2024 European Control Conference (ECC)
June 25-28, 2024. Stockholm, Sweden

Nonlinear Data-Driven Moment Matching in Reproducing Kernel Hilbert Spaces*

Alessio Moreschini[1], Matteo Scandella[2], Thomas Parisini[3]

Abstract— **The continuously increasing amount of noisy data demands the development of accurate and efficient models for analysis, modeling, and control. In this article, we propose a novel data-driven moment matching method which employs Tikhonov regularization in the Reproducing Kernel Hilbert Spaces (RKHSs). Specifically, considering a realistic scenario in which the system's plant is unknown and only noisy measured data are available, we provide an estimation of the moment of the unknown plant by solving a regularized optimization problem on RKHS. For, we first demonstrate that the estimation of the moment can be improved via tuning the regularization term, and further, we show under which condition the effect of the transient improves the performance of the estimation. Then, we construct a parameterized model characterized by a kernel-based output mapping. Finally, the proposed data-driven approach is validated and discussed by means of a DC-to-DC Ćuk converter driven by a Van der Pol oscillator.**

Index Terms— **Model reduction, Data-driven moment matching, Kernel-based modeling, Nonlinear system identification.**

I. Introduction

Due to the continuously increasing data volumes from simulation, control, and experimental measurements the study of large-scale systems became a prominent research area in systems and control theory [1]. Large-scale systems are dynamical systems described by numerous ordinary differential equations, which arise from various sources such as interconnected systems [2], [3], spatial discretization of partial differential equations [4], [5], or inherent system complexity [6], [7]. Computational challenges persist in integrating such systems that involve large quantities of data and massive datasets. Hence, to reduce the computational complexity of numerical simulations and facilitate the design of controllers, interpolatory model reduction methods aim to construct a reduced-order model that interpolates the transfer function of the large-scale model at selected interpolation points. Moment matching techniques, which fall under the category of interpolatory methods, are numerically reliable as they can be simply implemented by means of Krylov projectors to achieve interpolation without the need to explicitly evaluate the transfer function. This technique involves matching the moments of an underlying system with an interpolant, potentially of lower order. The breakthrough in extending moment matching to nonlinear dynamical systems came from the intuition that computing moments of a linear system is equivalent to solving a certain Sylvester equation [8]. Building on this insight, the notion of moment matching has been redefined in the time domain for linear and nonlinear systems [9]–[11]. In the time-domain analysis, assuming that a steady-state response exists, the concept of moment involves studying the output response of the underlying system driven by a signal generator defined by the desired interpolation points. However, with the increasing availability of high-dimensional data and advancements in computational power, data-driven model reduction has gained significant attention in recent years [10], [12]–[15]. In the moment matching framework, the problem of estimating the moment of an unknown system from input-output data was earlier considered in [10] employing the ordinary least squares approach. Yet, ordinary least-squares arguments undergo ill-conditioned optimization problems. Specifically, when the number of variables exceeds the number of observations the problem eventually leads to an infinite number of solutions. Further, the presence of noisy data in datasets may also impact the estimation of a meaningful moment of the unknown system. See [16, Sec 3.3] for a detailed discussion.

Contribution: In this paper the problem of estimating the moment of a nonlinear Multi-Input Multi-Output (MIMO) system with feedthrough is considered. Specifically, given noisy data obtained by measuring the output of a certain system we construct a data-driven moment matching method which employs Tikhonov regularization in the Reproducing Kernel Hilbert Spaces (RKHSs), see [16]–[20]. We estimate the moment function from a Hilbert space according to a data adherence criterion expressed using a regularized optimization problem composed of two terms: the empirical cost risk and a regularization term. Hence, taking advantage of the regularized term the optimization problem introduces further constraints which render the solution always uniquely determined. The advantage of employing RKHS is that the kernel method maps the input space of the data to a higher dimensional feature space, in which low-complexity models can be trained, resulting in efficient,

This work has been partially supported by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 739551 (KIOS CoE) and by the Italian Ministry for Research in the framework of the 2017 Program for Research Projects of National Interest (PRIN), Grant No. 2017YKXYXJ.

[1]Alesssio Moreschini is with the Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom. Email: a.moreschini@imperial.ac.uk

[2]Matteo Scandella was with the Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom. He is now with the Department of Management, Information and Production Engineering, University of Bergamo, via Marconi 5, 24044, Dalmine (BG), Italy. Email: matteo.scandella@unibg.it

[3]Thomas Parisini is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK, with the Department of Engineering and Architecture, University of Trieste, 34127 Trieste, Italy, and with the KIOS Research and Innovation Center of Excellence, University of Cyprus, CY-1678 Nicosia, Cyprus. Email: t.parisini@imperial.ac.uk

low-bias, low-variance models. Leveraging on this property and on a preliminary analysis of the measured system, we first show that the estimation of the moment on RKHSs can be improved by properly tuning the regularization term regardless of the choice of the kernel, and then we demonstrate how and under which condition the output transient impacts the performance of the performance of the estimation. Then, we propose a new parameterized model defined by a proper kernel-based output mapping. Furthermore, we provide an easily verifiable necessary condition on the RKHS in order to contain only functions that are suitable to be a moment of the system.

