The existence and uniqueness of solutions for kernel-based system identification

Khosravi, Mohammad; Smith, Roy S.

DOI
10.1016/j.automatica.2022.110728

Publication date
2023

Document Version
Final published version

Published in
Automatica

Automatica 148 (2023) 110728

Contents lists available at ScienceDirect

Automatica

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

The existence and uniqueness of solutions for kernel-based system identification ☆

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

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

ARTICLE INFO

Article history:
Received 24 April 2022
Received in revised form 28 June 2022
Accepted 8 September 2022
Available online xxxx

Keywords:
System identification
Kernel-based methods
Existence and uniqueness of solution
Integrable kernels

ABSTRACT

The notion of reproducing kernel Hilbert space (RKHS) has emerged in system identification during the past decade. In the resulting framework, the impulse response estimation problem is formulated as a regularized optimization defined on an infinite-dimensional RKHS consisting of stable impulse responses. The consequent estimation problem is well-defined under the central assumption that the convolution operators restricted to the RKHS are continuous linear functionals. Moreover, according to this assumption, the representer theorem hold, and therefore, the impulse response can be estimated by solving a finite-dimensional program. Thus, the continuity feature plays a significant role in kernel-based system identification. We show that this central assumption is guaranteed to be satisfied in considerably general situations, namely when the input signal is bounded, the kernel is an integrable function, and in the case of continuous-time dynamics, continuous. Furthermore, the strong convexity of the optimization problem and the continuity property of the convolution operators imply that the kernel-based system identification admits a unique solution. Consequently, it follows that kernel-based system identification is a well-defined approach.

1. Introduction

System identification, the theory of generating suitable abstract representations for dynamical systems based on measurement data, is a well-established research field (Zadeh, 1956). Due to the importance of mathematical models in various areas of science and technology, system identification is an active research area with numerous developed methodologies (Ahmadi & El Khadir, 2020; Khosravi & Smith, 2021a, 2021d; Ljung, 1999, 2010; Schoukens & Ljung, 2019). On the other hand, the concept of reproducing kernel Hilbert space (RKHS), initially introduced in Aronszajn (1950), has emerged in statistics, signal processing and numerical analysis (Berlinet & Thomas-Agnan, 2011; Cucker & Smale, 2002a; Kailath, 1971; Parzen, 1959, 1961; Wahba, 1990), and provided a solid foundation for estimation and interpolation problems. The inherent features of RKHSs, such as their fundamental relation to the positive semi-definite kernels and the Gaussian process (Kanagawa, Hennig, Sejdinovic, & Sriperumbudur, 2018; Kimeldorf & Wahba, 1970; Lukić & Beder, 2001), led to establishing various methodologies and opened numerous avenues of research in statistical learning theory (Cucker & Smale, 2002b).

In the seminal work of Pillonetto and De Nicolao (2010), the *kernel-based identification* methods are introduced by bringing the theory of RKHSs to the area of linear system identification, which led to a paradigm shift in the field (Ljung, Chen, & Mu, 2020). The kernel-based method unifies the identification theory of continuous-time systems and discrete-time systems, described either with a finite or an infinite impulse response, by formulating the identification problem as a regularized regression defined on a RKHS of stable systems, where the regularization term is specified based on the norm of employed RKHS (Pillonetto, Dinuzzo, Chen, De Nicolao, & Ljung, 2014). The resulting formulation addresses issues of model order selection, robustness, and bias–variance trade-off (Chiuso & Pillonetto, 2019; Khosravi & Smith, 2021e; Pillonetto et al., 2014). The cornerstone of a RKHS is the associated kernel function, which highlights the necessity of designing suitable kernels for system identification (Dinuzzo, 2015). The most frequently used kernels in the literature are tuned/correlated (TC), diagonal/correlated (DC), stable spline (SS), and their generalizations (Andersen & Chen, 2020; Chen, 2018a; Zorzi, 2021). Other forms of kernels and regularization matrices have been proposed, inspired by machine learning, system theory, harmonic analysis of stochastic processes,

☆ 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.2022.110728

and filter design methods (Chen, Andersen, Ljung, Chiuso, & Pillonetto, 2014; Marconato, Schoukens, & Schoukens, 2016; Zorzi & Chiuso, 2018). While in the classical identification methods, the complexity of models is described by the orders of system, which are integer variables determined based on metrics such as Akaike information criterion (Ljung, 1999), the model complexity in kernel-based approach is specified and regulated by the hyperparameters characterizing the kernel and the regularization weight, which are continuous variables to be tuned (Ljung et al., 2020). The estimation of hyperparameters can be performed using powerful and robust methods such as empirical Bayes, Stein unbiased risk estimator, and cross-validation (Mu, Chen, & Ljung, 2018a, 2018b, 2021; Pillonetto & Chiuso, 2015). Moreover, the kernel-based scheme allows the incorporation of various forms of side-information in the identification problem by designing appropriate kernel functions or imposing suitable constraints to the regression problem. The forms of this side-information, studied to date, include stability, relative degree, smoothness of the impulse response, resonant frequencies, external positivity, oscillatory behaviors, steady-state gain, internal positivity, exponential decay of the impulse response, structural properties, internal low-complexity, frequency domain features, and the presence of fast and slow poles (Chen, 2018b; Chen, Ohlsson, & Ljung, 2012; Darwish, Pillonetto, & Tóth, 2018; Everitt, Bottegal, & Hjalmarsson, 2018; Fujimoto, Maruta, & Sugie, 2017; Fujimoto & Sugie, 2018; Khosravi, Iannelli, Yin, Parsi and Smith, 2020; Khosravi & Smith, 2019, 2021b, 2021c, 2021f; Khosravi, Yin, Iannelli, Parsi and Smith, 2020; Marconato et al., 2016; Pillonetto, Chen, Chiuso, Nicolao, & Ljung, 2016; Prando, Chiuso, & Pillonetto, 2017; Risuleo, Bottegal, & Hjalmarsson, 2017; Risuleo, Lindsten, & Hjalmarsson, 2019; Zheng & Ohta, 2021). While kernel-based system identification has enjoyed considerable progress in the past decade, it is still a thriving area of research with state-of-the-art results and recent studies (Bisiacco & Pillonetto, 2020a, 2020b; Pillonetto, Chiuso, & De Nicolao, 2019; Pillonetto & Scampicchio, 2021; Scandella, Mazzoleni, Formentin, & Previdi, 2020, 2021). For example, the mathematical foundation of stable RKHSs is revisited in Bisiacco and Pillonetto (2020b), the sample complexity and the minimax properties of kernel-based methods are discussed in Pillonetto and Scampicchio (2021), and a long-standing question on the absolute summability of stable kernels is addressed in Bisiacco and Pillonetto (2020a).

The above-mentioned advantages of kernel-based methods stand on the assumption that the formulated regression problem is well-defined, i.e., the corresponding regularized optimization problem admits at least one solution. The base of this assumption is the continuity of convolution operators when they are restricted to the stable RKHSs (Dinuzzo, 2015; Pillonetto et al., 2014). Accordingly, one may ask about the conditions under which the continuity property holds. This paper shows that this central assumption is satisfied in certain but highly general situations, namely when the input signal is bounded, the kernel is an integrable function, and in the case of continuous-time dynamics, continuous. As a result, kernel-based system identification admits a unique solution according to the continuity of convolution operators and the strong convexity of the optimization problem, which also implies that the kernel-based approach is well-defined.

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, and the n-dimensional Euclidean space are denoted by $\mathbb{N}$, $\mathbb{Z}_+$, $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{R}^n$, respectively. Throughout the paper, $\mathbb{T}$ denotes either $\mathbb{Z}_+$ or $\mathbb{R}_+$, and $\mathbb{T}_\pm$ is defined as the set of scalars t where $t \in \mathbb{T}$ or $-t \in \mathbb{T}$. Furthermore, we consider the measurable space $(\mathbb{T}, \mathscr{F}_\mathbb{T})$, where $\mathscr{F}_\mathbb{T}$ is the Borel subsets of $\mathbb{R}_+$ when $\mathbb{T} = \mathbb{R}_+$, and $\mathscr{F}_\mathbb{T}$ is the power set of $\mathbb{Z}_+$ when $\mathbb{T} = \mathbb{Z}_+$. Accordingly, $\mathbb{T} \times \mathbb{T}$ is equipped with the product σ-algebra $\mathscr{F}_\mathbb{T} \otimes \mathscr{F}_\mathbb{T}$. Also, $\mathbb{R}$ is endowed with the Borel σ-algebra. The identity matrix/operator and the zero vector are denoted by $\mathbb{I}$ and $\mathbf{0}$, respectively. Given measurable space $\mathscr{X}$, we denote by $\mathbb{R}^\mathscr{X}$ as the space of measurable functions $\mathrm{v} : \mathscr{X} \to \mathbb{R}$. The element $\mathrm{v} \in \mathbb{R}^\mathscr{X}$ is shown entry-wise as $\mathrm{v} = (v_x)_{x\in\mathscr{X}}$, or $\mathrm{v} = (v(x))_{x\in\mathscr{X}}$. Depending on the context of discussion, $\mathscr{L}^\infty$ refers either to $\ell^\infty(\mathbb{Z})$ or $L^\infty(\mathbb{R})$. 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 $\|\cdot\|_p$. With respect to each $\mathrm{u} = (u_s)_{s\in\mathbb{T}_\pm} \in \mathscr{L}^\infty$ and $t \in \mathbb{T}_\pm$, the linear operator $\mathrm{L}_t : \mathscr{L}^1 \to \mathbb{R}$ is defined as $\mathrm{L}_t(\mathrm{g}) := \sum_{s\in\mathbb{Z}_+} g_s u_{t-s}$, when $\mathbb{T} = \mathbb{Z}_+$, and $\mathrm{L}_t(\mathrm{g}) := \int_{\mathbb{R}_+} g_s u_{t-s} \mathrm{d}s$, when $\mathbb{T} = \mathbb{R}_+$. Given sets $\mathcal{X}$ and $\mathcal{Y}$, where $\mathcal{Y} \subseteq \mathcal{X}$, the *indicator* function $\mathbf{1}_\mathcal{Y} : \mathcal{X} \to \{0, 1\}$ is defined as $\mathbf{1}_\mathcal{Y}(x) = 1$, when $x \in \mathcal{Y}$, and $\mathbf{1}_\mathcal{Y}(x) = 0$, otherwise. We say $f : \mathbb{R}_+ \to \mathbb{R}$ is a *simple* function, if there exist real numbers $a_1, \ldots, a_n \in \mathbb{R}$ and intervals $\mathrm{I}_1, \ldots, \mathrm{I}_n \subset \mathbb{R}_+$, such that $f(x) = \sum_{j=1}^n a_j \mathbf{1}_{\mathrm{I}_j}(x)$, for any $x \in \mathbb{R}_+$.

