

Kernel center adaptation in the reproducing kernel Hilbert space embedding method

Sai Tej Paruchuri¹  | Jia Guo² | Andrew Kurdila³

¹Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, Pennsylvania, USA

²Electrical and Computer Engineering, Georgia Tech, Atlanta, Georgia, USA

³Mechanical Engineering, Virginia Tech, Blacksburg, Virginia, USA

Correspondence

Sai Tej Paruchuri, Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, PA 18015, USA.
Email: saitejp@lehigh.edu

Funding information

No funding was used to conduct research presented in this article.

Summary

The performance of adaptive estimators that employ embedding in reproducing kernel Hilbert spaces (RKHS) depends on the choice of the location of basis kernel centers. Parameter convergence and error approximation rates depend on where and how the kernel centers are distributed in the state-space. In this article, we develop the theory that relates parameter convergence and approximation rates to the position of kernel centers. We develop criteria for choosing kernel centers in a specific class of systems by exploiting the fact that the state trajectory regularly visits the neighborhood of the positive limit set. Two algorithms, based on centroidal Voronoi tessellations and Kohonen self-organizing maps, are derived to choose kernel centers in the RKHS embedding method. Finally, we implement these methods on two practical examples and test their effectiveness.

KEYWORDS

adaptive estimation, centroidal Voronoi tessellations, Kohonen self-organizing maps, Lloyd's algorithm, persistence of excitation, reproducing kernel Hilbert space

1 | INTRODUCTION

Adaptive estimation of unknown nonlinearities appearing in dynamical systems is a topic that has been studied over the past four decades. The finite-dimensional versions of such problems are described in classical texts like References 1-3. These methods aim to estimate unknown terms appearing in the governing ordinary differential equations (ODEs). A common assumption in such problems is that all the states are available for measurement. Many of these methods also assume that the unknown function belongs to some hypothesis space of functions. The particular class of adaptive estimators studied in this article assumes that the hypothesis space is a reproducing kernel Hilbert space (RKHS). An RKHS \mathcal{H}_X is a Hilbert space of functions on the state-space $X = \mathbb{R}^d$ that is defined in terms of a symmetric, positive-definite kernel $\mathcal{K} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, which satisfies the reproducing property: for all $\mathbf{x} \in X$ and $f \in \mathcal{H}_X$, $(\mathcal{K}(\mathbf{x}, \cdot), f)_{\mathcal{H}_X} = \mathcal{E}_{\mathbf{x}} f = f(\mathbf{x})$.^{4,5} An example of an RKHS is the space generated by the Gaussian radial basis kernels that have the form $\mathcal{K}(x, y) := e^{-\zeta \|x-y\|^2}$, where ζ is positive. The additional structure induced by the kernel \mathcal{K} on $\mathcal{H}_{\mathbb{R}^d}$ enables the proof of crucial convergence results, even for the infinite-dimensional cases. The finite-dimensional version of the RKHS adaptive estimators has been studied by authors of References 6,7. However, the results for the infinite-dimensional adaptive estimation cases are new and were initially investigated by Bobade et al. in Reference 8.

In both the finite and infinite-dimensional cases, the unknown function f is assumed to be an element of the RKHS \mathcal{H}_X . In adaptive estimation techniques, a function estimate $\hat{f}(t, \cdot)$ is constructed. The goal is to develop a learning law (an ODE in the case of finite-dimensions or a distributed parameter system (DPS) in the case of infinite-dimensions) that defines

the evolution of the function estimate $\hat{f}(t, \cdot)$ such that it approaches the actual function f as $t \rightarrow \infty$. Note that as opposed to static learning theory, the estimate here is governed by a dynamical system. In the finite-dimensional case, the function estimate has the form $\hat{f}(t, \cdot) = \sum_i \alpha_i(t) \mathfrak{K}_{\mathbf{x}_i}(\cdot)$, where $\mathfrak{K}_{\mathbf{x}_i}(\cdot) := \mathcal{K}(\mathbf{x}_i, \cdot)$ with $\mathbf{x}_i \in \mathbb{R}^d$. We refer to $\mathfrak{K}_{\mathbf{x}_i} \in \mathcal{H}$ as the kernel function centered at \mathbf{x}_i . In the RKHS literature, $\mathfrak{K}_{\mathbf{x}_i}$ is referred to as the kernel section of \mathcal{K} . In the infinite-dimensional case, to enable practical implementation, the function estimate $\hat{f}(t, \cdot)$ is approximated by a finite-dimensional estimate of the form $\hat{f}_n(t, \cdot) = \sum_{i=1}^n \alpha_i(t) \mathfrak{K}_{\mathbf{x}_i}(\cdot)$. Since the infinite-dimensional learning law cannot be solved numerically, the evolution of the approximated finite-dimensional estimate is determined by integrating an approximate learning law (an ODE). In both finite- and infinite-dimensional RKHS based adaptive estimation techniques, the choice of the kernel centers \mathbf{x}_i affect the convergence of the function estimate to the actual function.

The general problem of kernel center selection is familiar in adaptive estimation, dynamical system identification, and control. Though some of the studies in these fields do not explicitly deal with kernel center selection, the center adaptation or the kernel adaptation problems are often indirectly addressed to improve performance. For instance, Lian et al. develop a self-organizing RBF network that tunes the RBF network parameters based on an adaptation law.⁹ They use this method for real-time approximation of dynamical systems. Han et al. describe a version of self-organizing RBF networks that use a growing and pruning algorithm in Reference 10. They illustrate the effectiveness of such networks and their variants for dynamical system identification and model predictive control.¹¹⁻¹⁴ Sanner and Slotine implement Gaussian networks for direct adaptive control in Reference 15. The neuro-control technique discussed in References 16,17 uses a fixed set of basis functions or kernel centers. On the other hand, in the controller using neural networks proposed in Reference 18, the kernel centers are chosen such that linear independence of $\mathfrak{K}_{\mathbf{x}_i}$ is maintained. As per the algorithm given in Reference 19, the kernel parameters are chosen to approximate the nonlinear inversion error over a compact set. Reference 20 presents the advantages of adapting the kernel parameters and presents a theory for static as well as dynamic problems.

In conventional system identification techniques, parameter convergence is not always guaranteed. In adaptive estimation, we ordinarily use sufficient conditions, referred to as *persistence of excitation* (PE) conditions, to ensure parameter convergence.¹⁻³ In most practical cases, the PE conditions are difficult to ensure *a priori*. They often do not play a constructive role in coming up with practical algorithms. For this reason, several authors have studied adaptive estimation methods which ensure parameter convergence without PE. In Reference 21, Chowdhary and Johnson show that if the chosen regressors evaluated at measured data are linearly independent, then we get parameter convergence. Kamalapurkar et al. extended this work in Reference 22 to relax the assumptions and developed a concurrent learning technique that implements a dynamic state-derivative estimator. Kingravi et al. in Reference 7 propose a real-time regressors update algorithm that uses the regressors linear independence test. In Reference 23, Modares et al. show that parameter convergence can be ensured by checking for linear independence of the filtered regressor. An alternative class of methods uses Gaussian processes for adaptive estimation and adaptive control.²⁴⁻²⁷ In these methods, the kernel centers are chosen at the points corresponding to the measured output data. An introduction to this theory with examples is given in Reference 28.

The conventional PE condition is linked to the richness of the regressor functions that are used to represent the unknown function. As mentioned earlier, such a notion is hard to interpret intuitively. However, in the RKHS embedding method, the modified PE conditions, studied in References 29,30, are directly related to the kernel center positions in the state-space. Studies on PE in RKHS show that only the kernel centers regularly visited by the state trajectory are persistently excited.^{6,31} Such a relationship makes it easier to ensure PE in RKHS for a certain class of dynamical systems. Furthermore, recent results have shown that the persistently exciting sets are contained in the positive limit sets for a particular class of RKHS adaptive estimators.³² These results clearly show us that the kernel centers play a crucial role in PE and the parameter convergence in RKHS based adaptive estimators.

Moreover, the results in Reference 33 establish that the accuracy of the RKHS embedding method can be shown to depend on the fill distance of samples in a manifold. As the fill distance decreases to zero, the finite-dimensional approximation of function estimate $\hat{f}_n(t, \cdot)$ converges to the infinite-dimensional function estimate $\hat{f}(t, \cdot)$. At the same time, it is also known that the condition number of the Gramian matrix that must be inverted to implement the RKHS embedding method is bounded below by the minimal separation of samples that define the space of approximants. These two observations suggest that strategies to control the distribution of kernel centers in practical simulations are needed.

The kernel center selection and adaptation techniques cited above do not explicitly consider these issues. The principal goal of this article is first to develop a set of intuitive criteria for kernel center selection based on the theory of adaptive estimation in RKHS. The second principal goal of this article is to develop two kernel center selection algorithms that

satisfy these criteria for certain classes of nonlinear systems. These algorithms apply to systems in which the neighborhoods of points in the positive limit sets are visited regularly by the state trajectory and are designed to be applied a priori to starting the RKHS based adaptive estimator. In the limited literature on adaptive estimation by RKHS embedding, such algorithms are yet to be explored to the best of the authors' knowledge. The first algorithm is based on constructing centroidal Voronoi tessellations (CVT) of a polygon surrounding the measured data. The second approach is based on Kohonen self-organizing maps. The advantages of these methods are as follows:

1. These algorithms choose kernel centers directly from the state-space. Such methods work for a large class of regressor functions or types of kernels that define the RKHS.
2. We do not need explicit equations for the persistently exciting sets, which is the case in most practical applications. It is hard to pick evenly distributed kernel centers in the persistently exciting set without such knowledge.
3. There is commercially available software for computing CVT and Kohonen self-organizing maps. This makes both methods simple to implement.

We organize the sections in this article as follows. In Section 2, we present the theory of adaptive estimation in infinite-dimensional RKHS and basic properties of persistence of excitation. We first prove that the infinite-dimensional PE condition implies uniform convergence of the parameter error in the PE sets (Corollary 1). This proof strengthens the results in References 29,30 in that it provides an intuitive insight into the implications of the PE condition in the infinite-dimensional RKHS embedding method. We also discuss the relation between the approximation rates and distribution of samples in the state-space. Finally, we develop a set of kernel center selection criteria based on the discussed theory and illustrate the effectiveness of the criteria using an example. In Section 3, we present the first method and theory of CVT-based kernel center selection. We also prove theorems on convergence in this section. Section 4 presents the method based on Kohonen self-organizing maps. Finally, we present two examples that illustrate the effectiveness of both methods in Section 5.

2 | RKHS EMBEDDING FOR ADAPTIVE ESTIMATION

This section aims to present the idea of RKHS based adaptive estimation rigorously and develop criteria for kernel center selection based on the theory.

2.1 | Reproducing kernel Hilbert space

A reproducing kernel Hilbert space \mathcal{H}_X is a Hilbert space associated with a symmetric, positive-definite kernel $\mathcal{K} : X \times X \rightarrow \mathbb{R}$. See References 4,5 for axiomatic definitions of what constitutes an admissible kernel. The kernel satisfies two properties, (1) $\mathcal{K}(\mathbf{x}, \cdot) \in \mathcal{H}$ for all $\mathbf{x} \in X$, and (2) the reproducing property: for all $\mathbf{x} \in X$ and $f \in \mathcal{H}_X$, $(\mathcal{K}(\mathbf{x}, \cdot), f)_{\mathcal{H}_X} = \mathcal{E}_{\mathbf{x}} f = f(\mathbf{x})$. Here, the notation $(\cdot, \cdot)_{\mathcal{H}_X}$ denotes the inner product associated with the Hilbert space \mathcal{H}_X . The term $\mathcal{E}_{\mathbf{x}}$ is the evaluation functional, which is a bounded linear operator. Throughout this article, we consider RKHS generated by kernels which satisfy the condition that $\mathcal{K}(\mathbf{x}, \mathbf{x}) \leq \bar{k} < \infty$. This condition implies that the RKHS is continuously embedded in the space of continuous functions $C(X)$.⁸ As a result of this assumption, we can extend some convergence in RKHS results to uniform convergence over compact subsets of X . Many reproducing kernels used in practice satisfy the above condition. Given a positive-definite kernel, the RKHS \mathcal{H}_X is generated by

$$\mathcal{H}_X := \overline{\text{span}\{\mathcal{K}(\mathbf{x}, \cdot) | \mathbf{x} \in X\}}.$$

Note that if the set X is infinite-dimensional, then the RKHS it generates is also infinite-dimensional. Given a subset $\Omega \subseteq X$, we define the associated RKHS $\mathcal{H}_\Omega \subseteq \mathcal{H}_X$ by

$$\mathcal{H}_\Omega := \overline{\text{span}\{\mathcal{K}(\mathbf{x}, \cdot) | \mathbf{x} \in \Omega\}}.$$

The above-mentioned reproducing property endows the RKHS with a structure that makes many calculations easy. A detailed list of properties of RKHS can be found in References 4,5. In this article, we are particularly interested in the

properties of projection operators that act on an RKHS. We let P_Ω be the \mathcal{H}_X orthogonal projection operator $P_\Omega : \mathcal{H}_X \rightarrow \mathcal{H}_\Omega$. From Hilbert space theory, we know that the operator P_Ω decomposes the Hilbert space \mathcal{H}_X into $\mathcal{H}_\Omega \oplus \mathcal{V}_\Omega$, where \mathcal{V}_Ω is the space of elements orthogonal to the elements of the space \mathcal{H}_Ω . Since the space \mathcal{H}_X is an RKHS, the reproducing property implies that for any $h \in \mathcal{V}_\Omega$, we have $h(\mathbf{x}) = 0$ for all $\mathbf{x} \in \Omega$. Another important property we use in this article is that for any discrete finite set Ω_n , the projection operator P_{Ω_n} coincides with the interpolation operator over Ω_n , that is, for all $h \in \mathcal{H}_X$, and $\mathbf{x} \in \Omega_n$, we have $h(\mathbf{x}) = (P_{\Omega_n} h)(\mathbf{x})$.³⁴

2.2 | Adaptive estimation in RKHS

Consider a nonlinear system governed by the ODE

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + Bf(\mathbf{x}(t)),$$

where $\mathbf{x}(t) \in \mathbb{R}^d$ is the state, $A \in \mathbb{R}^{d \times d}$ is a known Hurwitz matrix, $B \in \mathbb{R}^d$ is a known vector and $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is the unknown (nonlinear) function. Note, if the original system equations do not contain the term $A\mathbf{x}(t)$, we can add and subtract a known Hurwitz matrix and redefine the unknown nonlinear function to have the form shown above. As noted in Reference 8 and discussed in more detail there, more general systems can be addressed in the analysis that follows via analogy to the model problem above.

