A kernel-based approach to errors-in-variables identification of stable multivariable linear systems

A kernel-based approach to errors-in-variables identification of stable multivariable linear systems

Vito Cerone*, *Member, IEEE,* Edoardo Fadda*, Diego Regruto*, *Member, IEEE*

***Abstract*—In this paper, we present a kernel-based non-parametric approach to identifying stable multi-input multi-output linear systems in the presence of bounded noise affecting both the input and the output measurements. Firstly, we formulate the considered problem in terms of robust optimization techniques. Then, we show that the formulated robust optimization problem can be solved using semidefinite optimization. Since the involved optimization problem is computationally demanding, we also provide a result that allows the user to compute a bound on the approximation error introduced by considering reduced complexity models. We present some simulation examples to show the effectiveness of the proposed approach. Finally, we apply the proposed identification method to the dataset experimentally collected on a linear electronic filter.**

I. Introduction

In this work, we focus our attention on the identification of linear time-invariant (LTI) systems. Roughly speaking, we define *system identification* as the branch of science aimed at deriving a mathematical model of a dynamic system based on a set of experimental input-output data (see, e.g., [1], [2]). Since we thoroughly describe LTI systems by their impulse response, system identification in the case of such a class of systems can be formulated as the problem of estimating the impulse response. However, in general, direct estimation of the impulse response of an LTI system is a difficult task since it leads to the solution of an infinite dimensional, possibly ill-conditioned, inverse problem (see, e.g., [3]). A possible approach to overcome such a problem is to consider model classes where we describe the input-output mapping using a finite number of parameters, e.g., ARX, ARMAX, Box-Jenkins, to cite a few [1], [2]. State-space descriptions are, instead, considered in the context of subspace identification of MIMO LTI systems (see, e.g., the survey paper [4] and the references therein). The choice of a model in the above classes substantially simplifies the identification problem from a computational point of view, as witnessed by the contributions in the literature [1]. Unfortunately, the approaches mentioned above potentially suffer from a significant drawback: the choice of the number of parameters to identify or the order of the transfer function. On the one hand, we should increase the number of parameters to improve the estimate accuracy. On the other hand, too many parameters may cause over-fitting of the identification data set, leading to models with a reduced capability to predict the behaviour of the actual system under study when we consider different experimental conditions. We can use some statistical techniques from the literature to deal with this issue, known as model order or model structure selection. Popular approaches for the estimation of the order of input-output models in the transfer function form exploit the Akaike information criterion (AIC) [5] and the Bayesian information criterion (BIC) [6]. Bauer discusses the problem of order selection in the framework of subspace identification in [7].

Linear and nonlinear non-parametric identification approaches, inspired by results in the field of machine learning, have been recently proposed to avoid the model order/model structure selection issue (see, e.g., [3], [8]–[10]). Such approaches address the problem of finding the impulse response function that provides the best data fitting in an infinite-dimensional space. The book by Pillonetto et al. [11] provides a comprehensive overview of the results obtained along this line of research. The identification problem formulation leverages Reproducing Kernel Hilbert spaces (RKHSs) in such a context. An RKHS is a Hilbert space (i.e. an infinite dimensional vector space equipped with a scalar inner product complete with respect to the norm induced by the scalar product) in which point evaluation is a continuous linear functional. This characteristic plays a crucial role in the context of RKHSs since it leads to the formulation of the *Representer Theorem*, thanks to which we can transform the infinite-dimensional problem of computing the impulse response function into an equivalent finite-dimensional one. A remarkable property is that a kernel function can completely define these spaces. This property implies that all functions in the considered RKHS enjoy the same properties of the kernel function. This fact is noteworthy in system identification because essential properties, such as stability and causality of the estimated model, can be a-priori enforced by selecting a kernel function enjoying such properties. The reader is referred to the survey [3] and the references therein for a detailed discussion. Extensions of the kernel-based approach to vector-valued functions are discussed in paper [12], where the authors review different methods to design and learn suitable kernel functions for multiple output problems.

Most of the results on Kernel-based identification available in the literature assume that the collected data are noise–free or at most affected by additive output error only. Such methods cannot deal with the errors-in-variables (EIV) identification problem, which is common in actual practice, where measurement errors affect inputs and outputs. The interested reader can find a thorough review of available results on the identification of parametric EIV models in the recent book by Söderström

* V. Cerone and D. Regruto are with Department of Control and Computer Engineering, E. Fadda is with Department of Mathematical Sciences, Politecnico di Torino, 10129 Torino, Italy vito.cerone@polito.it, edoardo.fadda@polito.it, diego.regruto@polito.it

[13]. A stimulating contribution considering the problem of kernel-based identification in the errors-in-variables (EIV) case is paper [14], where the authors consider noisy and incomplete input-output data and propose a statistical approach (based on the expectation-maximization method) to solve the identification problem.

All the works mentioned above rely on a statistical description of the noise affecting the data. A common alternative to the stochastic description, inspired by the seminal work of Schweppe, in the late sixties [15], is the bounded-error or set-membership characterization where measurement errors are assumed to be unknown but bounded (UBB), i.e., measurement uncertainties are assumed to belong to a given bounded set. We can choose such a description when either a priori statistical information is unavailable or the errors are better characterized deterministically (e.g., systematic and class errors in measurement equipment, rounding, and truncation errors in digital devices).

This paper proposes an original kernel-based non-parametric approach for identifying multi-input multi-output (MIMO) errors-in-variables (EIV) LTI systems. According to the set-membership (or bounded error) paradigm (see, e.g., [16]–[19]), both input and output sequences are affected by additive noise only known to belong to a given ball. This work significantly extends the preliminary contribution presented by the authors in the conference paper [20]. First, we generalize the proposed approach to MIMO systems, while [20] focuses only on SISO systems. Furthermore, we provide a result that allows the user to compute a bound on the approximation error introduced while reducing the number of basis functions involved in the model. We can use such a bound to handle the complexity/accuracy trade-off and reduce the relevant computational burden required to solve the underlying optimization problems. Finally, we apply the proposed identification algorithm to a dataset experimentally collected on a SISO and a MIMO linear electronic filter, while in [20], we have considered only simulated data. We organize the paper as follows. Section II reviews notations and fundamental results on RKHS. Section III presents the problem formulation. In Section IV, we propose a robust optimization-based algorithm to compute the solution to the formulated problem. The key ingredient of the proposed solution is the formulation of the robust optimization problem in terms of an equivalent semidefinite programming problem. In Section V, we discuss the problem of reducing the computational complexity of the proposed algorithm. We derive a bound on the approximation error introduced by reducing the number of basis functions exploited in the identification. Section VI addresses the Kernel selection problem. In Section VII, we show the effectiveness of the proposed approach through some simulation examples. Section VIII applies the presented identification method to datasets experimentally collected.

II. Notation and Basic Results

This section reviews some basic definitions and results related to RKHS. For details, the reader is referred, e.g., to [21], [22]. In the paper, vectors without subscripts denote stacks of signal samples collected for different time instants t.

In the following, we consider Hilbert Spaces whose elements are functions $g : \mathcal{T} \subset \mathbb{R} \to \mathbb{R}$, and we use the symbol $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ to represent the inner product and the symbol $||\cdot||_{\mathcal{H}}$ to denote the norm induced by such an inner product.

Definition 1: **Reproducing Kernel Hilbert Space (RKHS)** A Hilbert Space $\mathcal{H}$ is an RKHS iff the point evaluation operator is a continuous linear functional, i.e.

$$\forall\ t\ \exists\ M_t :\ |g(t)| \leq M_t ||g||_{\mathcal{H}}\ \forall\ g \in \mathcal{H} \qquad \blacksquare$$

In view of *Riesz Representation Theorem* every continuous linear functional can be uniquely written using the inner product [22] . The following property holds since the point evaluation functional is in $\mathcal{H}^*$ from the definition of RKHS .

***Theorem** 1:* **Reproducing Property**
If $\mathcal{H}$ is a RKHS, then: $\forall t \in \mathcal{T}\ \exists\ K_t \in \mathcal{H}$ such that $g(t) = \langle K_t, g \rangle_{\mathcal{H}}$ The function K_t is called the representer of the evaluation functionals at time t. $\blacksquare$

Definition 2: **Reproducing Kernel**
We call reproducing kernel of $\mathcal{H}$ a function $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ defined as $K(t_1, t_2) = K_{t_1}(t_2)\ \forall\ t_1, t_2 \in \mathcal{T}$, where $K_{t_1}(t_2) = \langle K_{t_1}, K_{t_2} \rangle_{\mathcal{H}}$. $\blacksquare$

It can be proved (see, e.g., [21], [22]) that every function $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ that is symmetric and positive semi-definite, i.e., such that it satisfies the following two conditions

$$K(t_1, t_2) = K(t_2, t_1)\ \forall\ t_1, t_2\ \in\ \mathcal{T} \tag{1}$$

$$\sum_{i,j=1}^{n} c_i c_j K(t_i, t_j) \geq 0\ \forall\ n \in \mathbb{N},\ t_1, .., t_n \in \mathcal{T},\ c_1, .., c_n \in \mathbb{R} \tag{2}$$

defines a reproducing kernel. The following theorem highlights the importance of reproducing kernels.

***Theorem** 2:* **Moore-Aronszajn theorem ([22])** Every RKHS defines a unique reproducing kernel. Conversely, a reproducing kernel defines a unique RKHS whose elements are functions defined as $g = \sum_{i=1}^{\infty} \alpha_i K_{t_i}$. $\blacksquare$

***Remark** 1:* In the rest of the paper, we assume, without loss of generality, that $\{K(t_i, t)\}_i$ are linearly independent.

Now, we are in the position of stating the Representer Theorem.

***Theorem** 3:* **The Representer Theorem ([22])**
Given a set of N pair of points $(t_1, y_1), \ldots, (t_N, y_N)$ and a generic functional $V(\cdot) : \mathbb{R}^N \to \mathbb{R}$, a necessary condition for $g \in \mathcal{H}$ to be the minimizer of

$$V(g(t_1), \ldots, g(t_N), y) + \zeta ||g||^2_{\mathcal{H}} \tag{3}$$

is $g = \sum_{i=1}^{N} \alpha_i K_{t_i}$, for some N-tuple $(\alpha_1, \ldots, \alpha_N) \in \mathbb{R}^N$. $\blacksquare$

For self-consistency of the paper, we also report the proof of Theorem 3 since the reasoning exploited in the demonstration can be helpful to better understand some technical details of the results presented in the following sections.
The following version of the proof is taken from [23].

Proof: Define $\mathcal{H}_0$ as

$$\mathcal{H}_0 = \{g \ \in \ \mathcal{H} \ \ : \ \ g = \sum_{i=1}^{N} \alpha_i K_{t_i}\} \ \subset \ \mathcal{H}.$$

Let $\mathcal{H}_0^\perp$ be the space orthogonal to $\mathcal{H}_0$

$$\mathcal{H}_0^\perp = \{g^\perp \ \in \ \mathcal{H} \ \ : \langle g^\perp, g\rangle_{\mathcal{H}} = 0 \ \ \forall \ g \in \mathcal{H}_0 \ \}.$$

$\mathcal{H}_0$ is a subspace of finite dimension; hence it is closed. For this reason, we can write $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_0^\perp$ and further, we can express each function $g \in \mathcal{H}$ as $g = g_0 + g_0^\perp$ where $g_0 \in \mathcal{H}_0$ and $g_0^\perp \in \mathcal{H}_0^\perp$. By orthogonality, the following relation holds

$$||g||_{\mathcal{H}}^2 = ||g_0||_{\mathcal{H}}^2 + ||g_0^\perp||_{\mathcal{H}}^2.$$

Furthermore, by exploiting Definition 2, we can write

$$\begin{aligned} V(g(t_1), \dots, g(t_N), y) &= V(\langle g, K_{t_1}\rangle_{\mathcal{H}}, \dots, \langle g, K_{t_N}\rangle_{\mathcal{H}}, y) = \\ &= V(\langle g_0 + g_0^\perp, K_{t_1}\rangle_{\mathcal{H}}, \dots, \langle g_0 + g_0^\perp, K_{t_N}\rangle_{\mathcal{H}}, y) = \\ &= V(\langle g_0, K_{t_1}\rangle_{\mathcal{H}}, \dots, \langle g_0, K_{t_N}\rangle_{\mathcal{H}}, y). \end{aligned}$$

where the last equality follows from the fact that $g_0^\perp$ is orthogonal to all the elements in $\mathcal{H}_0$, particularly to K_{t_i}. Hence, if we calculate the function V for a generic $g \in \mathcal{H}$, we can write

$$\begin{aligned} &V(g(t_1), \dots g(t_N), y) + \zeta||g||_{\mathcal{H}}^2 = \\ &= V(g_0(t_1), \dots, g_0(t_N), y) + \zeta\big(||g_0||_{\mathcal{H}}^2 + ||g_0^\perp||_{\mathcal{H}}^2\big) \geq \\ &\geq V(g_0(t_1), \dots g_0(t_N), y) + \zeta||g_0||_{\mathcal{H}}^2. \end{aligned}$$

The last inequality proves that the minimizer of the functional must belong to the linear space $\mathcal{H}_0$. ■

***Remark** 2:* The Representer Theorem shows that, for RKHS, minimizing over a (possibly infinite-dimensional) Hilbert space boils down to minimizing over $\mathbb{R}^N$, where N is the number of sample points considered in the functional to be minimized.

