2024 American Control Conference (ACC)
July 8-12, 2024. Toronto, Canada

Distributed Estimation by Two Agents with Different Feature Spaces

Aneesh Raghavan and Karl Henrik Johansson

Abstract—**We consider the problem of estimation of a function by a system consisting of two agents and a fusion center. The two agents collect data comprising of samples of an independent variable and the corresponding value of a dependent variable. The objective of the system is to collaboratively estimate the function without any exchange of data among the members of the system. To this end, we propose the following framework. The agents are given a set of features using which they construct suitable function spaces to formulate and solve the estimation problems locally. The estimated functions are uploaded to a fusion space where an optimization problem is solved to fuse the estimates (also known as meta-learning) to obtain the system estimate of the mapping. The fused function is then downloaded by the agents to gather knowledge about the other agents estimate of the function. With respect to the framework, we present the following: a systematic construction of fusion space given the features of the agents; the derivation of an uploading operator for the agents to upload their estimated functions to a fusion space; the derivation of a downloading operator for the fused function to be downloaded. Through an example on least squares regression, we illustrate the distributed estimation architecture that has been developed.**

I. Introduction

A. Motivation

In statistical inference problems, observations about a given phenomenon can be obtained using different kind or multiple sensors. The observations are often collated to obtain a single data set. Each sensor output is considered as a *modality* associated with the data set. When the data set is used to estimate a mapping from a set of independent variables to a set of dependent variables, the resulting algorithms are known as *multimodal learning algorithms*. [1] provides an overview on to multimodal data fusion focusing on why it is needed and how it can be achieved. [2] is a survey paper on deep multimodal learning covering many aspects including comparison with conventional multimodal learning, fusion structures, and applications. Multi-modal learning has found applications in many areas including human activity recognition [3] and autonomous driving, [4] [5].

Kernel methods have played a central role in inference problems. For classical literature on the application of kernel methods to estimation, we refer to [6], [7], and, the references there in. In [8], [9], and, [10], identification of discrete time and continuous time systems using these methods have been reviewed and studied. In more recent times, multimodal kernel learning methods have been studied and algorithms referred to as "Multiple Kernel Learning" [11], have been developed and applied to problems in object recognition [12], disease detection [13], etc. Kernel methods for deep learning has been studied in [14]. Understanding deep learning through comparison with kernel based learning has been done in [15]. Hence, kernel methods which have been extensively used in classical inference problems, have evolved, and are relevant in contemporary inference problems as well.

Multimodal learning problems are usually studied through data collation, i.e., it is a centralized learning approach. The objective of this paper is to take a step towards achieving multi-modal learning through a distributed approach. One such scheme that has been studied in the literature in the context of IoTs, etc., is vertical federated learning, [16], [17], when the number of agents is large. However, distributed schemes with fewer number of agents and emphasis on the learning space itself has not received much attention. We note that in our previous work, [18], we considered a distributed regression problem with *noisy* data by two agents and a fusion center with the agents learning in the same space. For further discussion on motivation for the problem considered we refer to [19].

B. Problem Considered

Though we do not formally define data, information, knowledge in the context of inference problems, we differentiate between them as following. A set of measurable outcomes associated with an observable phenomenon or an experiment is referred to as *data*. Structured data, that is, data which could used to infer models or hidden patterns is referred to as *information*. The inferred models or patterns are referred to as knowledge. In the context of the experiment, the set of all possible models or patterns is referred to as the *knowledge space* (KS). As information is received sequentially, the knowledge about the observed phenomenon evolves in the KS. In the simplest setting, we can consider the estimation of a mapping from an independent variable (input) to a dependent variable (output). Information would correspond to pairs of input, output measurements. Knowledge corresponds to the function from the input to the output, while the function space where the inference problem is studied is interpreted as a knowledge space. Other examples of knowledge spaces include set of probability measures on a Borel σ algebra, probability measures on a orthoposet [20].

*Research supported by the Swedish Research Council (VR), Swedish Foundation for Strategic Research (SSF), and the Knut and Alice Wallenberg Foundation. The authors are with the Division of Decision and Control Systems, Royal Institute of Technology, KTH, Stockholm. Email: aneesh@kth.se, kallej@kth.se

979-8-3503-8265-5/$31.00 ©2024 AACC

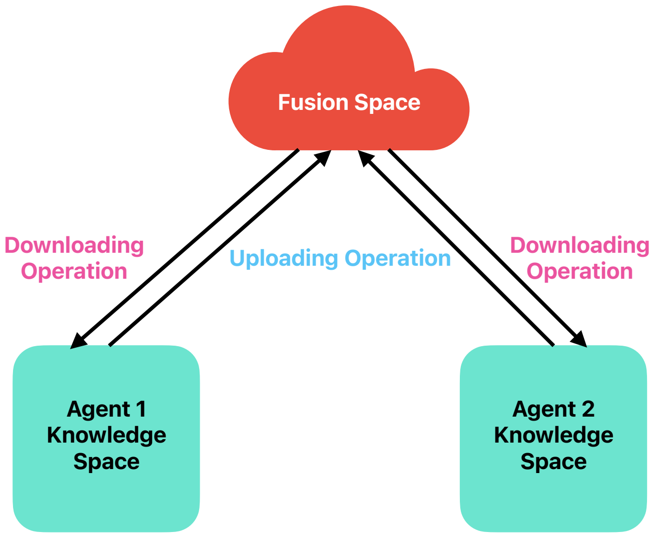

Fig. 1. Schematic for Distributed Estimation Architecture

The problem considered is as follows. We consider two agents and a fusion center. Each agent receives information comprising of samples of an independent variable and corresponding values of the dependent variable. The underlying phenomenon generating information for both the agents is the same. Given the information, the objective of the agents and the fusion center is to collaboratively estimate the mapping from the independent variable to the dependent variable *without* exchanging any information between the agents or the agents and the fusion center. Along with the samples, each agent is provided a set of features predominant in the information received by them.

