Rates of Convergence in a Class of Native Spaces for Reinforcement Learning and Control

Ali Bouland, Shengyuan Niu, Sai Tej Paruchuri, *Member, IEEE*, Andrew Kurdila, John Burns, *Life Fellow, IEEE*, and Eugenio Schuster, *Member, IEEE*

***Abstract*—This letter studies convergence rates for some value function approximations that arise in a collection of reproducing kernel Hilbert spaces (RKHS) $\mathsf{H}(\Omega)$. By casting an optimal control problem in a specific class of native spaces, strong rates of convergence are derived for the operator equation that enables offline approximations that appear in policy iteration. Explicit upper bounds on error in value function and control law approximations are derived in terms of power function $\mathcal{P}_{\mathsf{H},\mathsf{N}}$ for the space of finite dimensional approximants H_N in the native space $\mathsf{H}(\Omega)$. These bounds exhibit a distinctive geometric nature, refine and build upon some well-known, now classical results concerning the convergence of approximations of value functions.**

***Index Terms*—Reinforcement learning, optimal control, reproducing kernel.**

I. Introduction

CONSIDER a nonlinear system that is governed by the ordinary differential equations

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

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the input, and $f : \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ are known functions. We assume $f(0) = 0$ and are interested in a regulation problem that drives this system to the origin.

We seek an admissible state feedback function $\mu : x \mapsto \mu(x)$ that can stabilize the system presented above. Often, we restrict consideration to feedback functions μ that leave some subset of interest Ω positive invariant. In addition to stabilizing the system, a feedback function must then be continuous on Ω and satisfy $\mu(0) = 0$ to consider it admissible on the subset Ω of the state space. The cost associated with an admissible control policy μ is consequently defined as

$$V_\mu(x_0) = \int_0^\infty r(x(\tau), \mu(x(\tau)))d\tau, \tag{2}$$

where $r(x, \mu) = Q(x) + \mu^\mathsf{T} R\mu$, $Q(x)$ is a positive definite function and R is a symmetric positive definite matrix. Assuming that the function $V_\mu(x) : x \to V_\mu(x)$ is continuously differentiable, we can write down its differential Lyapunov-like equation in terms of the Hamiltonian as

$$\mathcal{H}(x, \mu, \nabla V) := \underbrace{(f(x) + g(x)\mu(x))^\mathsf{T}\nabla V_\mu(x)}_{:=(AV_\mu)(x)} + r(x, \mu(x)) = 0 \tag{3}$$

where ∇ denotes the gradient operator. The goal of optimal control is to choose a control policy μ^* such that $V_{\mu^*}(x_0)$ is minimized. The function V_{μ^*} is commonly referred to as the value function. Standard optimal control analysis [1], [2], [3] shows that the value function satisfies the Hamilton-Jacobi-Bellman (HJB) equation

$$0 = min_{\mu\in M(\Omega)}\mathcal{H}(x, \mu, \nabla V_{\mu^*}), \tag{4}$$

which is equivalent to $0 = \mathcal{H}(x^*, \mu^*, \nabla V_{\mu^*})$, where μ^* is given by $\mu^* = -\frac{1}{2}R^{-1}g^\mathsf{T}\nabla V_{\mu^*}$, and x^* is the optimal trajectory generated by μ^*. Once the HJB equation is solved for the optimal value function, the optimal controller can be found using this equation for μ^*.

In general, the HJB equation is a nonlinear partial differential equation that is difficult to solve, and the technical literature that studies this problem is vast. Among this collection of work, a few "now-classic" papers related to the study of Galerkin approximations are particularly relevant to this letter. These include the notable early efforts in [4], [5]. The highly cited work in [6] builds on the earlier work on Galerkin approximations to handle saturating actuators, which is subsequently used to form the theoretical foundation in [7] and many subsequent works [3], [8], [9], [10].

The treatises [3] and [10] give excellent accounts of the theory for reinforcement learning (RL) methods, and recent surveys include [8], [9]. One popular method of approximating the solution of the HJB equation is the actor-critic method. It entails an iterative approach of approximating the value function using the critic, then the actor uses the value approximation to get a control policy estimate, and the process

Manuscript received 15 September 2023; revised 16 November 2023; accepted 2 December 2023. Date of publication 15 December 2023; date of current version 22 January 2024. Recommended by Senior Editor J. Daafouz. *(Corresponding author: Ali Bouland.)*

Ali Bouland, Shengyuan Niu, and Andrew Kurdila are with the Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24060 USA (e-mail: bouland@vt.edu).

Sai Tej Paruchuri and Eugenio Schuster are with the Department of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, PA 18015 USA (e-mail: saitejp@lehigh.edu).

John Burns is with the Department of Mathematics, Virginia Tech, Blacksburg, VA 24060 USA (e-mail: jaburns@vt.edu).

Digital Object Identifier 10.1109/LCSYS.2023.3343439

repeats. A second common method is policy iteration (PI), which requires full knowledge of the system dynamics but allows an offline calculation of the optimal control law. The effectiveness of both methods relies on the convergence of the estimates of the value function. Recent works, such as [8], [11], [12], [13], have explored iteration convergence rates in terms of the iteration number but do not consider the explicit effects of approximation error on performance.

This letter explores the effects of approximation error, and derives bounds on the error between the estimates of the value function and the corresponding control law. These bounds are explicit in terms of the number of bases N used, and the geometric placement of centers that determines the bases.

