Received: 15 June 2021 | Revised: 21 March 2022 | Accepted: 7 May 2022

DOI: 10.1002/acs.3442

RESEARCH ARTICLE

WILEY

Adaptive estimation of external fields in reproducing kernel Hilbert spaces

Jia Guo[1] | **Michael E. Kepler**[2] | **Sai Tej Paruchuri**[1] | **Hoaran Wang**[1] | **Andrew J. Kurdila**[1] | **Daniel J. Stilwell**[2]

[1]Department of Mechanical Engineering, Virginia Tech, Blacksburg, Virginia, USA

[2]Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, Virginia, USA

Correspondence
Michael E. Kepler, Department of Electrical and Computer Engineering, Virginia Tech, 1185 Perry St., Blacksburg, VA 24060, USA.
Email: mkepler@vt.edu

Funding information
Office of Naval Research, Grant/Award Numbers: N00014-18-1-2627, N00014-19-1-2194

Summary

This article studies the distributed parameter system that governs adaptive estimation by mobile sensor networks of external fields in a reproducing kernel Hilbert space (RKHS). The article begins with the derivation of conditions that guarantee the well-posedness of the ideal, infinite dimensional governing equations of evolution for the centralized estimation scheme. Subsequently, convergence of finite dimensional approximations is studied. Rates of convergence in all formulations are established using history-dependent bases defined from translates of the RKHS kernel that are centered at sample points along the agent trajectories. Sufficient conditions are derived that ensure that the finite dimensional approximations of the ideal estimator equations converge at a rate that is bounded by the fill distance of samples in the agents' assigned subdomains. The article concludes with examples of simulations and experiments that illustrate the qualitative performance of the introduced algorithms.

KEYWORDS

distributed parameter systems, function estimation, learning systems, reproducing kernel Hilbert spaces

1 | INTRODUCTION

In this article we propose an infinite-dimensional, continuous-time technique for the adaptive estimation of external fields $g : \mathcal{X} \to \mathbb{R}$ with $\mathcal{X} \subset \mathbb{R}^d$. Here we use the term external field to describe a spatially varying function over a domain traversed by an estimating agent. Typical examples include unknown environments such as the depth map of a lake or a spatial field representing pollutant concentrations. In order to reconstruct and monitor the unknown external field, a team of sensing vehicles is deployed to take measurements at discrete sample locations, from which the continuous map is estimated.

An extensive body of literature exists that solves this problem using Gaussian process (GP) regression. The sparse online Gaussian process (SOGP) method proposed by Csató and Opper is an early attempt to overcome the high computational cost of GP regression as the number of samples grows.[1] Recent research in this area focuses on the scheme in which the samples are taken sequentially. Various approaches are incorporated to improve the efficiency of exploration, such as Bayesian optimization,[2] gradient-based methods,[3] and information-based methods.[4] Another approach for reducing the computation expense of GP regression is to leverage the multi-agent system and compute the regression in the distributed manner.[5-7]

The methods above are formulated using machine learning theory or Bayesian estimation, and they are most commonly understood in terms of regression in discrete time where samples are obtained from an independent and identically distributed (IID) random process in discrete time. In this article we view the problem from the perspective of online adaptive estimation in continuous time. The governing estimation problem is cast as a distributed parameter system (DPS) in an infinite dimensional state space. New results are derived for the adaptive estimation by a team of N agents of such external fields in a reproducing kernel Hilbert space (RKHS).

Throughout this article, we assume each sensor agent (labeled by superscripts $i = 1, \dots, N$) follows a trajectory $t \mapsto x_t^i \in \Omega^i$ in a domain $\Omega^i \subset \Omega \subseteq \mathbb{R}^d$ and gathers local samples $y_t^i = g(x_t^i)$ of the unknown scalar field. Here the subset $\Omega \subseteq \mathbb{R}^d$ is the domain of interest. We can think of Ω^i as the subdomain assigned to agent i. By $X_t := [x_t^1, \dots, x_t^N]^T$ we denote the assembly of all of the states of the agents in the team. In contrast to many estimation formulations based on (finite dimensional) optimization,[8] we assume the unknown function $g \in \mathcal{H}_\mathcal{X}$, with $\mathcal{H}_\mathcal{X}$ an RKHS of real-valued functions over $\mathcal{X}$. It is well-known that the evaluation operator $E_x : \mathcal{H}_\mathcal{X} \to \mathbb{R}$ and the multi-sampling operator $\mathbb{E}_X : \mathcal{H}_\mathcal{X} \to \mathbb{R}^N$ defined via

$$E_x : g \mapsto g(x) \in \mathbb{R}, \quad \mathbb{E}_X : g \mapsto [g(x^1), \dots, g(x^N)]^T \in \mathbb{R}^N, \tag{1}$$

are both linear and bounded as mappings on $\mathcal{H}_\mathcal{X}$. With the operator $\mathbb{E}_X$ defined as above, the sampling process of the agent team is modeled as $Y_t = \mathbb{E}_{X_t} g = [g(x_t^1), \dots, g(x_t^N)]^T$. In this article, the proposed scheme for adaptive estimation is expressed as an evolution in the infinite-dimensional RKHS $\mathcal{H}_\mathcal{X}$. In particular, we denote the collective estimate of the unknown function as $\hat{g}(t) := \hat{g}(t, \cdot) \in \mathcal{H}_\mathcal{X}$. The governing equation of the centralized update law is written as

$$\dot{\hat{g}}(t) = \gamma \mathbb{E}_{X_t}^*(Y_t - \mathbb{E}_{X_t}\hat{g}(t)) = \underbrace{\gamma \mathbb{E}_{X_t}^* \mathbb{E}_{X_t}}_{\mathcal{A}(t)}(g - \hat{g}(t)), \quad \text{with } \hat{g}(0) = \hat{g}_0 \in \mathcal{H}_\mathcal{X}, \tag{2}$$

where $\mathbb{E}_{X_t}^*$ denotes the adjoint of $\mathbb{E}_{X_t}$. This is a type of DPS on the generally infinite dimensional state space $\mathcal{H}_\mathcal{X}$. From a practical perspective, a realizable estimator is obtained in this article by constructing a finite dimensional approximation $\hat{g}_L(t)$ of the ideal estimate $\hat{g}(t)$ in terms of L basis elements.

The proposed approach generalizes some classical approaches to adaptive consensus estimation in continuous time, specifically those that make an assumption of fixed dimensionality at the outset. Such is the case if it is assumed that g is expressed in terms of a finite number of regressors, for instance, in a classical formulation of adaptive estimation in Euclidean spaces. We will see that the final, overall definition of a pragmatic realizable algorithm based on Equation (2) requires two primary steps. First, we derive conditions that ensure the existence and uniqueness of the ideal estimate $\hat{g}(t)$ that solves Equation (2) for all $t \in \mathbb{R}^+$. We address in detail the convergence of the ideal estimate $\hat{g}(t, \cdot)$ to the unknown function g as $t \to \infty$ by introducing a novel persistency of excitation (PE) condition in the RKHS $\mathcal{H}_\mathcal{X}$. This argument employs a definition of PE in the weak topology on the RKHS $\mathcal{H}_\mathcal{X}$ and is the first of its kind for this class of evolution problems in an RKHS. It is possible to view this strategy as an adaptation of the weak PE analyses that have been used in the study of DPS associated with some types of partial differential equations (PDEs) as in References 9-11. For the DPS in Equation (2) we exploit the uniform embedding of the RKHS space $\mathcal{H}_\mathcal{X} \hookrightarrow C(\mathcal{X})$ and the proof of output convergence to simplify some of the arguments of convergence in the weak topology in this case.

Second, we precisely describe the construction of finite dimensional approximants $\hat{g}_L(t)$ of the ideal estimate $\hat{g}(t)$: this approximation process leads to realizable algorithms. We introduce a specific strategy for the choice of the finite dimensional basis and investigate the factors that affect convergence of finite dimensional approximations of the ideal model. The strategy described in this article assumes that bases used for finite dimensional approximation are built using samples collected (previously) along the agent trajectories. In this sense it is possible to interpret the method of the article as a data-driven method, somewhat akin in philosophy to Koopman methods. As in the general class of Koopman methods, the original system is finite dimensional and nonlinear, but the DPS defining the estimator is infinite dimensional and linear. The difference here, of course, is that we formulate the estimation problem as one of online, continuous-time adaptive estimation using a DPS as opposed to an offline, discrete-time problem that is most often attacked using techniques from regression or optimization. Note that the term adaptive estimation is used since we can draw parallels between the proposed scheme and the class of adaptive parameter estimation techniques that are designed to simultaneously estimate the system states and the unknown parameters in a governing ODE equation.[12]

The theoretical considerations of the article culminate in a precise error bound that is powerful and intuitive. We show that the error in approximating the ideal team estimate is bounded by a power of the fill distance of the samples collected along the agent trajectories. To be specific, we suppose that agent i for $1 \leq i \leq N$ collects L samples $\Xi_L^i \subset \Omega^i$ from their assigned subdomain Ω^i, and we define the fill distance $h_{\Xi_L^i,\Omega^i}$ of the finite set Ξ_L^i in Ω^i as $h_{\Xi_L^i,\Omega^i} := \sup_{x\in\Omega^i} \min_{\xi_\ell\in\Xi_L^i} d_{\mathcal{X}}(x, \xi_\ell)$ where $d_{\mathcal{X}}$ is the metric on $\mathcal{X}$. We then guarantee that the realizable estimate, corresponding to the finite dimensional approximant, satisfies

$$\|\hat{g}(t) - \hat{g}_L(t)\| \leq \mathcal{O}\left(\max_{1\leq i\leq N} h^s_{\Xi_L^i,\Omega^i}\right).$$

where $s > 0$ is a parameter that describes the smoothness of the functions in $\mathcal{H}_{\mathcal{X}}$. The precise definition of s depends on the power function of the kernel that induces $\mathcal{H}_{\mathcal{X}}$, and we discuss later how to choose the bases and kernels that define $\mathcal{H}_{\mathcal{X}}$ and thus determine s. Overall then, the formulation described in this article enables us to improve on conclusions that are ordinarily achieved in finite dimensional estimation: we prove convergence of estimates of the unknown function g in terms of a norm defined on a RKHS of real-valued functions. The error of approximation converges (geometrically in terms of the smoothness parameter s) as a function of the fill distance of the samples in the assigned domains.

Apart from the obvious difference that it is more difficult to treat infinite dimensional estimation problems in general, there are two features that must be addressed that have no counterpart in estimation in Euclidean spaces. One difficulty is, of course, choosing the function space $\mathbb{F}$ used to measure convergence in practice. Since all norms are equivalent on $\mathbb{R}^d$, any norm is, theoretically speaking, as good as another in the study of the finite dimensional estimation problem. The selection of a particular norm might simplify proofs or be convenient for calculations, but derivation of convergence of estimates in one norm implies convergence in any norm. This is most definitely not the case in treating infinite dimensional estimation problems, and the choice of the topology on $\mathbb{F}$ comes to assume a critical role in these formulations. It is entirely conceivable that an estimation problem is well-posed and convergent in one choice of function space but not in another. Even if the estimates are convergent, the rates of convergence will depend generally on the specific function space choice $\mathbb{F}$. One of the contributions of this article is to formulate the consensus estimation problem in continuous time in such a way that a reasonable collection of spaces $\mathbb{F}$ can be selected in practice. This article describes a broad family of such methods.

The second difficulty, which is coupled to the first, is that any practical estimation method must be implementable. This demands that issues of approximation be addressed. Again, this has no counterpart in estimation in Euclidean spaces. It makes sense to discuss the rate of convergence *in time* of either finite or infinite dimensional estimation problems. However, for the infinite dimensional estimation problem, we must also address rates of convergence *in space*. As the proposed estimator $\hat{g}$ evolves in the generally infinite dimensional space $\mathbb{F}$, we obtain a computationally tractable estimator by projecting $\hat{g}$ onto a finite dimensional subspace of dimension L. The resulting estimator is of the form

$$\hat{g}_L(t,x) := \sum_{\ell=1}^{L} \alpha_{L,\ell}(t)\phi_{L,\ell}(x), \tag{3}$$

where $\{\phi_{L,\ell}\}_{\ell=1}^L \subset \mathbb{F}$ denote a set of user-specified basis functions and $\{\alpha_{L,\ell}(t)\}_{\ell=1}^L \subset \mathbb{R}$ denotes a set of time-varying coefficients. Indeed, estimators of the general form (3) have been studied in data-dependent applications within the context of spline theory,[13-15] kernel ridge regression,[16-18] and Gaussian process regression.[1,19,20] However, in these contexts the coefficients $\{\alpha_{L,\ell}(t)\}_{\ell=1}^L \equiv \{\alpha_{L,\ell}\}_{\ell=1}^L$ are time-invariant. When we say that we address rates of convergence *in space*, we mean that we describe the rate of convergence of the error $\|g - \hat{g}_L(t,\cdot)\|_{\mathbb{F}}$ as a function in the infinite dimensional space $\mathbb{F}$. It turns out that the selection of bases $\{\phi_{L,\ell}\}_{\ell=1}^L$ and the choice of the function space $\mathbb{F}$ are intimately connected here. In short then, we can state one general contribution of this article: this article formulates a specific strategy for choosing the space $\mathbb{F}$, bases $\{\phi_{L,\ell}\}_{\ell=1}^L$, and an evolution equation that defines a learning law such that convergence, and in some instances rates of convergence, as measured in $\|\cdot\|_{\mathbb{F}}$ can be deduced.

Even though the study of multi-agent centralized estimation in infinite dimensional problems is less mature than the corresponding study of finite dimensional problems, there are some contexts where aspects of this problem have been studied. When the environmental modeling task is carried out by a team of agents that communicate with each other, it can be advantageous to view the team of agents as a *mobile* wireless sensor network (WSN). The agents traverse the region of interest to collect measurements of the unknown external field, and these measurements are directly used to update a real-time model of the surrounding environment. One family of methods of this type consist of estimation problems where a discrete dynamic model is induced using optimization at each time step in a learning theory or regression framework. One of the often cited foundations of these methods include the body of work in References 21-23, which consist of single-agent formulations of discrete dynamics in a learning theory formulation. A good survey of the generalization to agent teams that communicate over networks can be found in References 24-26 and the references therein. As a whole, these methods construct discrete models of dynamics that are realized via the solution of optimization problems at each time step. The work by Predd et al.[24,27] investigates how the communication topology influences the convergence of such formulations of discrete dynamics via learning theory. While the work in References 24,27 studies the convergence of discrete dynamics when the information shared over the network takes the form of samples and pointwise errors, the authors in References 28,29 study discrete dynamics where the learning theory problem is formulated in terms of a saddle point problem. These methods differ from the approach in References 24,27 in that the information that is exchanged over the communication network consists of approximations of costates, not samples. This formulation also addresses rates of convergence (in space) of the discrete time estimates, a topic not covered in References 24,27. More recently, the work in References 5,30 gives examples where machine learning approaches that define discrete dynamics in time via the repeated solution of an optimization problems.

In addition to the discrete time formulation of adaptive estimation summarized above, there also is a body of work that studies infinite dimensional consensus estimation problems for various classes of DPSs generated by families of PDEs. Problems of this type are studied in the family of papers.[9-11] These formulations are quite general and are formulated for systems where the evolution laws *for the agents* are defined in terms of Gelfand triples, with the constituent operators satisfying some standard boundedness and coercivity properties for common DPS. In these papers the agent states evolve in an infinite dimensional state space, in contrast to the situation here. The system studied in this article is much simpler, and much of the analysis in this article is aimed at establishing just how much simpler questions of well-posedness are to establish for the system at hand. This is the thrust of Section 3.1.1 that makes use of the embedding $\mathcal{H}_\mathcal{X} \hookrightarrow C_\mathcal{X}$ to provide an arguably simpler proof of existence and uniqueness. Also, the discussion of rates of convergence in this article, obtained using certain types of bases defined in terms of the kernel that defines the RKHS, as of yet has no analog in these papers on PDE methods. Thus, while the system in this article is less general than the PDE systems, the conclusions are stronger. It is of interest to investigate whether similar rates of convergence could be deduced for the more complex systems where the agent states evolve in an infinite dimensional space (as well as the function estimate).

1.1 | Overview of the methodology

The centralized adaptive estimation strategy defines the collective estimate $\hat{g}(t) := \hat{g}(t, \cdot) \in \mathcal{H}_\mathcal{X}$ which evolves according to

$$\dot{\hat{g}}(t) = \gamma \mathbb{E}^*_{X_t}(Y_t - \mathbb{E}_{X_t}\hat{g}(t)) = \underbrace{\gamma \mathbb{E}^*_{X_t}\mathbb{E}_{X_t}}_{:=\mathcal{A}(t)}\tilde{g}(t), \quad \text{subject to } \hat{g}(0) = \hat{g}_0, \tag{4}$$

with $\mathbb{E}^*_X : \mathbb{R}^d \to \mathcal{H}_\mathcal{X}$ the adjoint of $\mathbb{E}_X$ for each $X \in \mathcal{X}^N$. The estimate $\hat{g}(t) \in \mathcal{H}_\mathcal{X}$ is referred to as the ideal, infinite dimensional estimate since the evolution law in (4) is not constrained to a particular finite dimensional space. It follows that the ideal estimation error $\tilde{g}(t) := g - \hat{g}(t)$ satisfies

$$\dot{\tilde{g}}(t) = -\mathcal{A}(t)\tilde{g} = -\gamma \sum_{i=1}^{N} E^*_{x^i_t} E_{x^i_t}\tilde{g} = -\gamma \sum_{i=1}^{N} \tilde{g}(x^i_t)\mathfrak{K}(x^i_t, \cdot). \tag{5}$$

Later, in Section 3.2 we discuss a finite dimensional, history-dependent approximation $\hat{g}_L$ of the ideal estimate $\hat{g}$. Note that there is one function estimate $t \mapsto \hat{g}(t)$ in this centralized approach, but it is constructed using samples taken by all agents. We consider the setting in which all agents can communicate with a single computational node that collects samples from various agents and builds an approximation of the solution of Equation (4). Indeed, if the communication bandwidth is sufficient and all agents have enough computing power, each agent i can build, using shared samples, the approximation $\hat{g}^i(t) \equiv \hat{g}(t)$ that is identical and obtained using the same equation.

