# Kernel-based Impulse Response Identification with Side-Information on Steady-State Gain

Khosravi, M.; Smith, R. S.

**DOI**
10.1109/TAC.2023.3243099

**Publication date**
2023

**Document Version**
Final published version

**Published in**
IEEE Transactions on Automatic Control

# Kernel-Based Impulse Response Identification With Side-Information on Steady-State Gain

M. Khosravi, *Member, IEEE*, and R. S. Smith, *Fellow, IEEE*

***Abstract*—In this article, we consider the problem of system identification when side-information is available on the steady-state gain (SSG) of the system. We formulate a general nonparametric identification method as an infinite-dimensional constrained convex program over the reproducing kernel Hilbert space (RKHS) of stable impulse responses. The objective function of this optimization problem is the empirical loss regularized with the norm of RKHS, and the constraint is considered for enforcing the integration of the SSG side-information. The proposed formulation addresses both the discrete-time and continuous-time cases. We show that this program has a unique solution obtained by solving an equivalent finite-dimensional convex optimization. This solution has a closed-form when the empirical loss and regularization functions are quadratic and exact side-information is considered. We perform extensive numerical comparisons to verify the efficiency of the proposed identification methodology.**

***Index Terms*—Kernel-based identification method, side-information, steady-state gain (SSG).**

## I. Introduction

System identification is a well-established research area on the theory and techniques of creating mathematical abstractions for the dynamical systems using their measurement data [1]. In various situations, identifying a dynamical system can be beyond a mere model fitting to the data, and additionally, we may need to include particular known characteristics of the system into the model. More precisely, together with the measurement data, we might be provided with certain so-called *side-information*, which is indeed a specific qualitative or quantitative knowledge to be incorporated in the identified model. Integrating side-information can improve the identification performance by rejecting spurious model candidates, which are common when the measurement data is scarce, highly noise-contaminated, or generated by insufficient excitation [2], [3], [4]. Various forms of side-information, such as stability, dissipativity, and region of attraction, are considered in identifying nonlinear dynamical systems [5], [6], [7], [8]. On the other hand, due to the importance of linear systems in practice, different sorts of side-information are included in their identification procedure, e.g., location of poles [9], structural properties [10], moments and derivatives of the transfer function [11], passivity [12], internal low complexity [13], positivity [14], stability [4], [15], [16], and generic frequency domain attributes [17].

The steady-state gain (SSG) information has particular importance from the control perspective, e.g., in closed-loop design and model predictive control [18], [19], [20]. This information may be obtained, in exact or approximate form, through the structure of the system, from the experimental or historical collected data, or by designing and performing suitable experiments. Hence, integrating the SSG side-information into the identified model is of particular interest. To this end, various heuristics are introduced based on the subspace identification approach [3], [21], [22], [23], [24]. Indeed, to identify a finite impulse response (FIR) model for the system, the subspace method can be employed in the multistep ahead prediction form [3], [21]. Following this and using a Bayesian approach, the SSG side-information can be encoded in the covariance of the prior distribution [3]. On the other hand, a frequentist framework is employed in [21], [22], [23], and [24], where the SSG side-information is incorporated by imposing linear constraints. Moreover, to leverage the previously mentioned advantages of the kernel-based approach, Bayesian FIR estimation methods are proposed in [25] and [26], where kernel-based priors are employed and the SSG side-information is integrated into the resulting estimation problem. The identification scheme in [25] first estimates the step response of the system, and then, the impulse response is obtained via a naïve discrete derivative calculation, which is prone to numerical imprecision and instability. On the other hand, while the method introduced in [26] improves the estimation performance approach in [25], the proposed formulation is incapable of including deterministic information on the SSG of the system. The abovementioned identification approaches are only applicable when a large set of high-quality data is available. Furthermore, they are limited to relatively short FIR estimation and fast decaying dynamics. Therefore, these estimation methodologies are not suitable for infinite impulse responses and continuous-time systems, particularly when the dynamics have a very slowly decaying impulse response and considerably long memory.

In this work, we develop a nonparametric kernel-based identification approach where the SSG side-information is integrated into the proposed scheme. We employ suitable infinite-dimensional RKHSs of stable impulse responses as the hypothesis space [4] to develop a unified formulation for the continuous-time and discrete-time cases. The identification problem is expressed as a constrained optimization where a generic regularized empirical loss is minimized subject to a suitably designed constraint encoding the available side-information on the SSG of the system. The resulting flexible formulation allows considering various issues, including statistical robustness and the presence of outliers. We show that the SSG linear functional is continuous on the employed RKHS, which implies that the problem is well-defined by guaranteeing the existence and uniqueness of the solution. For the initial infinite-dimensional formulation of the identification problem, we derive an equivalent finite-dimensional convex program with a unique solution. Accordingly, we introduce a tractable identification algorithm integrating the SSG side-information. Furthermore, we provide results for improving the computational complexity of the presented approach by obtaining the closed-form of quantities used in the algorithm. We perform extensive numerical simulations confirming the efficacy of the proposed identification method.

## II. Notation and Preliminaries

The set of natural numbers, the set of nonnegative integers, the set of real numbers, the set of nonnegative real numbers, the $n$-dimensional

Manuscript received 5 November 2022; accepted 30 December 2022. Date of publication 7 February 2023; date of current version 27 September 2023. Recommended by Associate Editor Paolo Rapisarda. *(Corresponding author: M. Khosravi.)*

M. Khosravi is with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands (e-mail: mohammad.khosravi@tudelft.nl).

R. S. Smith is with the Automatic Control Laboratory, ETH Zurich, Zurich, Switzerland (e-mail: rsmith@control.ee.ethz.ch).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2023.3243099.

Digital Object Identifier 10.1109/TAC.2023.3243099

Euclidean space and the space of $n$ by $m$ real matrices are denoted by $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}^n$ and $\mathbb{R}^{n\times m}$, respectively. The $i$th entry of vector $\mathrm{a}$ is denoted by $[\mathrm{a}]_{(i)}$, and the entry of matrix $\mathrm{A}$ at the $i$th row and the $j$th column is denoted by $[\mathrm{A}]_{(i,j)}$. 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 $\mathrm{u} = (u(x))_{x\in\mathscr{X}}$. Depending on the context of discussion, $\mathscr{L}^\infty$ refers either to $\ell^\infty(\mathbb{Z})$ or $L^\infty(\mathbb{R})$, i.e., the space of bounded signals. Similarly, $\mathscr{L}^1$ is either $\ell^1(\mathbb{Z}_+)$ or $L^1(\mathbb{R}_+)$, i.e., the space of stable impulse responses. Given kernel $\Bbbk : \mathbb{T}\times\mathbb{T} \to \mathbb{R}$, $\mathscr{H}_\Bbbk$ denotes the corresponding RKHS with inner product $\langle\cdot,\cdot\rangle_{\mathscr{H}_\Bbbk}$ and norm $\|\cdot\|_{\mathscr{H}_\Bbbk}$. We drop the subscript $\mathscr{H}_\Bbbk$ when it is clear from the context. Also, $\Bbbk_s$ denotes $\Bbbk(s,\cdot) \in \mathscr{H}_\Bbbk$, for any $s \in \mathbb{T}$. For $p \in \{1,\infty\}$, the norm in $\mathscr{L}^p$ is denoted by $\|\cdot\|_p$. Given $\mathscr{V} \subseteq \mathbb{X}$, the linear span of $\mathscr{V}$, denoted by 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. Give bounded signal $\mathrm{u} = (u_s)_{s\in\mathbb{T}_\pm} \in \mathscr{L}^\infty$ and $t \in \mathbb{T}_\pm$, the linear map $\mathrm{L}^{\mathrm{u}}_t : \mathscr{L}^1 \to \mathbb{R}$ is defined as $\mathrm{L}^{\mathrm{u}}_t(\mathrm{g}) := \sum_{s\in\mathbb{Z}_+} g_s u_{t-s}$, when $\mathbb{T} = \mathbb{Z}_+$, and $\mathrm{L}^{\mathrm{u}}_t(\mathrm{g}) := \int_{\mathbb{R}_+} g_s u_{t-s}\,\mathrm{d}s$, when $\mathbb{T} = \mathbb{R}_+$.

## III. Identification With SSG Side-Information

Let $\mathcal{S}$ be a stable LTI system with impulse response $\mathrm{g}^{(\mathcal{S})} := (g_t^{(\mathcal{S})})_{t\in\mathbb{T}} \in \mathbb{R}^{\mathbb{T}}$, where $\mathbb{T} := \mathbb{Z}_+$, for the case of discrete-time, and, $\mathbb{T} := \mathbb{R}_+$, for the case of continuous-time. The SSG of system $\mathcal{S}$ is $\ell_0(\mathrm{g}^{(\mathcal{S})})$, where $\ell_0$ is the linear operator defined on the space of stable impulse responses as

$$\ell_0(\mathrm{g}) := \begin{cases} \sum_{t\in\mathbb{Z}_+} g_t, & \text{if } \mathbb{T} = \mathbb{Z}_+ \\ \int_{\mathbb{R}_+} g_t\,\mathrm{d}t, & \text{if } \mathbb{T} = \mathbb{R}_+ \end{cases} \tag{1}$$

for any $\mathrm{g} = (g_t)_{t\in\mathbb{T}} \in \mathscr{L}^1$. Let $\mathrm{u} = (u_t)_{t\in\mathbb{T}}$ be a bounded input signal for system $\mathcal{S}$, and the corresponding output be measured at time instants $\mathscr{T} := \{t_i | i = 1,\dots,n_\mathscr{D}\}$, where $n_\mathscr{D} \in \mathbb{N}$ denotes the number of measurement samples. From the definition of $\mathrm{L}^{\mathrm{u}}_t$, the measured output of the system at $t \in \mathscr{T}$, denoted by $y_t$, is

$$y_t := \mathrm{L}^{\mathrm{u}}_t\left(\mathrm{g}^{(\mathcal{S})}\right) + w_t, \qquad t \in \mathscr{T} \tag{2}$$

where $\{w_t | t \in \mathscr{T}\}$ are the measurement uncertainty. Consequently, we are provided with the set of input–output data $\mathscr{D} = \{(u_t, y_t) | t \in \mathscr{T}\}$. In addition to $\mathscr{D}$, suppose that we know the SSG of the system. Accordingly, one may ask whether the given SSG side-information is naturally encoded in the identification of system $\mathcal{S}$. We elucidate this issue in the following demonstrative numerical example.

