Regularized Identification with Internal Positivity Side-Information

Khosravi, Mohammad; Smith, Roy S.

DOI
10.1137/23M1556095

Publication date
2025

Document Version
Final published version

Published in
SIAM Journal on Control and Optimization

Citation (APA)
Khosravi, M., & Smith, R. S. (2025). Regularized Identification with Internal Positivity Side-Information. *SIAM Journal on Control and Optimization*, *63*(1), 26-56. https://doi.org/10.1137/23M1556095

SIAM J. CONTROL OPTIM.
Vol. 63, No. 1, pp. 26–56

REGULARIZED IDENTIFICATION WITH INTERNAL POSITIVITY SIDE-INFORMATION*

MOHAMMAD KHOSRAVI† AND ROY S. SMITH‡

Abstract. In this paper, we present an impulse response identification scheme that incorporates the internal positivity side-information of the system. The realization theory of positive systems establishes specific criteria for the existence of a positive realization for a given transfer function. These transfer function criteria are translated to a set of suitable conditions on the shape and structure of the impulse responses of positive systems. Utilizing these conditions, the impulse response estimation problem is formulated as a constrained optimization in a reproducing kernel Hilbert space equipped with a stable kernel, and suitable constraints are imposed to encode the internal positivity side-information. The optimization problem is infinite-dimensional with an infinite number of constraints. An equivalent finite-dimensional convex optimization in the form of a convex quadratic program is derived. The resulting equivalent reformulation makes the proposed approach suitable for numerical simulation and practical implementation. A Monte Carlo numerical experiment evaluates the impact of incorporating the internal positivity side-information in the proposed identification scheme. The effectiveness of the proposed method is demonstrated using data from a heating system experiment.

Key words. system identification, positive systems, kernel methods

MSC codes. 93B30, 93C28, 46E22

DOI. 10.1137/23M1556095

1. Introduction.

In various dynamical systems, the characteristic variables are constrained to be nonnegative or bounded, either by the nature of their definition or according to the physics of the underlying system. In the broad sense, a system described only by such nonnegative variables is called a *positive system* [25]. Examples of such systems include dynamics describing charges in RC circuits, populations for certain species of animals or bacteria, temperatures in the buildings, mass flows in compartmental systems, concentration of pathogens, level of traffic and congestion in networks and roads, prices of stocks and goods, pressure of fluids, and many other quantities of interest that are always nonnegative [9, 25, 33, 46, 62]. For an illustrative practical case, see the heating system experiment in section 7. Depending on our perspective—i.e., whether the positivity feature is considered as an input-output property or a state-space characteristic—we have two central notions of positivity in the system theory literature [25, 31]: internal positivity and external positivity, where the main focus of the literature is on the former one. In externally positive systems, the nonnegativity of the input signal implies the same feature for the output signal. Meanwhile, in internally positive systems, the state trajectory and output signal are nonnegative when the initial state and input signal are nonnegative.

Positive systems have received extensive attention in the past decades owing to being omnipresent in various fields of science and their wide range of applications [6, 54, 68]. Luenberger pioneered the system theoretic approach to positive systems with his seminal work [48] in the 1980s.

*Received by the editors February 28, 2023; accepted for publication (in revised form) August 15, 2024; published electronically January 6, 2025.
https://doi.org/10.1137/23M1556095
†Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, Netherlands (mohammad.khosravi@tudelft.nl).
‡Automatic Control Lab, ETH Zürich, Zürich 8092 Switzerland (rsmith@control.ee.ethz.ch).

Since then, various aspects of system theory have been tackled for positive systems, e.g., realization theory [6], controllability and reachability [26, 65], model reduction [56], observability and observer design [3], stability [13, 20, 34], positive stabilization [59, 61], fault detection and estimation [50], decentralized and distributed control [17, 22], and large-scale positive systems [23, 55].

Concerning the identification problem [4, 67], there are two aspects when the underlying system is known to be internally positive. First, respecting the positivity property can be an essential issue in some applications, such as implementing model predictive control. Accordingly, any accurate mathematical modeling approach is expected to include this feature and construct an internally positive system. The second aspect is informed system identification [1, 37, 38, 39] and concerns utilizing the internal positivity side-information and integrating features of this knowledge in the model to improve the estimation accuracy. Indeed, disregarding the positivity information can lead to models that are not physically interpretable and explainable or behaviors that contradict our expectations [54, 64]. While positive systems have been extensively researched from various viewpoints [55], their identification problem is not well studied, especially with regard to these aspects. For instance, a set of conditions is introduced in [5] for identifying compartmental models, which is a particular case of internal positivity feature. In [16], assuming the output sequence of data is a Poisson process, a maximum likelihood approach is presented for third-order positive systems with distinct real poles. In [64], a particular situation is considered where state variable measurements are provided, in addition to input-output data, and the stability and scalability issues are discussed. Since internally positive systems are also externally positive, side-information on internal positivity implies external positivity. Accordingly, from the perspective of informed system identification, one can consider external positivity as partial information to be integrated into the model. To this end, one may employ the external positive system identification methods [32, 70]. For example, since the externally positive systems are precisely those with the nonnegative impulse response, a kernel-based nonparametric maximum a posteriori approach is introduced in [69, 70] for estimating a nonnegative finite impulse response (FIR). In this approach, the covariance of the prior distribution is specified by the stable kernels, while the mean is designed arbitrarily as an exponentially decaying FIR. Note that, in the FIR identification approaches with an external positivity constraint [32, 69, 70], the complete information of internal positivity is not exploited. The kernel-based methods [2, 8, 41, 42, 52], which resolve the issues of bias-variance trade-off, robustness, and model order selection [36, 40, 47, 53], also provide a proper framework for the integration of various sorts of side-information into the model, including the stability of the system, the smoothness of the impulse response, time constants, and resonant frequencies [10, 11, 14, 24, 28, 43, 44, 45, 49, 57, 58, 72]. This is mainly done through appropriate formulation of the identification problem [43] or based on suitable kernel design [71]; e.g., harmonic analysis of nonstationary Gaussian processes is employed in [72] for kernel design, allowing the incorporation of information about frequency content and decay rates and improving implementation efficiency by providing a practical method for approximating the kernel matrix. Together with the realization theory of positive systems [25], the kernel-based framework can be a suitable foundation for impulse response identification of positive systems.

This paper extends our previous work [37] and presents an identification method that integrates the internal positivity side-information in the estimated impulse response. From the realization theory of positive systems [25], we know that the impulse response of an internally positive system has a specific form; i.e., it has a dominant nonnegative part, where the corresponding transfer function has structured poles, and a residual part. This specific form can be translated to a set of structural constraints on the impulse response. Accordingly, the estimation problem is expressed in the form of a constrained optimization in a stable reproducing kernel Hilbert space, where suitable constraints are imposed to encode the internal positivity side-information. Alhough this problem is initially formulated in an infinite-dimensional space and with an infinite number of constraints, we derive an equivalent finite-dimensional convex optimization in the form of a convex quadratic program (QP). We evaluate the impact of incorporating the internal positivity side-information and assess the performance of the proposed identification scheme through a Monte Carlo numerical experiment. The efficacy of the proposed positive system identification technique is confirmed using data from a thermal dynamics experiment.

2. Notation. In this paper, the set of natural numbers, the set of integers, the set of nonnegative integers, the set of real numbers, the set of nonnegative real numbers, n-dimensional Euclidean space, and the set of n by m matrices are denoted, respectively, by $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}_+$, $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}^n$, and $\mathbb{R}^{n\times m}$. The positive orthant of $\mathbb{R}^n$ is denoted by $\mathbb{R}^n_+$. The identity matrix and zero matrix are denoted by $\mathbb{I}$ and $\mathbf{0}_n$, respectively. Also, the n-dimensional zero vector and the all-ones vector are denoted by 0_n and 1_n, respectively. When the dimension is clear from the context, we drop the subscript. For $p \in [1,\infty)$, the p-norm of vector $\mathrm{h} = (h_s)_{s=0}^{\infty} \in \mathbb{R}^{\mathbb{Z}_+}$ is defined as $\|\mathrm{h}\|_p = (\sum_{s\geq 0} |h_s|^p)^{\frac{1}{p}}$, and the ∞-norm of h is defined as $\|\mathrm{h}\|_\infty = \sup_{s\geq 0} |h_s|$. The space of vectors $\mathrm{h} \in \mathbb{R}^{\mathbb{Z}_+}$ with finite p-norm is denoted by ℓ^p. Given bounded signal $\mathrm{u} = (u_s)_{s\in\mathbb{Z}_+}$ and $t \in \mathbb{Z}$, the linear map $\mathrm{L}_t : \ell^1 \to \mathbb{R}$ is defined as $\mathrm{L}_t(\mathrm{g}) = \sum_{s=0}^{\infty} g_s u_{t-s}$ for any $\mathrm{g} = (g_s)_{s=0}^{\infty} \in \ell^1$. Given a subset $\mathcal{C} \subset \mathcal{X}$, the function $\delta_{\mathcal{C}} : \mathcal{X} \to \{0, +\infty\}$ is defined as $\delta_{\mathcal{C}}(x) = 0$ if $x \in \mathcal{C}$ and $\delta_{\mathcal{C}}(x) = \infty$, otherwise. The set of polynomials in x with maximum degree n and real coefficients is denoted by $\mathbb{R}_n[x]$. For transfer function G, $r(G)$ denotes its spectral radius.

3. System identification with internal positivity side-information: Problem statement and mathematical formulation.

3.1. Positive system identification: Problem statement. Let $\mathrm{g}^{(\mathcal{S})} := (g_t^{(\mathcal{S})})_{t=0}^{\infty}$ be the impulse response of stable and causal system $\mathcal{S}$ and $G^{(\mathcal{S})}(z) := \sum_{t=0}^{\infty} g_t^{(\mathcal{S})} z^{-t}$ be the corresponding transfer function. We call impulse response $\mathrm{g}^{(\mathcal{S})}$, or equivalently the system $\mathcal{S}$, *internally positive* if there exists a realization such that the state trajectory and the output remain nonnegative given that the initial state and the input are nonnegative (see Definition 3.1). Suppose that a bounded signal u is applied to the input of system $\mathcal{S}$. Let y_t denote the measured output at time instant $t \in \mathscr{T}$, where $\mathscr{T} := \{t_i \mid i = 0, \dots, n_{\mathscr{D}}-1\}$ for a given $n_{\mathscr{D}} \in \mathbb{N}$. In other words, we have that

$$y_t := \mathrm{L}_t(\mathrm{g}^{(\mathcal{S})}) + w_t, \qquad t \in \mathscr{T}, \tag{3.1}$$

where w_t denotes the uncertainty in the output measured at time instant t for $t \in \mathscr{T}$. Accordingly, we have a set of input-output measurement data denoted by $\mathscr{D}$. Based on the given setting, we introduce the following impulse response identification problem.

PROBLEM 1. *Using data $\mathscr{D}$, estimate the impulse response of $\mathrm{g}^{(\mathcal{S})}$ given the side-information that $\mathrm{g}^{(\mathcal{S})}$ is internally positive.*

In addressing this problem, the main concern is the appropriate integration of the available internal positivity side-information into the impulse response identification problem. To this end, we need to exploit suitable conditions inducing the desired positivity feature. In the sequel, these conditions are discussed, and the estimation problem is formulated accordingly.

3.2. Positive system identification: Mathematical formulation. In the realization theory of positive systems, sufficient conditions are introduced under which the transfer function of a system admits a so-called positive realization. We employ these conditions together with the notion of stable reproducing kernel Hilbert spaces (RKHSs) for bridging to the impulse response identification of stable positive systems.

DEFINITION 3.1 (see [25]). *The impulse response* $\mathrm{g}^{(\mathcal{S})} \in \ell^1$ *is said to be* internally positive*, or simply* positive*, if there exists a realization for system* $\mathcal{S}$ *as*

$$\begin{cases} \mathrm{x}_{t+1} = \mathrm{A}\mathrm{x}_t + \mathrm{b}u_t \\ y_t = \mathrm{c}\mathrm{x}_t + du_t \end{cases} \qquad \forall\ t \in \mathbb{Z}, \tag{3.2}$$

where $\mathrm{A} \in \mathbb{R}^{n_\mathrm{x} \times n_\mathrm{x}}$, $\mathrm{b} \in \mathbb{R}^{n_\mathrm{x}}$, $\mathrm{c} \in \mathbb{R}^{1 \times n_\mathrm{x}}$, $d \in \mathbb{R}$, *and* $n_\mathrm{x} \in \mathbb{N}$ *such that* $\mathrm{x}_0 \in \mathbb{R}_+^{n_\mathrm{x}}$ *and* $u_t \in \mathbb{R}_+$*, for all* $t \geq 0$*, implies that* $\mathrm{x}_t \in \mathbb{R}_+^{n_\mathrm{x}}$ *and* $y_t \in \mathbb{R}_+$ *for each* $t \geq 0$*. The realization* (3.2) *with this property is called a* positive realization *for* $\mathrm{g}^{(\mathcal{S})}$*, or equivalently, for* $G^{(\mathcal{S})}$*. Moreover, the set of stable internally positive impulse responses is denoted by* $\mathscr{P}$*.*

The internal positivity enforces a specific attribute on $\mathrm{g}^{(\mathcal{S})}$ according to Kronecker's theorem given below.

THEOREM 3.2 (see [27]). *With respect to impulse response* $\mathrm{g} = (g_t)_{t=0}^{\infty} \in \ell^1$*, define Hankel operator* $\mathrm{Hankel}(\mathrm{g}) : \ell^\infty \to \ell^\infty$ *with entrywise representation in the standard basis of* ℓ^∞ *as follows:*

$$\mathrm{Hankel}(\mathrm{g}) := \begin{bmatrix} g_0 & g_1 & g_2 & \cdots \\ g_1 & g_2 & & \\ g_2 & & & \\ \vdots & & & \end{bmatrix} = \left[g_{i+j-2}\right]_{i,j=1}^{\infty}. \tag{3.3}$$

Then, $G(z) := \sum_{t=0}^{\infty} g_t z^{-t}$ *is a rational function if and only if the rank of* $\mathrm{Hankel}(\mathrm{g})$ *is finite; i.e., we have that* $\mathrm{rank}(\mathrm{Hankel}(\mathrm{g})) := \dim\{\mathrm{Hankel}(\mathrm{g})\mathrm{v} \,\big|\, \mathrm{v} \in \ell^\infty\} < \infty$*. We call impulse response* g finite Hankel rank *when this property is satisfied.*

According to (3.2), we know that $G^{(\mathcal{S})}(z) = \mathrm{c}(z\mathbb{I} - \mathrm{A})^{-1}\mathrm{b} + d$ is a rational function. Therefore, due to Theorem 3.2, the internal positivity of $\mathrm{g}^{(\mathcal{S})}$ implies that

$$\mathrm{rank}\Big(\mathrm{Hankel}\big(\mathrm{g}^{(\mathcal{S})}\big)\Big) < \infty. \tag{3.4}$$

The realization (3.2) needs to have the special structure introduced in the next theorem.

THEOREM 3.3 (see [25]). *The impulse response* $\mathrm{g}^{(\mathcal{S})} \in \ell^1$ *is internally positive if and only if there exists a realization as in* (3.2) *such that the entries of* $\mathrm{A}, \mathrm{b}, \mathrm{c}$*, and* d *are nonnegative. Moreover, the internal passivity of impulse response* $\mathrm{g}^{(\mathcal{S})} \in \ell^1$ *implies that*

$$g_t^{(\mathcal{S})} \geq 0 \qquad \forall t \in \mathbb{Z}_+. \tag{3.5}$$

Let $\overline{\mathscr{P}}$ denote the set of stable externally positive impulse responses; i.e., $\overline{\mathscr{P}} := \{\mathrm{g} = (g_t)_{t=0}^{\infty} \in \ell^1 \,\big|\, g_t \geq 0\ \forall t \in \mathbb{Z}_+\} = \ell^1 \cap \mathbb{R}_+^{\mathbb{Z}_+}$. Theorem 3.3 implies that any internally positive impulse response is externally positive as well. Accordingly, due to (3.4), we have that $\mathscr{P} \subset \underline{\mathscr{P}}$, where the set of impulse responses $\underline{\mathscr{P}}$ is defined as

$$\underline{\mathscr{P}} := \overline{\mathscr{P}} \cap \Big\{ \mathrm{g} \in \ell^1 \,\Big|\, \operatorname{rank}(\operatorname{Hankel}(\mathrm{g})) < \infty \Big\}. \tag{3.6}$$

Meanwhile, from the next example, which is a modified version of an example given in [6], one can see that the inclusion in $\mathscr{P} \subset \underline{\mathscr{P}}$ is strict; i.e., $\mathscr{P} \neq \underline{\mathscr{P}}$.

Example 1. Let $\rho \in (0,1)$ and ω be an irrational real number. Define the impulse response $\mathrm{g} = (g_t)_{t=0}^{\infty}$ as $g_t = \rho^t(1+\cos(2\pi\omega t))$ for all $t \in \mathbb{Z}_+$. One can see that g is a nonnegative impulse response with the following transfer function:

$$G(z) = \frac{1}{1-\rho z^{-1}} + \frac{1-\rho\cos w\ z^{-1}}{1-2\rho\cos w\ z^{-1} + \rho^2 z^{-2}}. \tag{3.7}$$

Therefore, we have that $\mathrm{g} \in \underline{\mathscr{P}}$. However, there is no positive realization for the impulse response g [25]. Accordingly, due to Definition 3.1, we know that g is not internally positive; i.e., $\mathrm{g} \notin \mathscr{P}$ and $\mathscr{P} \neq \underline{\mathscr{P}}$.

Based on the above discussion, conditions (3.4) and (3.5) are necessary but not sufficient for $\mathrm{g}^{(\mathcal{S})}$ being an internally positive impulse response. Using the following theorem, we derive sufficient conditions for the internal positivity of g, which are used later to formulate the identification problem of internally positive systems.

THEOREM 3.4 (see [25]). *Let $\mathrm{g} \in \ell^1$ be a nonnegative impulse response and G be the corresponding transfer function. If G is a strictly proper rational function with a unique dominant pole $\rho \in (0,1)$, then there exists a positive realization for G.*

With respect to each $\rho \in (0,1)$, define $\mathscr{P}_\rho \subset \ell^1$ as the set of nonnegative impulse responses satisfying (3.4) such that we have that

$$\text{there exists } a \in (0,\infty), \quad \lim_{t\to\infty} \rho^{-t} g_t = a; \tag{3.8}$$

i.e., $\lim_{t\to\infty} \rho^{-t} g_t$ is well defined and equal to a positive real scalar a. Furthermore, we define $\mathscr{P}_{(0,1)}$ as $\mathscr{P}_{(0,1)} = \cup_{\rho\in(0,1)} \mathscr{P}_\rho$. Based on Theorem 3.4, we have the following corollary for $\mathscr{P}_\rho$ and $\mathscr{P}_{(0,1)}$.