1.2 | Our contributions

The primary result is stated in Theorem 1: a unique global solution $t \mapsto \hat{g}(t)$ to Equation (4) exists on all $\mathbb{R}^+$ provided the RKHS $\mathcal{H}_\mathcal{X}$ is continuously embedded in the space of continuous functions, $\mathcal{H}_\mathcal{X} \hookrightarrow C(\mathcal{X})$. The result is obtained by following standard fixed-point arguments that are familiar in the study of ODEs. This result is qualitatively quite similar to the related problem addressed in References 31,32 where RKHS embedding is used to frame the adaptive estimation of a nonlinear function appearing in the ODEs with respect to the state $x(t)$. Here, unlike,[31,32] the unknown function is an external field and is not a function that influences the agent trajectory $t \mapsto x^i(t)$ for agent i.

Once conditions are derived that guarantee the existence and uniqueness of the governing infinite dimensional equations, the remainder of the article concentrates on the study of convergence of the estimates to the unknown function $g \in \mathcal{H}_\mathcal{X}$. The analysis of convergence begins with the study of when the ideal, centralized estimate $\hat{g}(t)$ that solves Equation (4) converges to the unknown function g as $t \to \infty$. As in other adaptive estimation problems in $\mathbb{R}^d$ for ODEs, or in some problems for PDEs, this step relies on the suitable definition of persistency of excitation conditions in an RKHS. Again, such PE conditions have been studied for the related problem of estimating an unknown scalar function appearing in the governing ODEs in References 31,32, but not for the estimation of an external field.

In contrast to prior work, in this article we explore weak PE conditions that are stated in terms of the weak topology on $\mathcal{H}_\mathcal{X}$ for adaptive estimation of an external scalar field. The strategy based on the weak topology is qualitatively similar to that in Reference 33 for the study of certain classes of PDEs as in References 9-11, but the details of the proofs must be modified to exploit properties of the RKHS and the RKHS embedding formulation. This characterization of convergence is made precise by defining a weak norm $|\cdot|_w$ on $\mathcal{H}_\mathcal{X}$ and defining a linear subspace $W \subseteq \mathcal{H}_\mathcal{X}$ that consists of functions that, in some sense, "fail to be weakly persistently excited." It is later shown that any solution $\tilde{g}(t)$ of the governing ideal error equation (5) is contained in $\overline{B_r(0)}$, the closed ball in $\mathcal{H}_\mathcal{X}$ of radius r centered at the origin, for a suitably large $r > 0$. By introducing the weak distance from g to $\hat{W} := W \bigcap \overline{B_r(0)}$ as

$$d_w(g, \hat{W}) := \inf_{h \in \hat{W}} |g - h|_w,$$

we show in Theorem 3 that the ideal error $\tilde{g}(t) := g - \hat{g}(t, \cdot)$ satisfies

$$\lim_{t \to \infty} d_w(\tilde{g}(t), \hat{W}) = 0.$$

That is, asymptotically in time, the ideal error $\tilde{g}(t)$ approaches the set of functions that fail to be weakly persistently excited in $\mathcal{H}_\mathcal{X}$. Moreover, convergence in the weak topology to W enables some further, intuitive understanding of the persistency condition. If the entire space $\mathcal{H}_\mathcal{X}$ is weakly persistently excited, which means that $\hat{W} = \{0\}$, then we prove in Theorem 1 that we have the pointwise convergence

$$\lim_{t \to \infty} \hat{g}(t, x) = g(x),$$

for all $x \in \mathcal{X}$.

A practical (computationally tractable) estimator is obtained by approximating $\hat{g}$ in a finite dimensional space $\mathcal{H}_L$. Thus, the second phase of studying the convergence of estimates to the unknown function g is rooted in the construction of $\hat{g}_L(t)$. The total error $\tilde{g}_L(t) := g - \hat{g}_L(t)$ is then bounded as

$$\underbrace{\|\tilde{g}_L(t)\|}_{\text{total error}} := \|g - \hat{g}_L(t)\| \leq \underbrace{\|g - \hat{g}(t)\|}_{\|\tilde{g}(t)\| = \text{ideal error}} + \underbrace{\|\hat{g}(t) - \hat{g}_L(t)\|}_{\|\overline{g}_L(t)\| = \text{approximation error}}.$$

From the analysis of the above error bound, we state conditions that guarantee the ideal error converges to zero as $t \to \infty$. The finite dimensional estimate $\hat{g}_L(t)$ can be shown, for suitable choices of the finite dimensional spaces $\mathcal{H}_L$, to converge to the ideal estimate $\hat{g}(t)$ as $L \to \infty$. Rates of convergence that measure how quickly the approximation error $\|\overline{g}_L(t)\|$ converges to zero as the dimension of $\mathcal{H}_L$ increases are derived in Theorem 4 and Corollary 2. The sharpest results are given in Corollary 2. In the corollary, we suppose that each agent gathers discrete samples at the inputs $\Xi_L^i \subset \Omega^i \subseteq \Omega \subset \mathcal{X}$, and that these inputs are used to construct the space of approximants

$$\mathcal{H}_L := \text{span}\{\mathfrak{K}(\xi^i_{L,\ell},\cdot) \in \mathcal{H}_\mathcal{X} \mid \xi^i_{L,\ell} \in \Xi^i_L, 0 \le \ell \le L-1, 1 \le i \le N\}.$$

Note that in practice the samples are collected at the locations Ξ^i_L along the trajectory of agent i, and it is possible to interpret the approximations process as a history-dependent method of building bases from trajectories. It is assumed that the agents control their trajectories $t \mapsto x^i(t)$, so that the full collection of samples Ξ_L become increasingly dense in Ω as $L \to \infty$. The distribution of the samples Ξ_L in the domain of interest Ω is measured by the fill distance $h_{\Xi_L,\Omega}$ defined as

$$h_{\Xi_L,\Omega} := \sup_{x\in\Omega} \min_{\xi_\ell \in \Xi_L} d_\mathcal{X}(x,\xi),$$

and a corresponding definition holds for the fill distance $h_{\Xi^i_L,\Omega^i}$ of the local samples Ξ^i_L in the local domain Ω^i. Corollary 2 states conditions that ensure that the approximation error satisfies

$$\|\overline{g}_L(t)\|^2 = \|\hat{g}(t) - \hat{g}_L(t)\|^2 \le C_T \max_{1\le i\le N} h^s_{\Xi^i_L,\Omega^i},$$

for $t \in [0,T]$. In this equation the exponent s depends on the type of kernel used to define the space $\mathcal{H}_\mathcal{X}$. The coefficient C_T depends on T, so the approximation error $\overline{g}_L(t)$ converges uniformly in time over any bounded interval [0, T].

The remainder of this article is organized as follows. Section 2 reviews notation, symbols, and a few common conventions and definitions, as well as some standard definitions of function spaces used in this article. A survey of reproducing kernels and their native spaces is given in Section 2.2. Section 3 introduces the ideal centralized estimation method. Section 3.1.1 gives sufficient conditions that guarantee the existence and uniqueness of solutions, Section 3.1.2 analyzes its output convergence, and Section 3.1.3 introduces the weak PE conditions that suffice to prove that the ideal error converges to zero as $t \to \infty$. Finite dimensional approximations are introduced in Section 3.2, while Sections 4 and 5 discuss some example computations and the conclusions.

2 | BACKGROUND

2.1 | Notation and symbols

We denote by $\mathbb{R}, \mathbb{R}^+, \mathbb{N}, \mathbb{N}_0$ the real numbers, the non-negative real numbers, positive integers, and non-negative integers, respectively. The relationship $a \lesssim b$ means that there is a constant c, which does not depend on a, b, such that $a \le c \cdot b$. The symbol $\gtrsim$ is defined similarly. When $\mathcal{X}, \mathcal{Y}$ are normed vector spaces, we write $\mathcal{X} \hookrightarrow \mathcal{Y}$ to mean that $\mathcal{X}$ is continuously embedded in $\mathcal{Y}$. That is, $\mathcal{X}$ is a subset of $\mathcal{Y}$ and there is a constant $C > 0$ such that $\|u\|_\mathcal{Y} \le C\|u\|_\mathcal{X}$.

A number of standard spaces of real-valued functions are used in this article. The usual norm on the space of real-valued functions that are k times continuously differentiable is denoted

$$\|f\|_{C^k} := \sum_{|\alpha|\le k} \left| \frac{\partial^\alpha f}{\partial x^\alpha} \right|,$$

where $\alpha := (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d_0$ is a multiindex with length $|\alpha| := \sum_{i=1,\dots,d} \alpha_i$. The spaces of Lebesgue μ-integrable functions are likewise needed: in the usual way we define

$$\|f\|_{L^p_\mu} := \begin{cases} \left(\int_\mathcal{X} |f(x)|^p \mu(dx)\right)^{1/p} & 1 \le p < \infty, \\ \mu - \text{ess sup}|f(x)| & p = \infty, \end{cases}$$

where μ is a measure on $\mathcal{X}$. When μ is the Lebesgue measure, we simply put $\mu(dx) = dx$ and write $\|f\|_{L^p}$.

In addition to real-valued functions, we will also consider spaces of functions of time that take values in a Hilbert space $\mathcal{H}$. The space of continuous, $\mathcal{H}$-valued functions of time is equipped with the norm

$$\|f\|_{C(I,\mathcal{H})} := \sup_{t\in I} \|f(t)\|_\mathcal{H},$$

when $I \subseteq \mathbb{R}^+$ is an interval. The Lebesgue spaces of $\mathcal{H}$-valued functions over an interval I are defined in terms of the norm

$$\|f\|_{L^p(I,\mathcal{H})} := \begin{cases} \left(\int_I \|f(\tau)\|_{\mathcal{H}}^p d\tau\right)^{1/p} & 1 \le p < \infty, \\ \operatorname*{ess\,sup}_{\tau \in I} \|f(\tau)\|_{\mathcal{H}} & p = \infty, \end{cases}$$

with I an interval $I \subseteq \mathbb{R}^+$. As discussed in detail in the next section, in this article $\mathcal{H}_\mathcal{X}$ is always an RKHS space of real-valued functions over the set $\mathcal{X}$.

2.2 | Reproducing kernel Hilbert spaces

A RKHS $\mathcal{H}_\mathcal{X}$ over a set $\mathcal{X}$ is a Hilbert space that is defined in terms of reproducing kernel $\mathfrak{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that is symmetric, continuous, and positive definite. The kernel is positive definite when $\sum_{1 \le i,j \le m} \mathfrak{K}(x_i, x_j)\alpha_i\alpha_j \ge 0$ for any finite number of points $\{x_i \in \mathcal{X}\}_{i=1}^m$ and scalar coefficients $\{\alpha_i\}_{i=1}^m$, which holds when the Grammian matrix $[\mathfrak{K}(x_i, x_j)] \in \mathbb{R}^{m \times m}$ is positive definite. Common reproducing kernels include the Sobolev–Matern kernels, squared exponential kernels, and multi-quadratic kernels.[34,35] For any $x \in \mathcal{X}$, we denote the function $\mathfrak{K}(x, \cdot) \in \mathcal{H}_\mathcal{X}$ by $\mathfrak{K}_x(\cdot) := \mathfrak{K}(x, \cdot)$ and refer to it as a kernel basis function centered at x. The RKHS defined in terms of the kernel $\mathfrak{K}$ is precisely the closed linear span

$$\mathcal{H}_\mathcal{X} = \overline{\text{span}\{\mathfrak{K}_x | x \in \mathcal{X}\}},$$

where the closure is taken with respect to the pre-inner product $\langle \mathfrak{K}_x, \mathfrak{K}_y \rangle_{\mathcal{H}_\mathcal{X}} = \mathfrak{K}(x, y)$ that is defined for all $x, y \in \mathcal{X}$.

A necessary and sufficient condition for a Hilbert space $\mathcal{H}_\mathcal{X}$ to be a RKHS is that for any $x \in \mathcal{X}$ the evaluation functional $E_x : \mathcal{H}_\mathcal{X} \to \mathbb{R}$ is a bounded operator.[36] The evaluation functional E_x associated with the point $x \in \mathcal{X}$ is defined as $E_x f = f(x)$ for each $f \in \mathcal{H}_\mathcal{X}$, and so boundedness requires that $|E_x f| = |f(x)| \le c_x \|f\|_{\mathcal{H}_\mathcal{X}}$ for some constant $c_x > 0$ that may depend on x. In this article we only consider RKHSs that are continuously embedded in $C(\mathcal{X})$, the space of continuous functions on $\mathcal{X}$, and we write $\mathcal{H}_\mathcal{X} \hookrightarrow C(\mathcal{X})$ to denote the embedding. Such an embedding implies that there is a positive constant C such that $\|f\|_{C(\mathcal{X})} \le C\|f\|_{\mathcal{H}_\mathcal{X}}$ for all $f \in \mathcal{H}_\mathcal{X}$. The existence of a constant $\overline{k} > 0$ such that $\mathfrak{K}(x, x) \le \overline{k}^2$ for all $x \in \mathcal{X}$ is sufficient to ensure the embedding $\mathcal{H}_\mathcal{X} \hookrightarrow C(\mathcal{X})$ holds. All kernels considered in this article satisfy this condition. In some situations, notably in the application of Barbalat's lemma in Section 3.1.2, we require the continuous embedding of $\mathcal{H}_\mathcal{X}$ in a space of smoother functions $\mathcal{H}_\mathcal{X} \hookrightarrow C^1(\mathcal{X}) \hookrightarrow C(\mathcal{X})$.

The reproducing property is the defining feature of a RKHS $\mathcal{H}_\mathcal{X}$, which states that for any $f \in \mathcal{H}_\mathcal{X}$ and any $x \in \mathcal{X}$, it holds that

$$f(x) = E_x f = \langle \mathfrak{K}_x, f \rangle_{\mathcal{H}_\mathcal{X}}.$$

Throughout this article we also employ the adjoint of the evaluation operator $E_x^* : \mathbb{R} \to \mathcal{H}_\mathcal{X}$ which satisfies $E_x^* \alpha = \alpha \mathfrak{K}_x$ for all $\alpha \in \mathbb{R}$. In the following discussions, we will frequently use the two operators together in the form of

$$E_x^* E_x g = E_x^*[g(x)] = g(x)\mathfrak{K}_x,$$

which emphasizes that $E_x^* E_x g$ is a function in the RKHS. If we evaluate this function at a point y, we obtain $(E_x^* E_x g)(y) = g(x)\mathfrak{K}_x(y) = g(x)\mathfrak{K}(x, y)$. When we consider a team of N agents located at $X = \{x^1, \dots, x^N\} \in \mathbb{R}^d$, we denote the evaluation operator for the multiple locations by

$$\mathbb{E}_X g := [E_{x^1} g, \dots, E_{x^N} g]^\mathsf{T} = [g(x_t^1), \dots, g(x_t^N)]^\mathsf{T}.$$

Using the definition of the adjoint, it can be verified that the adjoint $\mathbb{E}_X^*$ that maps from $\mathbb{R}^N$ to $\mathcal{H}_\mathcal{X}$ satisfies $\mathbb{E}_X^* \alpha = \sum_{i=1}^N \alpha_i \mathfrak{K}(x_t^i, \cdot)$, for each $\alpha \in \mathbb{R}^N$. Similarly, the operator $\mathbb{E}_X^* \mathbb{E}_X$ that maps from $\mathcal{H}_\mathcal{X}$ to $\mathcal{H}_\mathcal{X}$ is derived as

$$\mathbb{E}_X^* \mathbb{E}_X g = \mathbb{E}_X^* \left([g(x^1), \dots, g(x^N)]^\mathsf{T}\right) = \sum_{i=1}^N g(x^i)\mathfrak{K}_{x^i}.$$

TABLE 1 Notation summary

Symbol	Description
$\mathcal{X}$	Input domain
$\mathfrak{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$	Kernel function
$\mathcal{H}_{\mathcal{X}}$	RKHS of functions, induced by $\mathfrak{K}$
$\overline{k} \in \mathbb{R}$	Upper bound on $\|\mathfrak{K}(x, \cdot)\|_{\mathcal{H}_{\mathcal{X}}}$ for all $x \in \mathcal{X}$.
$g \in \mathcal{H}_{\mathcal{X}}$	Function of interest
$\hat{g} \in \mathcal{H}_{\mathcal{X}}$	Ideal (infinite dimensional) approximation of g
$\tilde{g} = g - \hat{g} \in \mathcal{H}_{\mathcal{X}}$	Ideal approximation error
$X_t = \{x_t^1, \dots, x_t^N\} \subset \mathcal{X}$	Sampling locations of N agents at time instance t
$\mathbb{E}_X : \mathcal{H}_{\mathcal{X}} \to \mathbb{R}^N$	Evaluation operator defined such that $\mathbb{E}_X f = f(X) \in \mathbb{R}^N$ for any $X = \{x_i\}_{i=1}^N$ and any $f \in \mathcal{H}_{\mathcal{X}}$
$\mathbb{E}_X^* : \mathbb{R}^N \to \mathcal{H}_{\mathcal{X}}$	Adjoint of $\mathbb{E}_X$
$\gamma \in \mathbb{R}^+$	Adaptation rate parameter
$\mathcal{A}(t) := \gamma \mathbb{E}_X^* \mathbb{E}_X$	
$\tilde{Y}(t) = \tilde{g}(X_t) \in \mathbb{R}^N$	Ideal sampling error at time instance t.
$L \in \mathbb{R}$	Dimension of finite-dimensional approximation
$\Xi_L = \{\xi_1, \dots, \xi_L\} \subset \mathcal{X}$	Set of input locations
$\mathcal{H}_L \subset \mathcal{H}_{\mathcal{X}}$	Finite-dimensional subspace spanned by $\{\mathfrak{K}(\xi, \cdot)\}_{\xi \in \Xi_L}$
$\hat{g}_L \in \mathcal{H}_L$	Finite-dimensional approximation of $\hat{g}$
$W \subseteq \mathcal{H}_{\mathcal{X}}$	Set of functions that are not being weakly persistently excited; see Equation (9) for formal definition
Π_L	$\mathcal{H}_{\mathcal{X}}$-orthogonal projection operator on to $\mathcal{H}_L$
$Q_L := I - \Pi_L$	
$h_{\Xi_L, \Omega}$	Fill distance of the samples Ξ_L with respect to the subspace $\Omega \subseteq \mathcal{X}$
$\alpha_L \in \mathbb{R}^L$	Coefficients used to construct $\hat{g}_L$.