*Example.* Consider continuous-time system $\mathcal{S}$ described by the following transfer function $G^{(\mathcal{S})}(s) = \frac{s+2}{s^2+s+2}$, with the step response denoted by $\mathrm{s}^{(\mathcal{S})}$. The system is initially at rest, and then, it is actuated by a random switching pulse signal in the time interval [0, 100]. The output of system is uniformly measured with the sampling frequency of 2 Hz and the signal-to-noise ratio (SNR) of 20 dB. Let the SSG of the system be given, i.e., we know that $G^{(\mathcal{S})}(0) = 1$. The impulse response of the system can be estimated using *direct* and *indirect* methods [27]. In the direct approach, we use the `tfsrivc` function provided by CONTSID Toolbox [28] with the known order of the system. Let $\hat{\mathrm{g}}_1$ and $\hat{\mathrm{s}}_1$, respectively, denote the impulse response and the step response of the resulting identified model. Also, we identify the system indirectly by employing MATLAB's `n4sid` function to estimate a discrete-time model, and subsequently, the continuous-time impulse response is obtained from a linear interpolation of the discrete-time estimate. Let the resulting impulse and step responses be denoted by $\hat{\mathrm{g}}_2$ and $\hat{\mathrm{s}}_2$, respectively. As shown in Fig. 1, the steady-state values for $\hat{\mathrm{s}}_1$ and $\hat{\mathrm{s}}_2$ are, respectively, 0.85 and 1.22, meaning that the SSGs have a

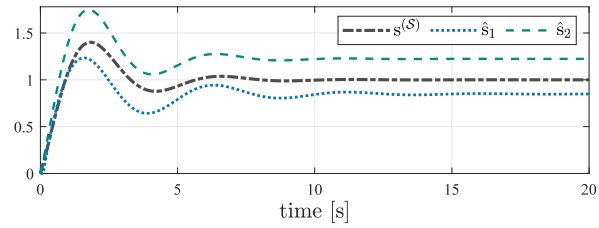


Fig. 1. Step responses for system $\mathcal{S}$ and the estimated models.

15% and a 22% error. Consequently, one can observe that the estimated models do not take into account the SSG side-information.

Motivated by this example, the main problem discussed in this article is the identification with side-information on the SSG of the system, i.e., we address the following identification problem.

*Problem 1:* Given data $\mathscr{D}$, estimate the impulse response of stable system $\mathcal{S}$ satisfying the side-information $\ell_0(\mathrm{g}^{(\mathcal{S})}) \in [\underline{\delta}, \overline{\delta}]$, where $\underline{\delta}$ and $\overline{\delta}$ are given bounds for the SSG of system $\mathcal{S}$.

Compared to the abovementioned example, designed to elaborate on the rationale of our discussion and its importance, Problem 1 addresses the more common scenarios in practice where the available side-information on the SSG is imprecise and provided in the form of interval $[\underline{\delta}, \overline{\delta}]$. When the SSG of the system is known to be exactly equal to $\delta$ that might be any arbitrary value in $\mathbb{R}$, we set $\underline{\delta} = \overline{\delta} = \delta$.

## IV. Identification Problem: Existence and Uniqueness of the Solution

Let $\Bbbk : \mathbb{T}\times\mathbb{T} \to \mathbb{R}$ be an integrable Mercer kernel, which means that $\Bbbk$ is continuous and absolutely integrable when $\mathbb{T} = \mathbb{R}_+$, and, absolutely summable when $\mathbb{T} = \mathbb{Z}_+$. Accordingly, we take stable RKHS $\mathscr{H}_\Bbbk$ as the hypothesis space of the identification problem [4]. Let $\mathcal{I} := \{i_k | k = 1,\dots,n_\mathcal{I}\}$ be a subset of $\{1,\dots,n_\mathscr{D}\}$, $\mathrm{y}_\mathcal{I}$ be the vector defined as $\mathrm{y}_\mathcal{I} = [y_{t_i}]_{i\in\mathcal{I}}$, and $\ell : \mathbb{R}^{n_\mathcal{I}} \times \mathbb{R}^{n_\mathcal{I}} \to \mathbb{R}_+$ be a given convex function. Accordingly, we define the *generalized loss function*, $\mathcal{E}_\ell : \mathscr{H}_\Bbbk \to \mathbb{R}_+$ as follows:

$$\mathcal{E}_\ell(\mathrm{g}) := \ell\left(\left[\mathrm{L}^{\mathrm{u}}_{t_i}(\mathrm{g})\right]_{i\in\mathcal{I}}, \mathrm{y}_\mathcal{I}\right), \qquad \forall \mathrm{g} \in \mathscr{H}_\Bbbk. \tag{3}$$

The introduced loss function $\mathcal{E}_\ell$ has a general form covering a wide range of objectives, e.g., defining $\ell$ based on the Huber function allows addressing the issue of outliers in the measurement data [1]. Also, when $\mathcal{I} = \{1,\dots,n_\mathscr{D}\}$ and function $\ell$ is the square of Euclidean distance, the empirical loss $\mathcal{E}_\ell$ becomes the sum of squared errors. We additionally define a regularization term, enforcing desired attributes such as stability, based on the norm in $\mathscr{H}_\Bbbk$. More precisely, given strictly increasing convex function $\rho : \mathbb{R}_+ \to \mathbb{R}_+$, the regularization function $\mathcal{R} : \mathscr{H}_\Bbbk \to \mathbb{R}_+$ is defined as $\mathcal{R}(\mathrm{g}) = \rho(\|\mathrm{g}\|)$, for any $\mathrm{g} \in \mathscr{H}_\Bbbk$. Accordingly, the objective function for the identification problem, $\mathcal{J} : \mathscr{H}_\Bbbk \to \mathbb{R}_+$, is defined as $\mathcal{J}(\mathrm{g}) := \mathcal{E}_\ell(\mathrm{g}) + \lambda\mathcal{R}(\mathrm{g})$, for any $\mathrm{g} \in \mathscr{H}_\Bbbk$, where $\lambda > 0$ is the regularization weight. Note that the regularization term also addresses the overfitting issue, enhances the numerical performance, and improves the bias-variance tradeoff.

Define the set $\mathscr{G}_\Bbbk([\underline{\delta}, \overline{\delta}]) \subseteq \mathscr{H}_\Bbbk$ as $\mathscr{G}_\Bbbk([\underline{\delta}, \overline{\delta}]) := \{\mathrm{g} \in \mathscr{H}_\Bbbk \,|\, \ell_0(\mathrm{g}) \in [\underline{\delta}, \overline{\delta}]\}$. The elements of $\mathscr{G}_\Bbbk([\underline{\delta}, \overline{\delta}])$ are exactly the ones satisfying the SSG side-information. Therefore, the identification Problem 1 is formulated as the following optimization problem:

$$\begin{aligned} \min_{\mathrm{g}\in\mathscr{H}_\Bbbk} \quad & \mathcal{E}_\ell(\mathrm{g}) + \lambda\mathcal{R}(\mathrm{g}) \\ \text{s.t.} \quad & \mathrm{g} \in \mathscr{G}_\Bbbk\left([\underline{\delta}, \overline{\delta}]\right). \end{aligned} \tag{4}$$

The existence and uniqueness of the solution of optimization problem (4) depend on the topological properties of set $\mathscr{G}_\Bbbk([\underline{\delta}, \overline{\delta}])$, which is characterized primarily by operator $\ell_0 : \mathscr{H}_\Bbbk \to \mathbb{R}$. To address this concern, we need the following theorem, which plays multiple additional central

roles in this work. The proof of this theorem and other main results are moved to the Appendix.

*Theorem 1:* Let $\Bbbk : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ be an integrable Mercer kernel, and $\varphi_0 = (\varphi_{0,t})_{t\in\mathbb{T}}$ be defined as following:

$$\varphi_{0,t} = \begin{cases} \sum_{s\in\mathbb{Z}_+} \Bbbk(t,s), & \text{if } \mathbb{T} = \mathbb{Z}_+ \\ \int_{\mathbb{R}_+} \Bbbk(t,s)\mathrm{d}s, & \text{if } \mathbb{T} = \mathbb{R}_+ \end{cases} \tag{5}$$

for any $t \in \mathbb{T}$. Then, $\varphi_0$ is well-defined, $\varphi_0 \in \mathscr{H}_\Bbbk$, and $\ell_0(\mathrm{g}) = \langle \varphi_0, \mathrm{g}\rangle$, for all $\mathrm{g}$ in $\mathscr{H}_\Bbbk$. Furthermore, we have that

$$\|\varphi_0\|^2 = \begin{cases} \sum_{s,t\in\mathbb{Z}_+} \Bbbk(t,s), & \text{if } \mathbb{T} = \mathbb{Z}_+ \\ \int_{\mathbb{R}_+\times\mathbb{R}_+} \Bbbk(t,s)\mathrm{d}s\mathrm{d}t, & \text{if } \mathbb{T} = \mathbb{R}_+. \end{cases} \tag{6}$$

Theorem 1 says that integrability of Mercer kernel $\Bbbk$ implies that the SSG linear operator $\ell_0 : \mathscr{H}_\Bbbk \to \mathbb{R}$ is continuous.

*Assumption 1:* For the integrable Mercer kernel $\Bbbk$, there exists $\tau \in \mathbb{T}$ such that $\ell_0(\Bbbk_\tau) \neq 0$.

The next theorem describes topological properties of set $\mathscr{G}_\Bbbk([\underline{\delta}, \overline{\delta}])$, and together with Theorem 1, provides the necessary foundation to verify that Problem 1, formulated as in (4), is well-defined.

*Theorem 2:* Let Assumption 1 hold. Then, for any $\underline{\delta}$ and $\overline{\delta}$ such that $-\infty \leq \underline{\delta} \leq \overline{\delta} \leq \infty$, the set $\mathscr{G}_\Bbbk([\underline{\delta}, \overline{\delta}])$ is a non-empty, closed and convex subset of $\mathscr{H}_\Bbbk$.

To show that (4) admits a unique solution in $\mathscr{H}_\Bbbk$ and introduce an equivalent convex finite-dimensional program, we need an additional assumption and lemma establishing certain features for the objective function in (4).

*Assumption 2:* For any $\tau \in \mathscr{T}$, operator $\mathrm{L}^{\mathrm{u}}_\tau : \mathscr{H}_\Bbbk \to \mathbb{R}$ is bounded.

When $\mathbb{T} = \mathbb{R}_+$ and $\mathrm{u}$ is a *piecewise continuous* (PWC) function as in (15), one can show the continuity of $\mathrm{L}^{\mathrm{u}}_\tau$, for any $\tau \in \mathbb{R}_+$, based on an argument similar to the proof of Theorem 1. For the case of $\mathbb{T} = \mathbb{Z}_+$, one can easily see that Assumption 2 holds if the system is initially at rest, or more generally, when $(u_t)_{t\leq t_{n_\mathscr{D}-1}}$ is finitely nonzero. Given this assumption, we have the following lemma due to Riesz representation theorem [29] and the reproducing property of kernel.

