T}}} |\mathcal{F}_\omega(\mathrm{g})| \leq 1$. More precisely, these constraints do not guarantee that for all $\omega \in \Omega_{\mathbb{T}} \backslash \mathscr{P}$ we have $|\mathcal{F}_\omega(\mathrm{g})| \leq 1$ as well. Accordingly, in order to approximate problem (15), we take $\epsilon > 0$ and consider the following problem as our approximation

$$\begin{aligned} \min_{\mathrm{g}\in\mathcal{H}_{\Bbbk}} \quad & \sum_{t\in\mathscr{T}} (\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) - y_t)^2 + \lambda \|\mathrm{g}\|_{\mathcal{H}_{\Bbbk}}^2, \\ \text{s.t.} \quad & |\mathcal{F}_\omega(\mathrm{g})|^2 \leq 1 - \epsilon, \quad \forall \omega \in \mathscr{P}. \end{aligned} \tag{29}$$

In the following, we provide appropriate conditions on partition $\mathscr{P}$ and $\epsilon > 0$ for ensuring that the solution of (29) satisfies $\sup_{\omega\in\Omega_{\mathbb{T}}} |\mathcal{F}_\omega(\mathrm{g})| \leq 1$. These conditions depend on the bandwidth of system g and the rate of change of $|\mathcal{F}_\omega(\mathrm{g})|$ with respect to ω. Accordingly, we need to introduce necessary definitions and preliminaries.

Definition 5. Let $\mathrm{g} \in \mathcal{H}_{\Bbbk}$ be a given impulse response. With respect to g, we define function $m_{\mathrm{g}} : \Omega_{\mathbb{T}} \to \mathbb{R}$ as

$$m_{\mathrm{g}}(\omega) = |\mathcal{F}_{\omega}(\mathrm{g})|^2. \tag{30}$$

Given impulse response g, the function m_{g} shows the squared magnitude of transfer function corresponding to g at different frequencies. The following lemma introduces a bound for the rate of changes of m_{g} in terms of its Lipschitz constant, which in turn can be expressed in terms of the properties of the regularization kernel. This will be used later in the analysis of problem (29).

Lemma 9. *Assume that $\mu_0, \mu_1 < \infty$. Then, for any $\mathrm{g} \in \mathcal{H}_{\Bbbk}$ and $\omega_1, \omega_2 \in \Omega_{\mathbb{T}}$, we have*

$$|m_{\mathrm{g}}(\omega_2) - m_{\mathrm{g}}(\omega_1)| \leq L_{\mathrm{g}}|\omega_2 - \omega_1|, \tag{31}$$

where $L_{\mathrm{g}} := 4\mu_0\mu_1\|\mathrm{g}\|^2_{\mathcal{H}_{\Bbbk}}$.

Proof. See Appendix D in Khosravi and Smith (2021c). □

For the case of discrete-time systems, i.e., $\mathbb{T} = \mathbb{Z}_+$, the frequency range $\Omega_{\mathbb{T}}$ is the bounded interval $[0, \omega_{\max}]$, where $\omega_{\max} = \pi$. In order to introduce analogous of $\omega_{\max}$ for the continuous-time systems, we need the next assumption.

Assumption 1. For the case $\mathbb{T} = \mathbb{R}_+$, we have

$$\lim_{\omega\to\infty} \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Bbbk(s,t)\mathrm{e}^{-\mathrm{j}\omega(s-t)}\mathrm{d}s\mathrm{d}t = 0. \tag{32}$$

This assumption says that the bandwidth of kernel $\Bbbk$ is bounded. Moreover, from (32), we know that there exists $\omega_{\max} \in \mathbb{R}_+$ such that

$$\left|\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Bbbk(s,t)\mathrm{e}^{-\mathrm{j}\omega(s-t)}\mathrm{d}s\mathrm{d}t\right| \leq \frac{\lambda}{\sum_{t\in\mathscr{T}} y_t^2}, \tag{33}$$

for all $\omega > \omega_{\max}$. In the following, it is shown that the frequency value $\omega_{\max}$ plays the role of an upper bound for the bandwidth of the system of interest. The next theorem shows that for TC kernels Assumption 1 holds. One may show similar results for other stable kernels in the literature.

Theorem 10. *Let $\beta > 0$ and $\Bbbk : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ be the TC kernel $\Bbbk(s,t) = \mathrm{e}^{-\beta\max(s,t)}$. Then, we have*

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Bbbk(s,t)\mathrm{e}^{-\mathrm{j}\omega(s-t)}\mathrm{d}s\mathrm{d}t = \frac{2}{\omega^2+\beta^2}. \tag{34}$$

Proof. We know that $\max(s,t) = \frac{1}{2}\big(s+t+|s-t|\big)$. Accordingly, using change of variable $\tau = s-t$, it follows that

$$\begin{aligned}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} &\Bbbk(s,t)\mathrm{e}^{-\mathrm{j}\omega(s-t)}\mathrm{d}s\mathrm{d}t \\ &= \int_{\mathbb{R}_+} \mathrm{e}^{-\beta t}\int_{-t}^{\infty} \mathrm{e}^{-\beta\tau^+-\mathrm{j}\omega\tau}\mathrm{d}\tau\mathrm{d}t,\end{aligned} \tag{35}$$

where $\tau^+ := \frac{1}{2}\big(\tau+|\tau|\big)$. Also, we have

$$\begin{aligned}\int_{-t}^{\infty}\mathrm{e}^{-\beta\tau^+-\mathrm{j}\omega\tau}\mathrm{d}\tau &= \int_{-t}^{0}\mathrm{e}^{-\mathrm{j}\omega\tau}\mathrm{d}\tau + \int_{0}^{\infty}\mathrm{e}^{-\beta\tau-\mathrm{j}\omega\tau}\mathrm{d}\tau \\ &= \left[-\frac{1}{\mathrm{j}\omega}+\frac{\mathrm{e}^{\mathrm{j}\omega t}}{\mathrm{j}\omega}\right] + \frac{1}{\beta+\mathrm{j}\omega} \\ &= \frac{\mathrm{e}^{\mathrm{j}\omega t}}{\mathrm{j}\omega} - \frac{\beta}{\mathrm{j}\omega(\beta+\mathrm{j}\omega)}.\end{aligned}$$

Subsequently, by replacing the right-hand side of this equation in (35) and simplifying the integrals, we obtain

$$\begin{aligned}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} &\Bbbk(s,t)\mathrm{e}^{-\mathrm{j}\omega(s-t)}\mathrm{d}s\mathrm{d}t \\ &= -\frac{1}{\mathrm{j}\omega(-\beta+\mathrm{j}\omega)} - \frac{1}{\mathrm{j}\omega(\beta+\mathrm{j}\omega)} = \frac{2}{\beta^2+\omega^2}.\end{aligned}$$

This concludes the proof. □

Based on the discussion above, we can present the main theorem of this section, which concerns optimization problem (29). One should compare this theorem with the analogous argument in Corollary 7 about optimization problem (15).

Theorem 11. *Assume that μ_0, μ_1, defined in (16), are finite. Let $\mathscr{P}$ be a given partition of $[0, \omega_{\max}]$ and $\epsilon > 0$. Then, optimization problem (29) is a convex program with a unique solution $\mathrm{g}^{(\star,\epsilon)}$. For $\mathrm{g}^{(\star,\epsilon)}$, we have*

$$\|\mathrm{g}^{(\star,\epsilon)}\|_{\mathcal{H}_{\Bbbk}} \leq \lambda^{-\frac{1}{2}}\|\mathrm{y}\|. \tag{36}$$

where y is the vector defined as $\mathrm{y} = [y_t]_{t\in\mathscr{T}}$. Moreover, if $\mathrm{mesh}(\mathscr{P}) \leq \frac{2\epsilon}{L}$, where

$$L := \frac{1}{\lambda}4\mu_0\mu_1\sum_{t\in\mathscr{T}} y_i^2, \tag{37}$$

then we have

$$\|G^{(\star,\epsilon)}\|_{\mathcal{H}_\infty} = \sup_{\omega\in\Omega_{\mathbb{T}}} |\mathcal{F}_{\omega}(\mathrm{g}^{(\star,\epsilon)})| \leq 1. \tag{38}$$

where $G^{(\star,\epsilon)}$ is the transfer function corresponding to $\mathrm{g}^{(\star,\epsilon)}$.

Proof. Due to the definition of the set $\mathscr{G}_{\Bbbk}$ in (21), we know that the feasible set of optimization problem (29) is $\mathscr{G}_{\Bbbk}\big((1-\epsilon)^{\frac{1}{2}}, \mathscr{P}\big)$. Subsequently, problem (29) can be written as

$$\min_{\mathrm{g}\in\mathscr{G}_{\Bbbk}((1-\epsilon)^{\frac{1}{2}},\mathscr{P})} \sum_{t\in\mathscr{T}}(\mathrm{L}_i^{\mathrm{u}}(\mathrm{g})-y_i)^2 + \lambda\|\mathrm{g}\|^2_{\mathcal{H}_{\Bbbk}}. \tag{39}$$

Then, according to Theorem 6, the optimization problem (39) as well as (29), is a convex program with unique solution, denoted by $\mathrm{g}^{(\star,\epsilon)}$, which satisfies (36).

Let $\omega \in [0, \omega_{\max}]$. If $\omega \in \mathscr{P}$, then $|\mathcal{F}_{\omega}(\mathrm{g}^{(\star,\epsilon)})| \leq (1-\epsilon)^{\frac{1}{2}} \leq 1$. If $\omega \notin \mathscr{P}$, then there exists $i \in \{1,\ldots,n\}$, such that $\omega \in (\omega_{i-1}, \omega_i)$. If $|\mathcal{F}_{\omega}(\mathrm{g}^{(\star,\epsilon)})| > 1$, then, due to Lemma 9 and $|\mathcal{F}_{\omega_{i-1}}(\mathrm{g}^{(\star,\epsilon)})|^2, |\mathcal{F}_{\omega_i}(\mathrm{g}^{(\star,\epsilon)})|^2 \leq 1-\epsilon$, we know that

$$\begin{aligned}\epsilon &< \left| |\mathcal{F}_{\omega_i}(\mathrm{g}^{(\star,\epsilon)})|^2 - |\mathcal{F}_{\omega}(\mathrm{g}^{(\star,\epsilon)})|^2 \right| \\ &= |m_{\mathrm{g}^{(\star,\epsilon)}}(\omega_i) - m_{\mathrm{g}^{(\star,\epsilon)}}(\omega)| \leq L_{\mathrm{g}^{(\star,\epsilon)}}|\omega_i-\omega|, \\ \epsilon &< \left| |\mathcal{F}_{\omega}(\mathrm{g}^{(\star,\epsilon)})|^2 - |\mathcal{F}_{\omega_{i-1}}(\mathrm{g}^{(\star,\epsilon)})|^2 \right| \\ &= |m_{\mathrm{g}^{(\star,\epsilon)}}(\omega) - m_{\mathrm{g}^{(\star,\epsilon)}}(\omega_{i-1})| \leq L_{\mathrm{g}^{(\star,\epsilon)}}|\omega-\omega_{i-1}|,\end{aligned} \tag{40}$$

where $L_{\mathrm{g}^{(\star,\epsilon)}} = 4\mu_0\mu_1\|\mathrm{g}^{(\star,\epsilon)}\|^2_{\mathcal{H}_{\Bbbk}}$. Therefore, since $|\omega_i - \omega_{i-1}| \leq \mathrm{mesh}(\mathscr{P})$, we have that $2\epsilon < L_{\mathrm{g}^{(\star,\epsilon)}}\mathrm{mesh}(\mathscr{P})$. Subsequently, it follows that