***Remark** 3:* We point out that the proof of this theorem was obtained without considering the nature of the function $V(g_0(t_1), \dots g_0(t_N), y)$.

Moreover, the following corollary holds.

Corollary 4: The minimizers $g_1, \dots, g_I \in \mathcal{H}_1, \dots, \mathcal{H}_I$ of the problem

$$\min_{g_1,\dots,g_I} \Big[V(g_1(t_1), \dots, g_1(t_N), \dots, g_I(t_1), \dots, g_I(t_N), y) + \zeta(||g_1||_{\mathcal{H}_1}^2 + \dots + ||g_I||_{\mathcal{H}_I}^2)\Big]$$

can be written as $g_j = \sum_{i=1}^{N} \alpha_i^j K_{t_i}^j \ \ \forall \ j = 1 \dots I$, for I N-tuples $(\alpha_1^1, \dots, \alpha_N^1) \in \mathbb{R}^N, \dots, (\alpha_1^I, \dots, \alpha_N^I) \in \mathbb{R}^N$. ■

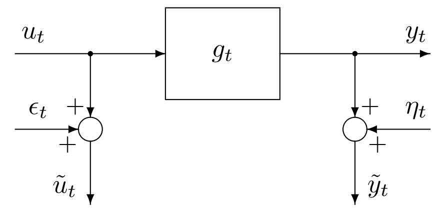

Fig. 1. Errors-in-variables basic setup for non-parametric LTI identification

Proof: The proof of Corollary 4 is obtained by applying the same reasoning exploited in the proof of Theorem 3 to each function $g_j, \forall \ j \ = \ 1, 2, \dots, I$. ■

In the next sections, we show that *Theorem 3* is the key result for deriving a suitable formulation of system identification problems in the framework of RKHS.

III. Problem Formulation

To simplify notation, in the rest of the paper we represent time dependence using subscripts (i.e., we write g_t instead of $g(t)$). Let us consider the discrete-time multiple input multiple output (MIMO) LTI dynamic system shown in Fig. 1, where $u_t \ \in \ \mathbb{R}^p$, $y_t \ \in \ \mathbb{R}^q$, and g_t is the continuous-time impulse response function.

Since the problem of identifying MIMO systems with p inputs and q outputs is equivalent to the identification of the q MISO transfer functions that map the p inputs to every single output, in the rest of the paper, we consider, without loss of generality, the problem of estimating the parameters of MISO systems described by the following general input-output mapping

$$y_t = \sum_{s=0}^{\infty} u_{1,t-s} g_{1,s} + \dots + \sum_{s=0}^{\infty} u_{p,t-s} g_{p,s} \tag{4}$$

where $g_1 = [g_{1,1}, \dots, g_{1,t}, \dots]$, ..., $g_p = [g_{p,1}, \dots, g_{p,t}, \dots]$ are the p impulse response functions describing the effect of the p inputs on the output y, and each element of g_j is $g_{j,t} = g_j(t)$.

We assume that $u_{jt} \ = \ 0$ for all $t \ < \ 0$; therefore the convolutions in (4) go from $s = 0$ to $s = t$. Measurements of the input and output data sequences are assumed to be corrupted by additive noises $\epsilon_{j,t}$ and η_t respectively

$$\tilde{u}_{j,t} = u_{j,t} + \epsilon_{j,t}, \ \forall j = 1, \dots, p \qquad \tilde{y}_t = y_t + \eta_t. \tag{5}$$

The error signals $\epsilon_j \ = \ [\epsilon_{j,1} \dots \epsilon_{j,N}]$, $\forall j \ = \ 1, \dots p$, and $\eta \ = \ [\eta_1 \dots \eta_N]$ are bounded and assumed to belong to the following balls:

$$\epsilon_j \in \mathcal{B}_{\rho_j} = \{\epsilon_j : \ ||\epsilon_j||_2^2 \leq \ \rho_j\} \ \forall j = 1, \dots, p \tag{6}$$

$$\eta \in \mathcal{B}_{\rho_q} = \{\eta : \ ||\eta||_2^2 \leq \ \rho_q\}, \tag{7}$$

where $\rho_1, \dots, \rho_p$ and ρ_q are known real constants. In common practice, we can infer the value of such error bounds from available measurement error information provided by the data sheets of the equipment used to collect the data. In applications where such information is not directly available, we must perform experiments to characterize the measurement device

and estimate such bounds. We point out that a similar problem also arises when considering a statistical description of the noise, where we have to infer the probability distribution of the measurement error or at least some of its moments.
The system to identify is only known to be LTI. No structural information (e.g., model order) is assumed to be *a-priori* available, and the model to identify is not constrained to belong to a finite-dimensional model class (e.g., FIR, ARX). Therefore, according to the kernel-based nonparametric identification paradigm (see, e.g., [3], [8]–[10]), the following nonparametric LTI identification problem is considered here

$$\hat{g} = \arg\min_{g_1,\ldots,g_p \in \mathcal{H}} \left\{ \sum_{t=0}^{N} \Big(y_t - \hat{y}_t(g_1,\ldots,g_p) \Big)^2 + \zeta J(g) \right\} \quad (8)$$

where $\mathcal{H}$ is a generic reproducing kernel Hilbert space, $\hat{y}$ is the output of the model to be estimated, $J(g)$ is a regularization term, and ζ is a positive real constant to weight the regularization term. Constraining the impulse response to lay inside an RKHS has several advantages (the reader can find a detailed discussion in the survey paper [3] and the reference therein). The main property of those spaces is that we can define them through a function called kernel. This property implies that all functions in the considered RKHS enjoy the same properties of the kernel function, thank to which, we can *a priori* enforce essential properties (such as stability and causality) by selecting the kernel function suitably. For a thorough discussion of the properties that can be enforced by adequately choosing the kernel function, we refer the reader to work [24].

By means of equations (4) and (5), we can write the estimation Problem (8) as

$$\hat{g}_1,\ldots,\hat{g}_p = \arg\min_{g_1,\ldots,g_p \in \mathcal{H}} \left\{ \sum_{t=0}^{N} \Big(\tilde{y}_t - \eta_t - \sum_{j=0}^{p}\sum_{s=0}^{t} \tilde{\epsilon}_{j,t-s} g_{j,s} \Big)^2 + \zeta J(g) \right\} \quad (9)$$

where, $\tilde{\epsilon}_{j,t} = \big(\tilde{u}_{j,t} - \epsilon_{j,t}\big)$ and

$$\tilde{\epsilon}_{j,t} \in \tilde{\mathcal{B}}_{\rho_j} = \{\tilde{\epsilon}_j : \tilde{\epsilon}_{j,t} = \big(\tilde{u}_{j,t} - \epsilon_{j,t}\big),\ ||\epsilon_j||_2^2 \leq\ \rho_j\} \quad \forall j = 1,2\ldots,p. \quad (10)$$

To simplify notation, in the rest of the paper we use the short form $\tilde{\epsilon} \in \tilde{\mathcal{B}}$ instead of (10). Since η_t and $\tilde{\epsilon}_{j,t},\ \forall j = 1,2,\ldots,p$, are uncertain variables only known to belong to the balls (7) and (10), we estimate the impulse response function by solving the following robust estimation problem, where we look for the minimum of the worst-case error between the actual unknown output of the system and the output of the estimated model

$$\hat{g}_1^r,\ldots,\hat{g}_p^r = \arg\min_{g_1,\ldots,g_p \in \mathcal{H}} \max_{\substack{\tilde{\epsilon}\in\tilde{\mathcal{B}},\\ \eta_t \in \mathcal{B}_{\rho_2}}} \left\{ \sum_{t=0}^{N} \Big(\tilde{y}_t - \eta_t - \sum_{j=1}^{p}\sum_{s=0}^{t} \tilde{\epsilon}_{j,t-s} g_{j,s} \Big)^2 + \zeta J(g) \right\}, \quad (11)$$

where $\hat{g}_1^r,\ldots,\hat{g}_p^r$ is the robust version of the estimate defined in Problem (9). For the sake of simplicity, we introduce the notation

$$\max_{\tilde{\epsilon}_{j,t},\eta_t} F(\tilde{\epsilon}_{j,t},\eta_t) = \max_{\tilde{\epsilon}\in\tilde{\mathcal{B}},\eta_t\in\mathcal{B}_{\rho_2}} F(\tilde{\epsilon}_{j,t},\eta_t),$$

where $F(\tilde{\epsilon}_{j,t},\eta_t)$ is a generic functional.

IV. A Robust Optimization Approach

In this section, we propose an original robust convex optimization approach to solve Problem (11).

For simplicity and without loss of generality, we refer to the following simplified version of the problem

$$\hat{g}_1^r,\ldots,\hat{g}_p^r = \arg\min_{g_1,\ldots,g_p \in H} \max_{\tilde{\epsilon}_{j,t},\eta_t} \sum_{t=0}^{N} \Big(\tilde{y}_t - \eta_t - \sum_{j=0}^{p}\sum_{s=0}^{t} \tilde{\epsilon}_{j,t-s} g_{j,s} \Big)^2, \quad (12)$$

obtained from (11) by removing the regularization term. We discuss the extension to the regularized version of the problem in Remark 5, reported at the end of Section IV.

The first step towards the derivation of the proposed algorithm is the following result, derived on the basis of the properties of RKHS.

***Result** 1:* The minimum of optimization Problem (12) is given by

$$\min_{\alpha^j\in\mathbb{R}^N, j=1,\ldots,p} \max_{\tilde{\epsilon}_{j,t},\eta_t} \sum_{t=0}^{N} \Big(\tilde{y}_t - \eta_t - \sum_{j=0}^{p}\sum_{s=0}^{t} \tilde{\epsilon}_{j,t-s} \sum_{i=0}^{N} \alpha_i^j K_s^{i,j} \Big)^2 \quad (13)$$

where $K_t^{i,j} = K_{t_i}(t)^j$ and $K_{t_i}(t)^j$ is the evaluation at time t of the kernel function used for the j-th input. Furthermore, the minimizer has the following form for each t

$$g_{j,t} = \sum_{i=0}^{N} \alpha_i^j K_t^{i,j}. \quad (14)$$

Proof: First, we note that the argument inside the $\arg\min$ operator of (12) depends only on the point evaluation of function $g_j\ \forall\ j = 1,2,\ldots,p$. Therefore, by applying Corollary 4, we can prove that the minimizer to Problem (12) has the form (14). Equation (13) is obtained by direct substitution of (14) into the functional of Problem (12). ■

Thanks to Result 1 and the fact that $\tilde{\epsilon}_{j,t-s}$ does not depend on i, we can write the robust optimization to be solved as follows

$$[\hat{\alpha}^1,\ldots,\hat{\alpha}^p] = \arg\min_{\alpha^j\in\mathbb{R}^N, j=1,\ldots,p} \max_{\tilde{\epsilon}_{j,t},\eta_t} \mathcal{J} = \arg\min_{\alpha^j\in\mathbb{R}^N, j=1,\ldots,p} \max_{\tilde{\epsilon}_{j,t},\eta_t} \sum_{j=0}^{p}\sum_{t=0}^{N} \Big(\tilde{y}_t - \eta_t - \sum_{s=0}^{t}\sum_{i=0}^{N} \alpha_i^j \tilde{\epsilon}_{j,t-s} K_s^{i,j} \Big)^2, \quad (15)$$

where $\mathcal{J}$ is implicitly defined in (15). Before stating the main results of the paper, we rewrite the functional $\mathcal{J}$ of (15) in a suitable equivalent form. To this aim, we first explicitly write the terms of the summation in (15) for some sample $t = 0,1,\ldots$

$$\tilde{y}_0 - \eta_0 - \sum_{j=1}^{p} \tilde{\epsilon}_{j,0} \sum_{i=0}^{N} \alpha_i^j K_0^{i,j},$$

$$\tilde{y}_1 - \eta_1 - \sum_{j=1}^{p} \tilde{\epsilon}_{j,1} \sum_{i=0}^{N} \alpha_i^j K_0^{i,j} - \sum_{j=1}^{p} \tilde{\epsilon}_{j,0} \sum_{i=0}^{N} \alpha_i^j K_1^{i,j},\ldots$$

By grouping the terms for different t into suitable matrices and arrays, we obtain the following equivalent description

$$\mathcal{J} = \left|\left|\tilde{y} - \eta - E^1\alpha^1 - \cdots - E^p\alpha^p\right|\right|_2^2,$$

where $E^j \in \mathbb{R}^{N\times N}$ has the following form

$$E^j = \begin{bmatrix} \tilde{\epsilon}_{j,0}K_0^{0,j} & \dots & \tilde{\epsilon}_{j,0}K_0^{N,j} \\ \tilde{\epsilon}_{j,1}K_0^{0,j} + \tilde{\epsilon}_{j,0}K_1^{0,j} & \dots & \tilde{\epsilon}_{j,1}K_0^{N,j} + \tilde{\epsilon}_{j,0}K_1^{N,j} \\ \dots & \dots & \dots \\ \sum_{i=1}^{N}\tilde{\epsilon}_{j,N-i}K_i^{0,j} & \dots & \sum_{i=1}^{N}\tilde{\epsilon}_{j,N-i}K_i^{N,j} \end{bmatrix}.$$