Organization: In Section II we recall the notion of moment matching for nonlinear systems. In Section III we make a preliminary analysis of the measured system. In Section IV we present our data-driven moment matching on RKHSs. Specifically, we first compute the moment estimation by employing Tikhonov regularization on RKHSs and then construct a parametrized model whose output is a function of the RKHS. In Section V, we analyze the effect of the transient in the estimation. In Section VI, we provide an easily verifiable needed condition for the RKHS to be suitable for this application. In Section VII we apply the proposed method to the estimation problem of the moment of the DC-to-DC Ćuk converter driven by a Van der Pol oscillator. Finally, Section VIII concludes the paper.

Notation.: We denote by $\mathbb{R}$ and $\mathbb{N}$ the fields of real and natural numbers, respectively ($0 \in \mathbb{N}$). The set of vectors having n rows with real-valued entries is denoted by $\mathbb{R}^n$, and the set of matrices having n rows and m columns with real-valued entries is denoted by $\mathbb{R}^{n\times m}$. Given $n \in \mathbb{N}$ and a vector $x \in \mathbb{R}^n$, $|x|$ is the Euclidean norm of x. Given $n, m \in \mathbb{N}$, $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix, $\mathbf{1}_{n\times m} \in \mathbb{R}^{n\times m}$ is the matrix where every entry is equal to one, and $\mathbf{0}_{n\times m} \in \mathbb{R}^{n\times m}$ is the matrix where every entry is equal to zero. For compactness, when clear from the context, we denote with 0 the zero matrix of appropriate dimension. Given a symmetric matrix S, $\lambda_{\max}[S]$ and $\lambda_{\min}[S]$ denote the largest and smallest eigenvalue of S, respectively. Given a matrix S, $|S|_2$ denotes the induced 2-norm of S, *i.e.* $|S|_2 = \sqrt{\lambda_{\max}[SS^\top]}$. All mappings are assumed smooth, if not otherwise stated.

II. Moment Matching for Nonlinear Systems

Consider a MIMO continuous-time nonlinear dynamical system of order $n_{\mathrm{x}} \in \mathbb{N}$ with $n_{\mathrm{u}} \in \mathbb{N}$ inputs and $n_{\mathrm{y}} \in \mathbb{N}$ outputs described by equations of the form

$$\dot{x}(t) = f(x(t), u(t)), \qquad x(0) = x_0, \tag{1a}$$

$$y(t) = h(x(t), u(t)), \tag{1b}$$

with $x(t) \in \mathbb{R}^{n_{\mathrm{x}}}$, $u(t) \in \mathbb{R}^{n_{\mathrm{u}}}$, $y(t) \in \mathbb{R}^{n_{\mathrm{y}}}$, $x_0 \in \mathbb{R}^{n_{\mathrm{x}}}$ and smooth mappings $f : \mathbb{R}^{n_{\mathrm{x}}} \times \mathbb{R}^{n_{\mathrm{u}}} \to \mathbb{R}^{n_{\mathrm{x}}}$ and $h : \mathbb{R}^{n_{\mathrm{x}}} \times \mathbb{R}^{n_{\mathrm{u}}} \to \mathbb{R}^{n_{\mathrm{y}}}$ such that $f(0, 0) = 0$ and $h(0, 0) = 0$. To define the notion of moment for system (1) we consider a *signal generator* described by equations of the form

$$\dot{\omega}(t) = s(\omega(t)), \qquad \omega(0) = \omega_0, \tag{2a}$$

$$u(t) = \ell(\omega(t)), \tag{2b}$$

in which $\omega(t) \in \Omega$ and $u(t) \in \mathbb{R}^{n_{\mathrm{u}}}$, with $\Omega \subset \mathbb{R}^{n_\omega}$ a sufficiently small open, connected, invariant neighborhood containing the origin, whereas the smooth mappings $s : \Omega \to \mathbb{R}^{n_\omega}$ and $\ell : \Omega \to \mathbb{R}^{n_{\mathrm{u}}}$ are such that $s(0) = 0$ and $\ell(0) = 0$, respectively. The notion of time-domain moment for nonlinear systems has been defined in [9] in terms of the steady-state response of the cascade interconnection of system (1) with the signal generator (2), that is

$$\dot{\omega}(t) = s(\omega(t)), \qquad \omega(0) = \omega_0, \tag{3a}$$

$$\dot{x}(t) = f(x(t), \ell(\omega(t))), \qquad x(0) = x_0, \tag{3b}$$

$$y(t) = h(x(t), \ell(\omega(t))), \tag{3c}$$

Before introducing the notion of moment for nonlinear systems, we make the following assumptions.

Assumption 1. *The system* (1) *is minimal*[1]*, i.e. locally observable and locally accessible at the origin. The origin of $\dot{x}(t) = f(x(t), 0)$ is locally exponentially stable.*

Assumption 2. *The signal generator* (2) *is locally observable and neutrally stable*[2]*.*

Assumption 3. *There exists a mapping $\pi : \Omega \to \mathbb{R}^{n_{\mathrm{x}}}$ with $\pi(0) = 0$, locally defined in Ω, which is the unique analytic solution of the partial differential equation*

$$\frac{\partial \pi}{\partial \omega}(\omega)\, s(\omega) = f\big(\pi(\omega), \ell(\omega)\big). \tag{4}$$

Definition 1 (Moment)**.** *Consider system* (1) *and the signal generator* (2)*. The moment of system* (1) *at (s, ℓ) is defined as $h(\pi(\cdot), \ell(\cdot))$ where π is the unique solution of the partial differential equation* (4)*.*

Owing to the center manifold theorem [24], the standing assumptions have been used in [9] to give a description of the notion of moment in terms of steady-state output response.