$$\begin{aligned}\mathrm{mesh}(\mathscr{P}) > \frac{2\epsilon}{L_{\mathrm{g}^{(\epsilon)}}} &= \frac{2\epsilon}{4\mu_0\mu_1\|\mathrm{g}^{(\epsilon)}\|^2_{\mathcal{H}_{\Bbbk}}} \\ &\geq \frac{2\epsilon\lambda}{4\mu_0\mu_1\sum_{t\in\mathscr{T}} y_t^2} = \frac{2\epsilon}{L},\end{aligned} \tag{41}$$

which contradicts the specification that $\mathrm{mesh}(\mathscr{P}) \leq \frac{2\epsilon}{L}$. Therefore, we have $|\mathcal{F}_{\omega}(\mathrm{g}^{(\epsilon)})| \leq 1$. This shows that $\|G^{(\star,\epsilon)}\|_{\mathcal{H}_\infty} = \sup_{\omega\in[0,\omega_{\max}]}|\mathcal{F}_{\omega}(\mathrm{g}^{(\star,\epsilon)})| \leq 1$, which concludes the proof for the case $\mathbb{T} = \mathbb{Z}_+$ where we have $\omega_{\max} = \pi$. Now, we consider the case $\mathbb{T} = \mathbb{R}_+$ and let $\omega > \omega_{\max}$. Due to Lemma 3 and the Cauchy–Schwarz inequality, we know that

$$\begin{aligned}|\mathcal{F}_\omega(\mathrm{g}^{(\star,\epsilon)})|^2 &= |\langle \mathrm{g}^{(\star,\epsilon)}, \varphi_\omega^{(\mathrm{r})}\rangle_{\mathcal{H}_{\Bbbk}}|^2 + |\langle \mathrm{g}^{(\star,\epsilon)}, \varphi_\omega^{(\mathrm{i})}\rangle_{\mathcal{H}_{\Bbbk}}|^2 \\ &\le \|\mathrm{g}^{(\star,\epsilon)}\|_{\mathcal{H}_{\Bbbk}}^2 \left[\|\varphi_\omega^{(\mathrm{r})}\|_{\mathcal{H}_{\Bbbk}}^2 + \|\varphi_\omega^{(\mathrm{i})}\|_{\mathcal{H}_{\Bbbk}}^2\right].\end{aligned} \tag{42}$$

On the other hand, from (18), (14), and (33), we have

$$\begin{aligned}&\|\varphi_\omega^{(\mathrm{r})}\|_{\mathcal{H}_{\Bbbk}}^2 + \|\varphi_\omega^{(\mathrm{i})}\|_{\mathcal{H}_{\Bbbk}}^2 = \langle \varphi_\omega^{(\mathrm{r})}, \varphi_\omega^{(\mathrm{r})}\rangle_{\mathcal{H}_{\Bbbk}}^2 + \langle \varphi_\omega^{(\mathrm{i})}, \varphi_\omega^{(\mathrm{i})}\rangle_{\mathcal{H}_{\Bbbk}}^2 \\ &= \int_{\mathbb{R}_+} \varphi_{\omega,t}^{(\mathrm{r})} \cos(\omega t)\mathrm{d}t + \int_{\mathbb{R}_+} \varphi_{\omega,t}^{(\mathrm{i})} \sin(\omega t)\mathrm{d}t \\ &= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Bbbk(t,s)\cos(\omega s)\cos(\omega t)\mathrm{d}s\mathrm{d}t \\ &\quad + \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Bbbk(t,s)\sin(\omega s)\sin(\omega t)\mathrm{d}s\mathrm{d}t \\ &= \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Bbbk(t,s)\mathrm{e}^{-\mathrm{j}\omega(s-t)}\mathrm{d}s\mathrm{d}t \le \frac{\lambda}{\sum_{t\in\mathscr{T}} y_t^2},\end{aligned} \tag{43}$$

where the last equality is due to the fact that $\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Bbbk(t,s) \sin(\omega(s-t))\mathrm{d}s\mathrm{d}t = 0$ which is a result of $\Bbbk(t,s) = \Bbbk(s,t)$ and $\sin(\omega(s-t)) = -\sin(\omega(t-s))$. Accordingly, due to (36) and (42), we have $|\mathcal{F}_\omega(\mathrm{g}^{(\star,\epsilon)})|^2 \le 1$. This shows that $\|G^{(\star,\epsilon)}\|_{\mathcal{H}_\infty} = \sup_{\omega\in\Omega_{\mathbb{T}}} |\mathcal{F}_\omega(\mathrm{g}^{(\star,\epsilon)})| \le 1$ and concludes the proof. □

Due to Theorem 11, we know that (29) admits a unique solution. While this optimization problem is defined over the infinite-dimensional Hilbert space $\mathcal{H}_{\Bbbk}$, we can find its unique solution $\mathrm{g}^{(\star,\epsilon)}$ by solving an equivalent convex finite-dimensional program. This feature, which makes (29) a tractable problem, is due to the structure of the RKHS $\mathcal{H}_{\Bbbk}$ and the *Representer Theorem*, provided below.

Theorem 12 (*Representer Theorem, Dinuzzo & Schölkopf, 2012 and Schölkopf, Herbrich, & Smola, 2001*)**.** *Let $e : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ and $r : \mathbb{R}_+ \to \mathbb{R}$ be functions such that r is an increasing function. Also, let $\mathcal{H}$ be a Hilbert space with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$. Consider the optimization problem*

$$\min_{\mathrm{w}\in\mathcal{H}} \; e(\langle \mathrm{w}_1, \mathrm{w}\rangle_{\mathcal{H}}, \ldots, \langle \mathrm{w}_m, \mathrm{w}\rangle_{\mathcal{H}}) + r(\|\mathrm{w}\|_{\mathcal{H}}), \tag{44}$$

where $\mathrm{w}_1, \ldots, \mathrm{w}_m \in \mathcal{H}$ are given vectors. Then, if (44) admits a solution, it has also a solution in $\mathscr{W} := \mathrm{span}\{\mathrm{w}_i\}_{i=1}^m$.

In order to present the tractable finite-dimensional optimization problem equivalent to (29), additional definitions are required. Define $m := n_{\mathscr{U}} + 2n_{\mathscr{P}} + 2$ and the index sets $\mathcal{I}^{(\mathrm{u})} := \{0,1,\ldots,n_{\mathscr{U}}-1\}$, $\mathcal{I}^{(\mathscr{P})} := \{0,1,\ldots,n_{\mathscr{P}}\}$, and $\mathcal{I} := \{0,1,\ldots,m-1\}$. Let $\{\varphi_i\}_{i=0}^{m-1}$ be vectors defined as

$$\begin{cases} \varphi_i := \varphi_{t_i}^{(\mathrm{u})}, & \text{for } i \in \mathcal{I}^{(\mathrm{u})}, \\ \varphi_{n_{\mathscr{U}}+2j} := \varphi_{\omega_j}^{(\mathrm{r})}, & \text{for } j \in \mathcal{I}^{(\mathscr{P})}, \\ \varphi_{n_{\mathscr{U}}+2j+1} := \varphi_{\omega_j}^{(\mathrm{i})}, & \text{for } j \in \mathcal{I}^{(\mathscr{P})}. \end{cases} \tag{45}$$

Let $\Phi \in \mathbb{R}^{m\times m}$ be a symmetric matrix such that its entry at the ith row and the jth column is $\langle\varphi_{i-1},\varphi_{j-1}\rangle_{\mathcal{H}_{\Bbbk}}$, for $i,j = 1,\ldots,m$. One can see that Φ is the Gram matrix of vectors $\varphi_0,\ldots,\varphi_{m-1}$. Moreover, for $i \in \mathcal{I}^{(\mathrm{u})}$, we define vector $\mathrm{a}_i \in \mathbb{R}^m$ as the $(i+1)$th column of Φ. Similarly, vectors $\mathrm{b}_j \in \mathbb{R}^m$ and $\mathrm{c}_j \in \mathbb{R}^m$ are defined respectively as the $(n_{\mathscr{U}}+2j+1)$th and the $(n_{\mathscr{U}}+2j+2)$th column of Φ, for $j \in \mathcal{I}^{(\mathscr{P})}$. In the following, without loss of generality, we assume that $\varphi_0,\ldots,\varphi_{m-1}$ are linearly independent. Indeed, if for some $i \in \mathcal{I}$, the vector φ_i belongs to $\mathrm{span}\{\varphi_j | j \in \mathcal{I}\backslash\{i\}\}$, it does not provide any additional information and one can replace it with a linear combination of $\{\varphi_j | j \in \mathcal{I}\backslash\{i\}\}$. We can now present the theorem on the equivalent tractable finite-dimensional program.

Theorem 13. *The unique solution of optimization problem (29), $\mathrm{g}^{(\star,\epsilon)}$, is in the linear form*

$$\mathrm{g}^{(\star,\epsilon)} = x_0^\star\varphi_0 + \cdots + x_{m-1}^\star\varphi_{m-1}, \tag{46}$$

where $\mathrm{x}^\star := \left[x_0^\star, x_1^\star, \ldots, x_{m-1}^\star\right]^\mathsf{T} \in \mathbb{R}^m$ is the unique solution of the following convex program

$$\begin{aligned}\min_{\mathrm{x}\in\mathbb{R}^m} \quad & \sum_{i\in\mathcal{I}^{(\mathrm{u})}} (\mathrm{a}_i^\mathsf{T}\mathrm{x} - y_i)^2 + \lambda \mathrm{x}^\mathsf{T}\Phi\mathrm{x} \\ \text{s.t.} \quad & (\mathrm{b}_j^\mathsf{T}\mathrm{x})^2 + (\mathrm{c}_j^\mathsf{T}\mathrm{x})^2 \le 1-\epsilon, \quad \forall j \in \mathcal{I}^{(\mathscr{P})}.\end{aligned} \tag{47}$$

Proof. Define $\mathcal{J} : \mathcal{H}_{\Bbbk} \to \mathbb{R}\cup\{+\infty\}$ as

$$\mathcal{J}(\mathrm{g}) := \sum_{t\in\mathscr{T}} (\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) - y_t)^2 + \sum_{\omega\in\mathscr{P}} \delta_{\{|\mathcal{F}_\omega|^2\le 1-\epsilon\}}(\mathrm{g}). \tag{48}$$

Since $\mathscr{P} = \{\omega_0,\ldots,\omega_{n_{\mathscr{P}}}\}$ is a finite set, the summation in (48) is well-defined. We know that optimization problem (29) is equivalent to $\min_{\mathrm{g}\in\mathcal{H}_{\Bbbk}} \mathcal{J}(\mathrm{g}) + \lambda\mathcal{R}(\mathrm{g})$, where $\mathcal{R} : \mathcal{H}_{\Bbbk} \to \mathbb{R}$ is defined as $\mathcal{R}(\mathrm{g}) = \|\mathrm{g}\|_{\mathcal{H}_{\Bbbk}}^2$ (see Section 4.2). Let the function $e : \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ be defined such that for any $\mathrm{z} = (z_i)_{i=0}^{m-1} \in \mathbb{R}^m$ we have

$$e(\mathrm{z}) := \sum_{i\in\mathcal{I}^{(\mathrm{u})}} (z_i - y_{t_i})^2 + \sum_{j\in\mathcal{I}^{(\mathscr{P})}} \delta_{\mathscr{A}_j}(\mathrm{z}), \tag{49}$$

where, for each $j \in \mathcal{I}^{(\mathscr{P})}$, $\mathscr{A}_j \subseteq \mathbb{R}^m$ is the following set

$$\mathscr{A}_j := \left\{(z_i)_{i=0}^{m-1} \in \mathbb{R}^n \middle| z_{n_{\mathscr{U}}+2j}^2 + z_{n_{\mathscr{U}}+2j+1}^2 \le 1-\epsilon\right\}. \tag{50}$$

