

# Vector-Valued Reproducing Kernel Banach Spaces for Neural Networks and Operators

Sven Dummer<sup>\*†</sup>, Tjeerd Jan Heeringa<sup>\*†</sup> and José A. Iglesias<sup>†</sup>

## Abstract

Recently, there has been growing interest in characterizing the function spaces underlying neural networks. While shallow and deep scalar-valued neural networks have been linked to scalar-valued reproducing kernel Banach spaces (RKBS),  $\mathbb{R}^d$ -valued neural networks and neural operator models remain less understood in the RKBS setting. To address this gap, we develop a general definition of vector-valued RKBS (vv-RKBS), which inherently includes the associated reproducing kernel. Our construction extends existing definitions by avoiding restrictive assumptions such as symmetric kernel domains, finite-dimensional output spaces, reflexivity, or separability, while still recovering familiar properties of vector-valued reproducing kernel Hilbert spaces (vv-RKHS). We then show that shallow  $\mathbb{R}^d$ -valued neural networks are elements of a specific vv-RKBS, namely an instance of the integral and neural vv-RKBS. To also explore the functional structure of neural operators, we analyze the DeepONet and Hypernetwork architectures and demonstrate that they too belong to an integral and neural vv-RKBS. In all cases, we establish a Representer Theorem, showing that optimization over these function spaces recovers the corresponding neural architectures.

**Keywords:** Reproducing Kernel Banach Space, RKBS, vector-valued, neural networks, neural operator, DeepONet, Hypernetwork, representer theorem

## 1 Introduction

Neural networks are ubiquitous in applications such as computer vision, medical imaging, and scientific computing. Although they perform remarkably well, the fundamental mathematical understanding of neural networks is incomplete.

One approach to advance this understanding is to study neural networks from a function space perspective. This perspective characterizes networks as elements of a function space and rigorously analyzes the analytical properties of the space. For instance, representer theorems are investigated that show that neural network architectures solve supervised optimization problems in these infinite-dimensional function spaces. Examples include spaces inspired by variational splines [38, 39] and the variation (or Barron) space [53, 36, 4, 54, 44, 7, 24, 51, 6]. Both belong to the broader class of Reproducing Kernel Banach Spaces (RKBSs) [14, 56, 23, 30, 25], which strictly generalizes the more familiar Reproducing Kernel Hilbert Spaces (RKHS) [48]. In particular, variation/Barron spaces are examples of integral RKBSs [47]. In this paper, we focus on situations in which the outputs belong to a vector space which is possibly infinite-dimensional, and we refer to the resulting function spaces as vector-valued RKBS (vv-RKBS). Given an output vector space  $\mathcal{U}$  and input domain  $X$ , the integral vv-RKBSs are defined via a feature map  $\Phi: X \rightarrow \mathcal{L}(\mathcal{F}; \mathcal{U})$  with  $\mathcal{F} = \mathcal{M}(\Omega; \mathcal{U})$  the space of regular countably additive  $\mathcal{U}$ -valued measures of bounded variation on the weight domain  $\Omega$ . Specifically, the feature map and resulting functions are defined by integrating a feature function  $\phi: X \times \Omega \rightarrow \mathbb{R}$  against the measure  $\mu \in \mathcal{M}(\Omega; \mathcal{U})$ :

$$(A_{\Omega \rightarrow X} \mu)(x) = \Phi(x) \mu = \int_{\Omega} \phi(x, w) d\mu(w). \quad (1.1)$$

In the scalar case  $\mathcal{U} = \mathbb{R}$ , Spek et al. [47] showed that the space  $\mathcal{B}$  of functions of the form  $A_{\Omega \rightarrow X} \mu$  and the

---

<sup>\*</sup>Equal contribution.

<sup>†</sup>Mathematics of Imaging & AI, Department of Applied Mathematics, University of Twente, the Netherlands (t.j.heeringa@utwente.nl, s.c.dummer@utwente.nl, jose.iglesias@utwente.nl).

2020 Mathematics Subject Classification (MSC): 46E15, 68T07, 46G10, 46E22, 46B10, 26B40

space  $\mathcal{B}^\circ$  of functions defined via a measure  $\rho \in \mathcal{M}(X)$  and interchanging the roles of weight and input domains

$$(A_{X \rightarrow \Omega} \rho)(w) = \int_X \phi(x, w) d\rho(x) \quad (1.2)$$

are adjoint. The corresponding reproducing kernel  $K: X \times \Omega \rightarrow \mathbb{R}$  is given by  $\phi(x, w)$ , and satisfies

$$f(x) = \langle \phi(x, \cdot) | f \rangle_{\mathcal{B}}, \quad g(w) = \langle g | \phi(\cdot, w) \rangle_{\mathcal{B}}, \quad \langle g | f \rangle_{\mathcal{B}} = \int_{\Omega} \int_X \phi(x, w) d\mu(w) d\rho(x), \quad (1.3)$$

for all  $f = A_{\Omega \rightarrow X} \mu$  and  $g = A_{X \rightarrow \Omega} \rho$ . This generalizes the familiar reproducing property from Reproducing Kernel Hilbert Spaces to the Banach space setting.

Most existing theory focuses on vector-input, scalar-output networks. In contrast,  $\mathbb{R}^d$ -valued neural networks, and especially those with outputs in a general Banach space, are far less studied. Besides Bartolucci et al. [7], the few exceptions [51, 44, 39, 27] either map to  $\mathbb{R}^d$  or restrict activation functions. Except for Wang et al. [51], these works do not discuss the existence or role of a reproducing kernel, which is central to the RKBS perspective.

This gap is especially relevant for neural operator methods [11], which map functions to functions and naturally require Banach-valued outputs. Examples include the Fourier Neural Operator (FNO) [29], Graph Neural Operator (GNO) [1], Convolutional Neural Operator (CNO) [42], DeepONet [32], POD-DeepONet [33], NOMAD [43], and PCA-NET [9].

Many neural operators can be interpreted as conditional implicit neural representations (INRs) [46, 20]. In this framework, a network takes a spatial or spatiotemporal coordinate  $x \in \mathbb{R}^{d_x}$  and a conditioning vector  $z \in \mathbb{R}^{d_z}$  to represent a family of output functions via

$$f_\theta: \mathbb{R}^{d_x + d_z} \rightarrow \mathbb{R}^d. \quad (1.4)$$

The DeepONet, POD-DeepONet, NOMAD, and PCA-NET can immediately be interpreted as conditional INRs. The FNO can be viewed as a DeepONet when evaluated on a fixed discretized grid [28], and can hence be interpreted as an INR. The CNO [42] projects inputs to a bandlimited space, processes them, and upsamples back to another bandlimited space. These projection steps yield a kernel-method interpretation. Methods such as RONOM [21] and the approach of Batlle et al. [8] explicitly use RKHS structures and can also be seen as kernel-based INRs. This makes the INR structure a natural framework for investigating neural operators from a function-space perspective.

In this work, we provide a unified framework for these approaches by introducing a general vector-valued Reproducing Kernel Banach Space (vv-RKBS) framework that can map to arbitrary Banach spaces. Specifically, we define a kernel  $K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$ , where  $(\mathcal{U}, \mathcal{U}^\circ)$  form a dual pair and  $\text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$  extends the classical use of bounded linear operators  $\mathcal{L}(\mathcal{U})$  in vv-RKHS to the vv-RKBS setting.

This kernel structure differs from existing vv-RKBS constructions [57, 17, 31, 18], which require  $\Omega = X$  and makes them inapplicable to neural networks. For approaches dealing with finite-dimensional outputs  $\mathcal{U} = \mathbb{R}^d$  [17, 31], the kernel maps to the space of matrices  $\mathcal{L}(\mathcal{U}) = \mathcal{L}(\mathbb{R}^d)$ . Approaches that allow infinite-dimensional outputs [57, 18] instead map to bounded, but not necessarily linear, operators from  $\mathcal{U}$  or  $\mathcal{U}^*$  to  $\mathcal{U}$ . The latter does not capture the linear structure of the space  $\mathcal{L}(\mathcal{U})$  that is used in vv-RKHS. Furthermore, these approaches for infinite-dimensional output spaces rely on reflexivity and, in some cases, separability of the RKBS, its feature space  $\mathcal{F}$ , or the output space. These properties are used to construct the kernel and prove representer theorems. In contrast, our framework imposes no reflexivity or separability assumptions.

This allows us to handle, for instance, feature spaces  $\mathcal{F}$  that are not reflexive. In particular, to connect the vv-RKBS setting with neural networks and neural operators, we consider the integral vv-RKBS introduced by Bartolucci et al. [7] and defined in (1.1) using a feature space of vector-valued measures. Most existing works on  $\mathbb{R}^d$ - and vector-valued neural networks [51, 44, 39, 27] either map to  $\mathbb{R}^d$  or restrict to homogeneous activation functions. The integral vv-RKBS construction avoids these limitations as it supports arbitrary Banach output spaces, allows non-homogeneous activation functions, and still yields representer theorems. Beyond the work of Bartolucci et al., we also provide the explicit kernel structure for these spaces and establish a representer theorem for infinite-dimensional output spaces. In addition, to explore the function space structure of neural operators, we utilize the previously discussed connection with conditional INRs. In particular, we construct an integral vv-RKBS for hypernetworks and DeepONets and prove a representer theorem.

## 1.1 Contributions

Our main contributions can be summarized as follows:

- We extend the kernel definition of vv-RKBSs to allow asymmetric kernel domains and to avoid structural assumptions such as reflexivity and separability. In addition, we demonstrate that several classical properties of vv-RKHSs still extend to this more general setting.
- We relate our vv-RKBSs and their reproducing kernels to existing formulations in the scalar RKBS setting. In particular, we show that our definition either recovers the corresponding vector-valued extensions directly or, with additional assumptions, recovers stronger versions.
- We further develop the integral and neural vv-RKBS framework introduced in Bartolucci et al. [7] by analyzing its reproducing kernel structure and establishing a general representer theorem.
- We embed  $\mathbb{R}^d$ -valued neural networks, DeepONets, and hypernetworks into the integral vv-RKBS framework and obtain representer theorems for them. Together, these ensure a clear bridge between vector-valued neural architectures and the kernel structure of RKBSs.

## 1.2 Paper outline

Section 2 reviews vector-valued reproducing kernel Hilbert spaces (vv-RKHS) for comparison with our vector-valued reproducing kernel Banach space (vv-RKBS) definitions. Section 3.1 introduces the general definition of a vv-RKBS through bounded point evaluation functionals and feature maps. Section 3.2 then develops the notion of adjoint vv-RKBS pairs, which naturally incorporate reproducing kernels. Their main properties are examined in Section 3.3, where we show, among other results, that every space defined via bounded point evaluations or feature maps admits a reproducing kernel within some adjoint pair.

Section 4 turns to adjoint vv-RKBS pairs tailored to neural networks, namely the integral and neural vv-RKBSs. After recalling the scalar-valued setting, we extend to the vector-valued case, establish key properties, and prove a general representer theorem. Finally, Section 5 specializes this framework to  $\mathbb{R}^d$ -valued neural networks, DeepONets, and hypernetworks, and provides representer theorems for them.

In Appendix A we provide a table of notation to facilitate the reading of the paper. Appendix B contains some additional technical results that are used in the main text, but which would otherwise interrupt the narrative.

## 2 Vector-valued Reproducing Kernel Hilbert Spaces

The classical definition of Reproducing Kernel Hilbert Spaces (RKHS) in terms of bounded evaluation functionals can be extended to the vector-valued case.

**Definition 2.1** (Vector-valued Reproducing Kernel Hilbert space). Let  $\mathcal{H}$  be a Hilbert space of functions over a set  $X$  mapping to a Hilbert space  $\mathcal{U}$ .  $\mathcal{H}$  is a vector-valued Reproducing Kernel Hilbert space (vv-RKHS) if point evaluations are bounded functionals, i.e. for all  $x \in X$

$$\|f(x)\|_{\mathcal{U}} \leq C_x \|f\|_{\mathcal{H}} \quad (2.1)$$

holds for all  $f \in \mathcal{H}$  with the constant  $C_x \geq 0$  depending on  $x$  but not on  $f$ .  $\diamond$

Just like the scalar RKHS, the vv-RKHS has an equivalent definition in terms of a kernel.

**Definition 2.2** (Kernel definition vv-RKHS). A Hilbert space  $\mathcal{H}$  of functions over a set  $X$  mapping to a Hilbert space  $\mathcal{U}$  is a vv-RKHS if and only if there exists a kernel  $K: X \times X \rightarrow \mathcal{L}(\mathcal{U})$  such that for all  $x \in X$  and  $u \in \mathcal{U}$ :

$$K(x, \cdot)u \in \mathcal{H} \quad (2.2a)$$

$$\langle u|f(x) \rangle_{\mathcal{U}} = \langle K(x, \cdot)u|f \rangle_{\mathcal{H}} \quad (2.2b)$$

$\diamond$

**Theorem 2.1.** *Definitions 2.1 and 2.2 are equivalent.*

*Proof.* Consider the bilinear functional

$$T_x: \mathcal{U} \times \mathcal{H} \rightarrow \mathbb{R}, (u, f) \mapsto \langle u|f(x) \rangle_{\mathcal{U}}. \quad (2.3)$$

Fixing either argument yields a linear functional in the other, and these functionals are bounded since

$$\langle u|f(x) \rangle_{\mathcal{U}} \leq \|u\|_{\mathcal{U}} \|f(x)\|_{\mathcal{U}} \leq C_x \|f\|_{\mathcal{H}} \|u\|_{\mathcal{U}}. \quad (2.4)$$

Hence, the map  $f \mapsto T_x(u, f)$  defines a bounded linear functional on the Hilbert space  $\mathcal{H}$ . By the Riesz representation theorem, there exists  $K(x, \cdot)u \in \mathcal{H}$  such that  $T_x(u, f) = \langle K(x, \cdot)u|f \rangle_{\mathcal{H}}$ . Moreover, by linearity of  $T_x(u, f)$  in  $u$ , we get  $T_x(u_1 + \lambda u_2, f) = \langle K(x, \cdot)u_1 + \lambda K(x, \cdot)u_2|f \rangle_{\mathcal{H}}$ . Hence,  $K(x, \cdot)(u_1 + \lambda u_2) = K(x, \cdot)u_1 + \lambda K(x, \cdot)u_2$ , meaning that  $K(x, \cdot): \mathcal{U} \rightarrow \mathcal{H}$  is linear. Since  $\mathcal{H}$  is a Hilbert space of functions and  $K(x, \cdot)u \in \mathcal{H}$ , we can evaluate at another value  $y \in X$ . This gives us  $K(x, y): \mathcal{U} \rightarrow \mathcal{U}$ . Note that

$$\|K(x, \cdot)u\|_{\mathcal{H}}^2 = \langle K(x, \cdot)u|K(x, \cdot)u \rangle_{\mathcal{H}} = \langle u|K(x, x)u \rangle_{\mathcal{U}} \leq \|u\|_{\mathcal{U}} \|K(x, x)u\|_{\mathcal{U}} \leq \|u\|_{\mathcal{U}} C_x \|K(x, \cdot)u\|_{\mathcal{H}}. \quad (2.5)$$

Dividing both sides by  $\|K(x, \cdot)u\|_{\mathcal{H}}$  gives  $\|K(x, \cdot)u\|_{\mathcal{H}} \leq C_x \|u\|_{\mathcal{U}}$ . Using this yields

$$\|K(x, y)u\|_{\mathcal{U}} \leq C_y \|K(x, \cdot)u\|_{\mathcal{H}} \leq C_y C_x \|u\|_{\mathcal{U}}, \quad (2.6)$$

which shows that  $K(x, y) \in \mathcal{L}(\mathcal{U})$ . Summarizing, these observations regarding the Riesz representation yield the kernel definition.