3. Kernel-based system identification

Consider a stable LTI system $\mathcal{S}$ characterized by an impulse response $\mathrm{g}^{(\mathcal{S})} := (g_t^{(\mathcal{S})})_{t\in\mathbb{T}} \in \mathbb{R}^\mathbb{T}$, where $\mathbb{T}$ is $\mathbb{Z}_+$ or $\mathbb{R}_+$ respectively for the case that the system is discrete-time or continuous-time. Suppose the system $\mathcal{S}$ is actuated by a signal $\mathrm{u} \in \mathscr{L}^\infty$, and the resulting output signal is measured with measurement noise at $n_\mathscr{D}$ time instants $t_1, \ldots, t_{n_\mathscr{D}}$. Let the measured output of the system at time instant t_i, and the corresponding measurement uncertainty, be denoted by y_{t_i} and w_{t_i}, respectively. Due to the definition of operators $\{\mathrm{L}_t | t \in \mathbb{T}_\pm\}$, we know that

$$y_{t_i} = \mathrm{L}_{t_i}(\mathrm{g}^{(\mathcal{S})}) + w_{t_i}, \qquad i = 1, \ldots, n_\mathscr{D}. \tag{1}$$

Therefore, we are provided with a set of input–output measurement data denoted by $\mathscr{D}$. Accordingly, the impulse response identification problem is formalized as estimating $\mathrm{g}^{(\mathcal{S})}$, the impulse response of stable system $\mathcal{S}$, based on the measurement data. In the kernel-based identification framework, this problem is formulated as an impulse response estimation in a reproducing kernel Hilbert space (RKHS) endowed with a stable kernel. To introduce the main result of this paper, we need to discuss this paradigm briefly.

Definition 1 (*Berlinet & Thomas-Agnan, 2011*)**.** Consider symmetric function $\mathbb{k} : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$, that is assumed to be continuous if $\mathbb{T} = \mathbb{R}_+$. We say $\mathbb{k}$ is a *Mercer kernel* when we have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \mathbb{k}(t_i, t_j) a_j \geq 0, \tag{2}$$

for all $n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathbb{T}$, and $a_1, \ldots, a_n \in \mathbb{R}$. Furthermore, with respect to each $\tau \in \mathbb{T}$, the *section* of kernel $\mathbb{k}$ at τ is the function $\mathbb{k}_\tau : \mathbb{T} \to \mathbb{R}$ defined as $\mathbb{k}_\tau(\cdot) = \mathbb{k}(\tau, \cdot)$.

Remark 1. When the continuity assumption in Definition 1 is relaxed, $\mathbb{k}$ is referred to as a *positive-definite kernel*, or simply, a *kernel*.[1] Accordingly, the Mercer kernels are positive-definite kernels, which are continuous when $\mathbb{T} = \mathbb{R}_+$. One should note that the mentioned continuity feature plays a significant role in the main result of this paper presented in Section 4.

[1] In the literature, kernels are commonly assumed to be positive-definite without being explicitly mentioned.

Theorem 1 (Berlinet & Thomas-Agnan, 2011). *With respect to each positive-definite kernel $\Bbbk : \mathbb{T}\times\mathbb{T}\to\mathbb{R}$, a unique Hilbert space $\mathscr{H}_\Bbbk \subseteq \mathbb{R}^{\mathbb{T}}$ endowed with inner product $\langle\cdot,\cdot\rangle_{\mathscr{H}_\Bbbk}$ exists such that, for each $t\in\mathbb{T}$, one has*
(i) $\Bbbk_t \in \mathscr{H}_\Bbbk$, *and*
(ii) $\langle \mathrm{g}, \Bbbk_t\rangle_{\mathscr{H}_\Bbbk} = \mathrm{g}_t$, *for all* $\mathrm{g} = (\mathrm{g}_s)_{s\in\mathbb{T}} \in \mathscr{H}_\Bbbk$.

In this case, we say $\mathscr{H}_\Bbbk$ is the RKHS with kernel $\Bbbk$. *Moreover, the second feature is called the* reproducing property.

Due to Theorem 1, one can see that each RKHS is uniquely characterized by the corresponding Mercer kernel. Since the to-be-estimated impulse response is known to be stable in the bounded-input–bounded-output (BIBO) sense, the employed kernel $\Bbbk$ is required to guarantee that $\mathscr{H}_\Bbbk \subseteq \mathscr{L}^1$. The sufficient and necessary condition for this property is established by the following theorem.

Theorem 2 (Carmeli, De Vito, & Toigo, 2006; Chen & Pillonetto, 2018). *Consider the positive-definite measurable kernel $\Bbbk : \mathbb{T}\times\mathbb{T}\to\mathbb{R}$ and the corresponding RKHS $\mathscr{H}_\Bbbk$. Then, $\mathscr{H}_\Bbbk \subseteq \mathscr{L}^1$ if and only if, for any $\mathrm{u} = (u_s)_{s\in\mathbb{T}} \in \mathscr{L}^\infty$, one has*

$$\sum_{t\in\mathbb{Z}_+}\left|\sum_{s\in\mathbb{Z}_+} u_s \Bbbk(t,s)\right| < \infty, \tag{3}$$

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

$$\int_{\mathbb{R}_+}\left|\int_{\mathbb{R}_+} u_s \Bbbk(t,s)\mathrm{d}s\right|\mathrm{d}t < \infty, \tag{4}$$

when $\mathbb{T}=\mathbb{R}_+$. When this property holds, kernel $\Bbbk$ is called stable *and $\mathscr{H}_\Bbbk$ is said to be a* stable RKHS.

Given the stable kernel $\Bbbk$ and the measurement data, the *kernel-based impulse response estimation* problem is formulated as

$$\min_{\mathrm{g}\in\mathscr{H}_\Bbbk} \sum_{i=1}^{n_{\mathscr{D}}} (\mathrm{L}_{t_i}(\mathrm{g}) - y_i)^2 + \lambda\|\mathrm{g}\|^2_{\mathscr{H}_\Bbbk}, \tag{5}$$

where $\lambda > 0$ is the regularization weight. Based on the same arguments as in Wahba (1990, Theorem 1.3.1), one can describe the solution of (5) in terms of the sections of the kernel at $t_1,\dots,t_{n_{\mathscr{D}}}$. To this end, we need vector y defined as $\mathrm{y} = \left[y_{t_1},\dots,y_{t_{n_{\mathscr{D}}}}\right]^{\mathsf{T}} \in \mathbb{R}^{n_{\mathscr{D}}}$, and the *output kernel matrix* $\mathrm{O}\in\mathbb{R}^{n_{\mathscr{D}}\times n_{\mathscr{D}}}$ formed from the input signal and defined entry-wise as

$$[\mathrm{O}]_{(i,j)} = \begin{cases} \displaystyle\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \Bbbk(s,t)u_{t_i-s}u_{t_j-t}\,\mathrm{d}s\,\mathrm{d}t, & \text{if } \mathbb{T}_+ = \mathbb{R}_+,\\ \displaystyle\sum_{t\in\mathbb{Z}_+}\sum_{s\in\mathbb{Z}_+} \Bbbk(s,t)u_{t_i-s}u_{t_j-t}, & \text{if } \mathbb{T}_+ = \mathbb{Z}_+, \end{cases}$$

for each $i,j = 1,\dots,n_{\mathscr{D}}$.

Theorem 3 (*Representer Theorem for System Identification*, Pillonetto et al. (2014)). *Let $\mathrm{L}_{t_i} : \mathscr{H}_\Bbbk \to \mathbb{R}$ be a continuous linear operator, for each $i = 1,\dots,n_{\mathscr{D}}$. Then, the minimizer of (5) is $\mathrm{g}^\star = (\mathrm{g}^\star_t)_{t\in\mathbb{T}} \in \mathscr{H}_\Bbbk$ defined as*

$$\mathrm{g}^\star_t = \sum_{i=1}^{n_{\mathscr{D}}} c_i^\star \mathrm{L}_{t_i}(\Bbbk_t), \qquad \forall t\in\mathbb{T}, \tag{6}$$

where the vector $\mathrm{c}^\star = \left[c_1^\star,\dots,c_{n_{\mathscr{D}}}^\star\right]^{\mathsf{T}} \in \mathbb{R}^{n_{\mathscr{D}}}$ is

$$\mathrm{c}^\star = \left(\mathrm{O} + \lambda\mathbb{I}_{n_{\mathscr{D}}}\right)^{-1}\mathrm{y}, \tag{7}$$

and, $\mathbb{I}_{n_{\mathscr{D}}}$ denotes identity matrix of dimension $n_{\mathscr{D}}$.

The main assumption in Theorem 3 is the continuity of convolution operators $\mathrm{L}_{t_1},\dots,\mathrm{L}_{t_{n_{\mathscr{D}}}}$, which depends mainly on the input signal u and kernel $\Bbbk$. Accordingly, a natural question one may ask is *under what conditions are the convolution operators continuous*. Indeed, one should note that in Theorem 3, the convolution operators are restricted to $\mathscr{H}_\Bbbk \subseteq \mathscr{L}^1$, and consequently, the continuity of $\mathrm{L}_t : \mathscr{L}^1 \to \mathbb{R}$ does not imply that the restricted operator $\mathrm{L}_t : \mathscr{H}_\Bbbk \to \mathbb{R}$ is continuous as well. We address this continuity concern in the next section.

4. Continuity of convolution operators

The main result of this section is based on the notion of integrable kernels introduced below.

Definition 2 (Pillonetto et al., 2014). The positive-definite measurable kernel $\Bbbk : \mathbb{T}\times\mathbb{T}\to\mathbb{R}$ is said to be *integrable* if

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} |\Bbbk(s,t)|\,\mathrm{d}s\mathrm{d}t < \infty, \tag{8}$$

when $\mathbb{T}=\mathbb{R}_+$, or, if

$$\sum_{s\in\mathbb{Z}_+}\sum_{t\in\mathbb{Z}_+} |\Bbbk(s,t)| < \infty, \tag{9}$$

when $\mathbb{T}=\mathbb{Z}_+$.

The integrable kernels are the largest known interesting subclass of stable kernels in the context of kernel-based impulse response identification (Bisiacco & Pillonetto, 2020a, 2020b). To present the main theorem of this paper, we need to introduce additional lemmas. Before further proceeding, we first recall that Mercer kernels are positive-definite, and in the case of continuous-time dynamics, they are continuous.