COROLLARY 3.5. *For any $\rho \in (0,1)$, each impulse response $\mathrm{g}^{(\mathcal{S})} \in \mathscr{P}_\rho$ is internally positive; i.e., $\mathscr{P}_\rho \subset \mathscr{P}$. Moreover, we have that $\mathscr{P}_{(0,1)} \subset \mathscr{P}$.*

Proof. Let $\mathrm{g}^{(\mathcal{S})} \in \mathscr{P}_\rho$ and $G^{(\mathcal{S})}$ be the corresponding transfer function. We know that $\mathrm{g}^{(\mathcal{S})}$ is a nonnegative impulse response, which satisfies (3.4). Accordingly, due to Theorem 3.2, there exist $n_{\mathrm{x}} \in \mathbb{N}$, $\mathrm{A} \in \mathbb{R}^{n_{\mathrm{x}}\times n_{\mathrm{x}}}$, $\mathrm{b} \in \mathbb{R}^{n_{\mathrm{x}}}$, $\mathrm{c} \in \mathbb{R}^{1\times n_{\mathrm{x}}}$, and $d \in \mathbb{R}$ such that $G^{(\mathcal{S})}(z) = \mathrm{c}(z\mathbb{I} - \mathrm{A})^{-1}\mathrm{b} + d$. Note that, since $d = g_0^{(\mathcal{S})}$, we know that $d \geq 0$. Let $\mathrm{g} = (g_t)_{t=0}^{\infty}$ be the impulse response defined as $g_0 = 0$ and $g_t = g_t^{(\mathcal{S})}$ for $t \geq 1$. Also, let G be the transfer function corresponding to g. One can easily see that g is nonnegative, and also, we have that $G(z) = G^{(\mathcal{S})}(z) - d = \mathrm{c}(z\mathbb{I} - \mathrm{A})^{-1}\mathrm{b}$, which is a strictly proper rational transfer function. Since there exists $a > 0$ such that $\lim_{t\to\infty} \rho^{-t} g_t^{(\mathcal{S})} = a$, we know that $\lim_{t\to\infty} \rho^{-t}(g_t - a\rho^t) = 0$. Therefore, the spectral radius of the rational transfer function $G(z) - a(1-\rho z^{-1})^{-1}$ is less than ρ, and consequently, ρ is the unique dominant pole of $G(z)$. Hence, according to Theorem 3.4, $G(z)$ admits a positive realization; i.e., there exist $m_{\mathrm{x}} \in \mathbb{N}$, $\mathrm{A}_+ \in \mathbb{R}_+^{m_{\mathrm{x}}\times m_{\mathrm{x}}}$, $\mathrm{b}_+ \in \mathbb{R}_+^{m_{\mathrm{x}}}$, and $\mathrm{c}_+ \in \mathbb{R}_+^{1\times m_{\mathrm{x}}}$ such that $G(z) = \mathrm{c}_+(z\mathbb{I} - \mathrm{A}_+)^{-1}\mathrm{b}_+$. Therefore, we have $G^{(\mathcal{S})}(z) = \mathrm{c}_+(z\mathbb{I} - \mathrm{A}_+)^{-1}\mathrm{b}_+ + d$, which says that $G^{(\mathcal{S})}$ has a positive realization due to $d \geq 0$. Accordingly, $\mathrm{g}^{(\mathcal{S})}$ is internally positive and $\mathrm{g}^{(\mathcal{S})} \in \mathscr{P}$; i.e., $\mathscr{P}_\rho \subset \mathscr{P}$. From this result and the definition of $\mathscr{P}_{(0,1)}$, the last claim is directly implied. ∎

Note that $\mathscr{P}_{(0,1)}$ contains exactly the impulse responses satisfying conditions (3.4), (3.5), and (3.8). Hence, Corollary 3.5 says that any impulse response in ℓ^1 that satisfies these conditions is internally positive. Accordingly, one can employ (3.4), (3.5), and (3.8) in the identification problem to enforce internal positivity on the impulse response to be estimated. The next theorem further highlights the importance of positive systems $\mathscr{P}_{(0,1)}$.

THEOREM 3.6. *The set of impulse responses $\mathscr{P}_{(0,1)}$ is dense in $\mathscr{P}$ with respect to p-norm topology for any $p \in [1,\infty]$.*

Proof. Let $\varepsilon > 0$ and $\mathrm{g} \in \mathscr{P}$ with transfer function G. Since $\mathscr{P} \subset \ell^1$, each element of $\mathscr{P}$ is a stable impulse response, and therefore, we have that $r(G) < 1$. Let ρ and a be positive real scalars such that $\rho \in (r(G), 1)$ and $a < (1-\rho)\varepsilon$. Consider an impulse response $\mathrm{g}^{(\varepsilon)} = (g_t^{(\varepsilon)})_{t=0}^\infty$ with transfer function $G^{(\varepsilon)}$, where, for any $t \in \mathbb{Z}_+$, g_t is defined as $g_t^{(\varepsilon)} = g_t + a\rho^t$. For any $t \in \mathbb{Z}_+$, one has $g_t \geq 0$, and, since $a, \rho > 0$, it follows that $g_t^{(\varepsilon)} \geq 0$; i.e., $\mathrm{g}^{(\varepsilon)}$ is a nonnegative impulse response. Moreover, one can easily see that $G^{(\varepsilon)}(z) = G(z) + a(1-\rho z^{-1})^{-1}$. Since g belongs to $\mathscr{P}$, we know that G is a rational transfer function. Accordingly, due to Theorem 3.2, it follows that (3.4) holds for $\mathrm{g}^{(\varepsilon)}$. Moreover, $\rho > r(G)$ implies that $\lim_{t\to\infty} \rho^{-t} g_t = 0$. Subsequently, one has $\lim_{t\to\infty} \rho^{-t} g_t^{(\varepsilon)} = a$, and therefore, $\mathrm{g}^{(\varepsilon)} \in \mathscr{P}_\rho \subset \mathscr{P}_{(0,1)}$. For the case of $p = \infty$, we have that $\|\mathrm{g} - \mathrm{g}^{(\varepsilon)}\|_\infty = \sup_{t \in \mathbb{Z}_+} a\rho^t = a < (1-\rho)\varepsilon < \varepsilon$. Also, for $p \in [1,\infty)$, one can see that

$$\|\mathrm{g} - \mathrm{g}^{(\varepsilon)}\|_p = a \left(\sum_{t=0}^{\infty} \rho^{pt} \right)^{\frac{1}{p}} = \frac{a}{(1-\rho^p)^{\frac{1}{p}}} < \frac{(1-\rho)\varepsilon}{(1-\rho^p)^{\frac{1}{p}}} \leq \varepsilon,$$

where the last inequality is due to $\rho^p + (1-\rho)^p \leq 1$, which holds for any $\rho \in (0,1)$ and $p \in [1,\infty)$. ∎

With respect to each $\mathrm{g} = (g_t)_{t=0}^\infty \in \mathscr{P}_{(0,1)}$, one can define impulse response $\mathrm{h} = (h_t)_{t=0}^\infty$ such that $h_t = g_t - a\rho^t$ for $t \in \mathbb{Z}_+$, where ρ and a are the positive scalars introduced in (3.8). More precisely, according to (3.8), we know that $\lim_{t\to\infty} \rho^{-t} h_t = 0$, and since $\rho \in (0,1)$, property (3.8) implies that $\mathrm{h} = (h_t)_{t=0}^\infty$ is a stable impulse response dominated by

$$\mathrm{f}_\rho = (f_t)_{t=0}^\infty := (\rho^t)_{t=0}^\infty; \tag{3.9}$$

i.e., $\mathrm{g} = (g_t)_{t=0}^\infty$ can be decomposed into a stable impulse response and a first-order positive impulse response that dominates the stable one. Note that, when $\mathrm{g} = (g_t)_{t=0}^\infty$ corresponds to a finite-dimensional system, the transfer function defined by $\mathrm{h} = (h_t)_{t=0}^\infty$ has a finite number of poles, each with a magnitude strictly less than $\rho \in (0,1)$. Following the above discussion, to identify the internally positive impulse response $\mathrm{g} = (g_t)_{t=0}^\infty$, we need to estimate ρ, a, and the stable impulse response $\mathrm{h} = (h_t)_{t=0}^\infty$, dominated by $\mathrm{f}_\rho = (\rho^t)_{t=0}^\infty$, and meanwhile ensure that $\mathrm{g} = (g_t)_{t=0}^\infty$ satisfies properties (3.4) and (3.5). To this end, we also need a suitable hypothesis space for $\mathrm{h} = (h_t)_{t=0}^\infty$. Thus, we employ stable RKHSs [12, 52].

DEFINITION 3.7 (see [7, 12]). *The nonzero symmetric function $\mathbf{k} : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{R}$ is said to be a* kernel *if, for any $m \in \mathbb{N}$, $t_1, \ldots, t_m \in \mathbb{T}$ and $a_1, \ldots, a_m \in \mathbb{R}$, we have that $\sum_{i=1}^m \sum_{j=1}^m a_i \mathbf{k}(t_i, t_j) a_j \geq 0$. Moreover, the* section *of kernel $\mathbf{k}$ at $t \in \mathbb{Z}_+$ is denoted by $\mathbf{k}_t$ and defined as the function $\mathbf{k}(t, \cdot) : \mathbb{Z}_+ \to \mathbb{R}$. Furthermore, the* *positive kernel* $\mathbb{k}$ *is said to be* stable *if, for any* $\mathrm{u} = (u_t)_{t\in\mathbb{Z}_+} \in \ell^\infty$*, we have that* $\sum_{t\in\mathbb{Z}_+}|\sum_{s\in\mathbb{Z}_+}u_s\mathbb{k}(t,s)| < \infty$.

THEOREM 3.8 (see [7, 12]). *Given a kernel* $\mathbb{k} : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{R}$*, there exists a* unique *Hilbert space* $\mathcal{H}_{\mathbb{k}} \subseteq \mathbb{R}^{\mathbb{Z}_+}$ *endowed with inner product* $\langle\cdot,\cdot\rangle_{\mathcal{H}_{\mathbb{k}}}$ *and norm* $\|\cdot\|_{\mathcal{H}_{\mathbb{k}}}$*, called the RKHS* with kernel $\mathbb{k}$*, such that, for each* $t \in \mathbb{Z}_+$*, we have* (i) $\mathbb{k}_t \in \mathcal{H}_{\mathbb{k}}$ *and* (ii) $\langle \mathrm{g}, \mathbb{k}_t\rangle_{\mathcal{H}_{\mathbb{k}}} = g_t$ *for all* $\mathrm{g} = (g_t)_{t\in\mathbb{Z}_+} \in \mathcal{H}_{\mathbb{k}}$*. The second feature is called the* reproducing property*. Moreover,* $\mathcal{H}_{\mathbb{k}} \subset \ell^1$ *if and only if* $\mathbb{k}$ *is a stable kernel. In this case,* $\mathcal{H}_{\mathbb{k}}$ *is said to be a* stable RKHS.

Given a stable kernel $\mathbb{k}$, we take $\mathcal{H}_{\mathbb{k}}$ as the hypothesis space for the stable impulse response h. Considering the set of input-output data $\mathscr{D}$, we define the empirical loss function $\mathcal{E}_\rho : \mathbb{R} \times \mathcal{H}_{\mathbb{k}} \to \mathbb{R}_+$ as

$$\mathcal{E}_\rho(a,\mathrm{h}) := \sum_{i=0}^{n_{\mathscr{D}}-1} \left(y_{t_i} - a\mathrm{L}_{t_i}(\mathrm{f}_\rho) - \mathrm{L}_{t_i}(\mathrm{h})\right)^2, \tag{3.10}$$

where we assume the hyperparameter $\rho \in (0,1)$ is given. The estimation of ρ will be discussed later. We formulate the identification problem with internal positivity side-information as the following regularized optimization problem:

$$\begin{array}{cl} \min\limits_{a\in\mathbb{R},\mathrm{h}\in\mathcal{H}_{\mathbb{k}}} & \mathcal{E}_\rho(a,\mathrm{h}) + \lambda\,\|\mathrm{h}\|^2_{\mathcal{H}_{\mathbb{k}}} \\ \text{s.t.} & h_t + a\rho^t \geq 0 \quad \forall t \geq 0, \\ & \mathrm{rank}(\mathrm{Hankel}(\mathrm{h})) < \infty, \\ & a > 0, \end{array} \tag{3.11}$$

where $\lambda > 0$ is the regularization weight. Note that, similar to the standard problem formulation in the literature on kernel-based impulse response identification [53], the objective function in (3.11) is an empirical loss function regularized with the RKHS norm of h. This ensures the stability of h and also allows incorporating other features such as exponential decay and smoothness [10]. Furthermore, one should note that, from $\mathrm{g} = \mathrm{h} + \alpha\mathrm{f}_\rho$, we have that $\mathrm{Hankel}(\mathrm{g}) = \mathrm{Hankel}(\mathrm{h}) + \mathrm{Hankel}(\alpha\mathrm{f}_\rho)$, which implies that rank(Hankel(h)) is finite if and only if rank(Hankel(g)) is finite. More precisely, due to $\mathrm{rank}(\mathrm{Hankel}(\alpha\mathrm{f}_\rho)) = 1$, we know that

$$\begin{aligned} \mathrm{rank}(\mathrm{Hankel}(\mathrm{g})) &= \mathrm{rank}\big(\mathrm{Hankel}(\mathrm{h}) + \mathrm{Hankel}(\alpha\mathrm{f}_\rho)\big) \\ &\leq \mathrm{rank}(\mathrm{Hankel}(\mathrm{h})) + \mathrm{rank}(\mathrm{Hankel}(\alpha\mathrm{f}_\rho)) \\ &= \mathrm{rank}(\mathrm{Hankel}(\mathrm{h})) + 1. \end{aligned} \tag{3.12}$$

Similarly, we can show that

$$\mathrm{rank}(\mathrm{Hankel}(\mathrm{h})) \leq \mathrm{rank}(\mathrm{Hankel}(\mathrm{g})) + 1. \tag{3.13}$$

Thus, from (3.12) and (3.13), one can conclude that $\mathrm{rank}(\mathrm{Hankel}(\mathrm{h})) < \infty$ is equivalent to $\mathrm{rank}(\mathrm{Hankel}(\mathrm{g})) < \infty$. The next theorem says that the solution of (3.11) leads to an internally positive estimation of impulse response g. Before proceeding to the theorem, we need to introduce an assumption.

Assumption 1. There exist $C \in \mathbb{R}_+$ and $\rho_{\mathrm{d}} \in (0,\rho)$ such that we have $|\mathbb{k}(t,t)| \leq C\rho_{\mathrm{d}}^{2t}$ for any $t \in \mathbb{Z}_+$.

The primary objective of the condition introduced in Assumption 1 is to ensure the noted dominancy feature on the elements of $\mathcal{H}_{\mathbb{k}}$, the hypothesis space utilized for the estimation of $\mathrm{h} = (h_t)_{t=0}^\infty$.

THEOREM 3.9. *Let Assumption* 1 *hold,* a *and* $\mathrm{h}=(h_t)_{t=1}^{\infty}$ *be a solution pair for* (3.11)*, and the impulse response* $\mathrm{g}=(g_t)_{t=0}^{\infty}$ *be defined as* $g_t=h_t+a\rho^t$ *for any* $t\in\mathbb{Z}_+$*. Then,* g *is internally positive.*

Proof. Let G be the transfer function that corresponds to g. Due to (3.11), the rank of Hankel operator Hankel(h) is finite. Subsequently, according to Theorem 3.2, the transfer function corresponding to h, denoted by H, has finite order. On the other hand, we know that $G(z)=az^{-1}(1-\rho z^{-1})^{-1}+H(z)$. Therefore, the order of G is finite. Accordingly, due to Theorem 3.2, it follows that rank(Hankel(g)) $<\infty$; i.e., g satisfies (3.4). Also, according to the first constraint in (3.11), one can see that g is a nonnegative impulse response and that (3.5) holds for g. Moreover, from the reproducing property of the kernel, we know that $h_t=\langle \mathbb{k}_t,\mathrm{h}\rangle$ and $\|\mathbb{k}_t\|_{\mathcal{H}_{\mathbb{k}}}^2=\langle \mathbb{k}_t,\mathbb{k}_t\rangle=\mathbb{k}(t,t)$ for any $t\in\mathbb{Z}_+$. Hence, due to the Cauchy–Schwartz inequality and Assumption 1, we have that

$$|h_t|=|\langle \mathbb{k}_t,\mathrm{h}\rangle|\le\|\mathbb{k}_t\|_{\mathcal{H}_{\mathbb{k}}}\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}=\mathbb{k}(t,t)^{\frac{1}{2}}\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}\le C^{\frac{1}{2}}\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}\rho_{\mathrm{d}}^t \tag{3.14}$$

for any $t\in\mathbb{Z}_+$. Following this, one can see that

$$0\le\liminf_{t\to\infty}\rho^{-t}h_t\le\limsup_{t\to\infty}\rho^{-t}h_t\le\limsup_{t\to\infty}\rho^{-t}|h_t|\le\limsup_{t\to\infty}C^{\frac{1}{2}}\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}\rho_{\mathrm{d}}^t\rho^{-t}=0,$$

where the last equality is due to $\rho_{\mathrm{d}}\in(0,\rho)$. Hence, $\lim_{t\to\infty}\rho^{-t}h_t$ is well defined, and we have that $\lim_{t\to\infty}\rho^{-t}h_t=0$. Subsequently, due to the definition of g, it follows that $\lim_{t\to\infty}\rho^{-t}g_t=a$ and g satisfies (3.8). Therefore, g belongs to $\mathscr{P}_\rho$, and consequently, due to Corollary 3.5, g is internally positive. ∎

Remark 3.10. For a, h, and g introduced in Theorem 3.9, we have that

$$\mathcal{E}_\rho(a,\mathrm{h})=\sum_{i=0}^{n_{\mathscr{D}}-1}\big(y_{t_i}-\mathrm{L}_{t_i}(\mathrm{g})\big)^2; \tag{3.15}$$

i.e., in the cost function of (3.11), the first term is the sum of squared errors for the impulse response fitting when the dominant pole ρ is known.

Let $a_{\min}>0$ be a specified lower bound for a, which may potentially be very small. Accordingly, we can reformulate the identification problem as

$$\begin{aligned}\min_{a\in\mathbb{R},\mathrm{h}\in\mathcal{H}_{\mathbb{k}}}\quad & \mathcal{E}_\rho(a,\mathrm{h})+\lambda\,\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2\\ \text{s.t.}\quad & h_t+a\rho^t\ge 0\quad\forall t\ge 0,\\ & \mathrm{rank}(\mathrm{Hankel}(\mathrm{h}))<\infty,\\ & a\ge a_{\min}.\end{aligned} \tag{3.16}$$

Indeed, the last constraint in (3.11) is modified to $a\ge a_{\min}$ to ensure that $a>0$. The following section studies this problem and presents a practical method for obtaining its solution.

Remark 3.11. While the last constraint in (3.16) is introduced primarily to ensure that $a>0$, it also allows for avoiding specific technical issues arising from employing strict inequalities. In many cases, it is feasible to obtain a lower bound, possibly conservative and imprecise, on the coefficient of the dominant part of the impulse response that can be used as $a_{\min}$ in the identification problem. Alternatively, one may initially assign an arbitrary positive value to $a_{\min}$ and solve the final optimization problem in section 4. If the constraint $a\ge a_{\min}$ is inactive, the decided choice for $a_{\min}$ is not a limiting factor. Otherwise, we can scale $a_{\min}$ with a positive scalar strictly smaller than one and repeat the procedure till we reach the numerical solver precision level. From the perspective of numerical implementation, there is no practical distinction between $a>0$ and $a \geq a_{\min}$ when $a_{\min}$ is a positive real scalar at the precision level of numerical solvers; i.e., one may alternatively set $a_{\min}$ to this value.

4. Toward a tractable solution. In this section, we investigate the optimization problem (3.16), introduced for impulse response identification with internal positivity side-information. This optimization problem is in an infinite-dimensional space with an infinite number of constraints. In the following, we analyze this problem and provide a tractable approach for deriving its solution.