A. Summary of New Results

As is often carried out in RL [3], [10], we can motivate this letter strategy by recalling the structure of PI. When the feedback function μ_i is known, we define the differential operator A to be given by $(Av)(x) \coloneqq (f(x) + g(x)\mu(x))^{\mathsf{T}}\nabla v(x)$ and $b(x) = -r(x, \mu(x))$. We then define v_i as the solution to the partial differential equation

$$\begin{aligned}(Av_i)(x) = b(x) &\coloneqq -r(x, \mu_i(x)),\\ v_i(0) &= 0.\end{aligned} \tag{5}$$

When v_i is determined from the above equation, we can subsequently define a new feedback law μ_{i+1} from the identity

$$\mu_{i+1}(x) = -\frac{1}{2}R^{-1}g^{\mathsf{T}}(x)\nabla v_i(x). \tag{6}$$

Setting $i \to i+1$ and repeating these steps generates a sequence of iterates $\{(\mu_i, v_i)\}_{i\in\mathbb{N}}$ that approximate the optimal functions μ^* and V^* that satisfy the HJB equations [1], [3], [6].

This letter derives *rates of convergence* for approximations of the solution v of the partial differential equation $Av = b$, defined in (5), for a given $\mu(x)$. It also provides rates of convergence for the controller μ_{i+1} approximation error generated by the PI method. Under the hypothesis that the solution $v \in H$, where H is a reproducing kernel Hilbert space (RKHS), we describe precise conditions on the reproducing kernel $\mathfrak{K}$ associated with H that ensures

$$\|v - v_N\|_H \le O\left(\sup_{x\in\Omega}\sqrt{\mathfrak{K}(x,x) - \mathfrak{K}_N(x,x)}\right).$$

In the above inequality, v_N is an approximate solution contained in the finite dimensional space $H_N \coloneqq \operatorname{span}\{\mathfrak{K}(\cdot, \xi_i) \in H \mid \xi_i \in \Xi_N\}$ determined by the N centers $\Xi_N \coloneqq \{\xi_1, \ldots, \xi_N\} \subset \Omega$. In this equation $\mathfrak{K}_N$ is the known reproducing kernel of H_N. We emphasize the following:

1) The above bound makes explicit the relationship of the center locations Ξ_N to the error in solutions of the operator equation.
2) For some popular kernels it is possible to bound the above expression in terms of the fill distance $h_{\Xi_N,\Omega} \coloneqq \sup_{x\in\Omega}\inf_{\xi_i\in\Xi_N}\|x - \xi_i\|$ of centers Ξ_N in the set Ω,

$$\|v - v_N\|_H \le O(h^s_{\Xi_N,\Omega}),$$

where s is a parameter that measures the regularity of the kernel $\mathfrak{K}$. Thus, the rate of convergence of the approximation error depends on the *smoothness* of the basis and the *geometric distribution* of the centers in $\Xi_N \subset \Omega$ that define the basis.

II. Theoretical Foundations

A. Symbols and Definitions

In this letter $\mathbb{R}$ and $\mathbb{R}^+$ are the real numbers and nonnegative real numbers, respectively. The non-negative integers are denoted $\mathbb{N}_0$, while the positive integers are $\mathbb{N}$. When U, V are normed vector spaces, $\mathcal{L}(U, V)$ is the normed vector space of bounded linear operators from U to V, and we just write $\mathcal{L}(U)$ for $\mathcal{L}(U, U)$. The range of an operator T is denoted $R(T)$ and the nullspace of T is written $N(T)$. The Lebesgue spaces $L^p(\Omega)$ are equipped with the usual norms

$$\|f\|_{L^p(\Omega)} \coloneqq \begin{cases}\left(\int_\Omega |f(x)|^p dx\right)^{1/p} & 1 \le p < \infty,\\ \operatorname{ess\,sup}\{|f(x)| \mid x \text{ a.e. in } \Omega\} & p = \infty.\end{cases}$$

B. Reproducing Kernels and Native Spaces

A real-valued native space, denoted as $H(\Omega)$ over a set Ω, is defined using a reproducing kernel $\mathfrak{K}(\cdot,\cdot) : \Omega \times \Omega \to \mathbb{R}$. This kernel, a Mercer kernel, is continuous, symmetric, and of positive type, which means that for any collection $\Xi_N \subset \Omega$ of N points, the Gramian matrix $\mathbb{K}(\Xi_N, \Xi_N) \coloneqq [\mathfrak{K}(\xi_i, \xi_j)] \in \mathbb{R}^{N\times N}$ is positive semidefinite. Once such a kernel is selected, the native space $H(\Omega)$ is defined as the closed linear span of the kernel sections $\mathfrak{K}_x(\cdot) \coloneqq \mathfrak{K}(x, \cdot)$,

$$H(\Omega) \coloneqq \overline{\operatorname{span}\{\mathfrak{K}_x \mid x \in \Omega\}}. \tag{7}$$

A few properties of the evaluation functional $E_x : H(\Omega) \to \mathbb{R}$ play a particularly important role in this letter. By definition, the evaluation functional satisfies $E_x f \coloneqq f(x)$ for all $f \in H(\Omega)$, and it is a bounded operator from $H(\Omega) \to \mathbb{R}$. Every native space satisfies the reproducing formula that connects the evaluation functional to inner products via $E_x f = f(x) = (f, \mathfrak{K}_x)_H$ for all $f \in H(\Omega), x \in \Omega$. Moreover, since E_x is a bounded operator, its adjoint $E_x^* \coloneqq (E_x)^* : \mathbb{R} \to H(\Omega)$ is also a bounded linear operator. It is given by the formula $E_x^*\alpha \coloneqq \mathfrak{K}_x\alpha$ for all $\alpha \in \mathbb{R}, x \in \Omega$.