C. Contributions

We propose the following architecture, refer Figure 1: Given local information, each agent estimates the mapping from the input to the output by solving a least squares regression problem in its local KS. The functions estimated by both agents are uploaded to the fusion center. At the fusion center, a fusion problem as an optimization problem is formulated and solved to fuse the functions estimated by the agents. The fused function is considered as the function estimated by the system. The fused function is downloaded on to the local KSs by the agents and is considered as the final estimates of the agents.

With respect to the above frame work we prove the following: Given the features, we present the construction of the individual KSs of the agents as a reproducing kernel Hilbert space (RKHS). The fusion space is defined as the set of functions obtained through linear combination of functions in the the local KSs. We prove that the fusion space is also an RKHS whose kernel is the sum of the kernels of the agents. As a corollary, we obtain that the uploading operator used by the agents to upload functions from the local KS to the fusion space is a linear bounded operator. All functions in the fusion center might not be decipherable in the local KSs, a download operator is needed to suitably transform the fused function which can be interpreted in the local KS. We present a detailed construction of such an download operator and prove that it is linear and bounded. To illustrate the distributed estimation scheme, we present a numerical example. For further discussion novelty of the proposed solution and interpretation of the proposed architecture we refer to [19].

The paper is organized as follows. In Section II, we discuss the construction of the individual KSs, the fusion space, the derivation of the uploading and downloading operator. In Section III, we discuss the regression problem for the agents, its solution, and, the fusion problem in fusion space. In Section IV, we present a numerical example demonstrating the distributed estimation methodology. In Section V, we summarize the contributions of this paper and discuss future work. Notation: we use superscript for the agent, subscript for samples and summation indicies. We represent vectors obtained by concatenating smaller vectors in boldface. For a function $f \in V$, V vector space, we use the notation f when it is treated as a vector and the notation $f(\cdot)$ when it is treated as a function. The projection onto a subspace $\mathcal{M}$ of a Hilbert space H is denoted by $\Pi_{\mathcal{M}}$.

II. Construction of Knowledge Spaces

In this section, we discuss the construction of KSs for the individual agents and the fusion space. Let $\mathcal{X} \subset \mathbb{R}^d$. Let $(H, \langle\cdot,\cdot\rangle_H)$ be a Hilbert space of functions, $f : \mathcal{X} \to \mathbb{R}$. Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a function, $n \in \mathbb{N}$, $\{x_1, \ldots, x_n\} \subset \mathcal{X}$, and, $\mathbf{K} := (K(x_i, x_j))_{ij}$ be Gram (kernel) matrix of K with respect to $x_1, \ldots, x_n$.

Definition 1. The function $K(\cdot,\cdot)$ is said to be a positive definite kernel if the gram matrix generated by the function is a positive definite matrix for all n and $\{x_1, \ldots, x_n\} \subset \mathcal{X}$.

Definition 2. $(H, \langle\cdot,\cdot\rangle_H)$ is said to be a reproducing kernel Hilbert space (RKHS) with a positive definite kernel K, if,

- $K(\cdot, x) \in H,\ \forall x \in \mathcal{X}$,
- the reproducing property is satisfied

$$f(y) = \langle f(\cdot), K(\cdot, y)\rangle_H, f \in H, y \in \mathcal{X}.$$

A. Construction of Individual Knowledge Spaces for the Agents

The set of features for agent i is the set of functions, $\{\varphi_j^i(\cdot)\}_{j\in\mathcal{I}^i}$, where $\varphi_j^i : \mathcal{X} \to \mathbb{R}$ and $|\mathcal{I}^i| < \infty$. The knowledge space for agent i is the finite dimensional vector space, H^i, defined as:

$$H^i = \{f : f(\cdot) = \sum_{j\in\mathcal{I}^i} \alpha_j \varphi_j^i(\cdot), \{\alpha_j\}_{j\in\mathcal{I}^i} \subset \mathbb{R}\}.$$

The null vector for the space H^i is the function $\theta^i(\cdot)$ defined as $\theta^i(x) = 0, \forall x \in \mathcal{X}$. For agent i, the features are assumed to be linearly independent, i.e., $\sum_{j\in\mathcal{I}^i} \alpha_j \varphi_j^i = \theta^i$ if and only if $\alpha_j = 0,\ \forall j$. The function space, H^i, is equipped with the inner product, $\langle\cdot,\cdot\rangle_{H^i} : H^i \times H^i \to \mathbb{R}$, defined as follows. For, $f(\cdot) = \sum_{j\in\mathcal{I}^i} \alpha_j \varphi_j^i(\cdot), g(\cdot) = \sum_{j\in\mathcal{I}^i} \beta_j \varphi_j^i(\cdot)$,

$$\langle f(\cdot), g(\cdot)\rangle_{H^i} = \langle \sum_{j\in\mathcal{I}^i} \alpha_j \varphi_j^i(\cdot), \sum_{j\in\mathcal{I}^i} \beta_j \varphi_j^i(\cdot)\rangle_{H^i} := \sum_{j\in\mathcal{I}^i} \alpha_j \beta_j.$$

It can be verified that the above definition of inner product on H^i satisfies the axioms of a inner product on a vector space. The norm induced by the inner product is $||f||^2_{H^i} = \sum_{j\in\mathcal{I}^i} \alpha_j^2$. The kernel $K^i : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ is defined as,

$$K^i(x,y) = \sum_{j\in\mathcal{I}^i} \varphi^i_j(x)\varphi^i_j(y).$$

Let $\{x_1,\dots,x_n\} \subset \mathcal{X}$. Let $\mathbf{K^i} := (K^i(x_k,x_l))_{kl}$, be the Gram matrix of $K^i(\cdot,\cdot)$ with respect to $\{x_1,\dots,x_n\}$. Then for any $\boldsymbol{\alpha}\in\mathbb{R}^n$, $\boldsymbol{\alpha}^T\mathbf{K^i}\boldsymbol{\alpha} =$

$$\sum_{k=1}^n\sum_{l=1}^n \alpha_k\alpha_l K^i(x_k,x_l) = \sum_{k=1}^n\sum_{l=1}^n \alpha_k\alpha_l \sum_{j\in\mathcal{I}^i}\varphi^i_j(x_k)\varphi^i_j(x_l)$$
$$= \sum_{j\in\mathcal{I}^i}\Big(\sum_{k=1}^n \alpha_k\varphi^i_j(x_k)\Big)\Big(\sum_{l=1}^n \alpha_l\varphi^i_j(x_l)\Big) = ||f||^2_{H^i} \geq 0,$$