Theorem 1 (See [9])**.** *Consider system* (1) *and the signal generator* (2)*. Suppose Assumptions 1, 2, and 3 hold. Then the moment of system* (1) *at (s, ℓ) is in a one-to-one relation with the steady-state response of the output y of system* (3)*.*

Definition 2 (Moment Matching)**.** *A system described by equations*

$$\dot{\xi}(t) = \bar{f}(\xi(t), u(t)), \qquad \xi(0) = \xi_0, \tag{5a}$$

$$\bar{y}(t) = \bar{h}(\xi(t), u(t)), \tag{5b}$$

with $\xi(t) \in \mathbb{R}^{n_\xi}$ and $\bar{y}(t) \in \mathbb{R}^{n_{\mathrm{y}}}$ is called model of (1) *at (s, ℓ) if* (5) *has the same moment at (s, ℓ) as* (1)*. In this case, system* (5) *is said to achieve moment matching at (s, ℓ).*

[1] For the notion of local observability and local accessibility we refer to [21, Definition 2.10] and [21, Definition 2.11], respectively. For additional details we refer to [22, Chapter 3].

[2] The equilibrium $w = 0$ is a stable equilibrium (in the sense of Lyapunov) and each initial state ω_0 is stable in the sense of Poisson, see [23, Chapter 1].

III. Analysis on the Measured System

The definition of moment we have recalled relies upon the availability of the state-space model of the underlying system (1). In practice, solving the partial differential equation (4) can be computationally challenging even when the structure of the mapping f is perfectly known. However, the model of the system is usually uncertain, or even unknown, and the measurements collected by the output are generally noisy. Hence, in the remainder of this article we focus on the case in which the state-space model is unknown and only noisy measurements of the form

$$\forall t \in \mathbb{R}, \qquad z(t) := y(t) + e(t) \in \mathbb{R}^{n_y} \tag{6}$$

are available, where $y(t) \in \mathbb{R}^{n_y}$ is the output (3c) and $e(t) \in \mathbb{R}^{n_y}$ is an additive white noise. The noisy measurements of the system (3) can be equivalently rewritten as

$$z(t) = h(\pi(\omega(t)), \ell(\omega(t))) + \tau(t) + e(t)$$

where, for all $t \in \mathbb{R}$, the function $\tau : \mathbb{R} \to \mathbb{R}^{n_y}$ is such that

$$\tau(t) := h(x(t), \ell(\omega(t))) - h(\pi(\omega(t)), \ell(\omega(t))) \tag{7}$$

describes the output transient response.

Assumption 4. *The white noise $e(t) \in \mathbb{R}^{n_y}$ with variance $\Sigma \in \mathbb{R}^{n_y \times n_y}$ is such that $\mathbb{E}[e(t)] = \mathbf{0}_{n_y \times 1}$ for every $t \in \mathbb{R}$. Moreover, for every $t_1, t_2 \in \mathbb{R}$, e is such that*

$$\mathbb{E}\Big[e(t_1)e(t_2)^\top\Big] = \begin{cases} \Sigma & \text{if } t_1 = t_2, \\ \mathbf{0}_{n_y \times n_y} & \text{otherwise.} \end{cases} \tag{8}$$

The following statements are corollaries of Theorem 1 and are preliminary results which will be instrumental for characterizing the moment at (s, ℓ) of nonlinear systems of the form (1) from the measurement (6). In particular, Corollary 1.1 shows how the transient output response vanishes exponentially whereas Corollary 1.2 unveils the asymptotic behavior of the noisy measured output.

Corollary 1.1. *For some $\alpha > 0$ and $\delta > 0$, the output transient response* (7) *yields*

$$\forall t \in \mathbb{R}, \qquad |\tau(t)| \leq \delta e^{-\alpha t} |x(0) - \pi(\omega(0))|. \tag{9}$$

Corollary 1.2. *Suppose Assumption 4 holds. Then the noisy measurements* (6) *yield*

$$\lim_{t \to \infty} \mathbb{E}\Big[z(t)\Big] - h(\pi(\omega(t)), \ell(\omega(t))) = 0. \tag{10}$$

The main objective of this article is to devise an algorithm that, given the signal generator (2), constructs an approximation of a system that achieves moment matching at (s, ℓ) using measurements of the output z obtained from an experiment on the interconnected system (3). In particular, we assume that the mappings f, h, τ, and π are unknown. Therefore, we define the dataset

$$\mathcal{D} = \big\{(\bar{t}_i, \bar{\omega}_i, \bar{u}_i, \bar{z}_i)\big\}_{i=1}^N \subseteq \mathbb{R} \times \Omega \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_y} \tag{11}$$

where $N \in \mathbb{N}$ is the amount of data collected, for all $i \in \{1, \ldots, N\}$, $\bar{t}_i$ is the i-th sampling time, $\bar{\omega}_i = \omega(\bar{t}_i)$, $\bar{u}_i = u(\bar{t}_i)$, and $\bar{z}_i = z(\bar{z}_i)$. For simplicity, we assume that the sampling times are ordered, *i.e.* $\bar{t}_i \leq \bar{t}_{i+1}$ for all $i \in \{1, \ldots, N\}$, and define $\bar{q}_i := (\bar{\omega}_i, \bar{u}_i) \in \mathcal{Q} := \Omega \times \mathbb{R}^{n_u}$.