In this letter, we always assume that the kernel $\mathfrak{K}(\cdot,\cdot)$ is bounded on the diagonal. That is, it is assumed that there is a $\bar{\mathfrak{K}} > 0$ such that $\mathfrak{K}(x, x) \le \bar{\mathfrak{K}}^2$ for all $x \in \Omega$. This ensures that all the functions in $H(X)$ are bounded, and that the evaluation operator E_x is uniformly bounded $\|E_x\| \le \bar{\mathfrak{K}}$ for all $x \in \Omega$, and that we have the continuous embedding $H(\Omega) \hookrightarrow C(\Omega)$. Many popular kernels are bounded on the diagonal including the exponential, inverse multiquadric, Wendland, and Sobolev-Matérn kernels [14].

1) Derivatives in Native Spaces: When $\mathfrak{K}$ is a Mercer kernel, having smoothness $\mathfrak{K} \in C^{2s}(\Omega \times \Omega)$ with $s \in \mathbb{N}$, that defines the native space $H(\Omega)$, it is possible to express the action of the partial derivative operator D^α on functions in $H(\Omega)$ in terms of the partial derivatives of the kernel. Suppose we fix y and are interested in partial derivatives with respect to x. To compute partial derivatives of the kernel, we interpret a multiindex $\alpha = (\alpha_1, \ldots, \alpha_d, \alpha_{d+1}, \ldots, \alpha_{2d}) \in \mathbb{N}_0^{2d}$ as having all zeros in the last d entries, so that $\alpha \coloneqq (\alpha_1, \ldots, \alpha_d, 0, \ldots, 0) \in \mathbb{N}_0^{2d}$ and

$$\begin{aligned}\left(D^\alpha \mathfrak{K}\right)_x(y) &\coloneqq \left(D^\alpha \mathfrak{K}\right)(x, y)\\ &\coloneqq \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}, \ldots, \partial x_d^{\alpha_d}}\mathfrak{K}(x_1, \ldots, x_d, y_1, \ldots, y_d) \quad \forall x, y \in \Omega.\end{aligned}$$

Reference [15, Th. (1)] specifies necessary conditions for the kernel, which entail it being a Mercer kernel that is sufficiently smooth, to prove that the derivative operator is a bounded operator on the native space and is vital to proving the results of this letter. Specifically, under the hypotheses described in the theorem, we have $(D^\alpha h)(x) = ((D^\alpha \mathfrak{K})(x,\cdot), h)_{H(\Omega)} = (D^\alpha_x \mathfrak{K}, h)_{H(\Omega)}$ for all $x \in \Omega$, $h \in H(\Omega)$ and $\sum \alpha_i \le s$.

2) Approximation in Native Spaces: To approximate function in $H(\Omega)$, we define $H_N := \text{span}\{\mathfrak{K}_{\xi_i} \mid \xi_i \in \Xi_N\} \subseteq H(\Omega)$ the space of approximants constructed using kernel sections defined in terms of the N locations $\Xi_N \subset \Omega$. Let Π_N be the $H(\Omega)$-orthogonal projection onto H_N. It is known that we have the general bound

$$\epsilon_{N,f}(x) := |E_x(I - \Pi_N)f| \le \mathcal{P}_{H,N}(x)\|f\|_{H(\Omega)}$$

for all $f \in H(\Omega)$ and $x \in \Omega$, where the power function $\mathcal{P}_{H,N}(x)$ is defined by $\mathcal{P}_{H,N}(x) := \sqrt{\mathfrak{K}(x,x) - \mathfrak{K}_N(x,x)}$ for all $x \in \Omega$. The kernel $\mathfrak{K}_N(\cdot,\cdot)$ is the reproducing kernel of H_N with

$$\begin{aligned}\mathfrak{K}_N(x,y) &:= (\Pi_N \mathfrak{K}_x, \Pi_N \mathfrak{K}_y)_{H(\Omega)} \\ &= \mathfrak{K}^{\mathsf{T}}_{\Xi_N}(x)\mathbb{K}^{-1}(\Xi_N,\Xi_N)\mathfrak{K}_{\Xi_N}(y), \end{aligned} \tag{8}$$

where $\mathfrak{K}_{\Xi_N}(\cdot) = \begin{bmatrix}\mathfrak{K}_{\xi_1}(\cdot) & \cdots & \mathfrak{K}_{\xi_n}(\cdot)\end{bmatrix}^{\mathsf{T}}$. This expression is used in a few different places in this letter.

III. Offline Approximation in a Native Space

A. The Operator Framework in a Native Space

We carry out value function approximation and subsequent analysis by first posing (3) as an operator equation. We define the differential operator A as

$$(Av)(x) := (f(x) + g(x)\mu(x))^{\mathsf{T}}\nabla v(x) \quad \text{for all } x \in \Omega,$$

whenever v is sufficiently smooth. Note that the operator ∇ in the above equation is defined in the usual way, with

$$\nabla f := \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_d}\right)^{\mathsf{T}} := \left(D^{e_1}f, \ldots, D^{e_d}f\right)^{\mathsf{T}},$$

where $D^\alpha(\cdot)$ is defined in Section II-B1 for any multiindex $\alpha \in \mathbb{N}_0^d$ and e_k is the canonical multiindex obtained by setting the k^{th} entry to one and all other entries to zero. The next theorem expresses some mapping properties of the operator A essential to our approximation schemes below.