Lemma 4. *Let $\mathbb{T}=\mathbb{R}_+$ and $\Bbbk$ be an integrable Mercer kernel. Consider $\underline{\tau}$ and $\overline{\tau}$ such that $0\le\underline{\tau}<\overline{\tau}\le\infty$. Then, $\int_{[\underline{\tau},\overline{\tau}]}\Bbbk(\cdot,t)\mathrm{d}t$ is a well-defined function and belongs to $\mathscr{H}_\Bbbk$ for which we have*

$$\left\|\int_{[\underline{\tau},\overline{\tau}]}\Bbbk(\cdot,t)\mathrm{d}t\right\|^2_{\mathscr{H}_\Bbbk} = \int_{\underline{\tau}}^{\overline{\tau}}\int_{\underline{\tau}}^{\overline{\tau}}\Bbbk(s,t)\mathrm{d}s\mathrm{d}t. \tag{10}$$

Moreover, for each $\mathrm{g} = (\mathrm{g}_t)_{t\in\mathbb{R}_+} \in \mathscr{H}_\Bbbk$, the following holds

$$\int_{[\underline{\tau},\overline{\tau}]}\mathrm{g}_t\mathrm{d}t = \left\langle\int_{[\underline{\tau},\overline{\tau}]}\Bbbk(\cdot,t)\mathrm{d}t, \mathrm{g}\right\rangle_{\mathscr{H}_\Bbbk}. \tag{11}$$

Proof. See Appendix A.1. □

From this Lemma, we have the following corollary.

Corollary 5. *Let $\mathbb{T}=\mathbb{R}_+$ and $\Bbbk$ be an integrable Mercer kernel. Consider $\underline{\tau}_1$, $\underline{\tau}_2$, $\overline{\tau}_2$ and $\overline{\tau}_2$ such that $0\le\underline{\tau}_1<\overline{\tau}_1\le\infty$ and $0\le\underline{\tau}_2<\overline{\tau}_2\le\infty$. Then, we have*

$$\begin{aligned}\left\langle\int_{[\underline{\tau}_1,\overline{\tau}_1]}\Bbbk(\cdot,t)\mathrm{d}t, \int_{[\underline{\tau}_2,\overline{\tau}_2]}\Bbbk(\cdot,s)\mathrm{d}s\right\rangle_{\mathscr{H}_\Bbbk} &= \int_{[\underline{\tau}_1,\overline{\tau}_1]\times[\underline{\tau}_2,\overline{\tau}_2]}\Bbbk(t,s)\mathrm{d}s\mathrm{d}t\\ &= \int_{\underline{\tau}_1}^{\overline{\tau}_1}\int_{\underline{\tau}_2}^{\overline{\tau}_2}\Bbbk(t,s)\mathrm{d}s\mathrm{d}t\\ &= \int_{\underline{\tau}_2}^{\overline{\tau}_2}\int_{\underline{\tau}_1}^{\overline{\tau}_1}\Bbbk(t,s)\mathrm{d}t\mathrm{d}s.\end{aligned} \tag{12}$$

Proof. See Appendix A.2. □

The next lemma is the discrete-time version of Lemma 4.

Lemma 6. *Let $\mathbb{T}=\mathbb{Z}_+$ and kernel $\Bbbk$ be integrable. Consider $\underline{\tau},\overline{\tau}\in\mathbb{Z}_+$ such that $0\le\underline{\tau}\le\overline{\tau}\le\infty$. Then, $\sum_{\underline{\tau}\le t\le\overline{\tau}}\Bbbk(\cdot,t)$ is a well-defined function and belongs to $\mathscr{H}_\Bbbk$ for which we have*

$$\Big\|\sum_{\underline{\tau}\le t\le\overline{\tau}}\Bbbk(\cdot,t)\Big\|^2_{\mathscr{H}_\Bbbk}=\sum_{\underline{\tau}\le s,t\le\overline{\tau}}\Bbbk(s,t). \tag{13}$$

Moreover, for each $\mathrm{g}=(g_t)_{t\in\mathbb{Z}_+}\in\mathscr{H}_\Bbbk$, the following holds

$$\sum_{\underline{\tau}\le t\le\overline{\tau}}g_t=\Big\langle\sum_{\underline{\tau}\le t\le\overline{\tau}}\Bbbk(\cdot,t)\mathrm{d}t,\mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk}. \tag{14}$$

Proof. See Appendix A.3. □

Based on Definitions 1, 2, Lemma 4, and Corollary 5, we can present the main theorem of this paper.

Theorem 7 (*Continuity of Convolution Operators*)**.** *Let $\Bbbk$ be an integrable Mercer kernel. Then, for any $\mathrm{u}\in\mathscr{L}^\infty$ and $\tau\in\mathbb{T}$, the operator L_τ is continuous (bounded). Moreover, there exists $\varphi^{(u)}_\tau=(\varphi^{(u)}_{\tau,t})_{t\in\mathbb{T}}\in\mathscr{H}_\Bbbk$ such that $\mathrm{L}_\tau(\mathrm{g})=\langle\varphi^{(u)}_\tau,\mathrm{g}\rangle$, for any $\mathrm{g}\in\mathscr{H}_\Bbbk$. Furthermore, for any $t\in\mathbb{T}$, we have*

$$\varphi^{(u)}_{\tau,t}=\mathrm{L}_\tau(\Bbbk_t)=\begin{cases}\displaystyle\int_{\mathbb{R}_+}\Bbbk(t,s)u_{\tau-s}\mathrm{d}s, & \text{if } \mathbb{T}=\mathbb{R}_+,\\ \displaystyle\sum_{s\in\mathbb{Z}_+}\Bbbk(t,s)u_{\tau-s}, & \text{if } \mathbb{T}=\mathbb{Z}_+.\end{cases} \tag{15}$$

Proof. We discuss the proof for the cases of $\mathbb{T}=\mathbb{R}_+$ and $\mathbb{T}=\mathbb{Z}_+$.

Case I: Let $\mathbb{T}=\mathbb{R}_+$ and define $\mathrm{v}=(v_s)_{s\in\mathbb{R}_+}$ such that $v_s=u_{\tau-s}$, for any $s\in\mathbb{R}_+$. Accordingly, we have

$$\mathrm{L}_\tau(\mathrm{g})=\int_{\mathbb{R}_+}v_sg_s\mathrm{d}s, \tag{16}$$

for each $\mathrm{g}=(g_s)_{s\in\mathbb{R}_+}$. Note that $\mathrm{v}\in\mathscr{L}^\infty$, and hence, in $\mathscr{L}^\infty$, there exists a sequence of step functions $\mathrm{v}^{(n)}:=(v^{(n)}_s)_{s\in\mathbb{R}_+}$, $n=1,2,\ldots$, such that $\|\mathrm{v}^{(n)}\|_\infty\le\|\mathrm{v}\|_\infty$, for each $n\in\mathbb{N}$, and, $\lim_{n\to\infty}v^{(n)}_s=v_s$, for almost all $s\in\mathbb{R}_+$ (Stein & Shakarchi, 2009). For each $n\in\mathbb{N}$, due to the definition of step functions, we know that there exists $M_n\in\mathbb{N}$, intervals $\mathrm{J}^{(n)}_i\subseteq\mathbb{R}_+$, $i=1,\ldots,M_n$, and $a^{(n)}_i\in\mathbb{R}$, $i=1,\ldots,M_n$, such that

$$v^{(n)}_s=\sum_{i=1}^{M_n}a^{(n)}_i\mathbf{1}_{\mathrm{J}^{(n)}_i}(s),\qquad\forall s\in\mathbb{R}_+. \tag{17}$$

For each $n\in\mathbb{N}$, define $\mathrm{f}_n=(f_{n,t})_{t\in\mathbb{R}_+}$ as

$$\begin{aligned}\mathrm{f}_n&:=\int_{\mathbb{R}_+}v^{(n)}_s\Bbbk(\cdot,s)\mathrm{d}s\\&=\int_{\mathbb{R}_+}\sum_{i=1}^{M_n}a^{(n)}_i\mathbf{1}_{\mathrm{J}^{(n)}_i}(s)\Bbbk(\cdot,s)\mathrm{d}s,\end{aligned} \tag{18}$$

which is well-defined and belongs to $\mathscr{H}_\Bbbk$ according to Lemma 4. Accordingly, due to (17), for each $\mathrm{g}=(g_s)_{s\in\mathbb{R}_+}\in\mathscr{H}_\Bbbk$, we have

$$\begin{aligned}\int_{\mathbb{R}_+}g_sv^{(n)}_s\mathrm{d}s&=\sum_{i=1}^{M_n}a^{(n)}_i\int_{\mathrm{J}^{(n)}_i}g_s\mathrm{d}s\\&=\sum_{i=1}^{M_n}a^{(n)}_i\Big\langle\int_{\mathrm{J}^{(n)}_i}\Bbbk(\cdot,s)\mathrm{d}s,\mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk}\\&=\Big\langle\sum_{i=1}^{M_n}a^{(n)}_i\int_{\mathrm{J}^{(n)}_i}\Bbbk(\cdot,s)\mathrm{d}s,\mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk}\\&=\Big\langle\int_{\mathbb{R}_+}\sum_{i=1}^{M_n}a^{(n)}_i\mathbf{1}_{\mathrm{J}^{(n)}_i}(s)\Bbbk(\cdot,s)\mathrm{d}s,\mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk}\\&=\Big\langle\int_{\mathbb{R}_+}v^{(n)}_s\Bbbk(\cdot,s)\mathrm{d}s,\mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk}\\&=\langle\mathrm{f}_n,\mathrm{g}\rangle_{\mathscr{H}_\Bbbk},\end{aligned} \tag{19}$$

where the second equality is due to Lemma 4. Let ε be an arbitrary positive real scalar. Define l as

$$l:=\int_{\mathbb{R}_+\times\mathbb{R}_+}v_t\Bbbk(t,s)v_s\mathrm{d}t\mathrm{d}s, \tag{20}$$

and, for any $n\in\mathbb{N}$, l_n as

$$l_n:=\int_{\mathbb{R}_+\times\mathbb{R}_+}v^{(n)}_t\Bbbk(t,s)v_s\mathrm{d}t\mathrm{d}s. \tag{21}$$

For almost all $s,t\in\mathbb{R}_+$, we have

$$\lim_{n\to\infty}v^{(n)}_t\Bbbk(t,s)v_s=v_t\Bbbk(t,s)v_s. \tag{22}$$

Moreover, for any $n\in\mathbb{N}$, we know that

$$\big|v^{(n)}_t\Bbbk(t,s)v_s\big|\le\|\mathrm{v}\|^2_\infty|\Bbbk(t,s)|. \tag{23}$$

Since $\Bbbk$ is integrable, from the dominated convergence theorem, we have

$$\begin{aligned}\lim_{n\to\infty}l_n&=\lim_{n\to\infty}\int_{\mathbb{R}_+\times\mathbb{R}_+}v^{(n)}_t\Bbbk(t,s)v_s\mathrm{d}s\mathrm{d}t\\&=\int_{\mathbb{R}_+\times\mathbb{R}_+}\lim_{n\to\infty}v^{(n)}_t\Bbbk(t,s)v_s\mathrm{d}s\mathrm{d}t\\&=\int_{\mathbb{R}_+\times\mathbb{R}_+}v_t\Bbbk(t,s)v_s\mathrm{d}s\mathrm{d}t\\&=l.\end{aligned} \tag{24}$$