By defining the following matrices

$$A_0^j = \begin{bmatrix} K_0^{0,j} & \dots & K_0^{N,j} \\ K_1^{0,j} & \dots & K_1^{N,j} \\ \dots & \dots & \dots \\ K_N^{0,j} & \dots & K_N^{N,j} \end{bmatrix}, \dots, A_N^j = \begin{bmatrix} 0 & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & 0 \\ K_0^{0,j} & \dots & K_0^{N,j} \end{bmatrix},$$

matrix E^j can be rewritten as

$$E^j = \tilde{\epsilon}_{j,0}A_0^j + \tilde{\epsilon}_{j,1}A_1^j + \cdots + \tilde{\epsilon}_{j,N}A_N^j,$$

while, by substituting $\tilde{\epsilon}_{j,t} = \tilde{u}_{j,t} - \epsilon_{j,t}$, we get

$$E^j = \tilde{u}_{j,0}A_0^j + \cdots + \tilde{u}_{j,N}A_N^j - \epsilon_{j,0}A_0^j - \cdots - \epsilon_{j,N}A_N^j.$$

By defining $E_0^j = \tilde{u}_{j,0}A_0^j + \tilde{u}_{j,1}A_1^j + \cdots + \tilde{u}_{j,N}A_N^j$ we can write

$$E^j = E_0^j - \epsilon_{j,0}A_0^j - \cdots - \epsilon_{j,N}A_N^j.$$

Thanks to the introduction of a slack variable $\lambda \in \mathbb{R}$, we can rewrite optimization Problem (15) in its final form:

$$\begin{aligned} [\hat{\alpha}^1, \dots, \hat{\alpha}^p] = & \operatorname*{arg\,min}_{\alpha^j \in \mathbb{R}^N, j=1,\dots,p, \lambda\in\mathbb{R}} \lambda \\ \text{s.t. } & \left|\left|\tilde{y} - \eta - \sum_{j=0}^{p}(E_0^j - \epsilon_{j,0}A_0^j - \cdots - \epsilon_{j,N}A_N^j)\alpha^j\right|\right|_2^2 \le \lambda \\ & \forall\ \epsilon_j \in \mathcal{B}_{\rho_j},\ \eta \in \mathcal{B}_{\rho_q} \end{aligned} \quad (16)$$

Finally, we are in the position of stating the main result.

***Result** 2:* The solution to Problem (16) can be obtained through the solution to the convex semidefinite optimization problem (SDP) in (17) (reported at the top of next page) where:

$$M(\alpha^j) = \begin{bmatrix} \vdots & \vdots & \vdots \\ A_0^j\alpha^j & \vdots & A_N^j\alpha^j \\ \vdots & \vdots & \vdots \end{bmatrix}.$$

The proof of Result 2 is given in Appendix A. Note that the matrix defining the constraints of Problem (17) belongs to $\mathbb{R}^{(N(I+2)+1)\times(N(I+2)+1)}$.

***Remark** 4:* For SISO systems, Problem (17) simplifies to

$$\begin{aligned} \hat{\alpha} = \arg & \min_{\alpha\in\mathbb{R}^N, \lambda\in\mathbb{R}, \tau_0,\tau_1\in\mathbb{R}_+} \lambda \\ s.\ t.\ & \\ & \left[\begin{array}{c|c|c|c} \tau_0 I - I & +M(\alpha) & (\tilde{y} - E_0\alpha) & 0 \\ \hline +M(\alpha)^T & \tau_1 I & 0 & M(\alpha)^T \\ \hline (\tilde{y} - E_0\alpha) & 0 & \lambda - \tau_0\rho_0 - \tau_1\rho_1 & (\tilde{y} - E_0\alpha)^T \\ \hline 0 & M(\alpha) & (\tilde{y} - E_0\alpha) & I \end{array}\right] \ge 0. \end{aligned} \quad (18)$$

***Remark** 5:* As already pointed out at the beginning of this section, we still have to consider the regularisation terms in deriving the proposed approach. In this remark, we discuss how to incorporate Lasso and Ridge regularizations. If we include the Lasso regularization term, the problem becomes

$$\begin{aligned} \hat{\alpha} = & \operatorname*{arg\,min}_{\alpha\in\mathbb{R}^N, \lambda\in\mathbb{R}, \tau_0,\tau_1\in\mathbb{R}_+} \lambda + \zeta\sum_i |\alpha_i| \\ s.\ t.\ & \\ & \left[\begin{array}{c|c|c|c} \tau_0 I - I & +M(\alpha) & (\tilde{y} - E_0\alpha) & 0 \\ \hline +M(\alpha)^T & \tau_1 I & 0 & M(\alpha)^T \\ \hline (\tilde{y} - E_0\alpha) & 0 & \lambda - \tau_0\rho_0 - \tau_1\rho_1 & (\tilde{y} - E_0\alpha)^T \\ \hline 0 & M(\alpha) & (\tilde{y} - E_0\alpha) & I \end{array}\right] \ge 0 \end{aligned} \quad (19)$$

where $\zeta \in \mathbb{R}$ is the parameter used to tune the regularization.

We can consider the regularization term by adding linear constraints to the optimization problem, which still is a standard SDP problem, and all the presented results and discussion apply.

If we include a Ridge regularization term, we have to perform further transformations. In this case, the problem is given by

$$\begin{aligned} \hat{\alpha} = & \operatorname*{arg\,min}_{\alpha\in\mathbb{R}^N, \lambda\in\mathbb{R}, \tau_0,\tau_1\in\mathbb{R}_+} \lambda + \zeta||\alpha||^2 \\ s.\ t.\ & \\ & \left[\begin{array}{c|c|c|c} \tau_0 I - I & +M(\alpha) & (\tilde{y} - E_0\alpha) & 0 \\ \hline +M(\alpha)^T & \tau_1 I & 0 & M(\alpha)^T \\ \hline (\tilde{y} - E_0\alpha) & 0 & \lambda - \tau_0\rho_0 - \tau_1\rho_1 & (\tilde{y} - E_0\alpha)^T \\ \hline 0 & M(\alpha) & (\tilde{y} - E_0\alpha) & I \end{array}\right] \ge 0. \end{aligned} \quad (20)$$

To move the additional regularization term into constraints, we have to introduce a new slack variable λ_1 such that $\zeta||\alpha||_{\mathcal{H}}^2 \le \lambda_1$.

In this way, we can rewrite the objective function as $\lambda + \lambda_1$ and, by noticing that $||\alpha||_{\mathcal{H}}^2 = \alpha^T G\alpha$ where G is the Gram matrix, which is symmetric positive definite, we can transform the quadratic constraint into the following SDP constraint

$$\left[\begin{array}{c|c} I & G^{1/2}\alpha \\ \hline \alpha^T(G^{1/2})^T & \lambda_1/\zeta \end{array}\right] \ge 0. \quad (21)$$

V. Model complexity reduction

In this section, we present a method to reduce the computational burden of the approach proposed in Section IV. For clarity and without loss of generality, we only present a detailed derivation for the SISO case. We discuss a straightforward generalization to the MISO case at the end of the section.

Since the size of the LMI constraint in (18) depends on the length N of the collected data sequence, SDP Problem (18) might become practically unsolvable for large values of N. In this case, a possible approach to reduce the computational burden of (18) is using a number $d < N$ of basis functions, leading to a reduced complexity model. More specifically, SDP Problem (18) is solvable in polynomial time with fixed

$$[\hat{\alpha}^1, \dots, \hat{\alpha}^p] = \underset{\substack{\alpha^j \in \mathbb{R}^N, j=1,\dots,p, \lambda \in \mathbb{R}, \\ \tau_i \in \mathbb{R}_+, i=0,\dots,p}}{\arg\min} \lambda$$

s. t.

$$\left[\begin{array}{c|c|c|c|c|c}
\tau_0 I - I & M(\alpha^1) & \dots & M(\alpha^p) & \tilde{y} - (E_0^1 \alpha^1 + \dots + E_0^I \alpha^p) & 0 \\ \hline
M(\alpha^1)^T & \tau_1 I & \dots & 0 & 0 & M(\alpha^1)^T \\ \hline
M(\alpha^2)^T & 0 & \dots & 0 & 0 & M(\alpha^2)^T \\ \hline
\dots & \dots & \dots & \dots & \dots & \dots \\ \hline
M(\alpha^p)^T & 0 & \dots & \tau_p I & 0 & M(\alpha^p)^T \\ \hline
[\tilde{y} - (E_0^1 \alpha^1 + \dots + E_0^I \alpha^p)]^T & 0 & 0 & 0 & \lambda - \tau_0\rho_0 - \tau_1\rho_1 - \dots - \tau_p\rho_p & [\tilde{y} - (E_0^1 \alpha^1 + \dots + E_0^I \alpha^p)]^T \\ \hline
0 & M(\alpha^1) & \dots & M(\alpha^p) & \tilde{y} - (E_0^1 \alpha^1 + \dots + E_0^I \alpha^p) & I
\end{array}\right] \geq 0 \tag{17}$$

accuracy. In particular, by applying an interior point method, the complexity for each iteration is $\mathcal{O}(N^6)$ (see [25]). Thus, by using d basis functions instead of N, the complexity for each iteration becomes $\mathcal{O}(d^6)$. For example, by considering $d = \frac{3}{4}N$, we reduce the number of iterations to about 18% of those related to the case $d = N$ considered in the previous section.

However, this approach introduces an approximation error differently from the optimal solution presented in the previous section, which considers exactly N basis functions. To quantify such an approximation error, we consider the following definition.

Definition 3: We define the approximation error L as

$$L = || \sum_{i=1}^{N} \alpha_i K_i - \sum_{i=1}^{d_0} \beta_i K_i ||^2,$$

where α_i and β_i are the coefficients of the optimal solutions obtained with the full-order model exploiting all the N kernel basis functions and the reduced-complexity model obtained using only d_0 basis functions, respectively.
The following theorem provides a bound on the approximation error E.

***Theorem** 5:* Given Problem (16) for $p = 1$, let us assume that $|u_0| > \rho_0$ and that we obtain a solution with $d_1 < N$ kernel basis functions. Then, if we use $d_0 \leq d_1$ kernel basis functions, the approximation error is bounded as follows:

$$L = || \sum_{i=1}^{N} \alpha_i K_i - \sum_{i=1}^{d_0} \beta_i K_i ||^2 \leq \frac{||\tilde{y} - \hat{y}||^2 + \rho_q^2}{(\tilde{u}_0 - \mathrm{sign}(\tilde{u}_0)\rho_0)^2} + \sum_{i=d_0}^{d_1} |\gamma_i|^2 \tag{22}$$

for all $d_0 \leq d_1$; γ_i are the coefficients of the optimal solutions obtained by applying the Gram-Schmidt orthonormalization to the kernel basis functions.

Proof: From (1), the linear space generated from N sections of the kernel has dimension N. If we apply the Gram-Schmidt orthonormalization procedure to the kernel sections, we find an orthonormal basis. We call the elements of this basis $\{\phi_t^i\}_i$. Thanks to the introduction of the orthonormal basis $\{\phi_t^i\}_i$, the first d coefficients of the optimal solution obtainable with N basis functions are precisely those of the optimal solution to the problem obtained using the first d basis functions. Hence, through a change of basis, from $\{K_t^i\}_i$ to $\{\phi_t^i\}_i$, we obtain

$$|| \sum_{i=1}^{N} \alpha_i K_t^i - \sum_{i=1}^{d_0} \beta_i K_t^i ||^2 = || \sum_{i=1}^{N} \gamma_i \phi_t^i - \sum_{i=1}^{d_0} \gamma_i \phi_t^i ||^2 = \sum_{i=d_0}^{N} |\gamma_i|^2. \tag{23}$$

We cannot calculate this quantity because we assume that the optimal solution with N orthogonal functions is unavailable. Nevertheless, we can obtain a bound of this approximation error by observing that

$$\begin{aligned}
&|| \sum_{i=1}^{N} \gamma_i \phi_i - \sum_{i=1}^{d_0} \gamma_i \phi_i ||^2 = \\
&= || \sum_{i=1}^{N} \gamma_i \phi_i - \sum_{i=1}^{d_1} \gamma_i \phi_i + \sum_{i=1}^{d_1} \gamma_i \phi_i - \sum_{i=1}^{d_0} \gamma_i \phi_i ||^2 \\
&\leq || \sum_{i=1}^{N} \gamma_i \phi_i - \sum_{i=1}^{d_1} \gamma_i \phi_i ||^2 + || \sum_{i=1}^{d_1} \gamma_i \phi_i - \sum_{i=1}^{d_0} \gamma_i \phi_i ||^2 \\
&\leq || g - \sum_{i=1}^{d_1} \gamma_i \phi_i ||^2 + || \sum_{i=1}^{d_1} \gamma_i \phi_i - \sum_{i=1}^{d_0} \gamma_i \phi_i ||^2 = \\
&= || g - \sum_{i=1}^{d_1} \gamma_i \phi_i ||^2 + \sum_{i=d_0}^{d_1} |\gamma_i|^2
\end{aligned} \tag{24}$$

where we can approximate the first term using the objective function of the problem obtained by considering a base of d_1 elements. In fact,

$$\begin{aligned}
&||y - \hat{y}||^2 = ||u * g - u * \hat{g}||^2 = ||u * (g - \hat{g})||^2 = \\
&= || \sum_{s=0}^{t} u_{t-s}(g_t - \hat{g}_t) ||^2.
\end{aligned} \tag{25}$$