where $f(\cdot) = \sum_{j\in\mathcal{I}^i}\Big(\sum_{k=1}^n \alpha_k\varphi^i_j(x_k)\Big)\varphi^i_j(\cdot)$. The Gram matrix of $K^i(\cdot,\cdot)$ is positive definite for any $\{x_1,\dots,x_n\}\subset\mathcal{X}$, for all $n\in\mathbb{N}$. Thus, $K^i(\cdot,\cdot)$ is a positive definite kernel (From Definition 1, also refer [21]). We note that, $K^i(\cdot,y) = \sum_{j\in\mathcal{I}^i}\varphi^i_j(y)\varphi^i_j(\cdot) \in H^i$, with $\alpha^i_j = \varphi^i_j(y),\ \forall j\in\mathcal{I}^i$. For $f\in H^i$, $f(\cdot) = \sum_{j\in\mathcal{I}^i}\alpha_j\varphi^i_j(\cdot)$,

$$\begin{aligned}\langle f(\cdot), K^i(\cdot,y)\rangle_{H^i} &= \langle \sum_{j\in\mathcal{I}^i}\alpha_j\varphi^i_j(\cdot), \sum_{j\in\mathcal{I}^i}\varphi^i_j(y)\varphi^i_j(\cdot)\rangle_{H^i}\\ &= \sum_{j\in\mathcal{I}^i}\alpha_j\varphi^i_j(y) = f(y).\end{aligned}$$

The reproducing property is satisfied by $f\in H^i$ with kernel $K^i(\cdot,\cdot)$. From Definition 2, it follows that H^i is a RKHS with kernel $K^i(\cdot,\cdot)$.

B. Construction of the Fusion Space

The motivation for the construction of a fusion space is to build a function space where algebraic operations can be simultaneously performed on functions living in the local KSs of the agents. For further discussion on possible candidates of the fusion space and the reasoning for the following construction, we refer to [19].

Definition 3. The fusion space, H, is defined as $H = \{f : f = f^1 + f^2, f^1\in H^1, f^2\in H^2\}$.

In the following theorem, we characterize the fusion space as a RKHS after associating a suitable inner product and find an expression for the norm induced by the inner product.

Theorem II.1. *If $K^i(\cdot,\cdot)$ is the reproducing kernel of Hilbert space H^i, with norm $||\cdot||_{H^i}$, then $K(x,y) = K^1(x,y) + K^2(x,y)$ is the reproducing kernel of the space $H = \{f : f = f^1 + f^2 | f^i \in H^i\}$ with the norm:*

$$||f||^2_H = \min_{\substack{f^1+f^2=f,\\ f^i\in H^i}} ||f^1||^2_{H^1} + ||f^2||^2_{H^2}.$$

Proof. Let $H_{\prod} = H^1\times H^2$ denote the product space, with inner product $\langle (f^1,f^2),(g^1,g^2)\rangle_{H_{\prod}} = \langle f^1,g^1\rangle_{H^1} + \langle f^2,g^2\rangle_{H^2}$. Let $H = \{f : f = f^1+f^2, f^i\in H^i, i=1,2\}$. Clearly, H is a vector space whose null vector we denote by θ. Let $L : H_\Pi \to H$ be a operator defined as $L((f^1,f^2)) = f^1+f^2$. L is a linear operator and its null space, $\mathcal{N}(L) = \{(f^1,f^2)\in H_{\prod} : f^1+f^2 = \theta\}$ is a closed subspace as it is finite dimensional. The basis vectors for $H_{\prod}$ are given by $\{\varphi^1_j\times\theta^2\}_{j\in\mathcal{I}^1}\cup\{\theta^1\times\varphi^2_j\}_{j\in\mathcal{I}^2}$. Any vector in $H_{\prod}$, can be expressed as $(f^1,f^2) = (\sum_{j\in\mathcal{I}^1}\alpha^1_j\varphi^1_j, \sum_{j\in\mathcal{I}^2}\alpha^2_j\varphi^2_j)$ and thus $H_{\prod}$ is isomorphic to $\mathbb{R}^{\mathcal{I}^1+\mathcal{I}^2}$. The null space $\mathcal{N}(L)$ is isomorphic to the subspace,

$$\mathcal{N} = \Big\{\Big(\boldsymbol{\alpha^1},\boldsymbol{\alpha^2}\Big)\in\mathbb{R}^{\mathcal{I}^1+\mathcal{I}^2} : \sum_{j\in\mathcal{I}^1}\alpha^1_j\varphi^1_j + \sum_{j\in\mathcal{I}^2}\alpha^2_j\varphi^2_j = \theta,$$
$$\boldsymbol{\alpha^1} = \Big(\alpha^1_1,\dots\alpha^1_{\mathcal{I}^1}\Big), \boldsymbol{\alpha^2} = \Big(\alpha^2_1,\dots\alpha^2_{\mathcal{I}^2}\Big)\Big\}.$$

Since $\mathcal{N}(L)$ is a closed subspace, there exists a unique closed subspace $\mathcal{M}$ such that $H_{\prod} = \mathcal{M}\oplus\mathcal{N}(L)$. The mapping $L_{\mathcal{M}} = L\circ\Pi_{\mathcal{M}}$ (operator L restricted to subspace $\mathcal{M}$) is bijection from $\mathcal{M}$ to H. For any function $f\in H$, let $L^{-1}_{\mathcal{M}}(f) = (L^1(f), L^2(f))$, i.e., $(L^1(f),L^2(f))$ is the *unique* tuple of functions in $\mathcal{M}$ such that $L^1(f)\in H^1$, $L^2(f)\in H^2$ and $L^1(f)+L^2(f) = f$. We now define the inner product on H as follows:

$$\langle f,g\rangle_H = \langle L^1(f),L^1(g)\rangle_{H^1} + \langle L^2(f),L^2(g)\rangle_{H^2}.$$