Remark 1. *If the signal generator* (2) *is known, then $\bar{\omega}_i$ and $\bar{u}_i$ can be easily retrieved using the knowledge of (s, ℓ).*

IV. Data-Driven Moment Matching

In this section, we explain the methodology to obtain the approximation of a model that achieves moment matching using the dataset $\mathcal{D}$. In particular, the proposed procedure is divided into two phases: **(i)** firstly, we derive a method that derives an approximation $\hat{\mu}$ of the function $\mu(\omega, u) := h(\pi(\omega), u)$, for all $(\omega, u) \in \mathcal{Q}$; **(ii)** then, we propose a parametric model that matches the moment $\hat{\mu}$.

A. Kernel-Based Estimation of Moments

The objective of this section is to approximate the moment at (s, ℓ) of (1) employing regularized estimation methods and using only the measured dataset $\mathcal{D}$. We propose a data-driven moment matching resorting to the theory of RKHSs [17].

Definition 3 (See [18])**.** *A Hilbert space $\mathcal{H}$ of functions $\mu : \mathcal{Q} \to \mathbb{R}^{\mathbb{R}^{n_y}}$, is said to be a Reproducing Kernel Hilbert Space (RKHS) if and only if the functional that maps $\mu \in \mathcal{H}$ into $z^\top \mu(a) \in \mathbb{R}$ is continuous for any $z \in \mathbb{R}^{n_y}$ and $a \in \mathcal{Q}$.*

Let the estimator of the moment, $\hat{\mu}$, be obtained by solving the optimization problem

$$\hat{\mu} := \operatorname*{argmin}_{\mu \in \mathcal{H}} \sum_{i=1}^N \Big|\bar{z}_i - \mu(\bar{q}_i)\Big|^2 + \rho \big|\mu\big|^2_{\mathcal{H}}, \tag{12}$$

where $\mathcal{H}$ is a RKHS containing functions that map $\mathcal{Q}$ to $\mathbb{R}^{n_y}$, $|\cdot|_{\mathcal{H}}$ is the norm on $\mathcal{H}$, and $\rho > 0$ is a parameter to be tuned. Since the functions that belong in $\mathcal{H}$ are vector-valued function, $\mathcal{H}$ is a vector-valued RKHS. The cost function describing the optimization problem (12) is constructed by summing two terms. The first term is the empirical cost risk which fosters a good fit of the data by minimizing the error between the estimated moment and the measurement output as in [10]. Instead, the second term is a regularization term in the form of an operator which is used to penalize more complex functions, where the complexity is defined using the norm of the selected RKHS. The parameter ρ regulates the relation between the first and the second term. To solve the optimization problem (12), it is necessary to review the properties of the vector-valued RKHS.

Definition 4. *A function $k : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}^{n_y \times n_y}$ is called reproducing kernel if and only if: **(i)** given $a, b \in \mathcal{Q}$, $k(a, b) = k(b, a)^\top$; **(ii)** given $a \in \mathcal{Q}$, $k(a, a)$ is positive semi-definite; **(iii)** given $w \in \mathbb{N}$, $\{a_j : 1 \leq j \leq w\} \subset \mathcal{Q}$ and $\{c_j : 1 \leq j \leq w\} \subset \mathbb{R}^{n_y}$,*

$$\sum_{i=1}^w \sum_{j=1}^w c_i^\top k(a_i, a_j) c_j \geq 0.$$

Theorem 2 (See [18])**.** *There exist a unique reproducing kernel k such that*

$$\forall z \in \mathbb{R}^{n_y}, \forall a \in \mathcal{Q}, \forall \mu \in \mathcal{H}, \quad z^\top \mu(a) = \langle k_a z, \mu \rangle_{\mathcal{H}}$$

where $k_a : \mathbb{R}^p \to \mathcal{H}$ is a linear function such that

$$\forall z \in \mathbb{R}^p, \forall b \in \mathcal{Q}, \quad k(a,b)z = (k_a z)(b). \tag{13}$$

Theorem 2 is a generalization of the Moore-Aronszajn Theorem [17] for vector-valued RKHS. This theorem provides a way to define an RKHS by finding a function k that satisfies Definition 4. However, $\mathcal{H}$ may be an infinite-dimensional space, so if the regularization expression cannot be solved explicitly, it is impossible to search the entire space for a solution. Instead, the Representer theorem tells us that the solution of (12) lies in finite-dimensional subspaces spanned by a finite amount of elements of $\mathcal{H}$.

Theorem 3 (Representer Theorem, see [18], [19])**.** *The solution $\hat{\mu}$ of* (12) *is unique, and we have*

$$\hat{\mu} = \sum_{i=1}^{N} k_{\bar{q}_i} c_i \tag{14}$$

where $k_{\bar{q}_i}$ are as defined in Theorem 2 and $c_i \in \mathbb{R}^{n_y}$ are vectors such that $c = (K + \rho I_{Nn_y})^{-1}\bar{z}$ where $c := [c_1, \ldots, c_N] \in \mathbb{R}^{Nn_y \times 1}$, $\bar{z} := [\bar{z}_1, \ldots, \bar{z}_N] \in \mathbb{R}^{Nn_y \times 1}$ and $K \in \mathbb{R}^{Nn_y \times Nn_y}$ is the block matrix whose (i,j)-th block is $k(\bar{q}_i, \bar{q}_j) \in \mathbb{R}^{n_y \times n_y}$.