If we set $g_t - \hat{g}_t = \tilde{g}$ we can write (see, e.g., [26])

$$||y - \hat{y}||^2 = ||U\tilde{g}||^2 \geq \Lambda_{\min}(U^T U) ||\tilde{g}||^2. \tag{26}$$

where U is the Toeplitz matrix of the input i.e.

$$U = \begin{bmatrix} u_0 & 0 & \dots & \dots & 0 \\ u_1 & u_0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ u_N & \dots & \dots & \dots & u_0 \end{bmatrix}. \tag{27}$$

and $\Lambda_{\min}(UU^T)$ is the smallest eigenvalue of UU^T. Let us define $A = U^T U$. Since U has rank N, A is positive definite and, therefore, $\Lambda_{\min}(UU^T) = \Lambda_{\min}(A)$ is positive. Furthermore, since A is symmetric, we can write $A = V\Sigma V^T$, where Σ is a diagonal matrix whose entries are the eigenvalues of A and V is a unitary matrix. We can compute the diagonalisation of A as follows. First, we rewrite A as

$$A = U^T U = W^T(u_0 I)(u_0 I)W, \tag{28}$$

where

$$W = [w^1, w^2, \dots, w^N] = \begin{bmatrix} 1 & 0 & \dots\dots & 0 \\ \frac{u_1}{u_0} & 1 & 0\ \dots & 0 \\ \frac{u_2}{u_0} & \frac{u_1}{u_0} & \dots\dots & \dots \\ \frac{u_N}{u_0} & \frac{u_{N-1}}{u_0} & \dots\dots & 1 \end{bmatrix}. \tag{29}$$

Then, to obtain unitary vectors V, we rewrite (28) as follows:

$$A = V^T W_N^{-1}(u_0^2 I)W_N^{-1}V. \tag{30}$$

where $V = W_N W$ is the normalized W matrix, and

$$W_N = \begin{bmatrix} \frac{1}{||w^1||} & 0 & \dots\dots & 0 \\ 0 & \frac{1}{||w^2||} & 0\ \dots & 0 \\ \dots & \dots & \dots\dots & \dots \\ 0 & \dots & \dots\dots & 1 \end{bmatrix} \tag{31}$$

$$W_N^{-1} = \begin{bmatrix} ||w^1|| & 0 & \dots\dots & 0 \\ 0 & ||w^2|| & 0\ \dots & 0 \\ \dots & \dots & \dots\dots & \dots \\ 0 & \dots & \dots\dots & 1 \end{bmatrix}. \tag{32}$$

We finally obtain the diagonalization $A = V\Sigma V^T$, with $\Sigma = W_N^{-1}(u_0^2 I)W_N^{-1}$ and $V = W_N W$. The N eigenvalues of the matrix A are the entries of the diagonal matrix Σ given by $||w^1||^2 u_0^2 = (u_0^2 + u_1^2 + \dots + u_N^2)$, $||w^2||^2 u_0^2 = (u_0^2 + u_1^2 + \dots + u_{N-1}^2)$, $\dots$, u_0^2. Then, the smallest eigenvalue is given by

$$\Lambda_{\min}(UU^T) = u_0^2 = (\tilde{u}_0 - \epsilon_0)^2. \tag{33}$$

By putting together (33) and (26) we obtain the following bound

$$||(g - \hat{g})||^2 \le \frac{||y - \hat{y}||^2}{(\tilde{u}_0 - \epsilon_0)^2}. \tag{34}$$

Since $|u_0| \ge \rho_0$, the maximum value of such a bound is achieved for $\epsilon_0 = \text{sign}(\tilde{u}_0)\rho_0$. Furthermore, since y is not exactly measurable, we replace y with $\tilde{y} - \eta$ in (34), which takes to

$$||(g - \hat{g})||^2 \le \frac{||\tilde{y} - \eta - \hat{y}||^2}{(\tilde{u}_0 - \text{sign}(\tilde{u}_0)\rho_0)^2} \le \frac{||\tilde{y} - \hat{y}||^2 + \rho_q^2}{(\tilde{u}_0 - \text{sign}(\tilde{u}_0)\rho_0)^2}. \tag{35}$$

We can evaluate the second term of (35) by considering the difference between the two solutions after a change of basis. Finally, we derive (22) by putting together (35) and the last row of (24).

■

***Remark** 6:* We can easily handle the model complexity and accuracy trade-off with the proposed approach since d is the only parameter to be selected, thanks to the result reported in Theorem 7.

***Remark** 7:* The assumption $|u_0| > \rho_0$ is not restrictive since the amplitude of the input sequence is typically much greater than the noise in a real experimental setting.

We can extend Theorem 5 to the MISO case, leading to the following result, which we can prove through the exact reasoning reported in the proof of Theorem 5 for all the inputs.

***Result** 3:* For MISO systems, the following result holds

$$\begin{aligned} &||\sum_{j=1}^{p}\big(\sum_{i=1}^{N}\alpha_i K_i^j - \sum_{i=1}^{d_0}\beta_i K_i^j\big)||^2 \\ &\le \sum_{j=1}^{p}\Big(\frac{||\tilde{y} - \hat{y}||^2 + \rho_q^2}{(\tilde{u}_{j,0} - \text{sign}(\tilde{u}_{j,0})\rho_j)^2} + \sum_{i=d_0}^{d_1}|\gamma_i^j|^2\Big). \end{aligned} \tag{36}$$

Equation (36) can be obtained by applying the statement of Theorem 5 and by noting that $||v_1 + \dots + v_n|| \le ||v_1|| + \dots + ||v_n||$.

The relevance of inequalities (22) and (36) lies in the fact that by computing the optimal solution for a given number d_1 of kernel basis functions, we can directly compute the bounds on the approximation error corresponding to any choice of $d_0 = 1, 2, \dots, d_1 - 1$. Therefore, we can use inequalities (22) and (36) to select the number of basis functions d_0, which provides the best trade-off between model accuracy and complexity.

The idea underlying the approach presented in this section to reduce the number of kernel basis functions is similar, to some extent, to the *subset of regressors* approach proposed by Smola and Bartlett [27] in the context of Gaussian process regression [28]. More specifically, the authors of [27] propose a greedy randomized algorithm to select the minimum number of basis functions that provide an acceptable level of approximation of the maximum posterior probability (MAP). In this work, we use Theorem 5 to assess the minimum number of basis functions needed to ensure a prescribed bound on the approximation error of the robust optimal solution to problem (12). Other methods are available in the literature to reduce the computational load of kernel-based methods, such as [29], [30], [31] and [32] which, although presenting exciting results, do not consider RKHS in the EIV settings. Thus, they cannot be directly applied in the framework considered in this paper.

VI. Kernel Choice

The choice of the kernel is a crucial task, as acknowledged by many authors (see, e.g. [24], [33], [34] and [35]). In the machine learning field, most often one uses the class of radial basis kernels (RBF), i.e. $K(x_1, x_2) = g(||x_1 - x_2||)$, and the most popular Gaussian kernels, i.e. $K(x_1, x_2) = e^{-\rho||x_1 - x_2||^2}$, with $\rho > 0$.

In the context of system identification, the choice of the kernel depends on the available a-priori information on the system. In this work, we limit our attention to the case of stable and causal dynamical systems. Extension of the results to the case of unstable systems will be the subject of further investigations. This section discusses the properties of the specific kernels used in the examples reported in Sections VII, VIII and IX. A complete discussion on the vast and complex subject of kernel selection is outside the scope of the work.

We start by recalling two theorems (see [24] for proofs and additional details) useful to build kernels that provide spaces of causal and stable functions.

***Theorem** 6:* (see [24]) The RKHS $\mathcal{H}$ contains only causal impulse responses if and only if the reproducing kernel satisfies

$$K(t_1, t_2) = H(t_1)H(t_2)\tilde{K}(t_1, t_2), \tag{37}$$

where $H(t)$ is the Heaviside step function (i.e. $H(t) = 0$ if $t < 0$ and $H(t) = 1$ if $t \geq 0$.) and $\tilde{K}$ is a kernel defined for non negative values of the time variable.

This theorem states that the kernel must be null for negative values of the time variable to have a causal impulse response function.

Let us now consider the problem of enforcing stability. An LTI system is BIBO stable if and only if its impulse response is absolutely summable, i.e., $\sum_{t=-\infty}^{+\infty} |g_t| < +\infty$. It can be proved (see, e.g., [36]) that the following theorem holds.

***Theorem** 7:* A non negative causal kernel, i.e. $K(t_1, t_2)$ such that $K(t_1, t_2) \geq 0, \ \forall\, t_1,\ t_2 \in \mathcal{T}$, is stable if and only if

$$\sum_{t_1=0}^{\infty}\sum_{t_2=0}^{\infty} K(t_1, t_2) < \infty \tag{38}$$

***Remark** 8:* The Gaussian kernel $K(t_1, t_2) = e^{-\rho(t_1-t_2)^2}$, where $\rho > 0$ is a given constant, is widely used in approximation with kernel functions. However, in the continuous-time case, the integral $\int_{\mathbb{R}^+}\int_{\mathbb{R}^+} e^{-\rho(t_1-t_2)^2} dt_1 dt_2$ diverges, thus violating Theorem (7), which prevents its use in system identification in the presence of stability constraints.

In this work, we consider the following kernel

$$\tilde{K}(t_1, t_2) = H(t_1)H(t_2)\frac{e^{-\omega_0(t_1-t_2)^2}}{\omega_1(1 + t_1 + t_2)^4}, \tag{39}$$

where $\omega_0 > 0$, $\omega_1 > 0$ are user-defined parameters.

The following theorem proves that the proposed kernel is stable and causal.

***Theorem** 8:* The Kernel $\tilde{K}$ defined in (39) is stable and causal.

Proof: Causality is obtained by definition, by observing that (39) has the same form as (37) with $\tilde{K}(t_1, t_2) > 0$. In order to prove stability we have to verify condition (38). We see that

$$\sum_{t_1,t_2=0}^{\infty} \frac{e^{-\omega_0(t_1-t_2)^2}}{\omega_1(1 + t_1 + t_2)^4} \leq \sum_{t_1,t_2=0}^{\infty} \frac{1}{\omega_1(1 + t_1 + t_2)^4}, \tag{40}$$

and, thanks to the integral test for convergence (also known as Maclaurin – Cauchy test), we notice that

$$\sum_{t_1,t_2=0}^{\infty} \frac{1}{\omega_1(1 + t_1 + t_2)^4}$$

converges if and only if

$$\int_0^{\infty}\int_0^{\infty} \frac{dt_1 dt_2}{(1 + t_1 + t_2)^4}$$

converges. On the other hand, we observe that

$$\int_0^{\infty}\int_0^{\infty} \frac{dt_1 dt_2}{\omega_1(1 + t_1 + t_2)^4} = \frac{1}{6\omega_1} < +\infty.$$

■

The proposed kernel (39) is suitable for system identification because it is causal and stable, differently from the standard Gaussian kernel, from which it has been derived.

***Remark** 9:* A wide variety of kernels that impose causality and stability are available in the literature. Among the others, we mention the stable spline (SS) [8], the Tuned Correlated (TC), Diagonal Correlated (DC), and Orthonormal Basis Functions (OBFs) kernels [37]. In this work, we use the kernel presented in (39) because it performs better in the specific examples considered in Sections VII, VIII, and IX. However, we point out that the analysis of the effects of the kernel choice on the obtained performance is outside the scope of the presented contribution. Furthermore, we can apply the proposed original methodology for nonparametric identification of EIV-MIMO systems to any stable and casual kernel in the literature.

VII. Simulation examples

In this section, we show the effectiveness of the proposed approach through two simulation examples, reported in Subsection VII-A and Subsection VII-B below. In the first example, we consider the non-parametric identification of two SISO systems, while in the second, we examine four MISO systems. In all the considered examples, given the measured input $\tilde{u}_t$ and the measured output $\tilde{y}_t$ of the system, which are affected by bounded errors, we look for an estimate of the impulse response g_t using the following Lasso regularized version of Problem (16):

$$\begin{aligned}\hat{\alpha} = &\operatorname*{arg\,min}_{\alpha\in\mathbb{R}^N, \lambda\in\mathbb{R}} \ \lambda \\ &\text{s.t.} \left|\left|\tilde{y} - \eta - (E_0 - \epsilon_0 A_0 - \cdots - \epsilon_N A_N)\alpha\right|\right|_2^2 + ||\alpha||_1 \leq \lambda \\ &\forall\ \epsilon\ \in\ \mathcal{B}_{\rho_1},\ \eta\ \in\ \mathcal{B}_{\rho_2}\end{aligned} \tag{41}$$

All the results presented in the paper for the problem without regularization also apply to the Lasso regularized version, as already discussed in Section IV. According to the literature, we opt for the Lasso regularization because it promotes sparsity in the solution.