For the converse, observe that we can take  $C_x = \sqrt{\|K(x, x)\|_{\mathcal{L}(\mathcal{U})}} = \sqrt{\sup_{\|u\|_{\mathcal{U}} \leq 1} \|K(x, x)u\|_{\mathcal{U}}} < \infty$  since

$$\begin{aligned} \|f(x)\|_{\mathcal{U}} &= \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle u|f(x) \rangle_{\mathcal{U}}| \\ &= \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle K(x, \cdot)u|f \rangle_{\mathcal{H}}| \\ &\leq \|f\|_{\mathcal{H}} \sup_{\|u\|_{\mathcal{U}} \leq 1} \|K(x, \cdot)u\|_{\mathcal{H}} \\ &= \|f\|_{\mathcal{H}} \sup_{\|u\|_{\mathcal{U}} \leq 1} \sqrt{\langle K(x, \cdot)u|K(x, \cdot)u \rangle_{\mathcal{H}}} \\ &= \|f\|_{\mathcal{H}} \sup_{\|u\|_{\mathcal{U}} \leq 1} \sqrt{\langle K(x, x)u|u \rangle_{\mathcal{U}}} \\ &\leq \|f\|_{\mathcal{H}} \sup_{\|u\|_{\mathcal{U}} \leq 1} C_x \|u\|_{\mathcal{U}} \\ &= C_x \|f\|_{\mathcal{H}}. \end{aligned} \quad \square$$

The kernel  $K$  in Definition 2.2 satisfies

$$\sum_{i,j} \langle K(x_i, x_j)u_i|u_j \rangle_{\mathcal{U}} = \sum_{i,j} \langle K(x_i, \cdot)u_i|K(x_j, \cdot)u_j \rangle_{\mathcal{H}} = \left\| \sum_i K(x_i, \cdot)u_i \right\|_{\mathcal{H}}^2 \geq 0 \quad (2.7)$$

for all  $(x_i, u_i) \in X \times \mathcal{U}$ , and

$$\begin{aligned} \langle K(x, y)^*u|v \rangle_{\mathcal{U}} &= \langle u|K(x, y)v \rangle_{\mathcal{U}} = \langle K(y, \cdot)u|K(x, \cdot)v \rangle_{\mathcal{H}} \\ &= \langle K(x, \cdot)v|K(y, \cdot)u \rangle_{\mathcal{H}} = \langle v|K(y, x)u \rangle_{\mathcal{U}} = \langle K(y, x)u|v \rangle_{\mathcal{U}}. \end{aligned} \quad (2.8)$$

for all  $(x, u), (y, v) \in X \times \mathcal{U}$ . Hence, each vv-RKHS kernel is of positive type (see Definition 2.2 in Carmeli et al. [15]) and  $K(x, y)^* = K(y, x)$ . Each such kernel  $K: X \times X \rightarrow \mathcal{L}(\mathcal{U})$  defines a unique vv-RKHS of  $\mathcal{U}$ -valued functions (see Proposition 2.3 in Carmeli et al. [15]), which is similar to the classical result by Aronszajn [2] for scalar-valued RKHS. Therefore, any vv-RKHS induces a symmetric positive semi-definite kernel, and every symmetric positive semi-definite kernel induces a vv-RKHS. In particular, starting with  $\mathcal{H}_s$  a scalar-valued RKHS with kernel  $K_s: X \times X \rightarrow \mathbb{R}$  and  $\mathcal{U}$  any Hilbert space, one can canonically define a kernel

$$K(\cdot, \cdot) = K_s(\cdot, \cdot) \text{Id}_{\mathcal{U}} \quad (2.9)$$

which then induces a vv-RKHS  $\mathcal{H}$  of functions  $f: X \rightarrow \mathcal{U}$ .

Another useful vv-RKHS characterization that is often used in machine learning is in terms of feature maps [15].

**Definition 2.3** (Feature map definition vv-RKHS). A Hilbert space  $\mathcal{H}$  of functions from  $X$  to a Hilbert space  $\mathcal{U}$  is a vv-RKHS if and only if there exists a Hilbert space  $\mathcal{F}$  and a map  $\Phi: X \rightarrow \mathcal{L}(\mathcal{F}, \mathcal{U})$  such that for  $(A\mu)(x) = \Phi(x)\mu$ :

1.  $\mathcal{H} = \{A\mu: x \mapsto (A\mu)(x) \mid \mu \in \mathcal{F}\} \cong \mathcal{F}/\mathcal{N}(A)$ ,
2.  $\|f\|_{\mathcal{H}} = \inf (\|\mu\|_{\mathcal{F}} \mid \mu \in \mathcal{F}, f = A\mu)$ .

◇

### 3 Vector-valued Reproducing Kernel Banach Spaces

While functions in a vv-RKHS map to a Hilbert space, functions in a vector-valued reproducing kernel Banach space (vv-RKBS) map to a Banach space. This section first introduces the standard definitions of vector-valued RKBSs, which do not assume any kernel structure. We subsequently define adjoint pairs of vv-RKBSs to introduce the reproducing kernel structure. Finally, we discuss properties of these adjoint pairs that parallel known results for vv-RKHSs. In particular, we show that every vv-RKBS has a reproducing kernel for some adjoint pair.

#### 3.1 Vector-valued RKBS definition

An easy way to change a vv-RKHS into a vv-RKBS is to adjust Definition 2.1 from the Hilbert space setting to the Banach space setting.

**Definition 3.1** (Vector-valued Reproducing Kernel Banach space). Let  $\mathcal{B}$  be a Banach space of functions over a set  $X$  mapping to a Banach space  $\mathcal{U}$ .  $\mathcal{B}$  is a vector-valued Reproducing Kernel Banach space (vv-RKBS) if point evaluations are bounded functionals, i.e. for all  $x \in X$

$$\|f(x)\|_{\mathcal{U}} \leq C_x \|f\|_{\mathcal{B}} \quad (3.1)$$

holds for all  $f \in \mathcal{B}$  with the constant  $C_x \geq 0$  depending on  $x$  but not on  $f$ .

◇

Similarly, Definition 2.3, where the vv-RKHS is defined via a feature map, can be extended by replacing the Hilbert spaces in the vv-RKHS definition by Banach spaces. This yields a definition of vv-RKBS equivalent to Definition 3.1 [7].

**Definition 3.2** (Feature map definition vv-RKBS). A Banach space  $\mathcal{B}$  of functions from  $X$  to a Banach space  $\mathcal{U}$  is a vv-RKBS if and only if there exists a Banach space  $\mathcal{F}$  and a map  $\Phi: X \rightarrow \mathcal{L}(\mathcal{F}, \mathcal{U})$  such that for  $(A\mu)(x) = \Phi(x)\mu$ :

1.  $\mathcal{B} = \{A\mu: x \mapsto (A\mu)(x) \mid \mu \in \mathcal{F}\} \cong \mathcal{F}/\mathcal{N}(A)$ ,
2.  $\|f\|_{\mathcal{B}} = \inf (\|\mu\|_{\mathcal{F}} \mid \mu \in \mathcal{F}, f = A\mu)$ .

◇

**Remark 3.1.** The scalar-valued RKBS definition [6] follows by setting  $\|\cdot\|_{\mathbb{R}} = |\cdot|$  in Definition 3.1. In Definition 3.2, we interpret  $\mathcal{L}(\mathcal{F}, \mathbb{R})$  as the dual space  $\mathcal{F}^*$  and write  $\Phi(x)\mu = \langle \Phi(x) | \mu \rangle$ , where  $\langle \cdot | \cdot \rangle$  denotes the canonical duality pairing.

◇

#### 3.2 Adjoint vv-RKBS pairs

Due to the absence of the Riesz representation theorem, there is no immediate kernel-based definition of an RKBS. To address this, many works impose reproducing kernel assumptions. For scalar RKBSs, some adopt a  $\delta$ -dual approach, assuming that the closure of the point evaluation functionals is isometrically isomorphic to a Banach space of functions or another RKBS [55, 50]. Others assume the existence of a Banach space of functions or RKBS that is isometrically embedded in the dual space  $\mathcal{B}^*$  of the RKBS  $\mathcal{B}$  [56, 47]. A third group of works uses less strict assumptions and only assumes the existence of another Banach space of functions (e.g., an RKBS)  $\mathcal{B}^\circ$  and a continuous bilinear form on  $\mathcal{B} \times \mathcal{B}$  [23, 30, 25, 24]. Based on this form, they define a reproducing kernel property for  $\mathcal{B}$ , and optionally also for  $\mathcal{B}^\circ$ .

In all these cases, the reproducing property is required for the original space  $\mathcal{B}$ , but is not always assumed for the  $\delta$ -dual, the embedded function space in  $\mathcal{B}^*$ , or the space  $\mathcal{B}^\circ$  appearing in the bilinear form. In contrast, we focus on the case where both  $\mathcal{B}$  and its counterpart form an adjoint RKBS pair, meaning both are RKBSs

and possess reproducing properties. We begin with the scalar-valued setting. Here, adjoint RKBS pairs and the corresponding kernel are defined through a continuous bilinear form, specifically a dual pairing between the RKBS  $\mathcal{B}$  and another RKBS  $\mathcal{B}^\circ$ . We then extend this framework to vector-valued RKBSs. In the next section, we show that, under additional assumptions, the dual pairing approach recovers the aforementioned isometric isomorphism-based formulations.

**Definition 3.3** (Dual pair [22]). A dual pair of Banach spaces  $(\mathcal{B}, \mathcal{B}^\circ)$  is two Banach spaces  $\mathcal{B}$  and  $\mathcal{B}^\circ$  with a bilinear map

$$\langle \cdot | \cdot \rangle_{\mathcal{B}} : \mathcal{B}^\circ \times \mathcal{B} \rightarrow \mathbb{R}, (g, f) \mapsto \langle g | f \rangle_{\mathcal{B}} \quad (3.2)$$

that satisfies

$$\langle g | f \rangle_{\mathcal{B}} = 0 \quad \forall g \in \mathcal{B}^\circ \implies f = 0 \quad (3.3a)$$

$$\langle g | f \rangle_{\mathcal{B}} = 0 \quad \forall f \in \mathcal{B} \implies g = 0 \quad (3.3b)$$

This bilinear map is called the pairing corresponding to  $(\mathcal{B}, \mathcal{B}^\circ)$  or, when no confusion arises, simply the pairing. Furthermore, the pairing is called continuous if the bilinear map is continuous.  $\diamond$

**Definition 3.4** (Adjoint pair of scalar RKBS). Let  $\mathcal{B}$  be an RKBS of functions mapping from a set  $X$  to  $\mathbb{R}$  and let  $\mathcal{B}^\circ$  be an RKBS of functions mapping from a set  $\Omega$  to  $\mathbb{R}$ . Moreover, let  $\langle \cdot | \cdot \rangle_{\mathcal{B}} : \mathcal{B}^\circ \times \mathcal{B} \rightarrow \mathbb{R}$  be a continuous dual pairing and  $K : X \times \Omega \rightarrow \mathbb{R}$ . We call  $K$  the reproducing kernel for  $\mathcal{B}$  when  $K(x, \cdot) \in \mathcal{B}^\circ$  for all  $x \in X$  and

$$f(x) = \langle K(x, \cdot) | f \rangle_{\mathcal{B}} \quad (3.4)$$

for all  $x \in X$  and  $f \in \mathcal{B}$ . If additionally  $K(\cdot, w) \in \mathcal{B}$  for all  $w \in \Omega$  and

$$g(w) = \langle g | K(\cdot, w) \rangle_{\mathcal{B}} \quad (3.5)$$

for all  $w \in \Omega$  and  $g \in \mathcal{B}^\circ$ , then we call  $\mathcal{B}^\circ$  an adjoint RKBS of  $\mathcal{B}$  and  $(\mathcal{B}, \mathcal{B}^\circ)$  an adjoint pair of RKBSs.  $\diamond$

**Remark 3.2.** While some works use a dual pairing [30, 24], others disregard the non-degeneracy conditions (3.3a) and (3.3b) and consider only a (continuous) bilinear form [23, 25]. In principle, one can work with a general bilinear form instead of a duality pairing, since the non-degeneracy conditions are implicitly enforced by the reproducing properties.

Suppose there exists  $f \in \mathcal{B}$  such that  $\langle g | f \rangle_{\mathcal{B}} = 0$  for all  $g \in \mathcal{B}^\circ$ . Taking  $g = K(x, \cdot)$  yields  $f(x) = \langle K(x, \cdot) | f \rangle_{\mathcal{B}} = 0$  for all  $x \in X$ , and hence  $f = 0$ . This confirms that condition (3.3a) must hold. A similar argument shows that condition (3.3b) is also necessary.

Thus, any bilinear form defining a reproducing property must be a duality pairing. Nevertheless, we retain non-degeneracy as an explicit assumption in our definition to emphasize its necessity and to avoid the misconception that an arbitrary bilinear form is sufficient.  $\diamond$

Definition 3.4 only applies to scalar-valued RKBSs, as vector-valued functions  $f$  cannot be represented by  $\langle K(x, \cdot) | f \rangle_{\mathcal{B}} \in \mathbb{R}$ . To resolve this issue, we can take an approach similar to the vv-RKHS setting in Definition 2.2. Instead of considering the function values directly, the vv-RKHS considers the inner products. However, since we are dealing with functions  $f : X \rightarrow \mathcal{U}$  with  $\mathcal{U}$  a Banach space, inner products are not available in general. We therefore replace the inner products with a duality pairing.