We assume that the unknown function f lives in the RKHS \mathcal{H}_X , where $X = \mathbb{R}^d$ is the state-space of the system. In other words, we assume that the unknown f has the form $f(\cdot) = \sum_{i \in \mathbb{I}} \alpha_i \mathcal{K}_{\mathbf{x}_i}(\cdot)$ for some $\{\mathbf{x}_i\}_{i \in \mathbb{I}}$ with \mathbb{I} either finite or infinite. We now define an estimator model of the form

$$\dot{\hat{\mathbf{x}}}(t) = A\hat{\mathbf{x}}(t) + B\hat{f}(t, \mathbf{x}(t)),$$

where $\hat{\mathbf{x}}(t) \in \mathbb{R}^d$ is the state estimate and $\hat{f}(t, \mathbf{x}(t))$ is the function estimate. For each t , the function estimate $\hat{f}(t)$ is an element of the space \mathcal{H}_X . In this article, we assume full-state measurement. This assumption allows us to define a function estimate $\hat{f}(t)$ that depends on the actual states $\mathbf{x}(t)$. Note that the function estimate also explicitly depends on the time t . The goal of adaptive estimation is to ensure that $\hat{f}(t) \rightarrow f$ as $t \rightarrow \infty$. To achieve this, we define the evolution of the function estimate by the learning law

$$\dot{\hat{f}}(t) = \Gamma^{-1}(B\mathcal{E}_{\mathbf{x}(t)})^* P(\mathbf{x}(t) - \hat{\mathbf{x}}(t)),$$

where $\Gamma \in \mathbb{R}$, $\Gamma > 0$.^{8,29,30} The notation $(\cdot)^*$ represents the adjoint of an operator. Additionally, the term P is a symmetric positive-definite matrix in $\mathbb{R}^{d \times d}$ that solves the Lyapunov equation $A^T P + PA = -Q$, where $Q \in \mathbb{R}^{d \times d}$ is an arbitrarily chosen symmetric positive-definite matrix.

If we define the state and function errors as $\tilde{\mathbf{x}}(t) := \mathbf{x}(t) - \hat{\mathbf{x}}(t)$ and $\tilde{f}(t) := f - \hat{f}(t)$, the error evolution equations can be expressed as

$$\begin{Bmatrix} \dot{\tilde{\mathbf{x}}}(t) \\ \dot{\tilde{f}}(t) \end{Bmatrix} = \underbrace{\begin{bmatrix} A & B\mathcal{E}_{\mathbf{x}(t)} \\ -\Gamma^{-1}(B\mathcal{E}_{\mathbf{x}(t)})^* P & 0 \end{bmatrix}}_{\mathbb{A}(t)} \begin{Bmatrix} \tilde{\mathbf{x}}(t) \\ \tilde{f}(t) \end{Bmatrix}. \tag{1}$$

Note, in the above error equation, the term $\mathbb{A}(t)$ is a uniformly bounded linear operator, and the states $\{\tilde{\mathbf{x}}(t) \ \tilde{f}(t)\}^T$ evolve in the infinite-dimensional space $\mathbb{R}^d \times \mathcal{H}_X$. Thus, the structure of the RKHS enables us to pose the nonlinear finite-dimensional estimation problem into a linear infinite-dimensional problem.

Standard stability analysis using the Lyapunov's theorem and Barbalat's lemma shows that the norm of the state error $\|\tilde{\mathbf{x}}(t)\|_{\mathbb{R}^d} \rightarrow 0$ as $t \rightarrow \infty$.^{8,29,30} However, we need persistence of excitation conditions to prove convergence of the function error. The following subsection covers this.

2.3 | Parameter convergence, PE and positive limit sets

As mentioned earlier, persistence of excitation (PE) conditions are used to prove convergence of the function estimate to the actual function. Two different definitions of PE in RKHS are available in the recent literature on RKHS embedding methods.^{29,30} They are as follows.

Definition 1 (PE. 1). The trajectory $\mathbf{x} : t \mapsto \mathbf{x}(t) \in \mathbb{R}^d$ persistently excites the indexing set Ω and the RKHS \mathcal{H}_Ω provided there exist positive constants T_0, γ, δ , and Δ , such that for each $t \geq T_0$ and any $g \in \mathcal{H}_X$, there exists $s \in [t, t + \Delta]$ such that

$$\left| \int_s^{s+\delta} \mathcal{E}_{\mathbf{x}(\tau)} g d\tau \right| \geq \gamma \|P_\Omega g\|_{\mathcal{H}_X} > 0.$$

Definition 2 (PE. 2). The trajectory $\mathbf{x} : t \mapsto \mathbf{x}(t) \in \mathbb{R}^d$ persistently excites the indexing set Ω and the RKHS \mathcal{H}_Ω provided there exist positive constants T_0, γ , and Δ such that

$$\int_t^{t+\Delta} \left(\mathcal{E}_{\mathbf{x}(\tau)}^* \mathcal{E}_{\mathbf{x}(\tau)} g, g \right)_{\mathcal{H}_X} d\tau \geq \gamma \|P_\Omega g\|_{\mathcal{H}_X}^2 > 0,$$

for all $t \geq T_0$ and any $g \in \mathcal{H}_X$.

Note that the PE condition given in Definition 2 structurally resembles the classical PE conditions defined using regressors in finite-dimensional spaces.¹⁻³ Recall that the term P_Ω in the above definitions is the orthogonal projection operator that maps elements from \mathcal{H}_X to \mathcal{H}_Ω . The following theorem is a special case of the results from References 29,30 and shows how these two PE conditions are related. Note that the notion of parameter convergence in the infinite-dimensional case is given with respect to PE condition in Definition 1 only.

Theorem 1 (29,30). *The PE condition in Definition PE. 1 implies the one in Definition PE. 2. Further, if Ω is a discrete finite set, the state trajectory $t \mapsto \mathbf{x}(t)$ is uniformly continuous and maps to a compact set, and the family of functions defined by $\{g(\mathbf{x}(\cdot)) : t \mapsto g(\mathbf{x}(t)) | g \in \mathcal{H}_X, \|g\| = 1\}$ is uniformly equicontinuous, then the PE condition in Definition PE. 2 implies the one in Definition PE. 1.*

Furthermore, if the trajectory $\mathbf{x} : t \mapsto \mathbf{x}(t)$ persistently excites the RKHS \mathcal{H}_Ω in the sense of Definition PE. 1. Then

$$\lim_{t \rightarrow \infty} \|\tilde{\mathbf{x}}(t)\| = 0, \quad \lim_{t \rightarrow \infty} \|P_\Omega \tilde{f}(t)\|_{\mathcal{H}_X} = 0.$$

We can view the term $P_\Omega \tilde{f}(t)$ as an element of the space \mathcal{H}_Ω . Thus, the above statement implies that $P_\Omega \tilde{f}(t)$ converges to the zero element in the \mathcal{H}_Ω space. However, this statement does not imply the convergence or even the existence of the limit of $\tilde{f}(t) \in \mathcal{H}_X$. The statement $\lim_{t \rightarrow \infty} \|P_\Omega \tilde{f}(t)\|_{\mathcal{H}_X} = 0$ is hard to interpret intuitively. The following corollary of the above theorem gives us the intuition about where the convergence is achieved.

Corollary 1. *If the trajectory $\mathbf{x} : t \mapsto \mathbf{x}(t)$ persistently excites the set Ω and the RKHS \mathcal{H}_Ω in the sense of Definition PE. 1, then $\hat{f}(t)$ converges uniformly to f on the set Ω as $t \rightarrow \infty$.*

Proof. Suppose the projection operator P_Ω decomposes the function $\tilde{f}(t)$ into $\tilde{f}(t) = P_\Omega \tilde{f}(t) + v(t)$, where $P_\Omega(\tilde{f}(t)) \in \mathcal{H}_\Omega$ and $v(t) \in \mathcal{V}_\Omega$. Since $v(t, \mathbf{x}) = 0$ for all $\mathbf{x} \in \Omega$, we have $\tilde{f}(t) = P_\Omega \tilde{f}(t, \mathbf{x})$. Thus, for all $\mathbf{x} \in \Omega$, we have

$$|\tilde{f}(t, \mathbf{x})| = |P_\Omega \tilde{f}(t, \mathbf{x})| = |\mathcal{E}_x P_\Omega \tilde{f}(t)| \leq \|\mathcal{E}_x\| \|P_\Omega \tilde{f}(t)\|_{\mathcal{H}_X}.$$

But we have assumed in this article that the kernel \mathcal{K} that induces \mathcal{H}_X satisfies $\mathcal{K}(\mathbf{x}, \mathbf{x}) \leq \bar{k} < \infty$ for all $\mathbf{x} \in X$. Since the evaluation functional is consequently uniformly bounded, the above inequality holds for all $\mathbf{x} \in \Omega$. Taking the limit $t \rightarrow \infty$ and using Theorem 1 gives us the desired result. ■

The above corollary clearly shows that, if the PE condition holds and the kernel satisfies $\mathcal{K}(\mathbf{x}, \mathbf{x}) \leq \bar{k} < \infty$, then $\hat{f}(t, \mathbf{x}) \rightarrow f(\mathbf{x})$ for all $\mathbf{x} \in \Omega$. Now, we can ask the question of where to look for persistently exciting sets in the state space. Theorem 2 from Reference 32, which is given below, answers this question. The theorem assumes that the RKHS space separates closed sets.

Definition 3. We say the RKHS \mathcal{H}_X separates a set $A \subseteq X$ if for each $\mathbf{b} \notin A$, there is a function $f \in \mathcal{H}_X$ such that $f(\mathbf{a}) = 0$ for all $\mathbf{a} \in A$ and $f(\mathbf{b}) \neq 0$.

The RKHS generated by the Gaussian kernel, which is extensively used for RKHS based adaptive estimation and machine learning, does not satisfy the above condition for all closed sets. A detailed account for RKHS that separate closed sets can be found in Reference 35. In this article, we use the Sobolev-Matern kernels, which does satisfy the above condition for all closed sets. A sufficient condition for an RKHS to separate closed sets is that it contains a rich family of bump functions.

Theorem 2 (32). Let \mathcal{H}_X be the RKHS of functions over X and suppose that this RKHS includes a rich family of bump functions. If the PE condition in Definition PE. 2 holds for Ω , then $\Omega \subseteq \omega^+(\mathbf{x}_0)$, the positive limit set corresponding to the initial condition \mathbf{x}_0 .

This theorem gives us a necessary condition for a set to be persistently excited. The proof of the theorem uses contradiction. To give an overview of the proof, suppose $\mathbf{y} \in \Omega$ is not contained in the positive limit set. Then, there exists an open ball $B_\delta(\mathbf{y})$ of radius δ centered at \mathbf{y} such that it does not intersect the positive limit set. By hypothesis, the RKHS contains a rich family of bump functions. Thus, there is a bump function $b_{\mathbf{y}} \in \mathcal{H}_X$ such that $b_{\mathbf{y}}(\mathbf{y}) = 1$ and $b_{\mathbf{y}}(\mathbf{x}) = 0$ for all \mathbf{x} outside a compact set $C_{\mathbf{y}} \subset B_\delta(\mathbf{y})$. Since \mathbf{y} is not contained in the positive limit set by assumption, after a finite amount of time,

$$\int_t^{t+\Delta} \left(\mathcal{E}_{\mathbf{x}(\tau)}^* \mathcal{E}_{\mathbf{x}(\tau)} b_{\mathbf{y}}, b_{\mathbf{y}} \right)_{\mathcal{H}_X} d\tau = \int_t^{t+\Delta} |b_{\mathbf{y}}(\mathbf{x}(\tau))|^2 d\tau = 0.$$

On the other hand, since $\mathcal{K}(\mathbf{y}, \mathbf{y}) \leq \bar{k}^2$, we have

$$1 = b_{\mathbf{y}}(\mathbf{y}) = |(b_{\mathbf{y}}, \mathcal{K}(\mathbf{y}, \cdot))_{\mathcal{H}_X}| = |(b_{\mathbf{y}}, P_{\Omega} \mathcal{K}(\mathbf{y}, \cdot))_{\mathcal{H}_X}| \leq \|P_{\Omega} b_{\mathbf{y}}\|_{\mathcal{H}_X} \|\mathcal{K}(\mathbf{y}, \cdot)\|_{\mathcal{H}_X} \leq \bar{k} \|P_{\Omega} b_{\mathbf{y}}\|_{\mathcal{H}_X}.$$

From the definition of PE, it is clear that the assumption $\mathbf{y} \in \Omega$ is not contained in the positive limit set leads to a contradiction. A detailed proof can be found in Reference 32. While designing a adaptive estimator, this necessary condition can tell us where to look for persistently excited sets in the state-space.