*Lemma 3 ([4]):* Let Assumption 2 hold. Then, for each $\tau \in \mathscr{T}$, there exists $\varphi^{(\mathrm{u})}_\tau = (\varphi^{(\mathrm{u})}_{\tau,t})_{t\in\mathbb{T}} \in \mathscr{H}_\Bbbk$ such that $\mathrm{L}^{\mathrm{u}}_\tau(\mathrm{g}) = \langle \varphi^{(\mathrm{u})}_\tau, \mathrm{g}\rangle$, for any $\mathrm{g} \in \mathscr{H}_\Bbbk$. Furthermore, for any $t \in \mathbb{T}$, we have

$$\varphi^{(\mathrm{u})}_{\tau,t} = \begin{cases} \int_{\mathbb{R}_+} \Bbbk(t,s)u_{\tau-s}\mathrm{d}s, & \text{if } \mathbb{T} = \mathbb{R}_+ \\ \sum_{s\in\mathbb{Z}_+} \Bbbk(t,s)u_{\tau-s}, & \text{if } \mathbb{T} = \mathbb{Z}_+. \end{cases} \tag{7}$$

For $k = 1, \ldots, n_\mathcal{I}$, define $\varphi_k$ as $\varphi^{(\mathrm{u})}_{t_i}$ with $i = i_k$. Also, let $\Phi$ be matrix $[\langle\varphi_i, \varphi_j\rangle]^{n_\mathcal{I}, n_\mathcal{I}}_{i,j=0}$, $\mathrm{a}_0, \ldots, \mathrm{a}_{n_\mathcal{I}}$ be columns of $\Phi$, and $\mathrm{A}$ be matrix $[\mathrm{a}_1, \ldots, \mathrm{a}_{n_\mathcal{I}}]^\mathsf{T}$. Given these definitions and the above lemma and theorems, we can present the main result.

*Theorem 4:* Let Assumptions 1 and 2 hold. Then, the optimization problem (4) admits a *unique* solution as $\mathrm{g}^\star = \sum_{i=0}^{n_\mathcal{I}} x^\star_i \varphi_i$, where $\mathrm{x}^\star := [x^\star_i]^{n_\mathcal{I}}_{i=0}$ is the solution of following convex program:

$$\begin{aligned} \min_{\mathrm{x}\in\mathbb{R}^{n_\mathcal{I}+1}} \quad & \ell(\mathrm{A}\mathrm{x}, \mathrm{y}_\mathcal{I}) + \lambda\mathcal{R}\left((\mathrm{x}^\mathsf{T}\Phi\mathrm{x})^{\frac{1}{2}}\right) \\ \text{s.t.} \quad & \mathrm{a}_0^\mathsf{T}\mathrm{x} \in [\underline{\delta}, \overline{\delta}]\,. \end{aligned} \tag{8}$$

In the literature, it is common to employ sum of squared errors empirical loss and quadratic regularization function. The following corollary provides a closed-form solution for this case when we have exact side-information on the SSG of the system.

*Corollary 5:* Let the assumptions of Theorem 4 hold, $\mathcal{E}_\ell$ be the sum of squared errors, the regularization function be $\mathcal{R}(\mathrm{g}) = \|\mathrm{g}\|^2$, and the SSG of the system is known to be exactly $\delta \in \mathbb{R}$. Then, (4) has a *unique* solution $\mathrm{g}^\star$ as in Theorem 4. Moreover, there exists $\gamma^\star \in \mathbb{R}$ such that $[\mathrm{x}^{\star\mathsf{T}}, \gamma^\star]^\mathsf{T}$ is a solution of the following system of linear equations:

$$\begin{bmatrix} \mathrm{Q} & \mathrm{a}_0 \\ \mathrm{a}_0^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} \mathrm{x} \\ \gamma \end{bmatrix} = \begin{bmatrix} \mathrm{A}^\mathsf{T}\mathrm{y} \\ \delta \end{bmatrix} \tag{9}$$

where $\mathrm{Q} = \mathrm{A}^\mathsf{T}\mathrm{A} + \lambda\Phi$ and $\mathrm{y} = [y_{t_i}]^{n_\mathscr{D}}_{i=1} \in \mathbb{R}^{n_\mathscr{D}}$. Furthermore, when $\varphi_0, \ldots, \varphi_{n_\mathscr{D}}$ are linearly independent, we have

$$\mathrm{x}^\star = \mathrm{Q}^{-1}\mathrm{A}^\mathsf{T}\mathrm{y} + \frac{\delta - \mathrm{a}_0^\mathsf{T}\mathrm{Q}^{-1}\mathrm{A}^\mathsf{T}\mathrm{y}}{\mathrm{a}_0^\mathsf{T}\mathrm{Q}^{-1}\mathrm{a}_0}\mathrm{Q}\mathrm{a}_0. \tag{10}$$

## V. Optimization Problem: Settings and Algorithm

Based on Theorem 4, addressing identification problem 1, or equivalently, optimization problem (4), reduces to the convex program (8). To this end, we need $\mathrm{a}_0$, $\mathrm{A}$, and $\Phi$. Since $\mathrm{A}$ is a submatrix of $\Phi$ and $\mathrm{a}_0$ is the first column of $\Phi$, it suffices to obtain $\Phi$. The entries of $\Phi$ are inner products $\langle\varphi_i, \varphi_j\rangle$, for $i, j \in \{0, \ldots, n_\mathscr{D}\}$, which are improper double integrals when $\mathbb{T} = \mathbb{R}_+$, or infinite double summations when $\mathbb{T} = \mathbb{Z}_+$. While these calculations can be performed using numerical techniques, their values may be obtained analytically in certain but fairly general situations. The prominent case is when we employ standard kernels in the literature, such as *diagonally/correlated* (DC), *tuned/correlated* (TC), and, *stable spline* (SS), which are, respectively, defined, for any $s, t \in \mathbb{T}$, as

$$\Bbbk_{\mathrm{DC}}(s,t) = \alpha^{\max(s,t)}\gamma^{|s-t|} \tag{11}$$

$$\Bbbk_{\mathrm{TC}}(s,t) = \alpha^{\max(s,t)} \tag{12}$$

$$\Bbbk_{\mathrm{SS}}(s,t) = \alpha^{\max(s,t)+s+t} - \tfrac{1}{3}\alpha^{3\max(s,t)} \tag{13}$$

for any $s, t \in \mathbb{T}$, where $\alpha, \gamma \in \mathbb{R}$ are such that $\alpha \in (0,1)$, $|\gamma| \in \left(0, \sqrt{\alpha^{-1}}\right)$, if $\mathbb{T} = \mathbb{Z}_+$, and, $\gamma \in \left(0, \sqrt{\alpha^{-1}}\right)$, if $\mathbb{T} = \mathbb{R}_+$ [4]. Note that by setting $\rho = \sqrt{\alpha}\gamma$, the definition of DC kernel in [4] reduces to (11). Also, without loss of generality, we can drop the extra scaling factor introduced in [4].

### A. Optimization Problem Configuration: Discrete-Time Case

Let $\mathbb{T} = \mathbb{Z}_+$ and the system be initially at rest, i.e., $u_t = 0$ for $t < 0$. Also, let $\mathscr{T} = \{0, 1, \ldots, n_\mathscr{D} - 1\}$. Given an integrable kernel $\Bbbk : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{R}$ and an input $\mathrm{u}$, define matrices $\mathrm{K}$ and $\mathrm{T}_\mathrm{u}$ such that $[\mathrm{K}]_{(i,j)} = \Bbbk(i-1, j-1)$ and $[\mathrm{T}_\mathrm{u}]_{(i,j)} = u_{i-j}$, for $i, j = 1, 2, \ldots, n_\mathscr{D}$. From these definitions, we have the next theorem.

*Theorem 6:* Let $\varphi \in \mathbb{R}^{n_\mathscr{D}}$ be the column vector defined as $\varphi := [\varphi_{0,i}]^{n_\mathscr{D}-1}_{i=0}$. Then, we have

$$\Phi = \begin{bmatrix} \|\varphi_0\|^2 & \varphi\mathrm{T}_\mathrm{u}^\mathsf{T} \\ \mathrm{T}_\mathrm{u}\varphi & \mathrm{T}_\mathrm{u}\mathrm{K}\mathrm{T}_\mathrm{u}^\mathsf{T} \end{bmatrix}. \tag{14}$$

*Remark 1:* Appendix H in [30] provides $\varphi_0$ and $\|\varphi_0\|^2$ for the standard kernels introduced in (11), (12), and (13), when $\mathbb{T} = \mathbb{Z}_+$.

### B. Optimization Problem Configuration: Continuous-Time Case

The set of piecewise constant (PWC) functions is dense in $L^p(\mathbb{R})$, for $p \in [1, \infty)$, and also, any function in $L^\infty(\mathbb{R})$ is an almost everywhere the point-wise limit of a sequence of PWC functions [29]. In other words, any signal of interest can be approximated arbitrarily closely by PWC functions. Accordingly, for the case of $\mathbb{T} = \mathbb{R}_+$, one can assume that the input signal $\mathrm{u} = (u_t)_{t\in\mathbb{R}_+}$ is PWC. More precisely, there exist $n_\mathrm{s} \in \mathbb{N}$, real scalars $\xi_1, \ldots, \xi_{n_\mathrm{s}}$ and a finite increasing sequence $(s_0, s_1, \ldots, s_{n_\mathrm{s}})$ in $\mathbb{R}_+$ such that

$$u_t = \sum_{i=0}^{n_\mathrm{s}-1} \xi_{i+1}\mathbf{1}_{[s_i, s_{i+1})}(t) \quad \forall t \in \mathbb{R}_+. \tag{15}$$

In this case, a closed-form for $\varphi^{(\mathrm{u})}_\tau$ can be introduced. To this end, we require the function $\psi : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ defined as

$$\psi(t, a, b) := \int_a^b \Bbbk(t, s)\mathrm{d}s \tag{16}$$

for any $a, b, t \in \mathbb{R}_+$. This function is denoted by $\psi_{\rm TC}$, $\psi_{\rm DC}$, or $\psi_{\rm SS}$, respectively, when $\Bbbk$ is $\Bbbk_{\rm TC}$, $\Bbbk_{\rm DC}$, or $\Bbbk_{\rm SS}$.

*Theorem 7:* For any $t \in \mathbb{R}_+$, we have

$$\varphi^{(\rm u)}_{\tau,t} = \sum_{i=0}^{n_{\rm s}-1} \xi_{i+1} \psi(t, \bar{s}_{i+1}(\tau), \bar{s}_i(\tau)) \tag{17}$$

where, for $i = 0, \ldots, n_{\rm s}$, function $\bar{s}_i : \mathbb{R}_+ \to \mathbb{R}_+$ is defined as $\bar{s}_i(\tau) := \max(\tau - s_i, 0)$, for any $\tau \in \mathbb{R}_+$.

For the standard kernels defined in (11), (12), and (13), one can obtain the closed-form of $\varphi^{(\rm u)}_\tau$ using (17). To this end, we need $\psi_{\rm TC}$, $\psi_{\rm DC}$, and $\psi_{\rm SS}$, which are provided by the following theorem.