Also, let $r : \mathbb{R}_+ \to \mathbb{R}$ be the increasing function defined as $r(z) = \lambda z^2$. For $i \in \mathcal{I}^{(\mathrm{u})}$ and $j \in \mathcal{I}^{(\mathscr{P})}$, we know that $\mathrm{L}_{t_i}^{\mathrm{u}}(\mathrm{g}) = \langle\varphi_i,\mathrm{g}\rangle_{\mathcal{H}_{\Bbbk}}$, $\mathcal{F}_{\omega_j}^{(\mathrm{r})}(\mathrm{g}) = \langle\varphi_{n_{\mathscr{U}}+2j},\mathrm{g}\rangle_{\mathcal{H}_{\Bbbk}}$ and $\mathcal{F}_{\omega_j}^{(\mathrm{r})}(\mathrm{g}) = \langle\varphi_{n_{\mathscr{U}}+2j+1},\mathrm{g}\rangle_{\mathcal{H}_{\Bbbk}}$. Accordingly, we have

$$\mathcal{J}(\mathrm{g}) + \lambda\mathcal{R}(\mathrm{g}) = e(\langle\varphi_0,\mathrm{g}\rangle,\ldots,\langle\varphi_{m-1},\mathrm{g}\rangle) + r(\|\mathrm{g}\|).$$

From Theorem 11, it follows that $\min_{\mathrm{g}\in\mathcal{H}_{\Bbbk}} \mathcal{J}(\mathrm{g}) + \lambda\mathcal{R}(\mathrm{g})$ has a unique solution denoted by $\mathrm{g}^{(\star,\epsilon)}$. Therefore, due to Theorem 12, it has a solution which belongs to $\mathscr{W} = \mathrm{span}\{\varphi_i\}_{i=0}^{m-1}$. From the uniqueness of the solution of (29), it follows that $\mathrm{g}^{(\star,\epsilon)}$ belongs to $\mathscr{W}$. Therefore, in order to find $\mathrm{g}^{(\star,\epsilon)}$, we need to obtain the corresponding coefficients in the linear representation (46). Accordingly, we replace g in $\min_{\mathrm{g}\in\mathcal{H}_{\Bbbk}} \mathcal{J}(\mathrm{g}) + \lambda\mathcal{R}(\mathrm{g})$, or equivalently in (29), by $\mathrm{g} = \sum_{i=0}^{m-1} x_i\varphi_i$, and solve the problem for $\mathrm{x} := [x_i]_{i=0}^{m-1} \in \mathbb{R}^m$. Due to the linearity of inner product and the definition of matrix Φ, one has $\mathcal{R}(\mathrm{g}) = \mathrm{x}^\mathsf{T}\Phi\mathrm{x}$. Similarly, we have $\mathrm{L}_{t_i}^{\mathrm{u}}(\mathrm{g}) = \mathrm{a}_i^\mathsf{T}\mathrm{x}$, or each $i \in \mathcal{I}^{(\mathrm{u})}$, and, $\mathcal{F}_{\omega_j}^{(\mathrm{r})}(\mathrm{g}) = \mathrm{b}_j^\mathsf{T}\mathrm{x}$ and $\mathcal{F}_{\omega_j}^{(\mathrm{i})}(\mathrm{g}) = \mathrm{c}_j^\mathsf{T}\mathrm{x}$, for each $j \in \mathcal{I}^{(\mathscr{P})}$. Therefore, from (48), one has

$$\mathcal{J}(\mathrm{g}) = \mathcal{J}\left(\sum_{k=0}^{m-1} x_k\varphi_k\right) = \sum_{i=0}^{n_{\mathscr{U}}-1}(\mathrm{a}_i^\mathsf{T}\mathrm{x} - y_i)^2 + \sum_{j=0}^{n_{\mathscr{P}}} \delta_{\mathscr{B}_j}(\mathrm{x}),$$

where, for each $j \in \mathcal{I}^{(\mathscr{P})}$, $\mathscr{B}_j \subseteq \mathbb{R}^m$ is the set defined as

$$\mathscr{B}_j := \left\{\mathrm{x} = (x_i)_{i=0}^{m-1} \in \mathbb{R}^m \,\middle|\, (\mathrm{b}_j^\mathsf{T}\mathrm{x})^2 + (\mathrm{c}_j^\mathsf{T}\mathrm{x})^2 \le 1-\epsilon\right\}.$$

Accordingly, solving $\min_{\mathrm{g}\in\mathcal{H}_{\Bbbk}} \mathcal{J}(\mathrm{g}) + \lambda\mathcal{R}(\mathrm{g})$ reduces to

$$\min_{\mathrm{x}\in\mathbb{R}^m} \sum_{i=0}^{n_{\mathscr{U}}-1}(\mathrm{a}_i^\mathsf{T}\mathrm{x} - y_i)^2 + \sum_{j=0}^{n_{\mathscr{P}}} \delta_{\mathscr{B}_j}(\mathrm{x}) + \lambda\mathrm{x}^\mathsf{T}\Phi\mathrm{x}, \tag{51}$$

which is equivalent to (47). Since $\mathrm{x}^\mathsf{T}\Phi\mathrm{x} = \|\sum_{i=0}^{m-1} x_i\varphi_i\|^2$ and $\{\varphi_i\}_{i=0}^{m-1}$ are linearly independent, Φ is a positive definite matrix. Therefore, the cost function in (47) is strongly convex. Also, for each $j = 0,\ldots,n_{\mathscr{P}}$, we have

$$(\mathrm{b}_j^\mathsf{T}\mathrm{x})^2 + (\mathrm{c}_j^\mathsf{T}\mathrm{x})^2 = \mathrm{x}^\mathsf{T}(\mathrm{b}_j\mathrm{b}_j^\mathsf{T} + \mathrm{c}_j\mathrm{c}_j^\mathsf{T})\mathrm{x}, \tag{52}$$

and $\mathrm{b}_j\mathrm{b}_j^\top + \mathrm{c}_j\mathrm{c}_j^\top$ is a positive semi-definite matrix. Consequently, the feasible set in (47) is a convex and closed set. Moreover, since $\epsilon < 1$, we know that $\mathrm{g} = \mathbf{0}$ is a feasible point of (47). Therefore, the optimization problem (47) is a convex program with a unique solution. This concludes the proof. □

Before proceeding further, we present a corollary which is useful in the implementation of the proposed approach. Let $\overline{\mathscr{P}} := \{\overline{\omega}_i\}_{i=0}^{n_{\overline{\mathscr{P}}}} \subseteq \mathscr{P}$, $\mathcal{I}^{(\overline{\mathscr{P}})} := \{0, 1, \ldots, n_{\overline{\mathscr{P}}}\}$ and $\overline{m} := n_{\mathscr{Y}} + 2n_{\overline{\mathscr{P}}} + 2$. Analogously to Φ and $\{\varphi_i\}_{i=0}^{m-1}$, we define $\overline{\Phi}$ and $\{\overline{\varphi}_i\}_{i=0}^{\overline{m}-1}$. Also, for $i \in \mathcal{I}^{(\mathrm{u})}$ and $j \in \mathcal{I}^{(\overline{\mathscr{P}})}$, let $\bar{\mathrm{a}}_i$, $\bar{\mathrm{b}}_j$ and $\bar{\mathrm{c}}_j$ be defined in the same manner as a_i, b_j and c_j, respectively.

Corollary 14. *The optimization problem*

$$\begin{aligned} \min_{\mathrm{g}\in\mathcal{H}_{\Bbbk}} \quad & \sum_{t\in\mathscr{T}} (\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) - y_t)^2 + \lambda\,\|\mathrm{g}\|_{\mathcal{H}_{\Bbbk}}^2, \\ \text{s.t.} \quad & |\mathcal{F}_\omega(\mathrm{g})|^2 \le 1-\epsilon, \quad \forall \omega \in \overline{\mathscr{P}}. \end{aligned} \tag{53}$$

has a unique solution denoted by $\overline{\mathrm{g}}^{(\star,\epsilon)}$. This solution satisfies (36), and admits a parametric form as $\overline{\mathrm{g}}^{(\star,\epsilon)} = \sum_{i=0}^{\overline{m}-1} x_i\overline{\varphi}_i$ where $\mathrm{x} = [x_i]_{i=0}^{\overline{m}-1}$ is the solution of the following convex program

$$\begin{aligned} \min_{\mathrm{x}\in\mathbb{R}^{\overline{m}}} \quad & \sum_{i\in\mathcal{I}^{(\mathrm{u})}} (\bar{\mathrm{a}}_i^\top\mathrm{x} - y_i)^2 + \lambda\mathrm{x}^\top\overline{\Phi}\mathrm{x} \\ \text{s.t.} \quad & (\bar{\mathrm{b}}_j^\top\mathrm{x})^2 + (\bar{\mathrm{c}}_j^\top\mathrm{x})^2 \le 1-\epsilon, \quad \forall j \in \mathcal{I}^{(\overline{\mathscr{P}})}. \end{aligned} \tag{54}$$

Moreover, if $|\mathcal{F}_\omega(\overline{\mathrm{g}}^{(\star,\epsilon)})|^2 \le 1-\epsilon$, for all $\omega \in \mathscr{P}$, then $\overline{\mathrm{g}}^{(\star,\epsilon)}$ coincides with $\mathrm{g}^{(\star,\epsilon)}$.

Proof. Using similar lines of argument to the proof of Theorem 11, one can show the existence and uniqueness of the solution of (53). The parametric form of $\overline{\mathrm{g}}^{(\star,\epsilon)}$ and the fact that x is the solution of (54) can be concluded based on a proof similar to the proof of Theorem 13.

Since, for each $\omega \in \mathscr{P}$, one has $|\mathcal{F}_\omega(\overline{\mathrm{g}}^{(\star,\epsilon)})|^2 \le 1-\epsilon$, we know that $\overline{\mathrm{g}}^{(\star,\epsilon)}$ is feasible for optimization (29). Therefore, as $\mathrm{g}^{(\star,\epsilon)}$ is the optimal solution of (29), it follows that $\mathcal{E}_{\mathscr{D}}(\mathrm{g}^{(\star,\epsilon)}) + \lambda\mathcal{R}(\mathrm{g}^{(\star,\epsilon)}) \le \mathcal{E}_{\mathscr{D}}(\overline{\mathrm{g}}^{(\star,\epsilon)}) + \lambda\mathcal{R}(\overline{\mathrm{g}}^{(\star,\epsilon)})$, where $\mathcal{E}_{\mathscr{D}}$ is defined in (7). Also, we know that the feasible set in (29) is a subset of feasible set of (53). Accordingly, since $\overline{\mathrm{g}}^{(\star,\epsilon)}$ is the optimal solution for (53), we have $\mathcal{E}_{\mathscr{D}}(\overline{\mathrm{g}}^{(\star,\epsilon)}) + \lambda\mathcal{R}(\overline{\mathrm{g}}^{(\star,\epsilon)}) \le \mathcal{E}_{\mathscr{D}}(\mathrm{g}^{(\star,\epsilon)}) + \lambda\mathcal{R}(\mathrm{g}^{(\star,\epsilon)})$. Therefore, $\overline{\mathrm{g}}^{(\star,\epsilon)}$ is an optimizer of (29). Consequently, from the uniqueness of the solution of (29), we have $\overline{\mathrm{g}}^{(\star,\epsilon)} = \mathrm{g}^{(\star,\epsilon)}$, concluding the proof. □

Theorem 13 introduces the finite-dimensional convex program (47) as a tractable problem equivalent to optimization (29). The solution of (47) is discussed in the next section. However, we should first verify that solving optimization problem (29), for small enough $\epsilon > 0$, provides a close approximation to the solution of the main optimization problem (15). This is addressed by the next theorem.