We will use the shorthand notation $x_t := x(t)$ to denote the trajectory $t \mapsto x$ of a single agent, and $X_t := \{x_t^1, \dots, x_t^N\}^\top \in \mathcal{X}^N$ to denote the trajectory of the team of N agents. Notation used throughout this article is summarized in Table 1.

3 | CENTRALIZED ESTIMATION FROM AGENT NETWORKS

In this section, we study a centralized adaptive estimation approach. We begin by establishing the existence and uniqueness of the estimator $\hat{g}(t)$ and estimator error $\tilde{g}(t)$, which evolve in time according to (4) and (5), respectively. We then show that the ideal sampling error $\tilde{Y}(t) = \mathbb{E}_{X_t}\tilde{g}(t) \to \mathbf{0} \in \mathbb{R}^N$ as $t \to \infty$. We conclude the section by proposing a finite-dimensional approximation $\hat{g}_L(t)$ of $\hat{g}(t)$, and show that $\hat{g}_L(t) \to \hat{g}(t)$ as the dimension $L \to \infty$.

3.1 | The ideal (infinite dimensional) adaptive estimator

3.1.1 | Existence and uniqueness of solutions

We begin with a study to ensure the governing Equation (5) is well-posed. That is, we want to establish that for any $\hat{g}_0 \in \mathcal{H}_{\mathcal{X}}$ there is a unique trajectory $t \to \hat{g}(t) \in \mathcal{H}_{\mathcal{X}}$ that satisfies Equation (4), or equivalently find the error trajectory $t \to \tilde{g}(t) \in \mathcal{H}_{\mathcal{X}}$ that satisfies Equation (5). Such a proof is straightforward when we assume that the kernel $\mathfrak{K}$ is selected so that the continuous embedding

$$\mathcal{H}_{\mathcal{X}} \hookrightarrow C(\mathcal{X}), \tag{6}$$

holds. We do not state the most general form of an existence proof for the trajectory of the ideal estimate, but choose to study the existence of classical solutions in $C([0,T],\mathcal{H}_{\mathcal{X}})$. While some standard tools as in Daleckii and Krein[37] could be applied to consider the more general notion of Caratheodory solutions that take values in a Hilbert space, the presentation here emphasizes how assumption (6) enables strategies that are entirely analogous to the case of time-varying evolution in finite dimensional Euclidean spaces.

Theorem 1. *Let the trajectory $t \mapsto x^i(t)$ be continuous in $\mathcal{X} = \mathbb{R}^d$ for each $i = 1, \dots, N$. Suppose the RKHS $\mathcal{H}_{\mathcal{X}}$ is uniformly embedded in the space of continuous functions $C(\mathcal{X})$. Then the DPS governed by Equation (5) has a unique solution $\tilde{g} \in C([0,T],\mathcal{H}_{\mathcal{X}})$ over any bounded interval $[0,T] \subset [0,\infty)$.*

Proof. We employ a standard fixed point argument, one that is analogous to the finite dimensional case.[38(p. 657)] Define the integral operator S in the usual way so that for any $f \in \mathcal{H}_{\mathcal{X}}$

$$(Sf)(t) := f(0) + \int_0^t \mathcal{A}(\tau) f(\tau) d\tau.$$

We argue that there is a classical solution to Equation (5) on any finite interval $[t_0, T]$ by making a fixed point argument for S as an operator $S : C([t_0, t_0+\delta], \mathcal{H}_{\mathcal{X}}) \to C([t_0, t_0+\delta], \mathcal{H}_{\mathcal{X}})$ for some $\delta > 0$ such that $t_0 + \delta \le T$. Note that the space $C([t_0, t_0+\delta], \mathcal{H}_{\mathcal{X}})$ is a Banach space endowed with the supremum norm

$$\|g\|_{C([t_0,t_0+\delta],\mathcal{H}_{\mathcal{X}})} := \sup_{\tau \in [t_0,t]} \|g(\tau)\|_{\mathcal{H}_{\mathcal{X}}}.$$

Let $u, v \in C([t_0, t_0+\delta], \mathcal{H}_{\mathcal{X}})$, then $Su, Sv \in C([t_0, t_0+\delta], \mathcal{H}_{\mathcal{X}})$. We have for $t \le t_0 + \delta$,

$$\begin{aligned}\|Su(t) - Sv(t)\|_{\mathcal{H}_{\mathcal{X}}} &= \left\| \int_{t_0}^t \mathcal{A}(\tau)(u(\tau) - v(\tau)) d\tau \right\|_{\mathcal{H}_{\mathcal{X}}} \\ &\le \gamma N \overline{k}^2 \delta \|u - v\|_{C([t_0,t_0+\delta],\mathcal{H}_{\mathcal{X}})},\end{aligned}$$

where the last inequality follows from the fact that the operator $\mathcal{A}(t)$ is bounded uniformly in t. To show this, notice that for any fixed $f \in \mathcal{H}_{\mathcal{X}}$

$$\begin{aligned}\|\mathcal{A}(t) f\|_{\mathcal{H}_{\mathcal{X}}} = \left\| \gamma \mathbb{E}_{X_t}^* \mathbb{E}_{X_t} f \right\|_{\mathcal{H}_{\mathcal{X}}} &= \gamma \left\| \sum_{i=1}^N \left\langle f, \mathfrak{K}_{x_t^i} \right\rangle \mathfrak{K}_{x_t^i} \right\|_{\mathcal{H}_{\mathcal{X}}} \\ &\le \gamma \|f\|_{\mathcal{H}_{\mathcal{X}}} \sum_{i=1}^N \|\mathfrak{K}_{x_t^i}\|_{\mathcal{H}_{\mathcal{X}}}^2.\end{aligned}$$

By assumption (6), a constant $\overline{k}$ exists such that $\|\mathfrak{K}_x\|_{\mathcal{H}_{\mathcal{X}}} = \sqrt{\langle \mathfrak{K}_x, \mathfrak{K}_x \rangle} \le \overline{k}$ uniformly for all $x \in \mathcal{X}$. Thus, $\|\mathcal{A}(t)\| \le \gamma N \overline{k}^2$ holds for all $x \in \mathcal{X}$, which in turn implies $\|\mathcal{A}(t)\| \le \gamma N \overline{k}^2$ for all $t \ge 0$.

Now, choose $\delta < \alpha / (\gamma N \overline{k}^2)$ for some $\alpha < 1$, then

$$\begin{aligned}\|S(u - v)\|_{C([t_0,t_0+\delta],\mathcal{H}_{\mathcal{X}})} &= \sup_{\tau \in [t_0,t_0+\delta]} \|Su(\tau) - Sv(\tau)\|_{\mathcal{H}_{\mathcal{X}}} \\ &< \alpha \|u - v\|_{C([t_0,t_0+\delta],\mathcal{H}_{\mathcal{X}})},\end{aligned}$$

which implies the operator S is a contraction on $C([t_0, t_0+\delta], \mathcal{H}_{\mathcal{X}})$. By the Banach Fixed-point theorem, the operator S has a *unique* fixed point in $C([t_0, t_0+\delta], \mathcal{H}_{\mathcal{X}})$, which implies that the governing Equation (5) has a solution over $[t_0, t_0+\delta]$. Notice that the choice of δ does not depend on the initial condition $\hat{g}_0$, nor the initial time t_0. We can extend the solution to $[t_0, \infty)$ by repeatedly solving Equation (5) in the interval of length δ with the initial condition set as the final value of the previous interval.

To establish the uniqueness of the solution to (5), we use the Grönwall inequality, as stated in Reference 39 (ch. 3). Suppose that $\tilde{g}$ and $\tilde{h}$ both solve the governing Equation (5), then we have

$$\dot{\tilde{g}}(t) = -\mathcal{A}(t)\tilde{g}(t), \qquad \tilde{g}(0) = \tilde{g}_0,$$
$$\dot{\tilde{h}}(t) = -\mathcal{A}(t)\tilde{h}(t), \qquad \tilde{h}(0) = \tilde{h}_0.$$

If we subtract the two equations above and integrate over the interval $[t_0, t]$, we have

$$\tilde{g}(t) - \tilde{h}(t) = \tilde{g}_0 - \tilde{h}_0 - \int_{t_0}^{t} \mathcal{A}(\tau)(\tilde{g}(\tau) - \tilde{h}(\tau))d\tau.$$

This implies

$$\left\|\tilde{g}(t) - \tilde{h}(t)\right\|_{\mathcal{H}_\mathcal{X}} \le \left\|\tilde{g}_0 - \tilde{h}_0\right\|_{\mathcal{H}_\mathcal{X}} + \gamma N\overline{k}^2 \int_0^t \left\|\tilde{g}(\tau) - \tilde{h}(\tau)\right\|_{\mathcal{H}_\mathcal{X}} d\tau,$$

where we have used the uniform boundedness $\|\mathcal{A}(\tau)\| \le \gamma N\overline{k}^2$. Application of the Grönwall inequality yields

$$\left\|\tilde{g}(t) - \tilde{h}(t)\right\|_{\mathcal{H}_\mathcal{X}} \le \left\|\tilde{g}_0 - \tilde{h}_0\right\|_{\mathcal{H}_\mathcal{X}} e^{\int_0^t \gamma N\overline{k}^2 d\tau}.$$

If the trajectories $t \mapsto \tilde{g}(t) \in \mathcal{H}_\mathcal{X}$ and $t \mapsto \tilde{h}(t) \in \mathcal{H}_\mathcal{X}$ start at the same initial condition $\tilde{g}_0 = \tilde{h}_0$, then it follows that $\left\|\tilde{g}(t) - \tilde{h}(t)\right\|_{\mathcal{H}_\mathcal{X}} = 0$. Thus, we conclude the solution $\tilde{g}(t) = \tilde{h}(t)$ is unique. ■

Having shown the estimator in Equation (5) is well-posed, we proceed to study the behavior of the ideal sampling error $\tilde{Y}(t) := \mathbb{E}_{X_t}\tilde{g}(t) = Y(t) - \mathbb{E}_{X_t}\hat{g}(t) \in \mathbb{R}^N$ as $t \to \infty$.

3.1.2 | Output error convergence

This article can be interpreted as a way of lifting conventional adaptive estimation in Euclidean spaces, as typified by the popular References 12,40-42, to the infinite dimensional setting that uses an RKHS for representation of the function estimate. It is well-known that the proof of convergence asymptotically in time in the finite dimensional Euclidean case proceeds in two general steps. First, the state and output convergence of the estimation scheme is proven, and subsequently an analysis based on a suitably defined PE condition is used to prove parameter convergence. Sections 3.1.2 and 3.1.3 follow this general outline, albeit the RKHS is the state space of the governing DPS. In the next section, we discuss a weak PE condition. Here, in this section, we apply (a generalization of) Barbalat's lemma from Reference 43 to study the convergence of the ideal output error $t \mapsto \tilde{Y}(t)$ to zero as $t \to \infty$.

Theorem 2. *Suppose that $\|x^i(t)\| \le c_0$ and $\|\dot{x}^i(t)\| \le c_1$ for all $i = 1, \dots, N$ and $t \in [0, \infty)$ and that $\mathcal{H}_\mathcal{X} \hookrightarrow C^1(\mathcal{X})$. Then the output error $\tilde{Y}(t) := \mathbb{E}_{X_t}\tilde{g}(t)$ generated by the learning law in Equation (5) converges to zero,*

$$\lim_{t\to\infty} \tilde{Y}(t) = 0 \in \mathbb{R}^N.$$

Proof. Barbalat's lemma is a standard tool used in the analysis of the asymptotic behavior of nonautonomous systems. While there are a few equivalent versions of the lemma, for control-related problems it is common to use Barbalat's lemma written in the following manner. Suppose $\tilde{Y} \in L^\infty([0,\infty), \mathbb{R}^N) \cap L^2([0,\infty), \mathbb{R}^N)$, and $\dot{\tilde{Y}} \in L^\infty([0,\infty), \mathbb{R}^N)$, then

$$\lim_{t\to\infty} \tilde{Y}(t) = 0 \in \mathbb{R}^N.$$

We begin the application of the lemma by showing that $\tilde{Y} \in L^2([0,\infty), \mathbb{R}^N)$. Define the shorthand notation $\tilde{g}_t := \tilde{g}(t, \cdot) \in \mathcal{H}_\mathcal{X}$ and consider the quadratic function $V(\tilde{g}_t) = \frac{1}{2}\langle \tilde{g}_t, \tilde{g}_t\rangle_{\mathcal{H}_\mathcal{X}}$. Its derivative with respect to time along the trajectory $\tilde{g}_t$ is

$$
\begin{aligned}
\dot{V}(\tilde{g}_t) &= \frac{1}{2}\langle \dot{\tilde{g}}_t, \tilde{g}_t \rangle_{\mathcal{H}_\mathcal{X}} + \frac{1}{2}\langle \tilde{g}_t, \dot{\tilde{g}}_t \rangle_{\mathcal{H}_\mathcal{X}}, \\
&= \frac{1}{2}\langle -\mathcal{A}(t)\tilde{g}_t, \tilde{g}_t \rangle_{\mathcal{H}_\mathcal{X}} + \frac{1}{2}\langle \tilde{g}_t, -\mathcal{A}(t)\tilde{g}_t \rangle_{\mathcal{H}_\mathcal{X}}, \\
&= -\left\langle \gamma \mathbb{E}^*_{X_t}\mathbb{E}_{X_t}\tilde{g}_t, \tilde{g}_t \right\rangle_{\mathcal{H}_\mathcal{X}} = -\gamma \sum_{i=1}^{N} [\tilde{g}_t(x^i_t)]^2 \le 0.
\end{aligned}
$$

Thus, $V(\tilde{g}_t)$ is non-increasing, bounded above by $\frac{1}{2}\|\tilde{g}_0\|^2$, and bounded below by 0. This implies that the limit $\lim_{t\to\infty} V(\tilde{g}_t) = V_\infty$ exists. Since we have

$$
\begin{aligned}
\lim_{t\to\infty}\int_0^t \frac{d}{dt}V(\tilde{g}(\tau))d\tau &= \lim_{t\to\infty}\int_0^t -\sum_{i=1}^{N}[\tilde{g}(x^i(\tau))]^2 d\tau = V_\infty - V_0 \\
&= -\int_0^\infty \|\tilde{Y}(\tau)\|^2 d\tau,
\end{aligned}
$$

we conclude that $\tilde{Y} \in L^2([0,\infty),\mathbb{R}^N)$. Next, we show that $\tilde{Y} \in L^\infty([0,\infty),\mathbb{R}^N)$. Note that for any $f \in \mathcal{H}_\mathcal{X}$ and $x \in \mathcal{X}$, $E_x f = f(x) = \langle \mathfrak{K}_x, f \rangle$, by the reproducing property of the RKHS. The Cauchy–Schwarz inequality implies that $E_x f \le \|\mathfrak{K}_x\|\|f\|$. By assumption, $\|\mathfrak{K}_x\| \le \overline{k}$ for all $x \in \mathcal{X}$, and thus the evaluation operator E_x is uniformly bounded above by $\overline{k}$. Since $y^i(t) = \tilde{g}_t(x^i(t))$, it follows that

$$
|\tilde{y}^i(t)| \le \|E_{x^i_t}\|\|\tilde{g}_t\|_{\mathcal{H}_\mathcal{X}} \le \overline{k}\|\tilde{g}_0\|_{\mathcal{H}_\mathcal{X}},
$$

where the last inequality follows from the fact that $V(\tilde{g}_t) = \frac{1}{2}\|\tilde{g}_t\|^2$ is non-increasing. Therefore, we conclude that $\tilde{y}^i \in L^\infty([0,\infty),\mathbb{R})$, and hence $\tilde{Y} \in L^\infty([0,\infty),\mathbb{R}^N)$. Lastly, we would like to conclude the $\dot{\tilde{Y}} \in L^\infty([0,\infty),\mathbb{R}^N)$ so that we can apply Barbalat's lemma. This follows if we assume some regularity, as described in the following lemma.

