Annual Reviews in Control 58 (2024) 100969

Contents lists available at ScienceDirect

Annual Reviews in Control

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

Full Length Article

Nonparametric adaptive control in native spaces: A DPS framework (Part I)

Andrew J. Kurdila [a], Andrea L'Afflitto [b,*], John A. Burns [c], Haoran Wang [a]

[a] Department of Mechanical Engineering, Virginia Tech, 1145 Perry St., Blacksburg, 24060, VA, USA
[b] Grado Department of Industrial and Systems, Virginia Tech, 1145 Perry St., Blacksburg, 24060, VA, USA
[c] Interdisciplinary Center for Applied Mathematics, Virginia Tech, 765 West Campus Drive, Blacksburg, 24061, VA, USA

ARTICLE INFO

MSC:
93C40
93D21
93C10
47B32

Keywords:
Adaptive control systems
Adaptive stabilization
Nonlinear systems in control theory
Linear operators in reproducing kernel Hilbert spaces

ABSTRACT

This two-part work presents a novel theory for model reference adaptive control (MRAC) of deterministic nonlinear ordinary differential equations (ODEs) that contain functional, nonparametric uncertainties that reside in a native space. The approach is unique in that it relies on interpreting the closed-loop control problem for the ODE as a simple type of distributed parameter system (DPS), from which implementable controllers are subsequently derived. A thorough comparative analysis between the proposed framework and classical MRAC is performed. The limiting distributed parameter system, which underlies the proposed adaptive control framework, is derived and discussed in detail in this first part of the paper. The second part of this work will detail numerous finite-dimensional implementations of the proposed native space-based approach.

1. Introduction

This paper is the first of a two-part work that presents in a systematic and tutorial manner both the theoretical foundation and several specific algorithms for a general theory of *nonparametric* model reference adaptive control (MRAC), that is, a theory of MRAC systems for which matched uncertainties are not parameterized *a priori* by a finite number of real parameters. One of the key problems in adaptive control is that of steering the trajectories of plants affected by nonlinear, unknown dynamics. It is therefore essential to make provision for the plant's uncertainties and, hence, construct the adaptive control system only leveraging the available information. The weaker the assumptions on the functional uncertainties, the broader the applicability of a specific control system, or, equivalently, the easier the portability of a specific control system to a different plant. One of the goals of this paper is to propose a paradigm shift in the state-of-the-art of deterministic parametric adaptive control theory by presenting adaptive control systems formulated for a wider class of functional uncertainties.

Similarly to other recent forays in dynamical systems and control theory, the proposed approach builds on recent developments in approximation, statistical, and machine learning theory (DeVore, 1998; Rasmussen, 2003; Temlyakov, 2011). We have chosen to frame the development of a nonparameteric control theory in the language of reproducing kernel Hilbert spaces (RKHSs), also known as native spaces, since the advantages of this setting are well-documented in numerous recent studies of estimation in native spaces, Gaussian process estimation, and Bayesian estimation (Berlinet & Thomas-Agnan, 2011; Paulsen & Raghupathi, 2016; Saitoh & Sawano, 2016). In contrast to these latter approaches, which are most often cast in a stochastic setting, this paper focuses on the adaptive control of deterministic systems governed by ordinary differential equations (ODEs).

1.1. The need for a native space setting

A key assumption in the design of existing adaptive control systems for deterministic ODEs is that the unknown functional uncertainties are contained in some subspace contained in the essentially bounded functions. This assumption, which is tacitly made more often than not, is essential to guarantee satisfactory performance of the controller over some sufficiently large domain Ω in the state space $\mathbb{X} \triangleq \mathbb{R}^n$ that contains the closed-loop plant trajectory. At its foundation this paper studies a general approach that is based on powerful approximation error characterizations that are available for functional uncertainties contained in an RKHS. The techniques in this paper exploit one of the most important uses of a native spaces in applications: the construction of rigorous methods for building *scattered bases* used in approximations, and, hence, in the representation of unknown nonlinearities.

The theory of approximations using scattered bases has a long history (Wendland, 2004), and they have been used to great profit

* Corresponding author.
E-mail address: a.lafflitto@vt.edu (A. L'Afflitto).

https://doi.org/10.1016/j.arcontrol.2024.100969
Received 6 December 2023; Received in revised form 14 September 2024; Accepted 17 September 2024
Available online 24 September 2024

in applications where it is unfeasible or inconvenient to use approximations defined over regular grids; see Berlinet and Thomas-Agnan (2011), Paulsen and Raghupathi (2016), Saitoh and Sawano (2016) and their references for a detailed presentation on how scattered basis methods have been used in approximation, interpolation, regression, stochastic processes, and to approximate solutions of partial differential equations (PDEs), for instance. Of all the applications of such scattered bases, some of the most well-known ones are those that use an RKHS to solve certain canonical problems of estimation theory, particularly problems in statistical and machine learning theory. Good accounts of the conspicuous literature on these topics can be found in references such as Williams and Rasmussen (2006) on Gaussian and Bayesian estimation, Cucker and Zhou (2007) on statistical learning theory, or Smola (2002) on machine learning theory. Such bases have also played an important role in the definition of approximation methods as they arise in data-driven modeling and Koopman techniques for both system identification, learning theory, and control (Klus et al., 2018; Klus, Nüske, Peitz, Niemann, Clementi, & Schütte, 2020; Mauroy, Susuki, & Mezić, 2020).

From a practical viewpoint, the choice in this paper to use scattered bases in a native space for adaptive control problems follows from the fact that they are well-suited for choosing data-dependent bases from centers along a trajectory, which naturally gives rise to scattered bases that are not defined in terms of some regular grid in the state space. Such a strategy of choosing bases along a trajectory is pursued in adaptive control strategies such as in Choi and Farrell (2000), Chowdhary, How, and Kingravi (2012), Chowdhary, Kingravi, How, and Vela (2015), Farrell (1998), Kamalapurkar, Rosenfeld, and Dixon (2015), Paruchuri, Guo, and Kurdila (2020, 2022, 2023), Rosenfeld, Kamalapurkar, and Dixon (2019).

> **Box 1**: Parametric adaptive control systems characterize uncertainties using some fixed finite set of basis functions and require some adaptive mechanism to generate the corresponding coefficients. *Direct* adaptive control methods ensure convergence of the tracking error to zero, while these coefficients are not required to approximate the uncertainty. In addition to ensuring convergence of the tracking error to zero, *indirect* adaptive control methods provide coefficients that best approximate the nonlinear uncertainty in the space characterized by this basis. In both approaches, the basis functions are either defined *a priori* by the user or constructed automatically.

1.2. Parametric versus nonparametric estimation

This paper establishes an adaptive control framework that we define *nonparametric* to distinguish it from existing approaches, which we refer to as parametric. Although this paper does not include any offline estimation mechanism, to appreciate the concept of nonparametric control and present this argument, we make a brief digression in the area of parametric and nonparametric estimation.

A traditional method of parametric estimation for an unknown function f arises when, at the outset in formulating the estimation task, we assume that the uncertainty f depends on a finite collection of real parameters, so that $f(x) \triangleq f(x;\Theta_N)$ with $\Theta_N \triangleq \left[\theta_1,\dots,\theta_N\right]^{\mathrm{T}} \in \mathbb{R}^N$ for some *fixed* $N \in \mathbb{N}$. In this setting, given the functional shape of f *a priori*, knowledge of f is completely determined by knowledge of the coefficients Θ_N. In this sense, analysis of estimation methods is *coordinate-centric*. Specifically, the question of whether an estimate $\hat{f} \triangleq \hat{f}(x,\hat{\Theta}_N)$ generated by some algorithm is a good estimate of f is completely characterized by how closely the coordinate estimates $\hat{\Theta}_N$ approximate the true coordinates Θ_N in the finite-dimensional parameter space $\mathbb{R}^N$. It is frequently the case in a traditional problem of parametric estimation that the true coefficients belong to some subset of parameters $C_N \subset \mathbb{R}^N$, and a parametric estimation algorithm uses the set C_N both to generate estimates and to describe the *uncertainty class* for the estimation problem. The *robustness* of an estimation method is quantified by the size of the uncertainty class over which a guarantee of performance holds. However, since all norms are equivalent on $\mathbb{R}^N$, the choice of the norm on the coordinate space $\mathbb{R}^N$ plays no particular role in proving the convergence of parametric estimates that evolve in continuous or discrete time.

> **Box 2**: The proposed adaptive control framework is called *nonparametric* because it does not choose bases *ab initio*. Rather, nonlinear uncertainties are considered as elements of an infinite-dimensional native space.

In contrast to parametric methods, traditional nonparametric estimation methods do not require such detailed insights into the unknown structure of a function f. It is usually assumed that f lies in some uncertainty class $\mathcal{C}$ contained in a *hypothesis space* $\mathcal{H}$ of functions, with $f \in \mathcal{C} \subset \mathcal{H}$; the term hypothesis space is commonly used in discussions of approximation theory or learning theory (Cucker & Zhou, 2007; DeVore, 1998; Temlyakov, 2011), but to a lesser degree in control theory. Here, it is important to note that the hypothesis space $\mathcal{H}$ should be infinite-dimensional. Otherwise, the problem can be reduced to a problem of parametric estimation. In contrast to the parametric case, a conventional nonparametric estimation method introduces a *family of estimates* f_N of f indexed by the number of parameters, and convergence is studied *in the norm on* $\mathcal{H}$, or even some other weaker norm, as the number of parameters $N \rightarrow \infty$. The traditional goal of nonparametric estimation methods is to devise methods that give good estimates in terms of N for different choices of the uncertainty class $\mathcal{C}$ and hypothesis space $\mathcal{H}$, and to characterize the estimation error in terms of the number of parameters N. In contrast to parametric estimation methods, since the hypothesis space $\mathcal{H}$ is infinite-dimensional, the choice of the norm on $\mathcal{H}$ plays a pivotal role in describing the performance of nonparametric estimation methods since all choices of $\mathcal{H}$ do not yield equivalent norms. By definition, consideration of the family indexed by N as a whole is an intrinsic part of conventional nonparametric estimation methods and statistical learning theory (Vapnik, 1999).

> **Box 3**: The proposed nonparametric adaptive control framework is based on viewing the problem as defining a distributed parameter system (DPS). To obtain implementable controllers, the DPS is approximated in finite-dimensional native spaces. A key advantage of nonparametric adaptive control is that it allows assessing the controller's performance as a function of the dimension of the approximating space. A challenge, which is addressed in the second part of this two-paper work, is defining an approximating space that allows attaining user-defined levels of performance in the tracking error dynamics. The existing literature on parametric adaptive control does not describe a general, but precise, approximation strategy to guarantee conclusions of this type.

1.3. Parametric versus nonparametric adaptive control

One of the goals of this paper is to develop a *general nonparametric theory of adaptive control* for nonlinear ODEs that have the same spirit as nonparametric methods of function estimation but guarantee controller performance instead of fidelity of function estimates. Remarkably, the proposed approach does not reduce to the application of nonparametric estimation methods to the classical MRAC framework. Indeed, to remark on this point, we present *direct* MRAC systems only, that is, adaptive control systems that assure certain levels of trajectory tracking error performance, such as its uniform ultimate boundedness or its asymptotic convergence, and boundedness of the adaptive terms, but do not produce or rely on estimates of the unknown elements of the plant model. In the context of this paper, adaptive terms denote functions computed as solutions of partial differential equations in an infinite-dimensional setting. In the second paper of this two-part work, the problem of identifying adaptive terms is reduced to the problem of computing adaptive matrices as solutions of ODEs in a finite-dimensional setting aimed at approximating and realizing in problems of practical interest the results from the infinite-dimensional setting.

The authors would argue that the current practices for deterministic adaptive control of ODEs mostly can be interpreted as parametric methods. All of the standard references on adaptive control for ODEs, including the theory in such popular texts as Farrell and Polycarpou (2006), Ioannou and Sun (2012), Krstic, Kanellakopoulos, and Kokotovic (1995), Lavretsky and Wise (2012), Narendra and Annaswamy (1989), Sastry and Bodson (2011), Slotine and Li (1991), Tao (2003), build controllers for plants affected by some nonlinear uncertainty $f(x;\Theta_N)$, whose functional shape is given and whose number of unknown parameters N is fixed. In these cases, adaptive control synthesis and performance analysis are *coordinate-centric*. Of all these texts, Farrell and Polycarpou (2006), Lavretsky and Wise (2012) come closest to framing a general theory of adaptive control in the philosophy of nonparametric estimation of functions. However, once the uniform approximation assumption (see Section 3.1 below) is made, also, these references focus primarily on stability and convergence in time of coordinate expressions in Euclidean space for a fixed N. These works do not make guarantees of performance for different choices of the *functional uncertainty classes* C in a variety of infinite dimensional hypothesis spaces $\mathcal{H}$ as the number of parameters N varies. Furthermore, they do not describe how coordinate implementations for different N must be related to draw conclusions about performance as a function of N in various function spaces, nor do they connect the controller performance explicitly to the dimension N in some general theoretical framework.

To ease the exposition of the proposed nonparametric control framework and demonstrate its distinguishing features, this paper presents how classical MRAC can be ported into the nonparametric setting. The proposed theory has not been developed in a vacuum but should be understood as a natural culmination of several influences in adaptive control theory over the years. There exist papers that describe approaches that can be construed as nonparametric control methods in one sense or another. For instance, Glasov, Zybin, and Kosyanchuk (2019), Medvedev (2013a, 2013b) describe nonparametric methods for a variety of specific problems. These methods differ from those in this paper in that we seek a general theory that holds for a whole family of adaptive control methods.

Since this paper draws some arguments from nonparametric estimation theory, it can be argued that some of the results presented herein are relatively close to those developed in Boffi, Tu, and Slotine (2022), Chowdhary et al. (2015), Kingravi, Chowdhary, Vela, and Johnson (2012). These references either employ probabilistic methods or combine deterministic and stochastic analyses, and often include techniques that synthesize powerful results on Gaussian processes with (stochastic) Lyapunov stability arguments. The proposed framework, however, is purely deterministic. The incorporation of stochastic elements in a basic, general theory could provide valuable additional tools for the development of algorithms or analysis of performance. The plethora of results from Gaussian process estimation, like those used in Boffi et al. (2022), Chowdhary et al. (2015), Kingravi et al. (2012), would be a valuable addition. However, to keep the analysis in this paper general enough to account for a wide collection of methods and systems, but simultaneously relatively easy to describe in terms of conventional methods, we have elected only to rely on common analysis tools based on deterministic Lyapunov methods. The extension to a stochastic setting, one that develops an accompanying theory with the same qualitative performance bounds described in this paper, would be a substantial and nontrivial addition to this paper. We leave this task for future work.

1.4. Additional novel results

This work does not only propose a novel control framework, namely nonparametric adaptive control but also presents several additional noteworthy contributions to the state-of-the-art of adaptive control theory. For instance, in the second paper of this two-part work, the performance of the control systems is measured by a single performance index, which bounds the ultimate tracking error. This performance index is then used to introduce the notion of "nearly approximation optimal" control. To the authors' knowledge, there is no such standard and general metric to quantify the quality of robust adaptive controllers in the existing literature for ODEs. In existing works on adaptive control, the problem of assessing the quality of asymptotic behaviors reduces to proving the boundedness of all signals at all times and the uniform ultimate boundedness of the tracking error within some bounds that, in general, are functions of unknown quantities, and hence, are usually impossible to assess *a priori*. The approach presented in this paper is the first attempt to provide a systematic performance error metric systematically, for a large collection of hypothesis spaces.