Let $\mathcal{V}_{\mathbb{k}}$ be the Hilbert space $\mathbb{R} \times \mathcal{H}_{\mathbb{k}}$, which is endowed with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{V}_{\mathbb{k}}} : \mathcal{V}_{\mathbb{k}} \times \mathcal{V}_{\mathbb{k}} \to \mathbb{R}$ defined as

$$\langle (a_1, \mathrm{h}_1), (a_2, \mathrm{h}_2) \rangle_{\mathcal{V}_{\mathbb{k}}} = a_1 a_2 + \langle \mathrm{h}_1, \mathrm{h}_2 \rangle_{\mathcal{H}_{\mathbb{k}}} \tag{4.1}$$

for any $a_1, a_2 \in \mathbb{R}$ and $\mathrm{h}_1, \mathrm{h}_2 \in \mathcal{H}_{\mathbb{k}}$. Also, let $\mathscr{F} \subseteq \mathcal{H}_{\mathbb{k}}$ be the set of finite Hankel rank impulse responses in $\mathcal{H}_{\mathbb{k}}$; i.e., $\mathscr{F} = \{\mathrm{h} \in \mathcal{H}_{\mathbb{k}} \,|\, \mathrm{rank}(\mathrm{Hankel}(\mathrm{h})) < \infty\}$. We define function $\mathcal{J}_{\mathscr{F}} : \mathcal{V}_{\mathbb{k}} \to \mathbb{R} \cup \{+\infty\}$ as

$$\mathcal{J}_{\mathscr{F}}(a, \mathrm{h}) = \mathcal{E}_\rho(a, \mathrm{h}) + \sum_{s=0}^{\infty} \delta_{\mathscr{R}_s}(a, \mathrm{h}) + \delta_{\mathscr{F}}(\mathrm{h}) + \lambda \|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2, \tag{4.2}$$

where $\mathscr{R}_s \subseteq \mathcal{V}_{\mathbb{k}}$ is the set

$$\mathscr{R}_s := \left\{ \big(a, (h_t)_{t \in \mathbb{Z}_+}\big) \in \mathcal{V}_{\mathbb{k}} \,\middle|\, h_s + a\rho^s \geq 0, a \geq a_{\min} \right\} \tag{4.3}$$

for $s \in \mathbb{Z}_+$. From the definition of $\mathcal{J}_{\mathscr{F}}$, it follows easily that the optimization problem (3.16) is equivalent to

$$\inf_{(a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}}} \mathcal{J}_{\mathscr{F}}(a, \mathrm{h}). \tag{4.4}$$

For $(a, \mathrm{h}) = (a_{\min}, \mathbf{0})$, where $\mathbf{0}$ denotes the zero vector in $\mathcal{H}_{\mathbb{k}}$, one can easily see that

$$\mathcal{J}_{\mathscr{F}}(a_{\min}, \mathbf{0}) = \sum_{i=1}^{n} \big(y_{t_i} - a_{\min} \mathrm{L}_{t_i}(\mathrm{f}_\rho) \big)^2 < \infty. \tag{4.5}$$

Since, for any $(a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}}$, we have that $\mathcal{J}_{\mathscr{F}}(a, \mathrm{h}) \geq 0$, it follows that (4.4) is bounded. However, this argument does not guarantee the existence of a solution for (4.4). In the following, we show that, under mild conditions, the optimization problem (4.4) admits a solution when the kernel $\mathbb{k}$ meets certain criteria. Let function $\mathcal{J} : \mathcal{V}_{\mathbb{k}} \to \mathbb{R} \cup \{+\infty\}$ be defined as

$$\mathcal{J}(a, \mathrm{h}) = \mathcal{E}_\rho(a, \mathrm{h}) + \sum_{s=0}^{\infty} \delta_{\mathscr{R}_s}(a, \mathrm{h}) + \lambda \|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2, \tag{4.6}$$

and consider the optimization problem

$$\inf_{(a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}}} \mathcal{J}(a, \mathrm{h}). \tag{4.7}$$

One can easily see that $\mathcal{J}_{\mathscr{F}} = \mathcal{J} + \delta_{\mathscr{F}}$, which implies that

$$\mathcal{J}(a, \mathrm{h}) \leq \mathcal{J}_{\mathscr{F}}(a, \mathrm{h}) \qquad \forall\, (a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}}. \tag{4.8}$$

Consequently, if (4.7) has a solution (a^*, h^*) such that the operator Hankel(h^*) is finite rank, then (a^*, h^*) is a solution for (4.4) as well. In other words, the identification problem with internal positivity side-information introduced in (3.16) admits a solution. Hence, we need to study the solution behavior of (4.7). To this end, we require several technical assumptions.

Assumption 2. There exists $i \in \{0, 1, \dots, n_{\mathscr{D}} - 1\}$ such that $\mathrm{L}_{t_i}(\mathrm{f}_\rho) \neq 0$.

If Assumption 2 does not hold, then, for all $i \in \{0, 1, \dots, n_{\mathscr{D}} - 1\}$, we have that $\mathrm{L}_{t_i}(\mathrm{f}_\rho) = 0$, which means that the dominant pole is not excited by the input signal u. Accordingly, this assumption essentially says that the input signal excites the dominant pole.

Assumption 3. There exists $\underline{t} \leq 0$ such that $u_t = 0$ for any $t < \underline{t}$.

Assumption 3 is a technical assumption and introduced mainly for the sake of our mathematical arguments, i.e., to guarantee the continuity of linear operator L_{t_i}: $\mathcal{H}_{\mathbb{k}} \to \mathbb{R}$ for $i \in \{0, 1, \dots, n_{\mathscr{D}} - 1\}$. Indeed, this assumption holds in realistic situations, such as when the system is initially at rest. Based on these assumptions, we can show the existence and uniqueness for the solution of (4.7).

THEOREM 4.1. *Under Assumptions* 2 *and* 3, *optimization problem* (4.7) *admits a unique solution; i.e., there exist* $(a^*, \mathrm{h}^*) \in \mathcal{V}_{\mathbb{k}}$ *such that*

$$\mathcal{J}(a^*, \mathrm{h}^*) < \mathcal{J}(a, \mathrm{h}) \qquad \forall\, (a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}} \backslash \big\{(a^*, \mathrm{h}^*)\big\}. \tag{4.9}$$

Proof. Let set $\mathscr{R} \subset \mathcal{V}_{\mathbb{k}}$ be defined as $\mathscr{R} = \bigcap_{s=0}^{\infty} \mathscr{R}_s$. Accordingly, one can see that $\sum_{s=0}^{\infty} \delta_{\mathscr{R}_s} = \delta_{\mathscr{R}}$, and hence, we have that

$$\mathcal{J}(a, \mathrm{h}) = \mathcal{E}_\rho(a, \mathrm{h}) + \lambda \|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2 + \delta_{\mathscr{R}}(a, \mathrm{h}). \tag{4.10}$$

With respect to each $s \in \mathbb{Z}_+$, define set $\mathcal{Q}_s \subset \mathcal{V}_{\mathbb{k}}$ as $\mathcal{Q}_s := \{(a, (h_t)_{t \in \mathbb{Z}_+}) \in \mathcal{V}_{\mathbb{k}} \,|\, h_s + a\rho^s \geq 0\}$. For any $a \in \mathbb{R}$, $\mathrm{h} = (h_t)_{t \in \mathbb{Z}_+} \in \mathcal{H}_{\mathbb{k}}$, and $s \in \mathbb{Z}_+$, due to the reproducing property of the kernel, we have that

$$h_s + a\rho^s = \langle \mathrm{h}, \mathbb{k}_s \rangle_{\mathcal{H}_{\mathbb{k}}} + a\rho^s = \langle (a, \mathrm{h}), (\rho^s, \mathbb{k}_s) \rangle_{\mathcal{V}_{\mathbb{k}}}. \tag{4.11}$$

Therefore, we know that $\mathcal{Q}_s \subset \mathcal{V}_{\mathbb{k}}$ is a half-space, and hence, it is a nonempty, closed, and convex subset of $\mathcal{V}_{\mathbb{k}}$, for all $s \in \mathbb{Z}_+$. Note that $[a_{\min}, \infty) \times \mathcal{H}_{\mathbb{k}}$ is a nonempty, closed, and convex subset of $\mathcal{V}_{\mathbb{k}}$. One can see that $\mathscr{R} = \big(\cap_{s=0}^{\infty} \mathcal{Q}_s\big) \cap \big([a_{\min}, \infty) \times \mathcal{H}_{\mathbb{k}}\big)$, and also, we know that $(a_{\min}, \mathbf{0})$ belongs to $[a_{\min}, \infty) \times \mathcal{H}_{\mathbb{k}}$ and $\mathcal{Q}_s$ for each $s \in \mathbb{Z}_+$. Therefore, $\mathscr{R}$ is a nonempty, closed, and convex subset of $\mathcal{V}_{\mathbb{k}}$. Consequently, it follows that $\delta_{\mathscr{R}} : \mathcal{V} \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex, and lower semicontinuous function [51], where we have that $\delta_{\mathscr{R}}(a_{\min}, \mathbf{0}) = 0$. With respect to each $i \in \{0, \dots, n_{\mathscr{D}} - 1\}$, define φ_i as

$$\varphi_i := \sum_{s=0}^{t_i - \underline{t}} u_{t_i - s} \mathbb{k}_s = u_{t_i} \mathbb{k}_0 + u_{t_i - 1} \mathbb{k}_1 + \dots + u_{\underline{t}} \mathbb{k}_{t_i - \underline{t}}. \tag{4.12}$$

Since $\mathcal{H}_{\mathbb{k}}$ is a linear space that contains the sections of the kernel, we know that $\varphi_i \in \mathcal{H}_{\mathbb{k}}$ for each $i \in \{0, 1, \dots, n_{\mathscr{D}} - 1\}$. Moreover, from Assumption 3, the reproducing property of the kernel, and the linearity property of the inner product, it follows that

$$\mathrm{L}_{t_i}(\mathrm{h}) = \sum_{s=0}^{t_i - \underline{t}} h_s u_{t_i - s} = \sum_{s=0}^{t_i - \underline{t}} \langle \mathrm{h}, \mathbb{k}_s \rangle_{\mathcal{H}_{\mathbb{k}}} u_{t_i - s} = \left\langle \mathrm{h}, \sum_{s=0}^{t_i - \underline{t}} \mathbb{k}_s u_{t_i - s} \right\rangle_{\mathcal{H}_{\mathbb{k}}} = \langle \mathrm{h}, \varphi_i \rangle_{\mathcal{H}_{\mathbb{k}}}. \tag{4.13}$$

Consequently, we have that

$$a\mathrm{L}_{t_i}(\mathrm{f}_\rho)+\mathrm{L}_{t_i}(\mathrm{h})=a\mathrm{L}_{t_i}(\mathrm{f}_\rho)+\langle \mathrm{h},\varphi_i\rangle_{\mathcal{H}_{\mathbb{k}}}=\langle (a,\mathrm{h}),\psi_i\rangle_{\mathcal{V}_{\mathbb{k}}}, \tag{4.14}$$

where $\psi_i\in\mathcal{V}_{\mathbb{k}}$ is the vector defined as $\psi_i=(\mathrm{L}_{t_i}(\mathrm{f}_\rho),\varphi_i)$ for $i=0,1,\dots,n_{\mathscr{D}}-1$. Therefore, one can see that

$$\mathcal{J}(a,\mathrm{h})=\sum_{i=0}^{n_{\mathscr{D}}-1}\big(y_{t_i}-\langle (a,\mathrm{h}),\psi_i\rangle_{\mathcal{V}_{\mathbb{k}}}\big)^2+\lambda\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2+\delta_{\mathscr{R}}(a,\mathrm{h}).$$

Accordingly, since $\delta_{\mathscr{R}}$ is proper, convex, and lower semicontinuous and also due to the fact that $\mathcal{J}(a_{\min},\mathbf{0})=\sum_{i=1}^{n}\big(y_{t_i}-a_{\min}\mathrm{L}_{t_i}(\mathrm{f}_\rho)\big)^2<\infty$, we know that $\mathcal{J}$ is a proper, convex, and lower semicontinuous function. This implies that (4.7) has a solution [51]. Define bilinear operator $\mathrm{Q}:\mathcal{V}_{\mathbb{k}}\times\mathcal{V}_{\mathbb{k}}\to\mathbb{R}$ as

$$\mathrm{Q}\big((a_1,\mathrm{h}_1),(a_2,\mathrm{h}_2)\big)=\sum_{i=0}^{n_{\mathscr{D}}-1}\langle (a_1,\mathrm{h}_1),\psi_i\rangle_{\mathcal{V}_{\mathbb{k}}}\langle (a_2,\mathrm{h}_2),\psi_i\rangle_{\mathcal{V}_{\mathbb{k}}}+\lambda\langle \mathrm{h}_1,\mathrm{h}_2\rangle_{\mathcal{H}_{\mathbb{k}}}. \tag{4.15}$$

For any $(a,\mathrm{h})\in\mathcal{V}_{\mathbb{k}}$, one can easily see that

$$\mathrm{Q}\big((a,\mathrm{h}),(a,\mathrm{h})\big)=\sum_{i=0}^{n_{\mathscr{D}}-1}\langle (a,\mathrm{h}),\psi_i\rangle_{\mathcal{V}_{\mathbb{k}}}^2+\lambda\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2\geq 0. \tag{4.16}$$

Moreover, since λ is a positive real scalar, if $\mathrm{Q}\big((a,\mathrm{h}),(a,\mathrm{h})\big)=0$, then we need to have $\mathrm{h}=0$ and $\langle (a,\mathrm{h}),\psi_i\rangle_{\mathcal{V}_{\mathbb{k}}}=0$ for all $i=0,1,\dots,n_{\mathscr{D}}-1$, which implies that $a\mathrm{L}_{t_i}(\mathrm{f}_\rho)=0$. Subsequently, due to Assumption 2, we have that $a=0$; i.e., $(a,\mathrm{h})=(0,\mathbf{0})$. Based on this argument, we know that Q is a positive definite bilinear operator. Therefore, the function $f:\mathcal{V}_{\mathbb{k}}\to\mathbb{R}$, defined as $f(\mathrm{v})=\mathrm{Q}(\mathrm{v},\mathrm{v})$ for all $\mathrm{v}\in\mathcal{V}_{\mathbb{k}}$, is strictly convex [51]. Note that we have that

$$\mathcal{J}(a,\mathrm{h})=f(a,\mathrm{h})-2\mathrm{L}(a,\mathrm{h})+\sum_{i=0}^{n_{\mathscr{D}}-1}y_{t_i}^2+\delta_{\mathscr{R}}(a,\mathrm{h}), \tag{4.17}$$

where $\mathrm{L}:\mathcal{V}_{\mathbb{k}}\to\mathbb{R}$ is the bounded linear operator defined as

$$\mathrm{L}(a,\mathrm{h})=\sum_{i=0}^{n_{\mathscr{D}}-1}y_{t_i}\langle (a,\mathrm{h}),\psi_i\rangle_{\mathcal{V}_{\mathbb{k}}}. \tag{4.18}$$

Since f is strictly convex, L is linear, and $\delta_{\mathscr{R}}$ is convex, it follows that $\mathcal{J}:\mathcal{V}_{\mathbb{k}}\to\mathbb{R}$ is a strictly convex function, and consequently, the solution of optimization problem (4.7) is unique [51]. ∎

Due to Theorem 4.1, we know that the convex program (4.7) has a unique solution (a^*,h^*). Meanwhile, one should note that (4.7) is an infinite-dimensional optimization problem with an infinite number of constraints. Thus, obtaining the solution (a^*,h^*) is not straightforward. On the other hand, since h^* belongs to the set of stable impulse responses $\mathcal{H}_{\mathbb{k}}$ dominated by f_ρ, one may intuitively expect that $\mathrm{h}^*\in\mathscr{R}_m$ (see (4.3)) when $m\in\mathbb{Z}_+$ is large enough. In other words, the solution to the optimization problem (4.7) is determined by a finite number of constraints, and the remaining constraints are unnecessary. In order to formalize this idea, let function $\mathcal{J}_m:\mathcal{V}_{\mathbb{k}}\to\mathbb{R}\cup\{+\infty\}$ be defined as $\mathcal{J}_m(a,\mathrm{h})=\mathcal{E}_\rho(a,\mathrm{h})+\sum_{s=0}^{m}\delta_{\mathscr{R}_s}(a,\mathrm{h})+\lambda\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2$, and consider the following program:

$$\inf_{(a,\mathrm{h})\in\mathcal{V}_{\mathbb{k}}} \mathcal{J}_m(a,\mathrm{h}). \tag{4.19}$$

Note that (4.19) is equivalent to

$$\begin{aligned} \min_{a\in\mathbb{R},\mathrm{h}\in\mathcal{H}_{\mathbb{k}}} \quad & \sum_{i=0}^{n_{\mathscr{D}}-1}\big(y_{t_i}-a\mathrm{L}_{t_i}(\mathrm{f}_\rho)-\mathrm{L}_{t_i}(\mathrm{h})\big)^2+\lambda\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2 \\ \text{s.t.} \quad & h_t+a\rho^t\geq 0 \quad \forall\, t\in\{0,1,\dots,m\}, \\ & a\geq a_{\min}. \end{aligned} \tag{4.20}$$

The next theorem guarantees the existence and uniqueness of solution for optimization problem (4.19), or equivalently, for program (4.20).

THEOREM 4.2. *Under the assumptions of Theorem* 4.1*, for each* $m\in\mathbb{Z}_+$*, problem* (4.19) *admits a unique solution* $(a^{(m)},\mathrm{h}^{(m)})$.

Proof. Define set $\mathscr{P}^{(m)}\subset\mathcal{V}_{\mathbb{k}}$ as $\mathscr{P}^{(m)}=\bigcap_{s=0}^m\mathscr{R}_s$. By replacing $\mathscr{P}$ with $\mathscr{P}^{(m)}$ in the proof of Theorem 4.1 and then repeating the same steps, the claim follows. ◻

Once the existence and uniqueness for the solution of (4.19) are established by Theorem 4.2, a reasonable concern is the asymptotic behavior of $(a^{(m)},\mathrm{h}^{(m)})$, especially with respect to (a^*,h^*). The next theorem reveals this link, saying that the solution $(a^{(m)},\mathrm{h}^{(m)})$ coincides with (a^*,h^*) when m is large enough. Before proceeding further, we need to introduce additional definitions. Define a_0 as $a_0:=\operatorname{argmin}_{a\geq a_{\min}}\mathcal{J}(a,\mathbf{0})$. Since $\mathcal{J}(\cdot,\mathbf{0}):\mathbb{R}\to\mathbb{R}$ is a quadratic function because

$$\mathcal{J}(a,\mathbf{0})=\sum_{i=0}^{n_{\mathscr{D}}-1}\big(y_{t_i}-a\mathrm{L}_{t_i}(\mathrm{f}_\rho)\big)^2 \qquad \forall a\in\mathbb{R}, \tag{4.21}$$

one can easily see that

$$a_0=\min\left\{a_{\min},\left(\sum_{i=0}^{n_{\mathscr{D}}-1}\mathrm{L}_{t_i}(\mathrm{f}_\rho)^2\right)^{-1}\sum_{i=0}^{n_{\mathscr{D}}-1}y_{t_i}\mathrm{L}_{t_i}(\mathrm{f}_\rho)\right\}, \tag{4.22}$$

which is well defined following Assumption 2. Additionally, we define C_0 and m_0, respectively, as $C_0:=\mathcal{J}(a_0,\mathbf{0})$ and

$$m_0:=\min\left\{m\in\mathbb{Z}_+\ \middle|\ m\geq\frac{1}{2}\frac{\ln(C_0C)-\ln(a_{\min}^2\lambda)}{\ln(\rho)-\ln(\rho_{\mathrm{d}})}\right\}. \tag{4.23}$$

Note that m_0 is defined only based on the data and a priori known constants.

THEOREM 4.3. *Under Assumptions* 2 *and* 3*, the following statements hold:*

i) *Given Assumption* 1*, there exists* $m\in\mathbb{Z}_+$ *such that* $a^{(m)}=a^*$ *and* $\mathrm{h}^{(m)}=\mathrm{h}^*$.

ii) *For* $m\in\mathbb{Z}_+$*, one has that* $(a^{(m)},\mathrm{h}^{(m)})=(a^*,\mathrm{h}^*)$ *if and only if* $\mathcal{J}(a^{(m)},\mathrm{h}^{(m)})<\infty$.

iii) *If* $(a^{(m)},\mathrm{h}^{(m)})=(a^*,\mathrm{h}^*)$*, for a nonnegative integer* m*, then* $a^{(\overline{m})}=a^*$ *and* $\mathrm{h}^{(\overline{m})}=\mathrm{h}^*$ *for every* $\overline{m}\geq m$.