Therefore, there exist $N_\varepsilon\in\mathbb{N}$ such that $|l_n-l|\le\frac{1}{8}\varepsilon^2$, for each $n\ge N_\varepsilon$. Define $l_{n,m}$ as $l_{n,m}:=\langle\mathrm{f}_n,\mathrm{f}_m\rangle_{\mathscr{H}_\Bbbk}$, for each $m,n\in\mathbb{N}$. Accordingly, from (18), Corollary 5 and the linearity of integration and inner product, it follows that

$$\begin{aligned}l_{n,m}&=\sum_{i=1}^{M_n}\sum_{j=1}^{M_m}a^{(n)}_ia^{(m)}_j\Big\langle\int_{\mathrm{J}^{(n)}_i}\Bbbk(\cdot,s)\mathrm{d}s,\int_{\mathrm{J}^{(m)}_j}\Bbbk(\cdot,t)\mathrm{d}t\Big\rangle_{\mathscr{H}_\Bbbk}\\&=\sum_{i=1}^{M_n}\sum_{j=1}^{M_m}a^{(n)}_ia^{(m)}_j\int_{\mathrm{J}^{(n)}_i}\int_{\mathrm{J}^{(m)}_j}\Bbbk(s,t)\mathrm{d}s\mathrm{d}t\\&=\int_{\mathbb{R}_+\times\mathbb{R}_+}\sum_{i=1}^{M_n}\sum_{j=1}^{M_m}a^{(n)}_ia^{(m)}_j\mathbf{1}_{\mathrm{J}^{(n)}_i}(t)\mathbf{1}_{\mathrm{J}^{(m)}_j}(s)\Bbbk(s,t)\mathrm{d}s\mathrm{d}t.\end{aligned}$$

Therefore, due to the definition of $\mathrm{v}^{(n)}$ and $\mathrm{v}^{(m)}$, we have

$$l_{n,m}=\langle\mathrm{f}_n,\mathrm{f}_m\rangle_{\mathscr{H}_\Bbbk}=\int_{\mathbb{R}_+\times\mathbb{R}_+}v^{(n)}_s\Bbbk(s,t)v^{(m)}_t\mathrm{d}s\mathrm{d}t. \tag{25}$$

Accordingly, since $\|\mathrm{v}^{(n)}\|_\infty,\|\mathrm{v}^{(m)}\|_\infty\le\|\mathrm{v}\|_\infty$, one can see that

$$\begin{aligned}|l_{n,m}-l_n|&=\Big|\int_{\mathbb{R}_+\times\mathbb{R}_+}v^{(n)}_t\Big[\Bbbk(t,s)(v^{(m)}_s-v_s)\Big]\mathrm{d}t\mathrm{d}s\Big|\\&\le\|\mathrm{v}\|_\infty\int_{\mathbb{R}_+\times\mathbb{R}_+}\Big|\Bbbk(t,s)(v^{(m)}_s-v_s)\Big|\mathrm{d}t\mathrm{d}s.\end{aligned} \tag{26}$$

From $\|\mathrm{v}^{(n)}\|_\infty,\|\mathrm{v}^{(m)}\|_\infty\le\|\mathrm{v}\|_\infty$, we have

$$|\Bbbk(t,s)(v^{(m)}_s-v_s)|\le2\|\mathrm{v}\|_\infty|\Bbbk(t,s)|. \tag{27}$$

Moreover, for almost all $s\in\mathbb{R}_+$, we know that

$$\lim_{m\to\infty}|\Bbbk(t,s)(v^{(m)}_s-v_s)|=0. \tag{28}$$

Since $\Bbbk$ is integrable, from the dominated convergence theorem, it follows that

$$\lim_{m\to\infty}\int_{\mathbb{R}_+\times\mathbb{R}_+}|\Bbbk(t,s)(v_s^{(m)}-v_s)|\mathrm{d}s\mathrm{d}t=0. \tag{29}$$

Therefore, due to (26), there exists M_ε such that, for any $m\geq M_\varepsilon$, we have $|l_{n,m}-l_n|\leq\frac{1}{8}\varepsilon^2$. Accordingly, from triangle inequality, we have $|l_{m,n}-l|\leq\frac{1}{4}\varepsilon^2$, for any $m,n\geq K_\varepsilon:=\max\{M_\varepsilon,N_\varepsilon\}$. Subsequently, it follows that

$$\begin{aligned}\|\,\mathrm{f}_n-\mathrm{f}_m\,\|^2&=\langle\mathrm{f}_n,\mathrm{f}_n\rangle_{\mathscr{H}_\Bbbk}-2\langle\mathrm{f}_n,\mathrm{f}_m\rangle_{\mathscr{H}_\Bbbk}+\langle\mathrm{f}_m,\mathrm{f}_m\rangle_{\mathscr{H}_\Bbbk}\\&=l_{n,n}-2l_{n,m}+l_{m,m}\\&\leq(l+\frac{1}{4}\varepsilon^2)-2(l-\frac{1}{4}\varepsilon^2)+(l+\frac{1}{4}\varepsilon^2)=\varepsilon^2.\end{aligned} \tag{30}$$

Hence, for any $m,n\geq K_\varepsilon$, we have $\|\mathrm{f}_n-\mathrm{f}_m\|_{\mathscr{H}_\Bbbk}\leq\varepsilon$. Therefore, $\{\mathrm{f}_n\}_{n=1}^\infty$ is a Cauchy sequence in $\mathscr{H}_\Bbbk$, and there exists $\mathrm{f}=(f_s)_{s\in\mathbb{R}_+}\in\mathscr{H}_\Bbbk$ such that $\lim_{n\to\infty}\mathrm{f}_n=\mathrm{f}$. Accordingly, due to the reproducing property, for any $t\in\mathbb{R}_+$, we have

$$\lim_{n\to\infty}f_{n,t}=\lim_{n\to\infty}\langle\Bbbk_t,\mathrm{f}_n\rangle_{\mathscr{H}_\Bbbk}=\langle\Bbbk_t,\mathrm{f}\rangle_{\mathscr{H}_\Bbbk}=f_t. \tag{31}$$

For any $n\in\mathbb{N}$ and for almost all $s,t\in\mathbb{R}_+$, we have

$$\left|\Bbbk(t,s)v_s^{(n)}\right|\leq\|\mathrm{v}\|_\infty|\Bbbk(t,s)|, \tag{32}$$

and

$$\lim_{n\to\infty}\Bbbk(t,s)v_s^{(n)}=\Bbbk(t,s)v_s. \tag{33}$$

Accordingly, since $\Bbbk$ is integrable, from the dominated convergence theorem, (18) and (31), it follows that

$$\begin{aligned}f_t&=\lim_{n\to\infty}\int_{\mathbb{R}_+}\Bbbk(t,s)v_s^{(n)}\mathrm{d}s\\&=\int_{\mathbb{R}_+}\lim_{n\to\infty}\Bbbk(t,s)v_s^{(n)}\mathrm{d}s=\int_{\mathbb{R}_+}\Bbbk(t,s)v_s\mathrm{d}s,\end{aligned} \tag{34}$$

i.e., we have

$$\mathrm{f}=\int_{\mathbb{R}_+}\Bbbk(\cdot,s)v_s\mathrm{d}s. \tag{35}$$

For almost all $s\in\mathbb{R}_+$, we know that $\lim_{n\to\infty}g_sv_s^{(n)}=g_sv_s$. Moreover, one has that $|g_sv_s^{(n)}|\leq\|\mathrm{v}\|_\infty|g_s|$, for each $n\in\mathbb{N}$. Since $\mathrm{g}=(g_s)_{s\in\mathbb{R}_+}\in\mathscr{H}_\Bbbk$ and each element of $\mathscr{H}_\Bbbk$ is integrable, due to the dominated convergence theorem, (19) and $\lim_{n\to\infty}\mathrm{f}_n=\mathrm{f}$, we have

$$\begin{aligned}\int_{\mathbb{R}_+}g_sv_s\mathrm{d}s&=\lim_{n\to\infty}\int_{\mathbb{R}_+}g_sv_s^{(n)}\mathrm{d}s\\&=\lim_{n\to\infty}\langle\mathrm{f}_n,\mathrm{g}\rangle_{\mathscr{H}_\Bbbk}=\langle\mathrm{f},\mathrm{g}\rangle_{\mathscr{H}_\Bbbk}.\end{aligned} \tag{36}$$

Let $\varphi_\tau^{(\mathrm{u})}=(\varphi_{\tau,t}^{(\mathrm{u})})_{t\in\mathbb{R}_+}$ be defined such that for any $t\in\mathbb{R}_+$, we have

$$\varphi_{\tau,t}^{(\mathrm{u})}=\int_{\mathbb{R}_+}\Bbbk(t,s)u_{\tau-s}\mathrm{d}s, \tag{37}$$

i.e., $\varphi_\tau^{(\mathrm{u})}=\int_{\mathbb{R}_+}\Bbbk(\cdot,s)u_{\tau-s}\mathrm{d}s$. Due to (34) and the fact that $v_s=u_{\tau-s}$, for $s\in\mathbb{R}_+$, we know that $\varphi_\tau^{(\mathrm{u})}=\mathrm{f}\in\mathscr{H}_\Bbbk$. Accordingly, from (36), we have

$$\begin{aligned}\mathrm{L}_\tau(\mathrm{g})&=\int_{\mathbb{R}_+}g_su_{\tau-s}\mathrm{d}s\\&=\left\langle\int_{\mathbb{R}_+}\Bbbk(\cdot,s)u_{\tau-s}\mathrm{d}s,\mathrm{g}\right\rangle_{\mathscr{H}_\Bbbk}\\&=\langle\varphi_\tau^{(\mathrm{u})},\mathrm{g}\rangle_{\mathscr{H}_\Bbbk},\end{aligned} \tag{38}$$

which implies that L_τ is a continuous (bounded) operator on $\mathscr{H}_\Bbbk$. This concludes the proof for the case of $\mathbb{T}=\mathbb{R}_+$.