An additional relevant contribution of this work is that, exploiting the theory of native space embedding allows us to quantify explicitly and *a priori* the ultimate bounds on the trajectory tracking error. These bounds are given in terms of the power function associated with the RKHS, which is a metric of the error of approximating an uncertainty lying in an infinite-dimensional space using a finite-dimensional space. The authors are unaware of any similarly simple and explicit error bound for the host of commonly used bases in approximation-based control theory, see for example Farrell and Polycarpou (2006), Lavretsky and Wise (2012).

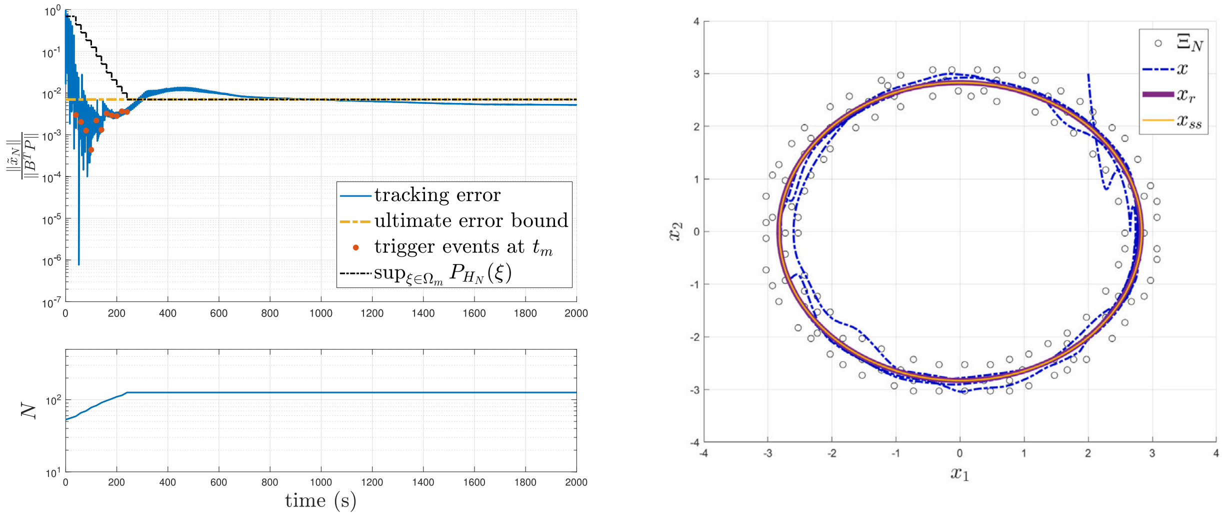
Also worthy of mention is the discussion on the proposed convex projection operator that is used to determine bounded adaptive gains evolving in Hilbert spaces. We provide an insightful study of the relationship between Fréchet derivatives, the convex projection operator, and the classical projection operator. Furthermore, within the context of the newly introduced convex projection operator in an RKHS, we retain all essential properties of the classical continuous projection operator employed in robust parametric MRAC.

1.5. Ramification of the proposed nonparametric MRAC framework

In this section, we outline connections and potential ramifications of the proposed work on nonparametric MRAC with three key trending topics, namely data-driven control methods, the use of neural networks in uncertainty characterization and control, and safety certification of adaptive controllers.

1.5.1. Implications for data-driven methods

Nonparametric control allows the development of novel data-driven methods for its ability to deduce conservative bounds on the tracking error as explicit functions of some measures of the user's ability to approximate the infinite-dimensional space of uncertainties using finite-dimensional parameterizations. As discussed in the second paper of this two-part work, the user's ability to approximate infinite-dimensional uncertainties can be measured by the number of centers for the kernels defining the native spaces, the distance between these kernels within some region, or the smoothness of these kernels. One of the substantial limitations of the current generation of approaches to approximation-based, parametric adaptive control is that the overall theory usually assumes that an *oracle* provides some highly structured information about the uncertainty. Indeed, this oracle defines the subspace used for approximations of the functional uncertainty in the feedback control law, and fixing this subspace amounts to selecting the dimension N and the associated basis functions that give a sufficiently accurate estimate of the uncertainty. Furthermore, existing parametric methods do not provide bounds on the closed-loop system's tracking error as a function of N. However, in the authors' opinion, one of the most substantial challenges in the current generation of adaptive control methods for nonlinear ODEs, which needs to be addressed to enable more systematic applications of adaptive control systems to poorly characterized plant models, is how to reduce the "information burden" on these oracles and improve the characterization of the controller's performance as



(a) (b)

Fig. 1. (a) Normalized tracking error and dimension $N(m)$ of the approximation space over time. The ultimate error is bounded by the approximation error $\sup_{\xi\in\Omega} \mathcal{P}_N(\xi)$, where $\mathcal{P}_N(\cdot)$ denotes the power function; for details, see Section 4.1 below. (b) A comparison of the state trajectory versus the reference model over the set of centers $\xi \in \Omega$ used after the final triggering event. The theoretical results of Powell, Kurdila, L'Afflitto, Wang, and Guo (2023) ensure the ultimate boundedness of the tracking error of the closed-loop system for the event-driven control law. The dimension of the finite-dimensional approximating space $N(t)$ approaches the effective dimension N_∞ as $t \to \infty$.

a function of the parameters characterizing approximations of the uncertainty space. The proposed results make a significant step in this direction by developing both theory and supporting algorithms, where the adaptive controller decides on the appropriate dimension N, the selection of bases, and the choice of subspaces.

To this end, since the proposed controllers are aware of their performance as functions of the number of bases N employed to approximate the infinite-dimensional space of uncertainties (see the second paper of this two-part work), data-driven methods can be employed to choose N as a function of user-defined ultimate bounds on the trajectory tracking error for a given control task. While a discussion of the attendant theory for such a basis augmentation is too lengthy for this paper, Fig. 1, which is extracted from Powell et al. (2023), depicts some initial efforts along these lines. In Powell et al. (2023), the authors use an event-driven MRAC control scheme with the goal of reducing the amount of information that an oracle must provide to formulate the adaptive control problem. The event trigger, which signals that it is time to enrich the basis, is defined explicitly in terms of the bounds that appear in the above table. As shown in Fig. 1, initially, we set $N(t = 0) = 0$ or $N(t = 0)$ very small, and then increase $N(t)$ as $t \to \infty$. The adaptive selection of bases ultimately generates the collection of scattered centers shown in Fig. 1(b). Knowledge about what subspaces constitute a good choice for the control task evolves over time, and, eventually, $\lim_{t\to\infty} N(t) = N_\infty < \infty$.

1.5.2. Implications for adaptive control via neural networks

The study of adaptive control of nonlinear ODEs has evolved synergistically with the development of the theory of approximations in terms of neural networks. Historically, most of the initial work along these lines uses single-layer neural network architectures. A general control theory-centric account of such architectures is given in Farrell and Polycarpou (2006), Lavretsky and Wise (2012). Recently, in light of the empirical evidence of the excellent performance of certain multi-layer neural networks (Daubechies, DeVore, Foucart, Hanin, & Petrova, 2022; DeVore, Hanin, & Petrova, 2021; Yarotsky, 2017), controllers that exploit deep neural networks have been developed in Joshi and Chowdhary (2019), Joshi, Virdi, and Chowdhary (2021), Le, Greene, Makumi, and Dixon (2021), Patil, Le, Greene, and Dixon (2021).

The proposed nonparametric adaptive control framework, with its explicit bounds expressed in terms of the power function, which is a measure of the approximation error, has the potential, to be developed in the future, to influence strongly the development of adaptive control strategies defined in terms of neural networks. Specifically, the proposed nonparametric adaptive control framework has the potential to be used in an *a priori* role in initial controller design and in a *a posteriori* role for advanced adaptive methods that refine the parameterization of the space of functional uncertainties as the control system operates. The proposed framework enables the comparison of different choices of bases, corresponding to different locations of centers, on the performance bounds on tracking error; for details see the second paper of this two-part work. Thus, we argue that this property gives an avenue for addressing a pervasive open question in adaptive control methods based on neural networks.

Recent papers on deep neural networks are motivated by the idea that the theory for single-layer networks is well-established and substantially complete. The authors believe that the framework presented in this paper may partly overturn this assessment. Although it is incontrovertible that the theory of adaptive controllers based on single-layer networks is older and more mature, the overwhelming majority of these approaches make the standard uniform approximation assumption at the outset of the theoretical studies of stability and convergence. In comparison, there are considerably fewer methods that do not make such an assumption *a priori*, and that adaptively add to the bases. Overall, the authors would not describe these methods as stating general theories *per se* since they do not describe a general way to make precise error estimates as a function of N for a variety of nonparametric functional uncertainty classes in a large family of function spaces. In this respect, a general nonparametric theory for adaptive control strategies based on single-layer networks is not yet complete. We believe that the results presented in this paper provide a mechanism to pursue such a general theory.

There is a rapidly growing body of literature on the expressiveness of deep neural networks (DNNs) (Daubechies et al., 2022; DeVore et al., 2021; Yarotsky, 2017). Existing results, however, do not necessarily produce practical algorithms for achieving such rates in an online setting. It is usually assumed that offline optimization, or training, processes generate realizations of such good estimates. Perhaps, just as importantly, the guarantees of the existence of estimates that achieve rates of convergence are made for bases defined over a compact set in $\mathbb{R}^n$, most often over $[0, 1]^n$. These rates of convergence historically have arisen in investigations of methods that excel for extremely large dimensions of approximants. On the other hand, if we consider recent efforts to develop nonparametric adaptive control methods, such as (Choi & Farrell, 2000; Chowdhary et al., 2012, 2015; Farrell, 1998), adaptive control methods are usually designed in terms of scattered bases that are relatively very low dimensional compared to deep neural networks and are defined in terms of samples along a trajectory. Our emphasis in this paper is on constructing nonparametric adaptive control methods with guaranteed or provable performance bounds in terms of N, for parsimonious collections of bases that are feasible for online computations. While the literature on the empirical performance of deep neural networks is, unarguably, exceptionally promising, approximation theorists may still lament the lack of understanding for these architectures and practical algorithms with guaranteed sharp error bounds; see the very recent Daubechies et al. (2022) for a discussion. In this work, we concentrate on cases where *explicit* performance bounds hold for approximants that are sufficiently low dimensional but useful for online approximations. The question of when and how DNNs can be effectively used with the nonparametric setting and goals of this paper is a complex topic, which should be studied carefully in future research.

1.5.3. Implications for safety certifications

Whereas the adaptive control technology has been successfully implemented on multiple, research-grade or very advanced industrial applications, one of the problems hindering the widespread of this technology is the difficulty in finding a universally accepted metric of performance and standard ways to capture the closed-loop system's performance in multiple scenarios, and, hence, certify their safety. One of the key elements of success for classical linear control systems, which still permeates a large number of commercial tools, was the wide array of tools, such as those available from the analysis of Bode and Nyquist plots, to certify their effectiveness. In the realm of adaptive control theory, because of their intrinsic nonlinear nature, such tools cannot be directly employed, and valid, universally accepted alternatives are unavailable.

As already mentioned, the proposed framework allows characterizing explicitly the trajectory error bounds as functions of the user-defined uncertainty classes approximating the actual nonlinear uncertainty. Furthermore, the notion of approximation theory optimal control systems presented in the second paper of this two-part work allows setting a standard metric of performance across control systems predicated assuming that the space of functional uncertainties affecting some plant model is a native space. For these reasons, the authors believe that the proposed framework will further advance the use of adaptive control in common industrial applications.

1.6. Outline of this publication

This paper is structured as follows. Section 2 provides a formal statement of the problem pursued in this paper. Section 3 shows some key features of classical parametric adaptive control, which will be compared to those of nonparametric adaptive control in the following section. In Section 3.1, we present the ubiquitous uniform approximation assumption that constitutes the foundation of nearly all MRAC methods for nonlinear ODEs; to a large extent, this assumption commits an analyst to a real-parametric approach. Successively, in Section 3.2, we review what is probably the most well-known adaptive law, namely the real parametric gradient method. We have chosen to focus on this adaptive law for the simplicity of its structure and of the comparative analysis between the classical parametric and the proposed nonparametric form of MRAC in the following section.

In Section 4, we present a detailed discussion on how nonparametric MRAC differs from parametric MRAC. Section 4.1 carefully discusses how the uniform approximation assumption evolves from the classical, parametric sense to the setting of native spaces, and presents essential tools of native spaces to quantify how approximation errors vary with the number of approximants. Section 4.2 describes the *nonparametric gradient law* and compares it to the classical, parametric one. This section also describes how using the general approach in this paper, it is possible to define rigorously a system that represents the limiting behavior of the control strategy as $N \to \infty$. Section 4.3 gives an account of how functional uncertainty can be described in native spaces and used to describe adaptive control problems. In particular, we present three key uncertainty classes. One of these classes is finite-dimensional and is the one typically used in parametric MRAC. The other two classes are infinite-dimensional and include the finite-dimensional one. These functional uncertainty classes are then used to introduce the notion of "nearly approximation optimal" control schemes. This step introduces a way of measuring the performance of a control scheme relative to a well-known standard, the error of best approximation in the uncertainty class. Finally, Section 4.4 describes how, in a nonparametric framework, it is possible to describe systematically how additional information in the form of *priors* can be used to improve conclusions about controller performance. In this context, the term "priors" refers to additional information about the functional uncertainty f that can be used to make better decisions about controller design. While the use of priors is a standard feature of approximation and learning theory, this topic is not addressed systematically in the existing practices of MRAC control for nonlinear ODEs.

Section 5 marks the beginning of the second part of this paper by showing how the classical MRAC architecture can be ported from the parametric to the nonparametric setting. The resulting governing equations form a limiting distributed parameter system (DPS) because they show the closed-loop system's dynamics as the number of approximating parameters N tends to ∞. The limiting DPS is an "ideal" system since the associated adaptive controller, which is used to define the DPS, contains adaptive gains associated to the functional uncertainty that reside in the uncertainty class in the generally infinite dimensional space $\mathcal{H}$. For this reason, the limiting controller cannot be implemented in practice. In Section 6, we draw conclusions and briefly anticipate key results presented in the second paper associated with this two-part work. These results allow to implementation the limiting DPS in finite-dimensional settings, and, hence, in numerical applications.

2. Problem statement

In this paper, we consider *plants* in the form

$$\dot{x}(t) = Ax(t) + B\left(u(t) + f(x(t))\right), \quad x(t_0) = x_0, \quad t \geq t_0, \tag{1}$$

where $x : [t_0, \infty) \to \mathbb{X}$ denotes the *plant trajectory*, $\mathbb{X} = \mathbb{R}^n$ denotes the *state space*, $u : [t_0, \infty) \to \mathbb{R}$ denotes the *control input*, the *system matrix* $A \in \mathbb{R}^{n\times n}$ is unknown, the *control influence operator* $B \in \mathbb{R}^n$ is known and such that the pair (A, B) is controllable, and $f : \mathbb{R}^n \to \mathbb{R}$ is the *nonlinear matched uncertainty*. The functional uncertainty f resides in the function space $\mathcal{H}$, which we take in this paper to be an RKHS of real, scalar-valued functions defined over the state space.