Theorem 15. *Let $\mathrm{L}^{\mathrm{u}} : \mathcal{H}_{\Bbbk} \to \mathbb{R}^{n_{\mathscr{Y}}}$ be a linear operator defined as $\mathrm{L}^{\mathrm{u}}(\mathrm{g}) := [\mathrm{L}_t^{\mathrm{u}}(\mathrm{g})]_{t\in\mathscr{T}}$, for any $\mathrm{g} \in \mathcal{H}_{\Bbbk}$. Then, we have*

$$\|\mathrm{g}^{(\star,\epsilon)} - \mathrm{g}^\star\|_{\mathcal{H}_{\Bbbk}}^2 \le 4\epsilon(\|\mathrm{L}^{\mathrm{u}}\|_{\mathcal{L}(\mathcal{H}_{\Bbbk},\mathbb{R}^{n_{\mathscr{Y}}})}^2 + 1)^{\frac{1}{2}}\|\mathrm{y}\|^2. \tag{55}$$

where y is the vector defined as $\mathrm{y} = [y_t]_{t\in\mathscr{T}}$. Moreover, we have $\lim_{\epsilon\to\mathbf{0}} \mathrm{g}^{(\star,\epsilon)} = \mathrm{g}^\star$ in $\mathcal{H}_{\Bbbk}$. Furthermore, if $\sup_{t\in\mathbb{T}} \Bbbk(t,t)^{\frac{1}{2}}$ is finite, one has $\mathrm{g}_t^{(\star,\epsilon)} \xrightarrow{\epsilon\to 0} \mathrm{g}_t^\star$, uniformly in $\mathbb{T}$.

Proof. Let $\mathscr{V}_{\Bbbk} := \mathbb{R}^{n_{\mathscr{Y}}} \times \mathcal{H}_{\Bbbk}$ be the Hilbert space endowed with the inner product defined as

$$\langle(\mathrm{x}_1,\mathrm{g}_1),(\mathrm{x}_2,\mathrm{g}_2)\rangle_{\mathscr{V}_{\Bbbk}} := \mathrm{x}_1^\top\mathrm{x}_2 + \lambda\langle\mathrm{g}_1,\mathrm{g}_2\rangle_{\mathcal{H}_{\Bbbk}}, \tag{56}$$

for all $(\mathrm{x}_1,\mathrm{g}_1)$ and $(\mathrm{x}_2,\mathrm{g}_2)$ in $\mathscr{V}_{\Bbbk}$. Given $\Omega \subseteq \Omega_{\mathbb{T}}$ and $\rho \in \mathbb{R}_+$, define set $\mathscr{U}_{\Bbbk}(\rho,\Omega) \subseteq \mathscr{V}_{\Bbbk}$ as

$$\mathscr{U}_{\Bbbk}(\rho,\Omega) := \big\{(\mathrm{x},\mathrm{g})\big|\mathrm{L}^{\mathrm{u}}(\mathrm{g}) - \mathrm{x} = \mathrm{y},\, |\mathcal{F}_\omega(\mathrm{g})| \le \rho,\, \forall\omega\in\Omega\big\},$$

where $\mathrm{y} := [y_t]_{t\in\mathscr{T}} \in \mathbb{R}^{n_{\mathscr{Y}}}$. Let $\mathscr{U}_\epsilon := \mathscr{U}_{\Bbbk}(1-\epsilon,[0,\pi])$ and $\mathscr{U}_{\mathscr{P}} := \mathscr{U}_{\Bbbk}(1-\epsilon,\mathscr{P})$. Accordingly, replacing η with ρ, (24) is equivalent to the following optimization problem

$$\min_{(\mathrm{x},\mathrm{g})\in\mathscr{U}_{\Bbbk}(\rho,\Omega)} \|(\mathrm{x},\mathrm{g})\|_{\mathscr{V}_{\Bbbk}}^2, \tag{57}$$

and therefore, it is a convex optimization with the same unique solution. Let $\mathrm{v}^{(\epsilon)} := (\mathrm{x}^{(\epsilon)},\mathrm{g}^{(\epsilon)})$ be the solution for (57) for $\mathscr{U}_\epsilon$ for $\epsilon \in [0,1)$. When $\epsilon = 0$, we simply write $\mathrm{v}^\star := (\mathrm{x}^\star,\mathrm{g}^\star)$. Also, let $\mathrm{v}^{(\star,\epsilon)} := (\mathrm{x}^{(\star,\epsilon)},\mathrm{g}^{(\star,\epsilon)})$ is the solution of (57) for $\mathscr{U}_{\mathscr{P}}$. These notations is consistent with our previous ones due to the equivalency of (24) and (57) and the uniqueness of the solution. One can easily see that $\mathscr{U}_\epsilon \subseteq \mathscr{U}_{\mathscr{P}}$. Also, based on an argument similar to those provided in the proof of Theorem 11, we know that $\mathscr{U}_{\mathscr{P}} \subseteq \mathscr{U}_0$. Accordingly, we have

$$\sup_{\mathrm{w}\in\mathscr{U}_{\mathscr{P}}} \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}} \le \sup_{\mathrm{w}\in\mathscr{U}_0} \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}},$$

$$\sup_{\mathrm{w}\in\mathscr{U}_0} \min_{\mathrm{v}\in\mathscr{U}_{\mathscr{P}}} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}} \le \sup_{\mathrm{w}\in\mathscr{U}_0} \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}}.$$

Let $\mathrm{v}_1 \in \mathscr{V}_{\Bbbk}$ be defined as

$$\mathrm{v}_1 := \mathrm{proj}_{\mathscr{U}_\epsilon}(\mathrm{v}^\star) = \mathrm{argmin}_{\mathrm{v}\in\mathscr{U}_\epsilon}\|\mathrm{v}-\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2. \tag{58}$$

Since $\mathrm{v}_1 \in \mathscr{U}_\epsilon$ and $\mathscr{U}_\epsilon \subseteq \mathscr{U}_0$, due to the definition of $\mathrm{v}^{(\epsilon)}$, we know that $\|\mathrm{v}^{(\epsilon)}\|_{\mathscr{V}_{\Bbbk}} \le \|\mathrm{v}_1\|_{\mathscr{V}_{\Bbbk}}$. Therefore, from triangle inequality and (58), we have

$$\begin{aligned} \|\mathrm{v}^{(\epsilon)}\|_{\mathscr{V}_{\Bbbk}} \le \|\mathrm{v}_1\|_{\mathscr{V}_{\Bbbk}} &\le \|\mathrm{v}_1 - \mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}} + \|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}} \\ &= \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}} + \|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}} \\ &\le \sup_{\mathrm{w}\in\mathscr{U}_0} \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}} + \|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}. \end{aligned}$$

Consequently, it follows that

$$0 \le \|\mathrm{v}^{(\epsilon)}\|_{\mathscr{V}_{\Bbbk}} - \|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}} \le \sup_{\mathrm{w}\in\mathscr{U}_0} \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}}.$$

Due to the convexity of $\mathscr{U}_0$ and $\mathscr{U}_\epsilon \subseteq \mathscr{U}_0$, we know that $\frac{1}{2}(\mathrm{v}^{(\epsilon)} + \mathrm{v}^\star) \in \mathscr{U}_0$, and therefore, we have

$$\|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}} \le \|\frac{1}{2}(\mathrm{v}^{(\epsilon)} + \mathrm{v}^\star)\|_{\mathscr{V}_{\Bbbk}} = \frac{1}{2}\|\mathrm{v}^{(\epsilon)} + \mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}. \tag{59}$$

Subsequently, one can see that

$$\begin{aligned} \|\mathrm{v}^{(\epsilon)} - \mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2 &= 2\|\mathrm{v}^{(\epsilon)}\|_{\mathscr{V}_{\Bbbk}}^2 + 2\|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2 - \|\mathrm{v}^{(\epsilon)} + \mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2 \\ &\le 2\|\mathrm{v}^{(\epsilon)}\|_{\mathscr{V}_{\Bbbk}}^2 + 2\|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2 - 4\|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2 \\ &= 2\|\mathrm{v}^{(\epsilon)}\|_{\mathscr{V}_{\Bbbk}}^2 - 2\|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2 \\ &= 2(\|\mathrm{v}^{(\star,\epsilon)}\|_{\mathscr{V}_{\Bbbk}} - \|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}})(\|\mathrm{v}^{(\star,\epsilon)}\|_{\mathscr{V}_{\Bbbk}} + \|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}) \\ &\le 2(\sup_{\mathrm{w}\in\mathscr{U}_0} \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}})(\|\mathrm{v}^{(\star,\epsilon)}\|_{\mathscr{V}_{\Bbbk}} + \|\mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}). \end{aligned}$$

Since $\mathbf{0}$ belongs to $\mathscr{U}_\epsilon$, $\mathscr{U}_{\mathscr{P}}$ and $\mathscr{U}_0$, one has $\|\mathrm{y}\| \le \|\mathrm{g}^{(\epsilon)}\|_{\mathcal{H}_{\Bbbk}}$, $\|\mathrm{y}\| \le \|\mathrm{g}^{(\star,\epsilon)}\|_{\mathcal{H}_{\Bbbk}}$ and $\|\mathrm{y}\| \le \|\mathrm{g}^\star\|_{\mathcal{H}_{\Bbbk}}$. Therefore, we have

$$\|\mathrm{v}^{(\epsilon)} - \mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2 \le 4(\sup_{\mathrm{w}\in\mathscr{U}_0} \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}})\|\mathrm{y}\|.$$

Similarly, one can show the following inequalities

$$\|\mathrm{v}^{(\epsilon)} - \mathrm{v}^{(\star,\epsilon)}\|_{\mathscr{V}_{\Bbbk}}^2 \le 4(\sup_{\mathrm{w}\in\mathscr{U}_{\mathscr{P}}} \min_{\mathrm{v}\in\mathscr{U}_\epsilon} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}})\|\mathrm{y}\|, \tag{60}$$

$$\|\mathrm{v}^{(\star,\epsilon)} - \mathrm{v}^\star\|_{\mathscr{V}_{\Bbbk}}^2 \le 4(\sup_{\mathrm{w}\in\mathscr{U}_0} \min_{\mathrm{v}\in\mathscr{U}_{\mathscr{P}}} \|\mathrm{v}-\mathrm{w}\|_{\mathscr{V}_{\Bbbk}})\|\mathrm{y}\|. \tag{61}$$

Now, let $\mathrm{v} = (\mathrm{x},\mathrm{g})$ be an arbitrary element of $\mathscr{U}_0$. Then, we know that $(1-\epsilon)\mathrm{g} \in \mathscr{G}(1-\epsilon,\Omega_{\mathbb{T}})$. Also, we have

$$\mathrm{L}^{\mathrm{u}}((1-\epsilon)\mathrm{g}) - (\mathrm{x} - \epsilon\mathrm{L}^{\mathrm{u}}(\mathrm{g})) = \mathrm{L}^{\mathrm{u}}(\mathrm{g}) - \mathrm{x} = \mathrm{y}.$$

Therefore $\mathrm{w} := (\mathrm{x}-\epsilon\mathrm{L}^{\mathrm{u}}(\mathrm{g}), (1-\epsilon)\mathrm{g})$ is an element of $\mathscr{U}_0$. Moreover, we have

$$\begin{aligned}\|\mathrm{v}-\mathrm{w}\|^2_{\mathcal{V}_{\Bbbk}} &= \epsilon(\|\mathrm{L}^{\mathrm{u}}(\mathrm{g})\|^2 + \|\mathrm{g}\|^2_{\mathcal{H}_{\Bbbk}})^{\frac{1}{2}} \\ &\le \epsilon(\|\mathrm{L}^{\mathrm{u}}\|^2_{\mathcal{L}(\mathcal{H}_{\Bbbk},\mathbb{R}^{n_{\mathscr{D}}})} + 1)^{\frac{1}{2}}\|\mathrm{g}\|^2_{\mathcal{H}_{\Bbbk}}.\end{aligned}$$