Theorem 1: Let the hypotheses of [15, Th. 1] hold and further suppose that μ and f_i, g_i for $1 \le i \le d$ are multipliers for $H(\Omega)$. Then

1) The operator $A : H(\Omega) \to L^2(\Omega)$ is bounded, linear, and compact.
2) The adjoint operator $A^* : L^2(\Omega) \to H(\Omega)$ has the representation

$$\begin{aligned}\left(A^*h\right)(y) &:= \int_\Omega (\nabla_x \mathfrak{K}(y))^{\mathsf{T}}(f(x) + g(x)\mu(x))h(x)dx \\ &:= \int_\Omega \ell^*(y,x)h(x)dx\end{aligned}$$

for any $y \in \Omega$ and $h \in L^2(\Omega)$.
3) Considered as an operator $A^* : L^2(\Omega) \to H(\Omega)$, the operator A^* is compact.

Proof: The proof can be found in [16]. ∎

As discussed in Section I, PI is based on the recursive solution of the operator equation $Av = b \in L^2(\Omega)$. If $b \in R(A)$, the above equation has a solution, and if $N(A) = 0$, it is unique. In any case the operator $(A|_{N(A)^\perp})^{-1} : R(A) \to N(A)^\perp$ is well-defined. However, the operator $(A|_{N(A)^\perp})^{-1} : N(A)^\perp \to (R(A), L^2(\Omega))$ is not bounded in general since $A : H(\Omega) \to L^2(\Omega)$ is compact. This complicates approximations.

A common way to approximate the solution of such an equation is to seek the minimum $v^* \in H(\Omega)$ of the offline optimization problem

$$v^* = \text{argmin}_{v \in H(\Omega)} J(v) := \frac{1}{2}\|Av - b\|^2_{L^2(\Omega)}. \tag{9}$$

When we rewrite the cost functional in the form

$$\begin{aligned}&\frac{1}{2}\|Av - b\|^2_{L^2(\Omega)} \\ &\quad = \frac{1}{2}(A^*Av, v)_{H(\Omega)} - (v, A^*b)_{H(\Omega)} + \frac{1}{2}(b,b)_{L^2(\Omega)},\end{aligned}$$

we can calculate its Frechet derivative $DJ(v) : H(\Omega) \to H(\Omega)$ that satisfies

$$(DJ(v), w)_{H(\Omega)} := (A^*Av - A^*b, w)_{H(\Omega)} = (A^*A\tilde{v}, w)_{H(\Omega)}$$

for all directions $w \in H(\Omega)$, with $A\tilde{v} := Av - b$. Therefore, a minimizer satisfies the operator equation $A^*Av = A^*b$, or

$$\mathcal{A}v = y \tag{10}$$

where $\mathcal{A} = A^*A : H(\Omega) \to H(\Omega)$ and $y = A^*b \in H(\Omega)$. Offline approximations of the solution of the above operator equation can be interpreted as approximations of the pseudoinverse solution $V^* := \mathcal{A}^\dagger b \equiv (A^*A)^{-1}A^*b$. The pseudoinverse operator $\mathcal{A}^\dagger$ is well-defined since $\mathcal{A}$ is self-adjoint, compact, and nonnegative [17].

B. Offline Approximations

Operator equation (10) is defined in terms of the bounded, linear, compact operator $\mathcal{A} : H(\Omega) \to R(A^*) := W(\Omega) \subseteq H(\Omega) \subset L^2(\Omega)$. Since $y = A^*b \in R(A^*) := W(\Omega)$, (10) always has a solution. It will be unique if $\mathcal{A}$ is injective, and in this case $\mathcal{A}^{-1}$ is a well-defined operator. However, when $\mathcal{A}^{-1}$ exists it is generally not a bounded operator.

Here we assume that bases used for approximation are defined in terms of kernel sections located at the N centers $\Xi_N := \{\xi_1, \ldots, \xi_N\} \subset \Omega$. We define the finite dimensional spaces of approximants

$$\begin{aligned}H_N &:= \text{span}\{\mathfrak{K}_{\xi_i}(\cdot) := \mathfrak{K}(\cdot,\xi_i) \mid \xi_i \in \Xi_N\} \subset H(\Omega), \\ L_N &:= \text{span}\{\ell_{\xi_i}(\cdot) := \ell(\cdot,\xi_i) \mid \xi_i \in \Xi_N\} \subset L^2(\Omega), \\ W_N &:= \text{span}\{w_{\xi_i}(\cdot) \mid \xi_i \in \Xi_N\} \subset W(\Omega) := R(A^*).\end{aligned}$$

We define $\ell(x,y) := \ell^*(y,x)$, where $l^*(y,x)$ is defined in Theorem 1. From this definition, we have $\ell_{\xi_i}(x) := (A\mathfrak{K}_{\xi_i})(x)$. So these bases satisfy the relations $\ell_{\xi_i} = A\mathfrak{K}_{\xi_i}$, and $w_{\xi_i} = A^*\ell_{\xi_i} = A^*A\mathfrak{K}_{\xi_i}$. We denote by $\Pi_N : H(\Omega) \to H_N$ the projection of $H(\Omega)$ onto H_N. We define the Galerkin approximation $v_N \in H_N$ of the solution $v \in H(\Omega)$ of (10) to be given by $v_N := (\Pi_N \mathcal{A}|_{H_N})^{-1}\Pi_N y := G_N y$. This is equivalent to the variational equations

$$\big(\mathcal{A}v_N - y, \mathfrak{K}_{\xi_i}\big)_{H(\Omega)} = 0 \quad \text{or,}$$
$$\big(A^*Av_N - A^*b, \mathfrak{K}_{\xi_i}\big)_{H(\Omega)} = 0 \quad \text{for } 1 \le i \le N.$$