We compare the estimate obtained by solving the robust Lasso regularized identification (RLR) Problem (41) with the estimate obtained by solving the following standard Lasso regularized (SLR) problem

$$\text{minimize}_{\alpha}\left|\left|\tilde{y} - E_0\alpha\right|\right|_2^2 + \zeta|\alpha|_1, \tag{42}$$

considered in [38]. While the proposed approach boils down to solving a semidefinite programming problem computed using the SeDuMi software [39], we can solve Problem (42) through

the standard MATLAB command `lasso`.
Although the theoretical results presented in the paper do not depend on the chosen kernel, we know that the choice of the kernel class may affect the properties of the estimated model. We select the kernel defined in (39) in the simulation examples. This choice is motivated by the fact that through this kernel, we obtain a casual and stable estimated model (see Section VI).

We evaluate the parameters ω_0, ω_1, and ζ through a cross-validation procedure and grid search. In particular, we choose ω_0 and ω_1 on a logarithmic scale in the interval $[10^{-2}, 10^3]$ and $[10^{-2}, 10^3]$, respectively. We select the regularization parameter ζ minimizing a Generalized Cross Validation (GCV) score [40] over a logarithmic grid in $[10^{-2}, 1]$. In particular, $\lambda = 10^{-2}, \omega_0 = 10^{-2}$ and $\omega_1 = 1$ yield the best results, while other parameter combinations achieve comparable results. In kernel-based methods, there are automated ways to tune the hyperparameters via the empirical Bayes approach (see [41]), i.e., maximizing the marginal likelihood. However, including such techniques in the framework considered here would require a rather complex modification of the settings, which could be considered the subject of future works. We refer the reader to [24] for a thorough discussion of kernel selection in the context of RKHS-based system identification.

The simulation input u_t is a white noise signal. The errors ϵ_t and η_t are independent random variables with uniform distribution in $[-\rho_u, \rho_u]$ and $[-\rho_y, \rho_y]$, respectively, where ρ_u and ρ_y take values to obtain the desired signal-to-noise ratio on the input and the output, defined as

$$SNR_u = 10 \log_{10} \frac{||u||_2^2}{\rho_u^2}, \qquad SNR_y = 10 \log_{10} \frac{||y||_2^2}{\rho_y^2}.$$

For each system considered in the examples reported in Subsection VII-A and Subsection VII-B, we compare the proposed approach and the standard one in sixteen simulation scenarios obtained considering four values of the SNR_u (10, 20, 40, 100 dB) and four values of the SNR_y (10, 20, 40, 100 dB). We consider the same value of ρ_u for all the inputs of the MISO system. To obtain a fair comparison of the proposed robust approach with the standard Lasso, we perform ten runs for each one of the sixteen experiments; for each run, we compute the quadratic impulse response estimation error given by $||g - \hat{g}||_2 = \sqrt{\sum_{t=0}^{N} (g_t - \hat{g}_t)^2}$. Then, we compute, for all the experiments, the average and the maximum errors over the ten runs.

A. Non-parametric identification of simulated SISO systems

In this subsection, we apply the proposed approach to the identification of two SISO systems, characterized, respectively, by the following two impulse response functions

$$g_1(t) = e^{-5t} + 2e^{-t/7}, \tag{43}$$

$$g_2(t) = (3t^2 + t)e^{-t}, \tag{44}$$

We simulate the SISO systems for $t \in [0, 10]$s, with a sampling time of 0.1s, obtaining input-output data sequences with length N = 100.

Tables I and II show the results for systems with impulse response (43) and (44) respectively. As expected, the quadratic estimation error increases with increasing noise. However, the proposed robust approach outperforms the standard one regarding both average and maximum estimation error for all the experiments. In particular, we point out that the maximum estimation error obtained with the proposed robust approach is up to 38 times lower than the one computed through the standard LASSO regularised estimator.

B. Non-parametric identification of simulated MISO systems

In this subsection, we apply the proposed approach to the identification of a MISO system with $p = 4$ inputs. The system is described by the following four impulse responses to be identified:

$$g_3(t) = (e^{-t}, t^2 e^{-t}), \tag{45}$$

$$g_4(t) = (\sin(t) e^{-t}, t^2 e^{-t}), \tag{46}$$

$$g_5(t) = (e^{-t} + e^{-3t}, (\sin(t) + \cos(t)) e^{-t}), \tag{47}$$

$$g_6(t) = (e^{-(t-3)^2},\ e^{-t}). \tag{48}$$

For MISO systems with p inputs, the solution of the convex semidefinite optimization Problem (17) requires the estimation of Np unknown by using N input-output samples. Although optimization Problem (17) makes perfect sense from a mathematical point of view, the obtained system model might result in a somewhat inaccurate description of the actual MISO system. Similar comments apply to the standard SLR estimator in (42). A possible approach to improve the estimation accuracy is collecting input-output data performing a number W of different experiments, each of length N. We can evaluate the sensitivity of the estimate accuracy to the number W of performed experiments by computing the ℓ_2 error $\left|\left|\tilde{y} - E_0 \alpha\right|\right|_2^2$ obtained for different values of W from 1 to p. We show the results obtained from such an analysis in Fig. 2, where we report the ℓ_2 errors obtained estimating the considered MISO systems for different values of the ratio W/p. As can be seen, the ℓ_2 error decreases as W increases, reaching a steady value in the limit for W which goes to p. The results reported in the rest of the paper refer to the case where we perform $W = p$ experiments.

We excite the systems under study with random uniformly distributed input signals of length $N = 100$ and collect the input-output noisy data sequences from time $t = 0$s to $t = 10$s with a sampling interval of 0.1 seconds.

***Remark** 10:* We must perform the W experiments starting from a system at rest (zero initial condition). In this way, the α variables entering the equations describing the different experiments are the same.

***Remark** 11:* If we must perform $W > 1$ different experiments, we have to replace Problem (17) with a similar optimization problem where we use $\sum_{w=1}^{W} \lambda_w$ as the objective function, and we extend the semidefinite constraints appearing in (17) in such a way as to consider all the W

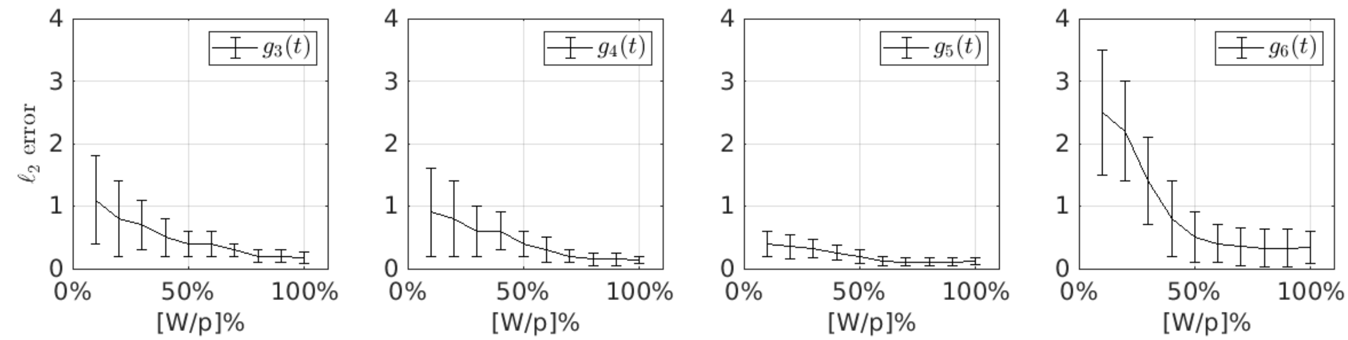

Fig. 2. Average ℓ_2 error for different W/p ratios, computed performing 100 random experiments for each value of W/p. The vertical lines represent the standard deviations.

different experiments.

The results reported in Tables III, IV, V and VI show that the proposed robust non-parametric approach significantly outperforms the standard non-parametric one with Lasso regularization when the measurement of the input and the output signals are affected by noise. The average decrease of the ℓ_2 mean error over all the instances is 73%, while the average decrease for the maximum error is 76%. From a computational point of view, the proposed approach is more expensive since it requires the solution of an SDP problem, while the standard lasso is a quadratic problem. Indeed, while the computation of the standard lasso regression takes less than one second for each run, the proposed approach needs more than ten seconds to compute the estimation of the impulse response function.

VIII. Non-parametric Identification of an Electronic Filter

In this section, we test the performance of the proposed approach on the experimental input-output data collected on a test bench SISO electronic filter. The considered impulse response is a transfer function of a third-order low-pass filter, with a couple of complex conjugated poles with a natural frequency of $120Hz$ and a damping factor of 0.5, and a real pole with a cutoff frequency of $160Hz$. A Sallen-Key circuit implements the complex conjugated poles, and an RC circuit gives the real pole. A signal generator excites the filter with a square wave input, and a National-Instruments PXI equipped with an NI–6221 DAQ board collects the measurements of the input and the output signals at a sample rate of $4kHz$. From the accuracy of the measurement equipment, we derive the upper bound on the measurement errors, $\Delta\xi = \Delta\eta = 0.003$V. We solve the underlying optimization problem using the software SeDuMi [42]. Fig. 3 shows the collected input-output data sequences of length $N = 500$. We use the first 187 samples (Identification Set) to perform the estimation and the remaining 313 samples (Validation Set) for validation. The obtained results, reported in Fig. 4, show that the estimated system accurately reproduces the behaviour of the real-world system.

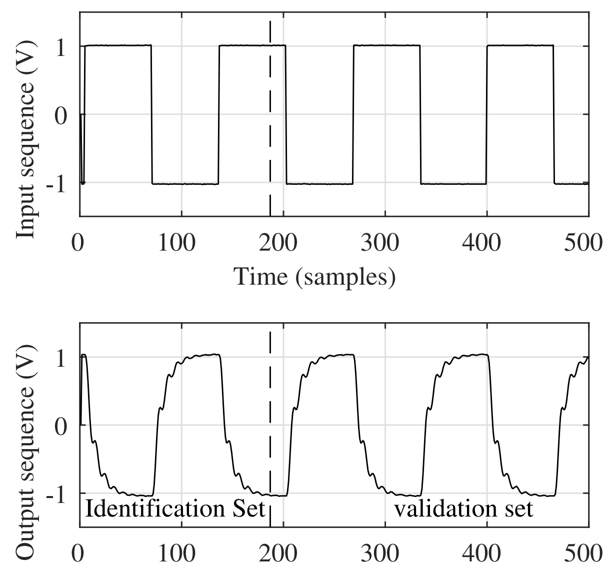

Fig. 3. Measured input and output data. The first 187 samples are used for identification and the last 313 are used for model validation.

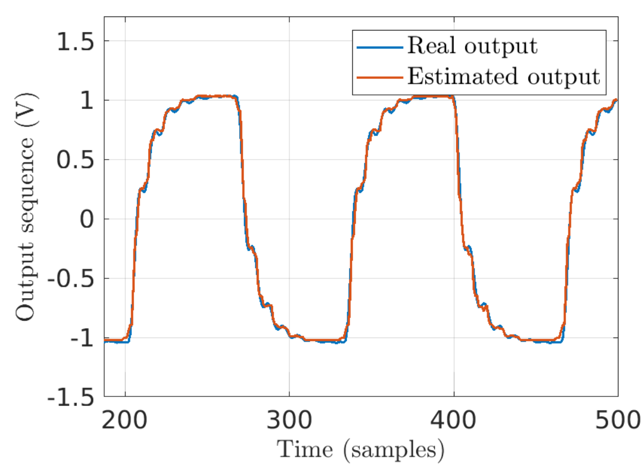

Fig. 4. Comparison between the measured output (blue) and the forecast given by the estimated impulse transfer function (red) on the validation data.

IX. Non-parametric Identification of a MIMO System

In this section, we apply the proposed approach to the input-output data collected on a 2-input, 2-output MIMO electronic filter, which we consider as an Electronic Process Simulator (EPS). This EPS is a purposely self-built electronic process simulator to test the main features of the MIMO identification procedure proposed in this paper. The experimental setting is the same as considered in [43].

The EPS is a real plant connected to the data acquisition equipment for collecting the measurements.

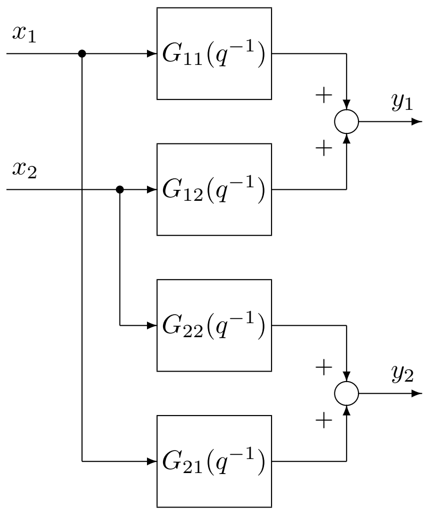

Fig. 5. Block-diagram description of the MIMO process simulator considered in the experimental test bench section.