Accordingly, it follows that

$$\begin{aligned}\sup_{\mathrm{w}\in\mathscr{U}_0}\min_{\mathrm{v}\in\mathscr{U}_\epsilon}\|\mathrm{v}-\mathrm{w}\|_{\mathcal{V}_{\Bbbk}} &\le \sup_{(\mathrm{x},\mathrm{g})\in\mathscr{U}_0} \epsilon(\|\mathrm{L}^{\mathrm{u}}\|^2_{\mathcal{L}(\mathcal{H}_{\Bbbk},\mathbb{R}^{n_{\mathscr{D}}})} + 1)^{\frac{1}{2}}\|\mathrm{g}\|^2_{\mathcal{H}_{\Bbbk}} \\ &\le \epsilon(\|\mathrm{L}^{\mathrm{u}}\|^2_{\mathcal{L}(\mathcal{H}_{\Bbbk},\mathbb{R}^{n_{\mathscr{D}}})} + 1)^{\frac{1}{2}}\|\mathrm{y}\|.\end{aligned} \tag{62}$$

Since $\mathscr{U}_\epsilon \subseteq \mathscr{U}_{\mathscr{P}} \subseteq \mathscr{U}_0$, we know that

$$\sup_{\mathrm{w}\in\mathscr{U}_0}\min_{\mathrm{v}\in\mathscr{U}_{\mathscr{P}}}\|\mathrm{v}-\mathrm{w}\|_{\mathcal{V}_{\Bbbk}} \le \sup_{\mathrm{w}\in\mathscr{U}_0}\min_{\mathrm{v}\in\mathscr{U}_\epsilon}\|\mathrm{v}-\mathrm{w}\|_{\mathcal{V}_{\Bbbk}}.$$

Subsequently, due to (61) and (62), we have

$$\|\mathrm{v}^{(\star,\epsilon)} - \mathrm{v}^\star\|^2_{\mathcal{V}_{\Bbbk}} \le \epsilon(\|\mathrm{L}^{\mathrm{u}}\|^2_{\mathcal{L}(\mathcal{H}_{\Bbbk},\mathbb{R}^{n_{\mathscr{D}}})} + 1)^{\frac{1}{2}}\|\mathrm{y}\|. \tag{63}$$

Consequently, it follows that

$$\begin{aligned}\|\mathrm{g}^{(\star,\epsilon)} - \mathrm{g}^\star\|^2_{\mathcal{H}_{\Bbbk}} &\le \|\mathrm{v}^{(\star,\epsilon)} - \mathrm{v}^\star\|^2_{\mathcal{V}_{\Bbbk}} \\ &\le 4\epsilon(\|\mathrm{L}^{\mathrm{u}}\|^2_{\mathcal{L}(\mathcal{H}_{\Bbbk},\mathbb{R}^{n_{\mathscr{D}}})} + 1)^{\frac{1}{2}}\|\mathrm{y}\|^2.\end{aligned}$$

This also shows that $\lim_{\epsilon\to\mathbf{0}}\mathrm{g}^{(\star,\epsilon)} = \mathrm{g}^\star$ in $\mathcal{H}_{\Bbbk}$. Note that, for any $t\in\mathbb{T}$, we have $g_t^{(\star,\epsilon)} = \langle \mathrm{g}^{(\star,\epsilon)}, \Bbbk_t\rangle_{\mathcal{H}_{\Bbbk}}$ and $g_t^\star = \langle \mathrm{g}^\star, \Bbbk_t\rangle_{\mathcal{H}_{\Bbbk}}$. Therefore, from the Cauchy–Schwarz inequality and the reproducing property, one has

$$\begin{aligned}|g_t^{(\star,\epsilon)} - g_t^\star| &= |\langle \mathrm{g}^{(\star,\epsilon)} - \mathrm{g}^\star, \Bbbk_t\rangle_{\mathcal{H}_{\Bbbk}}| \\ &\le \|\mathrm{g}^{(\star,\epsilon)} - \mathrm{g}^\star\|_{\mathcal{H}_{\Bbbk}}\|\Bbbk_t\|_{\mathcal{H}_{\Bbbk}}.\end{aligned}$$

Accordingly, from $\|\Bbbk_t\| = \Bbbk(t,t)^{\frac{1}{2}}$, it follows that

$$|g_t^{(\star,\epsilon)} - g_t^\star| \le \|\mathrm{g}^{(\star,\epsilon)} - \mathrm{g}^\star\|_{\mathcal{H}_{\Bbbk}} \sup_{t\in\mathbb{T}}\Bbbk(t,t)^{\frac{1}{2}}.$$

Since $\sup_{t\in\mathbb{T}}\Bbbk(t,t)^{\frac{1}{2}} < \infty$ and $\lim_{\epsilon\to\mathbf{0}}\|\mathrm{g}^{(\star,\epsilon)} - \mathrm{g}^\star\|_{\mathcal{H}_{\Bbbk}} = 0$, we have $g_t^{(\star,\epsilon)} \xrightarrow{\epsilon\to 0} g_t^\star$, uniformly in $\mathbb{T}$, and proof concludes. □

Remark 2. The property $\sup_{t\in\mathbb{T}}\Bbbk(t,t)^{\frac{1}{2}} < \infty$ is satisfied by the TC kernel and other common kernels in the literature, such as the diagonally/correlated (DC) kernel and the stable spline (SS) kernel (Pillonetto et al., 2014).

Remark 3. In the case of incorrect side-information, we have $\mathrm{g}^{(\mathscr{S})} \notin \mathscr{G}_{\Bbbk}(1,\Omega_{\mathbb{T}})$. Let $\mathrm{g}_\perp^{(\mathscr{S})} \in \mathcal{H}_{\Bbbk}$ be the projection of $\mathrm{g}^{(\mathscr{S})}$ on $\mathscr{G}_{\Bbbk}(1,\Omega_{\mathbb{T}})$, i.e., $\mathrm{g}_\perp^{(\mathscr{S})}$ is defined as

$$\mathrm{g}_\perp^{(\mathscr{S})} := \operatorname{argmin}_{\mathrm{g}\in\mathscr{G}_{\Bbbk}(1,\Omega_{\mathbb{T}})}\|\mathrm{g} - \mathrm{g}^{(\mathscr{S})}\|_{\mathcal{H}_{\Bbbk}}, \tag{64}$$

which exists uniquely (due to Theorem 4 and Peypouquet, 2015). Due to the definition of $\mathrm{g}_\perp^{(\mathscr{S})}$ and since $\mathrm{g}^\star, \mathrm{g}^{(\star,\epsilon)} \in \mathscr{G}_{\Bbbk}(1,\Omega_{\mathbb{T}})$ (see Theorem 11), we have

$$0 < \|\mathrm{g}^{(\mathscr{S})} - \mathrm{g}_\perp^{(\mathscr{S})}\|_{\mathcal{H}_{\Bbbk}} \le \|\mathrm{g}^{(\mathscr{S})} - \mathrm{g}^\star\|_{\mathcal{H}_{\Bbbk}}, \tag{65}$$

and

$$0 < \|\mathrm{g}^{(\mathscr{S})} - \mathrm{g}_\perp^{(\mathscr{S})}\|_{\mathcal{H}_{\Bbbk}} \le \|\mathrm{g}^{(\mathscr{S})} - \mathrm{g}^{(\star,\epsilon)}\|_{\mathcal{H}_{\Bbbk}}. \tag{66}$$

In other words, we have a systematic bias in the estimated impulse response $\mathrm{g}^\star$, and also, in its approximation $\mathrm{g}^{(\star,\epsilon)}$, for all $\epsilon > 0$. Accordingly, in this situation, not including the side-information may result in a more accurate identified model.

6. Optimization algorithm

Due to Theorem 13, the problem to be solved is

$$\begin{aligned}\min_{\mathrm{x}\in\mathbb{R}^m} \quad & \|\mathrm{A}\mathrm{x} - \mathrm{y}\|^2 + \lambda\mathrm{x}^{\mathsf{T}}\Phi\mathrm{x} \\ \text{s.t.} \quad & \mathrm{x}^{\mathsf{T}}(\mathrm{b}_j\mathrm{b}_j^{\mathsf{T}} + \mathrm{c}_j\mathrm{c}_j^{\mathsf{T}})\mathrm{x} \le 1-\epsilon, \quad \forall j\in\mathcal{I}^{(\mathscr{P})},\end{aligned} \tag{67}$$

where $\mathrm{A} := [\mathrm{a}_0,\ldots,\mathrm{a}_{n_{\mathscr{D}}-1}]^{\mathsf{T}}$ and $\mathrm{y} := [y_t]_{t\in\mathscr{T}}$. From the definition of the matrix A, we know that A is the first $n_{\mathscr{D}}$ rows of Φ. The vectors b_j and c_j are respectively the $(n_{\mathscr{D}}+2j+1)$th and the $(n_{\mathscr{D}}+2j+2)$th columns of Φ, for $j\in\mathcal{I}^{(\mathscr{P})}$. Therefore, matrix Φ is sufficient for establishing the optimization problem (67). To this end, for each $i,j\in\mathcal{I}$, we need to obtain the value of $\langle\varphi_i,\varphi_j\rangle_{\mathcal{H}_{\Bbbk}}$, which demands calculating an infinite double summation, when $\mathbb{T} = \mathbb{Z}_+$, or an improper double integral, when $\mathbb{T} = \mathbb{R}_+$. In general, to obtain these values one should employ numerical methods which are essentially inexact and also computationally demanding. On the other hand, one can derive these values analytically in specific but rather general situations, e.g., when the TC kernel is employed. The details are provided in Khosravi and Smith (2021c, Appendix G). Note that (67) is a convex *quadratically constrained quadratic program* program, for which there exist various efficient methods (Boyd & Vandenberghe, 2004). For example, we can utilize methods which are based on *log-barrier functions* (Boyd & Vandenberghe, 2004) and adapt them suitably to this application. For more details, see Appendix H in Khosravi and Smith (2021c).

In many situations, such as in the example given in Section 3, the constraint $|G(\mathrm{j}\omega)| = |\mathcal{F}_\omega(\mathrm{g})|^2 \le 1-\epsilon$ is not binding on the whole frequency range $\Omega_{\mathbb{T}}$. Accordingly, it is not required to impose this constraint for each $\omega\in\mathscr{P}$. Motivated by this fact and Corollary 14, we can introduce an iterative scheme. More precisely, let $\overline{\mathscr{P}}_0 := \emptyset$, and at iteration k, let $\overline{\mathscr{P}}_k$ be a given subset of $\mathscr{P}$. Consider the optimization problem (29) where only the constraints corresponding to the frequencies in $\overline{\mathscr{P}}_k$ are imposed, i.e., we have the following program

$$\begin{aligned}\min_{\mathrm{g}\in\mathcal{H}_{\Bbbk}} \quad & \sum_{t\in\mathscr{T}}(\mathrm{L}_t^{\mathrm{u}}(\mathrm{g}) - y_t)^2 + \lambda\|\mathrm{g}\|^2_{\mathcal{H}_{\Bbbk}}, \\ \text{s.t.} \quad & |\mathcal{F}_\omega(\mathrm{g})|^2 \le 1-\epsilon, \quad \forall\omega\in\overline{\mathscr{P}}_k.\end{aligned} \tag{68}$$

Due to Corollary 14, we know that (68) has a unique solution, denoted by g_k, which can be obtained by solving an equivalent finite-dimensional convex program as in (54). Given g_k, one can check whether the constraint $|\mathcal{F}_\omega(\mathrm{g}_k)|^2 \le 1-\epsilon$ is violated on the remaining frequencies in $\mathscr{P}$. Accordingly, one can obtain the following set

$$\Delta_k\mathscr{P} := \left\{\omega\in\mathscr{P}\backslash\overline{\mathscr{P}}_k \;\middle|\; |\mathcal{F}_\omega(\mathrm{g}_k)|^2 > 1-\epsilon\right\}, \tag{69}$$

Subsequently, we update the frequency set as $\overline{\mathscr{P}}_{k+1} := \overline{\mathscr{P}}_k \cup \Delta_k\mathscr{P}$, and proceed to the iteration $k+1$. The iterative scheme stops when $\Delta_k\mathscr{P} = \emptyset$. This happens either when $\overline{\mathscr{P}}_k = \mathscr{P}$ or the solution g_k satisfies the constraint $|\mathcal{F}_\omega(\mathrm{g}_k)|^2 \le 1-\epsilon$ for all $\omega\in\mathscr{P}$. Consequently, due to Corollary 14, we have $\mathrm{g}_k = \mathrm{g}^{(\star,\epsilon)}$ when the stopping condition is met. It is noteworthy that given the matrix Φ, one can extract $\overline{\Phi}_k$ as a sub-matrix of Φ, and subsequently, the corresponding vectors $\bar{\mathrm{a}}_i$, $\bar{\mathrm{b}}_j$ and $\bar{\mathrm{c}}_j$ in (54) are obtained as columns of $\overline{\Phi}_k$. This fact improves the computational tractability of the proposed approach. Algorithm 1 summarizes this iterative scheme.