To that end, we introduce a dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$  with pairing  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$ . Then, similar to the RKHS setting, we aim for a representation of the form

$$\langle u^\circ | f(x) \rangle_{\mathcal{U}} = \langle K_{\mathcal{U}^\circ}(x, \cdot) u^\circ | f \rangle_{\mathcal{B}} \quad (3.6)$$

for some kernel function  $K_{\mathcal{U}^\circ}(x, \cdot) u^\circ \in \mathcal{B}^\circ$  with  $K_{\mathcal{U}^\circ}(x, w) \in \mathcal{L}(\mathcal{U}^\circ)$  and  $w \in \Omega$ . On the adjoint side, we require

$$\langle g(w) | u \rangle_{\mathcal{U}} = \langle g | K_{\mathcal{U}}(\cdot, w) u \rangle_{\mathcal{B}} \quad (3.7)$$

for some  $K_{\mathcal{U}}(\cdot, w) u \in \mathcal{B}$  with  $K_{\mathcal{U}}(x, w) \in \mathcal{L}(\mathcal{U})$  and  $x \in X$ . In the vv-RKHS case, one has the identity

$$\langle f(x) | u \rangle_{\mathcal{U}} = \langle u | f(x) \rangle_{\mathcal{U}} = \langle K(x, \cdot) u | f \rangle_{\mathcal{H}} = \langle f | K(x, \cdot) u \rangle_{\mathcal{H}}, \quad (3.8)$$

so it is natural to take  $K_{\mathcal{U}} = K_{\mathcal{U}^\circ} = K$  in that setting. In general, however,  $K_{\mathcal{U}} \neq K_{\mathcal{U}^\circ}$  as they act on different spaces.

Moreover, in the vv-RKHS framework with a real-valued inner product, one has the identity

$$\langle K(y, x)u|v \rangle_{\mathcal{U}} = \langle u|K(x, y)v \rangle_{\mathcal{U}} = \langle K(x, \cdot)u|K(\cdot, y)v \rangle_{\mathcal{H}} =: K_b(x, y)(u, v), \quad (3.9)$$

where the first equality follows from the symmetry property  $K(x, y)^* = K(y, x)$ . This shows that combining the inner product on  $\mathcal{U}$  with a linear operator in either the first or second argument yields the same scalar output, which in turn can also be obtained via a specific bilinear form  $K_b(x, y): \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}$ . We aim to mimic this behavior when replacing the inner product with a duality pairing.

To formalize this structure and the reproducing property, we define the kernel as

$$K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\diamond). \quad (3.10)$$

Here,  $(\mathcal{U}, \mathcal{U}^\diamond)$  is a dual pair, and  $\text{Twin}(\mathcal{U}, \mathcal{U}^\diamond)$  is the space of twin operators which we introduce below and is based on the definition introduced in Diekmann et al. [19] for the more restrictive case in which  $(\mathcal{U}, \mathcal{U}^\diamond)$  is assumed to be norming. The space  $\text{Twin}(\mathcal{U}, \mathcal{U}^\diamond)$  can be viewed as a generalization of  $\mathcal{L}(\mathcal{U})$  (see Theorem 3.1), which is used in the vv-RKHS.

**Definition 3.5** (Twin operators). An operator  $T: \mathcal{U}^\diamond \times \mathcal{U} \rightarrow \mathbb{R}$  over a dual pair  $(\mathcal{U}, \mathcal{U}^\diamond)$  with pairing  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$  is a twin operator when it is a bounded bilinear map that defines both a linear operator  $T_{\mathcal{U}}: \mathcal{U} \rightarrow \mathcal{U}$  and linear operator  $T_{\mathcal{U}^\diamond}: \mathcal{U}^\diamond \rightarrow \mathcal{U}^\diamond$  via:

- $T_{\mathcal{U}}: \langle u^\diamond | T_{\mathcal{U}}u \rangle_{\mathcal{U}} = T(u^\diamond, u) \quad \forall u^\diamond \in \mathcal{U}^\diamond \quad \forall u \in \mathcal{U}$ , and
- $T_{\mathcal{U}^\diamond}: \langle T_{\mathcal{U}^\diamond}u^\diamond | u \rangle_{\mathcal{U}} = T(u^\diamond, u) \quad \forall u^\diamond \in \mathcal{U}^\diamond \quad \forall u \in \mathcal{U}$ .

The space of all twin operators  $T: \mathcal{U}^\diamond \times \mathcal{U} \rightarrow \mathbb{R}$  is denoted as  $\text{Twin}(\mathcal{U}, \mathcal{U}^\diamond)$  and endowed with norm

$$\|T\|_{\text{Twin}(\mathcal{U}, \mathcal{U}^\diamond)} = \sup_{\|u\|_{\mathcal{U}} \leq 1, \|u^\diamond\|_{\mathcal{U}^\diamond} \leq 1} |T(u^\diamond, u)| \quad (3.11)$$

**Remark 3.3.** In our setting, the twin operator  $T$  takes the role of  $K_b$  and the associated operators  $T_{\mathcal{U}}$  and  $T_{\mathcal{U}^\diamond}$  correspond to  $K_{\mathcal{U}}$  and  $K_{\mathcal{U}^\diamond}$ , respectively. Moreover,  $T_{\mathcal{U}^\diamond}$  can be interpreted as the adjoint of  $T_{\mathcal{U}}$ , and  $T_{\mathcal{U}}u$  plays the role of a Riesz representation. Specifically, for fixed  $u \in \mathcal{U}$ , the map  $u^\diamond \mapsto T(u^\diamond, u)$  defines a bounded linear functional on  $\mathcal{U}^\diamond$  by the boundedness of  $T$ . When  $\mathcal{U}^\diamond = \mathcal{U}$  and  $\mathcal{U}$  is a Hilbert space, the Riesz representation theorem guarantees the existence of an element  $T_{\mathcal{U}}u \in \mathcal{U}$  such that  $\langle u^\diamond | T_{\mathcal{U}}u \rangle_{\mathcal{U}} = T(u^\diamond, u)$ . Since  $T$  is linear in  $u$ , we have

$$\langle u^\diamond | T_{\mathcal{U}}(u + \lambda v) \rangle_{\mathcal{U}} = T(u^\diamond, u + \lambda v) = T(u^\diamond, u) + \lambda T(u^\diamond, v) = \langle u^\diamond | T_{\mathcal{U}}u + \lambda T_{\mathcal{U}}v \rangle_{\mathcal{U}}, \quad (3.12)$$

which implies that  $T_{\mathcal{U}}(u + \lambda v) = T_{\mathcal{U}}u + \lambda T_{\mathcal{U}}v$ , i.e.,  $T_{\mathcal{U}}$  is linear. Thus, in the RKHS setting, we know that a linear operator  $T_{\mathcal{U}}$  exists such that  $\langle u^\diamond | T_{\mathcal{U}}u \rangle_{\mathcal{U}} = T(u^\diamond, u)$  holds for all  $u^\diamond \in \mathcal{U}^\diamond = \mathcal{U}$  and  $u \in \mathcal{U}$ . However, in the Banach space setting, no such representation theorem holds in general, so we must assume the identity  $\langle u^\diamond | T_{\mathcal{U}}u \rangle_{\mathcal{U}} = T(u^\diamond, u)$  as part of the structure. Similar considerations apply for the operator  $T_{\mathcal{U}^\diamond}$ .  $\diamond$

**Remark 3.4.** When the pairing is continuous, an important twin operator is the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{U}}: \mathcal{U}^\diamond \times \mathcal{U} \rightarrow \mathbb{R}$  itself, with both induced linear operators being the identity on  $\mathcal{U}$  and  $\mathcal{U}^\diamond$ .  $\diamond$

Using this definition of the Twin operators, we can define the adjoint pair of vv-RKBSs.

**Definition 3.6** (Adjoint pair of vv-RKBS). Given a dual pair of Banach spaces  $(\mathcal{U}, \mathcal{U}^\diamond)$ , let  $\mathcal{B}$  be a vv-RKBS of functions mapping from a set  $X$  to  $\mathcal{U}$  and let  $\mathcal{B}^\diamond$  be a vv-RKBS of functions mapping from a set  $\Omega$  to  $\mathcal{U}^\diamond$ . Moreover, let  $\langle \cdot | \cdot \rangle_{\mathcal{B}}: \mathcal{B}^\diamond \times \mathcal{B} \rightarrow \mathbb{R}$  be a continuous dual pairing and  $K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\diamond)$ . We call  $K$  the reproducing kernel for  $\mathcal{B}$  when  $K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \in \mathcal{B}^\diamond$  for all  $x, u^\diamond \in X \times \mathcal{U}^\diamond$  and

$$\langle u^\diamond | f(x) \rangle_{\mathcal{U}} = \langle K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond | f \rangle_{\mathcal{B}} \quad (3.13)$$

for all  $x, u^\diamond \in X \times \mathcal{U}^\diamond$  and  $f \in \mathcal{B}$ . Moreover, if  $K_{\mathcal{U}}(\cdot, w)u \in \mathcal{B}$  for all  $(w, u) \in \Omega \times \mathcal{U}$  and

$$\langle g(w) | u \rangle_{\mathcal{U}} = \langle g | K_{\mathcal{U}}(\cdot, w)u \rangle_{\mathcal{B}} \quad (3.14)$$

for all  $(w, u) \in \Omega \times \mathcal{U}$  and  $g \in \mathcal{B}^\diamond$ , then we call  $\mathcal{B}^\diamond$  an adjoint vv-RKBS of  $\mathcal{B}$  and  $(\mathcal{B}, \mathcal{B}^\diamond)$  an adjoint pair of vv-RKBSs.  $\diamond$

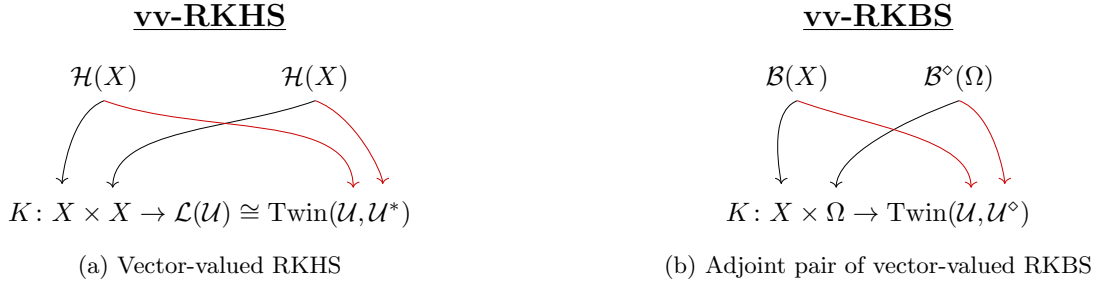


Figure 1: **From vv-RKHS to adjoint pair of vv-RKBS**. In a vv-RKHS, functions  $f: X \rightarrow \mathcal{U}$  take values in a Hilbert space  $\mathcal{U}$  and admit a reproducing kernel  $K: X \times X \rightarrow \mathcal{L}(\mathcal{U})$ . In the vv-RKBS setting, we instead use Banach spaces  $\mathcal{B}$  and  $\mathcal{B}^\circ$  for functions mapping to a dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$ , and replace inner products with duality pairings. This breaks the symmetry in the domain, replacing  $X \times X$  with  $X \times \Omega$ , and requires twin operators in place of  $\mathcal{L}(\mathcal{U})$  to accommodate the asymmetry.

**Remark 3.5.** We do not assume  $(\mathcal{U}, \mathcal{U}^\circ)$  to be a continuous dual pairing. The reason is that for the reproducing property of  $\mathcal{B}$  and the continuity of the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$ , it suffices that  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$  is continuous on the values attained by functions in  $\mathcal{B}$ . An analogous argument applies to the reproducing property of  $\mathcal{B}^\circ$ . Hence, full continuity of the dual pairing  $(\mathcal{U}, \mathcal{U}^\circ)$  is not required.  $\diamond$

Figure 1 summarizes the transition from the definition of a vv-RKHS to that of an adjoint pair of vv-RKBS. First, observe that the domain changes from  $X \times X$  to  $X \times \Omega$ , reflecting the fact that Banach spaces are not necessarily isomorphic to their duals. Moreover, instead of using the space of bounded linear operators  $\mathcal{L}(\mathcal{U})$ , we work with twin operators, which serve as a natural generalization of  $\mathcal{L}(\mathcal{U})$  as established in the theorem below.

**Theorem 3.1** (Twin operators generalize  $\mathcal{L}(\mathcal{U})$ ). *If  $\mathcal{U}$  is a reflexive space, then  $\text{Twin}(\mathcal{U}, \mathcal{U}^*) \cong \mathcal{L}(\mathcal{U})$ . Otherwise, we can only interpret  $\mathcal{L}(\mathcal{U})$  as a subset of  $\text{Twin}(\mathcal{U}, \mathcal{U}^*)$ .*

*Proof.* Define  $\mathcal{U}^\circ = \mathcal{U}^*$  and let  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$  be the canonical duality pairing. First, note that any  $T_{\mathcal{U}} \in \mathcal{L}(\mathcal{U})$  defines a bounded bilinear form via  $(u^*, u) \mapsto \langle u^* | T_{\mathcal{U}} u \rangle_{\mathcal{U}} = \langle T_{\mathcal{U}}^* u^* | u \rangle_{\mathcal{U}}$ , where  $T_{\mathcal{U}}^*$  is the adjoint of  $T_{\mathcal{U}}$ . Hence,  $\mathcal{L}(\mathcal{U})$  can always be seen as a subset of  $\text{Twin}(\mathcal{U}, \mathcal{U}^*)$ .

In case  $\mathcal{U}$  is reflexive, note that the set of bounded bilinear forms  $T: \mathcal{U}^* \times \mathcal{U} \rightarrow \mathbb{R}$  is isomorphic to the set of maps  $\tilde{T}: \mathcal{U} \rightarrow \mathcal{U}^{**}$ . In particular, the isomorphism is given by the mapping  $\kappa(T)(u) = T(\cdot, u) \in \mathcal{U}^{**}$  for  $u \in \mathcal{U}$  and its inverse is  $\kappa^{-1}(\tilde{T})(u^*, u) = (\tilde{T}(u))(u^*)$ . Furthermore,  $\kappa$  is isometric, because

$$\begin{aligned}
\|\kappa(T)\|_{\mathcal{L}(\mathcal{U}, \mathcal{U}^{**})} &= \sup_{\|u\|_{\mathcal{U}} \leq 1} \|\kappa(T)(u)\|_{\mathcal{U}^{**}} \\
&= \sup_{\|u\|_{\mathcal{U}} \leq 1, \|u^*\|_{\mathcal{U}^*} \leq 1} |\kappa(T)(u)(u^*)| \\
&= \sup_{\|u\|_{\mathcal{U}} \leq 1, \|u^*\|_{\mathcal{U}^*} \leq 1} |T(u^*, u)| \\
&= \|T\|_{\text{Twin}(\mathcal{U}, \mathcal{U}^*)}.
\end{aligned} \tag{3.15}$$

Since  $\mathcal{U}$  is reflexive, the evaluation map  $\iota: \mathcal{U} \rightarrow \mathcal{U}^{**}$  defined by  $u \mapsto \langle \cdot | u \rangle_{\mathcal{U}}$  is an isometry. Combining this with the isometric isomorphism  $\kappa$ , we obtain another isometric isomorphism

$$\tilde{\kappa}(T)(u) = \iota^{-1}(\kappa(T)(u)) \tag{3.16}$$

whose inverse is

$$\tilde{\kappa}^{-1}(T_{\mathcal{U}})(u^*, u) = \kappa^{-1}(\iota \circ T_{\mathcal{U}})(u^*, u) = \iota(T_{\mathcal{U}}(u))(u^*) \tag{3.17}$$

for  $T_{\mathcal{U}} \in \mathcal{L}(\mathcal{U})$ . Hence, every bounded bilinear operator corresponds uniquely to a  $T_{\mathcal{U}} \in \mathcal{L}(\mathcal{U})$  via

$$T(u^*, u) = \tilde{\kappa}^{-1}(T_{\mathcal{U}})(u^*, u) = \iota(T_{\mathcal{U}}(u))(u^*) = \langle u^* | T_{\mathcal{U}}(u) \rangle_{\mathcal{U}} = \langle T_{\mathcal{U}}^* u^* | u \rangle_{\mathcal{U}} \tag{3.18}$$

As  $\kappa$  is an isometry, this shows that  $\text{Twin}(\mathcal{U}, \mathcal{U}^*) \cong \mathcal{L}(\mathcal{U})$ .  $\square$

As a final remark on the twin operators, we show that, under some assumptions, the norm of a twin operator  $T \in \text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$  is linked to the norms of the two bounded linear operators  $T_{\mathcal{U}}$  and  $T_{\mathcal{U}^\circ}$ .