It is also worth noting that the Galerkin solution v_N above coincides with the Galerkin approximation of $Av = b$ in

$$\big(Av_N - b, \ell_{\xi_i}\big)_{L^2(\Omega)} = 0 \quad \text{for } 1 \le i \le N.$$

C. Coordinate Realizations

The study of the rates of convergence of the above approximations utilize coordinate representations of the operators. We need representations of the operator $A^*A{:}H(\Omega) \to H(\Omega)$. For A^*A we have

$$\begin{aligned}(A^*Av)(y) &:= \int_\Omega \ell^*(y,x)(\ell^*(\cdot,x), v)_{H(\Omega)}dx.\\ &= \int_\Omega \Big(\nabla_x \mathfrak{K}(x,y)^{\mathsf{T}}\psi(x)\psi(x)^{\mathsf{T}}\nabla_x \mathfrak{K}(x,\cdot), v\Big)_H dx,\end{aligned}$$

where $\psi(x) := f(x) + g(x)\mu(x)$.

The representation of the operators A^*A can now be used to determine the coordinate representations of the Galerkin approximations above. Define the matrix

$$\Phi(x, \Xi_N) := \begin{bmatrix} \frac{\partial \mathfrak{K}(x,\xi_1)}{\partial x_1} & \dots & \frac{\partial \mathfrak{K}(x,\xi_N)}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial \mathfrak{K}(x,\xi_1)}{\partial x_d} & \dots & \frac{\partial \mathfrak{K}(x,\xi_N)}{\partial x_d} \end{bmatrix} \in \mathbb{R}^{d\times N}.$$

Then for any two functions $v_N, w_N \in H_N$ with $v_N := \sum_{j=1}^N \alpha_j \mathfrak{K}_{\xi_j}$ and $w_N := \sum_{k=1}^N \beta_k \mathfrak{K}_{\xi_k}$, we have

$$\begin{aligned}&\big(A^*Av_N, w_N\big)_{H(\Omega)}\\ &\quad = \beta^{\mathsf{T}} \underbrace{\left(\int_\Omega \Phi(x,\Xi_N)^{\mathsf{T}}\psi(x)\psi(x)^{\mathsf{T}}\Phi(x,\Xi_N)dx\right)}_{[\int \ell^*(x,\xi_i)\ell(x,\xi_j)dx]}\alpha \quad (11)\end{aligned}$$

with $\alpha := [\alpha_1, \dots, \alpha_N]^{\mathsf{T}} \in \mathbb{R}^N$, $\beta := [\beta_1, \dots, \beta_N]^{\mathsf{T}} \in \mathbb{R}^N$.

D. Offline Rates of Convergence

Theorem 2: Let the hypothesis of Theorem 1 hold, and suppose that the unknown value function v satisfies the regularity condition $v = \mathcal{K}q$ for some fixed $q \in L^2(\Omega)$ where $\mathcal{K} : L^2(\Omega) \to H$ is the integral operator $v(x) = \int_\Omega \mathfrak{K}(x,\eta)q(\eta)d\eta$, and that the choice of centers Ξ_N ensures that an ideal "offline" persistence of excitation (PE) condition holds for the offline Galerkin approximations above. That is, there is a constant $\beta(N) > 0$ such that

$$\beta(N)I_N \le \int_\Omega \Phi(x,\Xi_N)^{\mathsf{T}}\psi(x)\psi(x)^{\mathsf{T}}\Phi(x,\Xi_N)dx$$

where I_N is the identity matrix on $\mathbb{R}^N$. Then the solution v_N of the Galerkin equations exists and is unique for all $N \in \mathbb{N}$. If the Galerkin method is convergent, then there is a constant $C > 0$ such that the solution v_N satisfies the error estimate

$$\begin{aligned}\|v - v_N\|_{H(\Omega)} &\le C \sup_{\xi\in\Omega} \mathcal{P}_{H,N}(\xi)\|v\|_{H(\Omega)}\\ &= C \sup_{\xi\in\Omega} \sqrt{\mathfrak{K}(\xi,\xi) - \mathfrak{K}_N(\xi,\xi)}\|\mathcal{K}^{-1}v\|_{L^2(\Omega)}.\end{aligned}$$

Proof: When we write $v_N := \sum_{j=1}^N \alpha_i \mathfrak{K}_{\xi_i}$, the Galerkin approximations give rise to the matrix equations

$$\begin{aligned}&\left[\int_\Omega \Phi(x,\Xi_N)^{\mathsf{T}}\psi(x)\psi(x)^{\mathsf{T}}\Phi(x,\Xi_N)dx\right]\alpha = \begin{Bmatrix}(A^*b)(\xi_1)\\ \vdots \\ (A^*b)(\xi_N)\end{Bmatrix}\\ &\quad = \int_\Omega \Phi(x,\Xi_N)^{\mathsf{T}}\psi(x)\ b dx \quad (12)\end{aligned}$$

with $\alpha = [\alpha_1, \dots, \alpha_N]^{\mathsf{T}} \in \mathbb{R}^N$. The representation in (11) makes clear that the offline PE condition ensures that the coefficient matrix is invertible. Also, the operator $G_N\mathcal{A} := (\Pi_N\mathcal{A}|_{H_N})^{-1}\Pi_N\mathcal{A}$ is a projection onto H_N since for any $p_N \in H_N$, we have

$$G_N\mathcal{A}p_N = (\Pi_N\mathcal{A}|_{H_N})^{-1}\Pi_N\mathcal{A}|_{H_N}p_N = p_N.$$