2.4 | Approximations, convergence rates, and sufficient condition

For practical implementation, we approximate the infinite-dimensional adaptive estimator equations given in the previous subsection. Let $\{\Omega_n\}_{n \in \mathbb{N}}$ be a nested sequence of finite subsets of Ω , Further, let $\{\mathcal{H}_{\Omega_n}\}_{n \in \mathbb{N}}$ be the corresponding subspaces of \mathcal{H}_{Ω} generated by the finite sets Ω_n . Now, define P_{Ω_n} as the orthogonal projection operator from \mathcal{H}_{Ω} to the subspace \mathcal{H}_{Ω_n} such that $\lim_{n \rightarrow \infty} P_{\Omega_n} f = f$ for all $f \in \mathcal{H}_{\Omega}$. With this definition of approximation, we write the finite-dimensional adaptive estimator model and the learning law as

$$\begin{aligned} \dot{\hat{\mathbf{x}}}_n(t) &= A \hat{\mathbf{x}}_n(t) + B \mathcal{E}_{\mathbf{x}(t)} P_{\Omega_n}^* \hat{f}_n(t), \\ \dot{\hat{f}}_n(t) &= \Gamma^{-1} \left(B \mathcal{E}_{\mathbf{x}(t)} P_{\Omega_n}^* \right)^* P \tilde{\mathbf{x}}_n(t) \end{aligned}$$

with $\tilde{\mathbf{x}}_n := \mathbf{x} - \hat{\mathbf{x}}_n$. The basis of the finite-dimensional RKHS \mathcal{H}_{Ω_n} is the set $\{\mathfrak{R}_{\mathbf{x}_i} | \mathbf{x}_i \in \Omega_n\}$. We now note that the finite-dimensional function estimate $\hat{f}_n(t)$ has the form $\hat{f}_n(t) := \sum_{i=1}^n \hat{\alpha}_i(t) \mathfrak{R}_{\mathbf{x}_i}$. Using the reproducing property of the kernel, we rewrite the above finite-dimensional learning law as

$$\dot{\hat{\boldsymbol{\alpha}}}(t) = \mathbb{K}^{-1} \Gamma^{-1} \mathcal{K}(\mathbf{x}_c, \mathbf{x}(t)) B^* P \tilde{\mathbf{x}}_n(t), \tag{2}$$

where $\hat{\boldsymbol{\alpha}}(t) := \{\hat{\alpha}_1(t), \dots, \hat{\alpha}_n(t)\}^T$, \mathbb{K} is the symmetric positive definite Grammian matrix whose ij th element is defined as $\mathbb{K}_{ij} := \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j)$, $\Gamma := \Gamma \mathbb{I}_n$ is the gain matrix, and

$$\mathcal{K}(\mathbf{x}_c, \mathbf{x}(t)) := \left\{ \mathcal{K}(\mathbf{x}_1, \mathbf{x}(t)), \dots, \mathcal{K}(\mathbf{x}_n, \mathbf{x}(t)) \right\}^T.$$

The new learning law defines the rate of evolution of the coefficients, as opposed to the old learning law which defines the rate of evolution of the function $\hat{f}_n(t)$. This step is essential for implementation purposes. We refer the reader to Reference 36 for the intermediate steps involved in the derivation. Note, the PE condition implies the convergence of the infinite-dimensional function estimate $\hat{f}(t)$ to f . It does not imply anything about the convergence of the approximation of the function estimate $\hat{f}_n(t)$ to f . On the other hand, the following theorem, proved in Reference 8, shows that the term $\hat{f}_n(t)$ to $\hat{f}(t)$ as $n \rightarrow \infty$.

Theorem 3 (8). *Suppose that $\mathbf{x} \in C([0, T], \mathbb{R}^d)$ and that the embedding $i : \mathcal{H}_\Omega \hookrightarrow C(\Omega)$ is uniform in the sense that*

$$\|f\|_{C(\Omega)} \equiv \|if\|_{C(\Omega)} \leq C\|f\|_{\mathcal{H}_\Omega}.$$

Then for any $T > 0$ and $t \in [0, T]$,

$$\begin{aligned} \|\hat{\mathbf{x}} - \hat{\mathbf{x}}_n\|_{C([0, T]; \mathbb{R}^d)} &\rightarrow 0, \\ \|\hat{f}(t) - \hat{f}_n(t)\|_{C([0, T]; \mathbb{R}^d)} &\rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$.

Thus, as we choose denser finite discrete sets Ω_n in Ω , the approximation of the function estimate $\hat{f}_n(t)$ gets closer to the function estimate $\hat{f}(t)$, which in turn converges to the actual function f as $t \rightarrow \infty$ if the PE condition holds. The above theorem does not explicitly tell us how to choose the set $\Omega_n \subseteq \Omega$. However, when the set Ω is a compact smooth Riemannian manifold embedded in \mathbb{R}^d with metric d_M , the rate at which $\hat{f}_n(t)$ converges to the $\hat{f}(t)$ depends on how the elements of the set Ω_n are distributed in the set Ω . This distribution is defined in terms of the *fill distance*

$$h_{\Omega_n, \Omega} := \sup_{x \in \Omega} \min_{\xi_i \in \Omega_n} d_M(x, \xi_i).$$

Theorem 4 (33). *Let $\Omega \subseteq X := \mathbb{R}^d$ be a k -dimensional smooth manifold, and let the native space \mathcal{H}_X be continuously embedded in a Sobolev space $W^{\tau, 2}(X)$ with $\tau > d/2$, so that $\|f\|_{W^{\tau, 2}(\mathbb{R}^d)} \lesssim \|f\|_{\mathcal{H}_X}$. Defines $s = \tau - (d - k)/2$ and let $0 \leq \mu \leq [s] - 1$. Then there is a constant h_Ω such that if $h_{\Omega_n, \Omega} \leq h_\Omega$, then for all $f \in \mathbf{R}_\Omega(\mathcal{H}_X)$ we have*

$$\|(I - P_{\Omega_n})\hat{f}(t)\|_{W^{\mu, 2}(\Omega)} \lesssim h_{\Omega_n, \Omega}^{s-\mu} \|\hat{f}(t)\|_{\mathbf{R}_\Omega(\mathcal{H}_X)}.$$

In the above theorem, the notation $\mathbf{R}_\Omega(\mathcal{H}_X)$ represents the restriction of the space \mathcal{H}_X to the set Ω , and the notation $a \lesssim b$ implies that there exists a positive constant c such that $a \leq cb$. This theorem requires a lot of technical details and we direct interested readers to Reference 33 for the detailed explanation of the rigorous theory and proofs. In this article, we are interested in the implications of the theorem. The theorem states that the fill distance $h_{\Omega_n, \Omega}$ defines the rate at which the norm of the error $\hat{f}(t) - \hat{f}_n(t)$ converges to zero.

2.4.1 | Sufficient condition

In all the discussion above, we assume that we have knowledge of the persistently excited set Ω . In many practical cases, it is impossible to determine this set exactly. However, there is a much more practical and intuitive way for selecting the kernel centers in the set Ω_n when the RKHS is generated by a strictly positive definite kernel. For the precise hypothesis of the theorem below, the reader should see Reference 31: the context of the following theorem is rather detailed.

Theorem 5 (6,31). *Suppose the RKHS is generated by a strictly positive definite kernel and the hypotheses of Reference 31 hold. Let $\epsilon < \frac{1}{2} \min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|$, where \mathbf{x}_i and \mathbf{x}_j are the kernel centers that belong to the set $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. For every $t_0 \geq 0$ and $\delta > 0$, define*

$$I_i := I_{i, \epsilon, \delta} := \{t \in [t_0, t_0 + \delta] : \|\mathbf{x}(t) - \mathbf{x}_i\| \leq \epsilon\}.$$

If there exists a $\delta = \delta(\epsilon)$ such that the measure of I_i is bounded below by a positive constant that is independent of t_0 and the kernel center x_i , and if the measure of $[t_0, t_0 + \delta]$ is less than or equal to δ , then the space \mathcal{H}_n is persistently exciting in the sense of PE 2.

Intuitively, the above theorem states that the neighborhoods of the points in the finite PE set Ω_n are visited by the state trajectory infinitely many times, and the time of visitation is bounded below in a certain sense. Note that the above sufficient condition implies PE 2. When the hypotheses of Theorem 1 hold, we can conclude that the sufficient condition given in Theorem 5 implies PE 1. While implementing the adaptive estimator, if we only know that the actual function $f \in \mathcal{H}_X$ with X an infinite set, (as usual) the sufficient condition given in Theorem 5 only implies ultimate boundedness of the function estimate instead of convergence, in particular when we use the dead zone gradient law. Notice that this ultimate bound actually implies a stronger result than the one in the conventional analysis in Euclidean space: here, the ultimate bound is explicit in terms of the approximation space error.³¹

2.5 | Center selection criteria

We discussed the various necessary and sufficient conditions for a set to be persistently excited in the previous subsections. The theory also illuminates whether we can achieve parameter convergence for a particular dynamical system. For instance, the state trajectory of a chaotic dynamical system does not regularly visit any points in the state-space. As a result, we cannot guarantee parameter convergence in this case. Furthermore, it is also clear that we can estimate the unknown function values only at the state-space regions that the state trajectory visits. Thus, selection of the kernel centers requires knowledge of the areas of state-space where the state trajectory concentrates over time. Based on the theory presented in the previous subsections, we list the following criteria for choosing the kernel centers.

- (C1) The kernel centers should be contained in or be as close as possible to the positive limit set based on Theorem 2.
- (C2) The kernel centers should be evenly distributed when possible. There are two reasons for selecting this criterion.
 - (i) The linear dependency of the kernels will be high if the centers are placed too close to each other. This will increase the condition number of the Grammian matrix in Equation (2).
 - (ii) On the other hand, if the centers are too far apart, the fill distance increases, which in turn reduces the approximation rates based on Theorem 4.
- (C3) The neighborhood of the centers should be visited by the state trajectory regularly. This is to satisfy the sufficient condition for PE based on Theorem 5.

2.6 | Example: The case when we have a priori knowledge of positive limit set

We test the above listed criteria on a simple practical example. We consider a nonlinear single-mode undamped piezoelectric oscillator³⁶ with no input to test the above criteria. The governing equations have the form

$$\underbrace{\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix}}_A = \underbrace{\begin{bmatrix} 0 & 1 \\ -\frac{\hat{K}}{M} & -\frac{C}{M} \end{bmatrix}}_A \underbrace{\begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}}_x + \underbrace{\begin{Bmatrix} 0 \\ -\frac{P}{M} \end{Bmatrix}}_u \underbrace{\ddot{z}(t)}_{u(t)} + \underbrace{\begin{Bmatrix} 0 \\ 1 \end{Bmatrix}}_B \underbrace{\left(-\frac{\hat{K}_{N_1}}{M} x_1^3(t) - \frac{\hat{K}_{N_2}}{M} x_1^5(t) \right)}_{f(x(t))}, \tag{3}$$

where M, \hat{K}, C, P are the modal mass, modal stiffness, modal damping, and modal input contribution term of the piezoelectric oscillator. The variables $\hat{K}_{N_1}, \hat{K}_{N_2}$ are the nonlinear stiffness terms. The terms x_1, x_2 , and z are the modal displacement, modal velocity and base displacement of the oscillator, respectively. The steps involved in deriving the above governing equations can be found in Reference 36. Typically, the magnitudes of the velocity and displacement values are not of the same order. In such cases, we have to use kernels that are skewed in a particular direction. Alternatively, we scale one of the states as $x_1 = S\tilde{x}_1$, where S is a positive constant. Note that, after scaling, we have $x(t) := \{\tilde{x}_1(t), x_2(t)\}^T$. In our simulations, we choose $M = 0.9745, \hat{K} = 329.9006, \hat{K}_{N_1} = -1.2901 \times 10^5$, and $\hat{K}_{N_2} = 1.2053 \times 10^9$. For the undamped, no input case, that is, $C = 0$ and $P = 0$, the total energy is conserved. In other words, the trajectory is always contained in

the limit set $\omega^+(\mathbf{x}_0)$, where $\mathbf{x}_0 \in \mathbb{R}^2$ is the initial condition. Note that any arbitrary discrete finite set in $\omega^+(\mathbf{x}_0)$ is visited by the state trajectory infinitely many times.