Proof. Part i) For any $a\geq a_{\min}$ and any $m\in\mathbb{Z}_+$, we have that

$$\mathcal{J}_m(a,\mathbf{0})=\mathcal{J}(a,\mathbf{0})=\sum_{i=1}^{n}\big(y_{t_i}-a\mathrm{L}_{t_i}(\mathrm{f}_\rho)\big)^2\in[0,\infty). \tag{4.24}$$

Note that $C_0 = \mathcal{J}_m(a_0, \mathbf{0})$ for any $m \in \mathbb{Z}_+$. If $C_0 = 0$, then $(a_0, \mathbf{0})$ needs to be a solution for optimization problems (4.7) and (4.19), where, because of their uniqueness, we have that $(a_0, \mathbf{0}) = (a^{(m)}, \mathrm{h}^{(m)}) = (a^*, \mathrm{h}^*)$. Now, we consider the case $C_0 > 0$. Based on the definition of m_0, one can easily see that

$$a_{\min}\rho^s - \lambda^{-\frac{1}{2}} C_0^{\frac{1}{2}} C^{\frac{1}{2}} \rho_{\mathrm{d}}^s \geq 0 \tag{4.25}$$

for any $s \geq m_0$. Let m be an arbitrary integer such that $m \geq m_0$, and consider the convex program (4.19) with unique solution $(a^{(m)}, \mathrm{h}^{(m)})$. We know that $h_s^{(m)} + a^{(m)}\rho^s \geq 0$ for each $s = 0, 1, \ldots, m$. On the other hand, for $s > m$, from the reproducing property and the Cauchy–Schwartz inequality, it follows that

$$h_s^{(m)} + a^{(m)}\rho^s = \langle \mathrm{h}^{(m)}, \mathbb{k}_s \rangle_{\mathcal{H}_{\mathbb{k}}} + a^{(m)}\rho^s \geq -\|\mathrm{h}^{(m)}\|_{\mathcal{H}_{\mathbb{k}}} \mathbb{k}(s,s)^{\frac{1}{2}} + a_{\min}\rho^s. \tag{4.26}$$

Note that, due to $\mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}) \leq \mathcal{J}_m(a_0, \mathbf{0})$, one has that $\lambda \|\mathrm{h}^{(m)}\|_{\mathcal{H}_{\mathbb{k}}}^2 \leq \mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}) \leq C_0$, which implies that $\|\mathrm{h}^{(m)}\|_{\mathcal{H}_{\mathbb{k}}} \leq \lambda^{-\frac{1}{2}} C_0^{\frac{1}{2}}$. Hence, according to (4.25), (4.26), and Assumption 1, we have that $h_s^{(m)} + a^{(m)}\rho^s \geq a_{\min}\rho^s - \lambda^{-\frac{1}{2}} C_0^{\frac{1}{2}} C^{\frac{1}{2}} \rho_{\mathrm{d}}^s \geq 0$, which implies that $h_s^{(m)} + a^{(m)}\rho^s \geq 0$ for all $s \in \mathbb{Z}_+$. Therefore, one can see that $\sum_{s=0}^{\infty} \delta_{\mathscr{R}_s}(a^{(m)}, \mathrm{h}^{(m)}) = 0$, and subsequently, we have that $\mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)})$. On the other hand, according to the definition of (a^*, h^*) and $(a^{(m)}, \mathrm{h}^{(m)})$, we know that $\mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}) \leq \mathcal{J}_m(a^*, \mathrm{h}^*)$ and $\mathcal{J}(a^*, \mathrm{h}^*) \leq \mathcal{J}(a^{(m)}, \mathrm{h}^{(m)})$. Accordingly, since $\mathcal{J}_m(a, \mathrm{h}) \leq \mathcal{J}(a, \mathrm{h})$ for all $(a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}}$, one can see that

$$\mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}) \leq \mathcal{J}_m(a^*, \mathrm{h}^*) \leq \mathcal{J}(a^*, \mathrm{h}^*) \leq \mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}). \tag{4.27}$$

Hence, we have that $\mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}_m(a^*, \mathrm{h}^*)$, and subsequently, due to Theorem 4.2, one has $(a^{(m)}, \mathrm{h}^{(m)}) = (a^*, \mathrm{h}^*)$. This concludes the proof of Part i).

Part ii) Consider the case that $\mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) < \infty$. This implies that $\sum_{s \geq 0} \delta_{\mathscr{R}_s}(a^{(m)}, \mathrm{h}^{(m)}) = 0$, and consequently, $\mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)})$. One can see that $\mathcal{J}_m(a, \mathrm{h}) \leq \mathcal{J}_{\overline{m}}(a, \mathrm{h}) \leq \mathcal{J}(a, \mathrm{h})$ for any $m \leq \overline{m}$ and each $(a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}}$. Accordingly, due to the definition of (a^*, h^*) and $(a^{(m)}, \mathrm{h}^{(m)})$, we have that

$$\mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}) \leq \mathcal{J}_m(a^*, \mathrm{h}^*) \leq \mathcal{J}(a^*, \mathrm{h}^*) \leq \mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}), \tag{4.28}$$

which implies that (a^*, h^*) is a solution for (4.19). Since this solution is unique according to Theorem 4.2, we need to have $(a^{(m)}, \mathrm{h}^{(m)}) = (a^*, \mathrm{h}^*)$. The converse is straightforward and concludes the proof of Part ii).

Part iii) From the previous part, we know that $\mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) < \infty$. Consequently, we have that $\sum_{s \geq 0} \delta_{\mathscr{R}_s}(a^{(m)}, \mathrm{h}^{(m)}) = 0$, which implies that $\mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}_{\overline{m}}(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}(a^{(m)}, \mathrm{h}^{(m)})$. Accordingly, due to the definition of $(a^{(m)}, \mathrm{h}^{(m)})$ and $(a^{(\overline{m})}, \mathrm{h}^{(\overline{m})})$, we have that

$$\mathcal{J}_{\overline{m}}(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}_m(a^{(m)}, \mathrm{h}^{(m)}) \leq \mathcal{J}_m(a^{(\overline{m})}, \mathrm{h}^{(\overline{m})}) \leq \mathcal{J}_{\overline{m}}(a^{(\overline{m})}, \mathrm{h}^{(\overline{m})}). \tag{4.29}$$

Therefore, we know that $(a^{(m)}, \mathrm{h}^{(m)})$ is the unique solution of optimization problem $\inf_{(a,\mathrm{h}) \in \mathcal{V}_{\mathbb{k}}} \mathcal{J}_{\overline{m}}(a, \mathrm{h})$, and consequently, we have that $(a^{(\overline{m})}, \mathrm{h}^{(\overline{m})}) = (a^{(m)}, \mathrm{h}^{(m)}) = (a^*, \mathrm{h}^*)$, where the second equality holds by assumption. This concludes the proof of Part iii) and the proof of Theorem 4.3. $\square$

The following observation is a direct result of Theorem 4.3.

COROLLARY 4.4. *Under the assumptions of Theorem* 4.3*, there exists a nonnegative integer* m^* *such that* $(a^{(m)},\mathrm{h}^{(m)}) = (a^*,\mathrm{h}^*)$ *if and only if* $m \geq m^*$*. Indeed, for any* $m < m^*$*, there exists* $s \in \mathbb{Z}_+$ *such that* $h_s^{(m)} + a^{(m)}\rho^s < 0$*. This implies that*

$$m^* = \min\left\{m \in \mathbb{Z}_+ \,\middle|\, \mathcal{J}(a^{(m)},\mathrm{h}^{(m)}) < \infty\right\}. \tag{4.30}$$

Moreover, for m_0 *introduced in* (4.23)*, we have that* $m^* \leq m_0$*.*

COROLLARY 4.5. *Let* $m \in \mathbb{Z}_+$ *be such that the impulse response* $\mathrm{g}^{(m)} := (h_s^{(m)} + a^{(m)}\rho^s)_{s=0}^\infty$ *is not nonnegative; i.e.,* $m < m^*$*. Then, based on the proof of Theorem* 4.3*, one can see that there exists* $\underline{s} \in \mathbb{Z}_+$ *such that* $g_{\underline{s}} < 0$ *and* $s \leq m_0$*, where* m_0 *is introduced in* (4.23)*.*

Due to Theorem 4.3 and Corollary 4.4, it suffices to consider only a finite number of constraints in optimization problem (4.7) as in (4.19). The feasible set of this optimization problem is of infinite dimension, which makes the problem intractable in the current format. In the remainder of this section, we derive a practical heuristic for obtaining the solution of (3.16) using the *representer theorem* [19, 35, 60] and an additional definition. Before proceeding to the next theorem, we define the matrices $\mathrm{O} \in \mathbb{R}^{n_{\mathscr{D}} \times n_{\mathscr{D}}}$, $\mathrm{L} \in \mathbb{R}^{n_{\mathscr{D}} \times (m+1)}$, and $\mathrm{K} \in \mathbb{R}^{(m+1)\times(m+1)}$, respectively, as

$$\begin{aligned} &\mathrm{O}(i,j) = \mathrm{L}_{t_{i-1}}(\mathrm{L}_{t_{j-1}}(\mathbb{k})), && 1 \leq i,j \leq n_{\mathscr{D}}, \\ &\mathrm{L}(i,j) = \mathrm{L}_{t_{i-1}}(\mathbb{k}_{j-1}), && 1 \leq i \leq n_{\mathscr{D}}, 1 \leq j \leq m+1, \\ &\mathrm{K}(i,j) = \mathbb{k}(i-1,j-1), && 1 \leq i,j \leq m+1. \end{aligned} \tag{4.31}$$

Also, the vectors $\mathrm{y} \in \mathbb{R}^{n_{\mathscr{D}}}$, $\mathrm{b} \in \mathbb{R}^{n_{\mathscr{D}}}$, and $\mathrm{c} \in \mathbb{R}^{m+1}$ are defined, respectively, as $\mathrm{y} := [y_{t_i}]_{i=0}^{n_{\mathscr{D}}-1}$, $\mathrm{b} := [\mathrm{L}_{t_i}(\mathrm{f}_\rho)]_{i=0}^{n_{\mathscr{D}}-1}$, and $\mathrm{c} := [\rho^j]_{j=0}^m$.

THEOREM 4.6. *Let Assumption* 3 *hold. Then, for any nonnegative integer* m*, there exists* $\mathrm{x}^{(m)} = [x_0^{(m)}, \ldots, x_{n_{\mathscr{D}}+m}^{(m)}]^\mathsf{T} \in \mathbb{R}^{n_{\mathscr{D}}+m+1}$ *such that the unique solution of* (4.20)*,* $\mathrm{h}^{(m)} = (h_t^{(m)})_{t=0}^\infty$*, admits the following parametric representation:*

$$h_t^{(m)} = \sum_{i=0}^{n_{\mathscr{D}}} x_i^{(m)} \mathrm{L}_{t_i}(\mathbb{k}_t) + \sum_{s=0}^{m} x_{n_{\mathscr{D}}+s}^{(m)} \mathbb{k}_s(t) \qquad \forall t \in \mathbb{Z}_+. \tag{4.32}$$

Moreover, $(a^{(m)},\mathrm{x}^{(m)})$ *is the solution of the following convex QP:*

$$\begin{aligned} \min_{a\in\mathbb{R},\ \mathrm{x}\in\mathbb{R}^{n_{\mathscr{D}}+m+1}} \quad & \left\|\mathrm{y} - \mathrm{b}a - \begin{bmatrix}\mathrm{O} & \mathrm{L}\end{bmatrix}\mathrm{x}\right\|^2 + \lambda \mathrm{x}^\mathsf{T}\begin{bmatrix}\mathrm{O} & \mathrm{L} \\ \mathrm{L}^\mathsf{T} & \mathrm{K}\end{bmatrix}\mathrm{x} \\ \text{s.t.} \quad & \begin{bmatrix}\mathrm{L}^\mathsf{T} & \mathrm{K}\end{bmatrix}\mathrm{x} + \mathrm{c}a \geq 0, \\ & a \geq a_{\min}. \end{aligned} \tag{4.33}$$

Proof. For $s = 0,\ldots,m$, let $\mathcal{A}_s$ be the set

$$\mathcal{A}_s = \left\{(a,x) \in \mathbb{R}^2 \,\middle|\, x + a\rho^s \geq 0, a \geq a_{\min}\right\}, \tag{4.34}$$

and define function $e : \mathbb{R}^{n_{\mathscr{D}}+m+1} \to \mathbb{R} \cup \{+\infty\}$ such that, for any $\mathrm{x} \in \mathbb{R}^{n_{\mathscr{D}}+m+1}$, we have

$$e(x_0,\ldots,x_{n_{\mathscr{D}}+m}) = \min_{a\in\mathbb{R}} \left[\sum_{i=0}^{n_{\mathscr{D}}-1} \left(y_{t_i} - a\mathrm{L}_{t_i}(\mathrm{f}_\rho) - x_i\right) + \sum_{s=0}^{m} \delta_{\mathcal{A}_s}(a, x_{n_{\mathscr{D}}+s})\right]. \tag{4.35}$$

Also, for $i = 0,\ldots,n_{\mathscr{D}}-1$ and $s = 0,\ldots,m$, let φ_i and $\varphi_{n_{\mathscr{D}}+s}$ be defined, respectively, as in (4.12) and $\varphi_{n_{\mathscr{D}}+s} = \mathbb{k}_s$. Due to the reproducing property and (4.13), we know that $\mathrm{L}_{t_i}(\mathrm{h}) = \langle \mathrm{h}, \varphi_i \rangle_{\mathcal{H}_{\mathbb{k}}}$ and $h_s = \langle \mathrm{h}, \mathbb{k}_s \rangle_{\mathcal{H}_{\mathbb{k}}}$ for any $\mathrm{h} = (h_s)_{s=0}^{\infty}$. Accordingly, due to (4.35), one can see that (4.19) is equivalent to the optimization problem

$$\min_{\mathrm{h}\in\mathcal{H}_{\mathbb{k}}} e\left(\langle \mathrm{h}, \varphi_0 \rangle_{\mathcal{H}_{\mathbb{k}}}, \dots, \langle \mathrm{h}, \varphi_{n_{\mathscr{D}}+m} \rangle_{\mathcal{H}_{\mathbb{k}}}\right) + \lambda \|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2 \tag{4.36}$$

with unique solution $\mathrm{h}^{(m)}$. Therefore, due to the representer theorem [19], we know that $\mathrm{h}^{(m)}$ belongs to the span of $\varphi_0, \varphi_1, \dots, \varphi_{n_{\mathscr{D}}+m}$; i.e., there exists $\mathrm{x}^{(m)} = [x_i^{(m)}]_{i=0}^{n_{\mathscr{D}}+m} \in \mathbb{R}^{n_{\mathscr{D}}+m+1}$ such that

$$\mathrm{h}^{(m)} = \sum_{i=0}^{n_{\mathscr{D}}+m} x_i^{(m)} \varphi_i = \sum_{i=0}^{n_{\mathscr{D}}-1} x_i^{(m)} \varphi_i + \sum_{s=0}^{m} x_{n_{\mathscr{D}}+s}^{(m)} \mathbb{k}_s. \tag{4.37}$$

Due to the reproducing property, we know that $h_t^{(m)} = \langle \mathrm{h}^{(m)}, \mathbb{k}_t \rangle_{\mathcal{H}_{\mathbb{k}}}$ for any $t \in \mathbb{Z}_+$. Accordingly, from (4.13), (4.37), the linearity property of the inner product, and the reproducing property, we have that

$$\begin{aligned} h_t^{(m)} &= \left\langle \sum_{i=0}^{n_{\mathscr{D}}-1} x_i^{(m)} \varphi_i + \sum_{s=0}^{m} x_{n_{\mathscr{D}}+s}^{(m)} \mathbb{k}_s, \mathbb{k}_t \right\rangle_{\mathcal{H}_{\mathbb{k}}} \\ &= \sum_{i=0}^{n_{\mathscr{D}}-1} x_i^{(m)} \langle \varphi_i, \mathbb{k}_t \rangle_{\mathcal{H}_{\mathbb{k}}} + \sum_{s=0}^{m} x_{n_{\mathscr{D}}+s}^{(m)} \langle \mathbb{k}_s, \mathbb{k}_t \rangle_{\mathcal{H}_{\mathbb{k}}} \\ &= \sum_{i=0}^{n_{\mathscr{D}}-1} x_i^{(m)} \mathrm{L}_{t_i}(\mathbb{k}_t) + \sum_{s=0}^{m} x_{n_{\mathscr{D}}+s}^{(m)} \mathbb{k}_s(t). \end{aligned} \tag{4.38}$$

Moreover, for $j = 0, \dots, n_{\mathscr{D}} - 1$, we have that

$$\begin{aligned} \mathrm{L}_{t_j}(\mathrm{h}^{(m)}) &= \mathrm{L}_{t_j}\left(\sum_{i=0}^{n_{\mathscr{D}}-1} x_i^{(m)} \varphi_i + \sum_{s=0}^{m} x_{n_{\mathscr{D}}+s}^{(m)} \mathbb{k}_s \right) \\ &= \sum_{i=0}^{n_{\mathscr{D}}-1} x_i^{(m)} \mathrm{L}_{t_j}(\mathrm{L}_{t_i}(\mathbb{k})) + \sum_{s=0}^{m} x_{n_{\mathscr{D}}+s}^{(m)} \mathrm{L}_{t_j}(\mathbb{k}_s). \end{aligned} \tag{4.39}$$

Considering optimization problem (4.20), which is equivalent to (4.19), we replace h with the parametric form given in (4.37). Hence, due to (4.31), (4.38), (4.39), and the definition of vectors b, c, and $\mathrm{h}^{(m)}$, the optimization problem (4.33) follows. ∎

Remark 4.7. Let the system be initially at rest and the sampling times be $\mathscr{T} = \{0, 1, \dots, n_{\mathscr{D}} - 1\}$. With respect to each $n_1, n_2 \in \mathbb{Z}_+$, we define matrix $\mathrm{K}_{n_1,n_2} \in \mathbb{R}^{n_1 \times n_2}$ such that $\mathrm{K}_{n_1,n_2}(i,j) = \mathbb{k}(i-1, j-1)$ for $i = 1, \dots, n_1$ and $j = 1, \dots, n_2$. Then, one can easily see that $\mathrm{O} = \mathrm{T_u} \mathrm{K}_{n_{\mathscr{D}},n_{\mathscr{D}}} \mathrm{T_u}^{\mathsf{T}}$, $\mathrm{L} = \mathrm{T_u} \mathrm{K}_{n_{\mathscr{D}},m+1}$, and $\mathrm{K} = \mathrm{K}_{m+1,m+1}$, where $\mathrm{T_u} \in \mathbb{R}^{n \times n}$ is the Toeplitz matrix defined as $\mathrm{T_u} = [u_{i-j}]_{i,j=1}^{n_{\mathscr{D}}}$.

Theorem 4.6 offers a practical way to solve problem (4.20). Due to Theorem 4.3 and Corollary 4.4, we know that this solution coincides with the solution of (4.6) provided that $m \geq m^*$. Nevertheless, compared to (4.6), the main optimization problem (4.2) has an additional constraint on the rank of resulting Hankel operator being finite. In the remainder of this section, we fill this gap by employing the notion of finite Hankel rank kernels.