The block diagram of the system is depicted in Fig. 5 and a detailed description of transfer functions G_{11}, G_{12}, G_{21} and G_{22} can be found in [43]. The chosen system inputs are two uncorrelated random input sequences of $N = 200$ samples, uniformly distributed in the interval $[-1, +1]$. A National–Instruments PXI, equipped with a NI–6221 DAQ board, excites the system and collects the input-output samples through custom software developed in LabVIEW. We collect the data with a sampling frequency (f_s) of $4kHz$. We refer the reader to [43] for a detailed discussion on the selection of the sampling frequency. SeDuMi is the software employed to solve the robust optimization problem discussed in Section IV. From the accuracy of the measurement equipment, we derive the upper bounds on the measurement errors, $\Delta\xi = \Delta\eta = 0.003$ V. The validation of the estimated model is carried out using the input-output data collected by exciting the EPS with the square-wave inputs depicted in the upper plot of Fig. 7. The results reported in Fig. 7 show that the output sequences obtained by simulating the model accurately reproduce the behavior of the real electronic circuit.

To show the effectiveness of the proposed approach in reducing the complexity of the identified model, we provide a comparison between the approximation error L and the proposed error bound in (36). In particular, we compute the left-hand side of (36) by taking the norm of the difference between the full-order solution computed with N basis functions and the one obtained with $d_0 < N$ basis functions for different values of d_0. Then, we evaluate the right-hand side for $d_1 = d_0 + 1$ to obtain the error bound. Fig. 6 shows the results for $d_0 = 1, \ldots, 90$, normalized with respect to $|| \sum_{j=1}^{p} \left(\sum_{i=1}^{N} \alpha_i K_i^j - \beta_i K_1^j \right) ||^2$. Those results show that the

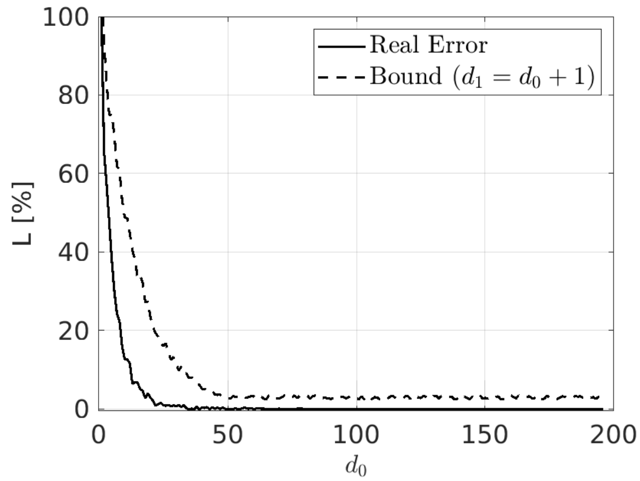

Fig. 6. Comparison between the real error (solid) and its bound computed by (36) for $d_1 = d_0 + 1$(dashed).

bound converges to a value about 3% greater than the true error L when we select a number of basis functions $d_0 = 50$.

TABLE I
COMPARISON BETWEEN THE STANDARD AND THE ROBUST NON-PARAMETRIC LASSO REGULARIZED ESTIMATORS (SLR AND RLR RESPECTIVELY) IN TERMS OF QUADRATIC ESTIMATION ERROR FOR IMPULSE RESPONSE FUNCTION g_1

		SLR ℓ_2 error		**RLR ℓ_2 error**	
SNR_u	SNR_y	**avg (std dev)**	**max**	**avg (std dev)**	**max**
10	10	17.5407(2.8208)	22.9767	0.9030(0.2566)	1.5639
10	20	0.7198(0.4013)	1.5283	0.3862(0.0611)	0.4755
10	40	0.4649(0.3491)	1.0426	0.1509(0.0159)	0.1776
10	100	0.4088(0.3011)	0.9706	0.1450(0.0190)	0.1630
20	10	18.0941(3.8884)	28.1115	0.8670(0.1483)	1.183
20	20	0.7331(0.3254)	1.2862	0.2992(0.0385)	0.369
20	40	0.5641(0.3059)	1.0273	0.1139(0.0076)	0.1251
20	100	0.5057(0.3458)	1.0304	0.1085(0.0071)	0.1174
40	10	18.3684(3.7057)	25.2336	0.8373(0.2361)	1.3171
40	20	0.6744(0.3543)	1.3752	0.3508(0.0607)	0.4552
40	40	0.4204(0.2921)	0.9123	0.0924(0.0067)	0.1042
40	100	0.3910(0.2815)	0.8944	0.0931(0.0082)	0.1066
100	10	17.4595(4.5478)	25.1049	0.8175(0.1998)	1.211
100	20	0.6644(0.2798)	1.0577	0.2893(0.0495)	0.359531
100	40	0.3752(0.2590)	0.9040	0.0698(0.0060)	0.0785
100	100	0.5554(0.3471)	1.0070	0.0898(0.0058)	0.1039

X. CONCLUSIONS

In this paper, we address the problem of robust non-parametric identification of linear systems when the collected input and output data are corrupted by measurement errors, only known to belong to bounded sets. By constraining the search for the system impulse response to a reproducing kernel Hilbert space, we show that we can exactly compute the solution to the estimation problem by solving a semidefinite programming problem. The proposed algorithm explicitly considers the noise affecting the input and the output sequence,

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2024.3410835

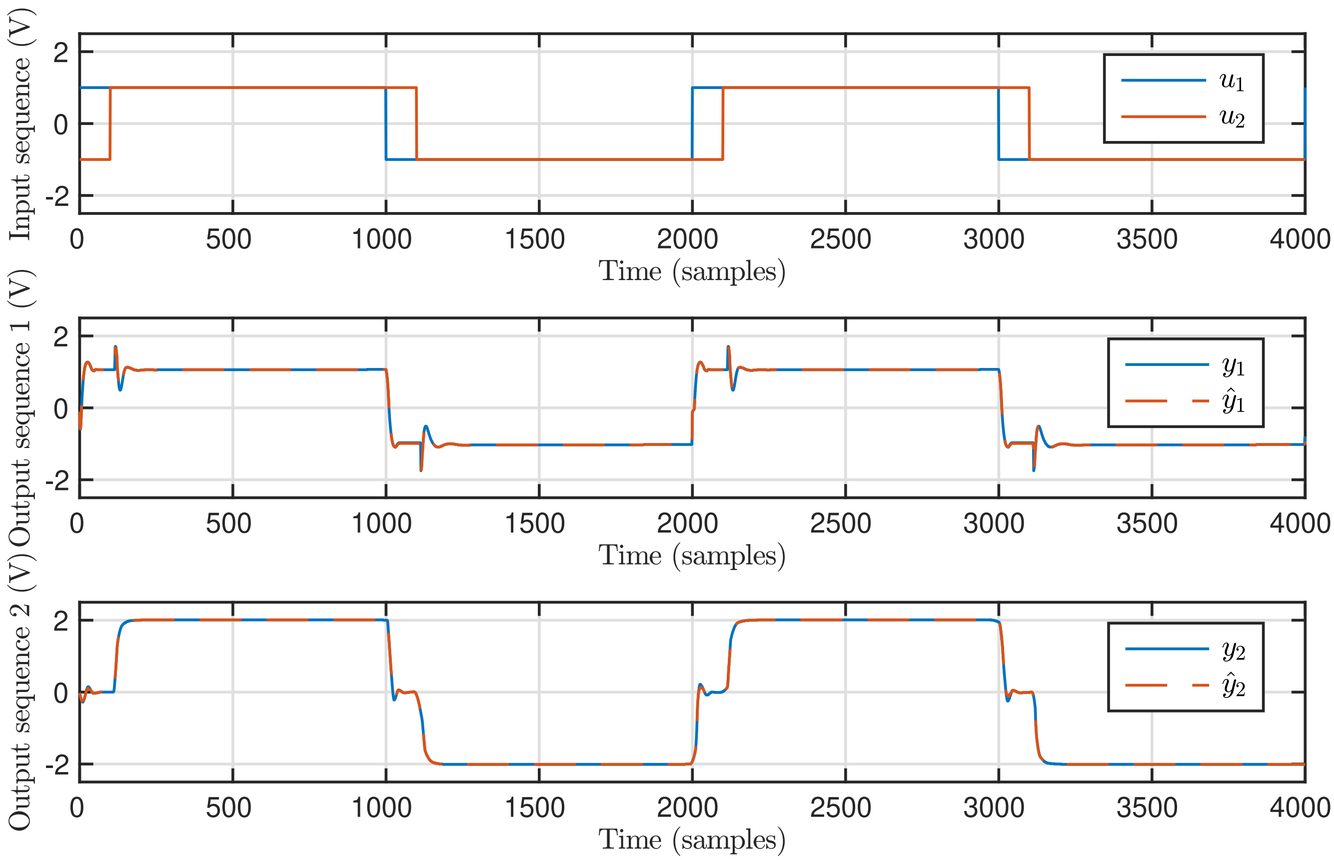

Fig. 7. Comparison between the circuit outputs (blue) and the outputs of the estimated MISO model (red) on the validation dataset. From top to bottom: applied input signals, comparison of outputs y_1, comparison of outputs y_2.

TABLE II
COMPARISON BETWEEN THE STANDARD AND THE ROBUST NON-PARAMETRIC LASSO REGULARIZED ESTIMATORS (SLR AND RLR RESPECTIVELY) IN TERMS OF QUADRATIC ESTIMATION ERROR FOR IMPULSE RESPONSE FUNCTION g_2

		SLR ℓ_2 error		**RLR ℓ_2 error**	
SNR_u	SNR_y	**avg (std dev)**	**max**	**avg (std dev)**	**max**
10	10	16.9993(2.4809)	21.2534	0.6110(0.1483)	0.8916
10	20	0.2450(0.0492)	0.3393	0.2288(0.0295)	0.2852
10	40	0.0632(0.0110)	0.0915	0.0300(0.0047)	0.0376
10	100	0.0591(0.0103)	0.0741	0.0148(0.0049)	0.0254
20	10	14.1659(3.2489)	19.7764	0.5546(0.1558)	0.7283
20	20	0.1973(0.0432)	0.2645	0.1855(0.0399)	0.2479
20	40	0.0593(0.0115)	0.0792	0.0227(0.0055)	0.0311
20	100	0.0560(0.0094)	0.0753	0.0041(0.0011)	0.0067
40	10	14.2956(4.0943)	24.3673	0.5851(0.2028)	1.054
40	20	0.2234(0.0545)	0.2888	0.2130(0.0416)	0.2577
40	40	0.0587(0.0072)	0.0732	0.0208(0.0013)	0.0232
40	100	0.0503(0.0058)	0.0580	0.002323(0.0004)	0.003
100	10	14.8368(6.6619)	33.1390	0.5848(0.1683)	0.8576
100	20	0.2144(0.0555)	0.3132	0.2049(0.0451)	0.2833
100	40	0.0586(0.0057)	0.0646	0.0227(0.0038)	0.0269
100	100	0.0550(0.0070)	0.0661	0.0019(0.0003)	0.0025

TABLE III
COMPARISON BETWEEN THE STANDARD AND THE ROBUST NON-PARAMETRIC LASSO REGULARIZED ESTIMATORS (SLR AND RLR RESPECTIVELY) IN TERMS OF QUADRATIC ESTIMATION ERROR FOR IMPULSE RESPONSE FUNCTION g_3

		SLR ℓ_2 error		**RLR ℓ_2 error**	
SNR_u	SNR_y	**avg (std dev)**	**max**	**avg (std dev)**	**max**
10	10	2.1299(1.3842)	5.3017	0.5222(0.1305)	0.6513
10	20	1.0398(0.5715)	1.7945	0.2416(0.1867)	0.5549
10	40	0.9568(0.5424)	1.8578	0.1349(0.1159)	0.3607
10	100	0.9408(0.5497)	1.8342	0.1235(0.1583)	0.5496
20	10	1.8107(1.3001)	5.1276	0.4567(0.1634)	0.7400
20	20	0.7334(0.5109)	1.7711	0.1166(0.0869)	0.2968
20	40	0.5496(0.3471)	1.3767	0.0258(0.0300)	0.0857
20	100	0.7446(0.4463)	1.4888	0.0598(0.0659)	0.2326
40	10	1.9025(1.4322)	4.7451	0.4690(0.1379)	0.6689
40	20	0.6423(0.3145)	1.3698	0.0822(0.0757)	0.2894
40	40	0.6975(0.4142)	1.3059	0.0189(0.0362)	0.1151
40	100	0.9104(0.7221)	2.1021	0.0064(0.0048)	0.0170
100	10	1.9078(2.009614)	7.1784	0.5484(0.1930)	0.8425
100	20	1.1066(0.3894)	1.6884	0.0605(0.0291)	0.1307
100	40	1.1192(1.1322)	3.1932	0.0012(0.0011)	0.0041
100	100	0.8993(0.7157)	2.3300	0.0007(0.0007)	0.0024

differently from the nonparametric approaches previously proposed in the literature, including the one based on the standard Lasso regularization. Although the min-max formulations might potentially suffer from conservativeness, the numerical results obtained in the performed simulation examples show that the proposed approach significantly outperforms a stan-

TABLE IV
COMPARISON BETWEEN THE STANDARD AND THE ROBUST NON-PARAMETRIC LASSO REGULARIZED ESTIMATORS (SLR AND RLR RESPECTIVELY) IN TERMS OF QUADRATIC ESTIMATION ERROR FOR IMPULSE RESPONSE FUNCTION g_4