*Theorem 8:* Define function $\eta : \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ as $\eta(s, \tau_1, \tau_2) = \min(\max(t, \tau_1), \tau_2)$, for any $s, \tau_1, \tau_2 \in \mathbb{R}_+$, and let $t, a, b \in \mathbb{R}_+$ such that $a \le b$. Then, we have

$$\psi_{\rm TC}(t,a,b) = (\eta(t,a,b) - a)\, \alpha^t + \frac{\alpha^b - \alpha^{\eta(t,a,b)}}{\ln(\alpha)} \tag{18}$$

$$\psi_{\rm DC}(t,a,b) = \frac{\gamma^{-a} - \gamma^{-\eta(t,a,b)}}{\ln(\gamma)} (\alpha\gamma)^t + \frac{(\alpha\gamma)^b - (\alpha\gamma)^{\eta(t,a,b)}}{\ln(\alpha\gamma)\gamma^t} \tag{19}$$

$$\begin{aligned}\psi_{\rm SS}(t,a,b) = {} & \frac{\alpha^{\eta(t,a,b)} - \alpha^a}{\ln(\alpha)} \alpha^{2t} + \frac{\alpha^{2b} - \alpha^{2\eta(t,a,b)}}{2\ln(\alpha)} \alpha^t \\ & - \frac{1}{3}(\eta(t,a,b) - a)\, \alpha^{3t} - \frac{\alpha^{3b} - \alpha^{3\eta(t,a,b)}}{9\ln(\alpha)}.\end{aligned} \tag{20}$$

Similar to the previous theorem, one can obtain the closed-form of $\varphi_0$ for the standard kernels (11), (12), and (13).

*Theorem 9:* For any $t \in \mathbb{R}_+$, we have

$$\varphi_{{\rm TC},0,t} = \left[ t - \frac{1}{\ln(\alpha)} \right] \alpha^t \tag{21}$$

$$\varphi_{{\rm DC},0,t} = -\left[ \frac{(1-\gamma^t)}{\ln(\gamma)} + \frac{1}{\ln(\alpha\gamma)} \right] \alpha^t \tag{22}$$

$$\varphi_{{\rm SS},0,t} = \left[ \frac{11}{18\ln(\alpha)} \alpha^t - \frac{1}{\ln(\alpha)} - \frac{t\alpha^t}{3} \right] \alpha^{2t}. \tag{23}$$

Given $\{\varphi_i\}_{i=0}^{n_{\mathscr{D}}}$, we can obtain $\langle \varphi_i, \varphi_j \rangle$, for $i, j = 0, \ldots, n_{\mathscr{D}}$. To this end, define functions $\nu : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ and $\bar{\nu} : \mathbb{R}_+ \to \mathbb{R}$, respectively, as

$$\nu(x,y) := \int_0^x \int_0^y \Bbbk(s,t) {\rm d}t {\rm d}s \qquad \forall x, y \in \mathbb{R}_+ \tag{24}$$

and

$$\bar{\nu}(x) := \int_0^x \int_0^\infty \Bbbk(s,t) {\rm d}t {\rm d}s \qquad \forall x \in \mathbb{R}_+. \tag{25}$$

When $\Bbbk$ is one of the standard kernels, we include a suitable subscript in $\nu$ and $\bar{\nu}$ to indicate the corresponding kernel. For each $i, j \in \{0, \ldots, n_{\rm s} - 1\}$, define functions $\kappa_{ij} : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ and $\bar{\kappa}_i : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ such that, for any $\tau, \tau_1, \tau_2 \in \mathbb{R}_+$, we have

$$\begin{aligned}\kappa_{ij}(\tau_1,\tau_2) = {} & \nu(\bar{s}_i(\tau_1), \bar{s}_j(\tau_2)) - \nu(\bar{s}_{i+1}(\tau_1), \bar{s}_j(\tau_2)) \\ & - \nu(\bar{s}_i(\tau_1), \bar{s}_{j+1}(\tau_2)) + \nu(\bar{s}_{i+1}(\tau_1), \bar{s}_{j+1}(\tau_2))\end{aligned} \tag{26}$$

and

$$\bar{\kappa}_i(\tau) = \bar{\nu}(\bar{s}_i(\tau)) - \bar{\nu}(\bar{s}_{i+1}(\tau)). \tag{27}$$

Based on these definitions, the following theorem presents the closed-form for $\left\langle \varphi_0, \varphi^{(\rm u)}_\tau \right\rangle$ and $\langle \varphi^{(\rm u)}_{\tau_1}, \varphi^{(\rm u)}_{\tau_2} \rangle$.

*Theorem 10:* For any $\tau, \tau_1, \tau_2 \in \mathbb{R}_+$, we have

$$\left\langle \varphi_0, \varphi^{(\rm u)}_\tau \right\rangle = \sum_{i=0}^{n_{\rm s}-1} \xi_{i+1} \bar{\kappa}_i(\tau) \tag{28}$$

$$\left\langle \varphi^{(\rm u)}_{\tau_1}, \varphi^{(\rm u)}_{\tau_2} \right\rangle = \sum_{i=0}^{n_{\rm s}-1} \sum_{j=0}^{n_{\rm s}-1} \xi_{i+1}\xi_{j+1}\kappa_{ij}(\tau_1,\tau_2). \tag{29}$$

**Algorithm 1:** System Identification with Steady-State Gain Side-Information.

1: **Input:** Set of data $\mathscr{D}$, integrable kernel $\Bbbk$, index set $\mathcal{I}$, convex function $\ell$, regularization weight $\lambda$, and, real scalar $\delta$ or interval $[\underline{\delta}, \bar{\delta}]$ for the steady-state gain.
2: Calculate matrix $\Phi$, and then obtain matrix A and vector $\mathrm{a}_0$.
▷ For discrete-time case, use Theorem 6 and Remark 1.
▷ For continuous-time case and PWC input, use (26), (27), Theorems 10, 11 and 12.
3: Obtain $\mathrm{x}^\star$ by solving optimization (8) or system of (9).
▷ If the steady-state gain is known to be $\delta$ and the empirical loss is the sum of squared errors, obtain $\mathrm{x}^\star$ by (10) or (9).
4: Calculate $\varphi_0$ by (5), or by Theorems 9 and [30, Theorem 15].
5: Calculate $\varphi_1, \ldots, \varphi_{n_{\mathcal{I}}}$ based on (7).
▷ For continuous-time case and PWC input, use Theorems 7 and 8.
6: Given $\mathrm{x}^\star$ and $\{\varphi_i\}_{i=0}^{n_{\mathcal{I}}}$, obtain $\mathrm{g}^\star = \sum_{i=0}^{n_{\mathcal{I}}} x_i^\star \varphi_i$.
7: **Output:** $\mathrm{g}^\star$ and $\mathrm{x}^\star$.

Due to (26) and (27), in order to employ Theorem 10 to calculate the entries of $\Phi$, we need to obtain functions $\nu$ and $\bar{\nu}$, which can be done in general using numerical techniques. Nevertheless, the closed-form of $\nu$ and $\bar{\nu}$ can be explicitly derived for the standard kernels.

*Theorem 11:* For any $x, y \in \mathbb{R}_+$, we have

$$\nu_{\rm TC}(x,y) = \frac{\min(x,y)\,(\alpha^x + \alpha^y)}{\ln(\alpha)} + \frac{2\left(1 - \alpha^{\min(x,y)}\right)}{\ln(\alpha)^2} \tag{30}$$

$$\bar{\nu}_{\rm TC}(x) = \frac{x\alpha^x \ln(\alpha) + 2(1-\alpha^x)}{\ln(\alpha)^2} \tag{31}$$

$$\begin{aligned}\nu_{\rm DC}(x,y) = {} & \frac{\left(1 - \gamma^{-\min(x,y)}\right)\left((\alpha\gamma)^x + (\alpha\gamma)^y\right)}{\ln(\gamma)\ln(\alpha\gamma)} \\ & + \frac{2 - 2\alpha^{\min(x,y)}}{\ln(\alpha)\ln(\alpha\gamma)}\end{aligned} \tag{32}$$

$$\bar{\nu}_{\rm DC}(x) = \frac{\alpha^x\gamma^x - \alpha^x}{\ln(\gamma)\ln(\alpha\gamma)} + \frac{2 - 2\alpha^x}{\ln(\alpha)\ln(\alpha\gamma)} \tag{33}$$

$$\begin{aligned}\nu_{\rm SS}(x,y) = {} & \frac{\left(\alpha^{\min(x,y)} - 1\right)\left(\alpha^{2x} + \alpha^{2y}\right)}{2\ln(\alpha)^2} \\ & - \frac{\min(x,y)\,(\alpha^{3x} + \alpha^{3y})}{9\ln(\alpha)} + \frac{7 - 7\alpha^{3\min(x,y)}}{27\ln(\alpha)^2}\end{aligned} \tag{34}$$

$$\bar{\nu}_{\rm SS}(x) = \frac{14 - 27\alpha^{2x} + 13\alpha^{3x}}{54\ln(\alpha)^2} - \frac{x\alpha^{3x}}{9\ln(\alpha)}. \tag{35}$$

Similar to the previous theorem, we can calculate the closed-form of $\|\varphi_0\|^2 = \langle \varphi_0, \varphi_0 \rangle$ for the standard stable kernels. The following theorem presents these closed-forms.

*Theorem 12:* We have $\|\varphi_{{\rm TC},0}\|^2 = 2(\ln(\alpha))^{-2}$, $\|\varphi_{{\rm DC},0}\|^2 = 2(\ln(\alpha)\ln(\alpha\gamma))^{-2}$, and $\|\varphi_{{\rm SS},0}\|^2 = 7(27\ln(\alpha))^{-2}$.

Based on the discussion above, we can derive the key elements of optimization (8) and solve Problem 1 to identify the impulse response of system. This procedure is summarized in Algorithm 1.

### C. Hyperparameter Tuning

To employ Algorithm 1, in addition to the set of data $\mathscr{D}$, an appropriate stable kernel $\Bbbk$ and the regularization weight $\lambda$ are required. The general form of the kernel depends on the shape and smoothness of the impulse response to be identified. Following specifying the type of kernel, in addition to $\lambda$, it is required to determine the hyperparameters $\theta_\Bbbk$ characterizing kernel $\Bbbk$. Therefore, we need to estimate the vector of hyperparameters $\theta := [\lambda, \theta_\Bbbk]$ in the admissible set $\Theta \subseteq \mathbb{R}^{n_\theta}$. For

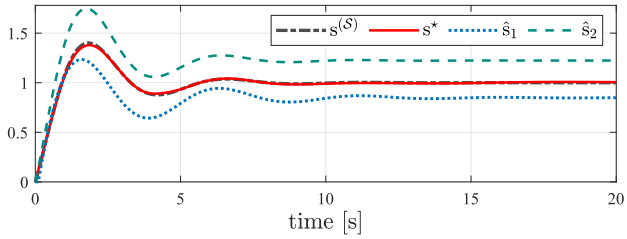


Fig. 2. Step responses for system $\mathcal{S}$ and the models estimated in Example 1. The step response $s^\star$ corresponds to the proposed method.