Since $K^i(\cdot,y)\in H^i$, it follows that $K(\cdot,y) = K^1(\cdot,y) + K^2(\cdot,y)\in H$. We claim that $L^i(K(\cdot,y)) = K^i(\cdot,y)$. Indeed, since $K(\cdot,y) = \sum_{j\in\mathcal{I}^1}\varphi^1_j(y)\varphi^1_j(\cdot) + \sum_{j\in\mathcal{I}^2}\varphi^2_j(y)\varphi^2_j(\cdot)$, to prove the claim it suffices to prove that the vector $\Big(\varphi^1_1(y),\dots,\varphi^1_{\mathcal{I}^1}(y),\varphi^2_1(y),\dots,\varphi^2_{\mathcal{I}^2}(y)\Big)$ is orthogonal to $\mathcal{N}$ for all $y\in\mathcal{X}$. That is, $\forall\Big(\boldsymbol{\alpha^1},\boldsymbol{\alpha^2}\Big)\in\mathcal{N}$,

$$\sum_{j\in\mathcal{I}^1}\alpha^1_j\varphi^1_j(y) + \sum_{j\in\mathcal{I}^2}\alpha^2_j\varphi^2_j(y) = 0,$$

which is true from the definition of $\mathcal{N}$. Thus,

$$\begin{aligned}&\langle f(\cdot),K(\cdot,y)\rangle_H = \langle L^1(f)(\cdot),L^1(K(\cdot,y))\rangle_{H^1} + \langle L^2(f)(\cdot),\\ &L^2(K(\cdot,y))\rangle_{H^2} = \langle L^1(f)(\cdot),K^1(\cdot,y))\rangle_{H^1} + \langle L^2(f)(\cdot),\\ &K^2(\cdot,y))\rangle_{H^2} = L^1(f)(y) + L^2(f)(y) = f(y),\end{aligned}$$

satisfying the reproducing property. From Definition 2, it follows that $(H,\langle\cdot,\cdot\rangle_H)$ is a RKHS with kernel $K(\cdot,\cdot)$. For $f\in H$, let $(f^1,f^2)\in H_{\prod}$ be such that $f = f^1+f^2 = L^1(f)+L^2(f)$. We note that $(L^1(f),L^2(f)) = \Pi_{\mathcal{M}}((f^1,f^2))$. Computing norm of f,

$$\begin{aligned}&||f||^2_H = \langle f,f\rangle_H = ||L^1(f)||^2_{H^1} + ||L^2(f)||^2_{H^2}\\ &||f^1||^2_{H^1} + ||f^2||^2_{H^2} = ||(f^1,f^2)||^2_{H_\Pi} = \underbrace{||\Pi_{\mathcal{M}}((f^1,f^2))}+\\ &\Pi_{\mathcal{N}(L)}((f^1,f^2))||^2_{H_\Pi}. \qquad = ||(L^1(f),L^2(f))||^2_{H_\Pi}\end{aligned}$$

Thus, for f^1,f^2 such that $f = f^1+f^2$, the minimum of $||f^1||^2_{H^1} + ||f^2||^2_{H^2}$ is achieved when $\Pi_{\mathcal{N}(L)}((f^1,f^2))||^2_{H_\Pi} = \theta$, i.e., $f^i = L^i(f)$, and, is equal to $||f||^2_H$. ☐

Given $f\in H^1$, we let $f^1 = f$ and $f^2 = 0$. From the Theorem II.1, we conclude that $||f||_H \leq ||f||_{H^1}$. Similarly $||f||_H \leq ||f||_{H^2}, f\in H^2$.

Corollary II.2. *The uploading operator from agent i's knowledge space, H^i, to the fusion space H, $\hat{L}^i : H^i \to H$, is $\hat{L}(f) = f$. $\hat{L}^i(\cdot)$, is linear and is bounded, $||\hat{L}^i|| = \sup\{||f||_H : f \in H^i, ||f||_{H^i} = 1\} \leq 1$.*

C. Retrieval of Individual Knowledge Spaces from Fusion Space

Given a function in the fusion space, it is not necessary that it belongs to both the KSs. Hence, it is mandatory to transform it to a form where it can be expressed using the features of the local KS. The objective of this subsection is to find an operation which would transform the function onto the individual KSs. For further discussion on the reasoning for the following construction, we refer to [19].

Lemma II.3. *For every $f \in H$, there exits $\{y_{k,f}\}_{k=1}^n$ and $\{\beta_{k,f}\}_{k=1}^n$ such that,*

$$f(\cdot)=\sum_{i=1,2}\sum_{j\in\mathcal{I}^i}\sum_{k=1}^{n}\beta_{k,f}\varphi_j^i(y_{k,f})\varphi_j^i(\cdot) = \sum_{k=1}^{n}\beta_{k,f}K(\cdot,y_{k,f}).$$

Proof. Given $f \in H$, we characterize $L^1(f)$ and $L^2(f)$. Let,

$$\Phi(y)=[\varphi_1^1(y),\ldots,\varphi_{\mathcal{I}^1}^1(y),\varphi_1^2(y),\ldots,\varphi_{\mathcal{I}^1}^2(y)]\in\mathbb{R}^{\mathcal{I}^1+\mathcal{I}^2},$$