Lemma 1. *Suppose that $\|x^i(t)\| \le c_0$ and $\|\dot{x}^i(t)\| \le c_1$ for all $i = 1, \dots, N$ and that $\mathcal{H}_\mathcal{X} \hookrightarrow C^1(\mathcal{X})$. Then $\dot{\tilde{Y}} \in L^\infty([0,\infty),\mathbb{R}^N)$.*

Proof. By definition

$$
\frac{d}{dt}(\tilde{y}^i(t)) = \sum_{j=1}^{d} \frac{\partial \tilde{g}}{\partial x^i_j}(x^i(t))\dot{x}^i_j(t) + \frac{\partial \tilde{g}}{\partial t}(t, x^i(t)).
$$

When we assume that $\mathcal{H}_\mathcal{X} \hookrightarrow C^1(\mathcal{X})$, we have

$$
\|f\|_{C^1(\mathcal{X})} := \|f\|_{C(\mathcal{X})} + \sum_{j=1}^{d} \left\| \frac{\partial f}{\partial x_j} \right\|_{C(\mathcal{X})} \lesssim \|f\|_{\mathcal{H}_\mathcal{X}},
$$

for any $f \in \mathcal{H}_\mathcal{X}$. This means that there exists a positive constant $\hat{c}_1$ such that

$$
\left| \sum_{j=1}^{d} \frac{\partial \tilde{g}}{\partial x^i_j}(x^i(t))\dot{x}^i_j(t) \right| \le c_1 \sum_{j=1}^{d} \sup_{\|x\|\le c_0} \left| \frac{\partial \tilde{g}}{\partial x^i_j}(x) \right| \le \hat{c}_1 \|\tilde{g}\|_{\mathcal{H}_\mathcal{X}} \le \hat{c}_1 \|\tilde{g}_0\|_{\mathcal{H}_\mathcal{X}},
$$

where the final inequality follows from $V(\tilde{g}_t) = \frac{1}{2}\|\tilde{g}_t\|^2_{\mathcal{H}_\mathcal{X}}$ being non-increasing. From the governing error equation (5), we see that

$$
\left| \frac{\partial \tilde{g}}{\partial t}(t, x^i(t)) \right| = \left| E_{x^i_t} \frac{\partial \tilde{g}}{\partial t}(t, \cdot) \right| \le \overline{c}\|\mathcal{A}(t)\tilde{g}(t)\|_{\mathcal{H}_\mathcal{X}} \le \overline{c}\gamma N \overline{k}^2 \|\tilde{g}_0\|_{\mathcal{H}_\mathcal{X}}.
$$

Combining both upper bounds, we see that $\dot{\tilde{y}}^i \in L^\infty([0,\infty),\mathbb{R})$ and therefore conclude that $\dot{\tilde{Y}} \in L^\infty([0,\infty),\mathbb{R}^N)$. ∎

Having shown $\tilde{Y} \in L^\infty([0,\infty),\mathbb{R}^N) \cap L^2([0,\infty),\mathbb{R}^N)$, and $\dot{\tilde{Y}} \in L^\infty([0,\infty),\mathbb{R}^N)$, we invoke Barbalat's lemma and conclude that the sampling error trajectory $t \mapsto \tilde{Y}(t)$ converges to $0 \in \mathbb{R}^N$ as $t \to \infty$. ∎

3.1.3 | Persistence of excitation

In this section, we turn to the study sufficient conditions, commonly referred to as persistence of excitation (PE) conditions, that ensure the convergence of the estimator error $\tilde{g}(t)$ to $0 \in \mathcal{H}_{\mathcal{X}}$ as $t \to \infty$. PE conditions have a long history in the study of adaptive estimation, particularly in the construction of observers for ODEs that evolve in Euclidean spaces. Again, this topic is covered in detail in standard textbooks like.[40-42,44] The study of the generalization of PE conditions to DPS that are associated with PDEs, which are built using Gelfand triples, can be found in Reference 33 and the references therein. More recently, References 31,32,45, written by a subset of the authors of this article, have introduced and studied PE conditions under a variety of circumstances for the RKHS embedding method. In this article we extend the analysis in References 31,32,45 by introducing weak PE conditions in the spirit of Reference 33, but here in the context of RKHS embedding. The persistency conditions here are assumed to hold in a weak topology, as opposed to the strong form introduced in References 31,32. To draw distinction with the earlier efforts in References 31,32, we begin by introducing simple modifications for multiple trajectories of the two definitions of persistency introduced in these references that are defined for a *single trajectory*.

Definition 1 (S-PE$_1$). The family of trajectories $t \to x^i(t)$ for $i = 1, \dots, N$ persistently excites the space $\mathcal{H}_{\mathcal{X}}$ in the sense of S-PE$_1$ if there exist positive constants Δ, T, and γ such that

$$\int_t^{t+\Delta} \left\langle \mathbb{E}^*_{X_\tau} \mathbb{E}_{X_\tau} g, g \right\rangle_{\mathcal{H}_{\mathcal{X}}} d\tau \geq \gamma \|g\|^2_{\mathcal{H}_{\mathcal{X}}}, \tag{7}$$

for all $t \geq T$ and any $g \in \mathcal{H}_{\mathcal{X}}$.

Alternatively, the following strong form is also introduced there.

Definition 2 (S-PE$_2$). The family of trajectories $t \to x^i(t)$ for $i = 1, \dots, N$ persistently excites the space $\mathcal{H}_{\mathcal{X}}$ in the sense of S-PE$_2$ if there exist positive constants $\Delta, \delta, T, \gamma$ such that for all $t \geq T$ and any $g \in \mathcal{H}_X$, there is an $s \in [t, t+\Delta]$ such that

$$\left| \int_s^{s+\delta} \mathbb{E}_{X_\tau} g d\tau \right| \geq \gamma \, \|g\|_{\mathcal{H}_{\mathcal{X}}}. \tag{8}$$

The above two definitions are closely related and are the infinite dimensional analogs of a few well-known forms of PE in Euclidean spaces introduced in Reference 41. See also Reference 33 for a discussion of generalizations to certain types of PDEs. The condition S-PE$_1$ in Definition 1 always implies the condition S-PE$_2$ in Definition 2. On the other hand, the converse implication requires some additional hypothesis. We refer the readers to Reference 31,32 for a detailed discussion of the proof of convergence when the above PE conditions are valid in a related estimation problem for system identification that uses the RKHS embedding method.

In this article, we explore how the general strategy described in Reference 33 of using PE conditions that hold in a weaker topology can be carried out for the RKHS embedding method in this article. Before we introduce the weak PE condition in an RKHS, we require some preliminary definitions. We define the set $W \subseteq \mathcal{H}_{\mathcal{X}}$ as

$$W = \left\{ g \in \mathcal{H}_{\mathcal{X}} \middle| \lim_{t \to \infty} \left| \int_t^{t+\Delta} \left\langle \mathbb{E}^*_{X_\tau} \mathbb{E}_{X_\tau} g, h \right\rangle_{\mathcal{H}_{\mathcal{X}}} d\tau \right| = 0, \ \forall \, h \in \mathcal{H}_{\mathcal{X}} \text{ and } \Delta > 0 \right\}. \tag{9}$$

For now, we can think of functions in W as not being weakly persistently excited, an interpretation that we will discuss in detail shortly.

Lemma 2. *The set W is a closed subspace of $\mathcal{H}_{\mathcal{X}}$ with respect to the norm of $\mathcal{H}_X$. If the family of trajectories $t \mapsto x^i(t)$ persistently excites the space $\mathcal{H}_{\mathcal{X}}$ in the sense of S-PE$_1$, then $W = \{0\}$.*

Before proceeding to the proof, note that this lemma concludes that $W = \{0\}$ provided $\mathcal{H}_{\mathcal{X}}$ is *strongly* persistently excited. In view of our comments above, this can be interpreted as saying that the only functions that "fail to be weakly persistently excited" is the zero function in this case. In general, the space W can consist of nonzero functions too.

Proof. We first prove the claim that W is a linear subspace. Suppose $f, g \in W$. Then we for any $\alpha, \beta \in \mathbb{R}$ have

$$\left|\int_t^{t+\Delta}\left\langle \mathbb{E}^*_{X_\tau}\mathbb{E}_{X_\tau}(\alpha f+\beta g, h)\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau\right| \leq |\alpha|\left|\int_t^{t+\Delta}\left\langle \mathbb{E}^*_{X_\tau}\mathbb{E}_{X_\tau}f, h\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau\right| + |\beta|\left|\int_t^{t+\Delta}\left\langle \mathbb{E}^*_{X_\tau}\mathbb{E}_{X_\tau}g, h\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau\right| \to 0,$$

as $t \to \infty$ for each $h \in \mathcal{H}_\mathcal{X}$ and $\Delta > 0$.

We next argue that the set W is norm closed for the case when the number of agents is $N = 1$. The general result for $N > 1$ follows similarly. Suppose the sequence $\{g_k\}_{k=1}^\infty \in W$ converges in norm to g. Then we have

$$G_k(\tau) := \left\langle E^*_{x^i_\tau}E_{x^i_\tau}g_k, h\right\rangle_{\mathcal{H}_\mathcal{X}} \to G(\tau) := \left\langle E^*_{x^i_\tau}E_{x^i_\tau}g, h\right\rangle_{\mathcal{H}_\mathcal{X}}, \tag{10}$$

as $k \to \infty$ for each $\tau \in \mathbb{R}^+$. Recall from the analysis of Theorem 2 that the evaluation operator E_x is uniformly bounded above by $\overline{k}$, and note that each function $G_k(\tau)$ is uniformly bounded independent of k since there is a constant $c_g > 0$ such that $\|g_k\|_{\mathcal{H}_\mathcal{X}} \leq c_g$ and hence $|G_k(\tau)| \leq \left\|E^*_{x^i_\tau}\right\|\left\|E_{x^i_\tau}\right\| \|g_k\|_{\mathcal{H}_\mathcal{X}} \|h\|_{\mathcal{H}_\mathcal{X}} \leq \overline{k}^2 c_g \|h\|_{\mathcal{H}_\mathcal{X}}$. By the Lebesgue dominated convergence theorem it follows that

$$\underbrace{\int_t^{t+\Delta}\left\langle E^*_{x^i_\tau}E_{x^i_\tau}g, h\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau}_{F(t)} = \int_t^{t+\Delta}\lim_{k\to\infty}\left\langle g_k, E^*_{x^i_\tau}E_{x^i_\tau}h\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau = \lim_{k\to\infty}\underbrace{\int_t^{t+\Delta}\left\langle g_k, E^*_{x^i_\tau}E_{x^i_\tau}h\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau}_{F_k(t)}.$$

Since each $g_k \in W$, it is trivial to see that for each k, $\lim_{t\to\infty} F_k(t) = 0$. On the other hand, now consider $|F_k(t) - F(t)|$. We have

$$\begin{aligned}|F_k(t) - F(t)| &\leq \left|\int_t^{t+\Delta}\left\langle g_k - g, E^*_{x^i_\tau}E_{x^i_\tau}h\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau\right| \\ &\leq \|g_k - g\|_{\mathcal{H}_\mathcal{X}}\int_t^{t+\Delta}\left\|E^*_{x^i_\tau}\right\|\left\|E_{x^i_\tau}\right\| \|h\|_{\mathcal{H}_\mathcal{X}} d\tau \\ &\leq \overline{k}^2\Delta \|g_k - g\|_{\mathcal{H}_\mathcal{X}} \|h\|_{\mathcal{H}_\mathcal{X}}.\end{aligned}$$

This proves that $F_k(t)$ converges uniformly in k to $F(t)$. Using the Moore-Osgood theorem, we get

$$\lim_{t\to\infty}\left|\int_t^{t+\Delta}\left\langle E^*_{x^i(\tau)}E_{x^i(\tau)}g, h\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau\right| = 0. \tag{11}$$

Since h and Δ were arbitrarily chosen, the above result holds for all $h \in \mathcal{H}_X$ and $\Delta > 0$. This proves that $g \in W$, and hence W is norm closed.

The analysis above in Equations (10) and (11) has been carried out under the assumption that the number agents $N = 1$. But modifications for $N > 1$ simply replaces the operator $E^*_{x^i_\tau}E_{x^i_\tau}$ with $\mathbb{E}^*_{X_\tau}\mathbb{E}_{X_\tau}$. Since the norm $\|\mathbb{E}_{X_\tau}\|$ is also uniformly bounded independent of time τ, all of the steps remain true for teams of agents, and it follows that W is norm closed for $N \geq 1$.

Finally, we prove that the subspace $W = \{0\}$ given the RKHS $\mathcal{H}_\mathcal{X}$ is strongly persistently excited. If the family of trajectories $t \mapsto x^i_\tau$ for $i = 1, \dots, N$ persistently excite the RKHS $\mathcal{H}_\mathcal{X}$ in the sense of S-PE$_1$, then it is clear from the definition of W that $0 \in W$. To see that W contains only the zero function, arbitrarily choose any $g_0 \in \mathcal{H}_\mathcal{X}$ with $g_0 \neq 0$. From the definition, if $\mathcal{H}_\mathcal{X}$ is S-PE$_1$, it follows that there exist positive constants Δ, T, and γ such that

$$\int_t^{t+\Delta}\left\langle \mathbb{E}^*_{X_\tau}\mathbb{E}_{X_\tau}g_0, g_0\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \geq \gamma\|g_0\|^2 > 0,$$

for all $t \geq T$. Hence,

$$\limsup_{t\to\infty}\left|\int_t^{t+\Delta}\left\langle \mathbb{E}_{X_\tau}^*\mathbb{E}_{X_\tau}g_0, g_0\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau\right| \neq 0,$$

implies $g_0 \notin W$. As the choice of $g_0 \neq 0$ was arbitrary, we conclude $W = \{0\}$. ∎

Lemma 2 provides some intuition about how we interpret the subspace W as "functions that are not PE" in a certain weak sense. The utility of this definition is that it can be used to give a description of the asymptotic behavior of the ideal error: we next show in what sense the ideal error is attracted to a particular subset of W. Let $\hat{W} := W \cap \overline{B_r(0)}$, where $\overline{B_r(0)} \subseteq \mathcal{H}_\mathcal{X}$ is the closed ball of radius r centered at 0. We choose r such that $\tilde{g}(t) \in \overline{B_r(0)}\ \forall\, t \geq 0$. Such a constant exists because the Lyapunov function $V(\tilde{g}_t) = \frac{1}{2}\langle \tilde{g}_t, \tilde{g}_t\rangle_{\mathcal{H}_\mathcal{X}}$ is non-increasing (see the proof of Theorem 2). In the theorem below, we show that the ideal error converges to $\hat{W}$ in a certain weak sense.

The proof that follows makes use of the "weak norm" $|\cdot|_w$ on a separable Hilbert space. In general, this norm is defined on the dual space of a separable normed vector space, see the discussion in problem 2.72 of Reference 46 or the proof of theorem I.3.11 in Reference 47. The construction holds for the separable Hilbert space $\mathcal{H}_\mathcal{X}$ using the fact that the Riesz mapping is an isometry from the dual of a Hilbert space to the space itself. Suppose that $\{h_n\}_{n\in\mathbb{N}}$ is a dense set in the separable Hilbert space $\mathcal{H}_\mathcal{X}$, with each $\|h_n\| = 1$ for $n \in \mathbb{N}$. We define the "weak norm"

$$|h|_w := \sum_{n=1}^{\infty} 2^{-n}|\langle h_n, h\rangle_\mathcal{H}|,$$

for each $h \in \mathcal{H}_\mathcal{X}$. To see that this expression defines a bona fide norm on the set $\mathcal{H}_\mathcal{X}$, and it does not depend on the particular choice of the dense set $\{h_n\}_{n\in\mathbb{N}}$, again see Reference 47. Since $|h|_w \leq \|h\|\sum_{n=1}^{\infty} 2^{-n} \leq C\|h\|$, we have the continuous embedding $(\mathcal{H}_\mathcal{X}, \|\cdot\|) \hookrightarrow (\mathcal{H}_\mathcal{X}, |\cdot|_w)$. It can be shown that the topology of $(\mathcal{H}, |\cdot|_w)$ relativized to any norm closed, bounded, convex set coincides with the weak topology of $\mathcal{H}_\mathcal{X}$ relativized to that set. Finally, we define the weak semidistance $d_w(g, \hat{W})$ from $g \in \mathcal{H}_\mathcal{X}$ to W as

$$d_w(g, \hat{W}) := \inf_{h\in\hat{W}} |g - h|_w.$$

In view of the properties of $\hat{W}$, the relative topology $(\hat{W}, |\cdot|_w)$ induced by the weak norm $|\cdot|_w$ on $\hat{W}$ is equivalent to the set $\hat{W}$ equipped with the topology it inherits from the usual weak topology on $\mathcal{H}_\mathcal{X}$, that is, $(\hat{W}, \text{weak}(\mathcal{H}_\mathcal{X}))$.

The asymptotic behavior of global solutions of the ideal error equations are described in the following theorem, which is the primary result of this section.