**Theorem 3.2** (Twin operator norm characterization for norming dual pairs). *A dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$  is called norming whenever*

$$\|u\|_{\mathcal{U}} = \sup_{\|u^\circ\|_{\mathcal{U}^\circ} \leq 1} |\langle u^\circ | u \rangle_{\mathcal{U}}| \quad (3.19a)$$

$$\|u^\circ\|_{\mathcal{U}^\circ} = \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle u^\circ | u \rangle_{\mathcal{U}}| \quad (3.19b)$$

Let  $(\mathcal{U}, \mathcal{U}^\circ)$  be a norming dual pair. If  $T \in \text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$ , then

$$\|T\|_{\text{Twin}(\mathcal{U}, \mathcal{U}^\circ)} = \|T_{\mathcal{U}}\|_{\mathcal{L}(\mathcal{U})} = \|T_{\mathcal{U}^\circ}\|_{\mathcal{L}(\mathcal{U}^\circ)} \quad (3.20)$$

where  $T_{\mathcal{U}} : \mathcal{U} \rightarrow \mathcal{U}$  and  $T_{\mathcal{U}^\circ} : \mathcal{U}^\circ \rightarrow \mathcal{U}^\circ$  are the bounded linear maps defined by  $T$  on  $\mathcal{U}$  and  $\mathcal{U}^\circ$ , respectively.

*Proof.* We only show  $\|T\|_{\text{Twin}(\mathcal{U}, \mathcal{U}^\circ)} = \|T_{\mathcal{U}}\|_{\mathcal{L}(\mathcal{U})}$  as  $\|T\|_{\text{Twin}(\mathcal{U}, \mathcal{U}^\circ)} = \|T_{\mathcal{U}^\circ}\|_{\mathcal{L}(\mathcal{U}^\circ)}$  is shown similarly.

$$\begin{aligned} \|T\|_{\text{Twin}(\mathcal{U}, \mathcal{U}^\circ)} &= \sup_{\|u\|_{\mathcal{U}} \leq 1, \|u^\circ\|_{\mathcal{U}^\circ} \leq 1} |T(u^\circ, u)| = \sup_{\|u\|_{\mathcal{U}} \leq 1, \|u^\circ\|_{\mathcal{U}^\circ} \leq 1} |\langle u^\circ | T_{\mathcal{U}} u \rangle_{\mathcal{U}}| \\ &= \sup_{\|u\|_{\mathcal{U}} \leq 1} \left( \sup_{\|u^\circ\|_{\mathcal{U}^\circ} \leq 1} |\langle u^\circ | T_{\mathcal{U}} u \rangle_{\mathcal{U}}| \right) = \sup_{\|u\|_{\mathcal{U}} \leq 1} \|T_{\mathcal{U}} u\|_{\mathcal{U}} = \|T_{\mathcal{U}}\|_{\mathcal{L}(\mathcal{U})}. \quad \square \end{aligned} \quad (3.21)$$

The kernel construction in Definition 3.6 differs from existing works on kernels for vv-RKBSs. Most of these works define a kernel on a symmetric domain  $X \times X$ . However, within this setting, the output space and the used assumptions vary considerably.

In Zhang et al. [57], the vv-RKBS  $\mathcal{B}$  and the output space  $\mathcal{U}$  are assumed to be uniformly convex and uniformly Fréchet differentiable, which in particular implies reflexivity. Under these assumptions, they introduce a compatible semi-inner product structure to construct a reproducing kernel  $K : X \times X \rightarrow B(\mathcal{U})$ . Here  $B(\mathcal{U})$  denotes the set of continuous, not necessarily linear, operators from  $\mathcal{U}$  to itself. Using this structure, they establish a general representer theorem for the adjoint element of the solution.

Lin et al. [31] and Chen et al. [17] relax the uniformity assumptions and work in non-reflexive  $\ell^1$  and group-lasso settings. However, they focus on  $\mathbb{R}^d$ -valued functions with kernels  $K : X \times X \rightarrow \mathbb{R}^{d \times d}$ . Similar to our approach, the reproducing property is achieved via two function spaces  $(\mathcal{B}, \mathcal{B}^\circ)$  linked by a bilinear form. Their representer theorem for the actual solution, however, requires a strong Lebesgue constant condition on the kernel as well as additional kernel admissibility conditions.

Combettes et al. [18] construct vv-RKBSs using a feature map approach and assume that both the output and feature spaces are separable real Banach spaces. Under reflexivity, strict convexity, and smoothness of the feature space, they show the existence of a reproducing kernel

$$K : X \times X \rightarrow B(\mathcal{U}^*, \mathcal{U}), \quad (3.22)$$

where  $B(\mathcal{U}^*, \mathcal{U})$  denotes operators (not necessarily linear) mapping bounded subsets of  $\mathcal{U}^*$  to bounded subsets of  $\mathcal{U}$ , and where the reproducing property additionally requires a duality mapping. They also prove a representer theorem that does not depend on the kernel structure itself, but still requires both the output and feature spaces to be separable and reflexive.

In contrast to the previously mentioned works, Wang et al. [51] go beyond the symmetric setting and, motivated by neural network architectures, define kernels  $K : X \times \Omega \rightarrow \mathbb{R}^d$ . Their construction, however, is specialized to finite-dimensional outputs, and the kernel's output does not directly correspond to a  $d \times d$  matrix as used in the vv-RKHS setting.

Compared to these existing approaches, our adjoint-pair definition treats both symmetric and asymmetric kernel domains, supports general Banach-valued outputs, and produces kernel outputs that generalize the bounded linear operators used in the vv-RKHS setting. Moreover, it avoids structural assumptions such as reflexivity and separability for the output and feature spaces. This will be crucial in Sections 4 and 5. There, we must work with the non-separable and non-reflexive space of vector measures as the feature space, as well as the non-separable and non-reflexive output spaces required for DeepONets and hypernetworks.

### 3.3 vv-RKBS Properties

As mentioned in the previous section, Definitions 3.4 and 3.6 do not fully align with the definitions based on the  $\delta$ -dual [55, 50] or those involving isometric embeddings into the dual space [56, 47]. The following theorem shows that these definitions become equivalent to Definitions 3.4 and 3.6 when additional conditions are imposed on the latter two. For clarity, we restrict the statement to the vector-valued case in Definition 3.6.

**Theorem 3.3.** *Let  $X, \Omega$  be sets,  $(\mathcal{U}, \mathcal{U}^\diamond)$  be a dual pair and let  $\mathcal{B}$  be a vv-RKBS containing functions mapping  $X$  to  $\mathcal{U}$ . Assume that  $\mathcal{B}^\diamond$  is a Banach space of functions mapping the set  $\Omega$  to  $\mathcal{U}^\diamond$ . Then the following statements are equivalent:*

1.  $\mathcal{B}^\diamond$  is isometrically embedded in  $\mathcal{B}^*$ .
2. There exists a duality pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}} : \mathcal{B}^\diamond \times \mathcal{B} \rightarrow \mathbb{R}$  and the norm  $\|\cdot\|_{\mathcal{B}^\diamond}$  satisfies  $\|g\|_{\mathcal{B}^\diamond} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}|$ .

Moreover, let  $\delta_{x, u^\diamond}(f) := \langle u^\diamond | f(x) \rangle_{\mathcal{U}}$  for  $f \in \mathcal{B}$ , let  $\Delta := \text{span}\{\delta_{x, u^\diamond} \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\}$ , and let the  $\delta$ -dual be defined as the closure  $\overline{\Delta}$  of  $\Delta$  under the  $\mathcal{B}^*$  norm. Then the following are also equivalent:

3.  $\Delta \subseteq \mathcal{B}^*$  and the  $\delta$ -dual  $\overline{\Delta} \subseteq \mathcal{B}^*$  is isometrically isomorphic to  $\mathcal{B}^\diamond$ .
4.
  - There exists a duality pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}} : \mathcal{B}^\diamond \times \mathcal{B} \rightarrow \mathbb{R}$ ,
  - the norm  $\|\cdot\|_{\mathcal{B}^\diamond}$  satisfies  $\|g\|_{\mathcal{B}^\diamond} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}|$ ,
  - there exists a map  $K_{\mathcal{U}^\diamond}$  that assigns to each  $(x, w) \in X \times \Omega$  a linear operator  $K_{\mathcal{U}^\diamond}(x, w) : \mathcal{U}^\diamond \rightarrow \mathcal{U}^\diamond$  such that for all  $u \in \mathcal{U}^\diamond$  and  $f \in \mathcal{B}$  the function  $K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \in \mathcal{B}^\diamond$  and  $\langle u^\diamond | f(x) \rangle_{\mathcal{U}} = \langle K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond | f \rangle_{\mathcal{B}}$ , and
  - the set  $\text{span}\{K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\}$  is dense in  $\mathcal{B}^\diamond$ .

*Proof.*

(1)  $\iff$  (2):

$\mathcal{B}^\diamond$  is isometrically embedded in  $\mathcal{B}^*$  if and only if there exists a mapping  $\iota : \mathcal{B}^\diamond \rightarrow \mathcal{B}^*$  such that  $\|g\|_{\mathcal{B}^\diamond} = \|\iota(g)\|_{\mathcal{B}^*} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle \iota(g) | f \rangle_{\mathcal{B}^*, \mathcal{B}}|$  with  $\langle \cdot | \cdot \rangle_{\mathcal{B}^*, \mathcal{B}}$  the canonical duality pairing on  $\mathcal{B}$ .

For (1)  $\Rightarrow$  (2), define the pairing between  $\mathcal{B}$  and  $\mathcal{B}^\diamond$  as  $\langle g | f \rangle_{\mathcal{B}} := \langle \iota(g) | f \rangle_{\mathcal{B}^*, \mathcal{B}}$ .

For (2)  $\Rightarrow$  (1), let  $\iota : \mathcal{B}^\diamond \rightarrow \mathcal{B}^*$  be defined via  $g \mapsto \langle g | \cdot \rangle_{\mathcal{B}} \in \mathcal{B}^*$ . Then

$$\|\iota(g)\|_{\mathcal{B}^*} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle \iota(g) | f \rangle_{\mathcal{B}^*, \mathcal{B}}| = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}| = \|g\|_{\mathcal{B}^\diamond}, \quad (3.23)$$

which shows that  $\mathcal{B}^\diamond$  is isometrically embedded in  $\mathcal{B}^*$ .

(3)  $\implies$  (4):

Let  $\iota : \mathcal{B}^\diamond \rightarrow \overline{\Delta}$  be the isometric isomorphism between the  $\delta$ -dual and  $\mathcal{B}^\diamond$ . Define the pairing  $\langle g | f \rangle_{\mathcal{B}} := \langle \iota(g) | f \rangle_{\mathcal{B}^*, \mathcal{B}}$ . Then, as in the proof of (1)  $\Rightarrow$  (2), we get  $\|g\|_{\mathcal{B}^\diamond} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}|$ .

Note that  $\iota^{-1}(\delta_{x, u^\diamond}) \in \mathcal{B}^\diamond$  for all  $(x, u^\diamond) \in X \times \mathcal{U}^\diamond$ . Since  $\delta_{x, u^\diamond + \lambda v^\diamond} = \delta_{x, u^\diamond} + \lambda \delta_{x, v^\diamond}$ , it follows that  $u^\diamond \mapsto \iota^{-1}(\delta_{x, u^\diamond}) \in \mathcal{B}^\diamond$  is linear. In particular, as  $\mathcal{B}^\diamond$  contains functions over  $\Omega$ , we have  $u^\diamond \mapsto \iota^{-1}(\delta_{x, u^\diamond})(w) \in \mathcal{U}^\diamond$  is linear for all  $w \in \Omega$ . Defining the linear operator  $K_{\mathcal{U}^\diamond}(x, w) := \iota^{-1}(\delta_{x, \cdot})(w)$  and using the definition of the pairing, we get:

$$\langle K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond | f \rangle_{\mathcal{B}} = \langle \iota^{-1}(\delta_{x, u^\diamond}) | f \rangle_{\mathcal{B}} = \langle \delta_{x, u^\diamond} | f \rangle_{\mathcal{B}^*, \mathcal{B}} = \langle u^\diamond | f(x) \rangle_{\mathcal{U}}. \quad (3.24)$$

This proves the reproducing property.

It remains to show that the set  $\{K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\}$  is dense in  $\mathcal{B}^\diamond$ . Since  $\iota(K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond) = \delta_{x, u^\diamond}$  and  $\iota$  is an isometry, taking closures with respect to the norms on both sides shows that the closure of  $\text{span}\{K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\}$  is isometrically isomorphic to  $\overline{\Delta}$ , which is isometrically isomorphic to  $\mathcal{B}^\diamond$ . This completes the proof.