Plant models in the same form as (1) are found in many engineering problems. For instance, in mechanical problems, the plant trajectory may consist of the set of generalized coordinates (*e.g.*, its position) and their derivatives. In these problems, the block structure of the matrix A can be represented as $A = \begin{bmatrix} 0_{k\times k} & I_k \\ -K_P & -K_D \end{bmatrix}$, where $k \in \mathbb{N}$ denotes the number of degrees of freedom. The top part of this matrix, namely $\begin{bmatrix} 0_{k\times k} & I_k \end{bmatrix}$, is known and contains the kinematic relationship between generalized coordinates and their derivatives. The bottom part of this matrix embodies linear visco-elastic effects, or, at least, their linear components, by means of the unknown matrices $K_P, K_D \in \mathbb{R}^{k\times k}$. Within the context of mechanical systems, the matrix B gives the relationship between forces and moments and the time derivatives of the generalized coordinates, and, hence, is a function of the usually constant inertial properties of the system such as mass or moments of inertia. Similar considerations can be made also for other classes of systems such as electro-mechanical or thermodynamic systems to name a few. Furthermore, in most problems of practical interest, where the relationship between generalized coordinates, their time derivatives, and quasi-velocities is not linear and, hence, cannot be defined by a constant matrix A as in (1), as it occurs, for instance, in the presence of rigid bodies rotating in the three-dimensional space, it is possible to apply feedback-linearizing control inputs that reduce the original nonlinear plant dynamics to the form (1).

For simplicity of exposition, we focus on a single input plant model. The multi-input case is relevant and applies to considerably larger classes of relevant problems. However, it adds some unnecessary complexity to the arguments presented in this paper, and we seek to emphasize what is novel to a large extent without the attendant bookkeeping for multi-inputs. By paying careful attention to the arguments exposed in this paper, results for the m-input problem can be deduced without excessive effort in cases where the functional uncertainty resides in the Cartesian product space $\mathcal{H}^m$. However, the extension to the full complexity of $\mathbb{R}^m$-valued native spaces as in Wang, Kurdila, L'Afflitto, Oesterheld, and Stilwell (2024) for $\mathbb{R}^m$-valued controllers, defined in terms of operator kernels, is substantially more abstract and left for future discussions.

The plant model (1) assumes that the nonlinear uncertainties are time-invariant, that is, explicit functions of the state only, and matched, that is, are premultiplied by the known matrix B. The unmatched uncertainties are linear in the state, that is, are captured by $Ax(t)$, $t \geq t_0$. Nonlinear unmatched uncertainties are neglected. Besides the fact that such a plant model is the one traditionally considered for classical MRAC systems and, as already discussed, can capture broad classes of problems of practical interest, such a model allows us to present the proposed control framework in a relevant, though simpler, context. Once the results in this paper are well established, they can be ported to more general classes of plant models, such as affine-in-control models, as described for example in Farrell and Polycarpou (2006). We leave the details of how to exploit the strategies in this paper for more general models for future study.

Box 4: To present the concept of nonparametric control, this paper presents a model reference adaptive control system tasked with steering the trajectory of the plant model (1) to the trajectory of the reference model (2). Despite classical MRAC, the matched uncertainty f is not characterized *ab initio* using a regressor vector of fixed length but is merely assumed to be an element of a native space.

Our goal is to find a *state-feedback control law* $\mu : \mathbb{X} \to \mathbb{R}$ such that the input $u(t) = \mu(x(t))$ steers the plant trajectory to a reference trajectory $x_r : [t_0, \infty) \to \mathbb{X}$ defined as the solution of the *reference model*

$$\dot{x}_r(t) = A_{\rm ref} x_r(t) + B_{\rm ref} r(t), \quad x_r(t_0) = x_{r,0}, \quad t \geq t_0, \tag{2}$$

where the *reference command input* $r : [t_0, \infty) \to \mathbb{R}$ is continuous and bounded, $A_{\rm ref} \in \mathbb{R}^{n\times n}$ is Hurwitz, $B_{\rm ref} \in \mathbb{R}^n$, the pair $(A_{\rm ref}, B_{\rm ref})$ is controllable, and the *matching conditions*

$$A_{\rm ref} = A + B\alpha^{\rm T}, \tag{3}$$

$$B_{\rm ref} = B\beta, \tag{4}$$

are verified by some $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$; note that α, in general, is unknown, whereas β must be known.

The problem of finding a feedback control law such that the trajectory of a poorly modeled plant follows the trajectory of some reference model is typical of the MRAC literature. The reference model should not be considered as a mere tool to generate the reference trajectory. Designing $A_{\rm ref}$ and $B_{\rm ref}$ implies designing those properties the plant dynamics should have at all times. The reference command input is to be interpreted as some user-defined input to the ideal model.

An example to appreciate the relationship between (1) and (2) is the following. The former may capture the dynamics of some aircraft, which is subject to poorly modeled dynamic and aerodynamic effects, and the latter may capture the aircraft's ideal dynamics subject to the pilot's input $r(\cdot)$. The matching uncertainties are not to be seen as *ad-hoc* constraints, which, as discussed in the following, are needed to derive the trajectory tracking error dynamics. The matching conditions substantially mean that the reference model needs to be chosen so that its dynamics can be mimicked by at least one control input, namely

$$u(t) = \alpha^{\rm T} x(t) + \beta r(t) - f(x(t)), \quad t \geq t_0, \tag{5}$$

should α and $f(\cdot)$ be known.

For a given control law $\mu(\cdot)$, let $x(t;\mu)$ denote the solution to (1) with $u(t) = \mu(x(t))$ so that the *trajectory tracking error corresponding to the control law* $\mu(\cdot)$ is given by

$$e(t) \triangleq e(t;\mu) \triangleq x(t) - x_r(t) = x(t;\mu) - x_r(t), \quad t \geq t_0. \tag{6}$$

The asymptotic tracking problem addressed in this paper reduces to finding $\mu(\cdot)$ so that

$$\lim_{t\to\infty} \|x(t;\mu) - x_r(t)\|_{\mathbb{R}^n} = 0. \tag{7}$$

3. The structure of real parametric adaptive control

In this section, we outline some key structural features of the theory that support classical MRAC schemes. Specifically, in Section 3.1, we discuss the role of the uniform approximation assumption on the definition of the functional uncertainties in the plant dynamics. In Section 3.2, we review one classical adaptive law, also known as parametric gradient learning law. This particular learning law is the foundation for many robust modifications that have subsequently been developed in MRAC methods. Developed in parallel to Section 4 below, this discussion allows appreciation of the fundamental differences between the proposed nonparametric control framework and classical real parametric control.

3.1. The role of the uniform approximation assumption

A key problem driving a great part of the MRAC literature in the past two decades lies in describing or characterizing the matched, functional uncertainty $f(\cdot)$ in (1). Deriving an MRAC control law without some very specific structural information about this term has been impossible this far.

A very large majority, if not the totality, of MRAC systems relies on the uniform approximation assumption. This structural assumption is used to justify a parameterization of $f(\cdot)$ by means of some known functions collected in a *regressor vector*, which are given *a priori* by the user or built automatically over time, and some unknown matrix of coefficients. Usually, this assumption is either explicitly or implicitly assumed to hold over some *approximation set* $\Omega \subseteq \mathbb{X}$, wherein the plant trajectory is expected to lie at all times, given the range of admissible initial conditions. As it will become apparent from Section 4 below, native space embedding adaptive control systems assume that functional uncertainties are elements of a native space and do not rely on a parameterization of such a space using a regressor vector, or an equivalent structure, given *a priori*.

Definition 3.1 (*Uniform Approximation Assumption*)**.** The scalar-valued function $f(\cdot)$ in (1) is approximated over the set $\Omega \subseteq \mathbb{X}$ by an ideal function

$$f_N^*(x) \triangleq \Phi_N^{\rm T}(x)\,\Theta_N^* \quad \text{for all } x \in \Omega, \tag{8}$$

such that

$$|f(x) - f_N^*(x)| \leq \|f(\cdot) - \Phi_N^{\rm T}(\cdot)\,\Theta_N^*\|_{L^\infty(\Omega)} \leq \epsilon \quad \text{for all } x \in \Omega, \tag{9}$$

where $\epsilon > 0$ denotes the *error tolerance*, the *regressor vector* $\Phi_N(\cdot) : \mathbb{X} \to \mathbb{R}^N$ is known and the vector of *ideal parameters* $\Theta_N^* \triangleq \left[\theta_1^*, \dots, \theta_N^*\right]^{\rm T} \in \mathbb{R}^N$ is unknown.

We emphasize that the uniform approximation assumption is a statement about the *offline approximation error* of the function $f(\cdot)$. This assumption is usually the starting point of an analysis of an adaptive control law in what is commonly known as approximation-based adaptive control theory for ODEs (Farrell & Polycarpou, 2006; Lavretsky & Wise, 2012). The error tolerance ϵ, the choice of basis functions or regressors $\phi_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, N$, such that $\Phi_N(x) \triangleq [\phi_1(x), \ldots, \phi_N(x)]^{\mathrm{T}}$ for all $x \in \Omega$, the number N of basis functions, and selection of the subset Ω constitute design choices since the feedback controllers depend on them. The uniform approximation assumption also usually entails the existence of a *hypothesis space* H, so that $\{\phi_i(\cdot)\}_{i=1}^N \subset H \subseteq L^\infty(\Omega)$ although this implication is seldom emphasized in the existing adaptive control setting. Since N is fixed, we refer to this body of work and general approach as *adaptive control theory in Euclidean spaces* or *real-parametric adaptive control theory*.

Usually, the offline approximation error tolerance ϵ is of the same order as the maximum magnitude of any noise or injected disturbance to which the system is subject. The choice of ϵ may also come from an understanding of the fidelity of the model from experiments or computational studies.

Even if ϵ and Ω are known *a priori*, the uniform approximation assumption should be interpreted as defining critical data or information provided by an *oracle*, namely the hypothesis space H, the number N, and the specific choice of basis $\{\phi_i(\cdot)\}_{i=1}^N$. Furthermore, even if ϵ was specified, since $f(\cdot)$ is unknown, it is generally a very difficult task to make an *a priori* choice of the appropriate set Ω, the number of basis functions N, and a good collection of basis functions $\Phi_N(\cdot)$ that ensure a uniform approximation error of magnitude ϵ or smaller. This proves challenging in no small part because such an appropriate set Ω must contain the trajectory of interest, which is not even known *a priori*. Employing larger sets Ω reduces the complexity of assuring that $x(t) \in \Omega$ for all $t \geq t_0$, but also increases the complexity of finding parameters that may verify the uniform approximation assumption. This situation should be contrasted to the task of selecting bases for approximation of partial differential equations (PDEs), where the domain over which approximations are sought is ordinarily known at the *outset*.

In MRAC systems, it is ordinarily true that a learning or adaptive law is introduced to govern the propagation of the real parameter $\hat{\Theta}_N : [t_0, \infty) \to \mathbb{R}^N$ that subsidize for the lack of knowledge of the ideal parameters Θ_N^*. In indirect MRAC, $\hat{\Theta}_N(\cdot)$ is designed to converge or, at least, closely approximate Θ_N^*. In direct MRAC, the convergence of $\hat{\Theta}_N(\cdot)$ to Θ_N^* is not required. For an overview of these two broad classes of techniques, see Farrell and Polycarpou (2006), Ioannou and Sun (2012), Krstic et al. (1995), Lavretsky and Wise (2012), Narendra and Annaswamy (1989), Sastry and Bodson (2011), Slotine and Li (1991).

3.2. The classical real parametric gradient law

In this section, we recall both the control law and the adaptive law that are often understood to define the iconic example of a classical MRAC system.

While relying on the uniform approximation assumption, many working implementations of adaptive control for ODEs are based on the real parametric gradient learning law, or one of its robust modifications. In particular, it follows from (1)–(4) that the trajectory tracking error dynamics is given by

$$\dot{e}_N(t) = A_{\mathrm{ref}} e_N(t) + B\left[u(t) - \alpha^{\mathrm{T}} x_N(t) - \beta r(t) + f_N^*(x_N(t))\right],$$
$$e_N(t_0) = x_0 - x_{\mathrm{ref},0}, \quad t \geq t_0, \tag{10}$$

where the trajectory tracking error (6) is denoted by $e_N(\cdot)$ in (10) to emphasize it dependence on the size N of the regressor vector.

The classic MRAC law is given by

$$\mu(t, x_N, \hat{\alpha}, \hat{\Theta}) \triangleq \hat{\alpha}^{\mathrm{T}} x_N + \beta r(t) - \Phi_N^{\mathrm{T}}(x_N)\hat{\Theta}_N,$$
$$(t, x_N, \hat{\alpha}, \hat{\Theta}) \in [t_0, \infty) \times \mathbb{X} \times \mathbb{R}^n \times \mathbb{R}^N, \tag{11}$$

where $x_N : [t_0, \infty) \to \Omega$ denotes the solution of (1) generated by $u_N(t) = \mu(t, x_N(t)\hat{\alpha}(t), \hat{\Theta}(t))$, the *adaptive gains* $\hat{\alpha} : [t_0, \infty) \to \mathbb{R}^n$ and $\hat{\Theta}_N : [t_0, \infty) \to \mathbb{R}^N$ verify the *adaptive laws* or *learning laws*

$$\dot{\hat{\alpha}}(t) = -\gamma_\alpha x_N(t) e_N^{\mathrm{T}}(t) P B, \quad \hat{\alpha}(t_0) = \hat{\alpha}_0, \tag{12}$$

$$\dot{\hat{\Theta}}_N(t) = \gamma \Phi_N\left(x_N(t)\right) e_N^{\mathrm{T}} P B, \quad \hat{\Theta}_N(t_0) = \hat{\Theta}_0, \tag{13}$$

the *adaptive rate matrices* $\Gamma_\alpha \in \mathbb{R}^{n\times n}$ and $\gamma \in \mathbb{R}^{N\times N}$ are symmetric and positive-definite, and $P \in \mathbb{R}^{n\times n}$ denotes the symmetric, positive-definite solution of the *algebraic Lyapunov equation*

$$-Q = A_{\mathrm{ref}}^{\mathrm{T}} P + A_{\mathrm{ref}} \tag{14}$$

with $Q \in \mathbb{R}^{n\times n}$ symmetric, positive-definite, and user-defined.

Considering the symmetric structure of the learning laws, in this section, for brevity, we let $\Phi_N(x_N)$ stand for $\left[x_N^{\mathrm{T}}, \Phi_N^{\mathrm{T}}(x_N)\right]^{\mathrm{T}}$, $\hat{\Theta}$ for $\left[\hat{\alpha}^{\mathrm{T}}, \hat{\Theta}^{\mathrm{T}}\right]^{\mathrm{T}}$, N for $n + N$, (13) for both (12) and (13), and $\mu(t, x_N, \hat{\alpha}, \hat{\Theta})$ for $\mu(t, \hat{\Theta}) = -\Phi_N^{\mathrm{T}}(x_N)\hat{\Theta}_N$. Finally, we refer to (13) as *the Euclidean or real parametric gradient learning law* since it defines a trajectory in the space of real parameters $\mathbb{R}^N$. A comprehensive discussion on how this learning law is derived is presented in Lavretsky and Wise (2012, Ch. 9).