The reproducing kernel k whereas characterizing the entire space $\mathcal{H}$ also completely defines the estimator $\hat{\mu} \in \mathcal{H}$ which exploits the RKHS norm as a regularization term. Thus, the choice of the reproducing kernel k has a crucial impact on the quality of the estimation of future output data. The most popular reproducing kernels can be found, *e.g.*, in [16, Sec. 6.6]. Nevertheless, for any reproducing kernel $k : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}^{n_y \times n_y}$, Theorem 3 equipped with (13) provides a way to evaluate the estimated moment $\hat{\mu} \in \mathcal{H}$. In particular, we have

$$\forall q \in \mathcal{Q}, \quad \hat{\mu}(q) = \sum_{i=1}^{N} (k_{\bar{q}_i} c_i)(q) = \sum_{i=1}^{N} k(\bar{q}_i, q) c_i,$$

and more compactly,

$$\forall q \in \mathcal{Q}, \quad \hat{\mu}(q) = S^\star(q)\bar{z}, \tag{15}$$

where

$$K^\star(q) := \Big[k(\bar{q}_1, q), \cdots, k(\bar{q}_N, q)\Big] \in \mathbb{R}^{n_y \times Nn_y},$$
$$S^\star(q) := K^\star(q)(K + \rho I_{Nn_y})^{-1} \in \mathbb{R}^{n_y \times Nn_y}.$$

B. Kernel-Based Parameterized Models

With the estimated moment function at hand, we need to construct a suitable parameterized model achieving moment matching. A model of minimum order of the system (1) achieving moment matching at (s, ℓ) that is when $n_\xi = n_\omega$ has been originally proposed in [9], and later considering feedthrough terms in [11]. In particular, the model is obtained by defining in (5) the following mappings

$$\bar{f}(\xi, u) = s(\xi) - \eta_1(\xi)\ell(\xi) + \eta_1(\xi)u, \tag{16a}$$
$$\bar{h}(\xi, u) = h(\pi(\xi), \ell(\xi)) - \eta_2(\xi)\ell(\xi) + \eta_2(\xi)u, \tag{16b}$$

where $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_y}$ is the output mapping of system (1) satisfying $h(0,0) = 0$, $s : \Omega \to \mathbb{R}^{n_\omega}$ and $\ell : \Omega \to n_u$ are given by the signal generator (2), $\pi : \mathbb{R}^{n_\xi} \to \mathbb{R}^{n_x}$ is the solution of the partial differential equation (4), and the mappings $\eta_1 : \mathbb{R}^{n_\xi} \to \mathbb{R}^{n_\xi \times n_u}$ and $\eta_2 : \mathbb{R}^{n_\xi} \to \mathbb{R}^{n_y \times n_u}$ are parameters of the family of reduced order models which can be used to assign prescribed properties. Yet, since the mapping h and π are unavailable for designing we let the parameterized model be defined by a kernel-based output mapping of the form

$$\bar{h}(\xi, u) = \hat{\mu}(\xi, \ell(\xi)) - \eta_2(\xi)\ell(\xi) + \eta_2(\xi)u,$$

where $\hat{\mu}$ is the unique solution of the optimization problem (12) as provided by Theorem 3.

V. Effect of the Transient

The presence of the output transient (7) may generally affect the quality of the estimation of the moment $\hat{\mu}$. In particular, in this case, $\bar{z}_i - \mu(\bar{q}_i) = \tau(\bar{t}_i) \neq 0$, for all $i \in \{1, \ldots, N\}$. Therefore, the first term of cost function (12) forces $\hat{\mu}$ to be similar to $\mu + \tau$ instead of the moment μ. Thus, to quantify the effect of the output transient on the estimation, we analyze the expected value of the squared deviation of $\hat{\mu}(q)$ from the *ideal* estimator, $\tilde{\mu}(q)$, that is

$$\forall q \in \mathcal{Q}, \quad \varepsilon(q) := \mathbb{E}\Big[\big|\hat{\mu}(q) - \tilde{\mu}(q)\big|^2\Big], \tag{17}$$

where $\tilde{\mu}(q)$ is the estimator which rules out the effect of the output transient, *i.e.* $\tilde{\mu}(q) = S^\star(q)(\bar{z} - \bar{\tau})$ for each $q \in \mathcal{Q}$.

Theorem 4. *For any given dataset $\mathcal{D}$ as in* (11) *and for all $q \in \mathcal{Q}$, the expected value $\varepsilon(q)$ in* (17) *is bounded, i.e.*

$$\exists \delta, \alpha > 0, \quad \varepsilon(q) \le |K^\star(q)|_2^2 \, \frac{N\delta^2}{\rho^2} e^{-2\alpha \bar{t}_1}. \tag{18}$$

The condition (18) yields a bound to the quantity (17) and trivially implies that the larger is $\bar{t}_1$ the better is the estimation of the moment as in the standard Least Square Estimation, see [10].

Now, to evaluate the effect of the output transient (7) on the estimation of $\hat{\mu}(q)$, we analyze the expected value of the difference between the squared deviation of $\hat{\mu}(q)$ from $\mu(q)$ and the squared deviation of $\tilde{\mu}(q)$ from $\mu(q)$. In particular, we define for all $q \in \mathcal{Q}$

$$\hat{\varepsilon}(q) := \mathbb{E}\Big[\big|\hat{\mu}(q) - \mu(q)\big|^2\Big], \quad \tilde{\varepsilon}(q) := \mathbb{E}\Big[\big|\tilde{\mu}(q) - \mu(q)\big|^2\Big],$$

where $\hat{\varepsilon}$ is the estimation error of the real estimator, $\hat{\mu}$, and $\tilde{\varepsilon}$ is the estimation error of the ideal estimator the estimation error of the real estimator, $\tilde{\mu}$. The following result can be proved.