(4)  $\implies$  (3):

Define the mapping  $\iota$  by  $\iota(K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond) := \delta_{x, u^\diamond}$ . Note that

$$\begin{aligned} \left\| \sum_{n=1}^N a_n K_{\mathcal{U}^\diamond}(x_n, \cdot)u_n^\diamond \right\|_{\mathcal{B}^\diamond} &= \sup_{\|f\|_{\mathcal{B}} \leq 1} \left| \left\langle \sum_{n=1}^N a_n K_{\mathcal{U}^\diamond}(x_n, \cdot)u_n^\diamond, f \right\rangle_{\mathcal{B}} \right| = \sup_{\|f\|_{\mathcal{B}} \leq 1} \left| \sum_{n=1}^N a_n \langle K_{\mathcal{U}^\diamond}(x_n, \cdot)u_n^\diamond, f \rangle_{\mathcal{B}} \right| \\ &= \sup_{\|f\|_{\mathcal{B}} \leq 1} \left| \sum_{n=1}^N a_n \delta_{x_n, u_n^\diamond}(f) \right| = \left\| \sum_{n=1}^N a_n \delta_{x_n, u_n^\diamond} \right\|_{\mathcal{B}^*}, \end{aligned} \quad (3.25)$$

for all  $a_n \in \mathbb{R}$ ,  $x_n \in X$  and  $u_n^\diamond \in \mathcal{U}^\diamond$  for any  $N \in \mathbb{N}$ . This shows that  $\iota: \text{span}\{K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\} \rightarrow \Delta$  is an isometric isomorphism. By the continuous extension theorem,  $\iota$  can be extended to an isometric isomorphism from  $\overline{\text{span}(K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond)}$  to  $\overline{\Delta}$ . Since  $\{K_{\mathcal{U}}(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\}$  is dense in  $\mathcal{B}^\diamond$ , it follows that  $\mathcal{B}^\diamond$  is isometrically isomorphic to  $\overline{\Delta}$ .  $\square$

The previous theorem demonstrates that using only a duality pairing is the weakest assumption one can make. This raises the question whether properties of the vv-RKHS extend to the adjoint pair of vv-RKBS under this minimal formulation, or whether additional assumptions, such as those in Theorem 3.3, are required.

An example where additional assumptions are necessary relates to the fact that, in the vv-RKHS setting, the set  $\{K(x, \cdot)u \mid (x, u) \in X \times \mathcal{U}\}$  is dense in the vv-RKHS. As the equivalence between (3) and (4) in Theorem 3.3 already illustrates, further assumptions are needed in the vv-RKBS setting to obtain a similar density property. Moreover, the theorem only establishes density in  $\mathcal{B}^\diamond$ , not in  $\mathcal{B}$ . In Wang et al. [50], this issue is also addressed; they show that, in the scalar case, assuming  $\mathcal{B}^\diamond$  is a pre-dual space leads to a corresponding density result in  $\mathcal{B}$ .

Another result that holds in the vv-RKHS setting and whose analogue we can ask for in the vv-RKBS setting is the existence of a kernel. In vv-RKHSs a reproducing kernel is always present, but in Definitions 3.1 and 3.2 no kernel arises directly, since the Riesz representation theorem is unavailable. It may therefore seem questionable to refer to these spaces as reproducing kernel spaces. Nevertheless, we show that every vv-RKBS as in these definitions is part of an adjoint pair of vv-RKBSs (Definition 3.6), which ensures the existence of a reproducing kernel. Moreover, the next theorem demonstrates that one can always choose an adjoint vv-RKBS satisfying the strongest assumptions of Theorem 3.3, in particular those in point 4. Thus, while these stronger assumptions are not strictly necessary, they can always be met in a suitable adjoint pair.

**Theorem 3.4** (Every vv-RKBS corresponds to an adjoint vv-RKBS pair). *Let  $\mathcal{B}$  be a vv-RKBS with output Banach space  $\mathcal{U}$ . Define  $\delta_{x, u^*}(f) := \langle u^* | f(x) \rangle_{\mathcal{U}}$  for  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$  the canonical duality pairing of the pair  $(\mathcal{U}, \mathcal{U}^*)$  with  $\mathcal{U}^*$  the continuous dual. Denote  $\Delta := \text{span}\{\delta_{x, u^*} \mid (x, u^*) \in X \times \mathcal{U}^*\}$ , and let the  $\delta$ -dual be defined as the closure  $\overline{\Delta}$  of  $\Delta$  under the  $\mathcal{B}^*$  norm. Define the Banach space*

$$\mathcal{B}^\diamond := \{g: \mathcal{B} \times \mathcal{U}^* \rightarrow \mathcal{U}^* \mid g(f, \tilde{u}^*) = g_*(f)\tilde{u}^* \text{ for some } g_* \in \overline{\Delta}\} \quad (3.26)$$

with  $\|g\|_{\mathcal{B}^\diamond} := \|g_*\|_{\mathcal{B}^*}$ . Then  $(\mathcal{B}, \mathcal{B}^\diamond)$  forms an adjoint pair of vv-RKBSs with duality pairing  $\langle g | f \rangle_{\mathcal{B}} := g_*(f)$  and kernel

$$K_{\mathcal{U}}(x, (f, \tilde{u}^*))u := f(x) \langle \tilde{u}^* | u \rangle_{\mathcal{U}}, \quad K_{\mathcal{U}^*}(x, (f, \tilde{u}^*))u^* := \langle u^* | f(x) \rangle_{\mathcal{U}} \tilde{u}^*, \quad (3.27)$$

$$K(x, (f, \tilde{u}^*))(u^*, u) = \langle u^* | f(x) \rangle_{\mathcal{U}} \langle \tilde{u}^* | u \rangle_{\mathcal{U}} = \delta_{x, u^*}(f) \langle \tilde{u}^* | u \rangle_{\mathcal{U}}.$$

Moreover, by definition,

$$\|g\|_{\mathcal{B}^\diamond} = \|g_*\|_{\mathcal{B}^*} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |g_*(f)| = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}|, \quad (3.28)$$

and the set

$$\text{span}\{K_{\mathcal{U}^*}(x, \cdot)\tilde{u}^* \mid (x, \tilde{u}^*) \in X \times \mathcal{U}^*\} = \text{span}\{g(f, \tilde{u}^*) := \delta_{x, u^*}(f)\tilde{u}^* \mid (x, u^*) \in X \times \mathcal{U}^*, (f, \tilde{u}^*) \in \mathcal{B} \times \mathcal{U}^*\} \quad (3.29)$$

is dense in  $\mathcal{B}^\diamond$ .

*Proof.* Note that, by definition,

$$\|g(f, \tilde{u}^*)\|_{\mathcal{U}^*} = \|g_*(f)\tilde{u}^*\|_{\mathcal{U}^*} = \|\tilde{u}^*\|_{\mathcal{U}^*} |g_*(f)| \leq \|\tilde{u}^*\|_{\mathcal{U}^*} \|f\|_{\mathcal{B}} \|g_*\|_{\mathcal{B}^*} = C_{f, \tilde{u}^*} \|g\|_{\mathcal{B}^\diamond} \quad (3.30)$$

where  $C_{f, \tilde{u}^*} = \|\tilde{u}^*\|_{\mathcal{U}^*} \|f\|_{\mathcal{B}}$ . This shows that  $\mathcal{B}^\diamond$  is a vv-RKBS.

Furthermore, assume  $\langle g|f \rangle_{\mathcal{B}} = 0$  for all  $g \in \mathcal{B}^\circ$ . Then  $\langle \delta_{x,u^*}|f \rangle_{\mathcal{B}} = \langle u^*|f(x) \rangle_{\mathcal{U}} = 0$  for all  $(x, u^*) \in X \times \mathcal{U}^*$ . This implies  $f(x) = 0$  for all  $x \in X$ , and hence  $f = 0$ . Conversely, if  $\langle g|f \rangle_{\mathcal{B}} = g_*(f) = 0$  for all  $f \in \mathcal{B}$ , then by definition  $g = 0$ . Combining these two facts shows that  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$  is a duality pairing.

We now check the reproducing properties. For  $f \in \mathcal{B}$ , the result immediately follows from the definition of the pairing

$$\langle K_{\mathcal{U}^*}(x, (\cdot, \cdot))u^*|f \rangle_{\mathcal{B}} = \delta_{x,u^*}(f) = \langle u^*|f(x) \rangle_{\mathcal{U}} \quad (3.31)$$

Similarly, for  $g \in \mathcal{B}^\circ$

$$\langle g|K_{\mathcal{U}}(\cdot, (f, \tilde{u}^*))u \rangle_{\mathcal{B}} = g_*(f) \langle \tilde{u}^*|u \rangle_{\mathcal{U}} = \langle g_*(f)\tilde{u}^*|u \rangle_{\mathcal{U}} = \langle g(f, \tilde{u}^*)|u \rangle_{\mathcal{U}} \quad (3.32)$$

Finally, the formula for  $K(x, (f, \tilde{u}^*))(u^*, u)$  follows by combining the reproducing properties of elements of  $\mathcal{B}^\circ$  with the specific form of  $K_{\mathcal{U}}$

$$\begin{aligned} K(x, (f, \tilde{u}^*))(u^*, u) &= \langle K_{\mathcal{U}^*}(x, (\cdot, \cdot))u^*|K_{\mathcal{U}}(\cdot, (f, \tilde{u}^*))u \rangle_{\mathcal{B}} \\ &= \langle \delta_{x,u^*}(f)\tilde{u}^*|u \rangle_{\mathcal{U}} \\ &= \delta_{x,u^*}(f) \langle \tilde{u}^*|u \rangle_{\mathcal{U}} \\ &= \langle u^*|f(x) \rangle_{\mathcal{U}} \langle \tilde{u}^*|u \rangle_{\mathcal{U}}. \end{aligned} \quad (3.33) \quad \square$$

Two other properties of the vv-RKHS are the uniqueness of the kernel and the fact that every symmetric positive semi-definite kernel defines a vv-RKHS. The next theorem shows that, under the weakest definition of an adjoint vv-RKBS pair, we still obtain (1) a unique kernel that satisfies a relation analogous to (3.9), (2) that every kernel satisfies a certain bound, and (3) that every kernel satisfying such a bound defines an adjoint vv-RKBS pair. However, to obtain a vv-RKBS with specific properties, additional assumptions are required. In particular, the next theorem also provides an example of additional conditions that yield an adjoint vv-RKBS pair satisfying  $\|g\|_{\mathcal{B}^\circ} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g|f \rangle_{\mathcal{B}}|$ .

**Theorem 3.5** (vv-RKBS kernel properties and kernels defining vv-RKBSs). *Let  $(\mathcal{U}, \mathcal{U}^\circ)$  be a dual pair. If  $(\mathcal{B}, \mathcal{B}^\circ)$  is an adjoint vv-RKBS pair with reproducing kernel  $K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$ , then the kernel  $K$  is unique, satisfies the equalities*

$$K(x, w)(u^\circ, u) = \langle K_{\mathcal{U}^\circ}(x, \cdot)u^\circ|K_{\mathcal{U}}(\cdot, w)u \rangle_{\mathcal{B}} = \begin{cases} \langle K_{\mathcal{U}^\circ}(x, w)u^\circ|u \rangle_{\mathcal{U}} \\ \langle u^\circ|K_{\mathcal{U}}(x, w)u \rangle_{\mathcal{U}}, \end{cases} \quad (3.34)$$

and the bounds

$$\|K_{\mathcal{U}}(x, w)u\|_{\mathcal{U}} \leq C_X(x)\|K_{\mathcal{U}}(\cdot, w)u\|_{\mathcal{B}} =: C_X(x)C_{\Omega \times \mathcal{U}}(w, u), \text{ and} \quad (3.35a)$$

$$\|K_{\mathcal{U}^\circ}(x, w)u^\circ\|_{\mathcal{U}} \leq C_\Omega(w)\|K_{\mathcal{U}^\circ}(x, \cdot)u^\circ\|_{\mathcal{B}^\circ} =: C_\Omega(w)C_{X \times \mathcal{U}^\circ}(x, u^\circ) \quad (3.35b)$$

for some functions  $C_X: X \rightarrow \mathbb{R}^+$  and  $C_\Omega: \Omega \rightarrow \mathbb{R}^+$ .

Conversely, if we are given a continuous dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$ , sets  $X, \Omega$ , and a map  $K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$  that satisfies

$$\|K_{\mathcal{U}}(x, w)u\|_{\mathcal{U}} \leq C_X(x)C_{\Omega \times \mathcal{U}}(w, u) \quad \text{and} \quad (3.36a)$$

$$\|K_{\mathcal{U}^\circ}(x, w)u^\circ\|_{\mathcal{U}} \leq C_\Omega(w)C_{X \times \mathcal{U}^\circ}(x, u^\circ) \quad (3.36b)$$

for some functions  $C_X: X \rightarrow \mathbb{R}^+$ ,  $C_\Omega: \Omega \rightarrow \mathbb{R}^+$ ,  $C_{\Omega \times \mathcal{U}}: \Omega \times \mathcal{U} \rightarrow \mathbb{R}^+$ , and  $C_{X \times \mathcal{U}^\circ}: X \times \mathcal{U}^\circ \rightarrow \mathbb{R}^+$ , then there exists a vv-RKBS pair  $(\mathcal{B}, \mathcal{B}^\circ)$  with domains  $X$  and  $\Omega$ , respectively, and  $K$  as reproducing kernel. Moreover, when additionally assuming

$$\|u^\circ\|_{\mathcal{U}^\circ} = \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle u^\circ|u \rangle_{\mathcal{U}}| \quad \text{and} \quad (3.37a)$$

$$\|K_{\mathcal{U}}(x, w)u\|_{\mathcal{U}} \leq C_X(x)C_\Omega(w)\|u\|_{\mathcal{U}}, \quad (3.37b)$$

one can define a vv-RKBS pair  $(\mathcal{B}, \mathcal{B}^\circ)$  satisfying  $\|g\|_{\mathcal{B}^\circ} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g|f \rangle_{\mathcal{B}}|$ .

*Proof.* Let  $(\mathcal{B}, \mathcal{B}^\circ)$  be an adjoint vv-RKBS pair. By first using the definition of a twin operator and subsequently the reproducing property, we obtain (3.34). Now, assume there is another kernel  $\tilde{K}$ . Again by the reproducing property:

$$0 = \langle g(w)|u \rangle - \langle g(w)|u \rangle = \left\langle g \left| \left( K_{\mathcal{U}}(x, w) - \tilde{K}_{\mathcal{U}}(x, w) \right) u \right. \right\rangle_{\mathcal{B}} \quad (3.38)$$

holds for all  $g \in \mathcal{B}^\circ$ . When  $g = K_{\mathcal{U}^\circ}(x, \cdot)u^\circ \in \mathcal{B}^\circ$  and we use the reproducing property  $\langle u^\circ | f(x) \rangle_{\mathcal{U}} = \langle K_{\mathcal{U}^\circ}(x, \cdot)u^\circ | f \rangle_{\mathcal{B}}$ , we get

$$0 = \left\langle K_{\mathcal{U}^\circ}(x, \cdot)u^\circ \left| \left( K_{\mathcal{U}}(\cdot, w) - \tilde{K}_{\mathcal{U}}(\cdot, w) \right) u \right\rangle_{\mathcal{B}} = \left\langle u^\circ \left| \left( K_{\mathcal{U}}(x, w) - \tilde{K}_{\mathcal{U}}(x, w) \right) u \right\rangle_{\mathcal{U}} \quad (3.39)$$

for all  $(x, w, u, u^\circ) \in X \times \Omega \times \mathcal{U} \times \mathcal{U}^\circ$ . By the non-degeneracy condition (3.3a) of the dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$ , we conclude  $K_{\mathcal{U}}(x, w) = \tilde{K}_{\mathcal{U}}(x, w)$ . Employing the same reasoning for  $K_{\mathcal{U}^\circ}$  shows that  $K_{\mathcal{U}^\circ} = \tilde{K}_{\mathcal{U}^\circ}$ . Hence, both  $K_{\mathcal{U}}$  and  $K_{\mathcal{U}^\circ}$  are unique. Thus, by (3.34), the full kernel is unique.