The performance of the control law (11) and of the learning law (13) can be assessed by analyzing the finite-dimensional, nonlinear, time-varying dynamical system

$$\begin{bmatrix} \dot{e}_N(t) \\ \dot{\tilde{\Theta}}_N(t) \end{bmatrix} = \begin{bmatrix} A_{\mathrm{ref}} & B\Phi_N^{\mathrm{T}}\left(x_N(t)\right) \\ -\gamma \Phi_N\left(x_N(t)\right) B^{\mathrm{T}} P & 0 \end{bmatrix} \begin{bmatrix} e_N(t) \\ \tilde{\Theta}_N(t) \end{bmatrix} + \begin{bmatrix} d(t) \\ 0 \end{bmatrix},$$
$$\begin{bmatrix} e_N(t_0) \\ \tilde{\Theta}_N(t_0) \end{bmatrix} = \begin{bmatrix} x_0 - x_{\mathrm{ref},0} \\ \tilde{\Theta}_0 \end{bmatrix}, \quad t \geq t_0, \tag{15}$$

where $\tilde{\Theta}_N(t) \triangleq \Theta_N^* - \hat{\Theta}_N(t)$ denotes the *parameter error*, and $d : [t_0, \infty) \to \mathbb{R}^n$ denotes a perturbation that depends on the *approximation error*

$$\epsilon_N(x_N(t)) \triangleq f\left(x_N(t)\right) - f_N^*\left(x_N(t)\right). \tag{16}$$

If $d(t) \equiv 0$, $t \geq t_0$, then Lyapunov stability of (15) and asymptotic convergence of $e_N(\cdot)$ to zero are guaranteed (Ioannou & Sun, 2012; Morgan & Narendra, 1977a, 1977b; Sastry & Bodson, 2011). If $d(t) \not\equiv 0$, $t \geq t_0$, as it occurs in most problems of practical interest, then boundedness of the adaptive gains in (15) is not guaranteed.

Over the past few decades, the study of *robust adaptive control methods* has developed a family of techniques that modify the basic gradient learning law to account for the possibility that $d(t) \not\equiv 0$, which occurs also in the presence of unmatched uncertainties. This topic is a central theme of the standard texts on real parametric adaptive control like Farrell and Polycarpou (2006), Ioannou and Sun (2012), Lavretsky and Wise (2012), Narendra and Annaswamy (1989), Sastry and Bodson (2011). Among these robust techniques, it is worthwhile recalling the σ-modification of MRAC, the e-modification of MRAC, various hard and soft deadzone algorithms, the use of the projection operator, robust back-stepping methods, and error bounding methods, among others.

4. The structure of RKHS embedding adaptive control

In this section, we describe the foundations of RKHS embedding and adaptive control under the assumption that the matched uncertainty is the element of an RKHS. Specifically, we focus on five structural features that distinguish the classical parametric adaptive control framework from the proposed nonparametric framework. Sections 4.1 and 4.2 below mirror Sections 3.1 and 3.2, respectively, and show how some features of classical MRAC can be lifted to an RKHS setting. Sections 4.3 and 4.4 below discuss the role of uncertainty classes and smoothness of uncertainties in the proposed nonparametric control framework, which do not find a counterpart in the existing literature on parametric control systems.

4.1. The approximation assumption in a native space

In typical applications of MRAC laws, to verify the uniform approximation assumption stated in Definition 3.1, analysts usually choose a basis for finite-dimensional approximations that is "rich enough" to give good pointwise error estimates in the set of essentially bounded functions $L^\infty(\Omega)$. It is assumed that, for all $x \in \Omega$, the collection of all the basis functions in $\bigcup_{N\geq 0}\bigcup_{i=1,\ldots,N}\phi_i(x)$ is dense in $L^\infty(\Omega)$. Such a guarantee is sometimes referred to as the *universal approximation theorem* (Park & Sandberg, 1991). For example, it is known that the uniform approximation assumption holds for polynomial bases over a compact subset of $\mathbb{R}^n$ by the Weierstrass theorem, and it holds for a host of single or multilayer neural networks. For an account of the range of possibilities to satisfy the uniform approximation assumption, see also Farrell and Polycarpou (2006), Lavretsky and Wise (2012). In particular, see Theorems 12.2 and 12.4 of Lavretsky and Wise (2012) for a few precise statements. However, while universal approximation theorems are popular and often cited in adaptive control approaches, they are rather weak in one important respect: they do not specify *how fast* these approximations converge as functions of N.

One of the most powerful features of native space embedding methods of adaptive control is that much stronger approximation guarantees are *automatically* available owing to some basic properties of native spaces. In the following, we present some of these key properties.

Definition 4.1 (*Admissible Kernel Functions*). Let $\mathfrak{K} : \mathbb{X}\times\mathbb{X} \to \mathbb{R}$ be a *kernel function*. This kernel function is *symmetric* if $\mathfrak{K}(x,y) = \mathfrak{K}(y,x)$ for all $x, y \in \mathbb{X}$. The kernel function $\mathfrak{K}(\cdot,\cdot)$ is *of non-negative type* if, for any integer $N \in \mathbb{N}$, any collection of N points $\Xi_N \triangleq \{\xi_1,\ldots,\xi_N\} \subset \mathbb{X}$, and any set of real coefficients $\{\alpha_1,\ldots,\alpha_N\} \subset \mathbb{R}$, it holds that

$$\sum_{i,j=1}^{N} \mathfrak{K}(\xi_i,\xi_j)\alpha_i\alpha_j \geq 0.$$

The kernel $\mathfrak{K}(\cdot,\cdot)$ is *admissible* if it is symmetric, non-negative, and that the matrix $[\mathfrak{K}(\xi_i,\xi_j)]$ is positive semi-definite for all choices of N centers $\Xi_N \triangleq \{\xi_i \mid 1 \leq i \leq N\} \subset \mathbb{X}$.

Definition 4.2 (*Reproducing Kernel Hilbert Spaces*). Given the admissible kernel function $\mathfrak{K} : \mathbb{X}\times\mathbb{X} \to \mathbb{R}$, a reproducing kernel Hilbert space (RKHS) is defined as

$$\mathcal{H} \triangleq \overline{\text{span}\{\mathfrak{K}_x \mid x \in \mathbb{X}\}}, \tag{17}$$

where $\mathfrak{K}_x(\cdot) \triangleq \mathfrak{K}(\cdot,x)$ and the closure is taken with respect to the inner product $\langle \mathfrak{K}_x, \mathfrak{K}_y\rangle_{\mathcal{H}} \triangleq \mathfrak{K}(x,y)$ for all $x, y \in \mathbb{X}$. Given an RKHS $\mathcal{H}$, the underlying $\mathfrak{K}(\cdot,\cdot)$ is known as the *reproducing kernel function*.

> **Box 5**: Native spaces, also known as reproducing kernel Hilbert spaces (RKHS), are Hilbert spaces defined as the closure of the span of admissible kernel functions. In this paper, functional uncertainties are considered elements of infinite-dimensional RKHSs.

Given a reproducing kernel function $\mathfrak{K}(\cdot,\cdot)$, $\mathfrak{K}_x(\cdot) \in \mathcal{H}$ is sometimes known as the *kernel section*. In the following, we refer to reproducing kernel functions and admissible kernels as kernels, unless specified otherwise.

Definition 4.2 means that the native space $\mathcal{H}$ consists of all functions that can be written as the limit in the norm in $\mathcal{H}$ as $N \to \infty$ of finite linear combinations of the functions $\{\mathfrak{K}_{\xi_{N,i}} \mid 1 \leq i \leq N\}$. In this paper, the reproducing kernel $\mathfrak{K}(\cdot,\cdot)$ is a user-defined function, and, often, the associated kernel section $\mathfrak{K}_x(\cdot)$ is a radial basis function centered at $x \in \mathbb{X}$ (Wendland, 2004). Before we further elaborate on Definition 4.2, it is worthwhile recalling the notion of orthogonal projection on a Hilbert space.

Definition 4.3 (*Orthogonal Projections*). Let $\mathcal{H}$ denote a Hilbert space. The operator $\Pi : \mathcal{H} \to \mathcal{H}$ is an orthogonal projection if

$$\langle f, \Pi g\rangle_{\mathcal{H}} = \langle \Pi f, \Pi g\rangle_{\mathcal{H}} = \langle \Pi f, g\rangle_{\mathcal{H}}, \tag{18}$$

for all $f, g \in \mathcal{H}$.

For the statement of the next result and throughout this paper, let $I \in \mathcal{H}$ denote the *identity operator*.

Theorem 4.1. *Let $\mathcal{H}$ denote a Hilbert space and let $\Pi : \mathcal{H} \to \mathcal{H}$ be an orthogonal projection. Then, $(I - \Pi)$ is an orthogonal projection. Furthermore, for all $f \in \mathcal{H}$, it holds that Πf is orthogonal to $(I-\Pi)f$, that is, $\langle \Pi f, (I-\Pi)f\rangle_{\mathcal{H}} = 0$.*

An additional result that is essential to the comprehension of the pivotal role of the orthogonal projection operator is the following. For the statement of this result, let $\text{dist} : \mathcal{H}\times\mathcal{H} \to \overline{\mathbb{R}}_+$ denote the distance between any two points in $\mathcal{H}$, and, as standard practice in functional analysis, let the norm on $\mathcal{H}$ be induced by the inner product, and the distance on $\mathcal{H}$ be induced by this norm. Finally, we define the distance between a generic point in $f \in \mathcal{H}$ and a subset $\mathcal{H}_N \subseteq \mathcal{H}$ as $\text{dist}(f,\mathcal{H}_N) \triangleq \inf_{f_N\in\mathcal{H}_N} \text{dist}(f, f_N)$. In the case of singleton sets, the distance between a point and a set reduces to the earlier definition of distance between any two points. Thus, the same notation is used for both notions.

Theorem 4.2. *Let $\mathcal{H}$ denote a Hilbert space, let $\Pi_N : \mathcal{H} \to \mathcal{H}$ be an orthogonal projection, and let $\mathcal{H}_N \subseteq \mathcal{H}$ be closed and linear. Then, $f_N = \Pi_N f$ if and only if $f_N \in \mathcal{H}_N$ and $\text{dist}(f_N,\mathcal{H}) = \text{dist}(f_N, f)$.*

This result, in practice, affirms that $\Pi_N f$ is the point in the subspace $\mathcal{H}_N$ that is closest to $f \in \mathcal{H}$. This notion extends the classical notion of orthogonal projection in $\mathbb{R}^n$.

> **Box 6**: Given a function f in an RKHS $\mathcal{H}$ and a user-defined tolerance ϵ, there exist N points in the state space, where f can be approximated with the desired tolerance. Such points, known as centers, can be used to define a finite-dimensional RKHS $\mathcal{H}_N$ that approximates the original space $\mathcal{H}$.

Let us resume our considerations about approximations in an RKHS. By the definition of the closed linear span, we deduce from (17) that, for any $f \in \mathcal{H}$ and $\epsilon > 0$, there exist $N \in \mathbb{N}$, a set of centers

$$\Xi_N \triangleq \left\{\xi_1,\ldots,\xi_N\right\} \subset \Omega \subset \mathbb{X}, \tag{19}$$

and a collection of coefficients $\Theta^* \triangleq \{\theta_1^*,\ldots,\theta_N^*\}$ such that

$$\left\| f - \sum_{i=1}^{N} \mathfrak{K}_{\xi_i}\theta_i^* \right\|_{\mathcal{H}} \leq \epsilon. \tag{20}$$

In other words, if $f \in \mathcal{H}$ and we set $\epsilon > 0$, then there always exists a set of centers Ξ_N and a subspace

$$\mathcal{H}_N \triangleq \text{span}\left\{\mathfrak{K}_{\xi_i} \mid \xi_i \in \Xi_N,\ 1 \leq i \leq N\right\} \tag{21}$$

such that the uniform approximation assumption outlined in Definition 3.1 holds for $f_N^* \triangleq \sum_{i=1}^{N}\mathfrak{K}_{\xi_i}\theta_i^*$. In fact, we can use the $\mathcal{H}$-orthogonal projection $\Pi_N : \mathcal{H} \to \mathcal{H}_N$ to define a set of coefficients Θ_N^* that yield approximations in $\mathcal{H}_N$ that are optimal for this representation in that they have minimal error. This is why we claim that a kind of approximation assumption is easy to demonstrate in a native space from its definition.

All kernel functions considered in this paper are bounded on the diagonal. As discussed in the following, this property is essential to further elaborate on the approximation error using RKHSs.

Definition 4.4 (*Bounded on the Diagonal Kernels*). Let $\mathfrak{K}(\cdot,\cdot)$ be the admissible kernel that defines the native space $\mathcal{H}$. The kernel function is *bounded on the diagonal* if there is a constant $\bar{\mathfrak{K}} > 0$ such that

$$\mathfrak{K}(\xi,\xi) \triangleq \|\mathfrak{K}_\xi\|_{\mathcal{H}}^2 \leq \bar{\mathfrak{K}}^2, \quad \text{for all } \xi \in \mathbb{X}. \tag{22}$$

The property of boundedness on the diagonal is satisfied for many standard choices of kernels defined on $\mathbb{R}^n$, including the Gaussian, Sobolev-Matern, inverse multiquadric, or Wendland kernels (Wendland, 2004). The norm in $\mathcal{H}$ dominates the norm in $L^\infty(\Omega)$ in most cases of interest (Paulsen & Raghupathi, 2016; Saitoh & Sawano, 2016). To prove this fact, we introduce the *evaluation functional* $E_x : \mathcal{H} \to \mathbb{R}$ for $x \in \mathbb{X}$, which is defined so that

$$E_x f \triangleq f(x), \quad \text{for all } f \in \mathcal{H} \text{ and } x \in \mathbb{X}. \tag{23}$$

Box 7: All RKHS considered in this paper are bounded on the diagonal. A consequence of this property is that both the evaluation operator and its adjoint are uniformly bounded. Since these operators are linear, uniform boundedness also implies their continuity.

If the kernel underlying an RKHS is bounded on the diagonal, then the operator norm $\|E_x\| = \|E_x^*\|$ is *uniformly* bounded. Indeed, considering E_x as acting on $\mathcal{H}$ into $\mathbb{R}$, we have that $\|E_x\| \le \bar{\mathfrak{K}}$ in the operator norm since, by the Cauchy–Schwarz inequality,

$$|E_x f| = |\langle f, \mathfrak{K}_x\rangle_{\mathcal{H}}| \le \|\mathfrak{K}_x\|\|f\|_{\mathcal{H}} \le \sqrt{\mathfrak{K}(x,x)}\|f\|_{\mathcal{H}} \le \bar{\mathfrak{K}}\|f\|_{\mathcal{H}}, \tag{24}$$

for all $f \in \mathcal{H}$. This same line of reasoning implies that all continuous functions in $\mathcal{H}$ are bounded on $\mathbb{X}$, and, in fact, $C(\mathbb{X}) \hookrightarrow \mathcal{H}$.

Any real Hilbert space $\mathcal{H}$ with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ is an RKHS if and only if all of the linear evaluation functionals are bounded (Aronszajn, 1950; Wendland, 2004). In this case, it follows from (20) and (24) that

$$\frac{1}{\bar{\mathfrak{K}}}\left| f(x) - \sum_{i=1}^{N} \theta_i^* \mathfrak{K}_{\xi_i}(x)\right| \le \left\| f - \sum_{i=1}^{N} \theta_i^* \mathfrak{K}_{\xi_i}\right\|_{\mathcal{H}} \le \epsilon, \quad \text{for all } f \in \mathcal{H}. \tag{25}$$

Box 8: The use of RKHSs allows quantifying *explicitly in terms of N* the pointwise approximation error throughout the state space for approximations using N basis functions.

We conclude from these observations that being able to prove (20) through RKHS theory allows addressing the performance of the uniform approximation assumption in terms of (9). The kernels employed in this paper always allow drawing this conclusion. In fact, as anticipated, we choose kernels such that $\|E_x\| \le \bar{\mathfrak{K}}$ for all $x \in \mathbb{X}$ in the operator norm for some constant $\bar{\mathfrak{K}} > 0$.

4.2. The nonparametric gradient learning law

In this section, we present the first, and most simple adaptive control system in the nonparametric setting. In particular, we discuss how, by interpreting the nonlinear uncertainty in (1) as a type of functional uncertainty, the adaptive law can be transposed from the parametric setting to the nonparametric setting.