$y \in \mathcal{X}$. Let $\hat{\mathcal{M}}$ be the span of $\{\Phi(y)\}_{y\in\mathcal{X}}$. We note that $\mathbb{R}^{\mathcal{I}^1+\mathcal{I}^2} = \hat{\mathcal{M}} \oplus \mathcal{N}$ [19]. This implies that $\hat{\mathcal{M}}$ is isomorphic to $\mathcal{M}$. Thus, every vector in $\mathcal{M}$ can be expressed as

$$\Big(\sum_{j\in\mathcal{I}^1}\gamma_j^1\big(\varphi_j^1\times\theta^2\big)+\sum_{j\in\mathcal{I}^2}\gamma_j^2\big(\theta^1\times\varphi_j^2\big)\Big)=\Big(\sum_{j\in\mathcal{I}^1}\gamma_j^1\varphi_j^1,$$
$$\sum_{j\in\mathcal{I}^2}\gamma_j^2\varphi_j^2\Big)=\Big(\sum_{j\in\mathcal{I}^1}\sum_{k=1}^{n}\beta_k\varphi_j^1(y_k)\varphi_j^1,\sum_{j\in\mathcal{I}^2}\sum_{k=1}^{n}\beta_k\varphi_j^2(y_k)\varphi_j^2\Big),$$

where $\boldsymbol{\gamma} = [\gamma_1^1,\ldots,\gamma_{\mathcal{I}^1}^1,\gamma_1^2,\ldots,\gamma_{\mathcal{I}^2}^2] \in \bar{\mathcal{M}}$ and $\gamma_j^i = \sum_{k=1}^n \beta_k\varphi_j^i(y_k)$. Thus, for any $f \in H$, $(L^1(f), L^2(f)) = \Big(\sum_{k=1}^n \beta_{k,f}K^1(\cdot,y_{k,f}), \sum_{k=1}^n \beta_{k,f}K^2(\cdot,y_{k,f})\Big)$. This implies that, $f = L^1(f) + L^2(f) = \sum_{k=1}^n \beta_{k,f}K(\cdot,y_{k,f})$. □

Lemma II.4. *Given the RKHS, $(H,\langle\cdot,\cdot\rangle_H)$, with kernel $K(\cdot,\cdot)$ and the kernels $K^i(\cdot,\cdot)$, $i = 1,2$, such that $K(x,y) = K^1(x,y) + K^2(x,y)$, we define operators, $\bar{L}^i : H \to H$, as*

$$\bar{L}^i(f)(x) = \langle f(\cdot), K^i(\cdot,x)\rangle_H,\ \textit{for}, i = 1,2.$$

Then, $\bar{L}^i$ is linear, symmetric, positive and bounded, $||\bar{L}^i|| \leq 1$.

Proof. From the linearity of the inner product with respect to the first argument, it follows that $\bar{L}^i$ is linear. From equation set (1), we note that $\bar{L}^i$ is positive, i.e., $\langle\bar{L}^i(f), f\rangle_H \geq 0,\ \forall f \in H$. From equation set(2), we note that $\bar{L}^i$ is symmetric, i.e., $\langle\bar{L}^i(f), g\rangle_H = \langle f, \bar{L}^i(g)\rangle_H,\ \forall f,g \in H$. Since $\bar{L}^1$ and $\bar{L}^2$ are symmetric, positive, and their sum is the identity operator ($\bar{L}^1 + \bar{L}^2 = \mathbb{I}$), $0 \leq ||\bar{L}^i|| \leq 1$. □

Theorem II.5. *Let $L : H \to H$ be a symmetric, positive, bounded operator. There exists a unique square root of operator L, $\sqrt{L} : H \to H$, i.e., $\sqrt{L}(\sqrt{L}(f)) = L(f)\forall f \in H$. $\sqrt{L}(\cdot)$ is linear, bounded, symmetric.*

Proof. Since L is symmetric, from the spectral theorem, it follows that (i) there exists an orthonormal basis of H, $\{\varphi_j\}_{j\in\mathcal{I}}$, which are eigenvectors of L ; (ii) the eigenvalues of L, $\{\lambda_j\}_{j=1}^{\mathcal{I}_\lambda}$, are real. Since L is positive, $\lambda_j \geq 0, 1 \leq j \leq \mathcal{I}_\lambda$. Let $f = \sum_{j=1}^{\mathcal{I}}\alpha_j\varphi_j, \{\alpha_j\} \subset \mathbb{R}$. $L(f) = L(\sum_{j=1}^{\mathcal{I}}\alpha_j\varphi_j) = \sum_{j=1}^{\mathcal{I}}\alpha_j\lambda_j\varphi_j$. From the construction of H, it follows that $\mathcal{I} \leq \mathcal{I}^1 + \mathcal{I}^2$. Since some of the eigenvalues of L could repeated, $\mathcal{I}_\lambda \leq \mathcal{I}$. The operator $\sqrt{L}$ is defined as $\sqrt{L}(\varphi_j) = \sqrt{\lambda_j}\varphi_j$, and, imposing linearity $\sqrt{L}(f) = \sum_{j=1}^{\mathcal{I}}\alpha_j\sqrt{\lambda_j}\varphi_j$.

$$\sqrt{L}(\sqrt{L}(f)) = \sqrt{L}(\sum_{j=1}^{\mathcal{I}}\alpha_j\sqrt{\lambda_j}\varphi_j) = \sum_{j=1}^{\mathcal{I}}\alpha_j\lambda_j\varphi_j = L(f)$$

$$\langle\sqrt{L}(f), f\rangle_H=\langle\sum_{j=1}^{\mathcal{I}}\alpha_j\sqrt{\lambda_j}\varphi_j,\sum_{l=1}^{\mathcal{I}}\alpha_l\varphi_l\rangle_H=\sum_{j=1}^{\mathcal{I}}\alpha_j^2\sqrt{\lambda_j} \geq 0$$

$$\langle\sqrt{L}(f), g\rangle_H = \langle\sum_{j=1}^{\mathcal{I}}\alpha_j\sqrt{\lambda_j}\varphi_j,\sum_{l=1}^{\mathcal{I}}\beta_l\varphi_l\rangle_H = \sum_{j=1}^{\mathcal{I}}\alpha_j\beta_j\sqrt{\lambda_j}$$
$$= \langle\sum_{j=1}^{\mathcal{I}}\alpha_j\varphi_j,\sum_{l=1}^{\mathcal{I}}\beta_l\sqrt{\lambda_l}\varphi_l\rangle_H = \langle f, \sqrt{L}(g)\rangle_H.$$