This procedure generates a strictly increasing sequence of sets $\emptyset = \overline{\mathscr{P}}_0 \subset \overline{\mathscr{P}}_1 \subset \overline{\mathscr{P}}_2 \subset \cdots \subset \mathscr{P}$. Since $\mathscr{P}$ is a finite set, this sequence is also finite and therefore, Algorithm 1 stops after finite number of iterations.

Remark 4. The partition $\mathscr{P}$ can be designed according to Theorem 11 by employing $\lambda = \vartheta_\lambda\lambda_{\mathrm{unc}}$ in (37), where λ_{unc} is the regularization weight obtained by solving the unconstrained problem, and $\vartheta_\lambda\in(0,1)$ is a small scaling coefficient. One should note that the bound in Theorem 11 is introduced mainly for technical reasons, i.e., it essentially says that if partition $\mathscr{P}$ is sufficiently fine, then constraint $|\mathcal{F}_\omega(\mathrm{g}^{(\star,\epsilon)})| \le 1$ is satisfied at

Algorithm 1 Kernel-Based Identification with Frequency Domain Side-Information $\|G^{(\mathcal{S})}\|_{\mathcal{H}_\infty} \leq 1$

input: data set $\mathscr{D}$, partition set $\mathscr{P}$, kernel $\mathbb{k}$, regularization weight λ, tolerance $\epsilon \in (0,1)$, matrix Φ, unconstrained estimation g_0
$k \leftarrow 0$ and $\mathrm{g}_k \leftarrow \mathrm{g}_0$.
$\overline{\mathscr{P}}_k \leftarrow \emptyset$ and get $\Delta_k\mathscr{P}$ due to (69).
while $\Delta_k\mathscr{P} \neq \emptyset$ **do**
  $\overline{\mathscr{P}} \leftarrow \overline{\mathscr{P}}_k \cup \Delta\mathscr{P}_k$.
  update $\overline{\Phi}$, $\bar{\mathrm{a}}_i$ for $i \in \mathcal{I}^{(\mathrm{u})}$, $\bar{\mathrm{b}}_j$ and $\bar{\mathrm{c}}_j$, for $j \in \mathcal{I}^{(\overline{\mathscr{P}})}$.
  solve optimization problem (54) and $\mathrm{g}_k \leftarrow \sum_i x_i\overline{\varphi}_i$.
  $\mathscr{P}_k \leftarrow \overline{\mathscr{P}}$ and get $\Delta_k\mathscr{P}$ due to (69).
end
Output: $\mathrm{g}^{(\star,\epsilon)}$

all frequencies $\omega \in \Omega_{\mathbb{T}}$. As a result, in practical implementations, we only need to employ $\mathscr{P}$ with small enough mesh($\mathscr{P}$). Accordingly, for a given $\mathscr{P}$, we solve the estimation problem (29), and then, verify the suitability of $\mathscr{P}$ by checking inequality $\max_{\omega\in\Omega_{\mathbb{T}}} |\mathcal{F}_\omega(\mathrm{g}^{(\star,\epsilon)})| \leq 1$, which is equivalent to a single variable optimization problem. If the inequality is not satisfied, we augment $\mathscr{P}$ by including additional frequency points around the locations where the inequality is violated, and subsequently, obtain a finer partition $\mathscr{P}_{\mathrm{new}}$ such that $\mathrm{mesh}(\mathscr{P}_{\mathrm{new}}) \leq \vartheta_{\mathscr{P}}\,\mathrm{mesh}(\mathscr{P})$, where $\vartheta_{\mathscr{P}} \in (0,1)$ is a predefined scalar. Following this, we replace $\mathscr{P}$ in (29) with $\mathscr{P}_{\mathrm{new}}$ and repeat the discussed steps. According to Theorem 11, this procedure terminates after a finite number of iterations.

7. Numerical examples

In this section, we provide numerical examples demonstrating the performance of the proposed scheme in Algorithm 1.

Example 1. We consider the settings of the example given in Section 3. We set $\epsilon = 10^{-5}$ and take partition set $\mathscr{P} = \{\omega_i | i = 0,\ldots,n_{\mathscr{P}}\}$ for interval $[0,\pi]$ such that $\omega_i = \frac{\pi}{n_{\mathscr{P}}} i$, for $i = 0,\ldots,n_{\mathscr{P}}$, with $n_{\mathscr{P}} = 3141$. We employ the TC kernel and apply Algorithm 1. In order to tune the hyperparameters α and λ, we utilize Bayesian optimization with a lower-confidence-bound acquisition function (Srinivas, Krause, Kakade, & Seeger, 2012) to find the hyperparameters minimizing an objective function defined based on a cross-validation procedure. More precisely, for a choice of hyperparameters, the first 100 points of $\mathscr{D}$ are used for training the model and following this, the cost function to be optimized by Bayesian optimization is defined as the validation error calculated using the remaining points of data. For more details, see Appendix I in Khosravi and Smith (2021d). Starting from $\overline{\mathscr{P}}_0 = \emptyset$, we estimate g_0 as the solution of the unconstrained problem. This solution violates the constraints on $\overline{\mathscr{P}}_1 := \{\omega_i | i = 36,\ldots,140\}$. Proceeding from this partition set, we obtain g_1 which satisfies the constraints for all of the frequencies in $\mathscr{P}$. Therefore, $\Delta_1\mathscr{P} = \emptyset$ and the algorithm terminates with $\mathrm{g}^{(\star,\epsilon)} = \mathrm{g}_1$.

Fig. 3 shows the transfer function of $G^{(\star,\epsilon)}$ along with the estimated models $\hat{G}_1$ and $\hat{G}_2$, obtained in Section 3, and also the estimated model $\hat{G}_3$ resulted from the method in Abe et al. (2016) which is briefly reviewed later in this section. One can see that the result of proposed scheme satisfies the side-information constraint, and also fits better to the true transfer function $G^{(\mathcal{S})}$. For quantitative evaluation and comparison of the estimated impulse response, we use *coefficient of determination*, also known as *R-squared*, which is defined as following

$$\mathrm{fit}(\mathrm{g}) = 100 \times \left(1 - \frac{\|\mathrm{g} - \mathrm{g}^{(\mathcal{S})}\|_2}{\|\mathrm{g}^{(\mathcal{S})}\|_2}\right), \tag{70}$$

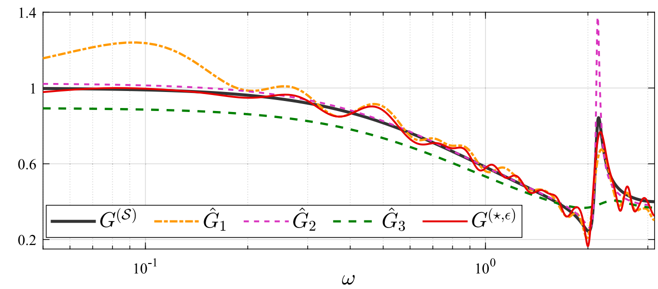

Fig. 3. The transfer function of system (4), the model estimated using the proposed approach, $G^{(\star,\epsilon)}$, and the estimated models $\hat{G}_1$, $\hat{G}_2$ and $\hat{G}_3$.

where g is the estimated impulse response. Here, we have $\mathrm{fit}(\hat{\mathrm{g}}_1) = 87.13\%$, $\mathrm{fit}(\hat{\mathrm{g}}_2) = 87.15\%$, and $\mathrm{fit}(\hat{\mathrm{g}}_3) = 79.08\%$, where $\hat{\mathrm{g}}_i$ is the impulse responses corresponding to $\hat{G}_i$, for $i = 1,2,3$. On the other hand, we have $\mathrm{fit}(\mathrm{g}^{(\star,\epsilon)}) = 90.84\%$ which shows an improvement in the estimation as well as satisfying the given side-information. Note that while $\hat{G}_3$ also satisfies the side-information constraint, the proposed method performs significantly better in terms of fitting performance. △

The method in Abe et al. (2016) is a variant of the subspace identification method incorporating frequency domain side-information, and is denoted by $\mathrm{FDI}_{\mathrm{sub}}$ in the following. In this approach, initially a sequence of state variables is estimated, and then, the subspace identification method is formulated with matrix inequalities, coming from the Kalman–Yakubovich–Popov (KYP) lemma, imposing the side-information. The resulting nonlinear program is reduced to a convex one using appropriate transformations. Compared to the proposed method, this approach can be employed in multi-input–multi-output (MIMO) cases and produces state-space models, which can be directly used in applications such as controller design.

Example 2. In this example, we perform numerical experiments to compare the proposed method with (Abe et al., 2016). We generate randomly four sets of 150 stable systems using `drss` MATLAB's function with orders in the range of $\{10,\ldots,30\}$ and poles not larger than 0.98. The systems are normalized with their $\mathcal{H}_\infty$-norm, and then, actuated with a realization of standard random white Gaussian signal $\mathrm{u} = (u_t)_{t=0}^{n_{\mathscr{D}}-1}$, i.e., $u_0,\ldots,u_{n_{\mathscr{D}}-1}$ are i.i.d. samples of $\mathcal{N}(0,1)$. The output of system is corrupted with additive Gaussian noise. The variance of output noise is chosen such that we have 10 dB, 20 dB, 30 dB and 40 dB signal-to-noise ratio (SNR) in the respective sets of the systems. Similar to the previous example, we tune the hyperparameters of both the proposed approach and the $\mathrm{FDI}_{\mathrm{sub}}$ using Bayesian optimization (Srinivas et al., 2012) and cross-validation. The settings for the proposed approach are similar to ones employed in Example 1. The box plots of results are shown in Fig. 4. While, we can see that the performance of the both of these approaches improves when SNR increases, the kernel-based approach significantly outperforms the subspace-based method $\mathrm{FDI}_{\mathrm{sub}}$.