From the triangle inequality we have the pointwise bound

$$\begin{aligned}\|v - v_N\|_H &\le \|v - G_N\mathcal{A}p_N\|_H + \|G_N\mathcal{A}p_N - G_N\mathcal{A}v\|_H\\ &\le \|v - p_N\|_H + \|G_N\mathcal{A}(v - p_N)\|_H\\ &\le \Big(1 + \tilde{C}\Big)\|v - p_N\|_H\end{aligned}$$

for any $p_N \in H_N$. In this inequality we have used the fact that in a convergent Galerkin scheme the matrix $G_N\mathcal{A}$ is uniformly bounded in N: there is a constant $\tilde{C} > 0$ such that $\|G_N\mathcal{A}\| \le \tilde{C}$ for all $N > 0$ [18]. We choose $p_N := \Pi_N v$. The theorem now follows from the characterizations of projection/interpolation errors in terms of the power function in a native space discussed in Section II-B2. We have

$$\begin{aligned}\|v - v_N\|_H &\le \Big(1 + \tilde{C}\Big)\|(I - \Pi_N)v\|_H\\ &\le \Big(1 + \tilde{C}\Big)\|\mathcal{P}_{H,N}\|_{L^2(\Omega)}\|q\|_{L^2(\Omega)}.\end{aligned}$$

The last line stems from the proof of Theorem 11.23 in [14, Sec. 11.5]. Alternatively, we have

$$\|v - v_N\|_H \le \Big(1 + \tilde{C}\Big)\sqrt{|\Omega|}\sup_{\xi\in\Omega}\mathcal{P}_{H,N}(\xi)\|q\|_{L^2(\Omega)}$$

for all $v = \mathcal{K}q$ with $q \in L^2(\Omega)$. ■

Observations: We make several observations about how the result above compares to existing results.

(1) We say that the offline PE condition in Theorem 2 is ideal since it involves the integration over Ω that cannot usually be carried out in closed form.

(2) It is important to allow that the constant $\beta(N)$ in the offline PE condition above depends on the dimension N. This form of the PE condition could alternatively be written as

$$\hat{\beta}(N)\|v\|^2_{H(\Omega)} \le \big(A^*Av, v\big)_{H(\Omega)} \quad \text{for all } v \in H_N$$

for another constant $\hat{\beta}(N)$ that depends on N. But we know that $A^*A : H(\Omega) \to H(\Omega)$ is compact from Theorem 1 above. If the PE condition above holds for a constant $\hat{\beta}$ that does not depend on N, we could conclude that $(A^*A)^{-1}$ is a bounded linear operator. But since A^*A is compact, this is only true when $H(\Omega)$ is finite dimensional.

(3) The right hand side in the above error bound is explicit since we know $\mathfrak{K}_N$ as given in (8).

(4) Using normalized regressors is popular practice, as summarized in [3], [10]. This is useful when regressors may be

unbounded, such as when using polynomial regressors [6], [7]. For the sake of obtaining simple analysis and error bounds, we do not use the normalized form. Here, regressors are always bounded when $H(\Omega)$ is defined in terms of a kernel $\mathfrak{K}(\cdot,\cdot)$ that is bounded on the diagonal. We also assume that the controller μ that is implicit in the operator equation $Av = b$ generates a trajectory that lies in the compact set Ω. Again, this choice is made for illustrating strong error bounds in the simplest possible form. For some standard kernel spaces, the error bounds in Theorem 2 can alternatively be bounded from above in terms of the fill distance $h_{\Xi_N,\Omega}$ of centers Ξ_N in Ω, which is defined in Section I-A.

Corollary 1: Let the hypothesis in Theorem 2 hold and further suppose that the kernel $\mathfrak{K}$ that defines H is given as in [14, Table XI.1] or [19, Table I]. Then if the domain Ω is sufficiently smooth, we have

$$\|v - v_N\|_H \leq O\left(\sqrt{\mathcal{F}(h_{\Xi_N})}\right)$$

for a known function $\mathcal{F}$ defined in [14, Table II.1] or [19, Table I].

Proof: The proof can be found in [16]. ■

For instance, for the Sobolev-Matérn kernels of smoothness $r > 0$, as used in the numerical examples, we have

$$\|v - v_N\|_{L^\infty(\Omega)} \leq \|v - v_N\|_H \leq O\left(h_{\Xi_N}^{\nu - d/2}\right), \qquad (13)$$

where ν is a smoothness parameter and d is the dimension of the space in which Ω is contained. Thus, the approximation error converges at a rate that is bounded above by the fill distance raised to the smoothness parameter. The following theorem links the value approximation error with the controller approximation error.

Theorem 3: Under the same assumptions in Theorem 2 and Corollary 1, for the next estimate $\mu_{i+1,N}$ of the control policy iteration μ_{i+1} in equation (6), we have

$$\|\mu_{i+1,N} - \mu_{i+1}\|_{C(\Omega)} \leq \gamma \|v_{i,N} - v_i\|_H \leq O\left(\sqrt{\mathcal{F}(h_{\Xi_N})}\right), \quad (14)$$

where γ is a constant that depends on the kernel choice and the set of centers.