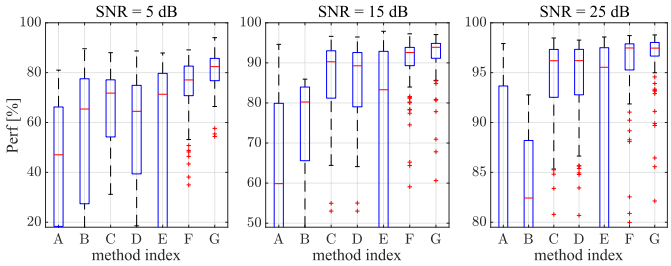


Fig. 3. Box plots compare methods discussed in Example 2 via fitting metric (37) for different SNR levels. One can see the proposed method (G) shows superior performance. Note that the $y$-axis range has been adjusted to improve visibility in each of these box plots.

this purpose, we use a cross-validation mechanism equipped with a Bayesian optimization heuristic [31]. More precisely, the index set of data is partitioned into disjoint sets $\mathcal{I}_{\rm T}$ and $\mathcal{I}_{\rm V}$ to be utilized respectively for training and validation. According to the prediction error on the validation data, *the model evaluation metric* $v : \Theta \to \mathbb{R}$, is defined as

$$v(\theta) = \frac{1}{|\mathcal{I}_{\rm V}|} \sum_{i \in \mathcal{I}_{\rm V}} \left(y_{t_i} - {\rm L}^{\rm u}_{t_i}(\mathrm{g}(\theta))\right)^2 \tag{36}$$

where $\mathrm{g}(\theta)$ is the impulse response identified using the proposed identification technique given the training data and the hyperparameters $\theta$. Then, $\theta$ is estimated as $\hat{\theta} := \operatorname{argmin}_{\theta \in \Theta} v(\theta)$. The dependency of model evaluation metric $v$ to the vector of hyperparameters is in a black-box oracle form. To solve this optimization problem, we use a Bayesian optimization algorithm, such as GP-LCB, which is available through MATLAB's `bayesopt` function [31].

## VI. Numerical Examples

In this section, we demonstrate and compare the performance of the proposed method through several numerical examples.

*Example 1:* For the example provided in Section III, we employ the proposed identification scheme presuming that the SSG of the system is given, i.e., we know that $\ell_0(\mathrm{g}^{(\mathcal{S})}) = 1$. Accordingly, we apply Algorithm 1 given the set of measurement data $\mathscr{D}$ and $\delta = 1$. We utilize $\Bbbk_{\rm TC}$ introduced in (12) and tune its hyperparameters based on the cross-validation mechanism introduced in Section V-C. To this end, the first 80% of data points are chosen for training and the remaining 20% for validation. Given this partitioning of the measurement data, the hyperparameters $\theta = [\lambda, \alpha]$ are estimated as $\operatorname{argmin}_{\theta \in \Theta} v(\theta)$, where $v$ is the model evaluation metric defined in (36). We solve this optimization problem using GP-LCB Bayesian optimization approach [31]. Fig. 2 shows the step response corresponding to $\mathrm{g}^\star$ together with the step responses of the models $\hat{\mathrm{g}}_1$ and $\hat{\mathrm{g}}_2$, estimated in Section III. For the estimated impulse response $\mathrm{g}^\star$, we have $\ell_0(\mathrm{g}^\star) = 1.00$. To evaluate and compare the estimated impulse responses, we employ the following performance metric:

$$\mathrm{Perf}(\mathrm{g}) = 100 \times \left(1 - \frac{\|\mathrm{g} - \mathrm{g}^{(\mathcal{S})}\|_2}{\|\mathrm{g}^{(\mathcal{S})}\|_2}\right), \quad \forall \mathrm{g} \in \mathscr{H}_{\Bbbk}. \tag{37}$$

For the estimated impulse responses, we have $\mathrm{Perf}(\hat{\mathrm{g}}_1) = 70.88\%$, $\mathrm{Perf}(\hat{\mathrm{g}}_2) = 58.60\%$, and $\mathrm{Perf}(\mathrm{g}^\star) = 95.84\%$. Accordingly, one can see that the proposed scheme outperforms the existing methods in terms of performance metric $\mathrm{Perf}(\cdot)$ and the precision of resulting SSG. △

*Example 2:* In this example, we compare the performance of the proposed identification method with the existing schemes through a Monte Carlo analysis. To this end, with respect to each $(n, r)$ in $\{16, 17, \ldots, 25\} \times \{0.8, 0.82, \ldots, 0.96\}$, we employ MATLAB's `drss` function to randomly generate a discrete-time LTI system with order $n$ and spectral radius $r$. We normalize these systems with their $\mathcal{H}_2$-norm and set them initially at rest. For each of these systems, a random zero-mean white Gaussian input signal with length $n_{\mathscr{D}} = 200$ is generated using MATLAB's `idinput` function. By applying these input signals to the systems, we obtain their noiseless output signals. We consider three levels of SNR: high, medium, and low, which are 5, 15, and 25 dB, respectively. For the additive measurement uncertainty, we generate a zero-mean white Gaussian signal for each of these SNR levels and each output signal. The noiseless output signal is then corrupted with the corresponding additive noise signals, and the resulting noisy output is measured at time instants $t_i = i$, for $i = 0, 1, \ldots, 199$. As a result, we have 100 sets of input–output data for each of the aforementioned SNR values. We employ the these input–output datasets and the following identification methods to estimate the impulse response of corresponding systems.

A. This method is a modified subspace approach incorporating steady-state features of output [22].
B. This method estimates impulse response by solving a constrained optimization problem formulated based on a behavioral approach [24].
C. This method considers the interpretation of subspace identification as the optimal multistep ahead prediction and modifies it to a constrained least-squares problem where the imposed equality constraint models the SSG information [21].
D. This method is a general Bayesian variant of the optimal multistep ahead predictor interpretation of subspace identification approach, where SSG information is integrated into the covariance of the prior distribution [3].
E. In this method, the step response of system is first estimated by a kernel-based Bayesian approach, and then, the FIR is calculated using discrete derivative [25].
F. This method estimates an FIR model based on a kernel-based Bayesian approach where the SSG information is enforced on the total summation of the estimated FIR [26].
G. The last method is the scheme proposed in this article and summarized in Algorithm 1.

The kernel-based methods E, F, and G employ the same kernel type (11) to give a fair comparison. To evaluate and compare the estimation performances of these methods, we employ the metric introduced in (37). The results are shown and compared in Fig. 3 .

*Discussion:* From Fig. 3, we observe that the proposed identification scheme demonstrates better estimation performance than other methods. Indeed, methods A, B, C, and D are based on the subspace approach and prediction error minimization. Meanwhile, the proposed scheme is a kernel-based method in which the stability of the system is included and the model complexity tuning is implemented based on the effective approach of estimating continuous regularization hyperparameters rather than selecting an integer order [4], [32]. The methods E and F estimate an FIR model for the system, which can be inexact when the spectral radius of the system is close to one, and the impulse response of system can not be approximated well by a short FIR. Moreover, the discrete derivative employed in method F makes the estimation prone to numerical sensitivity, particularly when the data is noisy. On the other hand, the proposed scheme estimates directly the impulse response without any truncation and inexact numerical procedures such as discrete derivatives. △

## VII. Conclusion

We have addressed the impulse response identification problem when side-information on the SSG of the system is provided. The problem is formulated as a generic nonparametric identification in the form of an infinite-dimensional constrained convex program over the RKHS of integrable impulse responses, where the imposed linear constraints enforce the integration of the given side-information into the solution. We have shown that the SSG is a bounded operator over the employed RKHS, which guarantees the existence and uniqueness of the solution. Using the representer theorem, the optimization problem is reduced to a tractable finite-dimensional convex program. Compared to the existing methods, the proposed identification approach can utilize non-uniform measurement samples and integrate SSG side-information into the continuous-time models directly. Through an extensive Monte Carlo numerical experiment, we have verified that the introduced methodology outperforms the benchmark approaches. The observed performance superiority is due to the direct estimation of the impulse response without truncation and without inexact numerical procedures such as discrete derivatives.

## Appendix

### 1) Proof of Theorem 1

Let $\mathbb{T}=\mathbb{R}_+$, $r\in\mathbb{R}_+$, and $I_r:=\int_0^r\int_0^r \Bbbk(s,t)\mathrm{d}s\mathrm{d}t$, which is well-defined since $\Bbbk$ is an integrable kernel. Define $\mathrm{f}_n^{(r)}:=(f_{n,s}^{(r)})_{s\in\mathbb{R}_+}\in\mathscr{H}_\Bbbk$ as $f_{n,s}^{(r)}=\frac{r}{n}\sum_{i=0}^{n-1}\Bbbk(\frac{i}{n}r,s)$, for any $s\in\mathbb{R}_+$ and each $n\in\mathbb{N}$. From the reproducing property, for any $n,m\in\mathbb{Z}_+$, we know that

$$\begin{aligned}\left\langle \mathrm{f}_n^{(r)},\mathrm{f}_m^{(r)}\right\rangle - I_r &= \left\langle \frac{r}{n}\sum_{i=0}^{n-1}\Bbbk\left(\frac{i}{n}r,\cdot\right),\frac{r}{m}\sum_{j=0}^{m-1}\Bbbk\left(\frac{j}{m}r,\cdot\right)\right\rangle - I_r\\ &=\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\left[\Bbbk\left(\frac{i}{n}r,\frac{j}{m}r\right)\frac{r^2}{nm}-\int_{\frac{i}{n}r}^{\frac{i+1}{n}r}\int_{\frac{j}{m}r}^{\frac{j+1}{m}r}\Bbbk(s,t)\mathrm{d}s\mathrm{d}t\right]\\ &=\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\int_{\frac{i}{n}r}^{\frac{i+1}{n}r}\int_{\frac{j}{m}r}^{\frac{j+1}{m}r}\left[\Bbbk\left(\frac{i}{n}r,\frac{j}{m}r\right)-\Bbbk(s,t)\right]\mathrm{d}s\mathrm{d}t.\end{aligned}$$

Since $[0,r]^2$ is a compact region, continuity of $\Bbbk$ implies that $\Bbbk$ is uniformly continuous on $[0,r]^2$. Therefore, for any $\varepsilon>0$, there exists $\delta_\varepsilon\in\mathbb{R}_+$ such that we have $|\Bbbk(s_1,t_1)-\Bbbk(s_2,t_2)|\le\frac{1}{4}\varepsilon^2r^{-2}$, for any $s_1,s_2,t_1,t_2\in[0,r]$ where $|s_1-s_2|+|t_1-t_2|\le\delta_\varepsilon$. Accordingly, if $n,m\ge n_\varepsilon$, where $n_\varepsilon$ is the smallest positive integer larger than $\frac{1}{\delta_\varepsilon}2r$, one has

$$\left|\left\langle \mathrm{f}_n^{(r)},\mathrm{f}_m^{(r)}\right\rangle - I_r\right|\le\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}\int_{\frac{i}{n}r}^{\frac{i+1}{n}r}\int_{\frac{j}{m}r}^{\frac{j+1}{m}r}\frac{\varepsilon^2}{4r^2}\mathrm{d}s\mathrm{d}t=\frac{1}{4}\varepsilon^2. \quad (38)$$