Let $y(t) \triangleq E_{x(t)} f = f(x(t))$, $t \ge t_0$, for some unknown uncertain function $f \in \mathcal{H}$, where $\mathcal{H}$ is an RKHS. Thus, introduce a time-dependent, nonparametric *adaptive gain* $\hat{f}(t,\cdot) \in \mathcal{H}$ associated to f, and a *function adaptive error* $\tilde{f}(t,\cdot) \triangleq f - \hat{f}(t,\cdot)$. Finally, define the *output prediction* $\hat{y}(t)$, and the *output error* $\tilde{y}(t)$ as

$$\hat{y}(t) = E_{x(t)}\hat{f}(t,\cdot) = \hat{f}(t, x(t)), \tag{26}$$

$$\tilde{y}(t) = y(t) - \hat{y}(t) = E_{x(t)}\tilde{f}(t,\cdot), \tag{27}$$

respectively.

Let $\mathcal{H}$ be an RKHS and $F : \mathcal{H} \to \mathbb{R}$ be such that $F(f) = \langle f, f\rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$. The Fréchet derivative of $F(\cdot)$ at $f \in \mathcal{H}$ is given by the linear functional $D : \mathcal{H} \to \mathbb{R}$ such that

$$Dg = 2\langle g, f\rangle_{\mathcal{H}}, \quad \text{for all } g \in \mathcal{H}. \tag{28}$$

Now, define the *error functional*

$$J(\tilde{f}) \triangleq \frac{1}{2}\tilde{y}^{\mathrm{T}}\tilde{y} = \frac{1}{2}\langle E_{x(\cdot)}\tilde{f}, E_{x(\cdot)}\tilde{f}\rangle_{\mathbb{R}} = \frac{1}{2}\left\langle E_{x(t)}^* E_{x(t)}\tilde{f}, \tilde{f}\right\rangle_{\mathcal{H}} \tag{29}$$

where $E_{x(\cdot)}^* : \mathbb{R} \to \mathcal{H}$ denotes the *adjoint of* $E_{x(\cdot)}$, and the operator $E_{x(\cdot)}^* E_{x(\cdot)}$ is such that $E_{x(t)}^* E_{x(t)} : \mathcal{H} \to \mathcal{H}$ for each $t \in [t_0, \infty)$. Proceeding as in Ioannou and Fidan (2006, App. B), it follows from (28) that the nonparametric adaptive law for $\tilde{f}(t,\cdot)$ is given by

$$\frac{\partial \tilde{f}}{\partial t}(t,\cdot) = -\gamma DJ\left(\tilde{f}(t)\right) = -\gamma E_{x(t)}^* E_{x(t)}\tilde{f}(t,\cdot) = -\gamma E_{x(t)}^*\tilde{y}(t), \quad t \ge t_0, \tag{30}$$

where $\gamma > 0$ is a user-defined adaptive gain. It is worthwhile noting that the arguments in Ioannou and Fidan (2006, App. B) yield in finite dimensions, whereas the proposed arguments yield on the infinite-dimensional space $\mathcal{H}$. In Section 5, we formally prove the effectiveness of an MRAC system, whose nonparametric law has the *gradient law* (30) as a special case. For this reason, and to focus on the analogies and differences with the results in Section 3.2, we leave (30) substantially unproven.

It is apparent from (30) that, for all $t \in [t_0, \infty)$, the adjoint operator $E_{x(t)}^*$ of the evaluation functional $E_{x(t)}$ defines the generalized, nonparametric gradient learning law. As for the gradient law in the Euclidean space $\mathbb{R}^n$, the generalized gradient law implies that the nonparametric function error $\tilde{f}(t,\cdot)$ evolves in "the local direction of fastest decrease" of the functional $J(\tilde{f})$. In (30), however, the local descent direction is *not* a vector, but a direction in the function space $\mathcal{H}$.

Next, consider the plant dynamics (1) and assume that A is known. Applying the control input

$$u(t) = -\alpha^{\mathrm{T}} x(t) + \beta r(t) - E_{x(t)}\hat{f}(t,\cdot), \quad t \ge t_0, \tag{31}$$

where $\tilde{f}(t,\cdot)$ satisfies (30), the trajectory tracking error dynamics and the function adaptive error dynamics are captured by

$$\begin{Bmatrix} \dot{e}(t) \\ \dfrac{\partial \tilde{f}(t,\cdot)}{\partial t}\end{Bmatrix} = \begin{bmatrix} A_{\mathrm{ref}} & BE_{x(t)} \\ -\gamma E_{x(t)}^* B^{\mathrm{T}} P & 0\end{bmatrix}\begin{Bmatrix} e(t) \\ \tilde{f}(t,\cdot)\end{Bmatrix},$$

$$\begin{Bmatrix} e(t_0) \\ \tilde{f}(t_0,\cdot)\end{Bmatrix} = \begin{Bmatrix} x_0 - x_{r,0} \\ \tilde{f}_0(\cdot)\end{Bmatrix}, \quad t \ge t_0, \tag{32}$$

Eq. (32) defines a DPS since the equation for $\partial\tilde{f}(t,x)/\partial t$ is a PDE, not an ODE.

Box 9: The proposed nonparametric adaptive framework mirrors in several key elements the classical parametric adaptive framework. However, in the nonparametric approach, the adaptive law associated with the functional uncertainty is a distributed parameter system, whereas in the classical parametric adaptive framework, this adaptive law reduces to a matrix ordinary differential equation.

The DPS (32) should be carefully compared to (15) with $d(t) \equiv 0$, $t \ge t_0$, that arises in the real parametric theory. It is apparent how (15) evolves in a finite-dimensional space, namely, $\mathbb{R}^n \times \mathbb{R}^N$, whereas (32) evolves in a generally infinite-dimensional space, namely $\mathbb{R}^n \times \mathcal{H}$. For this reason, the control input (31) cannot be implemented in practice, and it is necessary to approximate (31) to obtain realizable control inputs; this point is addressed in the second paper of this two-part work. If we generate a collection of approximations of (32) as $N \to \infty$, then the limiting behavior of the closed-loop response $x_N(\cdot)$ that can be obtained in practice converges to that of (32) that, for this reason, (32) is said to define the *limiting DPS system*.

In conclusion, this section has shown that an additional structural difference between real parametric adaptive control and RKHS embedding methods lies in the fundamental role played by the *limiting DPS* in the latter approach. It is safe to say that there is no such simple analog of the limiting DPS in classical treatments of real parametric adaptive control theory since real parametric adaptive control selects an *a priori* parameterization of the functional uncertainty.

4.3. Functional uncertainty classes

Another structural difference between real parametric and nonparametric adaptive control techniques lies in their definition and use of functional uncertainty classes. A central tenet of robust control theory is that it guarantees both the stability of the closed-loop system and its performance within some user-defined levels for a wide variety of "nearby" uncertain systems. As we explain in this section, invoking the approximation assumption in terms of the norm on a native space, there are associated natural choices of functional uncertainty classes. The key idea is that the controller performance in native space embedding methods is measured relative to the offline rates of convergence of best approximations for these functional uncertainty classes.

Three functional uncertainty classes are explored in this paper. These classes are formally defined as follows.

Box 10: The proposed nonparametric adaptive framework allows addressing functional uncertainties within classes that are considerably broader than those that can be addressed employing classical parametric control frameworks.

Definition 4.5 (*Functional Uncertainty Classes*). Let $N \in \mathbb{N}$, $R > 0$, and $\epsilon > 0$. The *real parametric uncertainty class of radius* R is defined as

$$C_{\Phi_N,R} \triangleq \left\{ f = \Phi_N^{\mathrm{T}} \Theta_N \mid \Theta_N \in \mathbb{R}^N, \|f\| \le R \right\} \subset \mathrm{span}\{\Phi_N\}. \tag{33}$$

The *nonparametric uncertainty class of radius* R *and projection error less than* ϵ is defined as

$$C_{N,\epsilon,R} \triangleq \left\{ f \in \mathcal{H} \mid \|(I - \Pi_N) f\|_{\mathcal{H}} \le \epsilon, \|f\|_{\mathcal{H}} \le R \right\} \subset \mathcal{H}. \tag{34}$$

The *nonparametric uncertainty class of radius* R is defined as

$$C_R \triangleq \left\{ f \in \mathcal{H} \mid \|f\|_{\mathcal{H}} \le R \right\} \subset \mathcal{H}. \tag{35}$$

The set $C_{\Phi_N,R}$ is an example of a real parametric uncertainty class since its elements are completely characterized by the N real coefficients in Θ_N that appear in representations in $\mathcal{H}_N$. As discussed in Section 3, this class is employed either implicitly or explicitly in the overwhelming majority of existing adaptive control schemes. Membership in $C_{\Phi_N,R}$ is characterized by the inequality $\Theta_N^{\mathrm{T}} \mathbb{G}_N \Theta_N \le R^2$, where $\mathbb{G}_N \triangleq [\langle \phi_i, \phi_j \rangle_{\mathcal{H}}]_{(i,j)}$ denotes the *fixed Grammian matrix* and $\phi_i(\cdot)$, $i \in \{1, \dots, N\}$, the ith element of Φ_N.

The uncertainty class $C_{N,\epsilon,R}$ plays a pivotal role in this paper because it captures the uncertainty due to the complexity of capturing $f \in \mathcal{H}$ by projecting f into $\mathcal{H}_N \subset \mathcal{H}$. This class will play a key role in the second paper of this two-part work, where the practical implementation of the proposed nonparametric adaptive control framework will be put in place.

Finally, if R can be set arbitrarily large, then C_R captures the ideal uncertainty set as it comprises all uncertainties of a given magnitude or less. The uncertainty classes $C_{N,\epsilon,R}$ and C_R are said to be nonparametric. Indeed, if $\mathcal{H}$ is infinite-dimensional, then functions in these uncertainty classes cannot be parameterized using representations in terms of a finite set of scalars and a finite collection basis functions. If $\mathcal{H}$ is infinite-dimensional, then, irrespectively of how small we choose ϵ, we have that, for any N and R, $C_{N,\epsilon,R} \not\subset C_{\Phi_N,R}$. In general, it holds that

$$C_{\Phi_N,R} \subset C_{N,\epsilon,R} \subset C_R, \tag{36}$$

and, for the scope of this paper, N, R, and ϵ are to be considered as user-defined. As discussed in the second paper of this two-part work, in practical implementations, applying the proposed control systems, ultimate bounds on the tracking error can be estimated in terms of these parameters, whereas classical parametric adaptive control theory does not allow such explicit bounds. As discussed in Powell, Kurdila, L'Afflitto, Wang, and Guo (2024) and Kurdila, L'Afflitto, and Burns (2025, Ch. 6), employing data-driven methods, N, R, and ϵ can be chosen adaptively by the control system to ultimately drive the tracking error to zero or minimize some cost functions. However, this point is beyond the scope of this work.

Note that $C_{\Phi_N,R}$ is a compact, convex subset of the finite-dimensional normed space $\mathcal{H}_N \triangleq \mathrm{span}(\Phi_N) \subseteq \mathcal{H}$. However, $C_{N,\epsilon,R}$ and C_R are only closed, convex, norm-bounded subsets of the generally infinite dimensional space $\mathcal{H}$. Furthermore, while C_R and $C_{N,\epsilon,R}$ are closed and bounded, they are compact if and only if $\mathcal{H}$ is finite-dimensional.

Overall, guarantees of the performance of control strategies for systems having functional uncertainty in the classes $C_{\Phi_N,R}$, $C_{N,\epsilon,R}$, and C_R improve for smaller uncertainty classes. If we only consider the smallest functional uncertainty class $C_{\Phi_N,R}$, it is easy to derive excellent bounds on performance in a variety of situations using classical techniques from real parametric adaptive control theory (Lavretsky & Wise, 2012, Ch. 9, 11). However, general and precise guarantees, which are explicit in the number N, for example, on the performance of adaptive control methods when $f \in C_{N,\epsilon,R}$ or just $f \in C_R$ are unstudied. The second paper of this two-part work addresses this key point by correlating ultimate bounds on the trajectory tracking error to the parameters N, ϵ, and R that define the uncertainty classes $C_{\Phi_N,R}$, $C_{N,\epsilon,R}$, and C_R in which uncertainties are assumed to lay.

4.4. Smoothness, hypothesis spaces, and controller performance

The proposed nonparametric adaptive control approaches allow studying how the smoothness of the functional uncertainty f and the smoothness of the choice of basis influences the performance of the closed-loop control scheme. The authors are unaware of any analogous theoretical setting in real-parametric adaptive control for ODEs.

In the proposed setting, it possible to choose a native space so that it is filled with "not-so-smooth" functions that lie just inside $L^\infty(\Omega)$, where $\Omega \subset \mathbb{X}$ denotes some subset containing the closed-loop plant trajectory, or choose kernels that generate native spaces of very smooth functions. In this work, and especially in its second part, we will show that choosing the native space $\mathcal{H}$ in which to formulate the adaptive control problem has important implications on the closed-loop performance, and if we happen to know that the functional uncertainty lies in a smoother native space, then there can be good reasons to use such a smoother space for generating controllers.

5. The nonparametric control law and the limiting DPS

In this section, we provide MRAC laws for plants in the same form as (1) while elaborating on the results discussed in Section 4 in general and Section 4.2 in particular. Successively, we study sufficient conditions for the existence and uniqueness of solutions as well as forward completeness. In the following, we newly assume that A in (1) is unknown, and the nonlinear matched uncertainty f is contained in the native space $\mathcal{H}$ of real-valued functions over $\mathbb{X} \triangleq \mathbb{R}^n$. The native space $\mathcal{H}$ is defined in terms of the admissible kernel $\mathfrak{K} : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$.

5.1. MRAC laws employing RKHS to capture nonlinearities

As summarized in Section 4, the RKHS embedding technique initially views the function $f \in \mathcal{H}$ as a nonparametric, functional uncertainty, so that, by (23) and (17), it holds that

$$f(x) = E_x f = \langle \mathfrak{K}_x, f \rangle_{\mathcal{H}}, \quad \text{for all } x \in \mathbb{X}. \tag{37}$$

Thus, to emphasize that the original plant model in (1) is understood in the sense of RKHS embedding, we rewrite it as

$$\dot{x}(t) = Ax(t) + B\left(u(t) + E_{x(t)} f\right), \quad x(t_0) = x_0, \quad t \ge t_0, \tag{38}$$

or, alternatively, as

$$\dot{x}(t) = Ax(t) + B\left(u(t) + \left\langle \mathfrak{K}_{x(t)}, f \right\rangle_{\mathcal{H}}\right), \quad x(t_0) = x_0.$$

As in the usual real parametric MRAC for ODEs, we begin the study of native space embedding MRAC by assessing the matching conditions. We propose an *ideal control law*

$$\mu(t, x, \alpha, \mathfrak{K}_x, f) \triangleq \alpha^{\mathrm{T}} x + \beta r(t) - \langle \mathfrak{K}_x, f \rangle_{\mathcal{H}},$$

$$(t, x, \alpha, \mathfrak{K}_x, f) \in [t_0, \infty) \times \mathbb{X} \times \mathbb{R}^n \times \mathcal{H} \times \mathcal{H}, \tag{39}$$

and the *ideal control input*

$$u^*(t) = \mu(t, x(t), \alpha, \mathfrak{K}_x(t), f), \quad t \geq t_0, \tag{40}$$

where $x : [t_0, \infty) \to \mathbb{X}$ denotes a solution of (1) with control input $u^*(t)$ and α and β verify the matching condition (3) and (4), respectively. Note that, since A is unknown, α is unknown. Furthermore, f is unknown. For these reasons, (40) is considered an *ideal* control input. It is also worthwhile noting that the control law (39) allows for $x \in \mathbb{X}$, whereas the classical control law (11) requires that $x \in \Omega \subset \mathbb{X}$ or, alternatively, that the parameterization of the uncertainty by means of the regressor vector to yield for all $x \in \mathbb{X}$.