Theorem 3. *Let $\tilde{g} \in C([0,\infty), \mathcal{H}_\mathcal{X})$ be a solution of Equation 5 on $[0,\infty)$. Then we have*

$$\lim_{t\to\infty} d_w(\tilde{g}(t), \hat{W}) = 0. \tag{12}$$

Proof. As in the last proof, we carry out this proof for the case when $N = 1$ since the case for $N > 1$ follows analogously. We prove this theorem by contradiction. Suppose the implication is not true. Then there exists a sequence $t_k \to \infty$ and a constant $\eta > 0$ such that

$$d_w(\tilde{g}(t_k), \hat{W}) \geq \eta > 0. \tag{13}$$

Since $\overline{B_r(0)}$ is weakly sequentially compact, there exists a subsequence $\{t_l\} \equiv \{t_{k_l}\} \subseteq \{t_k\}$ and $\tilde{g}_\infty \in \overline{B_r(0)}$ such that $\tilde{g}(t_l) \to \tilde{g}_\infty$ weakly in $\mathcal{H}_\mathcal{X}$. But since $\overline{B_r(0)}$ is norm closed and convex, it is weakly closed and $\tilde{g}_\infty \in \overline{B_r(0)}$. Also, $\{\tilde{g}(t_l)\}_{l\in\mathbb{N}}$, is contained in the norm closed, convex, bounded set $\overline{B_r(0)}$, and the weak topology coincides with the $(\mathcal{H}, |\cdot|_w)$ topology over this set. Thus we have $|\tilde{g}(t_l) - \tilde{g}_\infty|_w \to 0$. This result together with Equation (13) imply that $\tilde{g}_\infty \notin W$. Consequently, we know that there exists an $h_0 \in \mathcal{H}_X$ and a positive $\Delta_0 > 0$ such that

$$\lim_{l\to\infty}\left|\int_{t_l}^{t_l+\Delta_0}\left\langle E_{x_\tau^i}^* E_{x_\tau^i}\tilde{g}_\infty, h_0\right\rangle_{\mathcal{H}_\mathcal{X}} d\tau\right| \neq 0.$$

Since

$$\left| \int_{t_l}^{t_l+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}_\infty, h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| \leq \overline{k}^2 \Delta_0 \|\tilde{g}_\infty\|_{\mathcal{H}_\mathcal{X}} \|h_0\|_{\mathcal{H}_\mathcal{X}},$$

we know that

$$\limsup_{l\to\infty} \left| \int_{t_l}^{t_l+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}_\infty, h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| = c_0 < \infty. \tag{14}$$

This implies that there exists a subsequence $\{t_m\} \equiv \{t_{l_m}\} \subseteq \{t_l\}$ and an integer M such that

$$\left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}_\infty, h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| > \frac{c_0}{2}, \quad \forall\, m \geq M. \tag{15}$$

Consider the inequality

$$\begin{aligned} \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}_\infty, h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| &\leq \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} (\tilde{g}_\infty - \tilde{g}(t_m)), h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| \\ &\quad + \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}(t_m), h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right|. \end{aligned}$$

Taking the limsup on both sides, we get

$$\frac{c_0}{2} < \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}(t_m), h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right|. \tag{16}$$

We also know that since $\tilde{y} \in L^2([0,\infty), \mathbb{R})$, we have as $m \to \infty$,

$$\begin{aligned} \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} g, h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| &= \left| \int_{t_m}^{t_m+\Delta_0} \left(\tilde{y}(\tau), E_{x^i_\tau} h_0\right) d\tau \right| \\ &\leq \int_{t_m}^{t_m+\Delta_0} \overline{k} |\tilde{y}(\tau)| \, \|h_0\|_{\mathcal{H}_\mathcal{X}} d\tau \\ &\leq \overline{k} \sqrt{\Delta_0} \|\tilde{y}\|_{L^2([t_m, t_m+\Delta_0), \mathbb{R})} \|h_0\| \to 0. \end{aligned}$$

Similarly, we conclude that

$$\begin{aligned} &\left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}(\tau), h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| - \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}(t_m), h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| \\ &\leq \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} (\tilde{g}(\tau) - \tilde{g}(t_m)), h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| \\ &= \int_{t_m}^{t_m+\Delta_0} (|\tilde{y}(\tau)| + |\tilde{y}(t_m)|) \, |E_{x^i_\tau} h_0| d\tau \\ &\leq \overline{k} \, \|h_0\|_{\mathcal{H}_\mathcal{X}} \left(\sqrt{\Delta} \|\tilde{y}\|_{L^2([t_m, t_m+\Delta_0), \mathbb{R})} + \Delta |\tilde{y}(t_m)| \right) \\ &\to 0 \text{ as } m \to \infty. \end{aligned}$$

Consequently, we have

$$\lim_{m\to\infty} \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}(\tau), h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right| = \lim_{m\to\infty} \left| \int_{t_m}^{t_m+\Delta_0} \left\langle E^*_{x^i_\tau} E_{x^i_\tau} \tilde{g}(t_m), h_0 \right\rangle_{\mathcal{H}_\mathcal{X}} d\tau \right|.$$

Based on the above results, we conclude that

$$\begin{aligned}\lim_{m\to\infty}\left|\int_{t_m}^{t_m+\Delta_0}\left\langle E^*_{x^i_\tau}E_{x^i_\tau}\tilde{g}(t_m),h_0\right\rangle_{\mathcal{H}_\mathcal{X}}d\tau\right| &\leq \lim_{m\to\infty}\left|\int_{t_m}^{t_m+\Delta_0}\left\langle E^*_{x^i_\tau}E_{x^i_\tau}(\tilde{g}(t_m)-\tilde{g}(\tau)),h_0\right\rangle_{\mathcal{H}_\mathcal{X}}d\tau\right| \\ &\quad+\lim_{m\to\infty}\left|\int_{t_m}^{t_m+\Delta_0}\left\langle E^*_{x^i_\tau}E_{x^i_\tau}\tilde{g}(\tau),h_0\right\rangle_{\mathcal{H}_\mathcal{X}}d\tau\right| \\ &\to 0 \text{ as } m\to\infty.\end{aligned}$$

The above statement contradicts the one in (15). Thus, we conclude

$$\lim_{t\to\infty} d_w(\tilde{g}(t),\hat{W}) = 0.$$

as desired. ∎

The analysis above makes clear that, in analogy to the weak PE conditions in Reference 33 and the references therein, the solutions of the ideal error equations in RKHS embedding converges weakly to $\hat{W}$ as $t\to\infty$. We say that the space $\mathcal{H}_\mathcal{X}$ is weakly persistently excited whenever $\hat{W}\equiv\{0\}$. The following lemma makes clear that while the notion of weak persistence might at first seem rather abstract, it has some practical implications that are quite intuitive.

Corollary 1. *If the family of trajectories $t\to x^i(t)$, with $i=1,\dots,N$, weakly persistently excites the space $\mathcal{H}_\mathcal{X}$, the pointwise error converges,*

$$\lim_{t\to\infty}\tilde{g}(t,x) = 0\in\mathbb{R},$$

for all $x\in\mathcal{X}$.

Proof. By assumption, we have $\hat{W}=\{0\}$, which implies that

$$\lim_{t\to\infty}|\tilde{g}(t)|_w = 0.$$

We know that the ideal error $\tilde{g}(t,\cdot)\in\mathcal{H}_\mathcal{X}$ lies in $\overline{B_r(0)}$ for a suitably chosen $r>0$ defined from the Lyapunov analysis in the proof of Theorem 2. But the topology that $\overline{B_r(0)}$ inherits as subset of $(\mathcal{H}_\mathcal{X},|\cdot|_w)$ is equivalent to the weak topology on $\overline{B_r(0)}$. This means that we have the weak convergence of $\tilde{g}(t)$ to $0\in\mathcal{H}_\mathcal{X}$ as $t\to\infty$. Since $\mathfrak{K}_x\in\mathcal{H}_\mathcal{X}$ for all $x\in\mathcal{X}$, by definition of weak convergence, we have $0=\lim_{t\to\infty}\langle\tilde{g}(t),k_x\rangle_{\mathcal{H}_\mathcal{X}}=\tilde{g}(t,x)$. ∎

3.2 | Finite dimensional approximations

Here we consider a few different ways to approximate the solutions $\hat{g}(t)$ of the ideal, infinite dimensional governing system in Equations (4) and (5). It is important to observe that an approximation of the ideal estimation process represented in Equation (4) requires two distinct contributions: a discretization of the continuous evolution law in time as well as the definition of a suitable finite dimensional approximant at each discrete time. To understand how the final centralized consensus estimate will be executed, we start by introducing an explicit linear-multistep integration scheme in the Hilbert space $\mathcal{H}_\mathcal{X}$. The explicit linear multi-step discrete integrator of order q for the estimator (4) takes the form

$$\begin{aligned}\hat{g}_{k+1} &= \hat{g}_k + h\gamma\sum_{s=1}^{q}a_s\mathbb{E}^*_{X_{k-s}}(Y_{k-s}-\mathbb{E}_{X_{k-s}}\hat{g}_{k-s}) \\ &= \hat{g}_k + h\gamma\sum_{s=1}^{q}a_s\sum_{i=1}^{N}\mathfrak{K}(x^i_{k-s},\cdot)(y^i_{k-s}-\hat{g}_{k-s}(x^i_{k-s}))\in\mathcal{H}_\mathcal{X},\end{aligned} \tag{17}$$

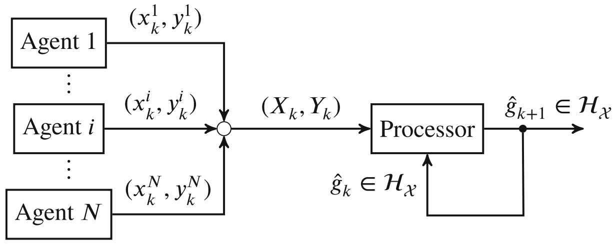

FIGURE 1 At each time index $k := t_k$, all agents communicate their samples to a central processor that, in principle, updates the centralized estimator $\hat{g}_k \in \mathcal{H}_\mathcal{X}$.

where h denotes the timestep, and the collection of coefficients $\mathbf{a} = \{a_s\}_{s=1}^q$ is determined by the particular multi-step method used. For example, if the Adams-Bashforth method with $q = 3$ is used then the set of coefficients is $\mathbf{a} = \frac{1}{12}\{23, -16, 5\}$. For more information on linear multistep methods, we refer the reader to Butcher.[48(ch. 2.24)] Note that the timestep h can vary and need not be held constant across all time indices.

By definition the recursion takes place in the (generally infinite dimensional) space $\mathcal{H}_\mathcal{X}$. The recursion scheme relies on past state-observation pairs, as well as, the past estimators. In principle, an implementation of the centralized estimator (17) can be carried out as follows. At each time step $k = 1, 2, \ldots$, every agent i of the team collects a sample and communicates the state-sample pair (x_k^i, y_k^i) to a central processor, which uses the collection of state-sample pairs from all the agents to update the estimator $\hat{g}_k$ as depicted in Figure 1.

Although the recursion in Equation (17) is closer to a realizable algorithm, it still remains ideal at best. The recursion defines the function $\hat{g}_{k+1}$ at the time step t_{k+1} in terms of the previous estimators $\hat{g}_k, \ldots, \hat{g}_{k-q} \subset \mathcal{H}_\mathcal{X}$, the previous samples $y_k^i, \ldots, y_{k-q}^i$, the previous states $x_k^i, \ldots, x_{k-q}^i \subset \mathcal{X}$, and the kernel functions $\mathfrak{K}(x_k^i, \cdot), \ldots, \mathfrak{K}(x_{k-q}^i, \cdot) \subset \mathcal{H}_\mathcal{X}$. Practical considerations, such as storage and numerical stability, require that we further project the functions of these equations onto some finite dimensional space $V_L := \text{span}\{v_{L,\ell} \in \mathcal{H}_\mathcal{X} \mid 1 \le \ell \le L\}$, where L denotes the number of basis elements for the subspace $V_L \subset \mathcal{H}_\mathcal{X}$. Generally, this is accomplished in terms of a family $\{P_L\}_{L\in\mathbb{N}}$ of uniformly bounded projection operators $P_L : \mathcal{H}_\mathcal{X} \to V_L$, which enable the discrete ideal law in Equation (17) to be written as

$$\begin{aligned} P_L\hat{g}_{k+1} &= P_L\left(\hat{g}_k + h\gamma\sum_{s=1}^q a_s \mathbb{E}^*_{X_{k-s}}(Y_{k-s} - \mathbb{E}_{X_{k-s}}\hat{g}_{k-s}) \in \mathcal{H}_\mathcal{X}\right) \\ &= P_L\left(\hat{g}_k + h\gamma\sum_{s=1}^q a_s \sum_{i=1}^N \mathfrak{K}(x_{k-s}^i, \cdot)(y_{k-s}^i - \hat{g}_{k-s}(x_{k-s}^i)) \in \mathcal{H}_\mathcal{X}\right). \end{aligned} \tag{18}$$

In the next section we describe some concrete ways that the finite dimensional space V_L will be chosen that depend on bases constructed from agent samples. The specific choice of spaces V_L and bases $\{v_{L,\ell}\}_{1\le\ell\le L}$ generate the final Equation (27) below that are suitable for implementation. In the next section, we detail how these bases and their spaces of approximants are constructed.

3.2.1 | Samples and history dependent bases

In this article we denote by Ξ_L^i the set of L samples collected, from the discrete initial time $\ell = 0$ up through the discrete time step $\ell = L - 1$, in the subdomain Ω^i by agent i for $i = 1, \ldots, N$ at times $\mathbb{T}_L := \{t_{L,\ell} \in \mathbb{R}^+ \mid 0 \le \ell \le L - 1\}$,

$$\Xi_L^i := \left\{\xi_{L,\ell} := x^i(t_{L,\ell}),\quad t_{L,\ell} \in \mathbb{T}_L\right\} \subset \Omega^i \subset X.$$

The entirety of samples collected by agent i is then given by $\Xi^i := \bigcup_{L\ge0}\Xi_L^i$, and it is further assumed that this collection is dense in the subset Ω^i, that is, $\Omega^i \subset \overline{\Xi^i}$. Correspondingly, we define the collection of all samples collected through discrete time t_{L-1} by all the agents as $\Xi_L := \bigcup_{i=1}^N \Xi_L^i$. We define Ω to be the set that is sampled by all the agents, which is given by

$$\Omega = \bigcup_{i=1}^{N} \Omega^i \subseteq X.$$

The full collection of samples that are collected over $[0, \infty)$ by all the agents is then $\Xi := \bigcup_{0 \le L < \infty} \Pi_L$. Clearly, it follows that $\Omega \subset \overline{\Xi}$. Associated with the sample sets Ξ_L^i and Ξ_L, we define the associated finite dimensional spaces H_L^i and H_L, respectively, that will be used to build approximants:

$$\begin{aligned} \mathcal{H}_L^i &:= \text{span}\left\{ \mathfrak{K}(\xi_{L,\ell}, \cdot) \mid \xi_{L,\ell}^i \in \Xi_L^i,\ 0 \le \ell \le L-1 \} \right\} \\ \mathcal{H}_L &:= \text{span}\left\{ \mathfrak{K}(\xi_{L,\ell}^i, \cdot) \mid \xi_{L,\ell}^i \in \Xi_L^i,\ 0 \le \ell \le L-1,\ 1 \le i \le N \right\}. \end{aligned}$$

We define Π_L and Π_L^i to be the $\mathcal{H}_\mathcal{X}$-orthogonal projections onto $\mathcal{H}_L$ and $\mathcal{H}_L^i$, respectively. The rates of convergence of the estimates will be studied by examining how the finite dimensional subspaces $\mathcal{H}_L$ and $\mathcal{H}_L^i$ approximate, respectively, the closed subspaces

$$\begin{aligned} \mathcal{H}_\Omega &:= \overline{\text{span}\{\mathfrak{K}(\xi, \cdot) \mid \xi \in \Omega\}} \subset \mathcal{H}_\mathcal{X}, \\ \mathcal{H}_{\Omega^i} &:= \overline{\text{span}\{\mathfrak{K}(\xi, \cdot) \mid \xi \in \Omega^i\}} \subset \mathcal{H}_\mathcal{X}. \end{aligned}$$

In the next section we describe how these bases are used in the construction of the finite dimensional approximation $\hat{g}_L(t) \in \mathcal{H}_L$ of the ideal estimate $\hat{g}(t) \in \mathcal{H}_\mathcal{X}$.

3.2.2 | Convergence of the FD, centralized estimator

We define the estimate $\hat{g}_L(t) \in \mathcal{H}_L$ as the solution of the evolution law

$$\dot{\hat{g}}_L(t) = \Pi_L[\mathcal{A}(t)\tilde{g}_L(t)] = \gamma \Pi_L \mathbb{E}_{X_t}^*(Y(t) - \mathbb{E}_{X_t}\hat{g}_L(t)), \tag{19}$$

where $\tilde{g}_L(t) = g - \hat{g}_L(t) \in \mathcal{H}_\mathcal{X}$. In the following theorem, we derive an expression that bounds the approximation error $\overline{g}_L(t) = \hat{g}(t) - \hat{g}_L(t)$. It will be used subsequently to prove that $\overline{g}_L(t) \to 0$ as $L \to \infty$ for each time t.

Theorem 4. *Let $\hat{g}$ be the solution of the infinite dimensional estimator Equation (4) and define the approximation error $\overline{g}_L(t) := \hat{g}(t) - \hat{g}_L(t)$. The approximation error is bounded by the expression*

$$\|\overline{g}_L(t)\|_{\mathcal{H}_\mathcal{X}}^2 \le \left(\|Q_L \hat{g}_0\|_{\mathcal{H}_\mathcal{X}}^2 + \gamma \left(\max_{1 \le i \le N} \sup_{x \in \Omega^i} \|Q_L \mathfrak{K}(x, \cdot)\|_{\mathcal{H}_\mathcal{X}}^2 \right) \|\tilde{Y}\|_{L^2((0,t),\mathbb{R}^N)}^2 \right) e^{ct} \tag{20}$$

with $c := 1 + 2\gamma N \overline{k}^2$ and $Q_L := I - \Pi_L$.