DEFINITION 4.8. *We call the kernel* $\mathbb{k} : \mathbb{Z}_+ \times \mathbb{Z}_+$ *a* finite Hankel rank *if, for each* $s \in \mathbb{Z}_+$*, the section of the kernel at* s *is a finite Hankel rank impulse response; i.e.,*

$$\operatorname{rank}(\operatorname{Hankel}(\mathbb{k}_s)) < \infty \quad \forall\, s \in \mathbb{Z}_+. \tag{4.40}$$

The standard stable kernels in the literature [18, 53] are *tuned/correlated* (TC), *diagonal/correlated* (DC), and *stable spline* (SS), which are, respectively, denoted by $\mathbb{k}_{\mathrm{TC}}$, $\mathbb{k}_{\mathrm{DC}}$, and $\mathbb{k}_{\mathrm{SS}}$ and, for any $s,t \in \mathbb{Z}_+$, defined as

$$\mathbb{k}_{\mathrm{TC}}(s,t) = \beta^{\max(s,t)} \quad \forall s,t \in \mathbb{Z}_+, \tag{4.41}$$

$$\mathbb{k}_{\mathrm{DC}}(s,t) = \beta^{\frac{s+t}{2}} \gamma^{|s-t|} \quad \forall s,t \in \mathbb{Z}_+, \tag{4.42}$$

$$\mathbb{k}_{\mathrm{SS}}(s,t) = \frac{1}{2}\beta^{s+t+\max(s,t)} - \frac{1}{6}\beta^{3\max(s,t)} \quad \forall s,t \in \mathbb{Z}_+, \tag{4.43}$$

where $\beta \in [0,1)$ and $\gamma \in [-1,1]$ are the corresponding hyperparameters. Furthermore, the second-order generalizations of the TC and DC kernels, respectively, denoted by $\mathbb{k}_{\mathrm{TC2}}$ and $\mathbb{k}_{\mathrm{DC2}}$, have been recently introduced in [71] as

$$\mathbb{k}_{\mathrm{TC2}}(s,t) = 2\beta^{\max(s,t)+1} + (1-\beta)(1+|t-s|)\beta^{\max(s,t)} \quad \forall s,t \in \mathbb{Z}_+, \tag{4.44}$$

$$\mathbb{k}_{\mathrm{DC2}}(s,t) = \frac{1}{1-\xi}\beta^{\max(s,t)}(1-(1-\beta)\xi^{|s-t|+1}) - \frac{\xi^2}{1-\xi}\beta^{\max(s,t)+1} \quad \forall s,t \in \mathbb{Z}_+ \tag{4.45}$$

with hyperparameters $\beta, \xi \in [0,1)$. As shown by the next theorem, these common kernels are finite Hankel rank.

THEOREM 4.9. *The finite support kernels,* $\mathbb{k}_{\mathrm{TC}}$*,* $\mathbb{k}_{\mathrm{DC}}$*,* $\mathbb{k}_{\mathrm{TC2}}$*,* $\mathbb{k}_{\mathrm{DC2}}$*, and* $\mathbb{k}_{\mathrm{SS}}$ *are finite Hankel rank kernels.*

Proof. Let c_{00} be the space of impulse responses that are finitely nonzero; i.e., for each $\mathrm{g} = (g_s)_{s=0}^{\infty}$, there exists $n_{\mathrm{g}} \in \mathbb{Z}_+$ such that $g_s = 0$ for all $s \geq n_{\mathrm{g}}$. Note that, for such $\mathrm{g} \in c_{00}$, we have that $\operatorname{Hankel}(\mathrm{g})\mathrm{v} \in \mathbb{R}^{n_{\mathrm{g}}} \times \{\mathbf{0}\}$ for any $\mathrm{v} \in \ell^\infty$. This implies that $\operatorname{rank}\big(\operatorname{Hankel}(\mathrm{g})\big) \leq n_{\mathrm{g}} < \infty$, and therefore, g is a finite Hankel rank impulse response. Let $\mathbb{k} : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{R}$ be a finite support kernel; i.e., there exists $n_{\mathbb{k}} \in \mathbb{Z}_+$ such that $\mathbb{k}(s,t) = 0$ when $s \geq n_{\mathbb{k}}$ or $t \geq n_{\mathbb{k}}$. One can easily see that $\mathbb{k}_t \in c_{00}$ for any $t \in \mathbb{Z}_+$. Therefore, we have that $\operatorname{rank}\big(\operatorname{Hankel}(\mathbb{k}_t)\big) \leq n_{\mathbb{k}} < \infty$, and consequently, $\mathbb{k}$ is a finite Hankel rank kernel.

Let $\mathrm{f}_\beta = (f_t)_{t=0}^{\infty}$ be the impulse response defined as $f_t = \beta^t$ for $t \in \mathbb{Z}_+$. One can easily see that $\{\operatorname{Hankel}(\mathrm{f}_\beta)\mathrm{v} \mid \mathrm{v} \in \ell^\infty\} = \{\mathrm{f}_\beta v \mid v \in \mathbb{R}\}$. Therefore, we have that $\operatorname{rank}\big(\operatorname{Hankel}(\mathrm{f}_\beta)\big) = 1$, and consequently, f_β is a finite Hankel rank impulse response. For the TC kernel introduced in (4.41) and $t \in \mathbb{Z}_+$, consider the section of $\mathbb{k}_{\mathrm{TC}}$ at t, i.e., $(\mathbb{k}_{\mathrm{TC},t})_{s=0}^{\infty}$. Note that we have $\mathbb{k}_{\mathrm{TC},t} = \mathrm{f}_\beta + \mathrm{g}$, where the impulse response $\mathrm{g} = (g_s)_{s=0}^{\infty}$ is defined by $g_s = \beta^{\max(s,t)} - \beta^s$ for $s \in \mathbb{Z}_+$. One can easily see that $g_s = 0$ for all $s \geq t$, which implies that $\mathrm{g} \in c_{00}$, and subsequently, $\operatorname{rank}\big(\operatorname{Hankel}(\mathrm{g})\big) < \infty$. Accordingly, since the rank of $\operatorname{Hankel}(\mathrm{f}_\beta)$ is finite, it follows that $\operatorname{rank}\big(\operatorname{Hankel}(\mathbb{k}_{\mathrm{TC},t})\big) < \infty$, and consequently, $\mathbb{k}_{\mathrm{TC}}$ is a finite Hankel rank kernel. Based on similar arguments, one can show the same result for $\mathbb{k}_{\mathrm{DC}}$, $\mathbb{k}_{\mathrm{TC2}}$, $\mathbb{k}_{\mathrm{DC2}}$, and $\mathbb{k}_{\mathrm{SS}}$ and conclude the proof. ∎

Based on the notion of finite Hankel rank kernel and our previous discussion, we can show when the solution of our estimation problem (3.16) can be obtained by solving (4.33).

THEOREM 4.10. *Under the assumptions of Theorem* 4.3*, if kernel* $\mathbb{k}$ *is finite Hankel rank, then the unique solution of* (4.19) *satisfies*

$$\operatorname{rank}\big(\operatorname{Hankel}(\mathrm{h}^{(m)})\big) < \infty. \tag{4.46}$$

Moreover, $(a^{(m)}, \mathrm{h}^{(m)})$ is a solution of (3.16) *provided that $m \geq m^*$.*

Proof. Due to Theorem 4.6 and (4.32), we know that

$$\mathrm{h}^{(m)} = \sum_{i=0}^{n_{\mathscr{D}}-1} x_i^{(m)} \left(\sum_{s=0}^{t_i - \underline{t}} u_{t_i - s} \mathbb{k}_s \right) + \sum_{s=0}^{m} x_{n_{\mathscr{D}}+s}^{(m)} \mathbb{k}_s, \tag{4.47}$$

where $\mathrm{x}^{(m)} = [x_0^{(m)}, \dots, x_{n_{\mathscr{D}}+m}^{(m)}]^\mathsf{T} \in \mathbb{R}^{n_{\mathscr{D}}+m+1}$ is the solution of (4.33) (see (4.12) and (4.37)). Rearranging the terms in (4.47), one can see that there exist real scalars $\overline{x}_0, \dots, \overline{x}_{\overline{t}}$, where $\overline{t} = \max\{m, t_{n_{\mathscr{D}}-1} - \underline{t}\}$, such that we have that $\mathrm{h}^{(m)} = \overline{x}_0 \mathbb{k}_0 + \cdots + \overline{x}_{\overline{t}} \mathbb{k}_{\overline{t}}$. Therefore, we know that $\operatorname{Hankel}(\mathrm{h}^{(m)}) = \sum_{s=0}^{\overline{t}} \overline{x}_s \operatorname{Hankel}(\mathbb{k}_s)$, and subsequently, it follows that $\operatorname{rank}(\operatorname{Hankel}(\mathrm{h}^{(m)})) \leq \sum_{s=0}^{\overline{t}} \operatorname{rank}(\operatorname{Hankel}(\mathbb{k}_s))$, which is finite. For $m \geq m^*$, we know that $(a^{(m)}, \mathrm{h}^{(m)})$ is the solution of (4.7). Accordingly, due to (4.46), we have that $\mathcal{J}_{\mathscr{F}}(a^{(m)}, \mathrm{h}^{(m)}) = \mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) < \infty$. Also, for any $(a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}}$, one can see that $\mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) \leq \mathcal{J}(a, \mathrm{h}) \leq \mathcal{J}_{\mathscr{F}}(a, \mathrm{h})$. Hence, it follows that $\mathcal{J}_{\mathscr{F}}(a^{(m)}, \mathrm{h}^{(m)}) \leq \mathcal{J}_{\mathscr{F}}(a, \mathrm{h})$ for any $(a, \mathrm{h}) \in \mathcal{V}_{\mathbb{k}}$; i.e., $(a^{(m)}, \mathrm{h}^{(m)})$ is a solution of (3.16). ∎

Based on the above discussion, in order to solve the optimization problem (3.16), it suffices to find the solution of QP (4.33), where $m \geq m^*$ and $\mathbb{k}$ is a given finite Hankel rank kernel such as $\mathbb{k}_{\mathrm{TC}}$, $\mathbb{k}_{\mathrm{DC}}$, $\mathbb{k}_{\mathrm{TC2}}$, $\mathbb{k}_{\mathrm{DC2}}$, or $\mathbb{k}_{\mathrm{SS}}$. Corollary 4.4 provides a bound for m^*. In some special cases, we can provide a more practical bound.

THEOREM 4.11. *Let the assumptions of Theorem* 4.3 *hold. If $\mathbb{k}$ is either the TC kernel* (4.41) *or the DC kernel* (4.42), *then we have that $m^* \leq t_{n_{\mathscr{D}}-1} - \underline{t} + 1$.*

Proof. Let $m = t_{n_{\mathscr{D}}-1} - \underline{t} + 1$, and consider $(a^{(m)}, \mathrm{h}^{(m)})$ the unique solution of (4.19). For $t = 0, 1, \dots, m$, we know that $h_t^{(m)} + a^{(m)} \rho^t \geq 0$. On the other hand, due to the definition of the TC kernel and (4.47), we have that $h_t^{(m)} = \beta^{t-m} h_m^{(m)}$ for any $t > m$. Note that, due to Assumption 1, we have that $\beta^s \leq C \rho_{\mathrm{d}}^s$ for all $s \in \mathbb{Z}_+$. This implies that $\beta \leq \rho_{\mathrm{d}}^2$. Therefore, since $\rho_{\mathrm{d}}^2 < \rho$, we have that

$$\begin{aligned} h_t^{(m)} + a^{(m)} \rho^t &= \beta^{t-m} h_m^{(m)} + a^{(m)} \rho^m \rho^{t-m} \\ &\geq \beta^{t-m} h_m^{(m)} + \beta^{t-m} a^{(m)} \rho^m = \beta^{t-m} (h_m^{(m)} + a^{(m)} \rho^m) \geq 0, \end{aligned} \tag{4.48}$$

where the last equality is due to $h_m^{(m)} + a^{(m)} \rho^m \geq 0$. Therefore, $(a^{(m)}, \mathrm{h}^{(m)})$ is feasible for (4.7), and subsequently, we have that $\mathcal{J}(a^{(m)}, \mathrm{h}^{(m)}) < \infty$. Hence, according to Theorem 4.3, we know that $(a^{(m)}, \mathrm{h}^{(m)}) = (a^*, \mathrm{h}^*)$, which implies that $m^* \leq m = t_{n_{\mathscr{D}}-1} - \underline{t} + 1$. Based on a similar argument, one can show the same result for the DC kernel. ∎

5. Numerical implementation algorithm. Based on the discussion in sections 3.2 and 4, to identify a system $\mathcal{S}$ with internal positivity side-information, we need to solve the convex QP (4.33). This optimization problem can be solved using standard off-the-shelf solvers such as MATLAB's `quadprog` or `cvx` supported by MOSEK [30]. Note that (4.33) depends on nonnegative integer parameter m, which is supposed to be larger or equal to the parameter m^* introduced in Corollary 4.4. One possible approach is to set the value of m to m_0, which is introduced in (4.23) and guaranteed to have the desired property. Also, one may take initially m equal to $t_{n_{\mathscr{D}}} - \underline{t} + 1$ and iteratively increase it until m exceeds m^* and a nonnegative impulse response is obtained. This is of special interest when a suitable QP solver with a

Algorithm 5.1 System identification with internal positivity side-information (PosiID)

Input: Set of data $\mathscr{D}$, finite Hankel rank stable kernel $\mathbb{k}$, dominant pole ρ, regularization weight λ, and $\Delta_m \in \mathbb{N}$.
$m \leftarrow t_{n_{\mathscr{D}}-1} - \underline{t} + 1$
while stopping/exiting condition is not met, **do**
  Calculate vectors $\mathrm{y} = [y_{t_i}]_{i=0}^{n_{\mathscr{D}}-1}$, $\mathrm{b} = [\mathrm{L}_{t_i}(\mathrm{f}_\rho)]_{i=0}^{n_{\mathscr{D}}-1}$, and $\mathrm{c} = [\rho^j]_{j=0}^m$.
  Obtain matrices O, L, and K as in (4.31) or by Remark 4.7
  Solve QP (4.33) for $a^{(m)}$ and $\mathrm{x}^{(m)}$.
  Obtain $\mathrm{h}^{(m)}$ based on (4.32), or equivalently, (4.47).
  $\mathrm{g}^{(m)} \leftarrow (a^{(m)}\rho^s + h_s^{(m)})_{s=0}^{\infty}$
  if $\mathrm{g}^{(m)}$ is nonnegative, **then** exit the loop;
  else, $m \leftarrow m + \Delta_m$
  end
end
Input: Internally positive impulse response g^* and also x^*, a^*, h^*.

warm-starting feature is available [29]. In this iterative approach, at each iteration m, one should check whether the estimated impulse response $\mathrm{g}^{(m)} = (g_s^{(m)})_{s=0}^{\infty}$ is nonnegative, or equivalently, $m \geq m^*$. According to Corollary 4.5, for verifying this stopping condition, it is enough to see whether $g_s^{(m)} \geq 0$ holds for $s \leq m_0$. Due to (4.23), we know that m_0 depends logarithmically on the parameters of problem, and thus, the size of m_0 is not prohibitively large in the practical examples. According to Theorem 4.11, when the TC or DC kernels are employed and the initial value of m is set to $t_{n_{\mathscr{D}}} - \underline{t} + 1$, the introduced iterative scheme takes only a single iteration. The outline of this approach is summarized in Algorithm 5.1.

In order to initialize Algorithm 5.1, in addition to the set of data $\mathscr{D}$, we need a suitable kernel $\mathbb{k}$ and also an estimation of the dominant pole ρ and the regularization weight λ. In general, deciding on the type of kernel depends on the shape and smoothness of the impulse response to be identified [53]. Once the type of kernel $\mathbb{k}$ is set, we need to estimate the vector of hyperparameters $\theta_{\mathbb{k}}$ that characterizes $\mathbb{k}$. Accordingly, vector θ defined as $\theta := [\rho, \lambda, \theta_{\mathbb{k}}] \in \Theta$ is the overall vector of hyperparameters to be determined, where Θ denotes the space of feasible hyperparameters. For estimating θ, we employ schemes with a *cross-validation* nature [66] equipped with a subsequent *Bayesian optimization* heuristic [63]. To this end, we need a *model evaluation metric* $v : \Theta \to \mathbb{R}$ described based on the measurement data and the identification strategy. Subsequently, the hyperparameters θ can be estimated as $\hat{\theta} := \operatorname{argmin}_{\theta \in \Theta} v(\theta)$. Depending on the choice of cross-validation scheme, the model evaluation function can be defined in various forms, including the ones discussed below.

- *Holdout cross-validation* (HCV): In this scheme, we split the index set of data, denoted by $\mathcal{I}$ and defined as $\mathcal{I} = \{0, 1, \ldots, n_{\mathscr{D}} - 1\}$, into two mutually disjoint subsets $\mathcal{I}_{\mathrm{T}}$ and $\mathcal{I}_{\mathrm{V}}$ for *training* and *validation*, respectively. The model evaluation metric $v_{\mathrm{HCV}} : \Theta \to \mathbb{R}$ is then defined based on the prediction error of the validation data as

$$v_{\mathrm{HCV}}(\theta) = \frac{1}{|\mathcal{I}_{\mathrm{V}}|} \sum_{i \in \mathcal{I}_{\mathrm{V}}} \big(y_{t_i} - \mathrm{L}_{t_i}(\mathrm{g}(\theta, \mathcal{I}_{\mathrm{T}}))\big)^2, \tag{5.1}$$

where $\mathrm{g}(\theta,\mathcal{I}_{\mathrm{T}})$ is the result of the proposed identification method using the training data (i.e., the data with index in $\mathcal{I}_{\mathrm{T}}$) and the hyperparameters $\theta\in\Theta$.

- *Leave-one-out cross-validation* (LOOCV): For each $i\in\mathcal{I}$, let $\mathrm{g}_{-i}(\theta)$ be the identification result obtained using the data with index set $\mathcal{I}\backslash\{i\}$ and the given hyperparameters $\theta\in\Theta$. Accordingly, the LOOCV model evaluation metric $v_{\mathrm{LOOCV}}:\Theta\to\mathbb{R}$ is defined as

$$v_{\mathrm{LOOCV}}(\theta)=\frac{1}{n_{\mathscr{D}}}\sum_{i=1}^{n_{\mathscr{D}}}\big(y_{t_i}-\mathrm{L}_{t_i}(\mathrm{g}_{-i}(\theta))\big)^2 \tag{5.2}$$

for any $\theta\in\Theta$.

- *Generalized cross-validation* (GCV): For any $i=0,1,\dots,n_{\mathscr{D}}-1$, let $\mathrm{g}_i(\theta)$ be the identification result given hyperparameters $\theta\in\Theta$ and using the data with index in $\mathcal{I}$, assuming that y_{t_i} is replaced with $\mathrm{L}_t(\mathrm{g}_{-i}(\theta))$. Define the function $\zeta:\Theta\to\mathbb{R}$ as

$$\zeta(\theta)=\sum_{i=0}^{n_{\mathscr{D}}-1}\frac{\mathrm{L}_{t_i}(\mathrm{g}_i(\theta))-\mathrm{L}_{t_i}(\mathrm{g}(\theta))}{\mathrm{L}_{t_i}(\mathrm{g}_{-i}(\theta))-y_{t_i}} \tag{5.3}$$

for any $\theta\in\Theta$, where $\mathrm{g}(\theta)$ is the identified impulse response using the data with index set $\mathcal{I}$ and the given hyperparameters $\theta\in\Theta$. Subsequently, the GCV model evaluation metric $v_{\mathrm{GCV}}:\Theta\to\mathbb{R}$ is defined, for any $\theta\in\Theta$, as

$$v_{\mathrm{GCV}}(\theta)=\frac{1}{\big(n_{\mathscr{D}}-\zeta(\theta)\big)^2}\sum_{i=0}^{n_{\mathscr{D}}-1}\big(y_{t_i}-\mathrm{L}_{t_i}(\mathrm{g}(\theta))\big)^2. \tag{5.4}$$

- *Approximate generalized cross-validation* (AGCV): In the AGCV scheme, the model evaluation metric $v_{\mathrm{AGCV}}:\Theta\to\mathbb{R}$ is defined as a modified version of v_{GCV}, where function $\zeta:\Theta\to\mathbb{R}$ in (5.4) is replaced with an approximation derived from the solution of (4.33). Further details are provided in [66].

Since the dependency of the mentioned model evaluation metrics on the hyperparameters θ has a black-box oracle form, we employ Bayesian optimization algorithms such as Gaussian Process - Lower Confidence Bound (GP-LCB) or similar alternatives [63]. These heuristics are readily available in MATLAB's `bayesopt` function.