Since we have a priori knowledge of the limit set $\omega^+(\mathbf{x}_0)$ for a given initial condition, we choose kernel centers in the set Ω and integrate the equations

$$\begin{aligned}\hat{\mathbf{x}}_n(t) &= A\hat{\mathbf{x}}_n(t) + B\hat{\boldsymbol{\alpha}}^T(t)\mathcal{K}(\mathbf{x}_c, \mathbf{x}(t)), \\ \dot{\hat{\boldsymbol{\alpha}}}(t) &= \mathbb{K}^{-1}\Gamma^{-1}\mathcal{K}(\mathbf{x}_c, \mathbf{x}(t))B^*P\hat{\mathbf{x}}_n(t)\end{aligned}$$

over the interval $[0, T]$ for some $T > 0$. In all our simulations, we use the Sobolev-Matern 3,2 kernel, which has the form

$$\mathcal{K}_{3,2}(\mathbf{x}, \mathbf{y}) = \left(1 + \frac{\sqrt{3}\|\mathbf{x} - \mathbf{y}\|}{l}\right) \exp\left(-\frac{\sqrt{3}\|\mathbf{x} - \mathbf{y}\|}{l}\right),$$

where l is the scaling factor of length.³⁷

To analyze the above-listed criteria's effectiveness, we tested the adaptive estimator with a random and a deterministic, uniform collection of kernel centers. We set $S = 0.02$, $l = 0.2$, $\Gamma = 0.001$, and $n = 40$. The states and the parameters are initialized at $\mathbf{x}_0 = \{1.5, 0\}^T$ and $\alpha_i(0) = 1$ for $i = 1, \dots, n$, respectively. For the uniform kernel center selection, we first calculate the distance between two adjacent kernel centers l_n when they are distributed uniformly in the positive limit set. Since we know the exact equation of the positive limit set,²⁹ we can calculate the total length and hence the length of the arc between two adjacent kernel centers. Given a kernel center, we choose the adjacent kernel center at a distance l_n . We repeat this procedure until we choose the required number of kernel centers that are distributed uniformly in the positive limit set. For choosing the kernel centers for the random case, we first ran the uniform center selection algorithm for $n = 48$ case, and then used the MATLAB function `randperm` to select $n = 40$ kernel centers randomly. Note that the MATLAB function `randperm` uses a uniform pseudorandom number generator algorithm.

Figures 1 and 2 show the pointwise error $|f(\mathbf{x}) - \hat{f}_n(T, \mathbf{x})|$ after running the adaptive estimator for $T = 2000$ s for a particular case of random and uniform selection of kernel centers. It is clear from the figures that the pointwise error is low in the case of uniform sampling. Figure 3 shows how the norm $\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}(t)\|_{\mathbb{R}^n}$ varies with time t for both the random and uniform center selection methods. It is clear from Figure 3 that the coefficient error norm converges rapidly to zero for the uniform centers case. For the random centers case, the error norm does not even start converging in the first 2000 s.

In the above problem, it is assumed that we have an explicit equation for the positive limit set $\omega^+(\mathbf{x}_0)$ for a given initial condition \mathbf{x}_0 . Furthermore, the state trajectory is contained in the set $\omega^+(\mathbf{x}_0)$. This makes it possible to choose kernel centers that are uniformly distributed. In most practical examples, we cannot derive an explicit expression for the set $\omega^+(\mathbf{x}_0)$. We only have samples of the state-trajectory that is contained in or converges to the positive limit set $\omega^+(\mathbf{x}_0)$. In the following two sections, we present kernel center selection algorithms that can be implemented when we do not have explicit equation of the positive limit set or when the state trajectory is not contained in the positive limit set. These algorithms are designed to be implemented a priori to starting the estimator.

3 | METHOD 1: BASED ON CVT AND LLOYD'S ALGORITHM

The first method we propose is based on building centroidal Voronoi tessellations (CVT) around the positive limit set. This method relies on samples taken in the positive limit set. However, we do not have to know the equation of the positive limit set explicitly, we only require that we know the limit set is contained in some known set. We implement this approach for systems where the state-trajectory is contained in the positive limit set or converges to the same in finite time. We assume that there is a dense sampling Ξ of the positive limit set, that is, $\bar{\Xi} = \omega^+(\mathbf{x}_0)$. Let $\{\Xi_m\}_{m=1}^{\infty}$ be a sequence of finite subsets of Ξ such that $\Xi_m \subset \Xi_{m+1}$ for all $m \in \mathbb{N}$ and $\cup_{m=1}^{\infty} \Xi_m = \Xi$, where $\Xi_m = \{\xi_1, \dots, \xi_{q_m}\}$. The term q_m represents the number of samples in the set Ξ_m . Given a set of samples Ξ_m , we construct a region Q_m that is assumed to enclose the positive limit set. Before we go into the details of implementation, let us take a look at the theory behind Voronoi partitions.

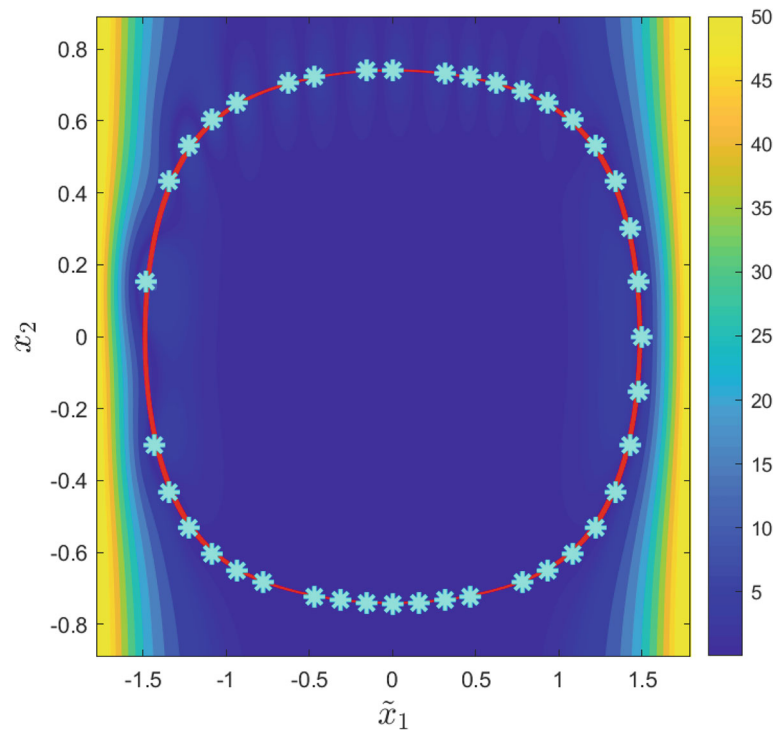


FIGURE 1 Random Centers—Pointwise error $|f(\mathbf{x}) - \hat{f}_n(T, \mathbf{x})|$. The marker * and the red line represent the kernel centers and the limit set, respectively

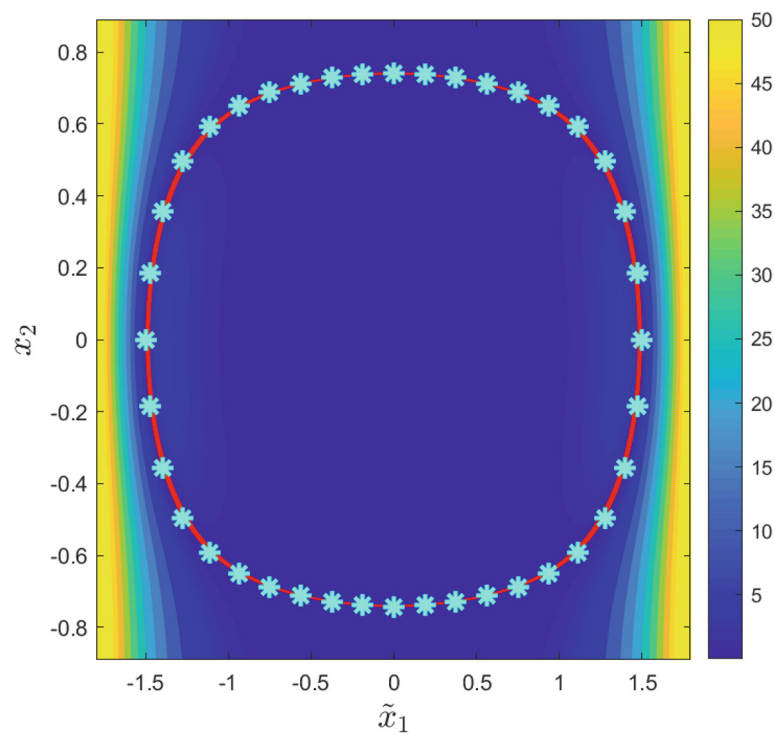


FIGURE 2 Uniform Centers—Pointwise error $|f(\mathbf{x}) - \hat{f}_n(T, \mathbf{x})|$. The marker * and the red line represent the kernel centers and the limit set, respectively

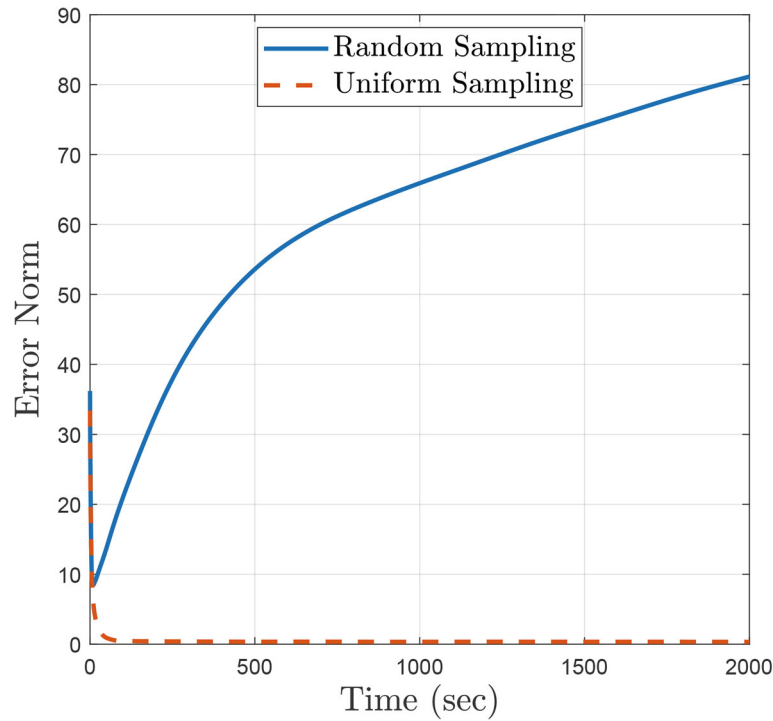


FIGURE 3 Variation of $\|\alpha - \hat{\alpha}(t)\|_{\mathbb{R}^n}$ with time

3.1 | Voronoi partition

Suppose the state-space X is endowed with the metric $d(\cdot, \cdot)$. In this article, we use the Euclidean metric. Let $Q_m \subseteq X$ be a convex polytope and let $P_m = \{\mathbf{p}_{m,1}, \dots, \mathbf{p}_{m,n_m}\}$ be a set of n_m points. The Voronoi partition $\mathcal{V}(P_m)$ generated by the set of points P_m is the collection of n_m polytopes, $P_{m,1}, \dots, P_{m,n_m}$, defined by

$$P_{m,i} = \{\mathbf{x} \in Q_m | d(\mathbf{x}, \mathbf{x}_i) \leq d(\mathbf{x}, \mathbf{x}_j), \text{ for } j = 1, \dots, n_m, j \neq i\}$$

for $i = 1, \dots, n_m$. An edge of the polytope $P_{m,i}$ is the region $P_{m,i} \cap P_{m,j}$ or $P_{m,i} \cap \partial Q_m$ for some $j \neq i$. We say that two polytopes are adjacent when they share a common edge. The notation ∂Q_m denotes the boundary of the region Q_m . We use the notation $\mathbb{E}(\mathcal{V}(P_m), Q_m)$ to denote the union of all edges of the polytopes in $\mathcal{V}(P_m)$. If $R \subseteq Q_m$, then $\mathbb{E}(\mathcal{V}(P_m), R) = \mathbb{E}(\mathcal{V}(P_m), Q_m) \cap R$. A particular class of Voronoi partitions are the *centroidal Voronoi partitions* or *centroidal Voronoi tessellations*, where each point generating the polytope is also its centroid. We use the notation $C_{P_{m,j}}$ to denote the centroid that generates the polytope $P_{m,j}$. Note, given a region $Y \subseteq X$ in the state-space, its centroid C_Y is defined as

$$C_Y = \frac{1}{M_Y} \int_Y \mathbf{y} \rho(\mathbf{y}) d\mathbf{y},$$

where $M_Y := \int_Y \rho(\mathbf{y}) d\mathbf{y}$ is the total mass of Y , and $\rho(\mathbf{y})$ is the mass density function over Y . When the polytope Q_m is convex, the partitions are also convex. This in turn implies that the centroid of each partition is contained inside the polytope. For a fixed number of partitions n_m , a convex polytope Q_m can have more than one centroidal Voronoi partition. While implementing this method for kernel center selection, the term n_m corresponds to the number of centers. The subscript m corresponds to the sampling subset Ξ_m . The number of kernel centers depends on the samples collected in this method.