Proof: The proof can be found in [16]. ■

IV. Numerical Simulations

In this section, we consider the nonlinear system in [7]:

$$\dot{x} = f(x) + g(x)u, \quad x \in R^2$$

where

$$f(x) = \begin{bmatrix} -x_1 + x_2 \\ -0.5x_1 - 0.5x_2\left(1 - (\cos(2x_1) + 2)^2\right) \end{bmatrix}$$

$$g(x) = \begin{bmatrix} 0 \\ \cos(2x_1) + 2 \end{bmatrix}$$

Using the typical cost function J associated with the linear quadratic regulator problem, we choose $R = 1$ and $Q = I_2$, that is, the 2×2 identity matrix. With this cost function, the value function is $V^*(x) = 0.5x_1^2 + x_2^2$, and the optimal control policy is $u^*(x) = -(\cos(2x_1)+2)x_2$. The simulations presented in [7] use polynomial bases whose finite dimensional span contains the unknown value function. Here, we use the RKHS bases to illustrate the theoretical results of this letter. Certainly, the theoretical bounds extend to cases where the value function is not spanned by a finite number of polynomial bases functions. Using (12) and a quadrature approximation, we solve for the coefficients α.

Then, the value function is approximated using $v_N := \sum_{j=1}^{N} \alpha_i \mathfrak{K}_{\xi_i}$, with Gaussian and Matérn kernels as defined in [20]. The simulations utilized routines provided in [21] for kernel based computations. The ideal control law is assumed to be known and is employed in this approximation. Our primary focus is to assess the accuracy of the value function approximation in an offline manner and to validate expected

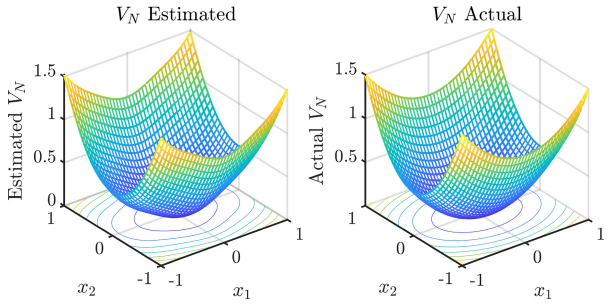

Fig. 1. The estimated and ideal value functions over the spatial domain.

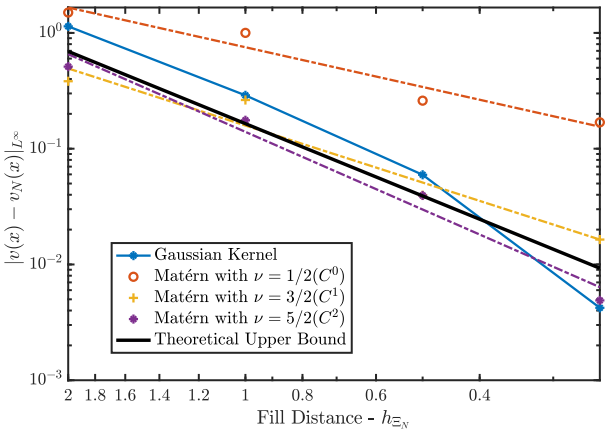

Fig. 2. Plot depicting the value function approximation error decay for Gaussian and Matérn kernels. The linear segments in the plot correspond to fitting a straight line to the logarithm of the data.

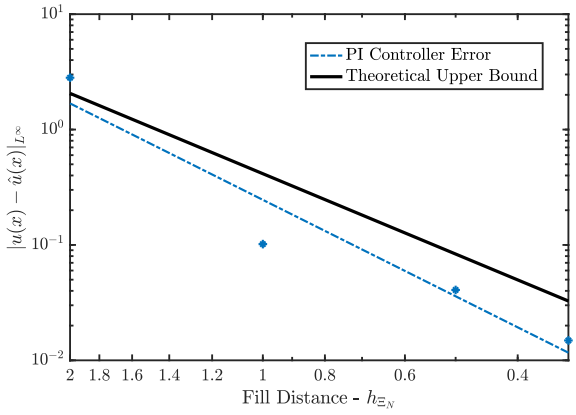

Fig. 3. In this plot, the PI algorithm was used to calculate the error between the estimated $\hat{u}(x)$ and optimal $u(x)$ for different fill distances.

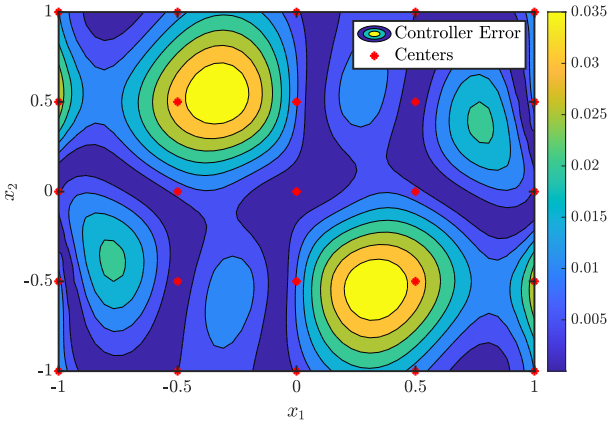

Fig. 4. A geometric representation of the error with centers. Errors are consistently smaller at the centers and larger away from the centers.