Case II: Let $\mathbb{T}=\mathbb{Z}_+$ and, similarly to the previous case, define $\mathrm{v}=(v_s)_{s\in\mathbb{Z}_+}$ as $v_s=u_{\tau-s}$, for any $s\in\mathbb{Z}_+$. One can easily see that $\|\mathrm{v}\|_\infty=\|\mathrm{u}\|_\infty$. For any $\mathrm{g}=(g_s)_{s\in\mathbb{Z}_+}\in\mathscr{H}_\Bbbk$, we know that $\sum_{s\in\mathbb{Z}_+}|g_s|<\infty$, which implies that $\mathrm{L}_\tau(\mathrm{g})=\sum_{s\in\mathbb{Z}_+}g_sv_s$ is absolutely convergent due to $\|\mathrm{v}\|_\infty=\|\mathrm{u}\|_\infty<\infty$. Let ε be an arbitrary positive real scalar. Since $\Bbbk$ is summable, there exists $N_\varepsilon\in\mathbb{N}$ such that

$$\sum_{s,t\geq N_\varepsilon+1}|\Bbbk(t,s)|\leq\frac{1}{\|\mathrm{v}\|_\infty^2}\varepsilon^2. \tag{39}$$

For any $n\in\mathbb{N}$, let $\mathrm{f}_n=(f_{n,t})_{t\in\mathbb{Z}_+}$ be defined as $\mathrm{f}_n=\sum_{s=0}^n\Bbbk(\cdot,s)v_s$. One can see that $\mathrm{f}_n\in\mathscr{H}_\Bbbk$. Let $n,m\in\mathbb{N}$ such that $n,m\geq N_\varepsilon$. Without loss of generality, assume $n\geq m$. Due to the reproducing property, we have

$$\begin{aligned}\|\mathrm{f}_n-\mathrm{f}_m\|_{\mathscr{H}_\Bbbk}^2&=\Big\|\sum_{s=n+1}^m\Bbbk(\cdot,s)v_s\Big\|_{\mathscr{H}_\Bbbk}^2\\&=\sum_{s,t\geq n+1}^m\Bbbk(s,t)v_sv_t\\&\leq\|\mathrm{v}\|_\infty^2\sum_{s,t\geq N_\varepsilon+1}|\Bbbk(s,t)|\\&\leq\varepsilon^2,\end{aligned} \tag{40}$$

i.e., $\|\mathrm{f}_n-\mathrm{f}_m\|_{\mathscr{H}_\Bbbk}\leq\varepsilon$. Therefore $\{\mathrm{f}_n\}_{n=1}^\infty$ is a Cauchy sequence in $\mathscr{H}_\Bbbk$, and there exists $\mathrm{f}=(f_t)_{t\in\mathbb{Z}_+}$ such that $\lim_{n\to\infty}\mathrm{f}_n=\mathrm{f}$. Note that, we have

$$f_t=\langle\Bbbk_t,\mathrm{f}\rangle_{\mathscr{H}_\Bbbk}=\lim_{n\to\infty}\langle\Bbbk_t,\mathrm{f}_n\rangle_{\mathscr{H}_\Bbbk}=\lim_{n\to\infty}f_{n,t}, \tag{41}$$

for any $t\in\mathbb{Z}_+$. Accordingly, from the reproducing property, one can see that

$$\begin{aligned}f_t&=\lim_{n\to\infty}\Big\langle\Bbbk_t,\sum_{s=0}^n\Bbbk(\cdot,s)v_s\Big\rangle_{\mathscr{H}_\Bbbk}\\&=\lim_{n\to\infty}\sum_{s=0}^n\Bbbk(t,s)v_s\\&=\sum_{s=0}^\infty\Bbbk(t,s)v_s,\end{aligned} \tag{42}$$

where the last equality is due to $\sum_{s=0}^\infty|\Bbbk(t,s)v_s|<\infty$, for any $t\in\mathbb{Z}_+$. Hence, we have $\mathrm{f}=\sum_{s=0}^\infty\Bbbk(\cdot,s)v_s$. For any $\mathrm{g}=(g_s)_{s\in\mathbb{Z}_+}\in\mathscr{H}_\Bbbk$, we know that $\sum_{s\in\mathbb{Z}_+}|g_s|\leq\infty$, which implies that $\mathrm{L}_\tau(\mathrm{g})=\sum_{s\in\mathbb{Z}_+}g_sv_s$ is absolutely convergent due to $\|\mathrm{v}\|_\infty=\|\mathrm{u}\|_\infty<\infty$. Therefore, one can see that

$$\begin{aligned}\sum_{s\in\mathbb{Z}_+}g_sv_s&=\lim_{n\to\infty}\sum_{s=0}^nv_s\langle\Bbbk_s,\mathrm{g}\rangle_{\mathscr{H}_\Bbbk}\\&=\lim_{n\to\infty}\Big\langle\sum_{s=0}^n\Bbbk(\cdot,s)v_s,\mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk}\\&=\lim_{n\to\infty}\langle\mathrm{f}_n,\mathrm{g}\rangle_{\mathscr{H}_\Bbbk}\\&=\langle\mathrm{f},\mathrm{g}\rangle_{\mathscr{H}_\Bbbk}.\end{aligned} \tag{43}$$

Let $\varphi_\tau^{(\mathrm{u})}=(\varphi_{\tau,t}^{(\mathrm{u})})_{t\in\mathbb{Z}_+}$ be defined as

$$\varphi_{\tau,t}^{(\mathrm{u})}=f_t=\sum_{s\in\mathbb{Z}_+}\Bbbk(t,s)v_s, \tag{44}$$

for any $t\in\mathbb{Z}_+$. Accordingly, we have

$$\mathrm{L}_\tau(\mathrm{g})=\sum_{s\in\mathbb{Z}_+}g_sv_s=\langle\varphi_\tau^{(\mathrm{u})},\mathrm{g}\rangle_{\mathscr{H}_\Bbbk}=\Big\langle\sum_{s\in\mathbb{Z}_+}\Bbbk(\cdot,s)v_s,\mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk},$$

for any $\mathrm{g}\in\mathscr{H}_\Bbbk$. This concludes the proof. □

Remark 2. For the case of $\mathbb{T} = \mathbb{R}_+$, note that the assumption on continuity of kernel $\Bbbk$ is required for the result presented in Theorem 7. This can be verified by inspecting the proof of Lemma 4 provided in Appendix A.1, which is the base of Theorem 7.

From Theorem 7, we have the following corollary.

Corollary 8. *Let $\Bbbk$ be an integrable Mercer kernel and $\mathrm{u} \in \mathscr{L}^\infty$. Then, the kernel-based impulse response estimation problem (5) admits a unique solution introduced in (6).*

Proof. From Theorem 7, it follows that the objective in (5) is function $\mathcal{J} : \mathscr{H}_\Bbbk \to \mathbb{R}$ defined as

$$\mathcal{J}(\mathrm{g}) = \sum_{i=1}^{n_\mathscr{D}} \left(\langle \varphi_{t_i}^{(\mathrm{u})}, \mathrm{g} \rangle_{\mathscr{H}_\Bbbk} - y_{t_i} \right)^2 + \lambda \|\mathrm{g}\|^2_{\mathscr{H}_\Bbbk}, \tag{45}$$

for any $\mathrm{g} \in \mathscr{H}_\Bbbk$. This implies that $\mathcal{J}$ is a quadratic continuous function. Since $\lambda > 0$, we know that $\mathcal{J}$ is strongly convex, and subsequently, coercive. Accordingly, from $\mathcal{J}(\mathbf{0}) = \|\mathrm{y}\|^2$, it follows that $\mathcal{J}$ is a proper continuous strongly convex function. Therefore, due to Peypouquet (2015, Theorem 2.19), we know that $\min_{\mathrm{g}\in\mathscr{H}_\Bbbk} \mathcal{J}(\mathrm{g})$ has a unique solution, which implies the existence and uniqueness for the solution of (5). The proof concludes from (15) and Wahba (1990, Theorem 1.3.1). $\square$

Remark 3. From Corollary 8, one can see that the presented result on the existence and uniqueness of the solution for the kernel-based impulse response estimation problem (5) depends only on the input signal and the kernel. The impact of other factors like measurement noise is on the exactness and statistical behavior of this solution.

5. Conclusion

The kernel-based system identification method stands on the central assumption that the convolution operators restricted to the chosen RKHS are continuous linear functionals. Current research work in the literature assumes, implicitly or explicitly, that this continuity property holds without elaborating the required conditions. In this work, we have addressed this long-standing question by specifying these conditions: the integrability of the kernel function and the boundedness of the input signal. For the case of continuous-time dynamics, we additionally need the continuity of the kernel. Furthermore, owing to the strong convexity of the optimization problem and the resulted continuity feature of the convolution operators, we have shown that the kernel-based approach is well-defined by guaranteeing the existence and uniqueness properties for the solution of the identification problem.

Appendix

A.1. Proof of Lemma 4

First we show the claims when $\overline{\tau} < \infty$, and then, extend the result to the general case.

Let $\Delta\tau$ be defined as $\Delta\tau = \overline{\tau} - \underline{\tau}$. For each $n \in \mathbb{N}$, define N_n and Δ_n respectively as $N_n := 2^n$ and $\Delta_n := 2^{-n}\Delta\tau$. Also, let function $\mathrm{f}_n := (f_{n,s})_{s\in\mathbb{R}_+} \in \mathscr{H}_\Bbbk$ be defined as

$$f_{n,s} = \Delta_n \sum_{i=1}^{N_n} \Bbbk(t_i^{(n)}, s), \qquad \forall s \in \mathbb{R}_+, \tag{A.1}$$

where $t_i^{(n)} := \underline{\tau} + (i-1)\Delta_n = \underline{\tau} + (i-1)2^{-n}\Delta\tau$, for $i = 1, \ldots, 2^n$. Let ε be an arbitrary positive real scalar. Since $\Bbbk$ is continuous, we know that it is uniformly continuous on compact region $[\underline{\tau}, \overline{\tau}] \times [\underline{\tau}, \overline{\tau}]$. Therefore, there exists positive real scalar δ such that for any $(s_1, t_1), (s_2, t_2) \in [\underline{\tau}, \overline{\tau}] \times [\underline{\tau}, \overline{\tau}]$ where $|s_1 - s_2| + |t_1 - t_2| \leq \delta$, we have

$$|\Bbbk(s_1, t_1) - \Bbbk(s_2, t_2)| \leq \frac{1}{4}\frac{\varepsilon^2}{\Delta\tau^2}. \tag{A.2}$$

Let δ_ε be the largest scalar in $(0, 1)$ with such property and define N_ε as the smallest integer such that

$$N_\varepsilon \geq \max(-\log_2(\frac{\delta_\varepsilon}{\Delta\tau}), 0) + 1. \tag{A.3}$$

Consider arbitrary integers n and m such that $n, m \geq N_\varepsilon$. From the reproducing property of the kernel, one can see that

$$\begin{aligned} \langle \mathrm{f}_n, \mathrm{f}_m \rangle_{\mathscr{H}_\Bbbk} &= \Big\langle \Delta_n \sum_{i=1}^{N_n} \Bbbk(t_i^{(n)}, \cdot), \Delta_m \sum_{j=1}^{N_m} \Bbbk(t_j^{(m)}, \cdot) \Big\rangle_{\mathscr{H}_\Bbbk} \\ &= \Delta_n \Delta_m \sum_{i=1}^{N_n} \sum_{j=1}^{N_m} \Bbbk(t_i^{(n)}, t_j^{(m)}). \end{aligned} \tag{A.4}$$