Therefore, one can easily see that $\|\mathrm{f}_n^{(r)}-\mathrm{f}_m^{(r)}\|\le\varepsilon$, for any $n,m\ge n_\varepsilon$. Accordingly, $\{\mathrm{f}_n^{(r)}\}_{n=1}^\infty$ is a Cauchy sequence in the Hilbert space $\mathscr{H}_\Bbbk$, which implies that there exists $\mathrm{f}^{(r)}=(f_s^{(r)})_{s\in\mathbb{R}_+}$ in $\mathscr{H}_\Bbbk$ such that $\{\mathrm{f}_n^{(r)}\}_{n=1}^\infty$ converges to $\mathrm{f}^{(r)}$. Due to the reproducing property, we know that $f_s^{(r)}-f_{n,s}^{(r)}=\langle\mathrm{f}^{(r)}-\mathrm{f}_n^{(r)},\Bbbk_s\rangle$, for any $s\in\mathbb{R}_+$. Thus, from the Cauchy–Schwartz inequality, we know

$$\lim_{n\to\infty}|f_s^{(r)}-f_{n,s}^{(r)}|\le\lim_{n\to\infty}\|\mathrm{f}^{(r)}-\mathrm{f}_n^{(r)}\|\,\|\Bbbk_s\|=0 \quad (39)$$

i.e., we have $\lim_{n\to\infty}f_{n,s}^{(r)}=f_s^{(r)}$. Furthermore, we know that

$$\left|f_{n,s}^{(r)}-\int_0^r\Bbbk(s,t)\mathrm{d}t\right|=\left|\sum_{i=0}^{n-1}\int_{\frac{i}{n}r}^{\frac{i+1}{n}r}\left[\Bbbk\left(\frac{i}{n}r,s\right)-\Bbbk(t,s)\right]\mathrm{d}t\right|$$

$$\le\sum_{i=0}^{n-1}\int_{\frac{i}{n}r}^{\frac{i+1}{n}r}\left|\Bbbk\left(\frac{i}{n}r,s\right)-\Bbbk(t,s)\right|\mathrm{d}t\le\sum_{i=0}^{n-1}\frac{1}{n}\frac{\varepsilon^2}{4r^2}=\frac{\varepsilon^2}{4r^2}.$$

Subsequently, we have $f_s^{(r)}=\lim_{n\to\infty}f_{n,s}^{(r)}=\int_0^r\Bbbk(s,t)\mathrm{d}t$, i.e., $\mathrm{f}^{(r)}=\int_0^r\Bbbk(\cdot,t)\mathrm{d}t$. Accordingly, from $\lim_{n\to\infty}\mathrm{f}_n^{(r)}=\mathrm{f}^{(r)}$, (38) and the definition of $I_r$, it follows that

$$\left\|\int_0^r\Bbbk(\cdot,t)\mathrm{d}t\right\|^2=\|\mathrm{f}^{(r)}\|^2=\lim_{n\to\infty}\langle\mathrm{f}_n^{(r)},\mathrm{f}_n^{(r)}\rangle=\int_0^r\int_0^r\Bbbk(s,t)\mathrm{d}s\mathrm{d}t. \quad (40)$$

Take an arbitrary $\mathrm{g}=(g_t)_{t\in\mathbb{R}_+}$ in $\mathscr{H}_\Bbbk$ and $t,\varepsilon\in\mathbb{R}_+$ such that $t+\varepsilon\in\mathbb{R}_+$. Since $\Bbbk$ is symmetric and continuous, and due to the Cauchy–Schwartz inequality and the reproducing property, we have

$$\begin{aligned}\lim_{\varepsilon\to0}|g_{t+\varepsilon}-g_t|&=\lim_{\varepsilon\to0}|\langle\Bbbk_{t+\varepsilon}-\Bbbk_t,\mathrm{g}\rangle|\le\lim_{\varepsilon\to0}\|\Bbbk_{t+\varepsilon}-\Bbbk_t\|\|\mathrm{g}\|\\ &=\lim_{\varepsilon\to0}\left[\Bbbk(t+\varepsilon,t+\varepsilon)-2\Bbbk(t,t+\varepsilon)+\Bbbk(t,t)\right]^{\frac{1}{2}}\|\mathrm{g}\|=0\end{aligned}$$

which implies the continuity of $\mathrm{g}=(g_t)_{t\in\mathbb{R}_+}$ as a function of $t$. Therefore, the Riemann integral of g on $[0,r]$ is well-defined. Accordingly, from the definition of $\mathrm{f}_n^{(r)}$, the reproducing property and $\lim_{n\to\infty}\mathrm{f}_n^{(r)}=\mathrm{f}^{(r)}$, it follows that

$$\begin{aligned}\int_0^r g_t\mathrm{d}t&=\lim_{n\to\infty}\frac{r}{n}\sum_{i=0}^{n-1}g\left(\frac{i}{n}r\right)=\lim_{n\to\infty}\langle\mathrm{f}_n^{(r)},\mathrm{g}\rangle\\ &=\langle\mathrm{f}^{(r)},\mathrm{g}\rangle=\left\langle\int_0^r\Bbbk(\cdot,t)\mathrm{d}t,\mathrm{g}\right\rangle.\end{aligned}$$

For $n\in\mathbb{N}$, set $r_n:=\min\{r\ge n|\int_r^\infty\int_r^\infty|\Bbbk(s,t)|\mathrm{d}s\mathrm{d}t\le\frac{1}{n}\}$, which is well-defined since $\Bbbk$ is continuous and integrable. Note that $\{r_n\}_{n=1}^\infty$ is an unbounded strictly increasing sequence. For any $n,m\in\mathbb{N}$, such that $n\le m$, we have

$$f_t^{(r_n)}-f_t^{(r_m)}=\int_0^{r_n}\Bbbk(s,t)\mathrm{d}s-\int_0^{r_m}\Bbbk(s,t)\mathrm{d}s=\int_{r_m}^{r_n}\Bbbk(s,t)\mathrm{d}s.$$

Accordingly, one can see that

$$\begin{aligned}\|\mathrm{f}^{(r_n)}-\mathrm{f}^{(r_m)}\|^2&=\langle\mathrm{f}^{(r_n)},\mathrm{f}^{(r_n)}-\mathrm{f}^{(r_m)}\rangle-\langle\mathrm{f}^{(r_m)},\mathrm{f}^{(r_n)}-\mathrm{f}^{(r_m)}\rangle\\ &=\int_0^{r_n}f_t^{(r_n)}-f_t^{(r_m)}\mathrm{d}t-\int_0^{r_m}f_t^{(r_n)}-f_t^{(r_m)}\mathrm{d}t\\ &=\int_{r_m}^{r_n}\int_{r_m}^{r_n}\Bbbk(s,t)\mathrm{d}s\mathrm{d}t\le\int_{r_m}^{r_n}\int_{r_m}^{r_n}|\Bbbk(s,t)|\mathrm{d}s\mathrm{d}t\le\frac{1}{m}.\end{aligned}$$

Thus, $\{\mathrm{f}^{(r_n)}\}_{n=1}^\infty$ is a Cauchy sequence, and, we know that there exists $\mathrm{f}=(f_s)_{s\in\mathbb{R}_+}\in\mathscr{H}_\Bbbk$ such that $\lim_{n\to\infty}\mathrm{f}^{(r_n)}=\mathrm{f}$. Subsequently, due to the reproducing property, for any $s\in\mathbb{R}_+$, we have

$$f_s=\langle\mathrm{f},\Bbbk_s\rangle=\lim_{n\to\infty}\langle\mathrm{f}^{(r_n)},\Bbbk_s\rangle=\lim_{n\to\infty}\int_0^{r_n}\Bbbk(s,t)\mathrm{d}t=\int_0^\infty\Bbbk(s,t)\mathrm{d}t$$

where the last equality holds due to $\Bbbk_s\in\mathscr{H}_\Bbbk$, $\mathscr{H}_\Bbbk\subset\mathscr{L}^1$, and the dominated convergence theorem. Note that f coincides with $\varphi_0$ defined by (5). Accordingly, due to $\lim_{n\to\infty}\mathrm{f}^{(r_n)}=\mathrm{f}$, (40), the definition of f, and the dominated convergence theorem, we have

$$\begin{aligned}\left\|\int_0^\infty\Bbbk(\cdot,t)\mathrm{d}t\right\|^2&=\|\mathrm{f}\|^2=\lim_{n\to\infty}\|\mathrm{f}^{(r_n)}\|^2\\ &=\lim_{n\to\infty}\int_0^{r_n}\int_0^{r_n}\Bbbk(s,t)\mathrm{d}s\mathrm{d}t=\int_0^\infty\int_0^\infty\Bbbk(s,t)\mathrm{d}s\mathrm{d}t.\end{aligned}$$

Similarly, for any $\mathrm{g} = (g_t)_{t\in\mathbb{R}_+} \in \mathscr{H}_{\Bbbk}$, one can see that

$$\left\langle \int_0^\infty \Bbbk(\cdot,t)\mathrm{d}t, \mathrm{g}\right\rangle = \langle \mathrm{f}, \mathrm{g}\rangle = \lim_{n\to\infty} \langle \mathrm{f}^{(r_n)}, \mathrm{g}\rangle$$
$$= \lim_{n\to\infty} \left\langle \int_0^{r_n} \Bbbk(\cdot,t)\mathrm{d}t, \mathrm{g}\right\rangle = \lim_{n\to\infty} \int_0^{r_n} g_t \mathrm{d}t = \int_0^\infty g_t \mathrm{d}t$$

where the last equality is due to the dominated convergence theorem and the fact that g is integrable. Therefore, we have $\ell_0(\mathrm{g}) = \langle \mathrm{f}, \mathrm{g}\rangle$, which concludes the proof for the case of $\mathbb{T} = \mathbb{R}_+$. The proof for the discrete-time is based on similar line of arguments (see [30, Appendix B] for more details). ■