Remark 5.1. To evaluate v_{HCV} and v_{AGCV}, one needs to solve a single QP or each. In contrast, the evaluation of v_{LOOCV} and v_{GCV} requires solving $n_{\mathscr{D}}$ and $2n_{\mathscr{D}}+1$ QPs, respectively. Therefore, the computational complexity of LOOCV and GCV is significantly higher compared to that of HCV and AGCV.

Remark 5.2. Note that, in the introduced problem formulation and the proposed identification scheme, no probabilistic assumption has been made on the variables. Accordingly, the hyperparameters are estimated using methods with a cross-validation nature. Extending the proposed methodology by employing a suitable probabilistic framework allows for utilizing other hyperparameter estimation techniques, such as Empirical-Bayes and Stein's unbiased risk estimator.

6. Further internal positivity side-information and extensions. In section 3.2, we employed positive system realization theory to formulate the identification problem with internal positivity side-information. The resulting optimization problem (3.16) is formulated using the fact that the transfer function of system $\mathcal{S}$ is in the form

$$G^{(\mathcal{S})}(z) = F^{(\mathcal{S})}(z) + H^{(\mathcal{S})}(z) = \frac{a}{1-\rho z^{-1}} + H^{(\mathcal{S})}(z); \tag{6.1}$$

i.e., the transfer function $G^{(\mathcal{S})}$ has a dominant part $F^{(\mathcal{S})}(z) := a(1-\rho z^{-1})^{-1}$ with $\rho \in (0,1)$ and a suppressed part $H^{(\mathcal{S})}$. Given that the impulse response of the system satisfies specific properties, the introduced formulation can be further extended to the following cases:

$$G^{(\mathcal{S})}(z) = F^{(\mathcal{S})}(z) + H^{(\mathcal{S})}(z) = \frac{N^{(\mathcal{S})}(z^{-1})}{(1-\rho z^{-1})^n} + H^{(\mathcal{S})}(z) \tag{6.2}$$

and

$$G^{(\mathcal{S})}(z) = F^{(\mathcal{S})}(z) + H^{(\mathcal{S})}(z) = \frac{N^{(\mathcal{S})}(z^{-1})}{1-\rho^n z^{-n}} + H^{(\mathcal{S})}(z), \tag{6.3}$$

where $N^{(\mathcal{S})} \in \mathbb{R}_{n-1}[z^{-1}]$ is a polynomial with degree less than n and $H^{(\mathcal{S})}$ is a transfer function with $r(H^{(\mathcal{S})}) < \rho$; i.e., the spectral radius of $H^{(\mathcal{S})}$ is less than ρ. Note that, according to [6, Theorem 9], the transfer function of a positive system with nonzero spectral radius is in form of (6.2) or (6.3), which are generalizations of (6.1). These cases correspond to the situations where, in addition to internal positivity, we may have further information on the dominant part of the impulse response of the system. In this section, we discuss these extensions. Before proceeding further, one should note that, according to [6, Theorem 8], when the transfer function $G^{(\mathcal{S})}$ has zero spectral radius ($r(G^{(\mathcal{S})}) = 0$), the impulse response $\mathrm{g}^{(\mathcal{S})}$ is internally positive if $\mathrm{g}^{(\mathcal{S})}$ is nonnegative. Furthermore, we know that $\mathrm{g}^{(\mathcal{S})}$ belongs to c_{00}; i.e., there exists $n_{\mathrm{g}} \in \mathbb{Z}_+$ such that $G^{(\mathcal{S})}(z) = \sum_{t=0}^{n_{\mathrm{g}}-1} g_t z^{-t}$ and $g_t = 0$ for all $t \geq n_{\mathrm{g}}$. Compared to the other cases of internal positivity side-information, the case of zero spectral radius provides the weakest information. Indeed, this knowledge only says the $\mathrm{g}^{(\mathcal{S})}$ is a nonnegative and finitely nonzero sequence and provides no further information about the behavior of the impulse response of system $\mathcal{S}$.

Remark 6.1. Based on the discussion above, the identification problem with the internal positivity side-information and the extra knowledge $r(G^{(\mathcal{S})}) = 0$ can be formulated as

$$\begin{array}{ll} \min\limits_{\mathrm{g}\in\mathcal{H}_{\mathbb{k}}} & \sum\limits_{i=0}^{n_{\mathscr{D}}-1} \big(y_{t_i} - \mathrm{L}_{t_i}(\mathrm{g})\big)^2 + \lambda \,\|\mathrm{g}\|_{\mathcal{H}_{\mathbb{k}}}^2 \\ \text{s.t.} & g_t \geq 0 \quad \forall t = 0,1,\dots,n_{\mathrm{g}}-1, \end{array} \tag{6.4}$$

where $\mathbb{k}$ is a kernel that is zero on $\mathbb{Z}_+^2 \setminus \{0,\dots,n_{\mathrm{g}}-1\}^2$. One can show that, for $\mathrm{g}^* = (g_t^*)_{t=0}^{\infty}$, the solution of (6.4), and $t = 0,1,\dots,n_{\mathrm{g}}-1$, we have that $g_t^* = \sum_{i=0}^{n_{\mathscr{D}}} x_i \mathrm{L}_{t_i}(\mathbb{k}_t) + \sum_{s=0}^{n_{\mathrm{g}}-1} x_{n_{\mathscr{D}}+s} \mathbb{k}_s(t)$, where $\mathrm{x} = [x_i]_{i=1}^{n_{\mathscr{D}}+n_{\mathrm{g}}-1}$ is the solution of the following convex QP:

$$\begin{array}{ll} \min\limits_{\mathrm{x}\in\mathbb{R}^{n_{\mathscr{D}}+n_{\mathrm{g}}}} & \big\|\mathrm{y} - [\mathrm{O}\ \mathrm{L}]\,\mathrm{x}\big\|^2 + \lambda \mathrm{x}^{\mathsf{T}} \begin{bmatrix} \mathrm{O} & \mathrm{L} \\ \mathrm{L}^{\mathsf{T}} & \mathrm{K} \end{bmatrix} \mathrm{x} \\ \text{s.t.} & [\mathrm{L}^{\mathsf{T}}\ \mathrm{K}]\,\mathrm{x} \geq 0. \end{array} \tag{6.5}$$

6.1. Nonsimple unique dominant pole. We first discuss the extension that corresponds to (6.2); i.e., the transfer function has a unique dominant pole with multiplicity n, where n can be larger than one. The mathematical proofs are omitted since they are similar to the ones given in section 4.

With respect to each $n \in \mathbb{N}$ and $\rho \in (0,1)$, let $\mathscr{F}_{\alpha,n}$ be the following set of impulse responses:

$$\mathscr{F}_{\alpha,n} = \left\{ \mathrm{f} = (f_t)_{t=0}^{\infty} \in \ell^1 \,\middle|\, \lim_{t\to\infty} t^{-n+1}\rho^{-t} f_t > 0, (1-\rho z^{-1})^n \sum_{t=0}^{\infty} f_t z^{-t} \in \mathbb{R}_{n-1}[z^{-1}] \right\}.$$

Also, define the impulse response sets $\mathscr{P}_{\alpha,n}$ and $\mathscr{P}_{(0,1),n}$, respectively, as

$$\mathscr{P}_{\alpha,n} = \left\{ \mathrm{g} = \mathrm{f} + \mathrm{h} \in \underline{\mathscr{P}} \,\middle|\, \mathrm{f} \in \mathscr{F}_{\alpha,n}, \mathrm{h} = (h_t)_{t=0}^{\infty} \in \ell^1, \lim_{t\to\infty} \rho^{-t} h_t = 0 \right\} \tag{6.6}$$

and $\mathscr{P}_{(0,1),n} = \cup_{\rho\in(0,1)} \mathscr{P}_{\alpha,n}$. According to [6, Theorem 11] and based on an argument similar to the proof of Corollary 3.5, one can show that, for any $\rho \in (0,1)$, the impulse responses in $\mathscr{P}_{\alpha,n}$ are internally positive. Indeed, $\mathscr{P}_{\alpha,n}$ is exactly the set of impulse responses of positive systems with dominant pole structure as in (6.2). Thus, to identify impulse response $\mathrm{g}^{(\mathcal{S})}$ with internal positivity side-information in the sense that $\mathrm{g}^{(\mathcal{S})} \in \mathscr{P}_{\alpha,n}$, we need to estimate f and h with the properties given in (6.6). One can see that each $\mathrm{f} = (f_t)_{t=0}^{\infty} \in \mathscr{F}_{\alpha,n}$ is uniquely characterized, in terms of real positive number a and vector $\mathrm{a} = [a_j]_{j=0}^{n-2} \in \mathbb{R}^{n-2}$, as $f_t = at^{n-1}\rho^t + \sum_{j=0}^{n-2} a_j t^j \rho^t$ for all $t \in \mathbb{Z}_+$. Subsequently, we redefine the empirical loss function $\mathcal{E}_{\rho,n} : \mathbb{R} \times \mathbb{R}^{n-2} \times \mathcal{H}_{\mathbb{k}} \to \mathbb{R}_+$ as

$$\mathcal{E}_{\rho,n}(a,\mathrm{a},\mathrm{h}) := \sum_{i=0}^{n_{\mathscr{D}}-1} \left[y_{t_i} - \mathrm{L}_{t_i}\left(a\mathrm{f}_{\rho,n-1} + \sum_{j=0}^{n-2} a_j \mathrm{f}_{\rho,j} \right) - \mathrm{L}_{t_i}(\mathrm{h}) \right]^2, \tag{6.7}$$

where, for $j = 0, \ldots, n-1$, the impulse response $\mathrm{f}_{\alpha,j}$ is defined as $\mathrm{f}_{\alpha,j} = (t^j\rho^t)_{t=0}^{\infty}$. According to (6.7), the identification problem (3.16) is updated to the following optimization problem:

$$\begin{array}{ll} \displaystyle\min_{a\in\mathbb{R},\mathrm{h}\in\mathcal{H}_{\mathbb{k}},\mathrm{a}\in\mathbb{R}^{n-1}} & \mathcal{E}_{\rho,n}(a,\mathrm{a},\mathrm{h}) + \lambda \|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2 + \varepsilon\|\mathrm{a}\|^2 \\ \text{s.t.} & h_t + \rho^t \left[at^{n-1} + \sum_{j=0}^{n-2} a_j t^j \right] \geq 0 \ \forall t \geq 0, \\ & \mathrm{rank}(\mathrm{Hankel}(\mathrm{h})) < \infty, \\ & a \geq a_{\min}, \end{array} \tag{6.8}$$

where $a_{\min} > 0$ is a given lower bound for a and $\varepsilon > 0$ is a regularization weight. Based on the same line of argument as in section 4, we can present a finite-dimensional convex QP equivalent to (6.8). Before proceeding further, we define matrices $\mathrm{B} \in \mathbb{R}^{n_{\mathscr{D}}\times n}$ and $\mathrm{C} \in \mathbb{R}^{(m+1)\times n}$ as follows:

$$\begin{array}{ll} \mathrm{B}(i,j) = \mathrm{L}_{t_{i-1}}(\mathrm{f}_{\rho,j-1}), & 1 \leq i \leq n_{\mathscr{D}}, 1 \leq j \leq n, \\ \mathrm{C}(i,j) = (i-1)^{j-1}\rho^{i-1}, & 1 \leq i \leq m+1, 1 \leq j \leq n. \end{array} \tag{6.9}$$

THEOREM 6.2. *Let Assumptions* 1, 2, *and* 3 *hold and* $\mathbb{k}$ *be a finite Hankel rank kernel. Also, with respect to each* $m \in \mathbb{Z}_+$, *let* $(a^{(m)}, \mathrm{a}^{(m)}, \mathrm{x}^{(m)})$ *be the solution of convex QP:*

$$\begin{array}{ll} \displaystyle\min_{\mathrm{x}\in\mathbb{R}^{n_{\mathscr{D}}+m+1}, a\in\mathbb{R}, \mathrm{a}\in\mathbb{R}^{n-1}} & \left\| \mathrm{y} - \mathrm{B}\begin{bmatrix}\mathrm{a}\\ a\end{bmatrix} - \begin{bmatrix}\mathrm{O} & \mathrm{L}\end{bmatrix}\mathrm{x} \right\|^2 + \lambda \mathrm{x}^{\mathsf{T}} \begin{bmatrix}\mathrm{O} & \mathrm{L}\\ \mathrm{L}^{\mathsf{T}} & \mathrm{K}\end{bmatrix} \mathrm{x} + \varepsilon\|\mathrm{a}\|^2 \\ \text{s.t.} & \begin{bmatrix}\mathrm{L}^{\mathsf{T}} & \mathrm{K}\end{bmatrix}\mathrm{x} + \mathrm{C}\begin{bmatrix}\mathrm{a}\\ a\end{bmatrix} \geq 0, \\ & a \geq a_{\min}, \end{array} \tag{6.10}$$

and the impulse response $\mathrm{h}^{(m)}$ *be defined according to* (4.32). *Then, there exists* m^* *such that* $(a^{(m)}, \mathrm{a}^{(m)}, \mathrm{h}^{(m)})$ *is a solution of* (6.8) *for any* $m \geq m^*$. *Moreover, for any* $m_1, m_2 \geq m^*$, *we have that* $(a^{(m_1)}, \mathrm{a}^{(m_1)}, \mathrm{h}^{(m_1)}) = (a^{(m_2)}, \mathrm{a}^{(m_2)}, \mathrm{h}^{(m_2)})$.

6.2. Multiple simple dominant poles. In this section, we introduce the extension corresponding to the case in (6.3); i.e., the dominant part of the transfer function has specially structured multiple simple dominant poles.

For any $n\in\mathbb{N}$ and any $\rho\in(0,1)$, define the impulse response sets $\mathscr{F}_\alpha^{(n)}$, $\mathscr{P}_\alpha^{(n)}$, and $\mathscr{P}_{(0,1)}^{(n)}$, respectively, as

$$\begin{aligned}\mathscr{F}_\alpha^{(n)} &= \left\{\mathrm{f}=(f_t)_{t=0}^\infty\in\underline{\mathscr{P}}\,\Big|\,\liminf_{t\to\infty}\rho^{-t}f_t>0,(1-\rho^n z^{-n})\sum_{t=0}^{\infty}f_t z^{-t}\in\mathbb{R}_{n-1}[z^{-1}]\right\},\\ \mathscr{P}_\alpha^{(n)} &= \left\{\mathrm{g}=\mathrm{f}+\mathrm{h}\in\underline{\mathscr{P}}\,\Big|\,\mathrm{f}\in\mathscr{F}_\alpha^{(n)},\mathrm{h}=(h_t)_{t=0}^\infty\in\ell^1,\liminf_{t\to\infty}\rho^{-t}|h_t|=0\right\},\end{aligned}\tag{6.11}$$

and $\mathscr{P}_{(0,1)}^{(n)}=\cup_{\rho\in(0,1)}\mathscr{P}_\alpha^{(n)}$. One can easily see that, for $n=1$, these sets coincide with $\mathscr{P}_\alpha$ and $\mathscr{P}_{(0,1)}$. Due to [6, Theorem 12] and by following the same line of argument as in the proof of Corollary 3.5, we can show that $\mathscr{P}_\alpha^{(n)}$ contains internally positive impulse responses for any $\rho\in(0,1)$. Accordingly, the identification with internal positivity side-information in the sense that the impulse response belongs to $\mathscr{P}_\alpha^{(n)}$ translates to the estimation of f and h with the properties given in (6.11). Note that, with respect to each $\mathrm{f}=(f_t)_{t=0}^\infty\in\mathscr{F}_\alpha^{(n)}$, there exist real scalars $a_0^{(\mathrm{r})},\ldots,a_{n-1}^{(\mathrm{r})}$ and $a_0^{(\mathrm{i})},\ldots,a_{n-1}^{(\mathrm{i})}$ such that, for any $t\in\mathbb{Z}_+$, we have that

$$f_t=\mathrm{real}\left(\sum_{k=0}^{n-1}(a_k^{(\mathrm{r})}+\mathrm{j}a_k^{(\mathrm{i})})\rho^t\omega^{kt}\right),0=\mathrm{imag}\left(\sum_{k=0}^{n-1}(a_k^{(\mathrm{r})}+\mathrm{j}a_k^{(\mathrm{i})})\rho^t\omega^{kt}\right),\tag{6.12}$$

where $\omega=\mathrm{e}^{\frac{2\pi}{n}}$; i.e., f is uniquely characterized in terms of vectors $\mathrm{a}^{(\mathrm{r})}=[a_0^{(\mathrm{r})},\ldots,a_{n-1}^{(\mathrm{r})}]^\mathsf{T}$ and $\mathrm{a}^{(\mathrm{i})}=[a_0^{(\mathrm{i})},\ldots,a_{n-1}^{(\mathrm{i})}]^\mathsf{T}$. Hence, we can reintroduce the empirical loss function $\mathcal{E}_\rho^{(n)}:\mathbb{R}^n\times\mathbb{R}^n\times\mathcal{H}_{\mathbb{k}}\to\mathbb{R}_+$ as follows:

$$\mathcal{E}_\rho^{(n)}(\mathrm{a}^{(\mathrm{r})},\mathrm{a}^{(\mathrm{i})},\mathrm{h}):=\sum_{i=0}^{n_{\mathscr{D}}-1}\left[y_{t_i}-\mathrm{L}_{t_i}\left(\sum_{k=0}^{n-1}\left(a_k^{(\mathrm{r})}\mathrm{f}_{\alpha,k}^{(\mathrm{r})}-a_k^{(\mathrm{i})}\mathrm{f}_{\alpha,k}^{(\mathrm{i})}\right)\right)-\mathrm{L}_{t_i}(\mathrm{h})\right]^2,$$

where $\mathrm{f}_{\alpha,k}^{(\mathrm{r})}:=\left(\rho^t\ \mathrm{real}(\omega^{kt})\right)_{t=0}^\infty$ and $\mathrm{f}_{\alpha,k}^{(\mathrm{i})}:=\left(\rho^t\ \mathrm{imag}(\omega^{kt})\right)_{t=0}^\infty$ for $k=0,\ldots,n-1$. Therefore, the identification problem (3.16) is modified to

$$\begin{aligned}\min_{\mathrm{a}^{(\mathrm{r})},\mathrm{a}^{(\mathrm{i})}\in\mathbb{R}^n,\mathrm{h}\in\mathcal{H}_{\mathbb{k}}}\quad & \mathcal{E}_\rho^{(n)}(\mathrm{a}^{(\mathrm{r})},\mathrm{a}^{(\mathrm{i})},\mathrm{h})+\lambda\,\|\mathrm{h}\|_{\mathcal{H}_{\mathbb{k}}}^2+\varepsilon\|\mathrm{E}\mathrm{a}^{(\mathrm{r})}\|^2+\varepsilon\|\mathrm{E}\mathrm{a}^{(\mathrm{i})}\|^2\\ \text{s.t.}\quad & h_t+\sum_{k=0}^{n-1}\left(a_k^{(\mathrm{r})}\mathrm{f}_{\alpha,k}^{(\mathrm{r})}-a_k^{(\mathrm{i})}\mathrm{f}_{\alpha,k}^{(\mathrm{i})}\right)\geq 0\quad\forall t\geq 0,\\ & \sum_{k=0}^{n-1}\left(a_k^{(\mathrm{i})}\mathrm{f}_{\alpha,k}^{(\mathrm{r})}+a_k^{(\mathrm{r})}\mathrm{f}_{\alpha,k}^{(\mathrm{i})}\right)=0\quad\forall t\geq 0,\\ & \liminf_{t\to\infty}\rho^{-t}\sum_{k=0}^{n-1}\left(a_k^{(\mathrm{r})}\mathrm{f}_{\alpha,k}^{(\mathrm{r})}-a_k^{(\mathrm{i})}\mathrm{f}_{\alpha,k}^{(\mathrm{i})}\right)\geq a_{\min},\\ & \mathrm{rank}(\mathrm{Hankel}(\mathrm{h}))<\infty,\end{aligned}\tag{6.13}$$