Define $\mathrm{I}_{i,j}^{(n,m)}$ as region $[t_i^{(n)}, t_i^{(n)} + \Delta_n) \times [t_j^{(m)}, t_j^{(m)} + \Delta_m)$, for $i = 1, \ldots, 2^n$ and $j = 1, \ldots, 2^m$. Also, let I be the value defined as

$$I := \int_{[\underline{\tau},\overline{\tau}]\times[\underline{\tau},\overline{\tau}]} \Bbbk(s, t)\mathrm{d}s\mathrm{d}t. \tag{A.5}$$

Note that I is a well-defined integral due to integrability of $\Bbbk$. From (A.4) and the triangle inequality, we have

$$\begin{aligned} &|\langle \mathrm{f}_n, \mathrm{f}_m \rangle_{\mathscr{H}_\Bbbk} - I| \\ &\quad = \left| \sum_{i=1}^{N_n} \sum_{j=1}^{N_m} \int_{\mathrm{I}_{i,j}^{(n,m)}} \Bbbk(t_i^{(n)}, t_j^{(m)}) - \Bbbk(s, t)\mathrm{d}s\mathrm{d}t \right| \\ &\quad \leq \sum_{i=1}^{N_n} \sum_{j=1}^{N_m} \int_{\mathrm{I}_{i,j}^{(n,m)}} \left|\Bbbk(t_i^{(n)}, t_j^{(m)}) - \Bbbk(s, t)\right|\mathrm{d}s\mathrm{d}t \\ &\quad \leq \sum_{i=1}^{N_n} \sum_{j=1}^{N_m} \frac{1}{4}\frac{\varepsilon^2}{\Delta\tau^2}\Delta_n\Delta_m \\ &\quad = \frac{1}{4}\varepsilon^2, \end{aligned} \tag{A.6}$$

where the inequality is due to (A.2). Subsequently, one can see that

$$I - \frac{1}{4}\varepsilon^2 \leq \langle \mathrm{f}_n, \mathrm{f}_m \rangle_{\mathscr{H}_\Bbbk} \leq I + \frac{1}{4}\varepsilon^2. \tag{A.7}$$

From (A.7), it follows that

$$\begin{aligned} \|\mathrm{f}_n - \mathrm{f}_m\|^2_{\mathscr{H}_\Bbbk} &= \langle \mathrm{f}_n, \mathrm{f}_n \rangle_{\mathscr{H}_\Bbbk} - 2\langle \mathrm{f}_n, \mathrm{f}_m \rangle_{\mathscr{H}_\Bbbk} + \langle \mathrm{f}_m, \mathrm{f}_m \rangle_{\mathscr{H}_\Bbbk} \\ &\leq (I + \frac{1}{4}\varepsilon^2) - 2(I - \frac{1}{4}\varepsilon^2) + (I + \frac{1}{4}\varepsilon^2) \\ &= \varepsilon^2, \end{aligned}$$

and, hence, we have $\|\mathrm{f}_n - \mathrm{f}_m\|_{\mathscr{H}_\Bbbk} \leq \varepsilon$. Therefore, $\{\mathrm{f}_n\}_{n=1}^\infty$ is a Cauchy sequence in $\mathscr{H}_\Bbbk$, and there exists $\mathrm{f} = (f_s)_{s\in\mathbb{R}_+} \in \mathscr{H}_\Bbbk$ such that $\lim_{n\to\infty} \|\mathrm{f}_n - \mathrm{f}\|_{\mathscr{H}_\Bbbk} = 0$. For any $s \in \mathbb{R}_+$, due to the Cauchy–Schwarz inequality and the reproducing property, we have

$$\begin{aligned} |f_{n,s} - f_s| &= |\langle \Bbbk_s, \mathrm{f}_n - \mathrm{f} \rangle_{\mathscr{H}_\Bbbk}| \\ &\leq \Bbbk(s, s)^{\frac{1}{2}} \|\mathrm{f}_n - \mathrm{f}\|_{\mathscr{H}_\Bbbk}, \end{aligned} \tag{A.8}$$

which implies that $\lim_{n\to\infty} f_{n,s} = f_s$. On the other hand, from (A.2), one can see that

$$\begin{aligned}\left|f_{n,s} - \int_{[\underline{\tau},\overline{\tau}]} \Bbbk(s,t)\mathrm{d}t\right| &= \left|\Delta_n \sum_{i=1}^{N_n} \Bbbk(t_i^{(n)}, s) - \int_{\underline{\tau}}^{\overline{\tau}} \Bbbk(s,t)\mathrm{d}t\right| \\ &= \left|\sum_{i=1}^{N_n} \int_{\mathrm{I}_i^{(n)}} \Bbbk(t_i^{(n)}, s) - \Bbbk(s,t)\mathrm{d}t\right| \\ &\leq \sum_{i=1}^{N_n} \int_{\mathrm{I}_i^{(n)}} \left|\Bbbk(t_i^{(n)}, s) - \Bbbk(s,t)\right|\mathrm{d}t \\ &\leq \sum_{i=1}^{N_n} \frac{1}{4}\frac{\varepsilon^2}{\Delta\tau^2}\Delta_n \\ &= \frac{1}{4}\frac{\varepsilon^2}{\Delta\tau},\end{aligned} \tag{A.9}$$

where, for $i = 1, \ldots, N_n$, interval $\mathrm{I}_i^{(n)}$ is defined as $[t_i^{(n)}, t_i^{(n)} + \Delta_n)$. Accordingly, we have

$$f_s = \lim_{n\to\infty} f_{n,s} = \int_{[\underline{\tau},\overline{\tau}]} \Bbbk(s,t)\mathrm{d}t, \tag{A.10}$$

which says that $\mathrm{f} = \int_{[\underline{\tau},\overline{\tau}]} \Bbbk(\cdot,t)\mathrm{d}t \in \mathscr{H}_\Bbbk$. Moreover, from $\lim_{n\to\infty} \mathrm{f}_n = \mathrm{f}$, we know that

$$\begin{aligned}\|\mathrm{f}\|^2_{\mathscr{H}_\Bbbk} &= \lim_{n\to\infty} \|\mathrm{f}_n\|^2_{\mathscr{H}_\Bbbk} \\ &= \lim_{n\to\infty} \langle \mathrm{f}_n, \mathrm{f}_n\rangle_{\mathscr{H}_\Bbbk}.\end{aligned} \tag{A.11}$$

Therefore, from (A.7) and the definition of f and I, it follows that

$$\begin{aligned}\left\| \int_{[\underline{\tau},\overline{\tau}]} \Bbbk(\cdot,t)\mathrm{d}t\right\|^2_{\mathscr{H}_\Bbbk} &= \lim_{n\to\infty} \langle \mathrm{f}_n, \mathrm{f}_n\rangle_{\mathscr{H}_\Bbbk} \\ &= \int_{[\underline{\tau},\overline{\tau}]\times[\underline{\tau},\overline{\tau}]} \Bbbk(s,t)\mathrm{d}s\mathrm{d}t.\end{aligned} \tag{A.12}$$

Note that, for any s such that $t + s \in \mathbb{R}_+$, due to the reproducing property of the kernel, we have

$$\begin{aligned}\|\Bbbk_{t+s} - \Bbbk_t\|^2_{\mathscr{H}_\Bbbk} &= \Bbbk(t+s, t+s) \\ &\quad - \Bbbk(t, t+s) - \Bbbk(t+s, t) + \Bbbk(t,t),\end{aligned} \tag{A.13}$$

which implies that $\lim_{s\to 0} \|\Bbbk_{t+s} - \Bbbk_t\|_{\mathscr{H}_\Bbbk} = 0$ due to continuity of the kernel. Meanwhile, for each $\mathrm{g} = (g_t)_{t\in\mathbb{R}_+} \in \mathscr{H}_\Bbbk$, from the Cauchy-Schwartz inequality and the reproducing property, one has

$$\begin{aligned}|g_{t+s} - g_t| &= |\langle \Bbbk_{t+s} - \Bbbk_t, \mathrm{g}\rangle_{\mathscr{H}_\Bbbk}| \\ &\leq \|\Bbbk_{t+s} - \Bbbk_t\|_{\mathscr{H}_\Bbbk} \|\mathrm{g}\|_{\mathscr{H}_\Bbbk}.\end{aligned} \tag{A.14}$$

Accordingly, we have $\lim_{s\to 0} |g_{t+s} - g_t| = 0$, which says that $\mathrm{g} = (g_t)_{t\in\mathbb{R}_+}$ is a continuous function of t. Hence, the integral of g on the interval $[\underline{\tau}, \overline{\tau}]$ exists, and we have

$$\begin{aligned}\int_{[\underline{\tau},\overline{\tau}]} g_t\mathrm{d}t &= \lim_{n\to\infty} \sum_{i=1}^{2^n} g\big(\underline{\tau} + (i-1)2^{-n}\Delta\tau\big)2^{-n}\Delta\tau \\ &= \lim_{n\to\infty} \sum_{i=1}^{N_n} g(t_i^{(n)})\Delta_n \\ &= \lim_{n\to\infty} \langle \mathrm{f}_n, \mathrm{g}\rangle_{\mathscr{H}_\Bbbk},\end{aligned} \tag{A.15}$$

where the last equality is due the definition of f_n in (A.1) and the reproducing property. Therefore, from $\lim_{n\to\infty} \mathrm{f}_n = \mathrm{f}$, one can see that

$$\begin{aligned}\int_{[\underline{\tau},\overline{\tau}]} g_t\mathrm{d}t &= \lim_{n\to\infty} \langle \mathrm{f}_n, \mathrm{g}\rangle_{\mathscr{H}_\Bbbk} \\ &= \langle \mathrm{f}, \mathrm{g}\rangle_{\mathscr{H}_\Bbbk} \\ &= \left\langle \int_{[\underline{\tau},\overline{\tau}]} \Bbbk(\cdot,t)\mathrm{d}t, \mathrm{g}\right\rangle_{\mathscr{H}_\Bbbk}.\end{aligned} \tag{A.16}$$

Now, we consider the case where $\overline{\tau} = \infty$. For each integer $n \geq \underline{\tau}$, let $\mathrm{h}_n = (h_{n,s})_{s\in\mathbb{R}_+}$ be function $\mathrm{h}_n = \int_{[\underline{\tau},n]} \Bbbk(\cdot,t)\mathrm{d}t$ which is well-defined and belongs to $\mathscr{H}_\Bbbk$. Let ε be an arbitrary positive real scalar. Since $\Bbbk$ is absolutely integrable, we know that

$$\lim_{\tau\to\infty} \int_\tau^\infty \int_\tau^\infty |\Bbbk(s,t)|\mathrm{d}s\mathrm{d}t = 0. \tag{A.17}$$