### 2) Proof of Theorem 2

Let $\delta$ be a real scalar such that $\underline{\delta} \le \delta \le \overline{\delta}$. We know that $\ell_0(\Bbbk_\tau) = \int_{\mathbb{R}_+} \Bbbk(\tau,t)\mathrm{d}t$ is non-zero. Define h as $\mathrm{h} := \frac{\delta}{\ell_0(\Bbbk_\tau)}\Bbbk_\tau \in \mathscr{H}_{\Bbbk}$. One can see that $\ell_0(\mathrm{h}) = \delta \in [\underline{\delta}, \overline{\delta}]$, and hence, $\mathrm{h} \in \mathscr{G}_{\Bbbk}([\underline{\delta}, \overline{\delta}])$. Thus, $\mathscr{G}_{\Bbbk}([\underline{\delta}, \overline{\delta}])$ is non-empty. From Theorem 1, we know that $\mathscr{G}_{\Bbbk}([\underline{\delta}, \overline{\delta}])$ is the intersection of $\{\mathrm{g} \in \mathscr{H}_{\Bbbk} | \langle \varphi_0, \mathrm{g}\rangle \ge \underline{\delta}\}$ and $\{\mathrm{g} \in \mathscr{H}_{\Bbbk} | \langle \varphi_0, \mathrm{g}\rangle \le \overline{\delta}\}$, which are closed half-spaces in $\mathscr{H}_{\Bbbk}$. Therefore, $\mathscr{G}_{\Bbbk}([\underline{\delta}, \overline{\delta}])$ is a convex and closed subset of $\mathscr{H}_{\Bbbk}$ as well. This concludes the proof. ■

### 3) Proof of Theorem 4

Define $\mathcal{J} : \mathscr{H}_{\Bbbk} \to \mathbb{R}_+ \cup \{+\infty\}$ as $\mathcal{J}(\mathrm{g}) := \mathcal{E}_\ell(\mathrm{g}) + \lambda\mathcal{R}(\mathrm{g}) + \delta_{\mathscr{F}}(\mathrm{g})$, for all $\mathrm{g} \in \mathscr{H}_{\Bbbk}$, where $\mathscr{G}$ is defined as $\mathscr{G}_{\Bbbk}([\underline{\delta}, \overline{\delta}])$ for the ease of notation. Note that $\min_{g\in\mathscr{H}_{\Bbbk}} \mathcal{J}(g)$ is equivalent to (4). For h introduced in the proof of Theorem 2, we know $\mathrm{h} \in \mathscr{G}$, i.e., $\delta_{\mathscr{G}}(\mathrm{h}) = 0$. Subsequently, we have $\mathcal{J}(\mathrm{h}) = \mathcal{E}_\ell(\mathrm{h}) + \lambda\rho(\|\mathrm{h}\|) < \infty$, which implies that $\mathcal{J}$ is a proper function. Due to Theorem 2, $\mathscr{G}$ is convex and closed. Accordingly, $\delta_{\mathscr{G}} : \mathscr{H}_{\Bbbk} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semi-continuous convex function [33]. Moreover, from Lemma 3, and also the continuity and convexity of function $\ell$, it follows that $\mathcal{E}_\ell : \mathscr{H}_{\Bbbk} \to \mathbb{R}_+$ is convex and continuous. The convexity of function $\rho$ implies its continuity. Furthermore, since $\rho$ is strictly increasing, we know that $\mathcal{R} : \mathscr{H}_{\Bbbk} \to \mathbb{R}_+$ is a strictly convex continuous function. Thus, $\mathcal{J} : \mathscr{H}_{\Bbbk} \to \mathbb{R} \cup \{+\infty\}$ is a *proper lower semi-continuous strictly convex* function, and subsequently, $\min_{g\in\mathscr{H}_{\Bbbk}} \mathcal{J}(g)$ admits a unique finite solution [33], denoted by $\mathrm{g}^\star$. From the definition of $\mathscr{G}$ as $\mathscr{G}_{\Bbbk}([\underline{\delta}, \overline{\delta}])$, Theorem 1, one has $\delta_{\mathscr{G}}(\mathrm{g}) = \delta_{[\underline{\delta},\overline{\delta}]}(\langle \varphi_0, \mathrm{g}\rangle)$, for all $\mathrm{g} \in \mathscr{H}_{\Bbbk}$. Also, according to Lemma 3, we know that $\mathcal{E}_\ell(\mathrm{g}) = \ell([\langle \varphi_i, \mathrm{g}\rangle]_{i=1}^{n_{\mathcal{I}}}, \mathrm{y}_{\mathcal{I}})$, for all $\mathrm{g} \in \mathscr{H}_{\Bbbk}$. Let function $e : \mathbb{R}^{n_{\mathcal{I}}+1} \to \mathbb{R}_+ \cup \{+\infty\}$ be defined as $e([z_i]_{i=0}^{n_{\mathcal{I}}}) = \ell([z_i]_{i=1}^{n_{\mathcal{I}}}, \mathrm{y}_{\mathcal{I}}) + \lambda\delta_{[\underline{\delta},\overline{\delta}]}(z_0)$, for any $[z_i]_{i=0}^{n_{\mathcal{I}}}$. Then, we have $e([\langle \varphi_i, \mathrm{g}\rangle]_{i=0}^{n_{\mathcal{I}}}) = \mathcal{E}_\ell(\mathrm{g}) + \lambda\delta_{\mathscr{G}}(\mathrm{g})$, for any $\mathrm{g} \in \mathscr{H}_{\Bbbk}$. Since $\min_{g\in\mathscr{H}_{\Bbbk}} \mathcal{J}(g)$ admits a solution, it has a solution in $\mathrm{span}\{\varphi_i\}_{i=0}^{n_{\mathcal{I}}}$ due to representer theorem (see [34] or [30, Appendix O]). According to the uniqueness of this solution, we know that $\mathrm{g}^\star$ belongs to $\mathrm{span}\{\varphi_i\}_{i=0}^{n_{\mathcal{I}}}$, i.e., $\mathrm{g}^\star$ has the mentioned parametric form. For each $\mathrm{g} \in \mathrm{span}\{\varphi_i\}_{i=0}^{n_{\mathcal{I}}}$, we know that there exists $\mathrm{x} = [x_i]_{i=0}^{n_{\mathcal{I}}}$ such that $\mathrm{g} = \sum_{j=0}^{n_{\mathcal{I}}} x_j\varphi_j$. Thus, from linearity of inner product and definition of $\Phi$, A, and $\mathrm{a}_0$, we have $\ell_0(\mathrm{g}) = \mathrm{a}_0^{\mathsf{T}}\mathrm{x}$, $[\mathrm{L}_{t_i}^{\mathrm{u}}(\mathrm{g})]_{i\in\mathcal{I}} = \mathrm{Ax}$, and $\|\mathrm{g}\|^2 = \mathrm{x}^{\mathsf{T}}\Phi\mathrm{x}$. Hence, we obtain convex program (8) by replacing g in (4) with its parametric form. ■

### 4) Proof of Theorem 6

From Lemma 3, the definition of matrices K and $\mathrm{T}_{\mathrm{u}}$, and since $u_t = 0$ for $t < 0$, one can see that

$$\langle \varphi_i, \varphi_j\rangle = \sum_{s=0}^{i-1}\sum_{t=0}^{j-1} \Bbbk(s,t)u_{j-1-t}u_{i-1-s} = [\mathrm{T}_{\mathrm{u}}\mathrm{K}\mathrm{T}_{\mathrm{u}}^{\mathsf{T}}]_{(i,j)} \tag{41}$$

for any $i, j \in \{1, \dots, n_{\mathscr{D}}\}$. Moreover, due to (5), we have

$$\langle \varphi_i, \varphi_0\rangle = \sum_{s=0}^{i-1} \varphi_{0,s}u_{i-1-s} = [\mathrm{T}_{\mathrm{u}}\varphi]_{(i)} \tag{42}$$

for any $i \in \{1, \dots, n_{\mathscr{D}}\}$. Following this, the claim concludes from the definition of matrix $\Phi$ and the fact that $\Phi = \Phi^{\mathsf{T}}$. ■

### 5) Proof of Theorem 7

From (7) and definition of u, we have

$$\varphi_{\tau,t}^{(\mathrm{u})} = \int_{\mathbb{R}_+} \Bbbk(t,s) \sum_{i=0}^{n_{\mathrm{s}}-1} \xi_{i+1}\mathbf{1}_{[s_i,s_{i+1})}(\tau - s)\mathrm{d}s$$
$$= \sum_{i=0}^{n_{\mathrm{s}}-1} \xi_{i+1} \int_{\bar{s}_{i+1}(\tau)}^{\bar{s}_i(\tau)} \Bbbk(t,s)\mathrm{d}s.$$

Thus, the claim is implied by the definition of $\psi$.

### 6) Proof of Theorem 8

First, we consider the case of TC kernel. For $t \le a$, we have

$$\psi_{\mathrm{TC}}(t,a,b) = \int_a^b \alpha^s \mathrm{d}s = \frac{\alpha^b - \alpha^a}{\ln(\alpha)}. \tag{43}$$

Also, for $t \ge b$, one has

$$\psi_{\mathrm{TC}}(t,a,b) = \int_a^b \alpha^t \mathrm{d}s = (b-a)\alpha^t. \tag{44}$$

Similar to the previous cases, when $t \in (a,b)$, we have

$$\psi_{\mathrm{TC}}(t,a,b) = \int_a^t \alpha^t \mathrm{d}s + \int_t^b \alpha^s \mathrm{d}s = (t-a)\alpha^t + \frac{\alpha^b - \alpha^t}{\ln(\alpha)}.$$

Accordingly, due to (43), (44), and the definition of $\eta$, one can see (18) holds. We have a similar argument for DC and SS kernels (see [30, Appendix J] for more details). ■

### 7) Proof of Theorem 9

From the definition of $\varphi_0$ and function $\psi$, we have

$$\varphi_{0,t} = \int_0^\infty \Bbbk(s,t)\mathrm{d}s = \lim_{b\to\infty} \int_0^b \Bbbk(s,t)\mathrm{d}s = \lim_{b\to\infty} \psi(t,0,b).$$

Thus, the theorem holds according to Theorem 8. ■

### 8) Proof of Theorem 10

Due to Theorem 1, Lemma 3 and (15), one has

$$\left\langle \varphi_0, \varphi_\tau^{(\mathrm{u})}\right\rangle = \int_0^\infty \varphi_{0,s} \sum_{i=0}^{n_{\mathrm{s}}-1} \xi_{i+1}\mathbf{1}_{[s_i,s_{i+1})}(\tau - s)\mathrm{d}s$$
$$= \int_0^\infty \int_0^\infty \Bbbk(s,t) \sum_{i=0}^{n_{\mathrm{s}}-1} \xi_{i+1}\mathbf{1}_{[s_i,s_{i+1})}(\tau - s)\mathrm{d}t\mathrm{d}s. \tag{45}$$

Thus, from the definition of $\bar{\nu}$ and $\bar{\kappa}_i$, $i = 0, \dots, n_{\mathrm{s}} - 1$, we have

$$\left\langle \varphi_0, \varphi_\tau^{(\mathrm{u})}\right\rangle = \sum_{i=0}^{n_{\mathrm{s}}-1} \xi_{i+1} \int_{\bar{s}_{i+1}(\tau)}^{\bar{s}_i(\tau)} \int_0^\infty \Bbbk(s,t)\mathrm{d}t\mathrm{d}s$$
$$= \sum_{i=0}^{n_{\mathrm{s}}-1} \xi_{i+1} \left(\bar{\nu}(\bar{s}_i(\tau)) - \bar{\nu}(\bar{s}_{i+1}(\tau))\right) = \sum_{i=0}^{n_{\mathrm{s}}-1} \xi_{i+1}\bar{\kappa}_i(\tau). \tag{46}$$