		SLR ℓ_2 error		**RLR ℓ_2 error**	
SNR_u	SNR_y	**avg (std dev)**	**max**	**avg (std dev)**	**max**
10	10	1.1843(0.8255)	3.0483	0.2548(0.0562)	0.3428
10	20	0.5751(0.5389)	2.0438	0.1159(0.0697)	0.2338
10	40	0.4381(0.3210)	1.0212	0.0566(0.0450)	0.1362
10	100	0.6957(0.4190)	1.3499	0.1473(0.1245)	0.4537
20	10	1.1944(0.6119)	2.3937	0.2991(0.0868)	0.4520
20	20	1.1737(0.6781)	2.1399	0.1131(0.1081)	0.3279
20	40	0.9289(0.8769)	2.7043	0.0643(0.0577)	0.1666
20	100	0.6652(0.4070)	1.3483	0.0843(0.0717)	0.2181
40	10	1.8531(1.5620)	5.7540	0.4154(0.1412)	0.7437
40	20	0.7200(0.5357)	2.0236	0.0812(0.0381)	0.1641
40	40	0.4422(0.2205)	0.9441	0.010741(0.004273)	0.020825
40	100	0.5944(0.3912)	1.3869	0.0142(0.0099)	0.0311
100	10	2.0088(2.0301)	7.4584	0.4509(0.1389)	0.7502
100	20	0.9289(0.7281)	2.4275	0.0769(0.0530)	0.2100
100	40	0.7370(0.5354)	1.7778	0.0082(0.0023)	0.0133
100	100	0.5978(0.4654)	1.6898	0.0078(0.0016)	0.0107

dard Lasso regularised non-parametric estimation approach. In the case of MISO systems with p inputs, the problem requires the estimation of Np unknown using N input-output samples. In this case, the system model obtained with the proposed approach might result in a somewhat inaccurate description of the actual MISO system. However, we can significantly improve the estimation accuracy by performing a number $W \in [2, p]$ of different experiments, each of length N. The proposed approach also proves effective on experimental data collected from electronic process simulators. Future development of this method may consider other optimisation settings like stochastic programming and analyse the problem structures to develop more efficient exact and heuristic solution methods.

XI. ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Professor A.L. Fradkov for the fruitful discussions on some technical aspects of the paper. The authors also thank the anonymous reviewers for their significant help in improving the overall quality of the paper.

REFERENCES

[1] T. Söderström and P. Stoica, *System Identification*. Upper Saddle River: Prentince Hall, 1989.
[2] L. Ljung, *System Identification, Theory for the User*. Upper Saddle River: Prentince Hall, 1999.
[3] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, pp. 657–682, 2014.
[4] S. J. Qin, "An overview of subspace identification," *Computers & Chemical Engineering*, vol. 30, no. 10, pp. 1502 – 1513, 2006.
[5] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, December 1974.
[6] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
[7] D. Bauer, "Order estimation for subspace methods." *Automatica*, vol. 37, no. 10, pp. 1561–1573, 2001.
[8] G. Pillonetto and G. De Nicolao, "A new kernel-based approach for linear system identification," *Automatica*, vol. 46, pp. 81–93, 2010.
[9] G. Pillonetto, M. Quang, and A. Chiuso, "A new kernel-based approach for nonlinear system identification," *IEEE Transactions on Automatic Control*, vol. 56, no. 12, pp. 2825–2840, 2011.
[10] T. Chen, M. Andersen, L. Ljung, A. Chiuso, and G. Pillonetto, "System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 2933–2945, 2014.
[11] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, *Regularized System Identification: Learning Dynamic Models from Data*. Springer International Publishing, 2022.
[12] M. A. Álvarez, L. Rosasco, and N. D. Lawrence, "Kernels for vector-valued functions: A review," *Foundations and Trends® in Machine Learning*, vol. 4, no. 3, pp. 195–266, 2012. [Online]. Available: http://dx.doi.org/10.1561/2200000036
[13] T. Söderström, *Errors-in-Variables Methods in System Identification*. Springer Cham, 2018.
[14] R. S. Risuleo, G. Bottegal, and H. Hjalmarsson, "Kernel-based system identification from noisy and incomplete input-output data," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, Dec 2016, pp. 2061–2066.
[15] F. Schweppe, "Recursive state estimation: unknown but bounded error and system input," *IEEE Transactions on Automatic Control*, vol. AC-13, pp. 556–558, 1968.
[16] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: an overview," *Automatica*, vol. 27(6), pp. 997–1009, 1991.
[17] M. Milanese, J. P. Norton, H. Piet-Lahanier, and E. Walter, Eds., *Bounding approaches to system identification*. New York: Plenum Press, 1996.
[18] V. Cerone, D. Piga, and D. Regruto, "Improved parameters bounds for set-membership EIV problems," *International Journal of Adaptive Control and Signal Processing*, vol. 57, no. 2, pp. 208–227, 2011.
[19] ——, "Set-membership error-in-variables identification through convex relaxation techniques," *IEEE Transactions on Automatic Control*, vol. 57, no. 2, pp. 517–522, 2012.
[20] V. Cerone, E. Fadda, and D. Regruto, "A robust optimization approach to kernel-based nonparametric error-in-variables identification in the presence of bounded noise," in *Proc. of American Control Conference ACC*, 2017.
[21] K. Bernd, "Book review of "Reproducing Kernel Hilbert Spaces and Applications, operator theory: Advances and applications", Daniel Alpay (Ed.), Birkhäuser Verlag, Basel-Boston-Berlin, 2003, Volume 143, 344 pages. ISBN 3-7643-0068-X." *Optimization*, vol. 53, no. 4, pp. 429–430, 2004.
[22] N. Aronszajn, "Theory of reproducing kernels," *Trans. Amer. Math. Soc.*, vol. 68, pp. 337–404, 1950.
[23] B. Schölkopf, R. Herbrich, and A. Smola, "A generalized representer theorem," *Computational Learning Theory. Lecture Notes in Computer Science.*, vol. 2111, pp. 416–426, 2001.
[24] F. Dinuzzo, "Kernels for linear time invariant system indentification," *SIAM J. control opt.*, vol. 53, no. 5, pp. 3299–3317, 2015.
[25] L. Vandenberghe, V. R. Balakrishnan, R. Wallin, A. Hansson, and T. Roh, *Interior-Point Algorithms for Semidefinite Programming Problems Derived from the KYP Lemma*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 195–238. [Online]. Available: https://doi.org/10.1007/10997703_12
[26] R. A. Horn and C. R. Johnson, "Positive definite and semidefinite matrices," in *Matrix Analysis*. Cambridge: Cambridge University Press, Oct. 2012, pp. 425–516.
[27] A. Smola and P. Bartlett, "Sparse greedy gaussian process regression," *Advances in Neural Information Processing Systems*, vol. 13, pp. 619–625, 2 2001.
[28] J. Quiñonero Candela, C. Rasmussen, and C. Williams, *Approximation Methods for Gaussian Process Regression*, ser. Neural Information Processing. Cambridge, MA, USA: MIT Press, Sep. 2007, pp. 203–223.
[29] F. P. Carli, A. Chiuso, and G. Pillonetto, "Efficient algorithms for large scale linear system identification using stable spline estimators," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 119 – 124, 2012, 16th IFAC Symposium on System Identification. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667015379386
[30] K. Zhang and J. T. Kwok, "Clustered nyström method for large scale manifold learning and dimension reduction," *IEEE Transactions on Neural Networks*, vol. 21, no. 10, pp. 1576–1587, Oct 2010.

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2024.3410835

[31] S. Fine and K. Scheinberg, "Efficient svm training using low-rank kernel representations," *J. Mach. Learn. Res.*, vol. 2, pp. 243–264, Mar. 2002.
[32] G. Pillonetto and B. M. Bell, "Bayes and empirical bayes semi-blind deconvolution using eigenfunctions of a prior covariance," *Automatica*, vol. 43, no. 10, pp. 1698 – 1712, 2007.
[33] T. Chen, H. Ohlsson, and L. Ljung, "On the estimation of transfer functions, regularizations and gaussian processes revisited," *Automatica*, vol. 48(8), pp. 1525–1535.
[34] T. Chen and L. Ljung, "On kernel structures for regularized system identification (i): a machine learning perspective," *IFAC-PapersOnLine*, vol. 48 (28), pp. 1035–1040.
[35] ——, "On kernel structures for regularized system identification (ii): a system theory perspective," *IFAC-PapersOnLine*, vol. 48 (28), pp. 1041–1046.
[36] C. Carmeli, E. D. Vito, and A. Toigo, "Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem." *Analysis and Applications*, vol. 4, pp. 377–408, 2006.
[37] 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*, vol. 87, pp. 318–329, 2018.
[38] R. Jahangir and S. T. Kumar, "Deconvolution and total least squares in finding the impulse response of an electromagnetic system from measured data." *IEEE Transactions on Antennas and Propagation*, vol. 43(4), pp. 416 – 421, May 1995.
[39] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB Toolbox for optimization over symmetric cones," *Optim. Methods Software*, vol. 11, no. 12, pp. 625–653, 1999.
[40] M. Hutchinson and F. Hoog, "Smoothing noisy data with spline functions," *Numerische Mathematik*, vol. 47, pp. 99–106, 03 1985.
[41] B. P. Carlin and T. A. Louis, "Bayes and empirical bayes methods for data analysis," *Statistics and Computing*, vol. 7, pp. 1573–1375, 1997.
[42] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optim. Methods Software*, vol. 11, no. 12, p. 625â635, 1999.
[43] V. Cerone, V. Razza, and D. Regruto, "Set-membership errors-in-variables identification of MIMO linear systems," *Automatica*, vol. 90, pp. 25 – 37, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109817306295
[44] A. L. Fradkov and V. A. Yakubovich, "The s-procedure and a duality realations in nonconvex problems of quadratic programming," *Vestnik Leningrad Univ. Math.*, vol. 1, no. 5, pp. 101–109, 1973.
[45] A. L. Fradkov, "Duality theorems for certain non convex extremal problems," *In Proceedings of the Vestnik Leningrad Univ. Math.*, 1979.

Vito Cerone (Member, IEEE) received the Laurea degree in electronic engineering and the Ph.D. degree in system engineering from Politecnico di Torino, Italy, in 1984 and 1989, respectively. From 1989 to 1990, he was a Research Fellow with the School of Electronic and Electrical Engineering, University of Birmingham, U.K. From 1989 to 1998, he was an Assistant Professor with the Department of Control and Computer Engineering, Politecnico di Torino, where was an Associate Professor, from 1998 to 2021, and has been a Full Professor, since 2021. He is the coauthor of Linear Quadratic Control: An Introduction. His current research interests include system identification, parameter estimation, optimization, and control with applications to automotive systems.

Edoardo Fadda received the Laurea degree in mathematical engineering and the Ph.D. degree in information technology and system engineering from the Politecnico di Torino, Torino, Italy, in 2014 and 2018, respectively. He is a Assistant Professor with the Dipartimento di Scienze Matematiche, Politecnico di Torino. Since 2018 he is member of the Isires's Scientific and Technical Committee (Istituto Italiano di Ricerca e Sviluppo). His main interests are optimization under uncertainty, statistics and their applications.

Diego Regruto (Member, IEEE) received the Laurea degree in electronic engineering and the Ph.D. degree in system engineering from Politecnico di Torino, Torino, Italy, in 2000 and 2004, respectively. He is currently an Associate Professor with Dipartimento di Automatica e Informatica, Politecnico di Torino. He chaired the IEEE CSS Technical Committee on System Identification and Adaptive Control (TC-SIAC), from January 2013 to December 2015. He has served as an Associate Editor for IEEE Transactions on Automatic Control, from January 2016 to December 2022, and the IEEE Conference Editorial Board, from June 2013 to December 2020. His current research interests include system identification, data-based robust control methods, and convex relaxation techniques for nonconvex optimization, with application to automotive, biomedical, and electrical engineering problems.

TABLE V
COMPARISON BETWEEN THE STANDARD AND THE ROBUST NON-PARAMETRIC LASSO REGULARIZED ESTIMATORS (SLR AND RLR RESPECTIVELY) IN TERMS OF QUADRATIC ESTIMATION ERROR FOR IMPULSE RESPONSE FUNCTION g_5

		SLR ℓ_2 error		**RLR ℓ_2 error**	
SNR_u	SNR_y	**avg (std dev)**	**max**	**avg (std dev)**	**max**
10	10	0.6321(0.3016)	1.1010	0.2956(0.0777)	0.4200
10	20	0.2704(0.2692)	0.9188	0.1235(0.1081)	0.3071
10	40	0.2497(0.1571)	0.5551	0.0368(0.0200)	0.0833
10	100	0.1625(0.2364)	0.8064	0.1013(0.0716)	0.2280
20	10	0.5096(0.2002)	0.7826	0.3490(0.1377)	0.6719
20	20	0.1771(0.1091)	0.4105	0.0737(0.0453)	0.1553
20	40	0.1243(0.1506)	0.4390	0.0315(0.0202)	0.0699
20	100	0.2144(0.1457)	0.5259	0.0334(0.0362)	0.1119
40	10	0.6361(0.4090)	1.3035	0.3410(0.0944)	0.4863
40	20	0.0918(0.0739)	0.2044	0.0461(0.0430)	0.1197
40	40	0.0777(0.0319)	0.1314	0.0048(0.0017)	0.0089
40	100	0.1583(0.1227)	0.4395	0.0247(0.0254)	0.0890
100	10	0.6159(0.1974)	0.8850	0.2757(0.0906)	0.4417
100	20	0.1795(0.1372)	0.4774	0.0974(0.0643)	0.2311
100	40	0.1496(0.1212)	0.3928	0.0028(0.0022)	0.0073
100	100	0.1328(0.1258)	0.4699	0.0230(0.0186)	0.0725

APPENDIX

Proof of Result 2

To prove Result 2, we first review the following Theorem related to the notion of $\mathcal{S}$-procedure (see, e.g., [44] and [45] for details).