Let $\tau_\varepsilon \geq \underline{\tau}$ be the smallest positive real scalar such that

$$\int_{\tau_\varepsilon}^\infty \int_{\tau_\varepsilon}^\infty |\Bbbk(s,t)|\mathrm{d}s\mathrm{d}t \leq \varepsilon^2, \tag{A.18}$$

and $n, m \in \mathbb{N}$ be arbitrary indices such that $n, m \geq \tau_\varepsilon$. Without loss of generality, assume $n \geq m$. Then, due to the discussion above and the triangle inequality, we have

$$\begin{aligned}\|\,\mathrm{h}_n - \mathrm{h}_m\,\|^2_{\mathscr{H}_\Bbbk} &= \left\| \int_m^n \Bbbk(\cdot,t)\mathrm{d}t\right\|^2_{\mathscr{H}_\Bbbk} \\ &= \int_m^n \int_m^n \Bbbk(s,t)\mathrm{d}s\mathrm{d}t \\ &\leq \int_m^n \int_m^n |\Bbbk(s,t)|\mathrm{d}s\mathrm{d}t \\ &\leq \int_{\tau_\varepsilon}^\infty \int_{\tau_\varepsilon}^\infty |\Bbbk(s,t)|\mathrm{d}s\mathrm{d}t.\end{aligned} \tag{A.19}$$

Accordingly, from (A.18), we know that $\|\mathrm{h}_n - \mathrm{h}_m\|_{\mathscr{H}_\Bbbk} \leq \varepsilon$, which implies that $\{\mathrm{h}_n\}_{n\in\mathbb{N}, n\geq\underline{\tau}}$ is a Cauchy sequence in $\mathscr{H}_\Bbbk$. Therefore, there exists $\mathrm{h} = (h_s)_{s\in\mathbb{R}_+} \in \mathscr{H}_\Bbbk$ such that $\lim_{n\to\infty} \mathrm{h}_n = \mathrm{h}$. Based on an argument similar to (A.8), one can show that $\lim_{n\to\infty} h_{n,s} = h_s$, for any $s \in \mathbb{R}_+$. Since $\Bbbk(s,\cdot) \in \mathscr{H}_\Bbbk$ and the elements of $\mathscr{H}_\Bbbk$ are integrable, due to the dominated convergence theorem, we have

$$\begin{aligned}h_s &= \lim_{n\to\infty} \int_{\underline{\tau}}^n \Bbbk(s,t)\mathrm{d}t \\ &= \lim_{n\to\infty} \int_{\underline{\tau}}^\infty \Bbbk(s,t)\mathbf{1}_{[\underline{\tau},n]}(t)\mathrm{d}t \\ &= \int_{\underline{\tau}}^\infty \Bbbk(s,t)\mathrm{d}t.\end{aligned} \tag{A.20}$$

In other words, one has $\mathrm{h} = \int_{\underline{\tau}}^\infty \Bbbk(\cdot,t)\mathrm{d}t$. Hence, from $\lim_{n\to\infty} \mathrm{h}_n = \mathrm{h}$ and the above discussion, we have

$$\begin{aligned}\left\| \int_{\underline{\tau}}^\infty \Bbbk(\cdot,t)\mathrm{d}t\right\|^2_{\mathscr{H}_\Bbbk} &= \lim_{n\to\infty} \|\mathrm{h}_n\|^2_{\mathscr{H}_\Bbbk} \\ &= \lim_{n\to\infty} \int_{\underline{\tau}}^n \int_{\underline{\tau}}^n \Bbbk(s,t)\mathrm{d}s\mathrm{d}t \\ &= \lim_{n\to\infty} \int_{\underline{\tau}}^\infty \int_{\underline{\tau}}^\infty \Bbbk(s,t)\mathbf{1}_{[\underline{\tau},n]^2}(s,t)\mathrm{d}s\mathrm{d}t \\ &= \int_{\underline{\tau}}^\infty \int_{\underline{\tau}}^\infty \Bbbk(s,t)\mathrm{d}s\mathrm{d}t,\end{aligned} \tag{A.21}$$

where the last equality is according to the dominated convergence theorem. Let $\mathrm{g} = (g_t)_{t\in\mathbb{R}_+}$ be an arbitrary element of $\mathscr{H}_\Bbbk$.

Based on same arguments as before, one can see that

$$\begin{aligned}\Big\langle \int_{\underline{\tau}}^{\infty} \Bbbk(\cdot,t)\mathrm{d}t, \mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk} &= \lim_{n\to\infty} \langle \mathrm{h}_n, \mathrm{g}\rangle_{\mathscr{H}_\Bbbk} \\ &= \lim_{n\to\infty}\Big\langle \int_{\underline{\tau}}^{n} \Bbbk(\cdot,t)\mathrm{d}t, \mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk} \\ &= \lim_{n\to\infty}\int_{\underline{\tau}}^{n} g_t\mathrm{d}t \\ &= \lim_{n\to\infty}\int_{\underline{\tau}}^{\infty} g_t\mathbf{1}_{[\underline{\tau},n]}(t)\mathrm{d}t \\ &= \int_{\underline{\tau}}^{\infty} g_t\mathrm{d}t,\end{aligned} \tag{A.22}$$

where the last equality is due to the dominated convergence theorem and the fact that g is integrable. □

A.2. Proof of Corollary 5

In (11), set $\underline{\tau}$, $\overline{\tau}$, and g respectively to $\underline{\tau}_1$, $\overline{\tau}_1$, $\mathrm{g} = \int_{[\underline{\tau}_2,\overline{\tau}_2]} \Bbbk(\cdot,s)\mathrm{d}s$. Accordingly, we have we

$$\Big\langle \int_{\underline{\tau}_1}^{\overline{\tau}_1} \Bbbk(\cdot,t)\mathrm{d}t, \int_{\underline{\tau}_2}^{\overline{\tau}_2} \Bbbk(\cdot,s)\mathrm{d}s\Big\rangle_{\mathscr{H}_\Bbbk} = \int_{\underline{\tau}_1}^{\overline{\tau}_1}\int_{\underline{\tau}_2}^{\overline{\tau}_2} \Bbbk(t,s)\mathrm{d}s\mathrm{d}t. \tag{A.23}$$

In this equation, since the kernel is integrable, the right-hand side is well-defined. Moreover, due to the Fubini theorem (Stein & Shakarchi, 2009), it equals to the other integrals in (12). □

A.3. Proof of Lemma 6

We know that $\Bbbk(\cdot,t) \in \mathscr{H}_\Bbbk$, for each $t = \underline{\tau},\ldots,\overline{\tau}$. If $\overline{\tau}$ is finite, one can see that $\sum_{\underline{\tau}\le t\le\overline{\tau}} \Bbbk(\cdot,t)$ belongs to $\mathscr{H}_\Bbbk$, hence, it is well-defined. Moreover, using the definition of norm and the reproducing property, one can show (13). Similarly, (14) is concluded from the reproducing property. Now, we consider the case $\overline{\tau} = \infty$. For $n \ge \underline{\tau}$, we define $\mathrm{f}_n \in \mathscr{H}_\Bbbk$ as $\mathrm{f}_n = (f_{n,s})_{s\in\mathbb{Z}_+} := \sum_{\underline{\tau}\le t\le n} \Bbbk(\cdot,t)$. Let ε be an arbitrary positive real scalar, and N_ε be the smallest non-negative integer such that $\sum_{s,t\ge N_\varepsilon} |\Bbbk(s,t)| \le \varepsilon$. Note that since $\Bbbk$ is integrable, there exist such N_ε for any positive ε. Now, let $n,m \in \mathbb{N}$ such that $n,m \ge N_\varepsilon$ and without loss of generality, we assume $n \ge m$. Based on the previous case, we know that

$$\begin{aligned}\|\mathrm{f}_n - \mathrm{f}_m\|_{\mathscr{H}_\Bbbk}^2 &= \Big\| \sum_{t=m+1}^{n} \Bbbk(\cdot,t)\Big\|_{\mathscr{H}_\Bbbk}^2 \\ &= \sum_{m+1\le s,t\le n} \Bbbk(s,t) \\ &\le \sum_{N_\varepsilon\le s,t} |\Bbbk(s,t)| \le \varepsilon^2.\end{aligned} \tag{A.24}$$

Accordingly, we have $\|\mathrm{f}_n - \mathrm{f}_m\|_{\mathscr{H}_\Bbbk} \le \varepsilon$, which implies that $\{\mathrm{f}_n\}_{n\ge\underline{\tau}}$ is a Cauchy sequence and hence convergent. Let $\mathrm{f} = (f_s)_{s\in\mathbb{Z}_+}$ denote the limit of this sequence. For any $s \in \mathbb{Z}_+$, we know that

$$|f_s - f_{n,s}| = |\langle \mathrm{f} - \mathrm{f}_n, \Bbbk_s\rangle_{\mathscr{H}_\Bbbk}| \le \Bbbk(s,s)^{\frac{1}{2}}\|\mathrm{f} - \mathrm{f}_n\|_{\mathscr{H}_\Bbbk}, \tag{A.25}$$

and consequently, we have $\lim_{n\to\infty} f_{n,s} = f_s$. Subsequently, since $\Bbbk(s,\cdot)$ is absolutely integrable, it follows that

$$f_s = \lim_{n\to\infty} \sum_{\underline{\tau}\le t\le n} \Bbbk(s,t) = \sum_{\underline{\tau}\le t} \Bbbk(s,t), \tag{A.26}$$

i.e., $\mathrm{f} = \sum_{\underline{\tau}\le t} \Bbbk(\cdot,t)$. Moreover, we have

$$\begin{aligned}\Big\| \sum_{\underline{\tau}\le t} \Bbbk(\cdot,t)\Big\|_{\mathscr{H}_\Bbbk}^2 &= \lim_{n\to\infty} \|\mathrm{f}_n\|_{\mathscr{H}_\Bbbk}^2 \\ &= \lim_{n\to\infty} \sum_{\underline{\tau}\le s,t\le n} \Bbbk(s,t) = \sum_{\underline{\tau}\le s,t} \Bbbk(s,t),\end{aligned} \tag{A.27}$$

where the last equality is due to the dominated convergence theorem and being $\Bbbk$ integrable. For any $\mathrm{g} = (g_t)_{t\in\mathbb{Z}_+} \in \mathscr{H}_\Bbbk$, we know that g is integrable, i.e., $\sum_{t\in\mathbb{Z}_+} |g_t| < \infty$. Therefore, from $\lim_{n\to\infty} \mathrm{f}_n = \mathrm{f}$, we have

$$\begin{aligned}\sum_{\underline{\tau}\le t} g_t &= \lim_{n\to\infty} \sum_{\underline{\tau}\le t\le n} g_t \\ &= \lim_{n\to\infty}\Big\langle \sum_{\underline{\tau}\le t\le n} \Bbbk(\cdot,t), \mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk} \\ &= \lim_{n\to\infty} \langle \mathrm{f}_n, \mathrm{g}\rangle_{\mathscr{H}_\Bbbk} \\ &= \langle \mathrm{f}, \mathrm{g}\rangle_{\mathscr{H}_\Bbbk} \\ &= \Big\langle \sum_{\underline{\tau}\le t} \Bbbk(\cdot,t), \mathrm{g}\Big\rangle_{\mathscr{H}_\Bbbk}.\end{aligned} \tag{A.28}$$

This concludes the proof. □

References

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

Andersen, Martin S., & Chen, Tianshi (2020). Smoothing splines and rank structured matrices: Revisiting the spline kernel. *SIAM Journal on Matrix Analysis and Applications*, *41*(2), 389–412.