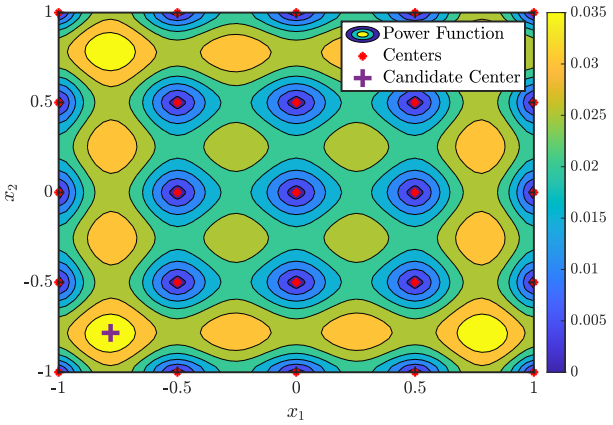

Fig. 5. Power function plot with the centers, and a depiction of where a candidate center, to be augmented, may be chosen based on the maximum of the power function over the domain.

convergence rates. As shown in Fig. 1, the approximated value function closely matches the optimal one. In Fig. 2, we see that as we decrease the fill distance (increase the number of centers), the approximation error decays as expected. Recall that the presented theoretical results apply to kernels of C^{2s} smoothness, which means it applies to the case with $\nu = 5/2$ only (refer to [20]). This is validated by the fact that the line corresponding to $\nu = 5/2$ in the figure is steeper than the theoretical upper bound described in (13).

Now, we begin with a stabilizing controller $\mu(x)$ and apply PI to approximate the optimal controller. Matérn kernel with $\nu = 5/2$ is used in these simulations. Furthermore, Fig. 3 shows the error between the ideal controller and the estimated controller is displayed for different fill distances. Again, the rate of error decay respects the limit predicted by (14). Fig. 4 is a geometric representation of the controller error plotted alongside the distribution of the centers. It is noteworthy that the error is generally smallest at the centers, and largest away from them. Based on the results of Theorem (3), one method to increase the number of bases adaptively is to position the next center at a location where the power function is largest. In Fig. 5, the power function and a candidate new basis are plotted with the centers.

V. Conclusion

In conclusion, this letter studies convergence rates for value function approximations that arise in a collection of RKHS. These rates can help in scenarios such as determining the placement of basis functions to achieve a required accuracy. These rates can also serve as the foundation for studies on rates of convergence for online actor-critic and RL methods. Future directions include developing bases adaption techniques based on the presented error estimates.

References

[1] D. Bertsekas, *Dynamic Programming and Optimal Control: Volume I*, vol. 1, Belmont, CA, USA: Athena Sci., 2012.
[2] D. Bertsekas, "Nonlinear programming," *J. Oper. Res. Soc.*, vol. 48, no. 3, pp. 334–334, 1997.
[3] F. L. Lewis and D. Liu, *Reinforcement Learning and Approximate Dynamic Programming for Feedback Control*. Hoboken, NJ, USA: Wiley, 2013.
[4] R. W. Beard and T. W. Mclain, "Successive Galerkin approximation algorithms for nonlinear optimal and robust control," *Int. J. Control*, vol. 71, no. 5, pp. 717–743, 1998.
[5] R. W. Beard, G. N. Saridis, and J. T. Wen, "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation," *Automatica*, vol. 33, no. 12, pp. 2159–2177, 1997.
[6] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.
[7] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.
[8] K. G. Vamvoudakis, Y. Wan, F. L. Lewis, and D. Cansever, *Handbook of Reinforcement Learning and Control*. Cham, Switzerland: Springer, 2021.
[9] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and autonomous control using reinforcement learning: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2042–2062, Jun. 2018.
[10] R. Kamalapurkar, P. Walters, J. Rosenfeld, and W. Dixon, *Reinforcement Learning for Optimal Feedback Control*. Cham, Switzerland: Springer, 2018.
[11] B. Kerimkulov, D. Siska, and L. Szpruch, "Exponential convergence and stability of Howard's policy improvement algorithm for controlled diffusions," *SIAM J. Control Optim.*, vol. 58, no. 3, pp. 1314–1340, 2020.
[12] F. Camilli and Q. Tang, "Rates of convergence for the policy iteration method for mean field games systems," *J. Math. Anal. Appl.*, vol. 512, no. 1, 2022, Art. no. 126138.
[13] M. L. Puterman, "On the convergence of policy iteration for controlled diffusions," *J. Optim. Theory Appl.*, vol. 33, pp. 137–144, Jan. 1981.
[14] H. Wendland, *Scattered Data Approximation*, vol. 17, Cambridge, U.K.: Cambridge Univ. Press, 2004.
[15] D.-X. Zhou, "Derivative reproducing properties for kernel methods in learning theory," *J. Comput. Appl. Math.*, vol. 220, nos. 1–2, pp. 456–463, 2008.
[16] A. Bouland, S. Niu, S. T. Paruchuri, A. Kurdila, J. Burns, and E. Schuster, "Rates of convergence in certain native spaces of approximations used in reinforcement learning," 2023, *arXiv:2309.07383*.
[17] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. vol. 375. Dordrecht, The Netherlands: Springer, 1996.
[18] R. Kress, V. Maz'ya, and V. Kozlov, *Linear Integral Equations*, vol. 82. New York, NY, USA: Springer, 1989.
[19] R. Schaback, "Error estimates and condition numbers for radial basis function interpolation," *Adv. Comput. Math.*, vol. 3, pp. 251–264, Apr. 1995.
[20] C. K. I. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2, Cambridge, MA, USA: MIT Press, 2006,
[21] R. Schaback, *MATLAB Programming for Kernel Based Methods*. MathWorks Comput. Softw. Corp., Natick, MA, USA, 2011.