Suppose $\tilde{L}$ is a positive semi-definite operator which is another square root for L. To prove uniqueness, it suffices to prove that $\tilde{L}(\phi_j) = \sqrt{L}(\phi_j)\forall\ j$, where $\{\phi_j\}_{j\in\mathcal{I}}$ is a basis for H. By the spectral theorem, $\tilde{L}$ posses a set of orthonormal eigenvectors,$\{\phi_j\}_{j\in\mathcal{I}}$, which form a basis for H. $L(\phi_j) = \tilde{L}\big(\tilde{L}(\phi_j)\big) = \tilde{L}(\bar{\lambda}_j\phi_j) = \bar{\lambda}_j^2\phi_j$. Thus, $\{\phi_j\}_{j\in\mathcal{I}}$ are eigenvectors for L. By definition of $\sqrt{L}$, $\{\phi_j\}_{j\in\mathcal{I}}$ are eigenvectors for $\sqrt{L}$ as well. As the eigenvectors $\{\phi_j\}_{j\in\mathcal{I}}$ could be ordered differently than $\{\varphi_j\}_{j\in\mathcal{I}}$, we denote the corresponding eigenvalues by $\hat{\lambda}_j$. Thus, $\sqrt{L}\big(\sqrt{L}(\phi_j)\big) = \hat{\lambda}_j^2\phi_j = L(\phi_j) = \bar{\lambda}_j^2\phi_j$. Since the operators are positive semidefinite, $\hat{\lambda}_j = \bar{\lambda}_j$. Hence, $\tilde{L}(\phi_j) = \sqrt{L}(\phi_j),\forall j$, i.e. $\tilde{L} = \sqrt{L}$. □

The above theorem, Theorem II.5, is well known in linear algebra and we mention the proof as the construction of the square root operator is essential for the proof of the theorem below.

Theorem II.6. *The linear space $\bar{H}^i = \{g : g = \sqrt{\bar{L}^i}(f), f \in H\}$ is a RKHS with kernel K^i. $\sqrt{\bar{L}^i}(\cdot)$ establishes an isometric isomorphism between $\mathcal{N}\big(\sqrt{\bar{L}^i}\big)^\perp$ and $\bar{H}^i$, and the norm, $||f||_{\bar{H}^i} = ||g||_H$, where $f = \sqrt{\bar{L}^i}g, g \in \mathcal{N}\big(\sqrt{\bar{L}^i}\big)^\perp$. Thus, the individual knowledge spaces can be retrieved from the fusion space.*

Proof. Since $\mathcal{N}\big(\sqrt{\bar{L}^i}\big)$ is closed subspace of H, $H = \mathcal{M}^i \oplus \mathcal{N}\big(\sqrt{\bar{L}^i}\big)$, where $\mathcal{M}^i = \mathcal{N}\big(\sqrt{\bar{L}^i}\big)^\perp$. $\sqrt{\bar{L}^i}()$ maps one-one from $\mathcal{M}^i$ to $\bar{H}^i$. Hence, $\mathcal{M}^i \subset \bar{H}^i$. Let $f \in \mathcal{N}\big(\sqrt{\bar{L}^i}\big)$ and $g \in H$. Then, $\langle f, \sqrt{\bar{L}^i}(g)\rangle_H = \langle\sqrt{\bar{L}^i}(f), g\rangle_H = 0$, i.e., $\sqrt{\bar{L}^i}(g) \perp f,\ \forall f \in \mathcal{N}\big(\sqrt{\bar{L}^i}\big)$, $\forall g \in H$. Hence, $\bar{H}^i \subset \mathcal{N}\big(\sqrt{\bar{L}^i}\big)^\perp = \mathcal{M}^i$. Therefore, $M^i = \bar{H}^i$. The inner product on $\bar{H}^i$ is defined as,

$$\langle f, g\rangle_{\bar{H}^i} = \langle\sqrt{\bar{L}^i}(\bar{f}), \sqrt{\bar{L}^i}(\bar{g})\rangle_{\bar{H}^i} = \langle\Pi_{\mathcal{M}^i}(\bar{f}), \Pi_{\mathcal{M}^i}(\bar{g})\rangle_H,$$

$$f=\sum_{j=1}^{n}\beta_j K(\cdot,y_j), \bar{L}^i(f)(x)=\langle\sum_{j=1}^{n}\beta_j K(\cdot,y_j),K^i(\cdot,x)\rangle_H=\sum_{j=1}^{n}\beta_j\langle K^i(\cdot,x),K(\cdot,y_j)\rangle_H=\sum_{j=1}^{n}\beta_j K^i(y_j,x)$$
$$=\sum_{j=1}^{n}\beta_j K^i(x,y_j).\ \langle\bar{L}^i(f),f\rangle_H=\langle\sum_{j=1}^{n}\beta_j K^i(\cdot,y_j),\sum_{k=1}^{n}\beta_k K(\cdot,y_k)\rangle_H=\sum_{j=1}^{n}\sum_{k=1}^{n}\beta_j\beta_k K^i(y_k,y_j)=||\bar{L}^i(f)||^2_{H^i}\geq 0. \quad (1)$$
$$g=\sum_{l=1}^{m}\delta_l K(\cdot,y_l), \langle\bar{L}^i(f),g\rangle_H=\langle\sum_{j=1}^{n}\beta_j K^i(\cdot,y_j),\sum_{l=1}^{m}\delta_l K(\cdot,y_l)\rangle_H=\sum_{j=1}^{n}\sum_{l=1}^{m}\beta_j\delta_l K^i(y_l,y_j).\langle f,\bar{L}^i(g)\rangle_H=\langle\sum_{j=1}^{n}\beta_j K(\cdot,y_j),$$
$$\sum_{l=1}^{m}\delta_l K^i(\cdot,y_l)\rangle_H=\sum_{j=1}^{n}\sum_{l=1}^{m}\beta_j\delta_l\langle K^i(\cdot,y_l),K(\cdot,y_j)\rangle_H=\sum_{j=1}^{n}\sum_{l=1}^{m}\beta_j\delta_l K^i(y_j,y_l)=\sum_{j=1}^{n}\sum_{l=1}^{m}\beta_j\delta_l K^i(y_l,y_j)=\langle\bar{L}^i(f),g\rangle_H. \quad (2)$$