Proposition 1. *For all $q \in \mathcal{Q}$, $\hat{\varepsilon}(q) < \tilde{\varepsilon}(q)$ if and only if*

$$\varepsilon(q) < 2\Big(\mu(q) - S^\star(q)\bar{\mu}\Big)^\top S^\star(q)\bar{\tau}. \qquad (19)$$

The inequality $\hat{\varepsilon}(q) < \tilde{\varepsilon}(q)$ implies that the effect of the output transient (7) improves the estimation $\varepsilon(q)$. In this respect, (19) establishes a necessary and sufficient condition to be verified in the estimation process.

VI. Reproducing Kernel Selection

The methodology presented in Section III stands for every reproducing kernel and corresponding RKHS. However, not all reproducing kernels are suitable for this application, as explained in the following Proposition.

Proposition 2. *Given Assumption 3, the RKHS $\mathcal{H}$ contains only valid moment functions only if $k(0,0) = \mathbf{0}_{n_y \times n_y}$.*

Therefore, when $n_y = 1$, a suitable reproducing kernel is the polynomial kernel with degree $d \in \mathbb{N}$ defined as $k(a,b) = \big(a^\top b\big)^d$, for all $a, b \in \mathcal{Q}$. Other examples can be found by exploiting the property that the multiplication of two kernels is a valid reproducing kernel [25, Prop. 13.2]. Hence, a valid reproducing kernel can be constructed by multiplying a suitable kernel with a generic kernel. For the non-scalar case, *i.e.* $n_y > 1$, a commonly used strategy to design the reproducing kernel is to use a separable kernel [19, Sec. 4]. In particular, if $k_1 : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ is a valid scalar reproducing kernel that satisfies the condition of Proposition 2, we define the vector-valued reproducing kernel as

$$\forall a, b \in \mathcal{Q}, \qquad k(a,b) := k_1(a,b) B_\alpha,$$

where $\alpha \in [0,1]$ is a parameter to be tuned, and $B_\alpha := \alpha \mathbf{1}_{n_y \times n_y} + (1-\alpha) I_{n_y}$. More detail on this type of separable reproducing kernel can be found in [19, Sec. 4.1].

Remark 2. *In this application, the domain $\mathcal{Q}$ of the functions inside the RKHS is composed of two easily separable parts because $\mathcal{Q} = \Omega \times \mathbb{R}^{n_u}$. Therefore, if the unknown model is known to not have a feedforward contribution to the output, it is possible to use a reproducing kernel $k : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}^{n_y \times n_y}$ such that there exists $\bar{k} : \Omega \times \Omega \to \mathbb{R}^{n_y \times n_y}$ with*

$$\forall \omega_a, \omega_b \in \Omega, u_a, u_b \in \mathbb{R}^{n_u}, \qquad k(q_a, q_b) = \bar{k}(\omega_1, \omega_2),$$

where $q_a = (\omega_a, u_a)$ and $q_b = (\omega_b, u_b)$.

Given $\omega_a, \omega_b \in \Omega$ and $u_a, u_b \in \mathbb{R}^{n_u}$, we define $q_a := (\omega_a, u_a) \in \mathcal{Q}$, and $q_b := (\omega_b, u_b) \in \mathcal{Q}$. Then, as an example, we define the reproducing kernel

$$k(q_a, q_b) = k_p(q_a, q_b) k_g(q_a, q_b) B_\alpha, \qquad (20)$$

where $k_p(q_a, q_b)$ is the polynomial kernel defined as $k_p(q_a, q_b) := \big(\tau_\omega \omega_b^\top \omega_a + \tau_u u_b^\top u_a\big)^d$, where $\tau_\omega \geq 0$, $\tau_u \geq 0$ and $d \in \mathbb{N}$ are parameters to be tuned. Whereas $k_g(q_a, q_b)$ is the Gaussian Kernel defined as $k_g(q_a, q_b) := \exp(-\gamma_\omega |\omega_a - \omega_b|^2 - \gamma_u |u_a - u_b|^2)$, where $\gamma_\omega \geq 0$ and $\gamma_u \geq 0$ are parameters to be tuned. The parameters τ_ω and γ_ω regulate the effect of ω whereas τ_u and γ_u regulate the effect u on the estimated moment.

VII. Numerical Example

In this section, we validate the proposed methodology by estimating the moment of an unknown DC-to-DC Ćuk converter from some noisy measured data. The DC-to-DC Ćuk converter, as considered in [26], is described by the differential equations

$$L_1 \frac{di_1}{dt} = uv_2 - v_2 + E, \qquad C_2 \frac{dv_2}{dt} = i_1 - ui_1 + ui_3,$$
$$L_3 \frac{di_3}{dt} = -uv_2 - v_4, \qquad C_4 \frac{dv_4}{dt} = i_3 - Gv_4,$$

where $i_1(t) \in \mathbb{R}_{\geq 0}$ and $i_3(t) \in \mathbb{R}_{\leq 0}$ are electrical currents, $v_2(t) \in \mathbb{R}_{\geq 0}$ and $v_4(t) \in \mathbb{R}_{\leq 0}$ are voltages, L_1, C_2, L_3, C_4, E, and G are positive parameters, and $u(t) \in (0,1)$ is a continuous signal which represents the slew rate of a PWM circuit. We assume that the model is unknown and only uniform noisy samples of the voltages v_2 and v_4 can be measured. We let the signal u be generated by a Van der Pol oscillator of the form (2) with states $\omega_1(t) \in \mathbb{R}$ and $\omega_2(t) \in \mathbb{R}$, and mappings

$$s(\omega) = \begin{pmatrix} \omega_2 \\ \omega_2 - \omega_1^2 \omega_2 - \omega_1 \end{pmatrix}, \qquad \ell(\omega) = 0.1\omega_2 + 0.3.$$