Proof. Observe that the approximation error $\overline{g}_L(t) := \hat{g}(t) - \hat{g}_L(t) = \tilde{g}_L(t) - \tilde{g}(t)$ propagates in time according to

$$\begin{aligned} \dot{\overline{g}}_L(t) &= \mathcal{A}(t)\tilde{g}(t) - \Pi_L[\mathcal{A}(t)\tilde{g}_L(t)] \\ &= \underbrace{\mathcal{A}(t)\tilde{g}(t) - \Pi_L[\mathcal{A}(t)\tilde{g}(t)]}_{(I-\Pi_L)\mathcal{A}(t)\tilde{g}(t)} + \underbrace{\Pi_L[\mathcal{A}(t)\tilde{g}(t)] - \Pi_L[\mathcal{A}(t)\tilde{g}_L(t)]}_{-\Pi_L\mathcal{A}(t)\overline{g}(t)}. \end{aligned} \tag{21}$$

Denote $Q_L := I - \Pi_L$. Then note that

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\langle \overline{g}_L(t), \overline{g}_L(t) \rangle_{\mathcal{H}_\mathcal{X}} &= \left\langle (I - \Pi_L)\mathcal{A}(t)\tilde{g}(t), \overline{g}(t) \right\rangle_{\mathcal{H}_\mathcal{X}} - \left\langle \Pi_L \mathcal{A}(t)\overline{g}(t), \overline{g}(t) \right\rangle_{\mathcal{H}_\mathcal{X}} \\ &\le \|Q_L \mathcal{A}(t)\tilde{g}(t)\|_{\mathcal{H}_\mathcal{X}} \|\overline{g}(t)\|_{\mathcal{H}_\mathcal{X}} + \|\Pi_L \mathcal{A}(t)\| \|\overline{g}_L(t)\|_{\mathcal{H}_\mathcal{X}}^2 \\ &\le \frac{1}{2}\|Q_L \mathcal{A}(t)\tilde{g}(t)\|_{\mathcal{H}_\mathcal{X}}^2 + \frac{1}{2}\|\overline{g}(t)\|_{\mathcal{H}_\mathcal{X}}^2 + n\overline{k}^2 \|\overline{g}_L(t)\|_{\mathcal{H}_\mathcal{X}}^2. \end{aligned}$$

Let $c := 1 + 2N\overline{k}^2$. Next, we integrate both sides from 0 to t to find that

$$\|\overline{g}_L(t)\|^2_{\mathcal{H}_\mathcal{X}} \leq \|\overline{g}_L(0)\|^2_{\mathcal{H}_\mathcal{X}} + \int_0^t \|Q_L\mathcal{A}(\tau)\tilde{g}(\tau)\|^2_{\mathcal{H}_\mathcal{X}} d\tau + c\int_0^t \|\overline{g}_L(\tau)\|^2_{\mathcal{H}_\mathcal{X}} d\tau.$$

By the assumption $\hat{g}_L(0) = \Pi_L\hat{g}(0)$, we have $\overline{g}_L(0) = Q_L\hat{g}(0)$. Then we apply Grönwall's inequality to obtain

$$\|\overline{g}_L(t)\|^2_{\mathcal{H}_\mathcal{X}} \leq \left(\|Q_L\hat{g}(0)\|^2_{\mathcal{H}_\mathcal{X}} + \int_0^t \|Q_L\mathcal{A}(\tau)\tilde{g}(\tau)\|^2_{\mathcal{H}_\mathcal{X}} d\tau \right) e^{ct}.$$

Next we focus on the rightmost term in the above inequality. We bound this term by writing

$$\begin{aligned}
\int_0^t \|Q_L\mathcal{A}(\tau)\tilde{g}(\tau)\|^2_{\mathcal{H}_\mathcal{X}} d\tau &= \gamma \int_0^t \left\|Q_L\mathbb{E}^*_{X_t}\tilde{Y}(\tau)\right\|^2_{\mathcal{H}_\mathcal{X}} d\tau \\
&\leq \gamma \int_0^t \sum_{i=1}^N \left\|Q_L\mathfrak{K}(x^i_\tau, \cdot)\tilde{y}^i(\tau)\right\|^2_{\mathcal{H}_\mathcal{X}} d\tau \\
&\leq \gamma \max_{1\leq i\leq N} \sup_{x\in\Omega^i} \|Q_L\mathfrak{K}(x,\cdot)\|^2_{\mathcal{H}_\mathcal{X}} \int_0^t \sum_{i=1}^N |\tilde{y}^i(\tau)|^2 d\tau.
\end{aligned}$$

This completes the proof of the theorem. ■

Our first result concerning the rate of convergence of approximations makes use of some standard theory concerning the interpolants in native spaces as summarized in chapter 11 of Wendland.[35] We now assume that the kernel $\mathfrak{K}(x,y) := r(\|x-y\|_2)$ for all $x, y \in \Omega \subseteq X$ with $r : \mathbb{R}^+ \to \mathbb{R}^+$ a radial basis function. We define the power function $P_{\mathfrak{K},Z} : \Omega \to \mathbb{R}^+$ for the kernel $\mathfrak{K}$ (see definition 11.2 in Reference 35) and a finite set $Z \subset \Omega$ having $J = \#(Z)$ points as

$$P_{\mathfrak{K},Z}(x) := \min_{\alpha\in\mathbb{R}^J} \left\| \mathfrak{K}(x,\cdot) - \sum_{j=1}^J \alpha_j \mathfrak{K}(z_j,\cdot) \right\|_{\mathcal{H}_\mathcal{X}}.$$

But we know by definition that

$$\|Q_L\mathfrak{K}(x,\cdot)\|_{\mathcal{H}_\mathcal{X}} = \|(I-\Pi_L)\mathfrak{K}(x,\cdot)\|_{\mathcal{H}_\mathcal{X}} = \min_{\alpha\in\mathbb{R}^L} \left\| \mathfrak{K}(x,\cdot) - \sum_{\ell=1}^L \alpha_\ell \mathfrak{K}(\xi_{L,\ell},\cdot) \right\|_{\mathcal{H}_\mathcal{X}},$$

and therefore we have

$$\sup_{x\in\Omega^i} \|Q_L\mathfrak{K}(x,\cdot)\|_{\mathcal{H}_\mathcal{X}} = \sup_{x\in\Omega^i} \|P_{\mathfrak{K},\Xi_L}\|_{\mathcal{H}_\mathcal{X}}. \tag{22}$$

In other words, the rightmost term in Equation (20) can be expressed in terms of the power function $P_{\mathfrak{K},\Omega_L}$ of the kernel $\mathfrak{K}$ and the samples by all the agents. The utility of making this identification is that specialists have carefully studied how to bound the power function for a wide variety of choices of kernel $\mathfrak{K}$ in terms of the fill distance. Recall that, for a finite set $Z \subset A$ of $J = \#(Z)$ points, the fill distance of Z in $A \subseteq X$ is

$$h_{Z,A} := \sup_{x\in A} \min_{z\in Z} d_\mathcal{X}(x,z).$$

We will need the global and local fill distances, defined respectively as

$$\begin{aligned}
h_{\Xi_L,\Omega} &:= \sup_{x\in\Omega} \min_{\xi\in\Xi_L} d_\mathcal{X}(x,\xi), \\
h_{\Xi^i_L,\Omega^i} &:= \sup_{x\in\Omega^i} \min_{\xi\in\Xi^i_L} d_\mathcal{X}(x,\xi).
\end{aligned}$$

Table 11.1 in Reference 35 summarizes how the power function $P_{\mathfrak{K},Z}$ for a finite subset $Z \subset A \subseteq X$ and a particular kernel $\mathfrak{K}$ is bounded in terms of a function F as in the expression

$$P^2_{\mathfrak{K},Z}(x) \lesssim F(h_{Z,A}). \tag{23}$$

The appropriate function F is given in the table for Gaussian, multiquadric, inverse multiquadric, thin plate splines, and Wendland compactly supported kernel functions. For example, when using thin plate splines with $r(\xi) := (-1)^{s+1}\xi^{2s}\log(\xi)$ and $s \in \mathbb{N}$, we have

$$P^2_{\mathfrak{K},\Xi^i_L}(x) \lesssim h^{2s}_{\Xi^i_L,\Omega^i}.$$

For the compactly supported Wendland kernels $r_{d,s}(\xi)$, it is known that

$$P^2_{\mathfrak{K},\Xi^i_L}(x) \lesssim h^{2s+1}_{\Xi^i_L,\Omega^i}.$$

Other expressions for F for the remaining types of kernels listed above can be found in Table 11.1 in Reference 35.

One of the important properties of the power function is that it enables pointwise bounds on projection error. By Theorem 11.4 of Reference 35, if Ω is open and the kernel $\mathfrak{K} \in C(\Omega \times \Omega)$, we have

$$|((I-\Pi_L)f)(x)| \le P_{\mathfrak{K},\Xi_L}(x)\|f\|_{\mathcal{H}_\mathcal{X}},$$

for all $x \in \Omega$ and $f \in \mathcal{H}_\mathcal{X}$. In fact, it is possible to similarly establish uniform bounds on the derivatives of the function f in terms of power functions.[35]

As a last preparation for the derivation of a rate of convergence in Corollary 2 below, we recall a standard mechanism for improving error estimates for smooth functions. Since we assume that the kernel $\mathfrak{K} : \Omega \times \Omega \to \mathbb{R}$ is positive definite, we can introduce an integral operator $T : L^2(\Omega,\mathbb{R}) \to \mathcal{H}_\mathcal{X}$ defined as

$$(Tf)(x) = \int_\Omega \mathfrak{K}(x,y)f(y)dy.$$

Intuitively, we can think of the integral operator as smoothing functions in $L^2(\Omega,\mathbb{R})$ to obtain functions contained in $\mathcal{H}_\mathcal{X}$. This operator enables a bound on $\|(I-\Pi_L)f\|_{\mathcal{H}_\mathcal{X}}$ for any $f \in T(L^2(\Omega,\mathbb{R})) \subseteq \mathcal{H}_\mathcal{X}$. Suppose $f = T(h)$ for some $h \in L^2(\Omega,\mathbb{R})$. By proposition 10.28 in Reference 35, we have that

$$\begin{aligned}\|(I-\Pi_L)f\|^2_{\mathcal{H}_\mathcal{X}} &= \langle (I-\Pi_L)f, f\rangle_{\mathcal{H}_\mathcal{X}} = \langle (I-\Pi_L)f, Th\rangle_{\mathcal{H}_\mathcal{X}} \\ &= ((I-\Pi_L)f, h)_{L^2(\Omega,\mathbb{R})} \le \|(I-\Pi_L)f\|_{L^2(\Omega,\mathbb{R})}\|T^{-1}f\|_{L^2(\Omega,\mathbb{R})} \\ &\le \|P_{\Xi_L,\Omega}\|_{L^2(\Omega,\mathbb{R})}\,\|(I-\Pi_L)f\|_{\mathcal{H}_\mathcal{X}}\,\|T^{-1}f\|_{L^2(\Omega,\mathbb{R})}.\end{aligned}$$

It follows, then, that $\|(I-\Pi_L)f\|_{\mathcal{H}_\mathcal{X}} \le \|P_{\Xi_L,\Omega}\|_{L^2(\Omega,\mathbb{R})}\|T^{-1}f\|_{L^2(\Omega)}$ whenever $f \in T(L^2(\Omega,\mathbb{R}))$.

We use these observations to derive a rate of convergence for the bound summarized in Theorem 4.

Corollary 2. *Suppose that the power functions $P_{\mathfrak{K},\Xi^i_L}$ for $i = 1,\dots,N$ and $P_{\mathfrak{K},\Xi_L}$ are bounded as in Equation (23) with $F(r) := r^p$ and $p \in \mathbb{N}$, that the domains Ω^i for $i = 1,\dots,N$ and Ω are open, the kernel $\mathfrak{K} \in C(\Omega\times\Omega)$ is positive definite, and the initial condition $\hat{g}_0 \in T(L^2(\Omega,\mathbb{R}))$. Then there is a constant $C > 0$ such that*

$$\|\overline{g}_L(t)\|^2_{\mathcal{H}_\mathcal{X}} \le C\left(1+\gamma\|\tilde{Y}\|^2_{L^2((0,t),\mathbb{R}^N)}\right)\max_{1\le i\le N} h^{2p}_{\Xi^i_L,\Omega^i},$$

for $t \in [0,T]$.

Proof. First, we note that by the observations above, we have

$$\begin{aligned}\|(I-\Pi_L)\hat{g}_0\|_{\mathcal{H}_\mathcal{X}} &\le \left(\int_\Omega P^2_{\mathfrak{K},\Xi_L}(\xi)d\xi\right)^{1/2}\|T^{-1}\hat{g}_0\|_{L^2(\Omega,\mathbb{R})} \\ &\lesssim h^p_{\Xi_L,\Omega}|\Omega|^{1/2}\|T^{-1}\hat{g}_0\|_{L^2(\Omega,\mathbb{R})}\end{aligned} \tag{24}$$

with $|\Omega|$ the measure of Ω. However, from the definition of the fill distance we have

$$\begin{aligned} h_{\Xi_L,\Omega} &= \sup_{x\in\Omega} \min_{\xi\in\Xi_L} d_{\mathcal{X}}(x,\xi), \\ &= \max_{1\le i\le N} \sup_{x\in\Omega^i} \min_{\xi\in\Xi_L} d_{\mathcal{X}}(x,\xi), \\ &\le \max_{1\le i\le N} \sup_{x\in\Omega^i} \min_{\xi\in\Xi_L^i} d_{\mathcal{X}}(x,\xi) = \max_{1\le i\le N} h_{\Xi_L^i,\Omega}. \end{aligned} \tag{25}$$

The conclusion of the corollary now follows from Equation (20) in Theorem 4 when we substitute the above bounds for $\|(I-\Pi_L)\hat{g}_0\|_{\mathcal{H}_{\mathcal{X}}}$ in Equation (24), the bound for $P_{\mathfrak{K},\Xi_L^i}$ in Equations (22) and (25). ■

3.2.3 | Implementation of the centralized estimation algorithm

In this section we discuss the implementation details of the finite dimensional approximation embodied in the evolution law

$$\begin{aligned} \dot{\hat{g}}_L(t) &= \Pi_L\left(\mathcal{A}(t)\tilde{g}_L(t)\right) = \gamma\Pi_L\mathbb{E}^*_{X_t}\left(Y(t)-\mathbb{E}_{X_t}\hat{g}_L(t)\right) \\ &= \gamma\Pi_L\sum_{i=1}^{N}\mathfrak{K}(x_t^i,\cdot)\left(y^i(t)-E_{x_t^i}\hat{g}_L(t)\right). \end{aligned} \tag{26}$$

Since by definition we have the expansion

$$\hat{g}_L(t) := \sum_{\ell=1}^{L}\alpha_{L,\ell}(t)\mathfrak{K}(\xi_{L,\ell},\cdot),$$

taking the inner product of the above equations with an arbitrary function $\mathfrak{K}(\xi_{L,j},\cdot)\in\mathcal{H}_L$ yields the system of ODEs that govern the coefficients $\{\alpha_{L,\ell}(t)\}_{\ell=1}^{L}$. For each $j=1,\ \dots\ ,L$ we obtain

$$\begin{aligned} &\sum_{\ell=1}^{L}\left\langle\mathfrak{K}(\xi_{L,j},\cdot),\mathfrak{K}(\xi_{L,\ell},\cdot)\right\rangle_{\mathcal{H}_{\mathcal{X}}}\dot{\hat{\alpha}}_{L,\ell} \\ &\quad = \gamma\sum_{i=1}^{N}\left\langle\mathfrak{K}(\xi_{L,j},\cdot),\mathfrak{K}(x_t^i,\cdot)\right\rangle_{\mathcal{H}_{\mathcal{X}}}\left(y^i(t)-\sum_{\ell=1}^{L}\left\langle\mathfrak{K}(\xi_{L,j},\cdot),\mathfrak{K}(x_t^i,\cdot)\right\rangle_{\mathcal{H}_{\mathcal{X}}}\alpha_{L,\ell}(t)\right) \\ &\sum_{\ell=1}^{L}\mathfrak{K}\left(\xi_{L,j},\xi_{L,\ell}\right)\dot{\hat{\alpha}}_{L,\ell} = \gamma\sum_{i=1}^{N}\mathfrak{K}\left(\xi_{L,j},x_t^i\right)\left(y^i(t)-\sum_{\ell=1}^{L}\mathfrak{K}\left(\xi_{L,j},x_t^i\right)\alpha_{L,\ell}(t)\right). \end{aligned}$$

We introduce the Grammian matrices

$$\begin{aligned} \mathbb{K}(\Xi_L,\Xi_L) &= [\mathfrak{K}(\xi_{L,j},\xi_{L,\ell})]\in\mathbb{R}^{L\times L}, \\ \mathbb{K}(\Xi_L,X_t) &= [\mathfrak{K}(\xi_{L,j},x_t^i)]\in\mathbb{R}^{L\times N}, \end{aligned}$$

and the ODEs are written in the succinct form

$$\dot{\hat{\alpha}}_L = \gamma\mathbb{K}^{-1}(\Xi_L,\Xi_L)\mathbb{K}(\Xi_L,X_t)\left(Y(t)-\mathbb{K}^T(\Xi_L,X_t)\hat{\alpha}_L(t)\right),$$

with $\hat{\alpha}_L := (\hat{\alpha}_{L,1},\ \dots\ ,\hat{\alpha}_{L,L})^T$. These equations can be integrated forward in time using any standard discrete integration method. For a linear multistep method of the type described in Equations (17) or (18), the final discrete evolution law is given by

$$\hat{\alpha}_{L,k+1} = \hat{\alpha}_{L,k} + h\gamma \sum_{s=1}^{q} a_s \mathbb{K}(\Xi_L, \Xi_L)^{-1} \mathbb{K}(\Xi_L, X_{k-s}) \left(Y_{k-s} - \mathbb{K}^T(\Xi_L, X_{k-s}) \hat{\alpha}_{L,k-s} \right). \tag{27}$$

Inspection of the above final, discrete evolution law makes clear that it depends on a fixed basis of $\mathcal{H}_L$ that depends on the samples defined in Ξ_L. Also, the rate of convergence at which $\hat{g}_L(t) \to \hat{g}(t)$ as $L \to \infty$ in Corollary 2 is also stated when the basis for $\mathcal{H}_L$ is held fixed for a given time interval. In practice, it is anticipated that the discrete estimation law above will be implemented over blocks of time during which the basis is held fixed, but the basis will be updated or enriched over time according to some strategy. This process of enrichment, detailed in the next paragraph, has been carried out in the examples in Section 4.