***Theorem** 9 ($\mathcal{S}$-procedure):* Given $Q_0, \ldots, Q_n \in \mathbb{S}^{m\times m}$, $s_0, \ldots, s_n \in \mathbb{R}^m$, and $r_0, \ldots, r_n \in \mathbb{R}$ If there exist real scalar $\tau_1, \ldots, \tau_n \geq 0$ such that

$$\begin{bmatrix} Q_0 & s_0 \\ s_0^T & r_0 \end{bmatrix} - \tau_1 \begin{bmatrix} Q_1 & s_1 \\ s_1^T & r_1 \end{bmatrix} - \cdots - \tau_n \begin{bmatrix} Q_n & s_n \\ s_n^T & r_n \end{bmatrix} \geq 0 \quad (49)$$

then

$$f(x) = x^T Q_0 x + 2s_0^T x + r_0 \geq 0 \quad (50a)$$

$$\forall x \text{ s.t.} \begin{cases} h_1(x) & = x^T Q_1 x + 2s_1^T x + r_1 \geq 0 \\ \ldots \\ h_n(x) & = x^T Q_n x + 2s_n^T x + r_n \geq 0 \end{cases} \quad (50b)$$

TABLE VI
COMPARISON BETWEEN THE STANDARD AND THE ROBUST NON-PARAMETRIC LASSO REGULARIZED ESTIMATORS (SLR AND RLR RESPECTIVELY) IN TERMS OF QUADRATIC ESTIMATION ERROR FOR IMPULSE RESPONSE FUNCTION g_6

		SLR ℓ_2 error		**RLR ℓ_2 error**	
SNR_u	SNR_y	**avg (std dev)**	**max**	**avg (std dev)**	**max**
10	10	2.8707(1.5081)	6.4479	1.1214(0.5407)	2.0398
10	20	1.2147(1.1787)	3.9958	0.4423(0.2824)	0.8619
10	40	0.6978(0.6402)	1.7597	0.4204(0.6601)	2.2019
10	100	1.0643(0.8568)	3.2304	0.5359(0.2693)	0.8731
20	10	2.0219(1.9959)	5.8395	0.6921(0.6053)	2.2719
20	20	0.9722(1.4278)	4.8174	0.4291(0.4325)	1.4152
20	40	1.1558(1.0563)	3.9067	0.1377(0.1945)	0.6350
20	100	0.6617(0.7019)	1.7948	0.1264(0.1155)	0.3657
40	10	1.4694(0.6889)	2.4222	0.4721(0.1447)	0.6769
40	20	1.2579(0.8028)	2.7536	0.0602(0.0313)	0.1295
40	40	1.1697(1.0937)	3.4585	0.0361(0.0410)	0.1460
40	100	0.4501(0.4361)	1.5498	0.0553(0.0566)	0.1942
100	10	1.8107(1.7221)	5.9515	0.7228(0.5092)	1.9440
100	20	0.5893(0.6101)	1.7999	0.0664(0.0522)	0.1729
100	40	1.8584(1.6024)	4.9292	0.0088(0.0104)	0.0321
100	100	1.0196(1.1562)	4.0697	0.0805(0.1032)	0.3327

where x^T is the transpose of vector x.

***Remark** 12:* Note that the $\mathcal{S}$-procedure stated in its classical form (as in Theorem 9) is a sufficient condition, i.e., (49) implies (50), while the reverse implication is not true in general. Therefore, if there is no $\tau_1, \dots, \tau_n \geq 0$ such that (49) holds, we cannot conclude that (50) does not hold. In fact, condition (49) can be satisfied by $\tau_1, \dots, \tau_n$ such that at least one of them is negative. It follows that to enforce that the condition of Theorem 9 is also necessary, it is enough to prove that the set $B = \{(\tau_1, \dots, \tau_n) \in \mathbb{R}^n$ such that $\tau_1 < 0 \vee \dots \vee \tau_n < 0$ and such that condition (49) holds$\}$ is the empty set.

Proof: **(Result 2)**

By rewriting the left-hand side of the first constraint of Problem (16) we obtain

$$\left|\left|\tilde{y} - \eta - \sum_{j=0}^{p}(E_0^j - \epsilon_{j,0}A_0^j - \dots - \epsilon_{j,N}A_N^j)\alpha^j\right|\right|_2^2 =$$

$$\Big|\Big|\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p) - \eta + \epsilon_{1,0}A_0^1\alpha^1 + \dots + \epsilon_{1,N}A_N^1\alpha^1$$

$$+ \dots + \epsilon_{p,0}A_0^p\alpha^p + \dots + \epsilon_{p,N}A_N^p\alpha^p\Big|\Big|_2^2 =$$

$$= ||\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)||^2 + ||\eta||^2$$

$$+(\epsilon_{1,0})^2||A_0^1\alpha^1||^2 + \dots + (\epsilon_{p,N})^2||A_N^p\alpha^p||^2$$

$$-2(\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p))^T\eta$$

$$+2\epsilon_{1,0}(\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p))^T A_0^1\alpha^1 + \dots$$

$$+2\epsilon_{p,N}(\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p))^T A_N^p\alpha^p$$

$$-2\epsilon_{1,0}\eta^T A_0^1\alpha^1 - \dots - 2\epsilon_{p,N}\eta^T A_N^p\alpha^p$$

$$+2\epsilon_{1,0}\epsilon_{1,1}(\alpha^1)^T A_0^T A_1(\alpha^1) + \dots + 2\epsilon_{1,N-1}\epsilon_{1,N}(\alpha^1)^T A_{N-1}^T A_N\alpha^1 + \dots$$

$$+2\epsilon_{p,0}\epsilon_{p,1}(\alpha^p)^T A_0^T A_1(\alpha^p) + \dots + 2\epsilon_{p,N-1}\epsilon_{p,N}(\alpha^p)^T A_{N-1}^T A_N\alpha^p + \dots$$

$$+2\epsilon_{1,0}\epsilon_{p,0}(\alpha^1)^T A_0^T A_0(\alpha^p) + \dots + 2\epsilon_{p-1,N}\epsilon_{p,N}(\alpha^{p-1})^T A_{N-1}^T A_N\alpha^p$$

which can be written in matrix form as in (53), with $M(\alpha)$ defined as in (17). In order to apply Theorem 9, we have to write (53) in the form of (50a) and to write $\mathcal{B}_{\rho_q}$, $\mathcal{B}_{\rho_j}$ $j = 1 \dots p$ in the form of (50b). For both cases we set the vector x equal to $[\eta, \epsilon^1, \dots, \epsilon^p]$. Eq. (53) has the form

$$x^T\Xi_0 x - 2\kappa_0^T x + \chi_0 \leq \lambda,$$

for suitable Ξ_0, κ_0, and χ_0 and we can obtain the same form of (50a) by simple algebraic manipulation

$$x^T Q_0 x + 2s_0^T x + r_0 \geq 0.$$

where $Q_0 = -\Xi_0$, $s_0 = \kappa_0$ and $r_0 = \chi_0 - \lambda$. Instead, the $\mathcal{B}_{\rho_q}$, $\mathcal{B}_{\rho_j}$ $j = 1 \dots p$ can be represented by

$$x^T\Xi_q x \leq \rho_q, \quad x^T\Xi_j x \leq \rho_j \quad \forall j = 1, \dots, p \tag{51}$$

for suitable block diagonal matrices Ξ_q, Ξ_j $j = 1, \dots, p$ each one of them having one block equal to the identity matrix and all the others equal to zero. Moreover, Eqs. (51) can be written as $-x^T\Xi_q x + \rho_q \geq 0$, $-x^T\Xi_j x + \rho_j \geq 0\, \forall j = 1, \dots, p$, thus in the same form of (50b) by setting $Q_1 = -\Xi_q$, $Q_{j+1} = -\Xi_j$, $r_1 = \rho_q$, $r_{j+1} = \rho_j$, $s_1 = s_{j+1} = 0, \forall j = 1, \dots, p$. Therefore, we can apply Eq. (49) of Theorem 9 as:

$$\begin{bmatrix} -\Xi_0 & s_0 \\ s_0^T & \lambda - r_0 \end{bmatrix} - \tau_q \begin{bmatrix} -\Xi_q & 0 \\ 0 & \rho_q \end{bmatrix} - \tau_1 \begin{bmatrix} -\Xi_1 & 0 \\ 0 & \rho_1 \end{bmatrix} - \dots - \tau_p \begin{bmatrix} -\Xi_p & 0 \\ 0 & \rho_p \end{bmatrix} \geq 0. \tag{52}$$

Now, we can apply the $\mathcal{S}$-procedure which states that if there exists $\tau_q, \tau_1, \dots, \tau_p \geq 0$ such that (54) is positive semi-definite, then (53) holds for each $\eta \in \mathcal{B}_{\rho_q}$, $\epsilon_j \in \mathcal{B}_{\rho_j}$ $j = 1 \dots p$. However, for a block matrix to be positive semi-definite, all the diagonal blocks must also be positive semi-definite. Hence, (54) can be positive semi-definite if and only if all τ_q, τ_j $j = 1, \dots, p$ are positive, which is the same as to prove that B defined in Remark 12 is an empty set. By splitting the linear part (with respect to α) of (54) and the quadratic one, we can derive (55). Then by applying the Shur complement, we obtain Problem (17) ∎

$$\begin{bmatrix} \eta \\ \epsilon^1 \\ \vdots \\ \epsilon^p \end{bmatrix}^T \left[\begin{array}{c|c|c|c} I & M(\alpha^1) & \dots & M(\alpha^p) \\ \hline M(\alpha^1)^T & M(\alpha^1)^T M(\alpha^1) & \dots & M(\alpha^1)^T M(\alpha^p) \\ \hline M(\alpha^2)^T & M(\alpha^2)^T M(\alpha^1) & \dots & M(\alpha^2)^T M(\alpha^p) \\ \hline \dots & \dots & \dots & \dots \\ \hline M(\alpha^p)^T & M(\alpha^p)^T M(\alpha^1) & \dots & M(\alpha^p)^T M(\alpha^p) \end{array}\right] \begin{bmatrix} \eta \\ \epsilon^1 \\ \vdots \\ \epsilon^p \end{bmatrix} - 2 \begin{bmatrix} \vdots \\ \tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p) \\ \vdots \\ \vdots \\ -M(\alpha)^T(\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)) \\ \vdots \end{bmatrix}^T \begin{bmatrix} \eta \\ \epsilon^1 \\ \vdots \\ \epsilon^p \end{bmatrix} \tag{53}$$

$$+ ||\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)||^2 \leq \lambda$$

$$\left[\begin{array}{c|c|c|c|c} \tau_0 I - I & M(\alpha^1) & \dots & M(\alpha^p) & \tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p) \\ \hline M(\alpha^1)^T & \tau_1 I - M(\alpha^1)^T M(\alpha^1) & \dots & M(\alpha^1)^T M(\alpha^p) & -M(\alpha_1)^T(\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)) \\ \hline M(\alpha^2)^T & M(\alpha^2)^T M(\alpha^1) & \dots & M(\alpha^2)^T M(\alpha^p) & -M(\alpha_2)^T(\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)) \\ \hline \dots & \dots & \dots & \dots & \dots \\ \hline M(\alpha^p)^T & M(\alpha^p)^T M(\alpha^1) & \dots & \tau_p I - M(\alpha^p)^T M(\alpha^p) & -M(\alpha_2)^T(\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)) \\ \hline \tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p) & -M(\alpha_1)^T(\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)) & \dots & \dots & \lambda - \tau_1\rho_1 - \dots - \tau_p\rho_p \\ & & & & -||\tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)||^2 \end{array}\right], \tag{54}$$

$$\left[\begin{array}{c|c|c|c|c} \tau_0 I - I & M(\alpha^1) & \dots & M(\alpha^p) & \tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p) \\ \hline M(\alpha^1)^T & \tau_1 I & \dots & 0 & 0 \\ \hline M(\alpha^2)^T & 0 & \dots & 0 & 0 \\ \hline \dots & \dots & \dots & \dots & \dots \\ \hline M(\alpha^p)^T & 0 & \dots & \tau_p I & 0 \\ \hline \tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p) & 0 & \dots & 0 & \lambda - \tau_1\rho_1 - \dots - \tau_p\rho_p \end{array}\right] \tag{55}$$

$$- \begin{bmatrix} 0 \\ M(\alpha^1) \\ \dots \\ M(\alpha^p) \\ \tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p) \end{bmatrix} I \Big[0,\ M(\alpha^1),\ \dots,\ M(\alpha^p),\ \tilde{y} - (E_0^1\alpha^1 + \dots + E_0^p\alpha^p)\Big]$$