where $a_{\min}>0$ is a small positive real scalar, $\varepsilon>0$ is a regularization weight, and $\mathrm{E}\in\mathbb{R}^{n\times n}$ is defined as $\mathrm{E}=\mathrm{diag}(0,1,1,\ldots,1)$. By an argument similar to section 4, we can derive an equivalent finite-dimensional convex QP for (6.13). Define matrices $\mathrm{V}_m\in\mathbb{R}^{m\times n}$, $\mathrm{B}^{(\mathrm{r})}\in\mathbb{R}^{n_{\mathscr{D}}\times n}$, and $\mathrm{B}^{(\mathrm{i})}\in\mathbb{R}^{n_{\mathscr{D}}\times n}$, respectively, as

$$\begin{aligned}\mathrm{V}_m(i,j) &= \omega^{(i-1)(j-1)}, && 1\leq i\leq m, 1\leq j\leq n,\\ \mathrm{B}^{(\mathrm{r})}(i,j) &= \mathrm{L}_{t_{i-1}}(\mathrm{f}_{\rho,j-1}^{(\mathrm{r})}), && 1\leq i\leq n_{\mathscr{D}}, 1\leq j\leq n,\\ \mathrm{B}^{(\mathrm{i})}(i,j) &= \mathrm{L}_{t_{i-1}}(\mathrm{f}_{\rho,j-1}^{(\mathrm{i})}), && 1\leq i\leq n_{\mathscr{D}}, 1\leq j\leq n\end{aligned}\tag{6.14}$$

for $m \in \mathbb{Z}_+ \cup \{\infty\}$. Also, let $\mathrm{V}_m^{(\mathrm{r})}$, $\mathrm{V}_m^{(\mathrm{i})}$, and D_m be defined, respectively, as $\mathrm{real}(\mathrm{V}_m)$, $\mathrm{imag}(\mathrm{V}_m)$, and $\mathrm{diag}(1,\rho,\ldots,\rho^{m-1})$. One can see that the first constraint in (6.13) is equivalent to $\mathrm{h} + \mathrm{D}_\infty \mathrm{V}_\infty^{(\mathrm{r})}\,\mathrm{a}^{(\mathrm{r})} - \mathrm{D}_\infty \mathrm{V}_\infty^{(\mathrm{i})}\,\mathrm{a}^{(\mathrm{i})} \geq 0$. The second constraint in (6.13) is

$$\mathrm{D}_\infty \mathrm{V}_\infty^{(\mathrm{r})}\,\mathrm{a}^{(\mathrm{i})} + \mathrm{D}_\infty \mathrm{V}_\infty^{(\mathrm{i})}\,\mathrm{a}^{(\mathrm{r})} = 0, \tag{6.15}$$

which implies that $\mathrm{V}_\infty^{(\mathrm{r})}\mathrm{a}^{(\mathrm{i})} + \mathrm{V}_\infty^{(\mathrm{i})}\mathrm{a}^{(\mathrm{r})} = 0$ due to $\rho > 0$. As $\omega^n = 1$, we know that V_m is an n-periodic Vandermonde matrix. Therefore, (6.15) is equivalent to $\mathrm{V}_n^{(\mathrm{r})}\mathrm{a}^{(\mathrm{i})} + \mathrm{V}_n^{(\mathrm{i})}\mathrm{a}^{(\mathrm{r})} = 0$. Similarly, one can show that the third constraint in (6.13) is equivalent to $\mathrm{V}_n^{(\mathrm{r})}\mathrm{a}^{(\mathrm{r})} + \mathrm{V}_n^{(\mathrm{i})}\mathrm{a}^{(\mathrm{i})} \geq a_{\min}\mathrm{1}_n$. Based on the discussion above and similar to those in section 4, we can present the finite-dimensional convex QP equivalent to (6.13).

THEOREM 6.3. *Let the assumptions of Theorem* 6.2 *hold. For any $m \in \mathbb{Z}_+$, let $(\mathrm{a}^{(\mathrm{r},m)}, \mathrm{a}^{(\mathrm{i},m)}, \mathrm{x}^{(m)})$ be the solution of the convex QP*

$$\begin{aligned}
&\min_{\substack{\mathrm{x}\in\mathbb{R}^{n_{\mathscr{D}}+m+1}\\ \mathrm{a}^{(\mathrm{r})},\,\mathrm{a}^{(\mathrm{i})}\in\mathbb{R}^n}} \left\|\mathrm{y}-\mathrm{B}^{(\mathrm{r})}\mathrm{a}^{(\mathrm{r})}-\mathrm{B}^{(\mathrm{i})}\mathrm{a}^{(\mathrm{i})}-\begin{bmatrix}\mathrm{O} & \mathrm{L}\end{bmatrix}\mathrm{x}\right\|^2+\lambda\mathrm{x}^{\mathsf{T}}\begin{bmatrix}\mathrm{O} & \mathrm{L}\\ \mathrm{L}^{\mathsf{T}} & \mathrm{K}\end{bmatrix}\mathrm{x}+\varepsilon\|\mathrm{E}\mathrm{a}^{(\mathrm{r})}\|^2+\varepsilon\|\mathrm{E}\mathrm{a}^{(\mathrm{i})}\|^2\\
&\text{s.t.}\quad \begin{bmatrix}\mathrm{L}^{\mathsf{T}} & \mathrm{K}\end{bmatrix}\mathrm{x}+\mathrm{D}_{m+1}\begin{bmatrix}\mathrm{V}_{m+1}^{(\mathrm{r})} & -\mathrm{V}_{m+1}^{(\mathrm{i})}\end{bmatrix}\begin{bmatrix}\mathrm{a}^{(\mathrm{r})}\\ \mathrm{a}^{(\mathrm{i})}\end{bmatrix}\geq 0,\\
&\qquad\ \mathrm{V}_n^{(\mathrm{r})}\mathrm{a}^{(\mathrm{i})}+\mathrm{V}_n^{(\mathrm{i})}\mathrm{a}^{(\mathrm{r})}=0,\\
&\qquad\ \mathrm{V}_n^{(\mathrm{r})}\mathrm{a}^{(\mathrm{r})}+\mathrm{V}_n^{(\mathrm{i})}\mathrm{a}^{(\mathrm{i})}\geq a_{\min}\mathrm{1}_n,
\end{aligned} \tag{6.16}$$

and define the impulse response $\mathrm{h}^{(m)}$ by (4.32). *Then, there exists m^* such that $(\mathrm{a}^{(\mathrm{r},m)}, \mathrm{a}^{(\mathrm{i},m)}, \mathrm{x}^{(m)})$ is a solution of* (6.8) *for each $m \geq m^*$. Moreover, for any integer $m_1, m_2 \geq m^*$, we have that $(\mathrm{a}^{(\mathrm{r},m_1)}, \mathrm{a}^{(\mathrm{i},m_1)}, \mathrm{h}^{(m_1)}) = (\mathrm{a}^{(\mathrm{r},m_2)}, \mathrm{a}^{(\mathrm{i},m_2)}, \mathrm{h}^{(m_2)})$.*

7. Numerical experiments.

In this section, we provide numerical and experimental examples to verify the efficacy and performance of the proposed method for impulse response identification with internal positivity side-information. The first example concerns the impact of incorporating internal positivity side-information on the estimation quality and provides a comparative analysis for the proposed identification scheme through a Monte Carlo analysis. The second example concerns the efficacy of the proposed identification scheme on a set of data collected from an experimental heating system.

7.1. Monte Carlo experiment.

Consider a system $\mathcal{S}$ described with the impulse response $\mathrm{g}^{(\mathcal{S})} = (g_t^{(\mathcal{S})})_{t=0}^{\infty}$ defined as $g_t^{(\mathcal{S})} = \rho^t(1+\beta^t\cos(2\pi\omega t))$ for all $t \in \mathbb{Z}_+$, where ρ and β are real scalars in $(0,1)$ and ω is an irrational real number in $(0,1)$.

One can easily see that $\mathrm{g}^{(\mathcal{S})}$ is a nonnegative impulse response with transfer function

$$G^{(\mathcal{S})}(z)=\frac{1}{1-\rho z^{-1}}+\frac{1-\rho\beta\cos w\ z^{-1}}{1-2\rho\beta\cos w\ z^{-1}+\rho^2\beta^2 z^{-2}}. \tag{7.1}$$

Therefore, according to Corollary 3.5, we know that $\mathrm{g}^{(\mathcal{S})}$ is internally positive.

7.1.1. Simulation configuration.

In this numerical experiment, we set $\rho = 0.98$, $\beta = 0.92$, and $\omega = \frac{1}{10}\pi^2$. Using MATLAB's `idinput` function, we generate a set of 400 random binary input signals, each with length of $n_{\mathscr{D}} = 200$. The system is initially at rest. The input signals are applied to the system, and the corresponding noiseless output is obtained. We consider three signal-to-noise ratio (SNR) levels of 0 dB, 5 dB, and 10 dB. With respect to each of these SNR levels and each output signal, we generate a zero-mean white Gaussian signal as the additive measurement uncertainty. The resulting noisy output is measured at time instants $t_i = i$ for $i = 0, 1, \ldots, 199$. Accordingly, with respect to each of the mentioned SNR levels, we have 120 sets of input-output data.

7.1.2. Comparison methods. For estimating the impulse response of system, we utilize the input-output data sets and the following identification methods:

A. The first method is based on the subspace approach implemented by MATLAB's `n4sid` and using the true order of system $\mathcal{S}$. Once we obtain an initial estimation $\tilde{\mathrm{g}}^{(1)}$, the result is projected on the positive orthant by $\hat{\mathrm{g}}^{(1)} = \max(\tilde{\mathrm{g}}^{(1)}, 0)$, where the max operation is performed coordinatewise.

B. In the second method, we employ the least-squares approach, and then, similar to method A, the projection on the positive orthant is applied to estimate a nonnegative finite impulse response. More precisely, $\hat{\mathrm{g}}^{(2)}$ is obtained by $\hat{\mathrm{g}}^{(2)} = \max(\tilde{\mathrm{g}}^{(2)}, 0)$, where $\tilde{\mathrm{g}}^{(2)} := \operatorname{argmin}_{\mathrm{g}\in\mathbb{R}^{n_{\mathrm{g}}}} \|\mathrm{T}_{\mathrm{u}}\mathrm{g} - \mathrm{y}\|^2$ and vector y and Toeplitz matrix T_{u} are, respectively, defined in Remark 4.7.

C. This method is based on a constrained least-squares approach, where the external positivity feature is enforced by setting the feasible set to the positive orthant. In other words, the impulse response is estimated as $\hat{\mathrm{g}}^{(3)} := \operatorname{argmin}_{\mathrm{g}\in\mathbb{R}_+^{n_{\mathrm{g}}}} \|\mathrm{T}_{\mathrm{u}}\mathrm{g} - \mathrm{y}\|^2$.

D. and E. The fourth and fifth methods are, respectively, similar to the second and third approaches, but with an additional kernel-based regularization term included in the corresponding optimization problems. For method D, one can employ MATLAB's `impulseest` function and then take the positive part of the resulting FIR. Also, method E essentially corresponds to (6.4), or equivalently, (6.5).

F. The sixth method is a Bayesian FIR estimation scheme for externally positive systems [69, 70]. This scheme is based on maximum a posteriori estimation, where the employed prior is a maximum entropy distribution with support on positive orthant and a kernel-based covariance.

PosiID. The last method is the scheme proposed in this paper and summarized in Algorithm 5.1 (see section 5).

One should note that, in all of the mentioned methods, the resulting impulse responses are nonnegative. In order to have a fair comparison, in the kernel-based methods D, E, F, and PosiID, the same kernel type (4.42) is employed.

7.1.3. Evaluation metrics and results. For evaluating the performances of these methods, we compare the resulting bias-variance trade-offs, as shown in Table 1. Moreover, for further quantitative comparison of the estimated impulse responses, we use *coefficient of determination*, or R-squared, which is denoted by fit and defined as

$$\mathrm{fit}(\hat{\mathrm{g}}) = 100 \times \left(1 - \frac{\|\hat{\mathrm{g}} - \mathrm{g}^{(\mathcal{S})}\|_2}{\|\mathrm{g}^{(\mathcal{S})}\|_2}\right), \tag{7.2}$$

where $\hat{\mathrm{g}}$ is the estimated impulse response. Figure 1 compares the resulting quality of fit for the different SNR levels.

7.1.4. Discussion. We have several observations from the Monte Carlo numerical experiment. As shown in Figure 1, all methods are outperformed by the proposed identification scheme, which maximally incorporates the internal positivity side-information. Indeed, the side-information helps exclude spurious model candidates and subsequently increases the accuracy of the estimation. The bias-variance

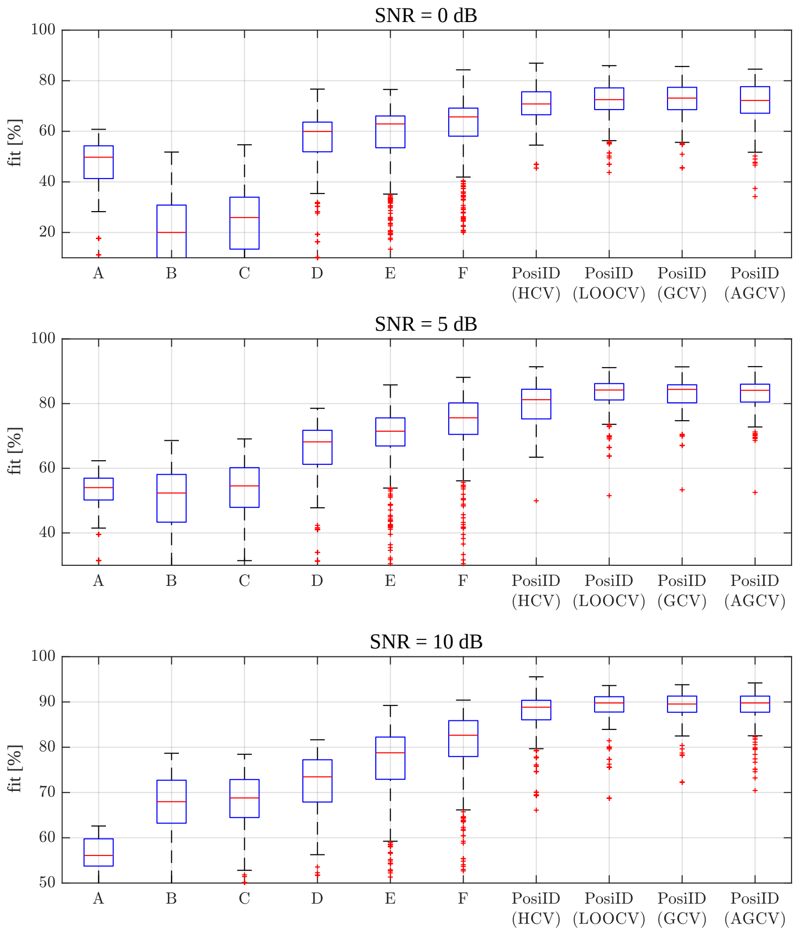

FIG. 1. *Box plots of the R-squared metric for the estimation results of the different methods and SNR levels.*

results presented in Table 1 confirm this fact. While each of these methods estimates a nonnegative impulse response and partially integrates positivity side-information, we can see that the level of integrated side-information by methods A, B, C, and D is less than that of PosiID; i.e., the proposed approach incorporates this information in the model maximally. Comparing methods B and C, one can see the former is a two-step procedure where the estimation is performed in the first step and nonnegativity of the impulse response is obtained in the second step, while the latter approach is a single-step procedure that considers impulse response nonnegativity during the estimation. On the other hand, according to the fitting results shown in Figure 1, C performs better than method B. For methods D and E, we have a similar argument. This observation highlights the importance of jointly considering the positivity with the impulse response estimation, as done by the proposed method. Method A knows the actual order of the system. However, according to the results presented in Figure 1 and Table 1, one can see that positivity is a more advantageous and stronger side-information for impulse response estimation, especially when it is incorporated with its maximum strength as it is done by PosiID. Finally, one can see that the kernel-based methods D, E, F, and PosiID have better estimation performance compared to the methods A, B, and C, which is expected [53].

TABLE 1

The bias, variance, and mean squared error (MSE) resulting from the identification methods listed in section 7.1. The last four columns correspond to the proposed approach, which integrates internal positivity.

	Method	A	B	C	D	E	F	PosiID (HCV)	PosiID (LOOCV)	PosiID (GCV)	PosiID (AGCV)
0 dB	**Bias**($\hat{g}$)	2.33	1.446	1.207	1.439	1.179	1.007	0.768	0.683	0.674	0.66
	Var($\hat{g}$)	3.02	22.09	16.68	4.256	4.260	3.835	2.057	1.994	1.960	2.104
	MSE($\hat{g}$)	8.46	24.18	18.14	6.327	5.651	4.849	2.646	2.460	**2.413**	2.539
5 dB	**Bias**($\hat{g}$)	2.26	0.829	0.802	1.080	0.950	0.790	0.482	0.375	0.378	0.377
	Var($\hat{g}$)	1.39	7.064	5.960	2.622	2.196	1.694	1.099	0.803	0.766	0.770
	MSE($\hat{g}$)	6.50	7.751	6.603	3.789	3.099	2.317	1.331	0.944	**0.909**	0.912
10 dB	**Bias**($\hat{g}$)	2.26	0.770	0.768	0.979	0.763	0.592	0.276	0.216	0.211	0.215
	Var($\hat{g}$)	0.70	2.863	2.602	1.606	1.257	0.971	0.472	0.368	0.332	0.330
	MSE($\hat{g}$)	5.82	3.456	3.192	2.564	1.839	1.323	0.548	0.415	**0.376**	**0.376**

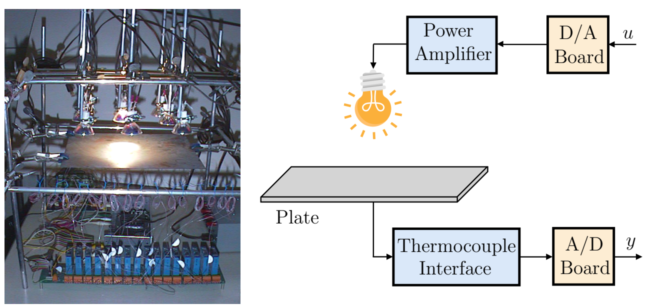

FIG. 2. *The experimental system (left) and the corresponding block-diagram schematic (right).*

7.2. Heating system experiment. In this example, we verify the efficacy of the proposed identification scheme on a set of data collected from an experimental nonlinear heating system [21].

Figure 2 shows the experiment configuration and the corresponding control and measurement schematic. In this experiment, a metal plate is heated up by a 300-watt Halogen lamp mounted almost 5 cm above the center of the plate. On the other side of the plate, a thermocouple is placed, measuring the temperature. The thermocouple is connected to a computer via an analog-to-digital (A/D) board for sampling and recording the temperature measurements. The lamp is supplied by a thyristor-based power amplifier driven by a digital-to-analog (D/A) board and controlled by a computer. The sampling time for control and data acquisition is $T_{\mathrm{s}} = 2\,\mathrm{s}$. Accordingly, we have a nonlinear discrete-time system from the input of the D/A board to the output of the A/D board, which is intuitively close to a linear positive system. The nonlinearity of the system is mainly due to the power amplifier [21]. Moreover, the system is subject to delay and disturbances from the ambient. To make the identification problem more challenging, we disregard the nonlinearity and external disturbances issues.