See Algorithm 1 for the implementation details of how the finite dimensional approximation is realized in practice. Before collecting any samples, a kernel $\mathfrak{K}$, together with any of its associated hyperparameters θ_g, must be specified. The agents then use their initial samples to form the estimator basis and coefficients. At each subsequent time index $k \in \mathbb{N}$, the team of agents sample the spatial field to obtain (X_k, Y_k). With this new set of samples in hand, we update the prediction coefficients using the discrete-time evolution law (27). Then, we iterate through the sampling location $x_k^i \in X_k$ for each $i = 1 : N$ and compute the novelty of x_k^i relative to Ξ_L to determine whether or not to add x_k^i to the basis set. As done in References 1,49,50, novelty of x_n^i, relative to Ξ_L, can be made precise in terms of the squared norm of the residual which arises from the orthogonal projection of $\mathfrak{K}_{x_k^i}$ onto $\mathcal{H}_L$ given by

$$\epsilon_k^i = \mathfrak{K}(x_k^i, x_k^i) - \mathbb{K}(x_k^i, \Xi_L)\mathbb{K}(\Xi_L, \Xi_L)^{-1}\mathbb{K}(\Xi_L, x_k^i). \tag{28}$$

Thus, if ϵ_k^i is greater than some user-defined threshold $\bar{\epsilon} \geq 0$, then we add x_k^i to Ξ_L. Given that the dimension of the subspace $\mathcal{H}_L$ has increased by one, the prediction coefficients must also increase in dimension. Letting $\mathbf{g} := g(\Xi_L)$ denote all the samples at the basis centers, we initialize the prediction coefficients such that $\alpha_{k+1} = \mathbb{K}(\Xi_L, \Xi_L)^{-1}\mathbf{g}$ whenever a new basis element is added. That is, we initialize the coefficients so as to interpolate g at the locations contained in Ξ.

For a given sampling iteration k, the computational complexity incurred to update the coefficients $\{\alpha\}_{\ell=1}^{L}$ is $\mathcal{O}(NL^2)$ provided the inverse $\mathbb{K}(\Xi_L, \Xi_L)^{-1}$ is cached in memory. If a new basis element is added, $\mathbb{K}(\Xi_L, \Xi_L)^{-1}$ must be updated. Using a rank-1 update, based on the Sherman–Morrison matrix inverse identity, this incurs a computational cost $\mathcal{O}(L^2)$; see Reference 51 for more details.

Algorithm 1. Centralized consensus estimation implementation

input: Estimator kernel $\mathfrak{K}$, estimator kernel hyperparameters $\theta_{\hat{g}}$, and novelty threshold $\bar{\epsilon}>0$
Agents collect initial samples: X_0, Y_0
Initialize basis and collection of associated samples: $\Xi_L = X_0$; $\mathbf{g} = Y_0$
Initialize prediction coefficients: $\alpha_1 = \mathbb{K}(\Xi_L, \Xi_L)^{-1}\mathbf{g}$
while Collecting samples **do**
  Sample spatial field: (X_k, Y_k)
  Update prediction coefficients via Equation (27): $\alpha_k \to \alpha_{k+1}$
  for $i = 1 : N$ **do**
    Compute novelty of x_k^i relative to Ξ_L via Equation (28): ϵ
    if $\epsilon > \bar{\epsilon}$ **then**
      Add basis element: $\{\Xi_L \cup \{x_k^i\}\} \to \Xi_L$; $\{\mathbf{g} \cup \{y_k^i\}\} \to \mathbf{g}$
      Initialize coefficients: $\alpha_{k+1} = \mathbb{K}(\Xi, \Xi)^{-1}\mathbf{g}$
    end if
  end for
  $k + 1 \to k$
end while

4 | NUMERICAL RESULTS

To illustrate the qualitative behavior of the finite-dimensional estimate $\hat{g}_L$, we choose the unknown function $g : \mathcal{X} \to \mathbb{R}$ that is an element of the RKHS $\mathcal{H}_\mathcal{X}$ induced by the 5/2 Matern kernel

$$\mathfrak{K}(x_i, x_j) = \sigma^2 \left(1 + \sqrt{5}r + \frac{5r^2}{3}\right) \exp(-\sqrt{5}r), \quad r = \frac{\|x_i - x_j\|_{\mathbb{R}^d}}{\ell},$$

with the hyperparameters $\theta_g = [\sigma, \ell] = [1, 1]$. The unknown function g is chosen to be a realization of a Gaussian process. We generate $g \in \mathcal{H}_\mathcal{X}$ by discretizing the domain $\mathcal{X} \equiv [0, 10] \times [0, 10]$ into a grid $\Xi_g \subset \mathcal{X}$ with a resolution of 0.25 units, and we write $g : \mathcal{X} \to \mathbb{R}$ as

$$g(\cdot) = \sum_{x_i \in \Xi_g} a_i \mathfrak{K}(x_i, \cdot),$$

where a_i is the ith element of $\mathbf{a} = \mathbb{K}(\Xi_g, \Xi_g)^{-1}\mathbf{g}$, and $\mathbf{g}$ is a vector that contains a particular realization over Ξ_g of a zero-mean Gaussian distribution whose covariance matrix is $\mathbb{K}(\Xi_g, \Xi_g)$. See Figure 2 for a visualization of g. We fix the realization g and use it for all subsequent numerical convergence studies.

The first set of simulations are designed so as to demonstrate that as we increase the number of basis centers $\Xi_L = \{\xi_i\}_{i=1}^L \subset \mathcal{X}$ used for the finite dimensional estimator $\hat{g}_L$, and the density of the sample points increases, the error measure $\|g - \hat{g}_L(t_f)\|_{C(\mathcal{X})}$ between the true function and the estimator $\hat{g}_L(t_f)$ at the final time t_f tends to 0. Since we have assumed that $\mathcal{H}_\mathcal{X} \hookrightarrow C(\mathcal{X})$, these simulations should satisfy $\|g - \hat{g}(t_f)\|_{C(\mathcal{X})} \lesssim \|g - \hat{g}(t_f)\|_{\mathcal{H}_\mathcal{X}} \to 0$. To this end, we perform 6 simulations corresponding to each plot in Figure 2, wherein the number of basis centers and corresponding fill distance h vary. Each simulation begins with all agents traversing a "lawnmower" path and collecting samples $g(\Xi_L) \in \mathbb{R}^L$ associated with the basis centers as seen in Figure 2. For simplicity, the subdomains $\{\Omega_i\}_{i=1}^4$ are arbitrarily chosen to divide the domain Ω equally. Indeed other approaches such as Voronoi-based decompositions could be used as well. The coefficients $\alpha \in \mathbb{R}^L$ of the finite dimensional approximation are initialized such that

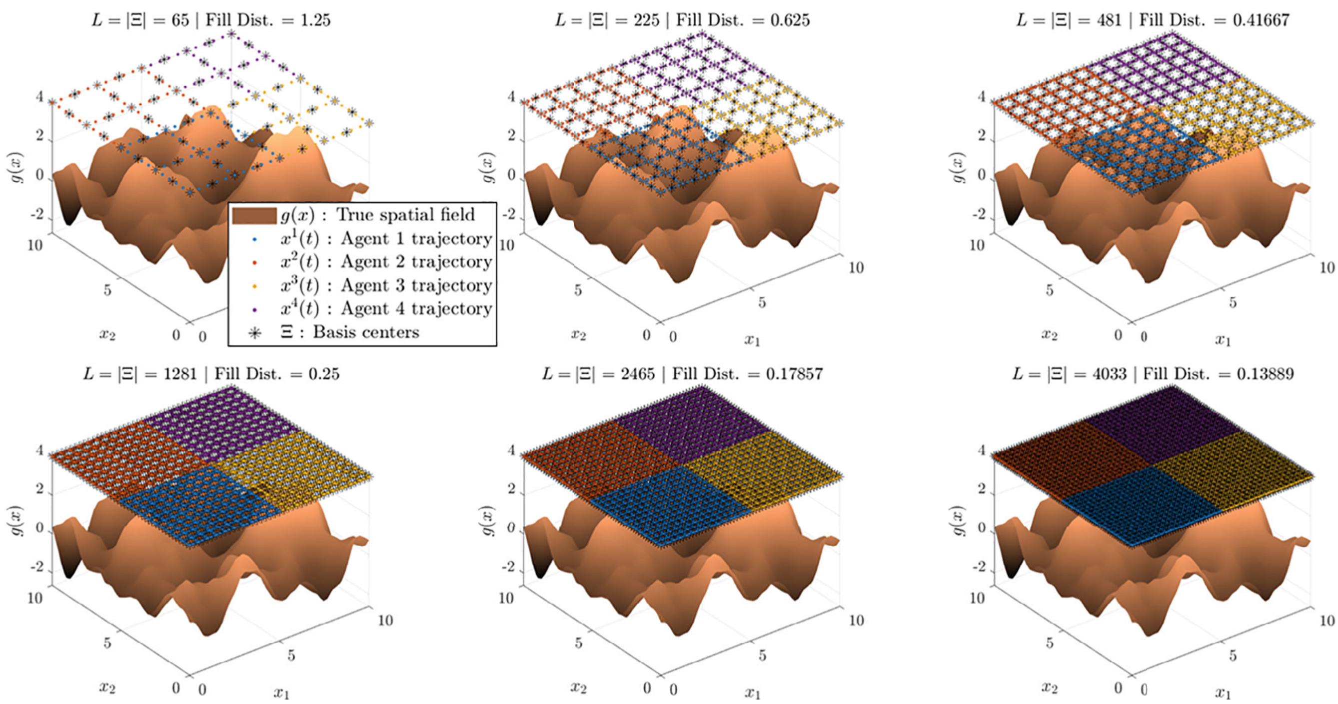

FIGURE 2 The true spatial field g induced by the 5/2 Matern kernel overlaid with the sample trajectories of the $N = 4$ agents and the kernel basis centers $\Xi_L \subset \mathcal{X}$ for the finite-dimensional estimator $\hat{g}_L$. To illustrate the influence of the number of basis functions and fill distance on $\hat{g}_L$, we perform a simulation corresponding to each plot.

$$\alpha_0 = \mathbb{K}(\Xi_L, \Xi_L)^{-1} g(\Xi_L),$$

and at each subsequent time step t_k (k=1,2,...), all agents collect the next sample along their lawnmower trajectory and the coefficients $\alpha_{t_{k-1}}$ are updated in accordance with the evolution law (27). Note that the step-size is fixed to 0.01 and $\gamma = 0.01$ in all experiments. Each simulation ends once the agents have completed their trajectories, and we denote this final time as t_f. In Figure 3A, we see that as $L = |\Xi_L|$ increases, the prediction error measure $\|g - \hat{g}_L(t_f)\|_{C(\mathcal{X})}$ tends to 0. For each $L \in \{65, 481, 1281, 2465, 4033\}$ there is an associated fill distance $h_{\Xi_L,\mathcal{X}}$. In Figure 3B we see that as the fill distance $h_{\Xi_L,\mathcal{X}}$ decreases, the prediction error measure tends to 0, which is consistent with Corollary 2. The black traces in plots (a) and (b) correspond to the ideal setting in which the hyperparameters $\theta_{\hat{g}}$ of the estimator align precisely with the hyperparameters θ_g of the true function g. In more practical settings, such as terrain or temperature mapping applications, θ_g is not typically known, and a practitioner may have to rely on rough estimate of the magnitude and length-scale of the spatial field to select $\theta_{\hat{g}}$. To emulate this scenario, we perform simulations for $\theta_{\hat{g}} = c\theta_g$ for various scale factors $c \in \mathbb{R}^+$. From the colored traces in Figure 3, we see that even if $\theta_{\hat{g}} \neq \theta_g$, the prediction error measure still tends to 0 as the fill distance approaches 0.

We also carried out a series of simulations to study the numerical conditioning of the approach in this article. Recall that the update law (27) requires the inversion of the Grammian matrix $\mathbb{K}(\Xi_L, \Xi_L)$. When the condition number of Grammian matrix is sufficiently large, the matrix can become numerically singular so that the inverse update effectively becomes unstable, despite using a rank-1 update.[51] We report the condition number of $\mathbb{K}(\Xi_L, \Xi_L)$ in Figure 3C and remark that no numerical instability was exhibited in these experiments.

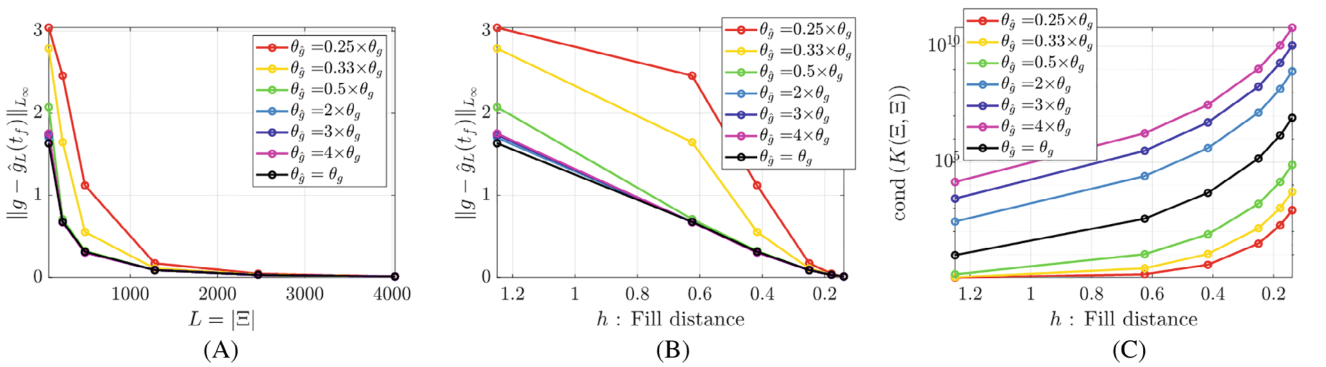

FIGURE 3 Prediction error measure $\|g - \hat{g}_L(t_f)\|_{C(\mathcal{X})}$ as (A) the number of basis centers $\Xi_L = \{\xi_i\}_{i=1}^L$ increases, and (B) as the fill distance h decreases. The black traces reflect the ideal scenario in which the hyperparameters $\theta_{\hat{g}}$ of the estimator are equal to those of the spatial field. The colored traces reflect the practical scenario in which θ_g is unknown, and the hyperparameters of the estimator are set to a scalar multiple of θ_g. The condition number cond($\mathbb{K}(\Xi_L, \Xi_L)$) of the Grammian matrix $\mathbb{K}(\Xi_L, \Xi_L)$ is reported in subplot (C).

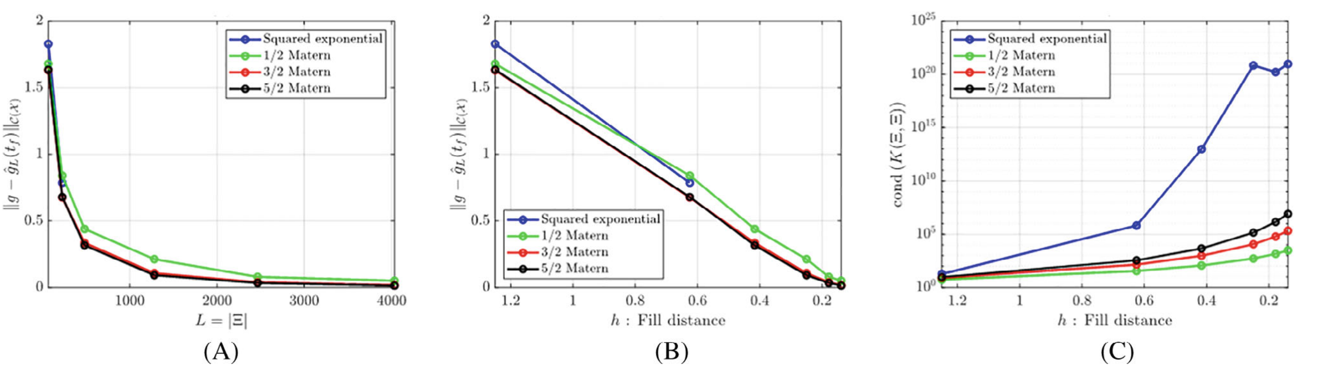

FIGURE 4 Prediction error measure $\|g - \hat{g}_L(t_f)\|_{C(\mathcal{X})}$ as (A) the number of basis centers $\Xi_L = \{\xi_i\}_{i=1}^L$ increases, and (B) as the fill distance h decreases. We also report the condition number of the Grammian matrice in (C). The black traces reflect the ideal scenario in which the kernel of the estimator is equal that of the spatial field. The colored traces serve to illustrate the behavior of the estimator when the kernel does not precisely match the kernel of g.