3.2 | Lloyd's algorithm

Lloyd's algorithm is used to construct the centroidal Voronoi tessellations for a given convex polytope Q_m and a fixed number of partitions n_m . It involves the following steps,

- (i) Choose an initial set of points P_m .
- (ii) Calculate the Voronoi partitions $\mathcal{V}(P_m)$ for the n_m points.
- (iii) Calculate the set of centroids $\{C_{P_{m,1}}, \dots, C_{P_{m,n_m}}\}$ of the Voronoi partitions.
- (iv) Set $P_m = \{C_{P_{m,1}}, \dots, C_{P_{m,n_m}}\}$ and go back to the second step.

The above set of steps are evaluated until convergence of centroids is achieved. The convergence of the algorithm for the convex case is proved in Reference 38.

3.3 | Implementation

The method we develop based on CVT and Lloyd's algorithm assumes that we have a finite sampling Ξ_m of the positive limit set $\omega^+(\mathbf{x}_0)$. We use this finite sampling Ξ_m to construct a region Q_m that encloses the positive limit set $\omega^+(\mathbf{x}_0)$. Thus, the working assumption is that while we do not know the exact form of $\omega^+(\mathbf{x}_0)$, we do know it is contained in Q_m . We then calculate the centroidal Voronoi partitions of the polygon and choose the kernel centers as the centroids of the partitions. In our implementation, we assume the mass density function as $\rho(\mathbf{q}) = 1$ for all $\mathbf{q} \in Q_m$ and $\rho(\mathbf{q}) = 0$ elsewhere. In the following discussion, we formalize this implementation.

Examples of the region Q_m for two different positive limit sets is shown in Figure 4. In the case (b) where the positive limit set $\omega^+(\mathbf{x}_0)$ is straight line, the region Q_m is nothing but the rectangle enclosing the set. For the case (a) where the positive limit set $\omega^+(\mathbf{x}_0)$ is a closed curve that is symmetric about the origin in the figure, the region Q_m is first formed by the joining the samples of the positive limit set to form a closed curve. The closed curve is then scaled to form a larger and smaller closed curves. We choose Q_m to be the region enclosed by the larger and smaller closed curves. As evident from Figure 4, the region Q_m is not always convex. Thus, the theory in the previous subsection is not strictly applicable. Let Q'_m be the convex hull of the polytope Q_m . We know that the Lloyd's algorithm converges for the convex case.³⁸ The mass density function is still equal to 1 on Q_m and 0 elsewhere. Suppose we choose n_m points in Q'_m and run the Lloyd's algorithm. As a result, we get a set of centroids P'_m that generate the centroidal Voronoi partition $\mathcal{V}(P'_m)$. Now we define the collection $\mathcal{V}(P_m) := \{P'_{m,1} \cap Q_m, \dots, P'_{m,n_m} \cap Q_m\}$. It is easy to see that $\mathcal{V}(P_m)$ is a centroidal Voronoi partition of the region Q_m generated by the centroids $P_m = P'_m$.

Thus, the Lloyd's algorithm indeed converges for the case in question. However, the polytopes in $\mathcal{V}(P_m)$ are not necessarily convex. And hence, the centroid $p_{m,i} \in P_m$ need not be contained in the polytope $P'_{m,i} \cap Q_m$ for $i = 1, \dots, n_m$. The centers need not even be contained in the region Q_m . This is certainly not desirable when implementing Lloyd's algorithm and CVT for problems like sensor location or multirobot coordination.³⁹ However, the goal of our problem is to choose kernel centers that are close to the positive limit set. In the following analysis, we show that with sufficient number of samples and careful selection of the region Q_m , we can often choose centers close to the positive limit set.

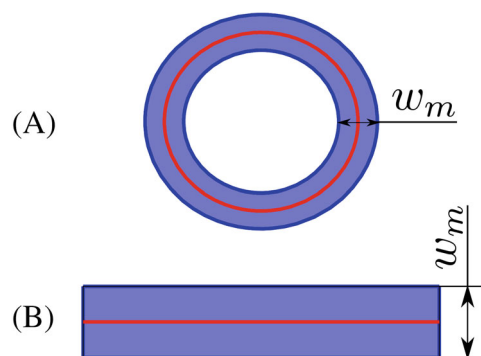


FIGURE 4 Examples of region Q_m constructed around the $\Xi_m \subseteq \omega^+(\mathbf{x}_0)$. In Subfigure (a), the samples are collected from a circle. Subfigure (b) shows the case where the samples are contained in a straight line. The red curves are formed by connecting the samples Ξ_m . The blue region represents the region Q_m

3.4 | Convergence for restricted cases

We restrict the following analysis to positive limit sets contained in \mathbb{R}^2 that are homeomorphic to a line or a circle. In other words, the positive limit set is an open or closed curve. With careful selection of Q_m , it is possible to show that we can choose kernel centers that approximate the positive limit set. The region Q_m is constructed such that the following conditions holds.

Condition 1. Associated with each Ξ_m is a region Q_m such that

1. the maximum width w_m of the region satisfies $w_m < r_m$, where $0 < r_m < r_{m-1}$ for all $m \in \mathbb{N}$,
2. the region Q_m is nested in Q_{m-1} for all $m \in \mathbb{N}$,
3. the sequence $\{r_m\}_{n=1}^{\infty}$ converges to 0,
4. for each r_m , there is an integer n_m such that the polytope $P_{m,j} \subseteq B_{cr_m}(C_{P_{m,j}})$ for all $j = 1, \dots, n_m$. Here, the term $B_{cr_m}(C_{P_{m,j}})$ is the closed ball of radius cr_m centered at the centroid $C_{P_{m,j}}$ that generates the polytope $P_{m,j}$ with c a fixed positive constant.

We can think of the maximum width w_m of the region Q_m given in Figure 4A as the Hausdorff distance between the inner and outer boundaries of the region Q_m . In the case of the region given in Figure 4B, the maximum width w_m corresponds to the Hausdorff distance between the two boundaries of the region Q_m that are parallel to the positive limit set.

Theorem 6. Suppose Condition 1 holds. Then $d_H(\omega^+(\mathbf{x}_0), P_m) \rightarrow 0$ as $m \rightarrow \infty$, where $d_H(\cdot, \cdot)$ is the Hausdorff distance, $\omega^+(\mathbf{x}_0)$ is the positive limit set and $P_m = \{C_{P_{m,1}}, \dots, C_{P_{m,n_m}}\}$ is the set of centroids that generate the CVT $\mathcal{V}(P_m)$.

Proof. We first note that the centroid of each polytope is contained in $B_{cr_m}(C_{P_{m,j}})$ since the ball is convex. Since the maximum width of the region w_m satisfies $w_m < r_m$, it is clear that $d_H(\omega^+(\mathbf{x}_0), Q_m) < r_m$. On the other hand, since the ball $B_{cr_m}(C_{P_{m,j}})$ contains the polytope $P_{m,j}$, we have $d_H(P_{m,j}, \{C_{P_{m,j}}\}) < cr_m$ for any $j = 1, \dots, n_m$. Note that the bound cr_m on $d_H(P_{m,j}, \{C_{P_{m,j}}\})$ is uniform. Also, recall that $Q_m = \bigcup_{j=1}^{n_m} P_{m,j}$, and $P_m = \bigcup_{j=1}^{n_m} \{C_{P_{m,j}}\}$. Thus, we have $d_H(Q_m, P_m) < cr_m$. Using triangle inequality, we get $d_H(\omega^+(\mathbf{x}_0), P_m) < (1+c)r_m$. Since $r_m \rightarrow 0$ as $m \rightarrow \infty$, we conclude that the centroids approach the positive limit set as $m \rightarrow \infty$. ■

The assumptions in the above theorem are very strong because of Condition 1. It is possible to relax some of the assumptions by considering the geometric properties of the partitions. But, from a practical standpoint, the maximum number of samples of the positive limit set is limited by the measurement equipment. This theorem provides a framework for an implementation that agrees with intuition—if new samples of the positive limit set are measured, choose Q_m such that r_m is reduced and number of kernel centers n_m are increased. For a given r_m , the number of kernel centers cannot be indefinitely increased. Consider the example in Figure 5. Due to numerical errors, the Lloyd's algorithm converges to a CVT in which the kernel centers do not lie on the positive limit set when n_m is large. On the other hand, the term r_m cannot be decreased indefinitely, since the region Q_m , built based on finite number of samples, may no longer contain the positive limit set. Thus, the number of samples collected restrict the effectiveness of this method.

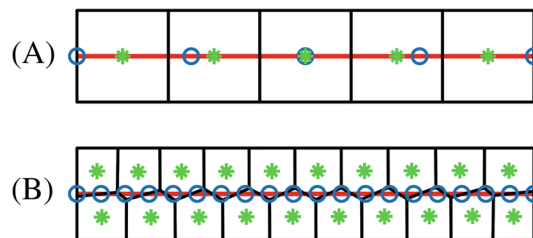


FIGURE 5 Increasing the number of kernel centers leads to completely different types of CVT while using the same Lloyd's algorithm. Case (a) considers 5 kernel centers. Case (b), on the other hand, considers 20 kernel centers. The markers o and * represent the initial positions and final converged positions of the kernel centers, respectively. The red line represents the limit set

To avoid CVTs that are similar to the one given in Figure 5B, we introduce the following condition. Let \bar{Q} represent the outer rectangle that is contained in \mathbb{R}^2 in Figure 5 and let $\bar{\mathcal{V}}_l$ represent the CVT made up of l horizontally stacked identical rectangles. Figure 5A depicts the CVT $\bar{\mathcal{V}}_5$ of \bar{Q} . The following condition inherently ensures that the kernel centers are evenly distributed in or near the positive limit set.

Condition 2. Let $l = 1, \dots, n_m$. For any possible l , consider an arbitrary collection of l polytopes $P_{m,i_1}, \dots, P_{m,i_l}$ in the partition $\mathcal{V}(P_m)$ such that each polytope is adjacent to at least one other polytope in the collection. The union of edges $\mathbb{E}(\mathcal{V}(P_m), P_{m,i_1} \cup \dots \cup P_{m,i_l})$ is homeomorphic to the union of edges $\mathbb{E}(\bar{\mathcal{V}}_l, \bar{Q})$ of the CVT $\bar{\mathcal{V}}_l$.

Algorithm 1. CVT based kernel center selection

Input: Ξ_m, n_m

Output: P_m

1. Initialize the constant r_m . Construct region Q_m such that the positive limit set $\omega^+(\mathbf{x}_0)$ is contained in Q_m .
 2. Choose n_m separate points in the convex hull of Q_m .
 3. Run the Lloyd's algorithm using the points chosen in Step 2 as the initial points.
 - (i) Calculate the Voronoi partitions $\mathcal{V}(P_m)$ for the n_m points.
 - (ii) Calculate the centroids $C_{P_{m,1}}, \dots, C_{P_{m,n_m}}$ of the Voronoi partitions $\mathcal{V}(P_m)$.
 - (iii) Set $P_m = \{C_{P_{m,1}}, \dots, C_{P_{m,n_m}}\}$ and go back to the Step 3 (i).
The above steps are repeated until convergence is achieved.
 4. If the CVT from Step 3 does not satisfy Condition 2, choose a constant s_m such that $s_m < r_m$. Set $r_m = s_m$ and go back to Step 2.
If the CVT satisfies Condition 2, choose the set of centroids of the CVT P_m as the kernel centers for the adaptive estimator.
-

Algorithm 1 shows the steps involved in implementing this method. Step 3 in the algorithm can be implemented using commercially available tools like MATLAB, which makes the algorithm extremely straightforward for implementation. The inputs to the algorithm are the samples Ξ_m and the number of kernel centers n_m . We iteratively choose r_m in the algorithm until Condition 2 is satisfied. The output of the algorithm is the set of kernel centers, which can be implemented in the adaptive estimator algorithm.

4 | METHOD 2: BASED ON KOHONEN SELF-ORGANIZING MAPS

The second approach presented in this article is based on Kohonen self-organizing maps (SOMs), which were first introduced by Teuvo Kohonen.⁴⁰ Self-organizing maps are typically used for applications like clustering data, dimensionality reduction, pattern recognition, and visualization. Thus, given a set of samples in the input space, these maps can be used to produce a collection of neurons on a low-dimensional manifold that represents the samples' distribution. In our problem, the input space is the state-space, and the samples are the state measurements. The neurons on the low-dimensional manifold are the kernel centers. The position of the kernel centers in the state-space are represented by the weight vectors that the SOM algorithm generates.

One of the critical features of self-organizing maps is that the underlying topology between the input space (the original dataset) and the output space is maintained. Intuitively, points that are close in the original dataset are mapped to neurons that are close to each other (in some predefined metric). For our problem, we want the kernel centers to be evenly spaced in the state-space in addition to being close to the measurement samples. To ensure this, we choose the initial set of kernel centers on a manifold that is homeomorphic to the positive limit set. This requires knowledge of the topology of the positive limit set. Before going over the details, let us take a look at the theory of Kohonen self-organizing maps.