Box 11: Similarly to the classical parametric framework, the nonparametric adaptive control framework requires the matching conditions to be verified to assure the existence of at least one ideal controller able to steer the plant trajectory toward the reference trajectory.

In the problem at hand, the matching conditions depend on the RKHS space $\mathcal{H}$ and the choice of kernel $\mathfrak{K}(\cdot, x)$, $x \in \mathbb{X}$, that appears in $\langle \mathfrak{K}_x, f \rangle_{\mathcal{H}}$. To ensure that it is possible for trajectories of (1) to track trajectories of the reference model (2), we set

$$\begin{aligned} A_{\mathrm{ref}} x_r(t) + B_{\mathrm{ref}} r(t) &= Ax(t) + B\left(\alpha^{\mathrm{T}} x(t) + \beta r(t)\right. \\ &\quad \left. - \left\langle \mathfrak{K}_{x(t)}, f \right\rangle_{\mathcal{H}} + \left\langle \mathfrak{K}_{x(t)}, f \right\rangle_{\mathcal{H}} \right), \\ &= \left(A + B\alpha^{\mathrm{T}}\right) x(t) + B\beta r(t), \quad t \geq t_0. \end{aligned}$$

By comparing coefficients, we recover the classical matching conditions (3) and (4). Furthermore, if (3) and (4) are verified, then the nonparametric feedback control law $\mu(\cdot)$ is compatible in the native space embedding method for any choice of RKHS $\mathcal{H}$ and kernel $\mathfrak{K}(\cdot,\cdot)$.

To design nonparametric MRAC laws that account for uncertainties in both A and f, we consider the nonparametric adaptive control input

$$u(t) = \mu(t, x(t), \hat{\alpha}(t), \mathfrak{K}_{x(t)}, \hat{f}(t,\cdot)), \quad t \geq t_0, \tag{41}$$

where the *adaptive gain* $\hat{\alpha}(t) : [t_0, \infty) \to \mathbb{R}^n$ verifies the *adaptive law*

$$\dot{\hat{\alpha}}(t) = -\Gamma_\alpha x(t) e^{\mathrm{T}}(t) P B, \quad \hat{\alpha}(t_0) = \hat{\alpha}_0, \tag{42}$$

and the *adaptive gain* $\hat{f}(t,\cdot) : [t_0, \infty) \to \mathcal{H}$ verifies the *ideal adaptive law*

$$\begin{aligned} \frac{\partial \hat{f}(t,\cdot)}{\partial t} &= \gamma_f \mathfrak{K}(\cdot, x(t)) e^{\mathrm{T}}(t) P B \\ &= \gamma_f E^*_{x(t)} B^{\mathrm{T}} P e(t), \qquad \hat{f}(t_0,\cdot) = \hat{f}_0(\cdot). \end{aligned} \tag{43}$$

In (42) and (43), $\Gamma_\alpha \in \mathbb{R}^{n\times n}$ is a user-defined, symmetric, and positive-definite matrix of adaptive rates, $x(\cdot)$ denotes the closed-loop plant trajectory, $e(\cdot)$ is defined in (6), and P denotes the symmetric and positive-definite solution of (14), and $\gamma_f > 0$ is a user-defined adaptive rate. It is worthwhile noting that (43) is a PDE, and, hence, the MRAC system given by (2) and (41)–(43) is considered nonparametric. The reason why (43) is referred to as an ideal adaptive law will become apparent in the second paper of this two-part work.

With the control input (41), it follows from (6) and (2) that the trajectory tracking error dynamics are captured by

$$\begin{aligned} \dot{e}(t) &= \dot{x}(t) - \dot{x}_r(t) \\ &= A_{\mathrm{ref}} e(t) - B\left(\tilde{\alpha}^{\mathrm{T}}(t) x(t) - \langle \mathfrak{K}_{x(t)}, \tilde{f}(t,\cdot)\rangle\right), \\ &\qquad e(t_0) = x_0 - x_{r,0}, \quad t \geq t_0. \end{aligned} \tag{44}$$

Collecting (44), (42), and (43), we can write the *limiting DPS error equations* in matrix operator form

$$\frac{\partial}{\partial t}\begin{Bmatrix} e(t) \\ \tilde{\alpha}(t) \\ \tilde{f}(t,\cdot) \end{Bmatrix} = \begin{bmatrix} A_{\mathrm{ref}} & -Bx^{\mathrm{T}}(t) & BE_{x(t)} \\ \Gamma_\alpha x(t) B^{\mathrm{T}} P & 0 & 0 \\ -\gamma_f \mathfrak{K}_{x(t)} B^{\mathrm{T}} P & 0 & 0 \end{bmatrix} \begin{Bmatrix} e(t) \\ \tilde{\alpha}(t) \\ \tilde{f}(t,\cdot) \end{Bmatrix}, \quad t \geq t_0, \tag{45}$$

with the same initial conditions as in (44), (42), and (43), where $\tilde{\alpha}(t) \triangleq \alpha - \hat{\alpha}(t)$ and $\tilde{f}(t,\cdot) \triangleq f - \hat{f}(t,\cdot)$. The limiting DPS error equations can be expressed also as

$$\frac{\partial}{\partial t}\begin{Bmatrix} e(t) \\ \tilde{\alpha}(t) \\ \tilde{f}(t,\cdot) \end{Bmatrix} = \begin{bmatrix} A_{\mathrm{ref}} & -Bx^{\mathrm{T}}(t) & BE_{x(t)} \\ \Gamma_\alpha x(t) B^{\mathrm{T}} P & 0 & 0 \\ -\gamma_f E^*_{x(t)} B^{\mathrm{T}} P & 0 & 0 \end{bmatrix} \begin{Bmatrix} e(t) \\ \tilde{\alpha}(t) \\ \tilde{f}(t,\cdot) \end{Bmatrix}. \tag{46}$$

Note that, similarly to (32), in (46), we do not have a matrix of real numbers but an operator since it contains the bounded linear operators $E_x : \mathcal{H} \to \mathbb{R}$ and $E^*_x : \mathbb{R} \to \mathcal{H}$. Remarkably, the state of (45), or, equivalently, (46), can be partitioned into three components. The first two components, $e(\cdot)$ and $\tilde{\alpha}(\cdot)$, reside in the finite-dimensional space $\mathbb{R}^n$ and the third component, $\tilde{f}(\cdot,\cdot)$, resides in the infinite-dimensional space $\mathcal{H}$. Thus, the closed-loop plant dynamics are still governed by an ODE. However, due to the coupling among closed-loop plant dynamics and adaptive gain dynamics, their stability properties need to be addressed at unison and using stability theory for DPS systems.

The next theorem proves the effectiveness of the proposed nonparametric MRAC approach. Specifically, the next result proves the boundedness of (46) and the convergence of the trajectory tracking error $e(\cdot)$ despite the uncertainties in A and f. For the statement of this theorem, recall that, by assumption, the scalar-valued reference command input $r(\cdot)$ is bounded on $[t_0, \infty)$, that is, $r \in L^\infty\left([t_0,\infty), \mathbb{R}\right)$.

Before we proceed to the analysis of the limiting DPS, we note one technical issue that arises when restricting attention to kernels that are bounded on the diagonal. This paper always chooses the kernel $\mathfrak{K}$ to be bounded on the diagonal over $\mathbb{X} \triangleq \mathbb{R}^n$ to simplify some proofs of well-posedness, convergence, and stability. However, as noted above, the native spaces induced by such kernels contain only functions bounded on $\mathbb{R}^n$. This would eliminate the study of many common systems for which f is unbounded over $\mathbb{X}$, such as polynomial uncertainties. In such cases, we eventually restrict the analysis to arbitrarily large compact sets over which the kernel is bounded on the diagonal. Such a restriction is not particularly constraining since A_{ref} is Hurwitz and $r(\cdot)$ is uniformly bounded, it follows from (2) that the reference trajectory is $x_r(\cdot)$ is bounded. Hence, in practice, it is possible to find a compact subset of $\mathbb{X}$ containing both $x_r(\cdot)$ and $x(\cdot)$ at all times. For now, we assume at the outset that $\mathfrak{K} : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ is bounded on the diagonal over all of $\mathbb{X}$.

Box 12: The proposed framework relies on Lyapunov-like arguments to certify the boundedness of the trajectory tracking error and of the adaptive gains at all times. Barbalat's lemma is employed to prove the asymptotic convergence of the trajectory tracking error to zero. Since the adaptive laws evolve over a DPS, despite the arguments underlying the Lyapunov analysis of classical adaptive control systems, forward completeness of the adaptive control system does not follow from arguments related to the compactness of level sets of the Lyapunov function.

Theorem 5.1. *Consider the limiting DPS (46), suppose that the kernel $\mathfrak{K}(\cdot,\cdot)$ that defines the native space $\mathcal{H}$ is bounded on the diagonal, and (46) is forward complete. Then, the trajectory of the limiting error DPS is uniformly bounded, and*

$$\lim_{t\to\infty} e(t) = 0$$

uniformly in $t_0 \in [0, \infty)$.

Proof. Consider the Lyapunov function candidate

$$\begin{aligned} V\left(e, \tilde{\alpha}, \tilde{f}\right) &\triangleq \langle Pe, e\rangle_{\mathbb{R}^n} + \left\langle \Gamma_\alpha^{-1}\tilde{\alpha}, \tilde{\alpha}\right\rangle_{\mathbb{R}} + \gamma_f^{-1}\left\langle \tilde{f}, \tilde{f}\right\rangle_{\mathcal{H}}, \\ &\qquad \left(e, \tilde{\alpha}, \tilde{f}\right) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{H}. \end{aligned} \tag{47}$$

Before we differentiate $V(\cdot,\cdot,\cdot)$ and apply Lyapunov-like arguments to the trajectories of (46), we make several observations. Since $\tilde{\alpha}(t) = \alpha - \hat{\alpha}(t)$, $t \geq t_0$, and α is constant, $\dot{\tilde{\alpha}}(t) = -\dot{\hat{\alpha}}(t)$. Per definition, $\tilde{f}(t,\cdot) = f - \hat{f}(t,\cdot)$, $t \geq t_0$, f is constant, and $\hat{f}(t,\cdot)$ is computed as the solution of (43), whose right-hand side is continuous. Indeed, if $t \mapsto x(t)$ is continuous on $[t_0,\infty)$, then it follows from (23) and the boundedness on the diagonal of $\mathfrak{K}(\cdot,\cdot)$ that $t \mapsto E_{x(t)}$ is continuous since E_x is uniformly bounded and linear. Furthermore, E_x^* is uniformly bounded and linear, and, hence, if $t \mapsto x(t)$ is continuous on $[t_0,\infty)$, then $t \mapsto E^*_{x(t)}$ is continuous. Furthermore, it follows from (6) that $e(t) = x(t) - x_r(t)$, $t \geq t_0$. Now, $t \mapsto x_r(t)$ is continuous on $[t_0,\infty)$ since the reference trajectory is the Carathéodory (hence, absolutely continuous) solution of (2), whose right-hand side is continuous. Thus, if $t \mapsto x(t)$ is continuous on $[t_0,\infty)$, then $t \mapsto e(t)$ is continuous. Finally, $t \mapsto x(t)$ is computed as the Carathéodory continuous solution of (38) with control input (41). Thus, the right-hand side of (43) is continuous on $[t_0,\infty)$, $\tilde{f} \in C^1([t_0,\infty),\mathcal{H})$, and $\dot{\tilde{f}}(t,\cdot) = -\dot{\hat{f}}(t,\cdot)$.

Differentiating (47) with respect to time, it follows from (28) that

$$\begin{aligned}\dot{V}\left(e(t),\tilde{\alpha}(t),\tilde{f}(t,\cdot)\right) &= \langle P\dot{e}(t), e(t)\rangle_{\mathbb{R}^n} + \langle Pe(t), \dot{e}(t)\rangle_{\mathbb{R}^n} \\ &\quad + 2\left(\left\langle \Gamma_\alpha^{-1}\tilde{\alpha}^{\mathrm{T}}(t), \dot{\tilde{\alpha}}(t)\right\rangle_{\mathbb{R}} + \gamma_f^{-1}\left\langle \tilde{f}(t,\cdot), \dot{\tilde{f}}(t,\cdot)\right\rangle_{\mathcal{H}}\right) \\ &= -\langle Qe(t), e(t)\rangle_{\mathbb{R}^n} \\ &\quad - 2\left\langle \tilde{\alpha}(t), x(t)e^{\mathrm{T}}(t)PB + \Gamma_\alpha^{-1}\dot{\hat{\alpha}}(t)\right\rangle_{\mathbb{R}} \\ &\quad + 2\left\langle \tilde{f}(t,\cdot), \mathfrak{K}_{x(t)}e^{\mathrm{T}}(t)PB - \gamma_f^{-1}\dot{\hat{f}}(t,\cdot)\right\rangle_{\mathcal{H}},\end{aligned} \tag{48}$$

for all $t \geq t_0$. Thus, along the trajectories of (46), it holds that

$$\dot{V}\left(e(t),\tilde{\alpha}(t),\tilde{f}(t,\cdot)\right) = -e^{\mathrm{T}}Qe(t) \leq 0, \quad t \geq t_0, \tag{49}$$

and, consequently, the trajectories of (46) are bounded over $[t_0,\infty)$ uniformly in $t_0 \geq 0$.

Since $V\left(e(t),\tilde{\alpha}(t),\tilde{f}(t,\cdot)\right)$, $t \in [t_0,\infty)$, is non-increasing, nonnegative, and bounded from below, we deduce that

$$\begin{aligned} &e \in L^\infty\left([t_0,\infty),\mathbb{R}^n\right), \quad \tilde{\alpha} \in L^\infty\left([t_0,\infty),\mathbb{R}^n\right), \\ &\tilde{\beta} \in L^\infty\left([t_0,\infty),\mathbb{R}\right), \quad \tilde{f} \in L^\infty\left([t_0,\infty),\mathcal{H}\right).\end{aligned}$$

Furthermore, by the Weierstrass theorem, the limit $V_\infty \triangleq \lim_{t\to\infty} V(t)$ exists and is finite. Thus, by integrating $\dot{V}\left(e(t),\tilde{\alpha}(t),\tilde{f}(t,\cdot)\right)$, and using the finite limit V_∞, we conclude that $e \in L^2\left([t_0,\infty),\mathbb{R}^n\right)$.