Moreover, in practical applications, the kernel of the RKHS to which the function g belongs is seldom known. To observe the effect of kernel selection on predictive performance, we perform the same simulations in the preceding paragraph, but rather than scale the hyperparameters $\theta_{\hat{g}}$ of the estimator, we fix the hyperparameters $\theta_{\hat{g}} = \theta_g$ and observe the predictive performance when the kernel of the estimator is the squared exponential kernel, the 1/2 Matern kernel, the 3/2 Matern kernel, and the 5/2 Matern kernel (the true kernel); see Figure 4. Here, the approximation error associated with the 3/2 and 1/2 Matern kernels approach 0. For the squared exponential kernel, results corresponding to fill distances of 0.41 and smaller were unobtainable because the Grammian matrix $\mathbb{K}(\Xi_L, \Xi_L)$ becomes numerically singular for high dimensions L. It is also noteworthy that all Matern kernels studied in Figure 4 exhibit significantly lower condition numbers compared to the squared exponential kernel having smaller fill distances (and correspondingly, high dimensions). At the very least, these studies suggest that further investigation is warranted to understand the practicality and numerical stability of the RKHS embedding method for various kernel choices. We also note that these observations regarding numerical stability and the potential for conditioning problems are well-known in other contexts for scattered data approximations. See Chapter 12 of Reference 35 for a general discussion of conditioning of Grammian matrices, as well as the References 52-54 for the development of preconditioning methods that address these stability issues.

In the previous empirical studies, the set of basis centers Ξ_L was fixed *a priori*. This final set of numerical simulations is designed to illustrate the behavior of the estimator when Ξ_L is iteratively updated as new measurements are collected, as outlined in Algorithm 1. The trajectory for the agents is generated by concatenating the first four lawnmower trajectories (associated with the labels $L = \{65, 225, 481, 1281\}$ depicted in Figure 2) so that the agents collect samples along the initially coarse grid and progressively refine the resolution of the sampling trajectory. We simulate the case when the estimator kernel hyperparameters $\theta_{\hat{g}}$ match those of the spatial field θ_g, as well as the case when $\theta_{\hat{g}} = c\theta_g$ for various scalar multiples $c \in \mathbb{R}^+$. In Figure 5A we observe the evolution of $L = |\Xi_L|$ over time. Note that we use the magnitude of the projection residual (28) as the novelty metric ϵ to decide whether or not to add the location $x \in \mathcal{X}$ of a new sample to Ξ_L.

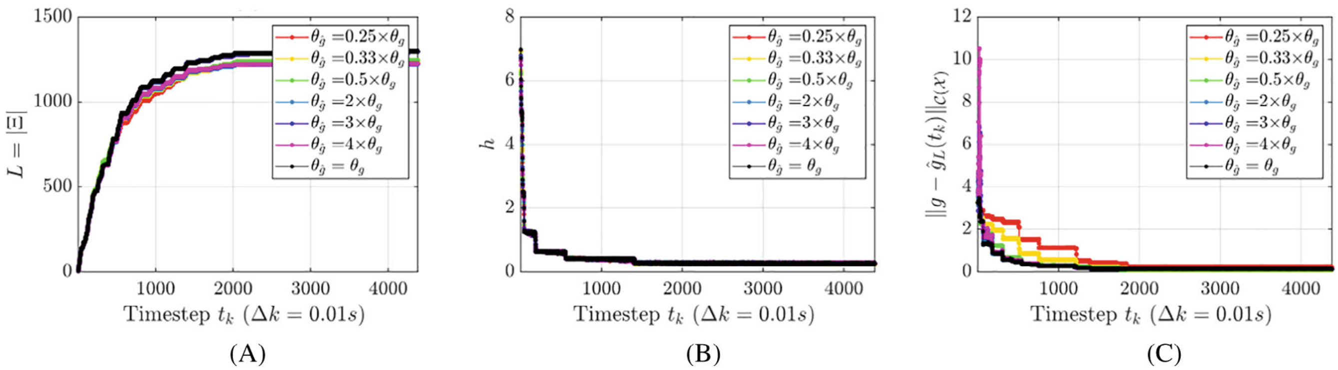

FIGURE 5 For the setting in which the basis centers are iteratively constructed, these plots depict the time evolution of (A) the number of basis centers contained in Ξ_L, (B) the fill distance $h_{\Xi_L,\mathcal{X}}$, and (C) the prediction error measure $\|g - \hat{g}_L(t_k)\|$

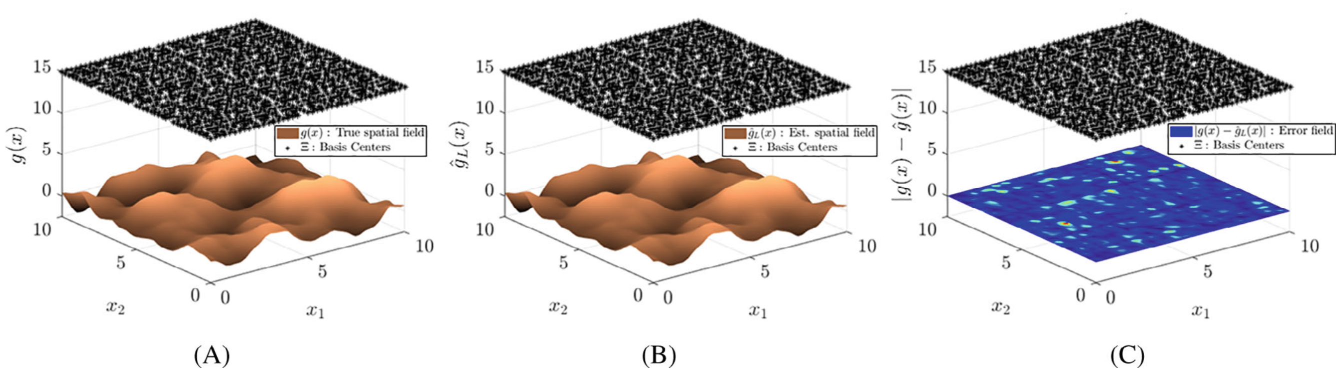

FIGURE 6 Depicted is the (A) true spatial field g, (B) estimated spatial field $\hat{g}_L(t_f)$, (C) error field $|g - \hat{g}_L(t_f)|$, each overlaid with the kernel basis centers contained in Ξ_L for the case when the hyperparameters of the estimator $\hat{g}_L$ coincide with those of g.

The novelty threshold $\bar{\epsilon}$ is tuned for each set of estimator hyperparameters so as to yield approximately the same number of basis functions. Observe that around timestep $t_k = 2000$, L begins to saturate along with the fill distance, shown in (b). In subplot (c), we see that as the fill distance decreases, the prediction error measure $\|g - \hat{g}_L(t_k)\|_{C(\mathcal{X})}$ tends to zero as expected. For the simulation in which $\theta_{\hat{g}} = \theta_g$, we plot, in Figure 6, the true spatial field, the estimated spatial field, and error field at the end of the simulation, each overlaid with the locations of the kernel centers Ξ_L. Qualitatively speaking, the estimated spatial field $\hat{g}_L$ agrees well with the true spatial field g, as evidenced by the small magnitude of the error field.

5 | CONCLUSIONS AND FUTURE WORK

In this article, we propose and study an online method for estimating the unknown external field using the multiagent system. Without parameterization, we cast the estimation problem in the infinite-dimensional RKHS of functions, and propose the evolution law in (4). We have shown that the evolution law (4) for the proposed infinite-dimensional estimator is well-posed and yields a unique solution $\hat{g}(t) \in \mathcal{H}_{\mathcal{X}}$. Moreover, convergence analysis is given for the pointwise ideal error $|g(x) - \hat{g}(t,x)|$ as $t \to \infty$ over the region that is persistently excited. The finite-dimensional estimator in RKH subspaces is then studied to approximate the ideal estimation result. We prove that the approximation error $\|\hat{g}(t) - \hat{g}_L(t)\|_{\mathcal{H}_{\mathcal{X}}}$ is bounded in terms of the fill distance h of basis centers which are used to construct the approximation subspace. As the dimension $L \to \infty$ and the fill distance $h \to 0$, the approximation error converges to zero. Numerical examples are given to verify the results in several different scenarios.

While the results in this article are given for the ideal centralized setting, it serves as the groundwork for the more practical, yet complex, multi-agent estimation settings. Consider, for example, the decentralized setting in which each agent $i = 1 : N$ is subject to various network topologies and must combine its local observations with information from neighboring agents to form its own individual estimator $\hat{g}^i(t)$. Of particular interest are conditions under which $\hat{g}^i(t) \to g$ as $t \to \infty$ for each $i = 1, \ldots, N$.

FUNDING INFORMATION

This work was supported by the National Defense Science and Engineering Graduate (NDSEG) fellowship program and the Office of Naval Research via Grants N00014-18-1-2627 and N00014-19-1-2194.

ACKNOWLEDGEMENTS

The authors thank the reviewers for their time and feedback.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

ORCID

Michael E. Kepler https://orcid.org/0000-0002-6145-6824
Sai Tej Paruchuri https://orcid.org/0000-0003-2372-3888

REFERENCES

1. Csató L, Opper M. Sparse on-line Gaussian processes. *Neural Comput*. 2002;14(3):641-668.
2. Bai S, Wang J, Chen F, Englot B. Information-theoretic exploration with Bayesian optimization; 2016:1816-1822; IEEE.
3. Lin TX, Al-Abri S, Coogan S, Zhang F. A distributed scalar field mapping strategy for Mobile robots; 2020:11581-11586; IEEE.
4. Ghaffari JM, Valls MJ, Dissanayake G. Gaussian processes autonomous mapping and exploration for range-sensing mobile robots. *Auton Robot*. 2018;42(2):273-290.
5. Pillonetto G, Schenato L, Varagnolo D. Distributed multi-agent Gaussian regression via finite-dimensional approximations. *IEEE Trans Pattern Anal Mach Intell*. 2018;41(9):2098-2111.
6. Kontoudis GP, Stilwell DJ. Decentralized nested Gaussian processes for multi-robot systems; 2021:8881-8887; IEEE.
7. Jang D, Yoo J, Son CY, Kim D, Kim HJ. Multi-robot active sensing and environmental model learning with distributed Gaussian process. *IEEE Robot Automat Lett*. 2020;5(4):5905-5912.
8. Sayed AH. Adaptation, learning, and optimization over networks. *Found trends Mach Learn*. 2014;7:311-801.
9. Demetriou MA, Nestinger SS. Adaptive consensus estimation of multi-agent systems; 2011:354-359; IEEE.
10. Demetriou MA. Adaptive consensus filters for collocated infinite dimensional systems; 2011:597-602; IEEE.
11. Sadikhov T, Demetriou MA, Haddad WM, Yucelen T. Adaptive estimation using multiagent network identifiers with undirected and directed graph topologies. *J Dyn Syst Meas Control*. 2014;136(2):021018-1 to 021018-9.

12. Ioannou PA, Sun J. *Robust Adaptive Control*. Courier Corporation; 2012.
13. Wahba Grace. Spline Models for Observational Data. SIAM; 1990.
14. Cox DD. Multivariate smoothing spline functions. *SIAM J Numer Anal*. 1984;21(4):789-813.
15. Silverman BW. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J R Stat Soc B Methodol*. 1985;47(1):1-21.
16. Saunders C, Gammerman A, Vovk V. Ridge regression learning algorithm in dual variables; 1998.
17. Schölkopf Bernhard, Smola Alexander J, Bach Francis, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press; 2002.
18. Cucker F, Smale S. On the mathematical foundations of learning. *Bull Am Math Soc*. 2002;39(1):1-49.
19. Williams CK, Rasmussen CE. *Gaussian Processes for Machine Learning*. Vol 2. MIT Press; 2006:302-303.
20. Krause A, Singh A, Guestrin C. Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies. *J Mach Learn Res*. 2008;9(2):235-284.
21. Smale S, Zhou D-X. Estimating the approximation error in learning theory. *Anal Appl*. 2003;1(01):17-41.
22. Smale S, Zhou D-X. Learning theory estimates via integral operators and their approximations. *Constr Approx*. 2007;26(2):153-172.
23. Cucker F, Zhou DX, eds. *Learning Theory: An Approximation Theory Viewpoint*. Cambridge University Press; 2007.
24. Predd JB, Kulkarni SB, Poor HV. Distributed learning in wireless sensor networks. *IEEE Signal Process Mag*. 2006;23(4):56-69.
25. Alsheikh MA, Lin S, Niyato D, Tan HP. Machine learning in wireless sensor networks: algorithms, strategies, and applications. *IEEE Commun Surv Tutor*. 2014;16(4):1996-2018.
26. Kumar DP, Amgoth T, Annavarapu CSR. Machine learning algorithms for wireless sensor networks: a survey. *Inf Fusion*. 2019;49:1-25.
27. Predd JB, Kulkarni SR, Poor HV. A collaborative training algorithm for distributed learning. *IEEE Trans Inf Theory*. 2009;55(4):1856-1871.
28. Deng Z, Gregory J, Kurdila A. Learning theory with consensus in reproducing kernel Hilbert spaces; 2012:1400-1405; IEEE.
29. Gregory JG. *A Rate of Convergence for Learning Theory with Consensus*. PhD thesis. Virginia Tech; 2015.
30. Xu Y, Choi J, Dass S, Maiti T. Efficient Bayesian spatial prediction with mobile sensor networks using Gaussian Markov random fields. *Automatica*. 2013;49(12):3520-3530.
31. Paruchuri ST, Guo J, Kurdila A. Sufficient conditions for parameter convergence over embedded manifolds using kernel techniques. arXiv preprint arXiv:2009.02866, 2020.
32. Guo J, Paruchuri ST, Kurdila AJ. Persistence of excitation in uniformly embedded reproducing kernel Hilbert (RKH) spaces; 2020:4539-4544; IEEE.
33. Baumeister J, Scondo W, Demetriou MA, Rosen IG. On-line parameter estimation for infinite-dimensional dynamical systems. *SIAM J Control Optim*. 1997;35(2):678-713.
34. Hangelbroek T, Narcowich FJ, Ward JD. Kernel approximation on manifolds I: bounding the Lebesgue constant. *SIAM J Math Anal*. 2010;42(4):1732-1760.
35. Wendland H. *Scattered Data Approximation*. Cambridge University Press; 2004.
36. Berlinet A, Thomas-Agnan C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media; 2011.
37. Daleckii JL, Krein M. *Stability of Solutions of Differential Equations in Banach Space*. Vol 43. American Mathematical Society; 2002.
38. Khalil Hassan K. *Nonlinear Systems*. Pearson; 1991.
39. Rugh WJ. *Linear System Theory*. Prentice-Hall, Inc; 1993.
40. Sastry S, Bodson M. *Adaptive Control: Stability, Convergence and Robustness*. Courier Corporation; 2011.
41. Narendra KS, Annaswamy AM. *Stable Adaptive Systems*. Courier Corporation; 2012.
42. Krstic M, Kokotovic PV, Kanellakopoulos I. *Nonlinear and Adaptive Control Design*. John Wiley & Sons, Inc.; 1995.
43. Farkas B, Wegner S-A. Variations on Barbălat's lemma. *Am Math Mon*. 2016;123(8):825-830.
44. Farrell JA, Polycarpou MM. *Adaptive Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. John Wiley & Sons; 2006.
45. Kurdila AJ, Guo J, Paruchuri ST, Bobade P. Persistence of excitation in reproducing kernel Hilbert spaces, positive limit sets, and smooth manifolds. arXiv Preprint arXiv:190912274, 2019.
46. Megginson RE. *An Introduction to Banach Space Theory*. Springer Science & Business Media; 2012.
47. Warga J. *Optimal Control of Differential and Functional Equations*. Academic Press; 2014.
48. Butcher JC. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons; 2016.
49. Gao T, Kovalsky SZ, Daubechies I. Gaussian process landmarking on manifolds. *SIAM J Math Data Sci*. 2019;1(1):208-236.
50. Kepler ME, Stilwell DJ. An approach to reduce communication for multi-agent mapping applications; 2020:4814-4820; IEEE.
51. Nguyen-Tuong D, Seeger M, Peters J. Model learning with local gaussian process regression. *Adv Robot*. 2009;23(15):2015-2034.
52. Chen K. *Matrix Preconditioning Techniques and Applications*. Vol 19. CambridgeUniversity Press; 2005.
53. Chen J. On the use of discrete laplace operator for preconditioning kernel matrices. *SIAM J Sci Comput*. 2013;35(2):A577-A602.
54. Stein ML, Chen J, Anitescu M. Difference filter preconditioning for large covariance matrices. *SIAM J Matrix Anal Appl*. 2012;33(1):52-72.

How to cite this article: Guo J, Kepler ME, Tej Paruchuri S, Wang H, Kurdila AJ, Stilwell DJ. Adaptive estimation of external fields in reproducing kernel Hilbert spaces. *Int J Adapt Control Signal Process*. 2022;36(8):1931-1957. doi: 10.1002/acs.3442