Suppose we have the set of samples $\Xi_m = \{\xi_{m,1}, \dots, \xi_{m,q_m}\}$. In the context of this article, the set Ξ_m is assumed to be the set of samples of the positive limit set $\omega^+(\mathbf{x}_0)$. Let n_m represent the number of kernel centers $\mathbf{p}_{m,1}, \dots, \mathbf{p}_{m,n_m}$ we want to choose. We associate the i th kernel center with a weight vector $\mathbf{p}_{m,i}(t) \in \mathbb{R}^d$ for $i = 1, \dots, n_m$. Note that the weight vectors depend on time and at any given instant in time t , the weight vector is an element of \mathbb{R}^d . The neighborhood function \mathcal{N}_j

defines neighbors of the center j . The choice of the neighborhood function depends on the topology we want to define on the kernel centers. The neurons (or the kernel centers) are often chosen in the form of a linear grid or a 2D grid, and the neighbors in such grids are naturally defined. The Kohonen self-organizing map's implementation involves the following steps. We first randomly choose a sample $\xi_{m,k}$ from the sample set Ξ_m , where $k \in \{1, \dots, q_m\}$. We then determine the winning neuron—the kernel center that is closest to the sample $\xi_{m,k}$. The winning neuron i at a given instant t is the one which satisfies the condition

$$d(\xi_{m,k}, \mathbf{p}_{m,i}(t)) \leq d(\xi_{m,k}, \mathbf{p}_{m,j}(t)) \quad (4)$$

for $j = 1, \dots, n_m$, where $d(\cdot, \cdot)$ is the Euclidean metric. We now update the weight vectors using the evolution equation

$$\frac{d\mathbf{p}_{m,j}(t)}{dt} = \beta_j(t) \mathcal{N}_j(t, i) (\xi_{m,k} - \mathbf{p}_{m,j}(t)) \quad (5)$$

for $j = 1, \dots, n_m$. In the above equation, $0 \leq \beta_j(t) < 1$ defines the rate of convergence of the center j . The neighborhood function determines which neighbors of the node i get updated. For convergence, we require that $\beta_j(t) \rightarrow 0$ and $\mathcal{N}_j(t, i) \rightarrow 0$ as $t \rightarrow \infty$, for any $i, j \in \{1, \dots, n_m\}$. While implementing this algorithm, we can observe the SOM goes through a *topological ordering phase* during which the grid of neurons try to match the patterns if the sample in the input space before convergence.

The self-organizing map algorithm is easy to implement. Researchers have studied and proved the algorithm's convergence for cases like the 1D linear array, when the nodes are arranged on a line. A review of some of the theoretical results on self-organized maps are given in Reference 41.

4.1 | Implementation

To implement Kohonen self-organizing maps for kernel center selection, we modify the above-discussed algorithm. In some dynamical systems, the trajectory approaches the positive limit set but is never contained in the set. In such cases, we only have measurements of the states and not the samples of positive limit set. Furthermore, arbitrary selection of state-samples might result in picking points away from the positive limit set. This in turn affects the convergence of the kernel centers to points inside the positive limit set. Hence, as opposed to choosing random samples $\xi_{m,j}$ from the set Ξ_m , we use the state measurement $\mathbf{x}(t)$ at a given time instant to determine the winning node. We replace the term $\xi_{m,j}$ with $\mathbf{x}(t)$ in Equations (4) and (5). This change enables us to implement this method for a more general class of systems in real-time.

A Kohonen self-organizing map algorithm gives a low-dimensional representation of all samples (which include the ones that are outside the limit set). On the other hand, the objective of our problem is to choose kernel centers on the positive limit set such that they are spaced as uniformly as possible. To ensure this, we choose the topology of the output space to match that of the positive limit set. In other words, we choose the initial kernel centers and the neighborhood function such that the topology is homeomorphic to the positive limit set. For example, if the positive limit set is a closed curve in \mathbb{R}^2 , the initial weight vectors can be points on the unit circle, and the neighborhood function can be defined as

$$\mathcal{N}_j(t, i) = \begin{cases} 1 & \text{if } j \in \mathcal{T}, \\ 0 & \text{if } j \notin \mathcal{T}, \end{cases} \quad (6)$$

where the set \mathcal{T} is defined as $\mathcal{T} = \{i-1, i, i+1\}$ for $i \neq 1, n_m$. For $i = 1$ and $i = n_m$, we choose $\mathcal{T} = \{n_m, 1, 2\}$ and $\mathcal{T} = \{n_m - 1, n_m, 1\}$, respectively.

On top of the above modifications, we enforce the condition that, when we have samples of the positive limit set, the number of kernel centers or neurons n_m should be strictly less than q_m , the number of samples in the set Ξ_m . When n_m is equal to q_m , the kernel centers can converge to the samples. In the case where the positive limit set is a closed curve, this can be interpreted as a solution to the traveling salesman problem.⁴² To avoid convergence to the samples, we impose the above dimensionality reduction condition.

Algorithm 2 shows the steps involved in implementing this method. We present the algorithm for the case where the positive limit set is a closed curve. However, the algorithm can be extended easily for other types of positive limit sets. The neighborhood function for this case, defined by Equation (6), is inherently accounted in the algorithm.

Algorithm 2. Kohonen SOM based kernel center selection—closed curve case

Input: $\mathbf{x}(t)$, q_m

Output: $\{\mathbf{p}_{m,1}(T), \dots, \mathbf{p}_{m,n_m}(T)\}$

1. Choose the number of kernel centers n_m such that $n_m < q_m$. If $p_m = 0$, choose a positive integer for n_m .
2. Choose β_j such that $0 \leq \beta_j(t) < 1$ for $t \in [0, \infty)$ and $\beta_j(t) \rightarrow 0$ as $t \rightarrow \infty$ for all $j = 1, \dots, n_m$.
3. Initialize the weight vectors $\mathbf{p}_{m,j}$ as the points on a circle contained inside the closed curve.
4. Implement the Kohonen SOM algorithm for $t \in [0, T]$ for some $T > 0$.
 - (i) At time t , determine the winning neuron i that satisfies the condition

$$d(\mathbf{x}(t) - \mathbf{p}_{m,i}(t)) \leq d(\mathbf{x}(t) - \mathbf{p}_{m,j}(t))$$

for $j = \{1, \dots, n_m\}$, where $d(\cdot, \cdot)$ is the Euclidean metric.

- (ii) Define the set \mathcal{T} as $\mathcal{T} = \{i-1, i, i+1\}$ for $i \neq 1, n_m$. For $i = 1$ and $i = n_m$, choose $\mathcal{T} = \{n_m, 1, 2\}$ and $\mathcal{T} = \{n_m - 1, n_m, 1\}$, respectively.
- (iii) Update the weight vectors based on

$$\frac{d\mathbf{p}_{m,j}(t)}{dt} = \begin{cases} \beta_j(t) (\mathbf{x}(t) - \mathbf{p}_{m,j}(t)) & \text{if } j \in \mathcal{T} \\ 0 & \text{if } j \notin \mathcal{T} \end{cases}$$

for $j = 1, \dots, n_m$. This update happens until next state measurement. Go back to Step 4 (i) after the update.

Recall that in the case of CVT based method presented in the previous section, the samples are contained in the positive limit set, which meant the trajectory was contained in the positive limit set or converged to the set in finite time. Since we use the state measurement for the Kohonen SOM based approach, we can relax some of the requirements of the CVT based method. It is sufficient for the trajectory to converge to the positive limit set as $t \rightarrow \infty$. However, it is important to choose $\beta_j(t)$ such that the state trajectory converges to the positive limit set faster than the rate at which $\beta_j(t) \rightarrow 0$. If this is violated, the kernel centers will not converge to the positive limit set.

Note, in the Lloyd's algorithm, the distance between any two kernel centers is inherently ensured to remain uniform by the algorithm. This can be attributed to the way partitions are defined and the selection of the mass density function. On the other hand, the distribution of the converged kernel centers from the Kohonen SOM based algorithm depends on the distribution of the sampled measurements. If the state measurements are concentrated on a particular neighborhood of the positive limit set, implementing Algorithm 2 will result in the kernel centers being concentrated in or near the neighborhood.

5 | NUMERICAL ILLUSTRATION OF CENTER SELECTION METHODS

We illustrate the effectiveness of the two approaches explained above for two examples in this section. The first example is the undamped piezoelectric oscillator example considered in Section 2.6. The positive limit set in this case is almost symmetric about the axis after scaling of the states. The second example is a nonlinear oscillator which has a nonsymmetric positive limit set. We implement the above discussed methods for both cases and use the resulting kernel centers in the adaptive estimators. We use MATLAB `lloydsAlgorithm` function, developed by Aaron T. Becker's Robot Swarm Lab, for implementing Step 3 of Algorithm 1.

The function expects the boundary of a polygon as input and hence we approximate the region Q_m using a polygon as shown in Figures 6 and 9. In the adaptive estimator simulations, we use the Sobolev-Matern 3,2 kernel given in Section 2.6.

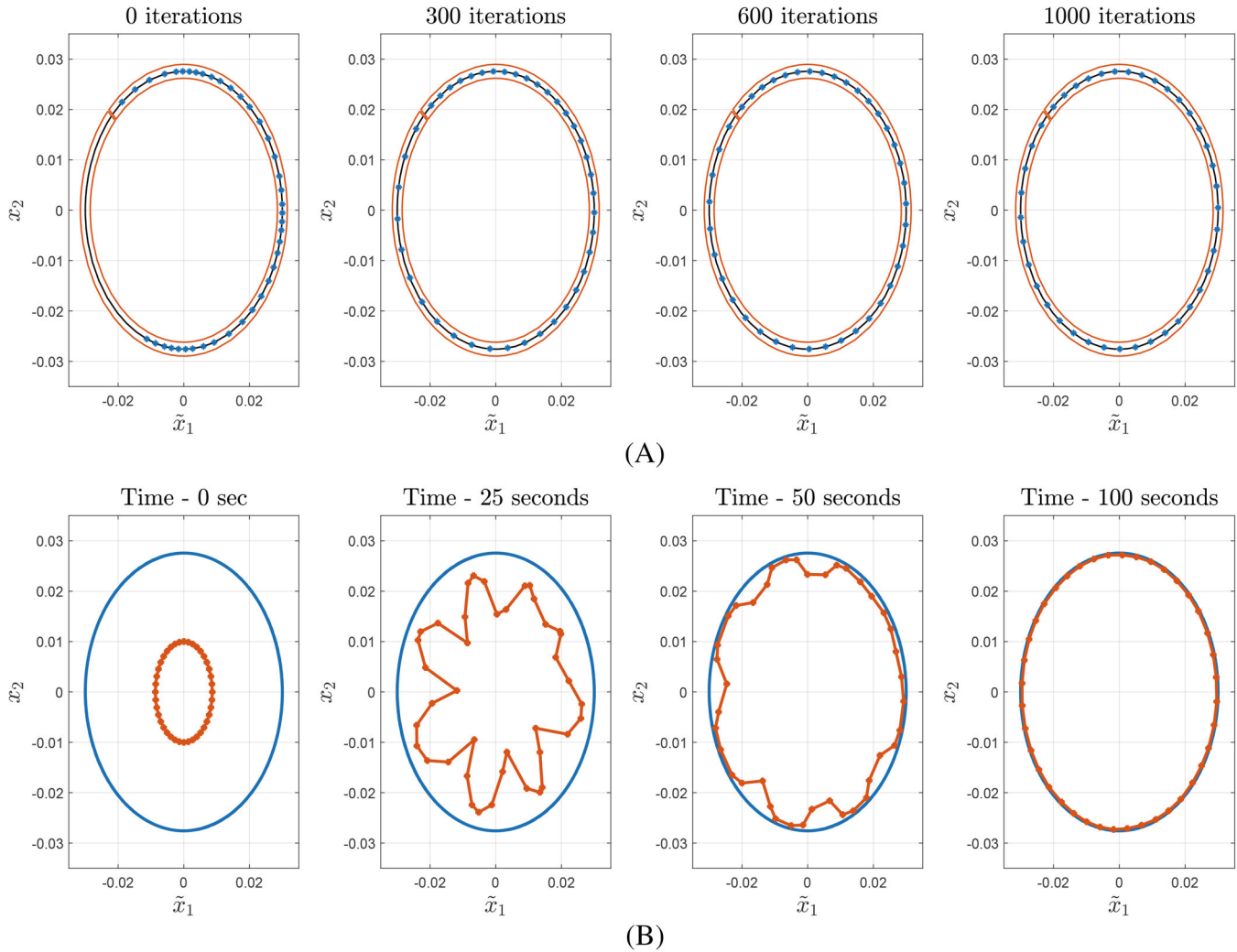


FIGURE 6 Algorithm outputs of Example 5.1. The marker * and the blue line represent the kernel centers and the limit set, respectively. (A) Algorithm 1 output; (B) Algorithm 2 output

5.1 | Example 1: Nonlinear piezoelectric oscillator

The first example we consider is the undamped nonlinear piezoelectric oscillator whose motion is governed by the Equation (3). We use the same values for the structural parameters as the ones used in the example in Section 2.6. We set the scaling factor $S = 0.02$ and initialized the states at $\mathbf{x}_0 = \{\tilde{x}_1(0), x_2(0)\}^T = \{0.03, 0\}^T$. Figure 6 shows how the kernel centers evolve while using Algorithms 1 and 2. We set the number of kernel centers as $n_m = 40$ for both of the algorithms. For implementing Algorithm 1, we first collect the set of samples Ξ_m of the positive limit set $\omega^+(\mathbf{x}_0)$. By connecting the samples in Ξ_m with straight lines, we form a closed curve which is represented by the blue line in Figure 6A. We then scale the closed curve by a factor of 1.1 and 0.9, thus forming concentric larger and smaller closed curves. We chose the region between these two closed curves as Q_m . Dividing the region Q_m as shown in Figure 6A results in a polygon, thus enabling us to use the `lloydAlgorithm` function in MATLAB. While implementing Algorithm 1, we chose $\beta_j(t) = 0.99$ for $t \leq 1000$ s and $\beta_j(t) = 0$ for $t > 1000$ s for all j . As evident from Figure 6, the CVT based approach and the Kohonen SOM based approach take 1000 iterations and 100 s, respectively to converge. It is clear that the kernel centers are more uniformly spaced than those picked arbitrarily in the example in Section 2.6. We subsequently use the converged kernel centers and simulate the adaptive estimator algorithm for $T = 300$ s. For the adaptive estimator, we set $l = 0.006$, $\Gamma = 0.001$ and initialized the parameters at $\alpha_i(t) = 0.0001$ for $i = 1 \dots, n_m$. Figures 7 and 8 shows the pointwise error $|f(\mathbf{x}) - \hat{f}(T, \mathbf{x})|$ obtained after using the kernel centers from the CVT and Kohonen SOM based approach. As expected, both the plots show that the error is $\mathcal{O}(10^{-4})$ over the positive limit set.