Next, we want to show that $\dot{e} \in L^\infty\left([t_0,\infty),\mathbb{R}^n\right)$. It follows from (44) that

$$\begin{aligned}\|\dot{e}(t)\|_{\mathbb{X}} &\leq \|A_{\mathrm{ref}}\|\|e\|_{L^\infty([t_0,\infty),\mathbb{R}^n)} + \|B\|\|e\|_{L^\infty([t_0,\infty),\mathbb{R}^n)}\|\tilde{\alpha}\|_{L^\infty([t_0,\infty),\mathbb{R}^n)} \\ &\quad + \|r\|_{L^\infty([t_0,\infty),\mathbb{R}^n)}\|\tilde{\beta}\|_{L^\infty([t_0,\infty),\mathbb{R}^n)} + \|B\|\|E_{x(t)}\| \underbrace{\|\tilde{f}(t,\cdot)\|_{\mathcal{H}}}_{\leq \|\tilde{f}(t,\cdot)\|_{L^\infty([t_0,\infty),\mathcal{H})}},\end{aligned} \tag{50}$$

for all $t \in [t_0,\infty)$. Now, since the kernel function is bounded on the diagonal, it follows from (24) that $\|E_{x(t)}\| < \bar{\mathfrak{K}} < \infty$, and, hence, $\dot{e} \in L^\infty\left([t_0,\infty),\mathbb{R}^n\right)$. Since $e \in L^\infty\left([t_0,\infty),\mathbb{R}^n\right) \cap L^2\left([t_0,\infty),\mathbb{R}^n\right)$, Barbalat's lemma implies that $\lim_{t\to\infty} e(t) = 0$ uniformly in $t_0 \geq 0$.

Several remarks on Theorem 5.1 and its proof are in order.

Remark 5.1. Since $\tilde{f}(t,\cdot) \in C^1([t_0,\infty),\mathcal{H})$, the composition rules on Fréchet derivatives ensure the identity

$$\frac{d}{dt}\left(\langle \tilde{f}(t,\cdot), \tilde{f}(t,\cdot)\rangle_{\mathcal{H}}\right) = 2\left\langle \tilde{f}(t,\cdot), \dot{\tilde{f}}(t,\cdot)\right\rangle_{\mathcal{H}}, \quad t \geq t_0, \tag{51}$$

which enabled (48).

Remark 5.2. Theorem 5.1 establishes that the DPS (46) behaves in a way that is analogous to the real parametric case captured by (10), (12), and (13) or, equivalently, by

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} e_N(t) \\ \tilde{\alpha}(t) \\ \tilde{\Theta}(t)\end{bmatrix} = \begin{bmatrix} A_{\mathrm{ref}} & -Bx(t) & B\Phi_N^{\mathrm{T}}(x_N(t)) \\ \Gamma_\alpha x_N(t)B^{\mathrm{T}}P & 0 & 0 \\ -\gamma\Phi_N(x_N(t))B^{\mathrm{T}}P & 0 & 0\end{bmatrix}\begin{bmatrix} e_N(t) \\ \tilde{\alpha}(t) \\ \tilde{\Theta}(t)\end{bmatrix},$$

$$\begin{bmatrix} e_N(t_0) \\ \tilde{\alpha}(t_0) \\ \tilde{\Theta}(t_0)\end{bmatrix} = \begin{bmatrix} x_0 - x_{r,0} \\ \tilde{\alpha}_0 \\ \tilde{\Theta}_0\end{bmatrix}, \quad t \geq t_0, \tag{52}$$

assuming *zero external disturbance*; for additional details, see Section 3.2.

Remark 5.3. A key passage in the proof of Theorem 5.1 is that the dual operator E_x^* can be understood as the operator given by multiplication by the kernel section $\mathfrak{K}_x \in \mathcal{H}$, that is, $E_x^*\alpha \triangleq \mathfrak{K}_x\alpha$ for all $\alpha \in \mathbb{R}$. This identity is crucial for approximations and implementations.

Theorem 5.1 relied on the key assumption whereby (46) is forward complete. In the next section, we provide sufficient conditions for this assumption to be verified.

5.2. Well-posedness of the limiting DPS

The Lyapunov analysis presented in Section 5.1 does not address the question of whether the limiting closed-loop control DPS (46) is well-posed, that is, solutions of (46) exist for all $t \in [t_0,\infty)$. Theorem 5.1 assumes that

$$t \mapsto z(t) \triangleq \left(e(t),\tilde{\alpha}(t),\tilde{f}(t,\cdot)\right) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{H} \triangleq \mathbb{Z} \tag{53}$$

exists for all $t \in [t_0,\infty)$. In classical MRAC for parametric systems, the well-posedness of solutions of the trajectory tracking error dynamics and of the adaptive laws follows from the Lipschitz continuity of the ODEs, which assures the existence of unique solutions, and the proof of definition of Lyapunov stability of ODEs, which implies the existence of these solutions for all times; for details, see Lavretsky and Wise (2012, Ch. 9). However, (46) is not an ODE, and the mechanisms that assure the well-posedness of ODEs do not apply.

In this section, we describe conditions that are sufficient to ensure the existence and uniqueness of solutions to the DPS (46) that describes the error dynamics. To address this point, we rewrite (46) as

$$\begin{aligned}\underbrace{\frac{\mathrm{d}}{\mathrm{d}t}\begin{Bmatrix} e(t) \\ \tilde{\alpha}(t) \\ \tilde{f}(t,\cdot)\end{Bmatrix}}_{\dot{z}(t)} &= \underbrace{\begin{bmatrix} A_{\mathrm{ref}} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0\end{bmatrix}\begin{Bmatrix} e(t) \\ \tilde{\alpha}(t) \\ \tilde{f}(t,\cdot)\end{Bmatrix}}_{\mathcal{A}z(t)} \\ &\quad + \underbrace{\begin{bmatrix} 0 & -B\left(e(t)+x_r(t)\right)^{\mathrm{T}} & BE_{(e(t)+x_r(t))} \\ \Gamma_\alpha x(t)B^{\mathrm{T}}P & 0 & 0 \\ -\gamma_f E^*_{(e(t)+x_r(t))}B^{\mathrm{T}}P & 0 & 0\end{bmatrix}\begin{Bmatrix} e(t) \\ \tilde{\alpha}(t) \\ \tilde{f}(t,\cdot)\end{Bmatrix}}_{\mathcal{G}(t,z(t))}\end{aligned} \tag{54}$$

with the same boundary conditions as in (46). With these definitions, the feedback control problem is an example of a nonlinear initial value problem, one that is described by a DPS in the state $z(\cdot)$ given by (53), and (54) takes the compact form

$$\dot{z}(t) = \mathcal{A}z(t) + \mathcal{G}(t,z(t)), \quad z\left(t_0\right) = z_0, \quad t \geq t_0. \tag{55}$$

The following theorem states sufficient conditions that ensure the existence, uniqueness, and continuous dependence of solutions of (55).

> **Box 13**: Forward completeness of the trajectory tracking error and adaptive gain dynamics follows as a consequence of boundedness on the diagonal of the admissible kernels underlying the RKHS and the embedding of this kernel into the space of Lipschitz continuous functions.

Theorem 5.2 (*Well-Posedness, Forward Completeness of the DPS*). *Consider the limiting error DPS given by (55) and the reference model given by (2). Suppose that the kernel $\mathfrak{K}(\cdot,\cdot)$ that defines the native space $\mathcal{H}$ is bounded on the diagonal, and that $\mathcal{H}$ is continuously embedded in the space of Lipschitz continuous functions $C^{0,1}(\mathbb{X})$, that is, $\mathcal{H} \hookrightarrow C^{0,1}(\mathbb{X})$. Then, for any $z_0 \in \mathbb{Z}$, (55) is forward complete with $z \in C^1([t_0,\infty),\mathbb{Z})$, that is, $t \mapsto z(t)$ is defined on $[t_0,\infty)$.*

Proof. The proof of the existence and uniqueness of a local solution follows by using Theorem 1.4 of Chapter 6 of Pazy (2012) for nonlinear systems having a nonlinear part that satisfies a local Lipschitz condition. This theorem requires (a) that the map $t \mapsto \mathcal{G}(t,z)$ is continuous in time for each fixed $z \in \mathbb{Z}$, and (b) that the map $z \to \mathcal{G}(t,z)$ is locally Lipschitz continuous, uniformly over bounded intervals of time. To verify these two conditions, consider

$$\mathcal{G}(t,z) = \begin{bmatrix} -B\tilde{\alpha}^{\mathrm{T}}\left(e+x_r(t)\right) - \tilde{\beta}r(t) + BE_{e+x_r(t)}\tilde{f}(t,\cdot) \\ \Gamma_\alpha\left(e+x_r(t)\right)e^{\mathrm{T}}PB \\ -\gamma_f\mathfrak{K}_{(e+x_r(t))}e^{\mathrm{T}}PB \end{bmatrix}.$$

By assumption, we know the reference command input $t \mapsto r(t)$ is continuous and uniformly bounded for all $t \in [t_0,\infty)$, and, since A_{ref} is Hurwitz, $t \mapsto x_r(t)$ is continuous and uniformly bounded. Next, since the kernel $\mathfrak{K}(\cdot,\cdot)$ is chosen such that $\mathcal{H} \hookrightarrow C^{0,1}(\mathbb{X}) \hookrightarrow C(\mathbb{X})$, the function $\tilde{f} \in \mathcal{H}$ is also continuous. Thus, the first two entries of $\mathcal{G}(t,z)$ are continuous in t for fixed $z \in \mathbb{Z}$.

To verify the continuity in time of the third component of $\mathcal{G}(t,z)$, let $t_\ell \in (t_0,\infty)$ be such that $t_\ell \to t$ as $\ell \to \infty$, and note that

$$\begin{aligned}
&\|\gamma_f\mathfrak{K}_{(e+x_r(t))}e^{\mathrm{T}}PB - \gamma_f\mathfrak{K}_{(e+x_r(t_\ell))}e^{\mathrm{T}}PB\,\|_{\mathcal{H}} \\
&\quad\le |\gamma_f e^{\mathrm{T}}PB|\|\mathfrak{K}_{(e+x_r(t))} - \mathfrak{K}_{(e+x_r(t_\ell))}\|_{\mathcal{H}} \\
&\quad\le |\gamma_f e^{\mathrm{T}}PB|\left[\mathfrak{K}\left(e+x_r(t),e+x_r(t)\right) + \mathfrak{K}\left(e+x_r\left(t_\ell\right),e+x_r\left(t_\ell\right)\right)\right. \\
&\qquad\qquad \left. -2\mathfrak{K}\left(e+x_r(t),e+x_r\left(t_\ell\right)\right)\right]^{\frac{1}{2}} \quad \text{for all } t \ge t_0.
\end{aligned} \tag{56}$$

Thus, the continuity of both $x_r(\cdot)$ and the kernel $\mathfrak{K}(\cdot,\cdot)$ implies the continuity in time of the third entry of $\mathcal{G}(t,z)$ in time t.

Next, we prove that $\mathcal{G} : [t_0,\infty)\times\mathbb{Z} \to \mathbb{Z}$ is locally Lipschitz continuous in $\mathbb{Z}$, uniformly in time over bounded intervals. That is, for every $t \ge t_0$ and for some $R > 0$, there exists a Lipschitz constant $L_{t,R} > 0$ such that

$$\|\mathcal{G}(\tau,y) - \mathcal{G}(\tau,z)\| \le L_{t,R}\|y-z\|, \tag{57}$$

for all $y,z \in \mathcal{B}_R(0) \triangleq \left\{z \in \mathbb{Z} \mid \|z\|_{\mathbb{Z}} < R\right\}$ and $\tau \in \left[t_0,t\right]$. To this goal, we proceed component-wise and leverage the triangle inequality to verify (57). Specifically, let $y \triangleq \{\tilde{e},\tilde{a},\tilde{g}\} \in \mathcal{B}_R(0) \subset \mathbb{Z}$ and let

$$\mathcal{G}(\tau,z) = \{\mathcal{G}_1(\tau,z),\mathcal{G}_2(\tau,z),\mathcal{G}_3(\tau,z)\}^{\mathrm{T}}.$$

For all $y,z \in \mathcal{B}_R(0)$, it holds that

$$\begin{aligned}
&\|\,\mathcal{G}_2(\tau,z) - \mathcal{G}_2(t,y)\,\| \\
&\quad= \left\|\Gamma_\alpha\left(e+x_r(\tau)\right)e^{\mathrm{T}}PB - \Gamma_\alpha\left(\tilde{e}+x_r(\tau)\right)\tilde{e}^{\mathrm{T}}PB\right\| \\
&\quad\le \|\Gamma_\alpha\|\Big[\|\left(e+x_r(\tau)\right)e^{\mathrm{T}}PB - \left(e+x_r(\tau)\right)\tilde{e}^{\mathrm{T}}PB\| \\
&\qquad\qquad + \|\left(e+x_r(\tau)\right)\tilde{e}^{\mathrm{T}}PB - \left(\tilde{e}+x_r(\tau)\right)\tilde{e}^{\mathrm{T}}PB\|\Big] \\
&\quad\le \|\Gamma_\alpha\|\|B^{\mathrm{T}}P\|\left[\|e+x_r(\tau)\|\,\|e-\tilde{e}\| + \|\tilde{e}\|\,\|e-\tilde{e}\|\right] \\
&\quad\le \|\Gamma_\alpha\|\|PB\|\left[2R+\bar{x}_r\right]\|e-\tilde{e}\|,
\end{aligned}$$

where $\bar{x}_r \triangleq \max_{t\in[t_0,\infty)}\|x_r(t)\|$, which proves the local Lipschitz continuity of $\mathcal{G}_2(t,\cdot)$.

The analysis of $\mathcal{G}_3(t,\cdot)$ proceeds similarly. Let $y,z \in \mathcal{B}_R(0)$ and note that

$$\begin{aligned}
&\|\,\mathcal{G}_3(t,z) - \mathcal{G}_3(t,y)\,\|_{\mathcal{H}} \\
&\quad= \gamma_f\|\mathfrak{K}\left(e+x_r(\tau),\cdot\right)e^{\mathrm{T}}PA - \mathfrak{K}\left(\tilde{e}+x_r(\tau),\cdot\right)\tilde{e}^{\mathrm{T}}PA\|_{\mathcal{H}}, \\
&\quad\le \gamma_f\Big[\left\|\mathfrak{K}\left(e+x_r(\tau),\cdot\right)e^{\mathrm{T}}PA - \mathfrak{K}\left(e+x_r(\tau),\cdot\right)\tilde{e}^{\mathrm{T}}PA\right\|_{\mathcal{H}} \\
&\qquad\qquad \left\|\mathfrak{K}\left(e+x_r(\tau),\cdot\right)\tilde{e}^{\mathrm{T}}PA - \mathfrak{K}\left(\tilde{e}+x_r(\tau),\cdot\right)\tilde{e}^{\mathrm{T}}PA\right\|_{\mathcal{H}}\Big], \\
&\quad\le \gamma_f\|PB\|\Big[\underbrace{\|\mathfrak{K}\left(e+x_r(\tau),\cdot\right)\|_{\mathcal{H}}}_{\le\bar{\mathfrak{K}}}\|e-\tilde{e}\| \\
&\qquad\qquad + \underbrace{\|\tilde{e}\|}_{\le R}\left\|\mathfrak{K}\left(e+x_r(\tau),\cdot\right) - \mathfrak{K}\left(\tilde{e}+x_r(\tau),\cdot\right)\right\|_{\mathcal{H}}\Big]
\end{aligned}$$

for each $t \in [t_0,\infty)$. Now, note that

$$\begin{aligned}
&\|\,\mathfrak{K}\left(e+x_r(\tau),\cdot\right) - \mathfrak{K}\left(\tilde{e}+x_r(\tau),\cdot\right)\|_{\mathcal{H}}^2 \\
&\quad= \mathfrak{K}\left(e+x_r(\tau),e+x_r(\tau)\right) + \mathfrak{K}\left(\tilde{e}+x_r(\tau),\tilde{e}+x_r(\tau)\right) \\
&\qquad - 2\mathfrak{K}\left(e+x_r(\tau),\tilde{e}+x_r(\tau)\right) \\
&\quad\le |\mathfrak{K}\left(e+x_r(\tau),e+x_r(\tau)\right) - \mathfrak{K}\left(e+x_r(\tau),\tilde{e}+x_r(\tau)\right)| \\
&\qquad\quad + |\mathfrak{K}\left(\tilde{e}+x_r(\tau),\tilde{e}+x_r(\tau)\right) - \mathfrak{K}\left(e+x_r(\tau),\tilde{e}+x_r(\tau)\right)| \\
&\quad\le \|\mathfrak{K}_{e+x_r(\tau)}\|_{C^{0,1}(\mathbb{X})}\|e-\tilde{e}\| + \|\mathfrak{K}_{\tilde{e}+x_r(\tau)}\|_{C^{0,1}(\mathbb{X})}\|e-\tilde{e}\| \\
&\quad\le 2\tilde{C}\bar{\mathfrak{K}}\|e-\tilde{e}\|
\end{aligned}$$

for each $t \in [t_0,\infty)$, where the constant $\tilde{C} > 0$ characterizes the space of Lipschitz continuous functions given by the embedding $\mathcal{H} \hookrightarrow C^{0,1}(\mathbb{X})$.