where $f=\sqrt{\bar{L}^i}(\bar{f}), \bar{f}\in H$ and $g=\sqrt{\bar{L}^i}(\bar{g}), \bar{g}\in H$. For $f(\cdot)=K(\cdot,y)$, $\bar{L}^i(f)(\cdot)=K^i(\cdot,y)$. Hence, $K^i(\cdot,y)\in\mathcal{R}(\bar{L}^i)\subseteq\bar{H}^i,\ \forall y\in\mathcal{X}$. Let $\{\psi^i_j\}$ be the eigenvectors of $\bar{L}^i$ with corresponding eigenvalues, $\{\lambda^i_j\}$. Then, from the construction of $\sqrt{\bar{L}^i}$ in Theorem II.5, $\{\psi^i_j\}$ are the eigenvectors of $\sqrt{\bar{L}^i}$ with corresponding eigenvalues, $\{\sqrt{\lambda^i_j}\}$. For any eigenvector, ψ^i_j, with $\lambda^i_j\neq 0$,

$$\langle\psi^i_j(\cdot),K^i(\cdot,y)\rangle_{\bar{H}^i}=$$
$$\langle\sqrt{\bar{L}^i}\Big(\frac{1}{\sqrt{\lambda^i_j}}\psi^i_j(\cdot)\Big),\sqrt{\bar{L}^i}\big(\sqrt{\bar{L}^i}(K(\cdot,y))\big)\rangle_{\bar{H}^i}$$
$$\overset{(a)}{=}\langle\frac{1}{\sqrt{\lambda^i_j}}\psi^i_j(\cdot),\Pi_{\mathcal{M}^i}(\sqrt{\bar{L}^i}(K(\cdot,y)))\rangle_H,\overset{(b)}{=}\langle\frac{1}{\sqrt{\lambda^i_j}}\psi^i_j(\cdot),$$
$$\sqrt{\bar{L}^i}(K(\cdot,y))\rangle_H\overset{(c)}{=}\langle\sqrt{\bar{L}^i}(\frac{1}{\sqrt{\lambda^i_j}}\psi^i_j(\cdot)),K(\cdot,y)\rangle_H$$
$$=\langle\psi^i_j(\cdot),K(\cdot,y)\rangle_H=\psi^i_j(y),$$

Thus, the reproducing property is satisfied by $\psi^i_j(\cdot)$. Since every function in $\mathcal{M}^i=\bar{H}^i$ can expressed as unique linear combination of $\{\psi^i_j\}$, and by the linearity of inner product it follows that $\langle f(\cdot),K^i(\cdot,y)\rangle_{\bar{H}^i}=f(y),\ \forall f\in\bar{H}^i$. The reasoning for the equalities are as follows, (a) by the definition of the inner product and $\psi^i_j\in\mathcal{M}^i$ as $\lambda^i_j\neq 0$, (b) $\sqrt{\bar{L}^i}(K^i(\cdot,y))\in\bar{H}^i=\mathcal{M}^i$, and (c) symmetry of $\sqrt{\bar{L}^i}$. From Definition 2, it follows that $\bar{H}^i$ is a RKHS with kernel $K^i(\cdot,\cdot)$. □

Corollary II.7. *The downloading operator from the fusion space H to agent i's knowledge space, H^i, is $\sqrt{\bar{L}^i}\circ\Pi_{\mathcal{M}^i}$. The downloading operator is linear and bounded.*

III. Regression and Fusion Problem

The knowledge spaces constructed in the previous section can be used to formulate many inference problems including regression, classification, etc. In this section, we consider the least squares regression problem and the fusion problem as an optimization problem.

A. Regression at the Agents

Given (input-output) information pairs, $\{(x^i_j,y^i_j)\}^m_{j=1}$, to agent i, the objective of the agent is to estimate a mapping from input to output. The estimation problem formulated as least squares regression problem is

$$\min_{f\in H^i}\sum_{j=1}^{m}(y^i_j-f(x^i_j))^2+\varrho^i||f||^2_{H^i}.$$

Let, $\mathbf{K^i}=(K^i(x^i_j,x^i_k))_{jk}=(\langle K^i(\cdot,x^i_k),K^i(\cdot,x^i_j)\rangle_{H^i})$, be the Gram matrix corresponding to agent i as defined in the beginning of Section II. It is well known from the representer theorem (refer [21]), that the solution for the above problem is given by $f^i(\cdot)=\sum_{l=1}^{m}\alpha^i_l K^i(\cdot,x^i_l)$ where $\boldsymbol{\alpha^i}=(\mathbf{K^i}^T\mathbf{K^i}+\varrho^i\mathbf{K^i})^{-1}\mathbf{K^i}^T\mathbf{y^i}$ and $\mathbf{y^i}=(y^i_1,\ldots,y^i_m)$.

B. Function Fusion Problem

The functions estimated by the agents are transmitted to the fusion center where the following fusion problem is considered. As presented in [18], we consider $\{\mathfrak{b}=\{b_k\}_{k\geq 1}\subset H\}$ which span H to define a dissimilarity measure between f^1 and f^2 as $d_{\mathfrak{b}}(f,g)=\sum_k\langle f-g,b_k\rangle^2_H$. The fusion problem is to find a linear combination of f^1 and f^2, f^*, such that the dissimilarity between f^1,f^* and f^2,f^* is minimized. The fusion problem as an optimization problem is

$$\min_{a,b\in\mathbb{R}} d_{\mathfrak{b}}(af^1+bf^2,f^1)+d_{\mathfrak{b}}(af^1+bf^2,f^2)+ \varrho||af^1+bf^2||^2_H.$$

The fused function is considered as the function estimated by the system. It is downloaded by the agents to compare (in the sense of norm) against their own estimates.

IV. Example

In this section, we demonstrate the application of the theory developed in the previous sections. We consider the estimation of a real valued cubic polynomial with real valued inputs. The coefficients of the polynomial where chosen at random. Agent 1 was provided with 20 samples of input-output data, where the input was restricted to the set $[-5,5]$. The inputs were uniformly spaced on the interval $[-5,5]$ and the corresponding outputs were obtained by providing the inputs to the true function. The features considered by Agent 1 were, $\varphi^1_1(x)=1,\varphi^1_2(x)=x$, and $\varphi^1_2(x)=x^2$, which implies that its kernel was $K^1(x,y)=1+xy+x^2y^2$. Agent 2 was also provided with 20 samples of input-output data where the input data was uniformly spaced and restricted to $[-10,-5] \cup [5,10]$, while the features considered by it where $\varphi_1^2(x) = x^2$ and $\varphi_2^2(x) = x^3$. Hence, $K^2(x,y) = x^2y^2 + x^3y^3$.