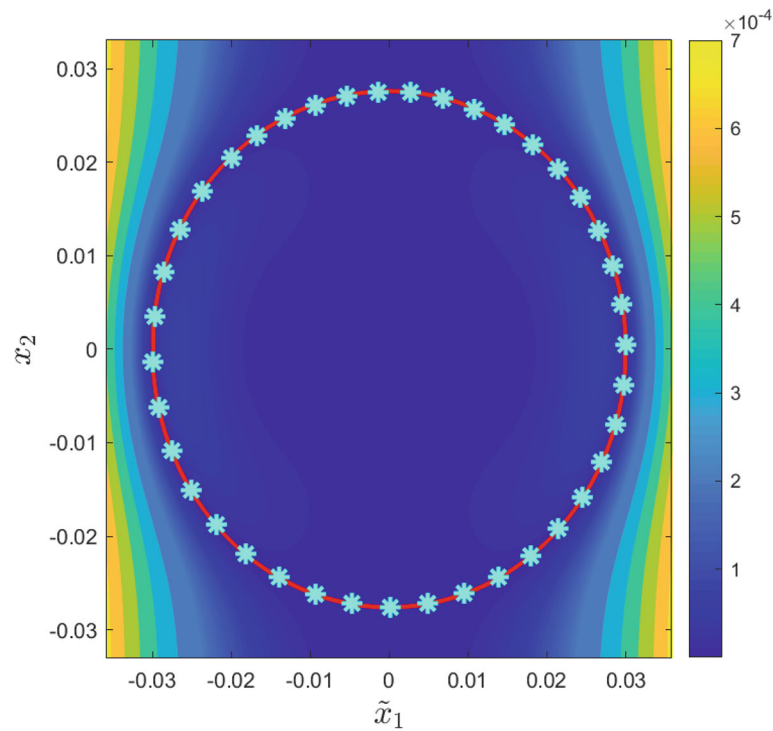


FIGURE 7 Kernel centers for Example 5.1 selected using Algorithm 1—Pointwise error $|f(\mathbf{x}) - \hat{f}_n(T, \mathbf{x})|$ obtained from adaptive estimator. The marker * and the red line represent the kernel centers and the limit set, respectively

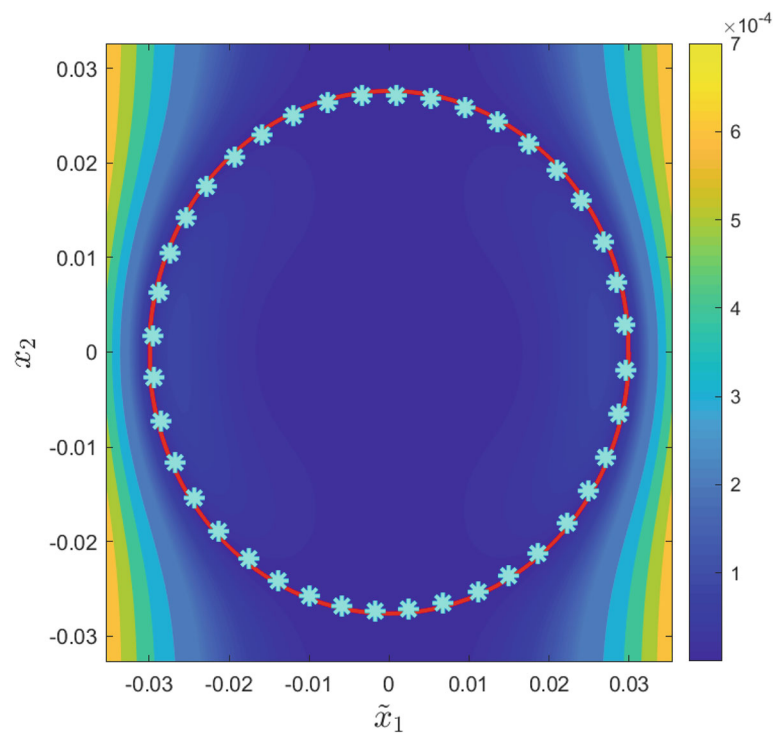


FIGURE 8 Kernel centers for Example 5.1 selected using Algorithm 2—Pointwise error $|f(\mathbf{x}) - \hat{f}_n(T, \mathbf{x})|$ obtained from adaptive estimator. The marker * and the red line represent the kernel centers and the limit set, respectively

5.2 | Example 2: Nonlinear oscillator

For the second example, we consider a nonlinear oscillator whose motion is governed by the equation

$$\begin{Bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{Bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -1 & 0.5 \end{bmatrix}}_A \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} + \underbrace{\begin{Bmatrix} 0 \\ 1 \end{Bmatrix}}_B \underbrace{(-x_1^2 x_2)}_{f(\mathbf{x}(t))}. \quad (7)$$

This system exhibits a more complex behavior than that in Example 5.1. First, the state trajectory is not contained in the positive limit set $\omega^+(\mathbf{x}_0)$, which is depicted as the blue, solid line in Figure 9. Note that the positive limit set is not symmetric. Refer Example 9.2.2 in Reference 43 for a detailed analysis of the nonlinear behavior of the oscillator. Here, we are interested in estimating the nonlinear function $f(\mathbf{x}(t)) = -x_1^2 x_2$.

Figure 9 shows the implementation of the CVT based and Kohonen SOM based kernel center selection methods for this problem. In both cases, we fixed number of kernel center as $n_m = 40$ and initialized the states at $\mathbf{x}_0 = \{x_1(0), x_2(0)\}^T = \{0, 2\}^T$. The polygon in Figure 9A for the CVT based approach is built similar to the method used for Example 5.1. For the Kohonen SOM approach, we set $\beta_j(t) = 0.99$ for $t \leq 1000$ s and $\beta_j(t) = 0$ for $t > 1000$ s for all j . As evident from the figures, the CVT and Kohonen SOM methods take 600 iterations and 200 s, respectively for convergence of the kernel

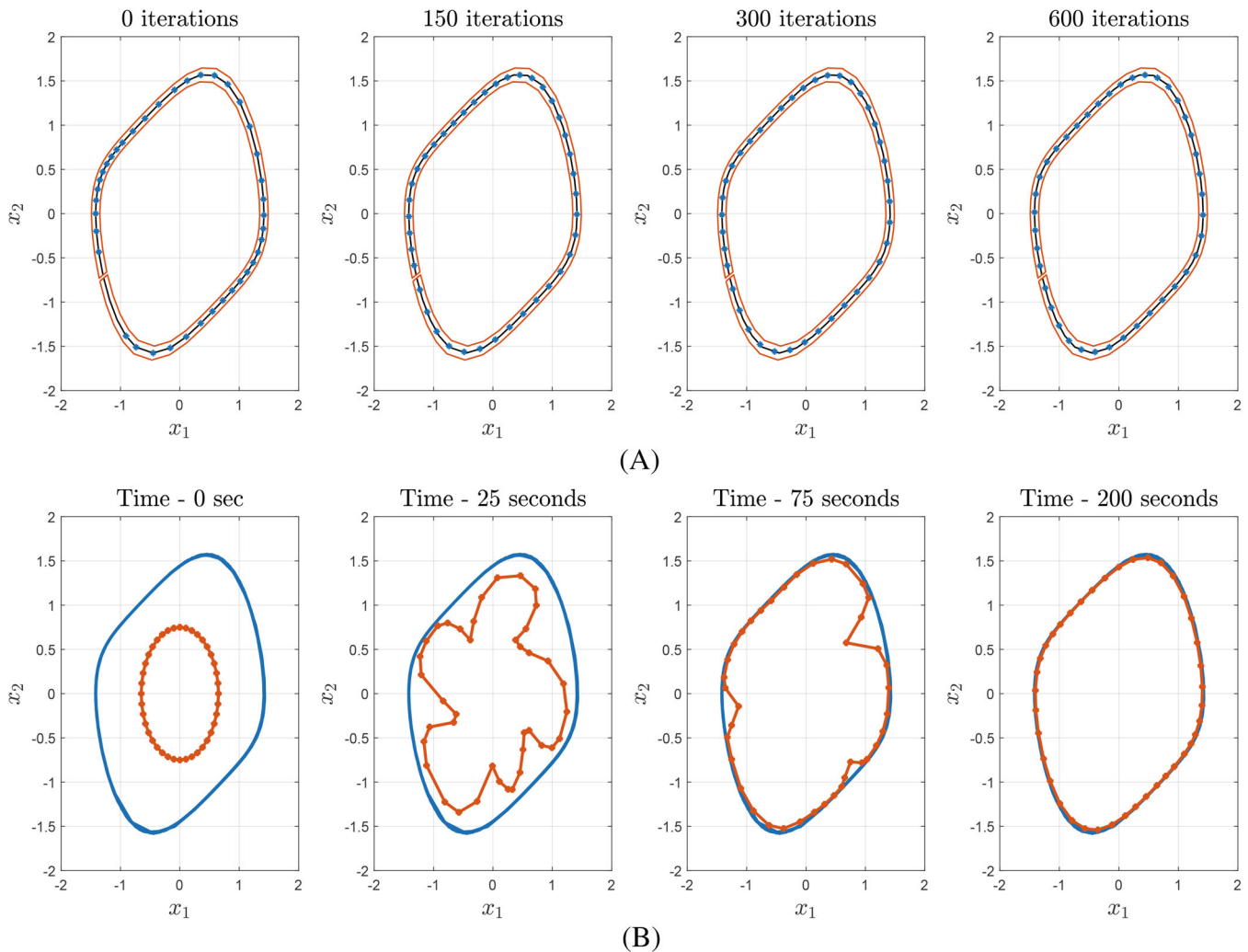


FIGURE 9 Algorithm outputs of Example 5.2. The marker * and the blue line represent the kernel centers and the limit set, respectively. (A) Algorithm 1 output; (B) Algorithm 2 output

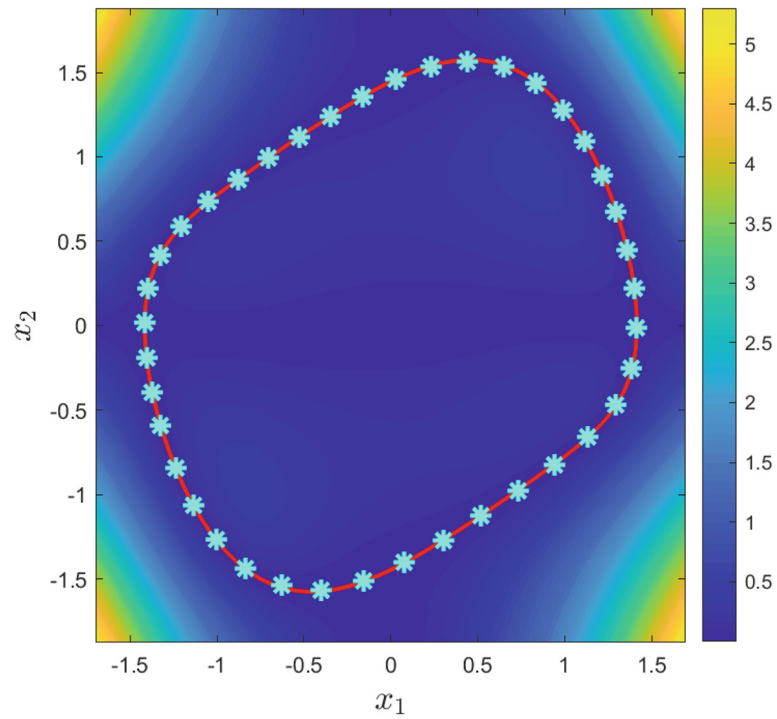


FIGURE 10 Kernel centers for Example 5.2 selected using Algorithm 1—Pointwise error $|f(\mathbf{x}) - \hat{f}_n(T, \mathbf{x})|$ obtained from adaptive estimator. The marker * and the red line represent the kernel centers and the limit set, respectively