To show the converse, we follow the approach in Heeringa et al. [24]. For a kernel  $K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$  satisfying (3.36a) and (3.36b) define the vector spaces:

$$\mathcal{V} := \text{span}\{K_{\mathcal{U}}(\cdot, w)u : X \rightarrow \mathcal{U} \mid (w, u) \in \Omega \times \mathcal{U}\}, \quad (3.40a)$$

$$\mathcal{V}^\circ := \text{span}\{K_{\mathcal{U}^\circ}(x, \cdot)u^\circ : \Omega \rightarrow \mathcal{U}^\circ \mid (x, u^\circ) \in X \times \mathcal{U}^\circ\} \quad (3.40b)$$

and the bilinear mapping

$$\left\langle \sum_j K_{\mathcal{U}^\circ}(x_j, \cdot)u_j^\circ \left| \sum_i K_{\mathcal{U}}(\cdot, w_i)u_i \right\rangle_{\mathcal{V}} := \sum_{i,j} K(x_j, w_i)(u_j^\circ, u_i) \quad (3.41)$$

between them. Equip  $\mathcal{V}$  with the norm

$$\|f\|_{\mathcal{V}} := \sup_{x \in X} \frac{\|f(x)\|_{\mathcal{U}}}{C_X(x)}. \quad (3.42)$$

This norm is well-defined since

$$\|f(x)\|_{\mathcal{U}} = \left\| \sum_i K_{\mathcal{U}}(x, w_i)u_i \right\|_{\mathcal{U}} \leq \sum_i \|K_{\mathcal{U}}(x, w_i)u_i\|_{\mathcal{U}} \leq C_X(x) \sum_i C_{\Omega \times \mathcal{U}}(w_i, u_i). \quad (3.43)$$

By definition, we also have

$$\|f(x)\|_{\mathcal{U}} \leq C_X(x) \|f\|_{\mathcal{V}} \quad (3.44)$$

for all  $f \in \mathcal{V}$ . Let  $\mathcal{B}$  be the completion of  $\mathcal{V}$  with respect to this norm. This makes  $\mathcal{B}$  a Banach space of functions, and hence, by the above inequality, a vv-RKBS.

For  $\mathcal{V}^\circ$ , define the norm as

$$\|g\|_{\mathcal{V}^\circ} := \max \left( \sup_{w \in \Omega} \frac{\|g(w)\|_{\mathcal{U}^\circ}}{C_{\Omega}(w)}, \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{V}^\circ, \mathcal{B}}| \right), \quad (3.45)$$

where the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{V}^\circ, \mathcal{B}}$  is the extension of the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{V}}$  from  $\mathcal{V}^\circ \times \mathcal{V}$  to  $\mathcal{V}^\circ \times \mathcal{B}$ . This ensures

$$\|g(w)\|_{\mathcal{U}^\circ} \leq C_{\Omega}(w) \|g\|_{\mathcal{V}^\circ}, \quad |\langle g | f \rangle_{\mathcal{B}}| \leq \|g\|_{\mathcal{V}^\circ} \|f\|_{\mathcal{B}}. \quad (3.46)$$

Let  $\mathcal{B}^\circ$  be the completion of  $\mathcal{V}^\circ$  with respect to  $\|\cdot\|_{\mathcal{V}^\circ}$ . Extending the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{V}}$  continuously via the density of  $\mathcal{V}, \mathcal{V}^\circ$  in the corresponding completions, yields a pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$  satisfying

$$|\langle g | f \rangle_{\mathcal{B}}| \leq \|g\|_{\mathcal{B}^\circ} \|f\|_{\mathcal{B}}. \quad (3.47)$$

To show that  $(\mathcal{B}, \mathcal{B}^\circ)$  is a dual pair, we verify non-degeneracy. First, note that the reproducing property on  $\mathcal{V}^\circ$  still holds on  $\mathcal{B}^\circ$  due to continuity of the duality pairing on  $(\mathcal{U}, \mathcal{U}^\circ)$ . Then, observe that

$$\forall f \in \mathcal{B} \setminus \{0\}: \langle g | f \rangle_{\mathcal{B}} = 0 \implies \forall (w, u) \in \Omega \times \mathcal{U}: \langle g(w) | u \rangle_{\mathcal{U}} = \langle g | K_{\mathcal{U}}(\cdot, w)u \rangle_{\mathcal{B}} = 0 \implies g = 0, \quad (3.48)$$

where the final implication follows from the non-degeneracy of the dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$ , i.e.,  $\langle g(w) | u \rangle_{\mathcal{U}} = 0$  for all  $u \in \mathcal{U}$  implies  $g(w) = 0$ . A similar argument establishes non-degeneracy condition (3.3a). Hence, we obtain a continuous dual pairing  $(\mathcal{B}, \mathcal{B}^\circ)$ , with  $\mathcal{B}$  and  $\mathcal{B}^\circ$  both vv-RKBSs. Thus, they form an adjoint pair of vv-RKBSs.

Now, suppose that (3.37a) and (3.37b) hold. Equip  $\mathcal{V}^\circ$  with the induced dual-norm

$$\|g\|_{\mathcal{V}^\circ} := \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}|. \quad (3.49)$$

Using  $\|g(w)\|_{\mathcal{U}^\circ} = \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle g(w)|u \rangle_{\mathcal{U}}|$  and the definition of the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$ , we get

$$\|g(w)\|_{\mathcal{U}^\circ} = \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle g(w)|u \rangle_{\mathcal{U}}| = \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle g|K_{\mathcal{U}}(\cdot, w)u \rangle_{\mathcal{B}}| \leq C_{\Omega}(w) \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g|f \rangle_{\mathcal{B}}| = C_{\Omega}(w) \|g\|_{\mathcal{V}^\circ}, \quad (3.50)$$

where the last inequality follows from

$$\|K_{\mathcal{U}}(\cdot, w)u\|_{\mathcal{B}} = \sup_{x \in X} \frac{\|K_{\mathcal{U}}(x, w)u\|_{\mathcal{U}}}{C_X(x)} \leq \sup_{x \in X} \frac{C_X(x)C_{\Omega}(w)\|u\|_{\mathcal{U}}}{C_X(x)} = C_{\Omega}(w)\|u\|_{\mathcal{U}} \leq C_{\Omega}(w). \quad (3.51)$$

Following the same reasoning as above, taking completions yields a vv-RKBS, a continuous duality pairing, and thus, a vv-RKBS pair with the required property.  $\square$

Finally, for a vv-RKHS, it is known that it is isomorphic [15, Proposition 2.7] to a scalar RKHS. Below, we show that this remains true even without any additional assumptions.

**Theorem 3.6** (Equivalence scalar RKBS (pair) and vv-RKBS (pair)). *Let  $(\mathcal{B}, \mathcal{B}^\circ)$  be a vv-RKBS dual pair over  $X, \Omega$  to a dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$  and with kernel  $K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$ . Define the spaces  $\mathcal{B}_s$  and  $\mathcal{B}_s^\circ$  as*

$$\mathcal{B}_s := \{\tilde{f}: X \times \mathcal{U}^\circ \rightarrow \mathbb{R} \mid \tilde{f}(x, u^\circ) = \langle u^\circ | f(x) \rangle_{\mathcal{B}}, f \in \mathcal{B}\}, \quad \|\tilde{f}\|_{\mathcal{B}_s} := \|f\|_{\mathcal{B}}, \quad (3.52a)$$

$$\mathcal{B}_s^\circ := \{\tilde{g}: \Omega \times \mathcal{U} \rightarrow \mathbb{R} \mid \tilde{g}(w, u) = \langle g(w) | u \rangle_{\mathcal{B}}, g \in \mathcal{B}^\circ\}, \quad \|\tilde{g}\|_{\mathcal{B}_s^\circ} := \|g\|_{\mathcal{B}^\circ} \quad (3.52b)$$

the pairing between them as

$$\langle \tilde{f} | \tilde{g} \rangle_{\mathcal{B}_s} := \langle f | g \rangle_{\mathcal{B}} \quad (3.53)$$

and the scalar kernel as

$$K_s: (X \times \mathcal{U}^\circ) \times (\Omega \times \mathcal{U}) \rightarrow \mathbb{R}, ((x, u^\circ), (w, u)) \mapsto K(x, w)(u^\circ, u) \quad (3.54)$$

Then  $(\mathcal{B}_s, \mathcal{B}_s^\circ)$  is an adjoint pair of RKBSs with kernel  $K_s$ ,  $\mathcal{B}_s \cong \mathcal{B}$ , and  $\mathcal{B}_s^\circ \cong \mathcal{B}^\circ$ .

*Proof.* First, we show  $\mathcal{B}_s$  is isometrically isomorphic to  $\mathcal{B}$ , where the isometric isomorphism between  $\mathcal{B}_s^\circ$  and  $\mathcal{B}^\circ$  follows from a similar argument and is hence omitted. By definition,  $f \rightarrow \tilde{f} := \langle \cdot | f(\cdot) \rangle_{\mathcal{U}}$  is a linear surjective map from  $\mathcal{B}$  to  $\mathcal{B}_s$ . To show injectivity (and hence bijectivity), assume we have  $f_1$  and  $f_2$  such that  $\langle u^\circ | f_1(x) \rangle_{\mathcal{U}} = \langle u^\circ | f_2(x) \rangle_{\mathcal{U}}$  for all  $(x, u^\circ) \in X \times \mathcal{U}^\circ$ . Hence, for a given  $x \in X$ ,  $\langle u^\circ | f_1(x) - f_2(x) \rangle_{\mathcal{U}} = 0$  for all  $u^\circ \in \mathcal{U}^\circ$ . By the nondegeneracy condition (3.3a) of the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$ , we get  $f_1(x) = f_2(x)$  for every  $x$ . As we are dealing with a Banach space of functions, this shows that  $f_1 = f_2$  and therefore shows injectivity. Finally, the isometric property follows from  $\|\tilde{f}\|_{\mathcal{B}_s} := \|f\|_{\mathcal{B}}$ .

What remains is to show that  $K_s$  is the kernel of the dual pair  $(\mathcal{B}_s, \mathcal{B}_s^\circ)$ . The combination of (3.34) and (3.54) implies that

$$K_s((\cdot, \bullet), (w, u)) = \langle \bullet | K_{\mathcal{U}}(\cdot, w)u \rangle_{\mathcal{U}} \in \mathcal{B}_s \quad (3.55a)$$

$$K_s((x, u^\circ), (\cdot, \bullet)) = \langle K_{\mathcal{U}^\circ}(x, \cdot)u^\circ | \bullet \rangle_{\mathcal{U}} \in \mathcal{B}_s^\circ \quad (3.55b)$$

and thus

$$\tilde{f}(x, u^\circ) = \langle u^\circ | f(x) \rangle_{\mathcal{U}} = \langle K_{\mathcal{U}^\circ}(x, \cdot)u^\circ | f \rangle_{\mathcal{B}} = \left\langle K_s((x, u^\circ), (\cdot, \bullet)) | \tilde{f} \right\rangle_{\mathcal{B}_s} \quad (3.56a)$$

$$\tilde{g}(w, u) = \langle g(w) | u \rangle_{\mathcal{U}} = \langle g | K_{\mathcal{U}}(\cdot, w)u \rangle_{\mathcal{B}} = \langle \tilde{g} | K_s((\cdot, \bullet), (w, u)) \rangle_{\mathcal{B}_s} \quad (3.56b)$$

where the second equality corresponds to the reproducing kernel property of the vv-RKBS pair  $(\mathcal{B}, \mathcal{B}^\circ)$  and the final equality follows from  $\langle \tilde{f} | \tilde{g} \rangle_{\mathcal{B}_s} := \langle f | g \rangle_{\mathcal{B}}$  and (3.55).  $\square$

## 4 Integral and neural RKBS

While the previous sections discussed vector-valued RKBSs in general, they did not yet address their connection to neural networks. In the scalar-valued case, the relevant RKBS is the scalar integral RKBS, particularly the subclass known as neural RKBSs. We begin by reviewing scalar integral and neural RKBSs, then extend the discussion to the vector-valued setting, and finally examine key properties of the vector-valued integral and neural RKBSs.

## 4.1 Scalar-valued integral and neural RKBS

The integral RKBS is defined via the feature map construction in Definition 3.2. In this setting, functions are constructed by integrating a feature function  $\phi \in C_0(X \times \Omega)$  against a Radon measure  $\mu \in \mathcal{M}(\Omega)$  or  $\rho \in \mathcal{M}(X)$ .