Discussion: The $\mathrm{FDI}_{\mathrm{sub}}$ approach (Abe et al., 2016) suffers from the well-known model order selection issue (Ljung et al., 2020). On the other hand, the proposed scheme is a derivation of kernel-based regularization methods, and therefore, tuning the complexity of model is performed by the powerful concept of estimating continuous regularization hyperparameters rather than picking an integer order based on a selection rule (Ljung et al., 2020; Pillonetto et al., 2014). Moreover, while the proposed approach works directly with input–output data, $\mathrm{FDI}_{\mathrm{sub}}$ employs an estimation of the state trajectory which leads to model estimation prone to high variance and noisy results. △

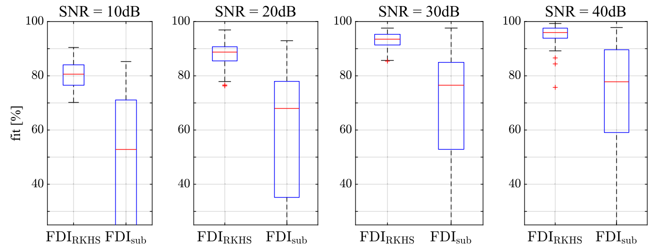

Fig. 4. The performance of $\mathrm{FDI_{sub}}$ (Abe et al., 2016) is compared with the proposed method, $\mathrm{FDI_{RKHS}}$.

The next example considers Problem 1 in the continuous-time case and demonstrates the role of proposed approach for the continuous-time systems.

Example 3. Consider the *unknown* continuous-time system $\mathcal{S}$ with transfer function $G^{(\mathcal{S})}(s)$ defined (Scandella, Mazzoleni, Formentin, & Previdi, 2021) as

$$G^{(\mathcal{S})}(s) = -\frac{2s^3 + 3.6s^2 + 2.095s + 0.396s}{0.461s^4 + 2.628s^3 + 4.389s^2 + 2.662s + 0.519}, \quad (71)$$

with side-information $\|G^{(\mathcal{S})}\|_{\mathcal{H}_\infty} \leq 1$. Let the system be initially at rest and actuated with a random switching pulse signal as shown in Fig. 5. Then, the output of system is measure at $n_{\mathscr{D}} = 250$ time instants $t_0, \ldots, t_{n_{\mathscr{D}}-1} \in [0, 10]$, where $t_k = kT_s + \delta_k$, for $k = 0, \ldots, n_{\mathscr{D}} - 1$, with $T_s = 0.04$ and $\delta_k \sim \mathrm{Uniform}([0, T_s])$. The output measurements, $\{y_{t_i}\}_{i=0}^{n_{\mathscr{D}}-1}$, are subject to additive white Gaussian noise such that the SNR is 20 dB. We interpolate the output measurement values at time instants $\bar{t}_k = (k+1)T_s$, for $k = 0, \ldots, n_{\mathscr{D}} - 1$, using shape-preserving piecewise cubic interpolation available in MATLAB's function `interp1` and `pchip` option. The SNR in the interpolated outputs, $\{\bar{y}_{t_i}\}_{i=0}^{n_{\mathscr{D}}-1}$, is 20.68 dB.

We identify the system through *direct* and *indirect* approaches.[2] For the direct approaches, we employ the interpolated data and CONTSID TOOLBOX (Garnier & Gilson, 2018) to obtain transfer function estimations $\hat{G}_1$ and $\hat{G}_2$, respectively, using `tfsrivc` and `rivc` functions with known orders of system. In indirect approaches, a discrete-time impulse response is estimated using the interpolated data, and then the continuous-time version is derived by the shape-preserving piecewise cubic interpolation method (Kahaner, Moler, & Nash, 1989). To this end, we use the subspace method (Abe et al., 2016) explained in Example 2 and also, the discrete-time version of the proposed method. Let the corresponding transfer functions be denoted respectively by $\hat{G}_{\mathrm{sub}}$ and $G_{\mathrm{ind}}^{(\star,\epsilon)}$. In obtaining $G_{\mathrm{ind}}^{(\star,\epsilon)}$, ϵ is set to 10^{-3} and the rest of settings are similar to Example 1. Starting from $\overline{\mathscr{P}}_0 = \emptyset$, the algorithm terminates in the second iteration with $\overline{\mathscr{P}}_1 := \{\omega_i | i = 50, \ldots, 76\}$.

In addition to the above methods, we identify the system in a direct approach based on the proposed algorithm for the case of continuous-time, and using the original uninterpolated measurement data. Accordingly, we apply Algorithm 1 where ϵ is chosen as 10^{-3}, the partition set is taken as $\mathscr{P} = \{\omega_i = 10^{-2}i | i = 0, \ldots, n_{\mathscr{P}} = 10^4\}$, a TC kernel is employed, and the hyperparameters are tuned similarly to Example 1. Initially, we have $\overline{\mathscr{P}}_0 = \emptyset$ where the resulting solution, g_0, violates the constraints on $\Delta\mathscr{P}_0 := \{\omega_i | i = 159, \ldots, 205\}$. The algorithm terminates in the second iteration. Denote the estimated impulse response and the corresponding transfer function respectively by $\mathrm{g}_{\mathrm{dir}}^{(\star,\epsilon)}$ and $G_{\mathrm{dir}}^{(\star,\epsilon)}$.

[2] The indirect identification of continuous-time systems consists of estimating a discrete-time model using conventional discrete-time transfer function estimation techniques, followed by converting the estimation result to a continuous-time model. On the other hand, direct identification methods estimate a continuous-time model directly from the measurement data (see Garnier, Wang, & Young, 2008 for further explanation.)

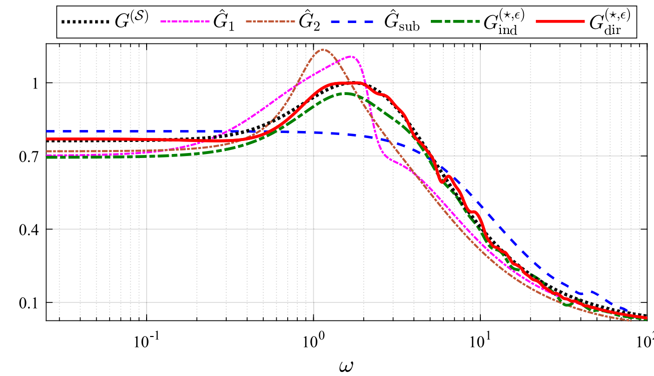

Fig. 5. The transfer function of system (71), the model estimated using the proposed approach, $G_{\mathrm{dir}}^{(\star,\epsilon)}$, and the estimated models $\hat{G}_1$, $\hat{G}_2$, $\hat{G}_{\mathrm{sub}}$ and $G_{\mathrm{ind}}^{(\star,\epsilon)}$.

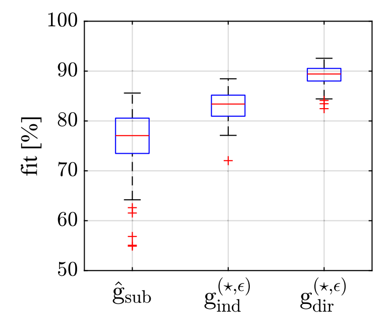

Fig. 6. Box plots of the R-squared metric for the estimation results of the methods discussed in Example 3.

In Fig. 5, the estimated transfer functions $\hat{G}_1$, $\hat{G}_2$, $\hat{G}_{\mathrm{sub}}$, $G_{\mathrm{ind}}^{(\star,\epsilon)}$ and $G_{\mathrm{dir}}^{(\star,\epsilon)}$ are shown and graphically compared with $G^{(\mathcal{S})}$. The side-information is satisfied only by $\hat{G}_{\mathrm{sub}}$, $G_{\mathrm{ind}}^{(\star,\epsilon)}$ and $G_{\mathrm{ind}}^{(\star,\epsilon)}$, and it is violated by $\hat{G}_1$ and $\hat{G}_2$. Indeed, we have $\|\hat{G}_1\|_{\mathcal{H}_\infty} = 1.11$ and $\|\hat{G}_2\|_{\mathcal{H}_\infty} = 1.17$. To evaluate quantitatively the estimation results, we employ *R-squared* metric defined in (70). The fitting results are $\mathrm{fit}(\hat{\mathrm{g}}_1) = 78.93\%$, $\mathrm{fit}(\hat{\mathrm{g}}_2) = 63.37\%$, $\mathrm{fit}(\hat{\mathrm{g}}_{\mathrm{sub}}) = 73.76\%$, $\mathrm{fit}(\mathrm{g}_{\mathrm{ind}}^{(\star,\epsilon)}) = 87.15\%$, and $\mathrm{fit}(\mathrm{g}_{\mathrm{dir}}^{(\star,\epsilon)}) = 91.86\%$, where $\hat{\mathrm{g}}_1$, $\hat{\mathrm{g}}_2$, $\hat{\mathrm{g}}_{\mathrm{sub}}$ and $\mathrm{g}_{\mathrm{ind}}^{(\star,\epsilon)}$ are the impulse responses corresponding to $\hat{G}_1$, $\hat{G}_2$, $\hat{G}_{\mathrm{sub}}$ and $G_{\mathrm{ind}}^{(\star,\epsilon)}$, respectively. Therefore, the proposed method significantly outperforms the other schemes and also satisfies the side-information.

For further comparison, we perform a Monte Carlo experiment with a set of 100 runs and settings similar to the numerical experiment above, where the SNR level is equal to 20 dB and the regular sampling is employed. We estimate the impulse response of the system in each run using the direct and indirect implementation of the proposed method and the previously mentioned subspace identification approach, i.e., we obtain impulse response estimations $\mathrm{g}_{\mathrm{dir}}^{(\star,\epsilon)}$, $\mathrm{g}_{\mathrm{ind}}^{(\star,\epsilon)}$ and $\hat{\mathrm{g}}_{\mathrm{sub}}$. Fig. 6 shows the boxplot comparing the estimation performance of these methods. Also, Table 1 provides the bias, variance and mean squared error (MSE) of the estimations.

Discussion: For the proposed method, we observe that the direct implementation significantly outperforms the indirect approach. This observation is in alignment with the literature on continuous-time system identification (Garnier et al., 2008). Moreover, one can see that the kernel-based approaches show better estimation performance compared to the subspace

Table 1
Bias, variance and MSE comparison between the continuous-time model estimation methods.

Method	$\mathrm{Bias}^2(\hat{g})$ [$\times 10^{-6}$]	$\mathrm{Var}(\hat{g})$ [$\times 10^{-6}$]	$\mathrm{MSE}(\hat{g})$ [$\times 10^{-6}$]
$\hat{G}_{\mathrm{sub}}$	6.39	7.97	14.36
$G_{\mathrm{ind}}^{(\star,\epsilon)}$	5.33	1.73	7.07
$G_{\mathrm{dir}}^{(\star,\epsilon)}$	2.25	0.63	2.88

method, as shown in Example 2. Furthermore, the proposed method can be implemented directly even when in the case of irregular output sampling. These features highlight the importance of the developed identification scheme for the estimation of continuous-time impulse responses. △

8. Conclusion

System identification with frequency domain side-information has been studied in this paper. We have employed an RKHS framework for both discrete-time and continuous-time dynamics. The problem is formulated as an infinite-dimensional optimization problem for fitting a stable impulse response to the data and satisfying an $\mathcal{H}_\infty$-norm constraint. The problem is well-defined and convex with a unique solution. We have proposed a finite-dimensional convex quadratically constrained quadratic program that tightly approximates the solution. It is shown that the approximation is uniformly exact, where this bound can be calculated a priori from the kernel hyperparameters and the measurement data. The effectiveness of the discussed method is illustrated by several numerical examples.

References

Abe, Y., Inoue, M., & Adachi, S. (2016). Subspace identification method incorporated with a priori information characterized in frequency domain. In *European control conference* (pp. 1377–1382). IEEE.

Ahmadi, A. A., & El Khadir, B. (2020). Learning dynamical systems with side information (short version). *Proceedings of Machine Learning Research*, *120*, 718–727.

Alenany, A., Shang, H., Soliman, M., & Ziedan, I. (2011). Improved subspace identification with prior information using constrained least squares. *IET Control Theory & Applications*, *5*(13), 1568–1576.

Antoulas, A. C. (2005). *Approximation of large-scale dynamical systems*. SIAM.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, *68*(3), 337–404.