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

Berlinet, Alain, & Thomas-Agnan, Christine (2011). *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media.

Bisiacco, Mauro, & Pillonetto, Gianluigi (2020a). Kernel absolute summability is sufficient but not necessary for RKHS stability. *SIAM Journal on Control and Optimization*, *58*(4), 2006–2022.

Bisiacco, Mauro, & Pillonetto, Gianluigi (2020b). On the mathematical foundations of stable RKHSs. *Automatica*, *118*, Article 109038.

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

Chen, Tianshi (2018a). Continuous-time DC kernel – a stable generalized first-order spline kernel. *IEEE Transactions on Automatic Control*, *63*(12), 4442–4447.

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

Chen, Tianshi, Andersen, Martin S., Ljung, Lennart, Chiuso, Alessandro, & Pillonetto, Gianluigi (2014). System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques. *IEEE Transactions on Automatic Control*, *59*(11), 2933–2945.

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

Chen, Tianshi, & Pillonetto, Gianluigi (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.

Cucker, Felipe, & Smale, Steve (2002a). Best choices for regularization parameters in learning theory: On the bias-variance problem. *Foundations of Computational Mathematics*, *2*(4), 413–428.

Cucker, Felipe, & Smale, Steve (2002b). On the mathematical foundations of learning. *American Mathematical Society*, *39*(1), 1–49.

Darwish, Mohamed Abdelmonim Hassan, Pillonetto, Gianluigi, & Tóth, Roland (2018). The quest for the right kernel in Bayesian impulse response identification: The use of OBFs. *Automatica*, *87*, 318–329.

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

Everitt, Niklas, Bottegal, Giulio, & Hjalmarsson, Håkan (2018). An empirical Bayes approach to identification of modules in dynamic networks. *Automatica*, *91*, 144–151.

Fujimoto, Yusuke, Maruta, Ichiro, & Sugie, Toshiharu (2017). Extension of first-order stable spline kernel to encode relative degree. *IFAC-PapersOnLine*, *50*(1), 14016–14021.
Fujimoto, Yusuke, & Sugie, Toshiharu (2018). Kernel-based impulse response estimation with a priori knowledge on the DC gain. *IEEE Control Systems Letters*, *2*(4), 713–718.
Kailath, Thomas (1971). RKHS approach to detection and estimation problems–I: Deterministic signals in Gaussian noise. *IEEE Transactions on Information Theory*, *17*(5), 530–549.
Kanagawa, Motonobu, Hennig, Philipp, Sejdinovic, Dino, & Sriperumbudur, Bharath K. (2018). Gaussian processes and kernel methods: A review on connections and equivalences. arXiv preprint arXiv:1807.02582.
Khosravi, Mohammad, Iannelli, Andrea, Yin, Mingzhou, Parsi, Anilkumar, & Smith, Roy S. (2020). Regularized system identification: A hierarchical Bayesian approach. *IFAC-PapersOnLine*, *53*(2), 406–411, IFAC World Congress 2020.
Khosravi, Mohammad, & Smith, Roy S. (2019). Kernel-based identification of positive systems. In *Conference on Decision and Control* (pp. 1740–1745).
Khosravi, Mohammad, & Smith, Roy S. (2021a). Convex nonparametric formulation for identification of gradient flows. *IEEE Control Systems Letters*, *5*(3), 1097–1102.
Khosravi, Mohammad, & Smith, Roy S. (2021b). Kernel-based identification with frequency domain side-information. arXiv preprint arXiv:2111.00410.
Khosravi, Mohammad, & Smith, Roy S. (2021c). Kernel-based impulse response identification with side-information on steady-state gain. arXiv preprint arXiv:2111.00409.
Khosravi, Mohammad, & Smith, Roy S. (2021d). Nonlinear system identification with prior knowledge on the region of attraction. *IEEE Control Systems Letters*, *5*(3), 1091–1096.
Khosravi, Mohammad, & Smith, Roy S. (2021e). On robustness of kernel-based regularized system identification. *IFAC-PapersOnLine*, *54*(7), 749–754, IFAC Symposium on System Identification.
Khosravi, Mohammad, & Smith, Roy S. (2021f). Regularized identification with internal positivity side-information. arXiv preprint arXiv:2111.00407.
Khosravi, Mohammad, Yin, Mingzhou, Iannelli, Andrea, Parsi, Anilkumar, & Smith, Roy S. (2020). Low-complexity identification by sparse hyperparameter estimation. *IFAC-PapersOnLine*, *53*(2), 412–417, IFAC World Congress 2020.
Kimeldorf, George S., & Wahba, Grace (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *The Annals of Mathematical Statistics*, *41*(2), 495–502.
Ljung, Lennart (1999). *System identification: Theory for the user*. Prentice Hall.
Ljung, Lennart (2010). Perspectives on system identification. *Annual Reviews in Control*, *34*(1), 1–12.
Ljung, Lennart, Chen, Tianshi, & Mu, Biqiang (2020). A shift in paradigm for system identification. *International Journal of Control*, *93*(2), 173–180.
Lukić, Milan, & Beder, Jay (2001). Stochastic processes with sample paths in reproducing kernel Hilbert spaces. *Transactions of the American Mathematical Society*, *353*(10), 3945–3969.
Marconato, Anna, Schoukens, Maarten, & Schoukens, Johan (2016). Filter-based regularisation for impulse response modelling. *IET Control Theory & Applications*, *11*(2), 194–204.
Mu, Biqiang, Chen, Tianshi, & Ljung, Lennart (2018a). Asymptotic properties of generalized cross validation estimators for regularized system identification. *IFAC-PapersOnLine*, *51*(15), 203–208.
Mu, Biqiang, Chen, Tianshi, & Ljung, Lennart (2018b). On asymptotic properties of hyperparameter estimators for kernel-based regularization methods. *Automatica*, *94*, 381–395.
Mu, Biqiang, Chen, Tianshi, & Ljung, Lennart (2021). On the asymptotic optimality of cross-validation based hyper-parameter estimators for regularized least squares regression problems. arXiv preprint arXiv:2104.10471.
Parzen, Emanuel (1959). *Statistical inference on time series by Hilbert space methods, I: Technical Report No. 23*, Department of Statistics, Stanford University.
Parzen, Emanuel (1961). An approach to time series analysis. *The Annals of Mathematical Statistics*, 951–989.
Peypouquet, Juan (2015). *Convex optimization in normed spaces: Theory, methods and examples*. Springer.
Pillonetto, Gianluigi, Chen, Tianshi, Chiuso, Alessandro, Nicolao, Giuseppe De, & Ljung, Lennart (2016). Regularized linear system identification using atomic, nuclear and kernel-based norms: The role of the stability constraint. *Automatica*, *69*, 137–149.
Pillonetto, Gianluigi, & Chiuso, Alessandro (2015). Tuning complexity in regularized kernel-based regression and linear system identification: The robustness of the marginal likelihood estimator. *Automatica*, *58*, 106–117.
Pillonetto, Gianluigi, Chiuso, Alessandro, & De Nicolao, Giuseppe (2019). Stable spline identification of linear systems under missing data. *Automatica*, *108*, Article 108493.
Pillonetto, Gianluigi, & De Nicolao, Giuseppe (2010). A new kernel-based approach for linear system identification. *Automatica*, *46*(1), 81–93.
Pillonetto, Gianluigi, Dinuzzo, Francesco, Chen, Tianshi, De Nicolao, Giuseppe, & Ljung, Lennart (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, *50*(3), 657–682.
Pillonetto, Gianluigi, & Scampicchio, Anna (2021). Sample complexity and minimax properties of exponentially stable regularized estimators. *IEEE Transactions on Automatic Control*.
Prando, Giulia, Chiuso, Alessandro, & Pillonetto, Gianluigi (2017). Maximum entropy vector kernels for MIMO system identification. *Automatica*, *79*, 326–339.
Risuleo, Riccardo Sven, Bottegal, Giulio, & Hjalmarsson, Håkan (2017). A nonparametric kernel-based approach to Hammerstein system identification. *Automatica*, *85*, 234–247.
Risuleo, Riccardo Sven, Lindsten, Fredrik, & Hjalmarsson, Håkan (2019). Bayesian nonparametric identification of Wiener systems. *Automatica*, *108*, Article 108480.
Scandella, Matteo, Mazzoleni, Mirko, Formentin, Simone, & Previdi, Fabio (2020). A note on the numerical solutions of kernel-based learning problems. *IEEE Transactions on Automatic Control*, *66*(2), 940–947.
Scandella, Matteo, Mazzoleni, Mirko, Formentin, Simone, & Previdi, Fabio (2021). Kernel-based identification of asymptotically stable continuous-time linear dynamical systems. *International Journal of Control*, 1–14.
Schoukens, Johan, & Ljung, Lennart (2019). Nonlinear system identification: A user-oriented road map. *IEEE Control Systems Magazine*, *39*(6), 28–99.
Stein, Elias M., & Shakarchi, Rami (2009). *Real analysis: Measure theory, integration, and Hilbert spaces*. Princeton University Press.
Wahba, Grace (1990). *Spline models for observational data*. SIAM.
Zadeh, L. (1956). On the identification problem. *IRE Transactions on Circuit Theory*, *3*(4), 277–281.
Zheng, Man, & Ohta, Yoshito (2021). Bayesian positive system identification: Truncated Gaussian prior and hyperparameter estimation. *Systems & Control Letters*, *148*, Article 104857.
Zorzi, Mattia (2021). A second-order generalization of TC and DC kernels. arXiv preprint arXiv:2109.09562.
Zorzi, Mattia, & Chiuso, Alessandro (2018). The harmonic analysis of kernel functions. *Automatica*, *94*, 125–137.

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, 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.