**Definition 4.1** (Scalar-valued integral RKBS). Given locally-compact Hausdorff  $X, \Omega$  and  $\phi \in C_0(X \times \Omega)$ , for  $\mu \in \mathcal{M}(\Omega)$  and  $\rho \in \mathcal{M}(X)$  define

$$f(x) = (A_{\Omega \rightarrow X}\mu)(x) := \int_{\Omega} \phi(x, w) d\mu(w), \quad (4.1)$$

$$g(w) = (A_{X \rightarrow \Omega}\rho)(w) := \int_X \phi(x, w) d\rho(x). \quad (4.2)$$

The integral RKBS is defined as an adjoint pair of scalar-valued RKBS  $(\mathcal{B}, \mathcal{B}^\circ)$

$$\mathcal{B} := \{f = A_{\Omega \rightarrow X}\mu \mid \mu \in \mathcal{M}(\Omega)\}, \quad \|f\|_{\mathcal{B}} := \inf_{f=A_{\Omega \rightarrow X}\mu} |\mu|(\Omega), \quad (4.3a)$$

$$\mathcal{B}^\circ := \{g = A_{X \rightarrow \Omega}\rho \mid \rho \in \mathcal{M}(X)\}, \quad \|g\|_{\mathcal{B}^\circ} := \sup_{w \in \Omega} |g(w)|, \quad (4.3b)$$

where, for every  $f \in \mathcal{B}$ ,  $g \in \mathcal{B}^\circ$ , and for any representative  $\mu \in \mathcal{M}(\Omega)$  of  $f$  and  $\rho \in \mathcal{M}(X)$  of  $g$ , the pairing is defined as

$$\langle g|f \rangle_{\mathcal{B}} = \langle \rho|f \rangle_{C_0(X)} = \langle \mu|g \rangle_{C_0(\Omega)} = \langle \mu \otimes \rho|\phi \rangle_{C_0(X \times \Omega)} = \int_{X \times \Omega} \phi(x, w) d(\rho \otimes \mu)(x, w), \quad (4.4)$$

where  $\rho \otimes \mu$  denotes the product measure. The kernel  $K: X \times \Omega \rightarrow \mathbb{R}$  is

$$K(x, w) = \phi(x, w) = \langle A_{X \rightarrow \Omega}\delta_x | A_{\Omega \rightarrow X}\delta_w \rangle_{\mathcal{B}} \quad (4.5)$$

In Spek et al. [47], it is proven that the pairing is well-defined, that  $(\mathcal{B}, \mathcal{B}^\circ)$  indeed forms an adjoint pair of RKBSs with kernel  $K$ , and that  $(\mathcal{B}^\circ)^*$  is isometrically isomorphic to  $\mathcal{B}$ ; that is,  $\mathcal{B}^\circ$  is a predual of  $\mathcal{B}$ . In particular, note

$$|\langle g|f \rangle_{\mathcal{B}}| = |\langle \mu|g \rangle_{C_0(\Omega)}| = \left| \int_{\Omega} g(w) d\mu(w) \right| \leq \left( \sup_{w \in \Omega} |g(w)| \right) |\mu|(\Omega) \implies \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g|f \rangle_{\mathcal{B}}| \leq \sup_{w \in \Omega} |g(w)| \quad (4.6)$$

where the last inequality follows by taking the infimum over  $\mu$  such that  $f = A_{\Omega \rightarrow X}\mu$ . Using this, and noting that  $g \in C_0(\Omega)$  attains its maximum at some  $w = w_{\max}$ , we find

$$\|g\|_{\mathcal{B}^\circ} = \sup_{w \in \Omega} |g(w)| = \left| \int_X \phi(x, w_{\max}) d\rho(x) \right| = |\langle \delta_{w_{\max}} | g \rangle_{C_0(\Omega)}| \leq \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g|f \rangle_{\mathcal{B}}| \leq \sup_{w \in \Omega} |g(w)| = \|g\|_{\mathcal{B}^\circ}. \quad (4.7)$$

Hence,  $\|g\|_{\mathcal{B}^\circ} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g|f \rangle_{\mathcal{B}}|$ . This shows that the integral RKBS has additional structure beyond that described in Definition 3.4. In particular, it corresponds to point (4) in Theorem 3.3, since it can also be shown that the  $K(x, \cdot) = \phi(x, \cdot)$  functions are dense in  $\mathcal{B}^\circ$ ; see Theorem 4.3 for a proof in the vector-valued setting.

Although the integral RKBS formulation in Definition 4.1 is closely related to neural networks, it does not inherently exhibit the characteristic structure of an affine transformation followed by a nonlinear activation. This structure can be recovered by selecting a specific form of  $\phi$  within the integral RKBS framework.

**Definition 4.2** (Scalar-valued neural RKBS). Let  $(\mathcal{V}, \mathcal{V}^\circ)$  be a dual pair of normed vector spaces, and assume  $X \subseteq \mathcal{V}$  is compact and  $\tilde{\Omega} \subseteq \mathcal{V}^\circ$  locally-compact Hausdorff. Let  $\Omega := \tilde{\Omega} \times \mathbb{R}$ , let  $\sigma: \mathbb{R} \rightarrow \mathbb{R}$  be a measurable activation function, and let  $\beta: \Omega \rightarrow \mathbb{R}$  be a measurable positive function such that  $\phi \in C_0(X \times \Omega)$ , where

$$\phi(x, w) = \phi(x, (\omega, b)) := \sigma(\langle \omega|x \rangle_{\mathcal{V}} + b) \beta(w) \quad (4.8)$$

with  $w = (\omega, b) \in \Omega$ .

If  $(\mathcal{B}, \mathcal{B}^\circ)$  is an integral RKBS pair with kernel  $\phi$ , then we refer to it as a neural RKBS pair.  $\diamond$

**Remark 4.1.** The function  $\beta$  in the above definition ensures that the kernel integrals in Equations (4.1) and (4.2) are well defined for commonly used activation functions. One could avoid introducing the weighting function  $\beta$  by modifying the definition of the integral RKBS. For example, we can restrict to measures for which the integrals are already well-defined. Another approach is taken in Barron spaces, a well-known example of a neural RKBS that also treats neural networks but incorporates  $\beta$  into the total variation norm of the measures instead [35, 34, 47]. Yet another option is to consider subspaces of measures satisfying moment conditions, which are dual to spaces of continuous functions with controlled growth [5].

Although such alternatives exist, we retain the  $C_0(X \times \Omega)$  assumption, as it simplifies the analysis without introducing significant limitations. In practice, exceedingly large weights are rarely desirable, so applying a continuous cut-off to zero for such values is both reasonable and effective.  $\diamond$

## 4.2 Vector-valued integral and neural RKBS

To extend the previous section from scalar outputs to outputs in a Banach space  $\mathcal{U}$ , two natural approaches would be to either make  $\phi$  a  $\mathcal{U}$ -valued function, or to replace the scalar-valued measures with  $\mathcal{U}$ -valued ones. We follow the second approach, as also done in Bartolucci et al. [7]. With the first option, it is challenging to ensure that the represented functions are expressive enough to take any value in a high- or even infinite-dimensional space  $\mathcal{U}$  at each domain point. It would require us to work explicitly with parameter spaces  $\Omega$  of comparable dimension and with measures defined on them. In contrast, by working with vector-valued measures, we can directly use any scalar-valued  $\phi$  while maintaining expressivity in terms of the attained values in  $\mathcal{U}$ . This approach generalizes the standard vv-RKHS construction starting from a scalar kernel of the form (2.9). Moreover, it allows us to work with spaces of measures on locally-compact parameter domains, which behave significantly better in terms of duality.

For a Banach space  $\mathcal{U}$  and  $\Omega$  locally-compact Hausdorff, we define  $\mathcal{M}(\Omega; \mathcal{U})$  to be the set of regular countably additive  $\mathcal{U}$ -valued vector measures over  $\Omega$  equipped with the Borel  $\sigma$ -algebra and with finite total variation. In this setting, a vector-valued generalization [37, Theorem 1] of the Riesz representation theorem provides an isometric isomorphism

$$C_0(\Omega; \mathcal{U})^* \cong \mathcal{M}(\Omega; \mathcal{U}^*), \quad (4.9)$$

where the pairing is an integral-like operation that is constructed via uniform approximation by simple functions, and is defined later in the paper. In the special case when  $\Omega$  is compact Hausdorff, this duality result is known as Singer's representation theorem [45]. In the context of Hilbert space-valued measures on second countable locally-compact Hausdorff domains, it also appears as Theorem 9 of Carmeli et al. [16].

**Definition 4.3** (Vector-valued integral and neural RKBS). Given locally-compact Hausdorff  $X, \Omega$ ,  $\phi \in C_0(X \times \Omega)$ , and a continuous dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$  of Banach spaces, for  $\mathcal{U}$ -valued measure  $\mu \in \mathcal{M}(\Omega; \mathcal{U})$  and  $\mathcal{U}^\circ$ -valued measure  $\rho \in \mathcal{M}(X; \mathcal{U}^\circ)$  define

$$f(x) = (A_{\Omega \rightarrow X} \mu)(x) := \int_{\Omega} \phi(x, w) d\mu(w), \quad (4.10)$$

$$g(w) = (A_{X \rightarrow \Omega} \rho)(w) := \int_X \phi(x, w) d\rho(x). \quad (4.11)$$

The integral vv-RKBS is an adjoint pair of vv-RKBS  $(\mathcal{B}, \mathcal{B}^\circ)$  defined as

$$\mathcal{B} := \{f = A_{\Omega \rightarrow X} \mu \mid \mu \in \mathcal{M}(\Omega; \mathcal{U})\}, \quad \|f\|_{\mathcal{B}} := \inf_{f = A_{\Omega \rightarrow X} \mu} \|\mu\|_{\mathcal{U}(\Omega)}, \quad (4.12a)$$

$$\mathcal{B}^\circ := \{g = A_{X \rightarrow \Omega} \rho \mid \rho \in \mathcal{M}(X; \mathcal{U}^\circ)\}, \quad \|g\|_{\mathcal{B}^\circ} := \sup_{w \in \Omega} \|g(w)\|_{\mathcal{U}^\circ}, \quad (4.12b)$$

where, for every  $f \in \mathcal{B}$ ,  $g \in \mathcal{B}^\circ$ , and for any representative  $\mu \in \mathcal{M}(\Omega; \mathcal{U})$  of  $f$  and  $\rho \in \mathcal{M}(X; \mathcal{U}^\circ)$  of  $g$ , the pairing is defined as

$$\langle g | f \rangle_{\mathcal{B}} = \int_{X \times \Omega} \phi(x, w) d\langle \rho | \mu \rangle_{\mathcal{U}}(x, w) \quad (4.13)$$

where  $\langle \rho | \mu \rangle_{\mathcal{U}}$  denotes the Hahn-Kolmogorov extension of the scalar-valued measure satisfying  $\langle \rho | \mu \rangle_{\mathcal{U}}(E \times F) = \langle \rho(E) | \mu(F) \rangle_{\mathcal{U}}$  for any Borel subsets  $E \subset X, F \subset \Omega$ . The kernel  $K: X \times \Omega \rightarrow \text{Twinn}(\mathcal{U}, \mathcal{U}^\circ)$  satisfies

$$\begin{aligned} K_{\mathcal{U}}(x, w)u &:= \phi(x, w)u, & K_{\mathcal{U}^\circ}(x, w)u^\circ &:= \phi(x, w)u^\circ \\ K(x, w)(u, u^\circ) &= \phi(x, w) \langle u^\circ | u \rangle_{\mathcal{U}}. \end{aligned} \quad (4.14)$$

If, in addition,  $(\mathcal{V}, \mathcal{V}^\circ)$  is a dual pair of normed vector spaces with  $X \subseteq \mathcal{V}$  compact,  $\tilde{\Omega} \subseteq \mathcal{V}^\circ$  locally-compact Hausdorff,  $\Omega := \tilde{\Omega} \times \mathbb{R}$ , and

$$\phi \in C_0(X \times \Omega), \quad \phi(x, (\omega, b)) := \sigma(\langle \omega | x \rangle_{\mathcal{V}} + b) \beta((\omega, b)), \quad (4.15)$$

for a measurable activation  $\sigma: \mathbb{R} \rightarrow \mathbb{R}$  and a measurable positive  $\beta: \Omega \rightarrow \mathbb{R}$ , then  $(\mathcal{B}, \mathcal{B}^\circ)$  is called a neural vv-RKBS pair.  $\diamond$

**Remark 4.2.** Since it is still based on the real-valued function  $\phi$  and leads to the kernel  $\phi(x, w) \langle u^\circ | u \rangle_{\mathcal{U}}$ , this definition generalizes the Hilbert space construction in (2.9).  $\diamond$

**Remark 4.3.** The Hahn–Kolmogorov extension theorem (see Theorem 1.7.8 in Tao [49]) is usually stated for nonnegative pre-measures, whereas here we are dealing with a signed set function. Moreover, for the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$  to be well-defined, the extension  $\langle \rho | \mu \rangle_{\mathcal{U}}$  must be of bounded variation, which is not addressed by the Hahn–Kolmogorov theorem. To ensure that the extension is both well-defined and of bounded total variation, we establish this result in Theorem B.1.  $\diamond$

**Remark 4.4.** Although all the operations are defined for bounded measurable  $\phi$ , the definition uses the standard  $\phi \in C_0(X \times \Omega)$  assumption, which already covers many cases. In particular, for  $\mathbb{R}^d$ -valued neural networks,  $X \subseteq \mathbb{R}^{d_x}$  can be taken compact, since real-world inputs are always bounded. With  $\Omega \subseteq \mathbb{R}^{d_\Omega}$  and bounded weights (e.g. through  $\beta$ ), this ensures  $\phi \in C_0(X \times \Omega)$ .

In neural operators, weights also lie in  $\Omega \subseteq \mathbb{R}^{d_\Omega}$ , but inputs are theoretically infinite-dimensional functions, making compactness a strong assumption. The finite-dimensional manifold hypothesis, however, suggests that data effectively lies on a finite-dimensional manifold. Moreover, in practice, inputs are given as meshes or point clouds. Many methods first project them to finite-dimensional (function) spaces [42, 21, 8, 9] or latent spaces [43, 26]. In all presented cases, the inputs or their projections lie in a finite-dimensional space, where compactness assumptions are more natural.

Given this and the simplification it provides for analysis, we retain the  $\phi \in C_0(X \times \Omega)$  assumption.  $\diamond$

It is not yet clear whether the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$  is independent of the specific choices of  $\mu$  and  $\rho$ . Before establishing that this is indeed the case, we introduce a specific integral operator.

Let  $(\mathcal{U}, \mathcal{U}^\circ)$  be a dual pair equipped with a continuous duality pairing and write a simple function  $g: \Omega \rightarrow \mathcal{U}^\circ$  as

$$g = \sum_{j=1}^m u_j^\circ \mathbb{1}_{B_j} \quad (4.16)$$

where the sets  $B_j \subseteq \Omega$  are disjoint. For  $\mu \in \mathcal{M}(\Omega; \mathcal{U})$ , we then define

$$\langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)} := \int_{\Omega} \langle g(w) | d\mu(w) \rangle_{\mathcal{U}} := \sum_{j=1}^m \langle u_j^\circ | \mu(B_j) \rangle_{\mathcal{U}}. \quad (4.17)$$

Since  $\langle \cdot | \cdot \rangle_{\mathcal{U}}: \mathcal{U}^\circ \times \mathcal{U} \rightarrow \mathbb{R}$  is continuous, there exists  $C > 0$  such that  $\langle u^\circ | u \rangle_{\mathcal{U}} \leq C \|u^\circ\|_{\mathcal{U}^\circ} \|u\|_{\mathcal{U}}$  for all  $u^\circ \in \mathcal{U}^\circ$  and  $u \in \mathcal{U}$ . Consequently, the linear operator  $g \mapsto \langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)}$  is bounded on the space of simple functions

$$\begin{aligned} \left| \langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)} \right| &= \left| \sum_{j=1}^m \langle u_j^\circ | \mu(B_j) \rangle_{\mathcal{U}} \right| \leq C \sum_{i=1}^m \|u_i^\circ\|_{\mathcal{U}^\circ} \|\mu(B_i)\|_{\mathcal{U}} \leq C \left( \max_{i=1, \dots, m} \|u_i^\circ\|_{\mathcal{U}^\circ} \right) \sum_{j=1}^m \|\mu(B_j)\|_{\mathcal{U}} \\ &= C \left( \sup_{w \in \Omega} \|g(w)\|_{\mathcal{U}^\circ} \right) \sum_{j=1}^m \|\mu(B_j)\|_{\mathcal{U}} \leq C \left( \sup_{w \in \Omega} \|g(w)\|_{\mathcal{U}^\circ} \right) |\mu|_{\mathcal{U}}(\Omega). \end{aligned} \quad (4.18)$$

where the last inequality follows from the definition of the total variation  $|\mu|_{\mathcal{U}}(\Omega)$ .