Benvenuti, L., De Santis, A., & Farina, L. (2002). On model consistency in compartmental systems identification. *Automatica*, *38*(11), 1969–1976.

Berlinet, A., & Thomas-Agnan, C. (2011). *Reproducing kernel hilbert spaces in probability and statistics*. Springer Science and Business Media.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Brogliato, B., Lozano, R., Maschke, B., & Egeland, O. (2007). *Dissipative systems analysis and control*. Springer.

Carmeli, C., De Vito, E., & Toigo, A. (2006). Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem. *Analysis and Applications*, *4*(4), 377–408.

Chen, T. (2018). On kernel design for regularized LTI system identification. *Automatica*, *90*, 109–122.

Chen, T., & Pillonetto, G. (2018). On the stability of reproducing kernel Hilbert spaces of discrete-time impulse responses. *Automatica*, *95*, 529–533.

Chiuso, A., & Pillonetto, G. (2019). System identification: A machine learning perspective. *Annual Review of Control, Robotics, and Autonomous Systems*, *2*, 281–304.

De Santis, A., & Farina, L. (2002). Identification of positive linear systems with Poisson output transformation. *Automatica*, *38*(5), 861–868.

Dinuzzo, F., & Schölkopf, B. (2012). The representer theorem for Hilbert spaces: A necessary and sufficient condition. In *Advances in neural information processing systems* (pp. 189–196).

Fazel, M., Pong, T. K., Sun, D., & Tseng, P. (2013). Hankel matrix rank minimization with applications to system identification and realization. *SIAM Journal on Matrix Analysis and Applications*, *34*(3), 946–977.

Garnier, H., & Gilson, M. (2018). CONTSID: A MATLAB toolbox for standard and advanced identification of black-box continuous-time models. *IFAC-PapersOnLine*, *51*(15), 688–693.

Garnier, H., Wang, L., & Young, P. C. (2008). Direct identification of continuous-time models from sampled data: Issues, basic solutions and relevance. In *Identification of continuous-time models from sampled data* (pp. 1–29). Springer.

Goethals, I., Van Gestel, T., Suykens, J., Van Dooren, P., & De Moor, B. (2003). Identification of positive real models in subspace identification by using regularization. *IEEE Transactions on Automatic Control*, *48*(10), 1843–1847.

Grussler, C., Umenberger, J., & Manchester, I. R. (2017). Identification of externally positive systems. In *Conference on decision and control* (pp. 6549–6554). IEEE.

Haddad, W. M., & Chellaboina, V. (2011). *Nonlinear dynamical systems and control*. Princeton University Press.

Hara, K., Inoue, M., & Sebe, N. (2019). Learning Koopman operator under dissipativity constraints. arXiv:1911.03884.

Hoagg, J. B., Lacy, S. L., Erwin, R. S., & Bernstein, D. S. (2004). First-order-hold sampling of positive real systems and subspace identification of positive real models. In *American control conference*: *Vol. 1*, (pp. 861–866). IEEE.

Inoue, M. (2019). Subspace identification with moment matching. *Automatica*, *99*, 22–32.

Kahaner, D., Moler, C., & Nash, S. (1989). *Numerical methods and software*. Prentice–Hall, Inc..

Khosravi, M., Iannelli, A., Yin, M., Parsi, A., & Smith, R. S. (2020). Regularized system identification: A hierarchical Bayesian approach. *IFAC-PapersOnLine*, *53*(2), 406–411, IFAC World Congress 2020.

Khosravi, M., & Smith, R. S. (2019). Kernel-based identification of positive systems. In *Conference on decision and control* (pp. 1740–1745).

Khosravi, M., & Smith, R. S. (2021a). Convex nonparametric formulation for identification of gradient flows. *IEEE Control Systems Letters*, *5*(3), 1097–1102.

Khosravi, M., & Smith, R. S. (2021c). Kernel-based identification with frequency domain side-information. arXiv preprint arXiv:2111.00410.

Khosravi, M., & Smith, R. S. (2021d). Kernel-based impulse response identification with side-information on steady-state gain. arXiv preprint arXiv:2111.00409.

Khosravi, M., & Smith, R. S. (2021e). Nonlinear system identification with prior knowledge on the region of attraction. *IEEE Control Systems Letters*, *5*(3), 1091–1096.

Khosravi, M., & Smith, R. S. (2021f). On robustness of kernel-based regularized system identification. *IFAC-PapersOnLine*, *54*(7), 749–754, IFAC Symposium on System Identification.

Khosravi, M., & Smith, R. S. (2021g). Regularized identification with internal positivity side-information. arXiv preprint arXiv:2111.00407.

Khosravi, M., & Smith, R. S. (2023). The existence and uniqueness of solutions for kernel-based system identification. *Automatica*, *148*, 110728.

Khosravi, M., Yin, M., Iannelli, A., Parsi, A., & Smith, R. S. (2020). Low-complexity identification by sparse hyperparameter estimation. *IFAC-PapersOnLine*, *53*(2), 412–417, IFAC World Congress 2020.

Koch, A., Berberich, J., & Allgöwer, F. (2020). Provably robust verification of dissipativity properties from data. arXiv:2006.05974.

Koch, A., Montenbruck, J. M., & Allgöwer, F. (2019). Sampling strategies for data-driven inference of input-output system properties. arXiv:1910.08919.

Lacy, S. L., & Bernstein, D. S. (2003). Subspace identification with guaranteed stability using constrained optimization. *IEEE Transactions on Automatic Control*, *48*(7), 1259–1263.

Ljung, L. (1999). *System identification: Theory for the user*. Prentice Hall.

Ljung, L. (2010). Perspectives on system identification. *Annual Reviews in Control*, *34*(1), 1–12.

Ljung, L., Chen, T., & Mu, B. (2020). A shift in paradigm for system identification. *International Journal of Control*, *93*(2), 173–180.

Ljung, L., & Singh, R. (2012). Version 8 of the MATLAB system identification toolbox. *IFAC-PapersOnLine*, *45*(16), 1826–1831.

Marconato, A., Schoukens, M., & Schoukens, J. (2016). Filter-based regularisation for impulse response modelling. *IET Control Theory & Applications*, *11*(2), 194–204.

Miller, D. N., & De Callafon, R. A. (2013). Subspace identification with eigenvalue constraints. *Automatica*, *49*(8), 2468–2473.

Mohan, K., & Fazel, M. (2010). Reweighted nuclear norm minimization with application to system identification. In *American control conference* (pp. 2953–2959). IEEE.

Müller, M. I., Valenzuela, P. E., Proutiere, A., & Rojas, C. R. (2017). A stochastic multi-armed bandit approach to nonparametric $\mathcal{H}_\infty$-norm estimation. In *Conference on decision and control* (pp. 4632–4637).

Okada, M., & Sugie, T. (1996). Subspace system identification considering both noise attenuation and use of prior knowledge. In *Conference on decision and control*: *Vol. 4*, (pp. 3662–3667). IEEE.

Peypouquet, J. (2015). *Convex optimization in normed spaces: Theory, methods and examples*. Springer.

Pillonetto, G., Chen, T., Chiuso, A., Nicolao, G. D., & Ljung, L. (2016). Regularized linear system identification using atomic, nuclear and kernel-based norms: The role of the stability constraint. *Automatica*, *69*, 137–149.

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.
Romer, A., Berberich, J., Köhler, J., & Allgöwer, F. (2019). One-shot verification of dissipativity properties from input–output data. *IEEE Control Systems Letters*, *3*(3), 709–714.
Romer, A., Montenbruck, J. M., & Allgöwer, F. (2017). Determining dissipation inequalities from input-output samples. *IFAC-PapersOnLine*, *50*(1), 7789–7794.
Scandella, M., Mazzoleni, M., Formentin, S., & Previdi, F. (2021). Kernel-based identification of asymptotically stable continuous-time linear dynamical systems. *International Journal of Control*, 1–14.
Schölkopf, B., Herbrich, R., & Smola, A. J. (2001). A generalized representer theorem. In *International conference on computational learning theory* (pp. 416–426). Springer.
Schoukens, J., & Ljung, L. (2019). Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine*, *39*(6), 28–99.
Shah, P., Bhaskar, B. N., Tang, G., & Recht, B. (2012). Linear system identification via atomic norm regularization. In *Conference on decision and control* (pp. 6265–6270).
Smith, R. S. (2014). Frequency domain subspace identification using nuclear norm minimization and Hankel matrix realizations. *IEEE Transactions on Automatic Control*, *59*(11), 2886–2896.
Srinivas, N., Krause, A., Kakade, S. M., & Seeger, M. W. (2012). Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, *58*(5), 3250–3265.
Umenberger, J., & Manchester, I. R. (2016). Scalable identification of stable positive systems. In *Conference on decision and control* (pp. 4630–4635). IEEE.
Umenberger, J., & Manchester, I. R. (2018). Specialized interior-point algorithm for stable nonlinear system identification. *IEEE Transactions on Automatic Control*, *64*(6), 2442–2456.
Willems, J. C. (1972). Dissipative dynamical systems part I: General theory. *Archive for Rational Mechanics and Analysis*, *45*(5), 321–351.
Yoshimura, S., Matsubayashi, A., & Inoue, M. (2019). System identification method inheriting steady-state characteristics of existing model. *International Journal of Control*, *92*(11), 2701–2711.
Zadeh, L. (1956). On the identification problem. *IRE Transactions on Circuit Theory*, *3*(4), 277–281.
Zames, G. (1966). On the input-output stability of time-varying nonlinear feedback systems part one: Conditions derived using concepts of loop gain, conicity, and positivity. *IEEE Transactions on Automatic Control*, *11*(2), 228–238.
Zheng, M., & Ohta, Y. (2021). Bayesian positive system identification: Truncated Gaussian prior and hyperparameter estimation. *Systems & Control Letters*, *148*, Article 104857.

Mohammad Khosravi is an assistant professor at Delft Center for Systems and Control (DCSC), Delft University of Technology. He received a B.Sc. in electrical engineering and a B.Sc. in mathematical sciences from the Sharif University of Technology, Tehran, Iran, in 2011. He obtained a postgraduate diploma in mathematics from ICTP, Trieste, Italy, in 2012. He was a research assistant in the mathematical biology group at Institute for Research in Fundamental Sciences, Iran, from 2012 to 2014. He received his MASc degree in electrical and computer engineering from Concordia University, Montreal, Canada, in 2016. He obtained his Ph.D. from the Swiss Federal Institute of Technology (ETH), Zürich, in 2022. He has won several awards, including the Gold Medal of the National Mathematics Olympiad, the Outstanding Student Paper Award in CDC 2020, the Silver Medal of ETH Zürich, and the Outstanding Reviewer Award for IEEE Journal of Control Systems Letters. His research interests involve data-driven and learning-based methods in modeling, model reduction, optimization, and control of dynamical systems and their applications in buildings, energy, industry, and thermodynamic and power systems.

Roy S. Smith is a professor of Electrical Engineering at the Swiss Federal Institute of Technology (ETH), Zürich. Prior to joining ETH in 2011, he was on the faculty of the University of California, Santa Barbara, from 1990 to 2010. His Ph.D. is from the California Institute of Technology (1990) and his undergraduate degree is from the University of Canterbury (1980) in his native New Zealand. He has been a long-time consultant to the NASA Jet Propulsion Laboratory and has industrial experience in automotive control and power system design. His research interests involve the modeling, identification, and control of uncertain systems. Particular control application domains of interest include chemical processes, flexible structure vibration, spacecraft and vehicle formations, aerodynamic control of kites, automotive engines, Mars aeromaneuvering entry design, building and energy hub control, and thermoacoustic machines. He is a Fellow of the IEEE and the IFAC, an Associate Fellow of the AIAA, and a member of SIAM.