The Lipschitz continuity of $\mathcal{G}_1(t,\cdot)$ follows as in the analysis of the Lipschitz continuity of $\mathcal{G}_2(t,\cdot)$ and $\mathcal{G}_3(t,\cdot)$, and most of its details are omitted for brevity. This analysis would ultimately show that

$$|\mathcal{G}_1(\tau,y) - \mathcal{G}_1(\tau,z)| \le \|B\|\left(\left(R+\bar{x}_r\right)\|\tilde{\alpha}-\tilde{a}\| + R\left(1+\tilde{C}\right)\|e-\tilde{e}\| + \|\tilde{f}-\tilde{g}\|_{\mathcal{H}}\right)$$

for each $t \in [t_0,\infty)$ and for all $y,z \in \mathcal{B}_R(0) \subset \mathbb{Z}$.

Having proven continuity of $\mathcal{G}(\cdot,\cdot)$ in its first argument and local Lipschitz continuity in its second argument, Theorem 1.4 in Chapter 6 of Pazy (2012) now guarantees that there is a unique mild solution over $[t_0,\infty)$ or a compact subset thereof containing t_0.

By proceeding as in the proof of Theorem 5.1, we can prove that $\tilde{f} \in C^1([t_0,T_{\max}),\mathcal{H})$ and solutions of (55) are bounded on $[t_0,T_{\max})$. Thus, we conclude that $T_{\max} = \infty$, and the mild solution $z \in C([t_0,\infty),\mathbb{Z})$. Finally, we can use Theorem 5 in Chapter 3 of Pazy (2012) to establish forward completeness in $C^1([t_0,\infty),\mathbb{Z})$. Since $\mathcal{A}$ is a bounded linear operator, it generates a C^0 semigroup.

Theorem 5.2 concludes this first paper of a two-part work. Together with Theorem 5.1, this result characterizes key properties of the limiting DPS (55) that is the foundation of the proposed nonparametric DPS framework. These properties are the existence and uniqueness of solutions to the trajectory tracking error dynamics and adaptive gain dynamics at all times, boundedness of these solutions, and asymptotic convergence of the trajectory error dynamics to zero. Whereas the proposed solution appears enticing, it does not apply to problems of practical interest since the adaptive law (43), and, hence, the coupled system given by the plant dynamics (1) with control input (41), the adaptive law (42), and (43) form a DPS. The next paper will address the key problem of approximating such a DPS. A particular aspect of this problem lies in the fact that, in principle, applying approximations of the limiting DPS, and hence, approximating the space of functional uncertainties $\mathcal{H}$, asymptotic convergence of the tracking error to zero may not be recovered. It is worthwhile remarking on how, if we set $x_r(t_0) = 0$ and $r(t) \equiv 0$ for all $t \ge t_0$, then the proposed results allow addressing the problem of stabilizing the solution of the nonlinear plant model (1) to zero.

6. Conclusion

This first paper of a two-part work introduced a novel control framework called nonparametric adaptive control. According to this framework, nonlinear uncertainties are not parameterized by a regressor vector, which is provided *a priori* or constructed in real-time, as it occurs in existing parametric adaptive control frameworks. Assuming that nonlinear uncertainties are elements of native spaces, the proposed framework allows controlling the plant dynamics without any explicit characterization of the unknown terms.

The proposed framework stems from a DPS that, by its nature, is not implementable in finite dimensions. The second part of this work provides several alternative solutions, all inspired by classical robust adaptive control techniques, to approximate the DPS underlying the proposed framework. The use of RKHSs and, in particular, their numerous tools to estimate approximation errors, will be essential to capture the performance of these approximations of the proposed DPS, whose behavior can be considered as a limiting case of a series of finite-dimensional approximations parameterized by the size N of the approximating space.

Worthy of mention is that the proposed approach, which extended classical, parametric, MRAC theory to a nonparametric setting, can be extended to other control techniques, such as robust control Lyapunov functions and backstepping control. These results, which are beyond the scope of this work, can be found in Wang, Scurlock, Kurdila, and L'Afflitto (2024) and Kurdila et al. (2025, Ch. 5).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was in part performed with the support of National Science Foundation through the Grant no. 2137159 and the US Army Research Lab under Grant no. W911QX2320001

References

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, *68*, 337–404. http://dx.doi.org/10.1090/S0002-9947-1950-0051437-7.

Berlinet, A., & Thomas-Agnan, C. (2011). *Reproducing kernel Hilbert spaces in probability and statistics*. Springer.

Boffi, N. M., Tu, S., & Slotine, J.-J. E. (2022). Nonparametric adaptive control and prediction: Theory and randomized algorithms. *Journal of Machine Learning Research*, *23*(1), 12841–12886.

Choi, J., & Farrell, J. (2000). Nonlinear adaptive control using networks of piecewise linear approximators. *IEEE Transactions on Neural Networks*, *11*(2), 390–401. http://dx.doi.org/10.1109/72.839009.

Chowdhary, G., How, J., & Kingravi, H. (2012). Model reference adaptive control using nonparametric adaptive elements. In *AIAA guidance, navigation, and control conference* (p. 5038). Minneapolis, MN.

Chowdhary, G., Kingravi, H. A., How, J. P., & Vela, P. A. (2015). Bayesian nonparametric adaptive control using Gaussian processes. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(3), 537–550. http://dx.doi.org/10.1109/TNNLS.2014.2319052.

Cucker, F., & Zhou, D. X. (2007). *Learning Theory: An approximation theory viewpoint*. Cambridge Press.

Daubechies, I., DeVore, R., Foucart, S., Hanin, B., & Petrova, G. (2022). Nonlinear approximation and (deep) ReLU networks. *Constructive Approximation*, *55*(1), 127–172. http://dx.doi.org/10.1007/s00365-021-09548-z.

DeVore, R. A. (1998). Nonlinear approximation. *Acta Numerica*, *7*, 51–150. http://dx.doi.org/10.1017/S0962492900002816.

DeVore, R., Hanin, B., & Petrova, G. (2021). Neural network approximation. *Acta Numerica*, *30*, 327–444. http://dx.doi.org/10.1017/S0962492921000052.

Farrell, J. (1998). Stability and approximator convergence in nonparametric nonlinear adaptive control. *IEEE Transactions on Neural Networks*, *9*(5), 1008–1020. http://dx.doi.org/10.1109/72.712182.

Farrell, J. A., & Polycarpou, M. M. (2006). *vol. 48, Adaptive approximation based control: Unifying neural, fuzzy and traditional adaptive approximation approaches*. John Wiley & Sons.

Glasov, V., Zybin, E., & Kosyanchuk, V. (2019). Nonparametric method for aircraft flight control. *vol. 476*, In *IOP conference series: Materials science and engineering*. (1), IOP Publishing, Article 012011.

Ioannou, P., & Fidan, B. (2006). *Adaptive control tutorial*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Ioannou, P. A., & Sun, J. (2012). *Robust adaptive control*. Courier Corporation.

Joshi, G., & Chowdhary, G. (2019). Deep model reference adaptive control. In *IEEEConference on decision and control* (pp. 4601–4608). http://dx.doi.org/10.1109/CDC40024.2019.9029173.

Joshi, G., Virdi, J., & Chowdhary, G. (2021). Asynchronous deep model reference adaptive control. In *Conference on robot learning* (pp. 984–1000). PMLR.

Kamalapurkar, R., Rosenfeld, J. A., & Dixon, W. E. (2015). State following (StaF) kernel functions for function approximation part II: Adaptive dynamic programming. In *American control conference* (pp. 521–526). IEEE, http://dx.doi.org/10.1109/ACC.2015.7170788.

Kingravi, H. A., Chowdhary, G., Vela, P. A., & Johnson, E. N. (2012). Reproducing kernel Hilbert space approach for the online update of radial bases in neuro-adaptive control. *IEEE Transactions on Neural Networks and Learning Systems*, *23*(7), 1130–1141. http://dx.doi.org/10.1109/TNNLS.2012.2198889.

Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., et al. (2018). Data-driven model reduction and transfer operator approximation. *Journal of Nonlinear Science*, [ISSN: 1432-1467] *28*(3), 985–1010. http://dx.doi.org/10.1007/s00332-017-9437-7.

Klus, S., Nüske, F., Peitz, S., Niemann, J.-H., Clementi, C., & Schütte, C. (2020). Data-driven approximation of the koopman generator: Model reduction, system identification, and control. *Physica D: Nonlinear Phenomena*, *406*, Article 132416. http://dx.doi.org/10.1016/j.physd.2020.132416.

Krstic, M., Kanellakopoulos, I., & Kokotovic, P. (1995). *Nonlinear and adaptive control design*. John Wiley & Sons.

Kurdila, A. J., L'Afflitto, A., & Burns, J. A. (2025). *Nonparametric adaptive control of nonlinear systems over native spaces*. London, UK: Springer, In press.

Lavretsky, E., & Wise, K. A. (2012). Robust adaptive control. In *Robust and adaptive control: With aerospace applications* (pp. 317–353). Springer.

Le, D. M., Greene, M. L., Makumi, W. A., & Dixon, W. E. (2021). Real-time modular deep neural network-based adaptive control of nonlinear systems. *IEEE Control Systems Letters*, *6*, 476–481. http://dx.doi.org/10.1109/LCSYS.2021.3081361.

Mauroy, A., Susuki, Y., & Mezić, I. (2020). *Koopman operator in systems and control*. Springer.

Medvedev, A. (2013a). Theory of nonparametric systems. Control-I. *Siberian Aerospace Journal*, *14*(2), 57–63.

Medvedev, A. (2013b). Theory of nonparametric systems. Control-II. *Siberian Aerospace Journal*, *14*(3), 85–90.

Morgan, A., & Narendra, K. (1977a). On the uniform asymptotic stability of certain linear nonautonomous differential equations. *SIAM Journal on Control and Optimization*, *15*(1), 5–24. http://dx.doi.org/10.1137/0315002.

Morgan, A., & Narendra, K. (1977b). On the stability of nonautonomous differential equations $\dot{x} = A + B(t)x$, with skew-symmetric matrix $B(t)$. *SIAM Journal on Control and Optimization*, *15*(1), 163–176. http://dx.doi.org/10.1137/0315013.

Narendra, K. S., & Annaswamy, A. M. (1989). *Stable adaptive systems*. Dover.

Park, J., & Sandberg, I. W. (1991). Universal approximation using radial-basis-function networks. *Neural Computation*, *3*(2), 246–257. http://dx.doi.org/10.1162/neco.1991.3.2.246.

Paruchuri, S. T., Guo, J., & Kurdila, A. (2020). Reproducing kernel Hilbert space embedding for adaptive estimation of nonlinearities in piezoelectric systems. *Nonlinear Dynamics*, *101*(2), 1397–1415. http://dx.doi.org/10.1007/s11071-020-05812-2.

Paruchuri, S. T., Guo, J., & Kurdila, A. (2022). Kernel center adaptation in the reproducing kernel Hilbert space embedding method. *International Journal of Adaptive Control and Signal Processing*, *36*(7), 1562–1583. http://dx.doi.org/10.1002/acs.3407.

Paruchuri, S. T., Guo, J., & Kurdila, A. (2023). Sufficient conditions for parameter convergence over embedded manifolds using kernel techniques. *IEEE Transactions on Automatic Control*, *68*(2), 753–765. http://dx.doi.org/10.1109/TAC.2022.3148716.

Patil, O. S., Le, D. M., Greene, M. L., & Dixon, W. E. (2021). Lyapunov-derived control and adaptive update laws for inner and outer layer weights of a deep neural network. *IEEE Control Systems Letters*, *6*, 1855–1860. http://dx.doi.org/10.1109/LCSYS.2021.3134914.

Paulsen, V. I., & Raghupathi, M. (2016). *vol. 152, An introduction to the theory of reproducing kernel Hilbert spaces*. Cambridge University Press.

Pazy, A. (2012). *vol. 44, Semigroups of linear operators and applications to partial differential equations*. Springer Science & Business Media.

Powell, N., Kurdila, A. J., L'Afflitto, A., Wang, H., & Guo, J. (2023). Newton Bases and Event-Triggered Adaptive Control in Native Spaces. *Proceedings of the 2023 Conference on Decision and Control, submitted*.

Powell, N., Kurdila, A. J., L'Afflitto, A., Wang, H., & Guo, J. (2024). Newton bases and event-triggered adaptive control in native spaces. In *2024 American control conference* (pp. 1586–1591). http://dx.doi.org/10.23919/ACC60939.2024.10644218.

Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning* (pp. 63–71). Springer.

Rosenfeld, J. A., Kamalapurkar, R., & Dixon, W. E. (2019). The state following approximation method. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(6), 1716–1730. http://dx.doi.org/10.1109/TNNLS.2018.2870040.

Saitoh, S., & Sawano, Y. (2016). *Theory of reproducing kernels and applications*. Springer.

Sastry, S., & Bodson, M. (2011). *Adaptive control: Stability, convergence and robustness*. Courier Corporation.

Slotine, J.-J. E., & Li, W. (1991). *vol. 199, Applied nonlinear control*. (1), Englewood Cliffs, NJ: Prentice Hall.

Smola, B. S. A. J. (2002). *Learning with Kernels: Support vector machines, regularization, optimization, and beyond*. MIT Press.

Tao, G. (2003). *vol. 37, Adaptive control design and analysis*. John Wiley & Sons.

Temlyakov, V. (2011). *vol. 20, Greedy approximation*. Cambridge, UK: Cambridge University Press.

Vapnik, V. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, *10*(5), 988–999. http://dx.doi.org/10.1109/72.788640.

Wang, H., Kurdila, A. J., L'Afflitto, A., Oesterheld, D., & Daniel J. Stilwell, M. (2024). Robust Model Reference Adaptive Control Based on Reproducing Kernel Hilbert Spaces. *Journal of Adaptive Control and Signal Processing*, Accepted.

Wang, H., Scurlock, B., Kurdila, A., & L'Afflitto, A. (2024). Robust, nonparametric backstepping control over reproducing kernel Hilbert spaces. In *Conference on decision and control*. Milan, Italy.

Wendland, H. (2004). *vol. 17*, *Scattered data approximation*. Cambridge University Press.

Williams, C. K., & Rasmussen, C. E. (2006). *vol. 2*, *Gaussian processes for machine learning*. (3), MIT press Cambridge, MA.

Yarotsky, D. (2017). Error bounds for approximations with deep ReLU networks. *Neural Networks*, *94*, 103–114. http://dx.doi.org/10.1016/j.neunet.2017.07.002.