Following the previous discussions, we construct a parameterized model of the form (16) with states $\xi_1(t) \in \mathbb{R}$ and $\xi_2(t) \in \mathbb{R}$, and mappings

$$\bar{f}(\xi, u) = \begin{pmatrix} \xi_2 \\ -\xi_2 - \xi_1 + 3\xi_1^2 + 20u - 10\xi_1^2 u - 6 \end{pmatrix},$$
$$\bar{h}(\xi) = \hat{\mu}(\xi, \ell(\xi)) = \begin{pmatrix} \hat{\mu}_1(\xi, \ell(\xi)) \\ \hat{\mu}_2(\xi, \ell(\xi)) \end{pmatrix},$$

where $\hat{\mu}_1(\xi, \ell(\xi))$ and $\hat{\mu}_2(\xi, \ell(\xi))$ identify the estimated moment of the DC-to-DC Ćuk converter obtained implementing the reproducing kernel (20).

We collected $N = 801$ data points of uniformly sampled noisy measurements of the voltages with sampling time $\bar{t}_k = 31 + \iota 0.01$, for all $\iota \in \{0, \ldots, N-1\}$. Following Theorem 4, we start to sample at $31\,s$ to decrease the effect of the transient on the noise. The identification procedure is carried out for 1001 Monte Carlo experiments to statistically validate the results. In each experiment, the parameters of the DC-to-DC Ćuk converter are set according to [26] whereas the initial conditions $x(0) = (i_1(0), v_2(0), i_3(0), v_4(0))$ are chosen randomly from a normal distribution with expected value $(0.5, 10, -1, -12)$ and variance $2I_4$. The noise affecting the measurements of the voltages are sampled from a normal distribution with $\mathbf{0}_{2\times 1}$-mean and variance $4I_2$. Both the Van der Pol oscillator used as signal generator and the parameterized model are initialized at $(0.5, 0.5)$. The identification is carried out using the reproducing kernel (20) with $\tau_u = \gamma_u = 0$ as the considered system is strictly proper. Since $\tau_u = 0$, we set $\tau_w = 1$ without loss of generality. For simplicity, we also set $d = 1$. Instead, the hyperparameters τ_ω, α and ρ, *i.e.* the regularization parameter of the cost function (12), are set using the data available using the empirical Bayes procedure [27, Sec. 5.4.1]. After the hyperparameters are selected the identified model is optimized

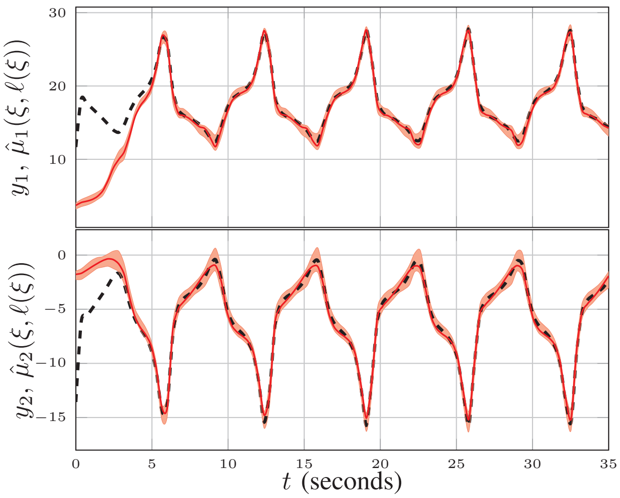

Fig. 1. Time history of the ideal moment (dashed line, black) of the DC-to-DC Ćuk converter, estimated moment of the estimation with the median performance of the Monte Carlo experiment (solid line, red) and the range of all the Monte Carlo experiments (light red area).

using the approach presented in [28] which boosts sparsity and performance of the estimated model.

The good fit of the data obtained by the proposed parameterized model and the outcomes of all the experiments are depicted in Figures 1. Specifically, Figure 1 compares the (ideal) values of the voltages $v_2(t)$ and $v_4(t)$ for $t \in [0, 35]$ with the estimated moments obtained from 1001 experiments and evaluated on the trajectories $\xi_1(t)$ and $\xi_1(t)$ of the parameterized model.

VIII. Conclusions

In this paper, we addressed the problem of estimating the moment of an unknown system from given noisy data. In this regard, we proposed a novel data-driven moment matching method resorting to the theory of RKHS. Data-driven moment matching in the RKHSs provides a promising approach for the estimation of moments in several practical scenarios, such as noisy data in the dataset, and has the advantage of identifying smoother moments from a given dataset. By incorporating Tikhonov regularization and RKHSs, we were able to overcome the limitations of ordinary least squares approaches, which can suffer from ill-conditioned optimization problems. We explored the conditions under which the effect of the output transient improves the performance of the moment estimation. We further introduced a new parameterized model and constructed a suitable reproducing kernel for the moment matching problem. Finally, the proposed approach is validated and illustrated on a DC-to-DC Ćuk converter driven by a Van der Pol oscillator.