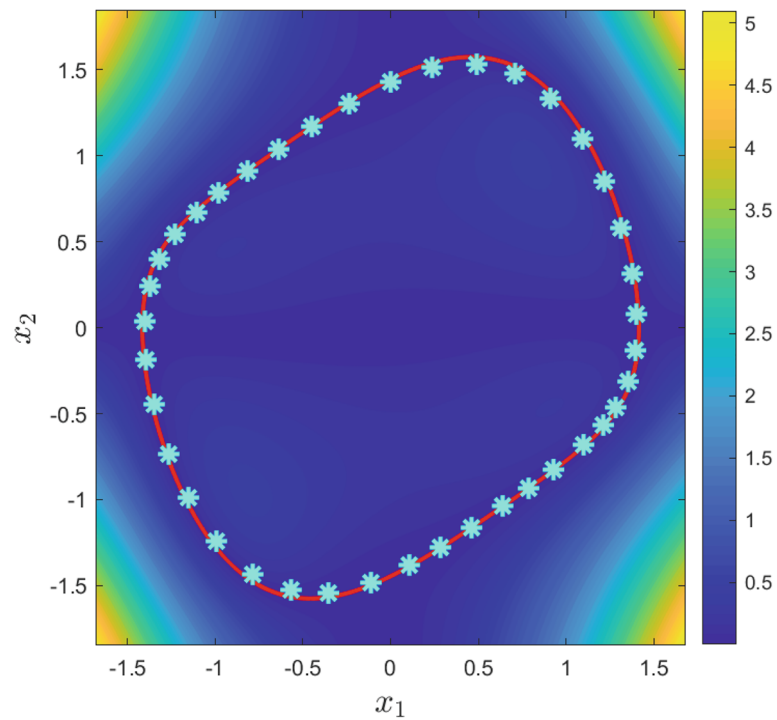



FIGURE 11 Kernel centers for Example 5.2 selected using Algorithm 2—Pointwise error $|f(\mathbf{x}) - \hat{f}_n(T, \mathbf{x})|$ obtained from adaptive estimator. The marker * and the red line represent the kernel centers and the limit set, respectively

centers. It is clear that the kernel centers from the CVT based algorithm are more uniformly placed than the output of the Kohonen SOM algorithm. This can be attributed to the fact the state measurement samples are not uniformly distributed and to the fact that the CVT method makes strong assumptions about the structure of Q_m . Since the distribution of the state measurement affects the results of the Kohonen SOM based approach, the kernel centers are not uniform in this case. However, when the kernel centers from these algorithms are implemented in the adaptive estimator, we obtain convergence on the positive limit set. Figures 10 and 11 show the pointwise error $|f(\mathbf{x}) - \hat{f}(T, \mathbf{x})|$ after implementing the adaptive estimator for $T = 300$ s using the kernel centers from the CVT and Kohonen SOM based kernel center selection approach, respectively. We set $l = 0.5$, $\Gamma = 0.001$ and initialized the parameters at $\alpha_i(t) = 0.0001$ for $i = 1 \dots, n_m$. As in Example 5.1, the error is the smallest over the positive limit set.

6 | CONCLUSION

In this article, we developed criteria for kernel center selection for use in the theory of infinite-dimensional adaptive estimation in reproducing kernel Hilbert spaces. We introduced two algorithms that use these criteria for kernel center selection. The first algorithm uses centroidal Voronoi tessellations and Lloyd's algorithm, and the second is based on Kohonen self-organizing maps. Both approaches provide a simple way to choose kernel centers for a specific class of nonlinear systems—systems in which state trajectory regularly visits the neighborhoods of the positive limit set. Furthermore, the proposed algorithms do not require explicit equations for the positive limit set and only assume a priori knowledge of its general features. In particular, the approach based on Lloyd's algorithm assumes that we can build a polytope in which the positive limit set is contained. On the other hand, the algorithm that uses Kohonen self-organizing maps requires knowledge of the positive limit set's topology. We illustrated the effectiveness of both algorithms using practical examples and numerical simulations. The approaches discussed in this article assume that approximations or estimates are constructed using a fixed number of kernel centers. It would be of great interest to develop techniques that iteratively add kernel centers in real-time while accounting for the persistence of excitation and fill-distance conditions.

ORCID

Sai Tej Paruchuri  <https://orcid.org/0000-0003-2372-3888>

REFERENCES

- Ioannou PA, Sun J. *Robust Adaptive Control*. Upper Saddle River, NJ: PTR Prentice-Hall; 1996;1(1).
- Sastry S, Bodson M. *Adaptive Control: Stability, Convergence and Robustness*. Dover Publications; 2011.
- Narendra KS, Annaswamy AM. *Stable Adaptive Systems*. Courier Corporation; 2012.
- Aronszajn N. Theory of reproducing kernels. *Trans Am Math Soc*. 1950;68(3):337-404. doi:10.2307/1990404
- Berlinet A, Thomas-Agnan C. *Reproducing kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media; 2011.
- Kurdila AJ, Narcowich FJ, Ward JD. Persistency of excitation in identification using radial basis function approximants. *SIAM J Control Optim*. 1995;33(2):625-642. doi:10.1137/S0363012992232555
- Kingravi HA, Chowdhary G, Vela PA, Johnson EN. Reproducing kernel Hilbert space approach for the online update of radial bases in neuro-adaptive control. *IEEE Trans Neural Netw Learn Syst*. 2012;23(7):1130-1141. doi:10.1109/TNNLS.2012.2198889
- Bobade P, Majumdar S, Pereira S, Kurdila AJ, Ferris JB. Adaptive estimation for nonlinear systems using reproducing kernel Hilbert spaces. *Adv Comput Math*. 2019;45(2):869-896. doi:10.1007/s10444-018-9639-z
- Lian J, Lee Y, Sudhoff SD, Zak SH. Self-organizing radial basis function network for real-time approximation of continuous-time dynamical systems. *IEEE Trans Neural Netw*. 2008;19(3):460-474. doi:10.1109/TNN.2007.909842
- Han H, Qiao J. A self-organizing fuzzy neural network based on a growing-and-pruning algorithm. *IEEE Trans Fuzzy Syst*. 2010;18(6):1129-1143. doi:10.1109/TFUZZ.2010.2070841
- Han H, Lu W, Hou Y, Qiao J. An adaptive-PSO-based self-organizing RBF neural network. *IEEE Trans Neural Netw Learn Syst*. 2018;29(1):104-117. doi:10.1109/TNNLS.2016.2616413
- Han HG, Qi C, Qiao JF. An efficient self-organizing RBF neural network for water quality prediction. *Neural Netw*. 2011;24(7):717-725. doi:10.1016/j.neunet.2011.04.006
- Han HG, Qiao JF, Chen QL. Model predictive control of dissolved oxygen concentration based on a self-organizing RBF neural network. *Control Eng Pract*. 2012;20(4):465-476. doi:10.1016/j.conengprac.2012.01.001
- Qiao JF, Han HG. Identification and modeling of nonlinear dynamical systems using a novel self-organizing RBF-based approach. *Automatica*. 2012;48(8):1729-1734. doi:10.1016/j.automatica.2012.05.034
- Sanner RM, Slotine JE. Gaussian networks for direct adaptive control. *IEEE Trans Neural Netw*. 1992;3(6):837-863. doi:10.1109/72.165588
- Volyanskyy KY, Haddad WM, Calise AJ. A new neuroadaptive control architecture for nonlinear uncertain dynamical systems: beyond σ - and e -modifications. Proceedings of the 47th IEEE Conference on Decision and Control; 2008; IEEE.

17. Kim YH, Lewis FL. Neural network output feedback control of robot manipulators. *IEEE Trans Robot Autom.* 1999;15(2):301-309. doi:10.1109/70.760351
18. Patino HD, Carelli R, Kuchen BR. Neural networks for advanced control of robot manipulators. *IEEE Trans Neural Netw.* 2002;13(2):343-354. doi:10.1109/72.991420
19. Nardi F. Neural Network Based Adaptive Algorithms for Nonlinear Control. PhD thesis. School of Aerospace Engineering, Georgia Institute of Technology; 2000.
20. Senanayake R, Tompkins A, Ramos F. Automorphing kernels for nonstationarity in mapping unstructured environments. In: Billard A, Dragan A, Peters J, Morimoto J, eds. *Proceedings of the 2nd Conference on Robot Learning. Volume 87 of Proceedings of Machine Learning Research.* PMLR; 2018:443-455.
21. Chowdhary G, Johnson E. Concurrent learning for convergence in adaptive control without persistency of excitation. Proceedings of the 49th IEEE Conference on Decision and Control (CDC); 2010; IEEE.
22. Kamalapurkar R, Reish B, Chowdhary G, Dixon WE. Concurrent learning for parameter estimation using dynamic state-derivative estimators. *IEEE Trans Automat Contr.* 2017;62(7):3594-3601. doi:10.1109/TAC.2017.2671343
23. Modares H, Lewis FL, Naghibi-Sistani M. Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks. *IEEE Trans Neural Netw Learn Syst.* 2013;24(10):1513-1525. doi:10.1109/TNNLS.2013.2276571
24. Chowdhary G, How J, Kingravi H. Model reference adaptive control using nonparametric adaptive elements. Proceedings of AIAA Guidance, Navigation, and Control and Co-located Conference. American Institute of Aeronautics and Astronautics; 2012
25. Chowdhary G, Kingravi HA, How JP, Vela PA. Bayesian nonparametric adaptive control of time-varying systems using Gaussian processes. Proceedings of the 2013 American Control Conference; 2013; IEEE.
26. Grande RC, Chowdhary G, How JP. Nonparametric adaptive control using Gaussian Processes with online hyperparameter estimation. Proceedings of the 52nd IEEE Conference on Decision and Control; 2013; IEEE.
27. Abdollahi A, Chowdhary G. Adaptive-optimal control under time-varying stochastic uncertainty using past learning. *Int J Adapt Control Signal Process.* 2019;33(12):1803-1824. doi:10.1002/acs.3061
28. Liu M, Chowdhary G, de Silva BC, Liu S, How JP. Gaussian processes for learning and control: a tutorial with examples. *IEEE Control Syst Mag.* 2018;38(5):53-86. doi:10.1109/MCS.2018.2851010
29. Guo J, Paruchuri ST, Kurdila AJ. Persistence of excitation in uniformly embedded reproducing kernel Hilbert (RKH) spaces. Proceedings of the 2020 American Control Conference; 2020:4539-4544; IEEE.
30. Guo J, Paruchuri ST, Kurdila AJ. Partial Persistence of Excitation in RKHS Embedded Adaptive Estimation; 2020. arXiv preprint arXiv:2002.07963.
31. Paruchuri ST, Guo J, Kurdila AJ. Sufficient conditions for parameter convergence over embedded manifolds using kernel techniques. *IEEE Trans Automat Contr.* 2022;1-1. doi:10.1109/TAC.2022.3148716
32. Kurdila AJ, Guo J, Paruchuri ST, Bobade P. Persistence of excitation in reproducing kernel Hilbert spaces, positive limit sets, and smooth manifolds; 2019. arXiv preprint arXiv:190912274.
33. Guo J, Paruchuri ST, Kurdila AJ. Approximations of the reproducing kernel Hilbert space (RKHS) embedding method over manifolds. Proceedings of the 59th IEEE Conference on Decision and Control; 2020; IEEE.
34. Wendland H. *Scattered Data Approximation.* Vol 17. Cambridge University Press; 2004.
35. De Vito E, Rosasco L, Toigo A. Learning sets with separating kernels. *Appl Comput Harmon Anal.* 2014;37(2):185-217. doi:10.1016/j.acha.2013.11.003
36. Paruchuri ST, Guo J, Kurdila A. Reproducing kernel Hilbert space embedding for adaptive estimation of nonlinearities in piezoelectric systems. *Nonlinear Dyn.* 2020;101(2):1397-1415.
37. Rasmussen CE. Gaussian processes in machine learning. In: Bousquet O, von Luxburg U, Rätsch G, eds. *Advanced Lectures on Machine Learning. ML 2003. Lecture Notes in Computer Science,* vol 3176. Berlin, Heidelberg: Springer; 2004. doi:10.1007/978-3-540-28650-9_4
38. Cortes J, Martinez S, Karatas T, Bullo F. Coverage control for mobile sensing networks. *IEEE Trans Robot Autom.* 2004;20(2):243-255.
39. Breitenmoser A, Schwager M, Metzger J, Siegwart R, Rus D. Voronoi coverage of non-convex environments with a group of networked robots. Proceedings of the 2010 IEEE International Conference on Robotics and Automation; 2010; IEEE.
40. Kohonen T. *Self-organization and Associative Memory.* Vol 8. Springer Science & Business Media; 2012.
41. Cottrell M, Fort JC, Pagès G. Theoretical aspects of the SOM algorithm. *Neurocomputing.* 1998;21(1):119-138. doi:10.1016/S0925-2312(98)00034-4
42. Brocki Ł, Koržinek D. Kohonen self-organizing map for the traveling salesperson problem. In: Jabłoński R, Turkowski M, Szewczyk R, eds. *Recent Advances in Mechatronics.* Springer; 2007:116-119.
43. Hubbard JH, West BH. *Differential Equations: A Dynamical Systems Approach: Ordinary Differential Equations.* Vol 5. Springer; 2013.

How to cite this article: Paruchuri ST, Guo J, Kurdila A. Kernel center adaptation in the reproducing kernel Hilbert space embedding method. *Int J Adapt Control Signal Process.* 2022;36(7):1562-1583. doi: 10.1002/acs.3407