With this set up, the regression problems (subsection III-A) were solved by the agents. These functions were uploaded to the fusion center. Function uploaded by Agent 1 and Agent 2 are plotted in Figure 2 ([19]) and Figure 3 ([19]) respectively. At the fusion center, the set $\mathfrak{b} = \{K(\cdot, \bar{x}_j)\}_{j=1}^{40}$ was considered, where $\{\bar{x}_j\}_{j=1}^{40}$ were randomly sampled from $[-10, 10]$. The function fusion problem (subsection III-B) was solved. The fused function is plotted in Figure 4 ([19]).

To download the fused function onto the individual KSs, we demonstrate the procedure outlined in subsection II-C. Suppose we choose the set of basis vectors for the space H as $\varphi_1(x) = 1, \varphi_2(x) = x, \varphi_3(x) = \sqrt{2}x^2, \varphi_4(x) = x^3$, then the kernel generated is $K(x,y) = 1 + xy + 2x^2y^2 + x^3y^3 = K^1(x,y) + K^2(x,y)$. With these basis vectors, the coefficients for $K^1(\cdot,y)$ are $[1, y, \frac{y^2}{\sqrt{2}}, 0]$, and for $K^2(\cdot,y)$ are $[0, 0, \frac{y^2}{\sqrt{2}}, y^3]$. The matrix representation of the operators $\bar{L}^i$ and $\sqrt{\bar{L}^i}$, L_M^i and $\sqrt{L_M^i}$, has be derived in [19].

The coefficients of the fused function with respect to basis chosen for H were obtained. The coefficients for downloaded functions were obtained through the operation $\boldsymbol{\alpha}^i = \sqrt{L_M^i}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ is the vector of coefficients corresponding to the fused function. The downloaded functions correspond to $\sum_{j=1}^4 \alpha_j^i \varphi_j$, where $\boldsymbol{\alpha}^i = [\alpha_1^i, \alpha_2^i, \alpha_3^i, \alpha_4^i]$. The function downloaded by Agent 1 and Agent 2 are plotted in Figures 2 and 3 respectively.

We observe that both the uploaded and downloaded function for Agent 1 do not estimate the true function well as it is missing the $\varphi_4(x) = x^3$ feature. Agent 2 is able to "better" estimate the function however the impact of missing data points (between $[-5,5]$) and missing features is visible. However, we note that the agents are able to exchange knowledge as the downloaded functions are "significantly" better than the uploaded functions. To compare the performance of the fusion procedure, we compare it against the centralized estimation method. In this method, the data collected by both agents is sent to the fusion center, where a regression problem (refer subsection III-A) is solved using the collective data. The function estimated using this method is plotted in Figure 4. We observe that the fused function is the best estimate among all estimates considered so far. For further discussion on the practicality of different agents using different kernels and the interpretation of results observed in this example we refer to [19].

V. Conclusion and Future Work

We considered the problem of function estimation by two agents and a fusion center given local data and different set of features. We presented the construction of suitable spaces for the estimation problems and fusion problem to be studied. We derived operators to transform functions across the spaces coherently. A distributed estimation scheme without exchange of data was presented to solve the problem considered. As future work, we are interested in: (i) developing a sequential collaborative learning scheme involving the agents and the fusion center; (ii) studying the consistency properties of such a scheme in the local KS and the fusion space; (iii) quantifying the transfer of knowledge from one agent to another.

References

[1] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[2] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE signal processing magazine*, vol. 34, no. 6, pp. 96–108, 2017.

[3] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley mhad: A comprehensive multimodal human action database," in *2013 IEEE workshop on applications of computer vision (WACV)*. IEEE, 2013, pp. 53–60.

[4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[5] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3–15, 2017.

[6] A. Berlinet and C. Thomas-Agnan, *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.

[7] I. Steinwart and A. Christmann, *Support vector machines*. Springer Science & Business Media, 2008.

[8] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung, "Kernel methods in system identification, machine learning and function estimation: A survey," *Automatica*, vol. 50, no. 3, pp. 657–682, 2014.

[9] T. Chen, M. S. Andersen, L. Ljung, A. Chiuso, and G. Pillonetto, "System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 2933–2945, 2014.

[10] M. Espinoza, J. A. Suykens, and B. De Moor, "Kernel based partially linear models and nonlinear identification," *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1602–1606, 2005.

[11] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *The Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

[12] S. S. Bucak, R. Jin, and A. K. Jain, "Multiple kernel learning for visual object recognition: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1354–1369, 2013.

[13] M. Donini, J. M. Monteiro, M. Pontil, J. Shawe-Taylor, and J. Mourao-Miranda, "A multimodal multiple kernel learning approach to alzheimer's disease detection," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2016, pp. 1–6.

[14] Y. Cho and L. Saul, "Kernel methods for deep learning," *Advances in neural information processing systems*, vol. 22, 2009.

[15] M. Belkin, S. Ma, and S. Mandal, "To understand deep learning we need to understand kernel learning," in *International Conference on Machine Learning*. PMLR, 2018, pp. 541–549.

[16] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[17] Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, and B. He, "A survey on federated learning systems: Vision, hype and reality for data privacy and protection," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[18] A. Raghavan and K. H. Johansson, "Distributed regression by two agents from noisy data," in *2023 European Control Conference (ECC)*. IEEE, 2023, pp. 1–6.

[19] ——, "Distributed estimation by two agents with different feature spaces," *arXiv preprint arXiv:2403.07749*, 2024.

[20] J. S. Baras and A. Raghavan, "Order effects of measurements in multi-agent hypothesis testing," *Communications in Information and Systems*, vol. 21, no. 2, pp. 173–208, 2021.

[21] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *The Annals of Statistics*, vol. 36, no. 3, pp. 1171–1220, 2008.