References

[1] A. Antoulas, *Approximation of large-scale dynamical systems*. SIAM, 2005.

[2] J. C. Willems, "The behavioral approach to open and interconnected systems," *IEEE Control Syst. Mag.*, vol. 27, no. 6, pp. 46–99, 2007.

[3] A. Moreschini, M. Bin, A. Astolfi, and T. Parisini, "On ϱ-passivity," in *Proc. 22nd IFAC World Congr.*, vol. 56, pp. 8556–8561, 2023.

[4] G. A. Baker, "Finite element methods for elliptic equations using nonconforming elements," *Math. Comput.*, vol. 31, no. 137, pp. 45–59, 1977.

[5] Y. Bar-Sinai, S. Hoyer, J. Hickey, and M. P. Brenner, "Learning data-driven discretizations for partial differential equations," *Proc. Natl. Acad. Sci.*, vol. 116, no. 31, pp. 15344–15349, 2019.

[6] M. Newman, *Networks: An Introduction*. Oxford University Press, 03 2010.

[7] A. Moreschini, S. Monaco, and D. Normand-Cyrot, "Dirac structures for a class of port-Hamiltonian systems in discrete time," *IEEE Trans. Autom. Control*, vol. 69, no. 3, pp. 1999–2006, 2024.

[8] K. Gallivan, A. Vandendorpe, and P. Van Dooren, "Sylvester equations and projection-based model reduction," *J. Comput. Appl. Math.*, vol. 162, no. 1, pp. 213–229, 2004.

[9] A. Astolfi, "Model reduction by moment matching for linear and nonlinear systems," *IEEE Trans. Autom. Control*, vol. 55, no. 10, pp. 2321–2336, 2010.

[10] G. Scarciotti and A. Astolfi, "Data-driven model reduction by moment matching for linear and nonlinear systems," *Automatica*, vol. 79, pp. 340–351, 2017.

[11] J. D. Simard, A. Moreschini, and A. Astolfi, "Parameterization of All Moment Matching Interpolants," in *Proc. Eur. Control Conf. (ECC)*, pp. 1–6, 2023.

[12] A. C. Ionita and A. C. Antoulas, "Data-driven parametrized model reduction in the Loewner framework," *SIAM J. Sci. Comput.*, vol. 36, no. 3, pp. A984–A1007, 2014.

[13] I. V. Gosea, S. Gugercin, and C. Beattie, "Data-driven balancing of linear dynamical systems," *SIAM J. Sci. Comput.*, vol. 44, no. 1, pp. A554–A582, 2022.

[14] A. Moreschini, J. D. Simard, and A. Astolfi, "Model Reduction for Linear Port-Hamiltonian Systems in the Loewner Framework," in *Proc. 22nd IFAC World Congr.*, vol. 56, pp. 9493–9498, 2023.

[15] A. Moreschini, J. D. Simard, and A. Astolfi, "Model Reduction in the Loewner Framework for Second-Order Network Systems On Graphs," in *Proc. 62nd IEEE Conf. Decis. Control (CDC)*, pp. 6713–6718, IEEE, 2023.

[16] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, "Regularization in Reproducing Kernel Hilbert Spaces," in *Regularized System Identification: Learning Dynamic Models from Data*, pp. 181–246, Springer, 2022.

[17] N. Aronszajn, "Theory of reproducing kernels," *Trans. Am. Math. Soc.*, vol. 68, pp. 337–404, May 1950.

[18] C. A. Micchelli and M. Pontil, "On Learning Vector-Valued Functions," *Neural Comput.*, vol. 17, no. 1, pp. 177–204, 2005.

[19] M. A. Álvarez, L. Rosasco, and N. D. Lawrence, "Kernels for Vector-Valued Functions: A Review," *Found. Trends Mach. Learn.*, vol. 4, no. 3, pp. 195–266, 2012.

[20] M. Scandella, A. Moreschini, and T. Parisini, "Kernel-Based Continuous-Time System Identification: A Parametric Approximation," in *Proc. 62nd IEEE Conf. Decis. Control (CDC)*, pp. 1492–1497, IEEE, 2023.

[21] G. Scarciotti and A. Astolfi, "Nonlinear model reduction by moment matching," *Found. Trends Syst. Control*, vol. 4, no. 3-4, pp. 224–409, 2017.

[22] H. Nijmeijer and A. van der Schaft, *Nonlinear dynamical control systems*. Springer, 1990.

[23] C. I. Byrnes, F. D. Priscoli, and A. Isidori, *Output regulation of uncertain nonlinear systems*. Springer Science & Business Media, 1997.

[24] J. Carr, *Applications of centre manifold theory*, vol. 35. Springer Science & Business Media, 2012.

[25] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: The MIT Press, 2018.

[26] H. Rodriguez, R. Ortega, and A. Astolfi, "Adaptive partial state feedback control of the DC-to-DC Cuk converter," in *Proc. Am. Control Conf. (ACC)*, pp. 5121–5126, IEEE, 2005.

[27] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press Ltd, 2006.

[28] M. Scandella, M. Mazzoleni, S. Formentin, and F. Previdi, "A note on the numerical solutions of kernel-based learning problems," *IEEE Trans. Autom. Control*, vol. 66, no. 2, pp. 940–947, 2020.