From Lemma 3 and due to (7) and (15), we know that

$$\begin{aligned}\langle\varphi_{\tau_1}^{(\mathrm{u})},\varphi_{\tau_2}^{(\mathrm{u})}\rangle &= \int_0^\infty \varphi_{\tau_2,s}^{(\mathrm{u})}\sum_{i=0}^{n_\mathrm{s}-1}\xi_{i+1}\mathbf{1}_{[s_i,s_{i+1})}(\tau_1-s)\mathrm{d}s\\ &= \int_0^\infty\int_0^\infty\Bigg(\Bbbk(s,t)\sum_{j=0}^{n_\mathrm{s}-1}\xi_{j+1}\mathbf{1}_{[s_j,s_{j+1})}(\tau_2-t)\\ &\quad\sum_{i=0}^{n_\mathrm{s}-1}\xi_{i+1}\mathbf{1}_{[s_i,s_{i+1})}(\tau_1-s)\Bigg)\mathrm{d}t\mathrm{d}s\\ &= \sum_{i=0}^{n_\mathrm{s}-1}\sum_{j=0}^{n_\mathrm{s}-1}\xi_{i+1}\xi_{j+1}\int_0^\infty\int_0^\infty\Bigg(\Bbbk(s,t)\\ &\quad\mathbf{1}_{[s_j,s_{j+1})}(\tau_2-t)\mathbf{1}_{[s_i,s_{i+1})}(\tau_1-s)\Bigg)\mathrm{d}t\mathrm{d}s.\end{aligned}\tag{47}$$

From (24), it follows that

$$\begin{aligned}&\int_0^\infty\int_0^\infty\Bbbk(s,t)\mathbf{1}_{[s_j,s_{j+1})}(\tau_2-t)\mathbf{1}_{[s_i,s_{i+1})}(\tau_1-s)\mathrm{d}t\mathrm{d}s\\ &= \int_{\bar{s}_{i+1}(\tau_1)}^{\bar{s}_i(\tau_1)}\int_{\bar{s}_{j+1}(\tau_2)}^{\bar{s}_j(\tau_2)}\Bbbk(s,t)\mathrm{d}t\mathrm{d}s\\ &= \nu\left(\bar{s}_i(\tau_1),\bar{s}_j(\tau_2)\right)-\nu\left(\bar{s}_{i+1}(\tau_1),\bar{s}_j(\tau_2)\right)\\ &\quad-\nu\left(\bar{s}_i(\tau_1),\bar{s}_{j+1}(\tau_2)\right)+\nu\left(\bar{s}_{i+1}(\tau_1),\bar{s}_{j+1}(\tau_2)\right).\end{aligned}\tag{48}$$

Thus, the claim is implied from (47) and (48). ■

### 9) Proof of Theorem 11

Since $\bar{\nu}(x)=\lim_{y\to\infty}\nu(x,y)$, it is enough to obtain $\nu(x,y)$. Without loss of generality we assume $x\le y$. Due to the definition of $\nu$ in (24), for TC kernel, we have

$$\begin{aligned}\nu_{\mathrm{TC}}(x,y) &= \int_0^x\int_0^s\alpha^s\mathrm{d}t\mathrm{d}s+\int_0^x\int_s^y\alpha^t\mathrm{d}t\mathrm{d}s\\ &= \int_0^x s\alpha^s\mathrm{d}s+\int_0^x\frac{\alpha^y-\alpha^s}{\ln(\alpha)}\mathrm{d}s\\ &= \frac{x\alpha^x\ln(\alpha)+1-\alpha^x}{\ln(\alpha)^2}+\frac{x\alpha^y}{\ln(\alpha)}-\frac{\alpha^x-1}{\ln(\alpha)^2}.\end{aligned}\tag{49}$$

Reordering the terms and replacing $x$ and $y$, respectively, with $\min(x,y)$ and $\max(x,y)$, we obtain $\nu_{\mathrm{TC}}(x,y)$. For $\Bbbk_{\mathrm{DC}}$ and $\Bbbk_{\mathrm{SS}}$, similar arguments hold (see [30, Appendix M] for more details). ■

### 10) Proof of Theorem 12

From (6) and dominated convergence theorem, we know $\|\varphi_0\|^2=\lim_{x,y\to\infty}\int_0^x\int_0^y\Bbbk(s,t)\mathrm{d}s\mathrm{d}t$. Thus, from the definition of $\nu$, we have $\|\varphi_0\|^2=\lim_{x,y\to\infty}\nu(x,y)$. From this, the proof is followed directly from Theorem 11. ■

## References


[1] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.

[2] T. A. Johansen, "Constrained and regularized system identification," *Model., Identification Control*, vol. 19, no. 2, pp. 109–116, 1998.

[3] P. Trnka and V. Havlena, "Subspace like identification incorporating prior information," *Automatica*, vol. 45, no. 4, pp. 1086–1091, 2009.

[4] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[5] J. Umenberger and I. R. Manchester, "Specialized interior-point algorithm for stable nonlinear system identification," *IEEE Trans. Autom. Control*, vol. 64, no. 6, pp. 2442–2456, Jun. 2019.

[6] M. Khosravi and R. S. Smith, "Nonlinear system identification with prior knowledge on the region of attraction," *IEEE Control Syst. Lett.*, vol. 5, no. 3, pp. 1091–1096, Jul. 2021.

[7] A. A. Ahmadi and B. El Khadir, "Learning dynamical systems with side information (short version)," *Proc. Mach. Learn. Res.*, vol. 120, pp. 718–727, 2020.

[8] M. Khosravi and R. S. Smith, "Convex nonparametric formulation for identification of gradient flows," *IEEE Control Syst. Lett.*, vol. 5, no. 3, pp. 1097–1102, Jul. 2021.

[9] D. N. Miller and R. A. De Callafon, "Subspace identification with eigenvalue constraints," *Automatica*, vol. 49, no. 8, pp. 2468–2473, 2013.

[10] P. Massioni and M. Verhaegen, "Subspace identification of circulant systems," *Automatica*, vol. 44, no. 11, pp. 2825–2833, 2008.

[11] M. Inoue, "Subspace identification with moment matching," *Automatica*, vol. 99, pp. 22–32, 2019.

[12] L. F. Rodrigues, L. P. Ihlenfeld, and G. H. da C. Oliveira, "A novel subspace identification approach with passivity enforcement," *Automatica*, vol. 132, 2021, Art. no. 109798.

[13] R. S. Smith, "Frequency domain subspace identification using nuclear norm minimization and Hankel matrix realizations," *IEEE Trans. Autom. Control*, vol. 59, no. 11, pp. 2886–2896, Nov. 2014.

[14] M. Khosravi and R. S. Smith, "Kernel-based identification of positive systems," in *Proc. Conf. Decis. Control*, 2019, pp. 1740–1745.

[15] S. L. Lacy and D. S. Bernstein, "Subspace identification with guaranteed stability using constrained optimization," *IEEE Trans. Autom. Control*, vol. 48, no. 7, pp. 1259–1263, Jul. 2003.

[16] T. Van Gestel, J. A. Suykens, P. Van Dooren, and B. De Moor, "Identification of stable models in subspace identification by using regularization," *IEEE Trans. Autom. Control*, vol. 46, no. 9, pp. 1416–1420, Sep. 2001.

[17] M. Khosravi and R. S. Smith, "Kernel-based identification with frequency domain side-information," *Automatica*, vol. 150, 2023, Art. no. 110813.

[18] P. Grosdidier, M. Morari, and B. R. Holt, "Closed-loop properties from steady-state gain information," *Ind. Eng. Chem. Fundam.*, vol. 24, no. 2, pp. 221–235, 1985.

[19] P. J. Campo and M. Morari, "Achievable closed-loop properties of systems under decentralized control: Conditions involving the steady-state gain," *IEEE Trans. Autom. Control*, vol. 39, no. 5, pp. 932–943, May 1994.

[20] Z. Q. Zheng and M. Morari, "Robust stability of constrained model predictive control," in *Proc. IEEE Amer. Control Conf.*, 1993, pp. 379–383.

[21] A. Alenany, H. Shang, M. Soliman, and I. Ziedan, "Improved subspace identification with prior information using constrained least squares," *IET Control Theory Appl.*, vol. 5, no. 13, pp. 1568–1576, 2011.

[22] S. Yoshimura, A. Matsubayashi, and M. Inoue, "System identification method inheriting steady-state characteristics of existing model," *Int. J. Control*, vol. 92, no. 11, pp. 2701–2711, 2019.

[23] Y. Abe, M. Inoue, and S. Adachi, "Subspace identification method incorporated with a priori information characterized in frequency domain," in *Proc. IEEE Eur. Control Conf.*, 2016, pp. 1377–1382.

[24] I. Markovsky and G. Mercère, "Subspace identification with constraints on the impulse response," *Int. J. Control*, vol. 90, no. 8, pp. 1728–1735, 2017.

[25] Y. Fujimoto and T. Sugie, "Kernel-based impulse response estimation with a priori knowledge on the DC gain," *IEEE Control Syst. Lett.*, vol. 2, no. 4, pp. 713–718, Oct. 2018.

[26] A. H. Tan and D. S. Ong, "Kernel-based impulse response estimation with prior DC gain using built-in self-scaling technique," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 7295–7305, Sep. 2020.

[27] H. Garnier, L. Wang, and P. C. Young, "Direct identification of continuous-time models from sampled data: Issues, basic solutions and relevance," in *Identification of Continuous-Time Models From Sampled Data*. London, U.K.: Springer, 2008, pp. 1–29.

[28] H. Garnier and M. Gilson, "CONTSID: A MATLAB toolbox for standard and advanced identification of black-box continuous-time models," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 688–693, 2018.

[29] H. Brézis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Berlin, Germany: Springer, 2011.

[30] M. Khosravi and R. S. Smith, "Kernel-based impulse response identification with side-information on steady-state gain," 2021, *arXiv:2111.00409*.

[31] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3250–3265, May 2012.

[32] L. Ljung, T. Chen, and B. Mu, "A shift in paradigm for system identification," *Int. J. Control*, vol. 93, no. 2, pp. 173–180, 2020.

[33] J. Peypouquet, *Convex Optimization in Normed Spaces: Theory, Methods and Examples*. Berlin, Germany: Springer, 2015. [Online]. Available: https://doi.org/10.1007/978-3-319-13710-0

[34] F. Dinuzzo and B. Schölkopf, "The representer theorem for Hilbert spaces: A necessary and sufficient condition," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 189–196.