The system is actuated by a piecewise constant input, and the output of the system is measured for 801 samples. The collected input-output data are shown in Figure 3, which is also available in the DAISY database [15]. We employ these data for identifying the system using the identification methods discussed in section 7.1 and then compare the results. Since the lamp failed near the end of experiment and

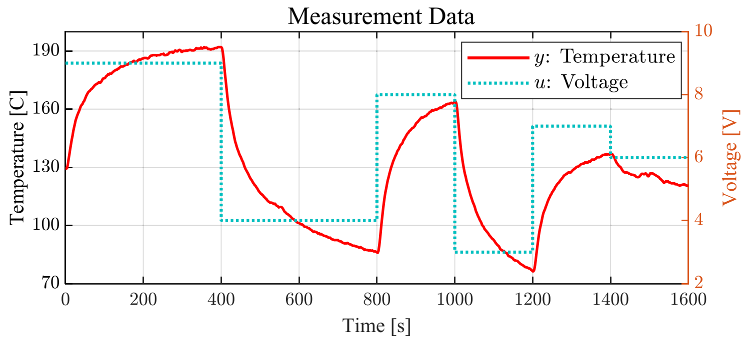

FIG. 3. *The measurement data corresponding to the experiments.*

TABLE 2
R-squared metric evaluations of test data for different identification methods.

Method	N + ARX	A	B	C	D	E	F	PosiID
Fit [%]	48.5	83.4	81.2	80.4	81.7	89.7	85.8	**92.2**

the tail of data has less fidelity, we discard the last 101 samples (200 seconds). We split the data into a *training set*, to be used for identification, and a *test set*, which is the base for comparing the identified models. The sample training set contains the first 500 measurement samples, and the next 200 data points belong to the test set. The quality of the estimated models is evaluated based on the *R-squared* metric, which measures the prediction precision on the test data and is defined as follows:

$$\text{fit}(\hat{\mathrm{g}}) = 100 \times \left(1 - \left[\frac{\sum_{501\le i\le 700}(y_{t_i} - \hat{y}_{t_i})^2}{\sum_{501\le i\le 700}(y_{t_i} - \overline{y})^2}\right]^{\frac{1}{2}}\right), \tag{7.3}$$

where $\hat{y}_s$ denotes the predicted output for time instant s and $\overline{y}$ is the average of output measurements in the test set. In [21], a Hammerstein model is derived, where the static nonlinear block is a sinusoidal map derived by curve-fitting and the linear block is an autoregressive exogenous (ARX) model with estimated coefficients. We denote this method by N + ARX. In the kernel-based methods D, E, F, and PosiID, we have employed the TC kernel (4.41). Also, for the methods that are estimating an FIR, we have set $n_{\mathrm{g}} = 200$. We evaluate the R-squared metric on the test data for this model and the ones estimated by the above methods. Table 2 reports fitting results where one can see that the proposed method provides more accurate fit. This is also confirmed by Figure 4, which compares the test data with the output signals predicted by methods N + ARX, E, F, and PosiID. It seems that for obtaining models with more accurate predictions, one should identify a model with nonlinear dynamics.

8. Conclusion. In this paper, we have considered the problem of impulse response identification when side-information is available on the internal positivity of the system. We have employed the realization theory of positive systems to introduce the identification scheme in which the positivity side-information is integrated into the identified model. The resulting formulation is in the form of a constrained optimization over an RKHS endowed with a stable kernel, where the constraints are suitably designed to incorporate the positivity side-information in the solution. We have bor-

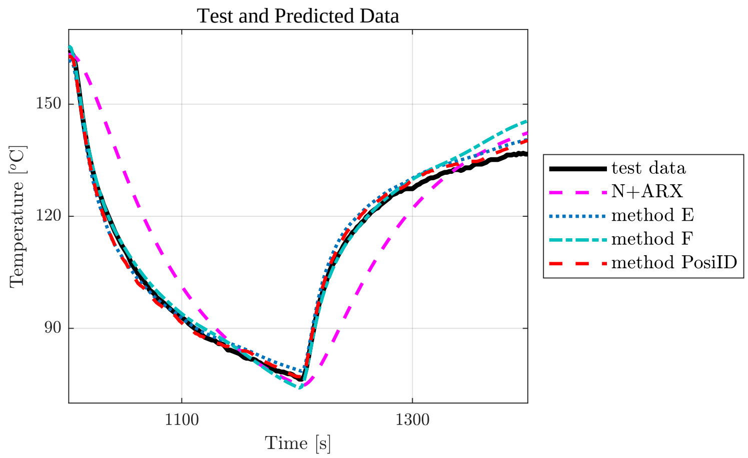

FIG. 4. *The figure compares the test measurement data and the predicted values.*

rowed techniques and tools from optimization theory in normed spaces to derive an equivalent finite-dimensional convex QP. This gives a computationally tractable identification scheme that incorporates the internal positivity side-information and has the well-known advantageous features of kernel-based methods. We have performed a Monte Carlo numerical experiment to compare the performance of the proposed approach with FIR identification methods considering only the external positivity feature. This has empirically studied the impact of integrating positivity side-information in terms of estimation bias, variance, and mean squared error. The results show that the proposed identification approach, which integrates internal positivity, outperforms the schemes considering only external positivity. We have observed that incorporating internal positivity side-information reduces the estimation bias and variance. This observation is expected since FIR external positivity implies the weakest form of information about an internally positive system and fails to exploit the complete information of internal positivity. We have further evaluated the effectiveness of the proposed identification scheme using data from a heating system experiment.

REFERENCES

[1] A. A. Ahmadi and B. El Khadir, *Learning dynamical systems with side information (short version)*, Proc. Mach. Learn. Res., 120 (2020), pp. 718–727, http://proceedings.mlr.press/v120/ahmadi20a.

[2] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, *Generalized system identification with stable spline kernels*, SIAM J. Sci. Comput., 40 (2018), pp. B1419–B1443, https://doi.org/10.1137/16M1070517.

[3] J. Back and A. Astolfi, *Design of positive linear observers for positive linear systems via coordinate transformations and positive realizations*, SIAM J. Control Optim., 47 (2008), pp. 345–373, https://doi.org/10.1137/060663891.

[4] R. Bellman, *Dynamic programming, system identification, and suboptimization*, SIAM J. Control, 4 (1966), pp. 1–5, https://doi.org/10.1137/0304001.

[5] L. Benvenuti, A. De Santis, and L. Farina, *On model consistency in compartmental systems identification*, Automatica, 38 (2002), pp. 1969–1976, https://doi.org/10.1016/S0005-1098(02)00107-3.

[6] L. Benvenuti and L. Farina, *A tutorial on the positive realization problem*, IEEE Trans. Automat. Control, 49 (2004), pp. 651–664, https://doi.org/10.1109/TAC.2004.826715.

[7] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Springer, Cham, 2011.

[8] M. Bisiacco and G. Pillonetto, *Kernel absolute summability is sufficient but not necessary for RKHS stability*, SIAM J. Control Optim., 58 (2020), pp. 2006–2022, https://doi.org/10.1137/19M1278442.

[9] R. F. Brown, *Compartmental system analysis: State of the art*, IEEE Trans. Bio-Mec. Eng., BME-27 (1980), pp. 1–11, https://doi.org/10.1109/TBME.1980.326685.

[10] T. Chen, *On kernel design for regularized LTI system identification*, Automatica, 90 (2018), pp. 109–122, https://doi.org/10.1016/j.automatica.2017.12.039.

[11] T. Chen, H. Ohlsson, and L. Ljung, *On the estimation of transfer functions, regularizations and Gaussian processes – Revisited*, Automatica, 48 (2012), pp. 1525–1535, https://doi.org/10.1016/j.automatica.2012.05.026.

[12] T. Chen and G. Pillonetto, *On the stability of reproducing kernel Hilbert spaces of discrete-time impulse responses*, Automatica, 95 (2018), pp. 529–533, https://doi.org/10.1016/j.automatica.2018.05.017.

[13] M. Colombino and R. S. Smith, *A convex characterization of robust stability for positive and positively dominated linear systems*, IEEE Trans. Automat. Control, 61 (2015), pp. 1965–1971, https://doi.org/10.1109/TAC.2015.2480549.

[14] M. A. H. Darwish, G. Pillonetto, and R. Tóth, *The quest for the right kernel in Bayesian impulse response identification: The use of OBFs*, Automatica, 87 (2018), pp. 318–329, https://doi.org/10.1016/j.automatica.2017.10.007.

[15] B. De Moor, P. De Gersem, B. De Schutter, and W. Favoreel, *DAISY: A database for identification of systems*, J. A, 38 (1997), pp. 4–5.

[16] A. De Santis and L. Farina, *Identification of positive linear systems with Poisson output transformation*, Automatica, 38 (2002), pp. 861–868, https://doi.org/10.1016/S0005-1098(01)00277-1.

[17] N. K. Dhingra, M. Colombino, and M. R. Jovanović, *Structured decentralized control of positive systems with applications to combination drug therapy and leader selection in directed networks*, IEEE Trans. Control Network Syst., 6 (2018), pp. 352–362, https://doi.org/10.1109/TCNS.2018.2820499.

[18] F. Dinuzzo, *Kernels for linear time invariant system identification*, SIAM J. Control Optim., 53 (2015), pp. 3299–3317, https://doi.org/10.1137/130920319.

[19] F. Dinuzzo and B. Schölkopf, *The representer theorem for Hilbert spaces: A necessary and sufficient condition*, in Advances in Neural Information Processing Systems 2012, 2012, pp. 189–196.

[20] X. Duan, S. Jafarpour, and F. Bullo, *Graph-theoretic stability conditions for Metzler matrices and monotone systems*, SIAM J. Control Optim., 59 (2021), pp. 3447–3471, https://doi.org/10.1137/20M131802X.

[21] G. Dullerud and R. Smith, *Sampled-data model validation: An algorithm and experimental application*, Int. J. Robust Nonlinear Control, 6 (1996), pp. 1065–1078, https://doi.org/10.1002/(SICI)1099-1239(199611)6:9/10⟨;1065::AID-RNC269⟩;3.0.CO;2-N.

[22] Y. Ebihara, D. Peaucelle, and D. Arzelier, *Decentralized control of interconnected positive systems using L1-induced norm characterization*, in 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), IEEE, 2012, pp. 6653–6658.

[23] Y. Ebihara, D. Peaucelle, and D. Arzelier, *Stability and persistence analysis of large scale interconnected positive systems*, in 2013 European Control Conference (ECC), IEEE, 2013, pp. 3366–3371.

[24] N. Everitt, G. Bottegal, and H. Hjalmarsson, *An empirical Bayes approach to identification of modules in dynamic networks*, Automatica, 91 (2018), pp. 144–151, https://doi.org/10.1016/j.automatica.2018.01.011.

[25] L. Farina and S. Rinaldi, *Positive Linear Systems: Theory and Applications*, Wiley, Hoboken, NJ, 2011.

[26] E. Fornasini and M. E. Valcher, *Reachability of a class of discrete-time positive switched systems*, SIAM J. Control Optim., 49 (2011), pp. 162–184, https://doi.org/10.1137/090757551.

[27] P. A. Fuhrmann, *A Polynomial Approach to Linear Algebra*, Springer, Cham, 2011.

[28] Y. Fujimoto, I. Maruta, and T. Sugie, *Extension of first-order stable spline kernel to encode relative degree*, IFAC-PapersOnLine, 50 (2017), pp. 14016–14021, https://doi.org/10.1016/j.ifacol.2017.08.2425.

[29] J. Gondzio, P. González-Brevis, and P. Munari, *Large-scale optimization with the primal-dual column generation method*, Math. Program. Comput., 8 (2016), pp. 47–82, https://doi.org/10.1007/s12532-015-0090-6.

[30] M. Grant and S. Boyd, *CVX: Matlab Software for Disciplined Convex Programming*, version 2.1, 2014, https://cvxr.com/cvx/.

[31] C. Grussler and A. Rantzer, *On second-order cone positive systems*, SIAM J. Control Optim., 59 (2021), pp. 2717–2739, https://doi.org/10.1137/20M1337454.
[32] C. Grussler, J. Umenberger, and I. R. Manchester, *Identification of externally positive systems*, in 2017 IEEE 56th Annual Conference on Decision and Control, IEEE, 2017, pp. 6549–6554.
[33] W. M. Haddad, V. Chellaboina, and Q. Hui, *Nonnegative and Compartmental Dynamical Systems*, Princeton University Press, Princeton, NJ, 2010.
[34] D. Hinrichsen and E. Plischke, *Robust stability and transient behaviour of positive linear systems*, Vietnam J. Math., 35 (2007), pp. 429–462.
[35] M. Khosravi, *Representer theorem for learning Koopman operators*, IEEE Trans. Automat. Control, 68 (2023), pp. 2995–3010, https://doi.org/10.1109/TAC.2023.3242325.
[36] M. Khosravi, A. Iannelli, M. Yin, A. Parsi, and R. S. Smith, *Regularized system identification: A hierarchical Bayesian approach*, IFAC-PapersOnLine, 53 (2020), pp. 406–411.
[37] M. Khosravi and R. S. Smith, *Kernel-based identification of positive systems*, in 2019 IEEE 58th Annual Conference on Decision and Control, IEEE, 2019, pp. 1740–1745, https://doi.org/10.1109/CDC40024.2019.9029276.
[38] M. Khosravi and R. S. Smith, *Convex nonparametric formulation for identification of gradient flows*, IEEE Control Syst. Lett., 5 (2021), pp. 1097–1102, https://doi.org/10.1109/LCSYS.2020.3000176.
[39] M. Khosravi and R. S. Smith, *Nonlinear system identification with prior knowledge on the region of attraction*, IEEE Control Syst. Lett., 5 (2021), pp. 1091–1096, https://doi.org/10.1109/LCSYS.2020.3005163.
[40] M. Khosravi and R. S. Smith, *On robustness of kernel-based regularized system identification*, IFAC-PapersOnLine, 54 (2021), pp. 749–754.
[41] M. Khosravi and R. S. Smith, *Diagonally Square Root Integrable Kernels in System Identification*, preprint, arXiv:2302.12929, 2023.
[42] M. Khosravi and R. S. Smith, *The existence and uniqueness of solutions for kernel-based system identification*, Automatica, 148 (2023), 110728, https://doi.org/10.1016/j.automatica.2022.110728.
[43] M. Khosravi and R. S. Smith, *Kernel-based identification with frequency domain side-information*, Automatica, 150 (2023), 110813, https://doi.org/10.1016/j.automatica.2022.110813.
[44] M. Khosravi and R. S. Smith, *Kernel-based impulse response identification with side-information on steady-state gain*, IEEE Trans. Automat. Control, 68 (2023), pp. 6401–6408, https://doi.org/10.1109/TAC.2023.3243099.
[45] M. Khosravi, M. Yin, A. Iannelli, A. Parsi, and R. S. Smith, *Low-complexity identification by sparse hyperparameter estimation*, IFAC-PapersOnLine, 53 (2020), pp. 412–417.
[46] U. Krause, *Positive Dynamical Systems in Discrete Time*, De Gruyter, Berlin, 2015.
[47] L. Ljung, T. Chen, and B. Mu, *A shift in paradigm for system identification*, Internat. J. Control, 93 (2020), pp. 173–180, https://doi.org/10.1080/00207179.2019.1578407.
[48] D G. Luenberger, *Introduction to Dynamic Systems; Theory, Models, and Applications*, Wiley, Hoboken, NJ, 1979.
[49] A. Marconato, M. Schoukens, and J. Schoukens, *Filter-based regularisation for impulse response modelling*, IET Control Theory Appl., 11 (2016), pp. 194–204, https://doi.org/10.1049/iet-cta.2016.0908.
[50] A. Oghbaee, B. Shafai, and S. Nazari, *Complete characterisation of disturbance estimation and fault detection for positive systems*, IET Control Theory Appl., 12 (2018), pp. 883–891, https://doi.org/10.1049/iet-cta.2017.0911.
[51] J. Peypouquet, *Convex Optimization in Normed Spaces: Theory, Methods and Examples*, Springer, Cham, 2015.
[52] G. Pillonetto and G. De Nicolao, *A new kernel-based approach for linear system identification*, Automatica, 46 (2010), pp. 81–93, https://doi.org/10.1016/j.automatica.2009.10.031.
[53] 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, 50 (2014), pp. 657–682, https://doi.org/10.1016/j.automatica.2014.01.001.
[54] A. Rantzer and M. E. Valcher, *A tutorial on positive systems and large scale control*, in 2018 IEEE Conference on Decision and Control, IEEE, 2018, pp. 3686–3697.
[55] A. Rantzer and M. E. Valcher, *Scalable control of positive systems*, Annu. Rev. Control. Robotics Auton. Syst., 4 (2021), pp. 319–341, https://doi.org/10.1146/annurev-control-061520-010621.
[56] T. Reis and E. Virnik, *Positivity preserving balanced truncation for descriptor systems*, SIAM J. Control Optim., 48 (2009), pp. 2600–2619, https://doi.org/10.1137/080734200.

[57] R. S. Risuleo, G. Bottegal, and H. Hjalmarsson, *A nonparametric kernel-based approach to Hammerstein system identification*, Automatica, 85 (2017), pp. 234–247, https://doi.org/10.1016/j.automatica.2017.07.055.

[58] R. S. Risuleo, F. Lindsten, and H. Hjalmarsson, *Bayesian nonparametric identification of Wiener systems*, Automatica, 108 (2019), 108480, https://doi.org/10.1016/j.automatica.2019.06.032.

[59] B. Roszak and E. J. Davison, *Necessary and sufficient conditions for stabilizability of positive LTI systems*, Syst. Control Lett., 58 (2009), pp. 474–481, https://doi.org/10.1016/j.sysconle.2009.02.003.

[60] B. Schölkopf, R. Herbrich, and A. J. Smola, *A generalized representer theorem*, in International Conference on Computational Learning Theory, Lecture Notes in Comput. Sci. 2111, Springer, Cham, 2001, pp. 416–426.

[61] B. Shafai, J. Chen, and M. Kothandaraman, *Explicit formulas for stability radii of nonnegative and Metzlerian matrices*, IEEE Trans. Automat. Control, 42 (1997), pp. 265–270, https://doi.org/10.1109/9.554408.

[62] R. Shorten, F. Wirth, and D. Leith, *A positive systems model of TCP-like congestion control: Asymptotic results*, IEEE/ACM Trans. Network., 14 (2006), pp. 616–629, https://doi.org/10.1109/TNET.2006.876178.

[63] 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. Inform. Theory, 58 (2012), pp. 3250–3265, https://doi.org/10.1109/TIT.2011.2182033.

[64] J. Umenberger and I. R. Manchester, *Scalable identification of stable positive systems*, in 2016 IEEE 55th Conference on Decision and Control, IEEE, 2016, pp. 4630–4635, https://doi.org/10.1109/CDC.2016.7798974.

[65] M. E. Valcher, *Reachability properties of continuous-time positive systems*, IEEE Trans. Automat. Control, 54 (2009), pp. 1586–1590, https://doi.org/10.1109/TAC.2009.2015556.

[66] G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia, 1990.

[67] D. Xiong, L. Chai, and J. Zhang, *Sparse system identification in pairs of pulse and Takenaka–Malmquist bases*, SIAM J. Control Optim., 58 (2020), pp. 965–985, https://doi.org/10.1137/18M1217474.

[68] Y. Zeinaly, J. H. van Schuppen, and B. D. Schutter, *Linear positive systems may have a reachable subset from the origin that is either polyhedral or nonpolyhedral*, SIAM J. Matrix Anal. Appl., 41 (2020), pp. 279–307, https://doi.org/10.1137/19M1268161.

[69] M. Zheng and Y. Ohta, *Positive FIR system identification using maximum entropy prior*, IFAC-PapersOnLine, 51 (2018), pp. 7–12, https://doi.org/10.1016/j.ifacol.2018.09.082.

[70] M. Zheng and Y. Ohta, *Bayesian positive system identification: Truncated Gaussian prior and hyperparameter estimation*, Systems Control Letters, 148 (2021), 104857, https://doi.org/10.1016/j.sysconle.2020.104857.

[71] M. Zorzi, *A second-order generalization of TC and DC kernels*, IEEE Trans. Automat. Control, 69 (2024), pp. 3835–3848, https://doi.org/10.1109/TAC.2023.3337056.

[72] M. Zorzi and A. Chiuso, *The harmonic analysis of kernel functions*, Automatica, 94 (2018), pp. 125–137, https://doi.org/10.1016/j.automatica.2018.04.015.