To extend this definition beyond simple functions, we consider the space of functions that are uniform limits of simple functions and equip this space with the supremum norm. The upper bound  $C \left( \sup_{w \in \Omega} \|g(w)\|_{\mathcal{U}^\circ} \right) |\mu|_{\mathcal{U}}(\Omega)$  remains computable for all functions in this space in which the simple functions are dense by definition. Hence, we can extend  $g \mapsto \langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)}$  continuously. For  $f: X \rightarrow \mathcal{U}$  and  $\rho \in \mathcal{M}(X; \mathcal{U}^\circ)$ , we define  $\langle \rho | f \rangle_{C_0(X; \mathcal{U})}$  in the same way.

**Remark 4.5.** The above construction coincides with the integral in Appendix A of Carmeli et al. [16]. They extend using the Hahn-Banach theorem to the space  $L^1(\Omega, |\mu|_{\mathcal{U}}; \mathcal{U}^\circ)$  of functions from  $\Omega$  to  $\mathcal{U}^\circ$  that are Bochner integrable with respect to the measure  $|\mu|_{\mathcal{U}}$ . While such an extension would also be possible in our setting, our simpler extension suffices as it immediately preserves the bound in (4.18), which is crucial for establishing the continuity of the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$ .  $\diamond$

Using the integral operators constructed above, we now show that the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$  is well-defined.

**Theorem 4.1** (Pairing of integral vv-RKBS is well-defined). *Let  $(\mathcal{U}, \mathcal{U}^\circ)$  be a dual pair of Banach spaces with a continuous pairing. Let  $(\mathcal{B}, \mathcal{B}^\circ)$  be a vector-valued integral RKBS, where functions in  $\mathcal{B}$  map from a set  $X$  to  $\mathcal{U}$  and functions in  $\mathcal{B}^\circ$  map from a set  $\Omega$  to  $\mathcal{U}^\circ$ . For any  $\mu \in \mathcal{M}(\Omega; \mathcal{U})$  and  $\rho \in \mathcal{M}(X; \mathcal{U}^\circ)$ , if  $f = A_{\Omega \rightarrow X} \mu$  and  $g = A_{X \rightarrow \Omega} \rho$ , then*

$$\langle g | f \rangle_{\mathcal{B}} = \langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)} = \langle \rho | f \rangle_{C_0(X; \mathcal{U})} \quad (4.19)$$

*In particular, the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$  does not depend on the specific representations  $\mu$  and  $\rho$  of  $f$  and  $g$ , respectively.*

*Proof.* As  $\phi \in C_0(X \times \Omega)$ , Theorem B.2 guarantees that for any  $\varepsilon > 0$ , there exist:

- pairwise disjoint Borel sets  $A_1, \dots, A_n \subseteq X$  with  $\bigcup_{i=1}^n A_i = X$ ,
- pairwise disjoint Borel sets  $B_1, \dots, B_m \subseteq \Omega$  with  $\bigcup_{j=1}^m B_j = \Omega$ ,

and a simple function of the form

$$\phi_\varepsilon(x, w) = \sum_{i=1}^n \sum_{j=1}^m a_{ij} \mathbb{1}_{A_i \times B_j}(x, w) \quad (4.20)$$

such that

$$\sup_{(x, w) \in X \times \Omega} |\phi_\varepsilon(x, w) - \phi(x, w)| \leq \varepsilon. \quad (4.21)$$

By uniform convergence, we have:

$$\langle g | f \rangle_{\mathcal{B}} = \int_{X \times \Omega} \phi(x, w) d \langle \rho | \mu \rangle_{\mathcal{U}}(x, w) = \lim_{\varepsilon \rightarrow 0} \int_{X \times \Omega} \phi_\varepsilon(x, w) d \langle \rho | \mu \rangle_{\mathcal{U}}(x, w). \quad (4.22)$$

The integral in the limit can be written equivalently as

$$\begin{aligned} \int_{X \times \Omega} \phi_\varepsilon(x, w) d \langle \rho | \mu \rangle_{\mathcal{U}}(x, w) &= \sum_{i=1}^n \sum_{j=1}^m a_{ij} \langle \rho(A_i) | \mu(B_j) \rangle_{\mathcal{U}} \\ &= \sum_{j=1}^m \left\langle \sum_{i=1}^n a_{ij} \rho(A_i) \middle| \mu(B_j) \right\rangle_{\mathcal{U}} = \langle \mu | g_\varepsilon \rangle_{C_0(\Omega; \mathcal{U}^\circ)}, \end{aligned} \quad (4.23)$$

by using the simple function

$$g_\varepsilon(w) := \sum_{j=1}^m \left( \sum_{i=1}^n a_{ij} \rho(A_i) \right) \mathbb{1}_{B_j}(w). \quad (4.24)$$

Hence,

$$\langle g | f \rangle_{\mathcal{B}} = \lim_{\varepsilon \rightarrow 0} \langle \mu | g_\varepsilon \rangle_{C_0(\Omega; \mathcal{U}^\circ)}. \quad (4.25)$$

We will now show that  $\lim_{\varepsilon \rightarrow 0} \langle \mu | g_\varepsilon \rangle_{C_0(\Omega; \mathcal{U}^\circ)} = \langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)}$ . For fixed  $\mu$ , the map  $g \mapsto \langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)}$  is bounded on the space of functions that can be expressed as uniform limits of simple functions, where this space is equipped with the supremum norm. Hence, it suffices to prove that  $g_\varepsilon \rightarrow g$  uniformly.

To see this, fix  $w \in B_j$ . Because the  $A_i$  are pairwise disjoint and cover  $X$ , we have:

$$\begin{aligned}
\|g_\varepsilon(w) - g(w)\|_{\mathcal{U}} &= \left\| \sum_{i=1}^n a_{ij} \rho(A_i) - \int_X \phi(x, w) d\rho(x) \right\|_{\mathcal{U}^\circ} \\
&= \left\| \sum_{i=1}^n \int_{A_i} (a_{ij} - \phi(x, w)) d\rho(x) \right\|_{\mathcal{U}^\circ} \\
&\leq \sum_{i=1}^n \left\| \int_{A_i} (a_{ij} - \phi(x, w)) d\rho(x) \right\|_{\mathcal{U}^\circ} \\
&\leq \sum_{i=1}^n \sup_{x \in A_i} |a_{ij} - \phi(x, w)| \cdot |\rho|_{\mathcal{U}^\circ}(A_i) \\
&= \sum_{i=1}^n \sup_{x \in A_i} |\phi_\varepsilon(x, w) - \phi(x, w)| \cdot |\rho|_{\mathcal{U}^\circ}(A_i) \\
&\leq \varepsilon \sum_{i=1}^n |\rho|_{\mathcal{U}^\circ}(A_i) = \varepsilon |\rho|_{\mathcal{U}^\circ}(X).
\end{aligned} \tag{4.26}$$

Since  $|\rho|_{\mathcal{U}^\circ}$  has finite total variation and is independent of  $\varepsilon$ , this shows:

$$\sup_{w \in \Omega} \|g_\varepsilon(w) - g(w)\|_{\mathcal{U}} \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0. \tag{4.27}$$

Hence,

$$\lim_{\varepsilon \rightarrow 0} \langle \mu | g_\varepsilon \rangle_{C_0(\Omega; \mathcal{U}^\circ)} = \langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)}. \tag{4.28}$$

Combining this with (4.23), we conclude:

$$\langle g | f \rangle_{\mathcal{B}} = \langle \mu | g \rangle_{C_0(\Omega; \mathcal{U}^\circ)}. \tag{4.29}$$

A completely analogous argument, interchanging the roles of  $\rho$  and  $\mu$ , shows that

$$\langle g | f \rangle_{\mathcal{B}} = \langle \rho | f \rangle_{C_0(X; \mathcal{U})}. \tag{4.30}$$

□

### 4.3 Properties of integral vv-RKBS

Having established that all components in the definition of the integral vector-valued RKBS are well-defined, the question remains whether all the requirements for an adjoint pair of vector-valued RKBS are satisfied. The following theorem confirms this, and in particular shows that the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$  inherits the continuity bound from the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$ .

**Theorem 4.2** (Integral vv-RKBS is adjoint pair of vv-RKBS). *The integral vv-RKBS  $(\mathcal{B}, \mathcal{B}^\circ)$  is an adjoint pair of vv-RKBS. In particular, if  $C > 0$  is the smallest constant such that the continuous pairing  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$  satisfies  $|\langle u^\circ | u \rangle_{\mathcal{U}}| \leq C \|u^\circ\|_{\mathcal{U}^\circ} \|u\|_{\mathcal{U}}$  for all  $(u, u^\circ) \in \mathcal{U} \times \mathcal{U}^\circ$ , then the pairing  $\langle \cdot | \cdot \rangle_{\mathcal{B}}$  is continuous with the bound*

$$|\langle g | f \rangle_{\mathcal{B}}| \leq C \|g\|_{\mathcal{B}^\circ} \|f\|_{\mathcal{B}} \tag{4.31}$$

for all  $(f, g) \in \mathcal{B} \times \mathcal{B}^\circ$ .

*Proof.* To prove that the integral vector-valued RKBS  $(\mathcal{B}, \mathcal{B}^\circ)$  forms an adjoint pair of vv-RKBSs, we verify the following points:

#### $\mathcal{B}$ and $\mathcal{B}^\circ$ are vv-RKBS:

The space  $\mathcal{B}^\circ$  is a vv-RKBS by construction, as it uses the supremum norm. For  $\mathcal{B}$ , consider an arbitrary  $x \in X$ . For any  $f = A_{\Omega \rightarrow X} \mu$  with  $\mu \in \mathcal{M}(\Omega; \mathcal{U})$ ,

$$\|f(x)\|_{\mathcal{U}} = \left\| \int_{\Omega} \phi(x, w) d\mu(w) \right\|_{\mathcal{U}} \leq \sup_{w \in \Omega} |\phi(x, w)| |\mu|_{\mathcal{U}}(\Omega), \tag{4.32}$$

where  $\sup_{w \in \Omega} |\phi(x, w)| < \infty$  due to  $\phi \in C_0(X \times \Omega)$ . Taking the infimum over all  $\mu$  such that  $f = A_{\Omega \rightarrow X} \mu$ , we obtain

$$\|f(x)\|_{\mathcal{U}} \leq \sup_{w \in \Omega} |\phi(x, w)| \|f\|_{\mathcal{B}} \tag{4.33}$$

Since  $x$  was arbitrary, all the point evaluations are continuous. Hence,  $\mathcal{B}$  is a vv-RKBS.

**$K$  satisfies the reproducing properties:**

Let  $f = A_{\Omega \rightarrow X} \mu \in \mathcal{B}$ . For any  $x \in X$  and  $u^\diamond \in \mathcal{U}^\diamond$ , we have

$$\langle u^\diamond | f(x) \rangle_{\mathcal{U}} = \int_{\Omega} \phi(x, w) d \langle u^\diamond | \mu \rangle_{\mathcal{U}}(w) = \int_{X \times \Omega} \phi(y, w) d \langle u^\diamond \delta_x | \mu \rangle_{\mathcal{U}}(y, w) = \langle \phi(x, \cdot) u^\diamond | f \rangle_{\mathcal{B}}, \quad (4.34)$$

which verifies the reproducing property for  $K_{\mathcal{U}}$ . The reproducing property for  $K_{\mathcal{U}^\diamond}$  follows by a similar argument. Furthermore, by the definition of a twin operator:

$$K(x, w)(u^\diamond, u) = \langle u^\diamond | K_{\mathcal{U}}(x, w)u \rangle_{\mathcal{U}} = \langle u^\diamond | \phi(x, w)u \rangle_{\mathcal{U}} = \phi(x, w) \langle u^\diamond | u \rangle_{\mathcal{U}} \quad (4.35)$$

**The pairing is non-degenerate:**

Assume  $g \in \mathcal{B}^\diamond$  is not the zero function. Then there exists a  $w \in \Omega$  such that  $g(w) \neq 0$ . Since  $(\mathcal{U}, \mathcal{U}^\diamond)$  is a dual pair, the non-degeneracy of  $\langle \cdot | \cdot \rangle_{\mathcal{U}}$  implies that there exists a  $u \in \mathcal{U}$  such that  $\langle g(w) | u \rangle_{\mathcal{U}} \neq 0$ . Using the reproducing property,

$$0 \neq \langle g(w) | u \rangle_{\mathcal{U}} = \langle g | K_{\mathcal{U}}(\cdot, w)u \rangle_{\mathcal{B}}. \quad (4.36)$$

Hence, there exists an  $f \in \mathcal{B}$  with  $\langle g | f \rangle_{\mathcal{B}} \neq 0$ , proving non-degeneracy in the second argument. A similar argument with the roles of  $f$  and  $g$  reversed shows the non-degeneracy in the first argument.

**The pairing is continuous:**

This follows immediately from (4.18) when taking the infimum over all  $\mu$  such that  $A_{\Omega \rightarrow X} \mu = f$ .  $\square$

The previous theorem demonstrates that the integral vv-RKBS forms an adjoint pair of vv-RKBS when using Definition 3.6. As highlighted in Theorem 3.3, this definition is the weakest one possible. The following theorem establishes that, under limited assumptions on the output dual pair  $(\mathcal{U}, \mathcal{U}^\diamond)$ , the integral vector-valued RKBS satisfies additional properties that correspond to the  $\delta$ -dual definition.

**Theorem 4.3** ( $\mathcal{B}^\diamond$  in the integral RKBS is a  $\delta$ -dual). *Let  $(\mathcal{U}, \mathcal{U}^\diamond)$  be a dual pair with continuous pairing and let  $(\mathcal{B}, \mathcal{B}^\diamond)$  be a vector-valued integral RKBS, where functions in  $\mathcal{B}$  map from a set  $X$  to  $\mathcal{U}$  and functions in  $\mathcal{B}^\diamond$  map from a set  $\Omega$  to  $\mathcal{U}^\diamond$ .*

*Then the set*

$$\{K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\} = \{\phi(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\} \quad (4.37)$$

*is dense in  $\mathcal{B}^\diamond$ .*

*Furthermore, if*

$$\|u^\diamond\|_{\mathcal{U}^\diamond} = \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle u^\diamond | u \rangle_{\mathcal{U}}|, \quad (4.38)$$

*then*

$$\|g\|_{\mathcal{B}^\diamond} = \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}|. \quad (4.39)$$

*Proof.* To prove the density, we apply Lemma B.1 to obtain pairwise disjoint sets  $A_1, \dots, A_n \subseteq X$  with  $\bigcup_{i=1}^n A_i = X$ , pairwise disjoint sets  $B_1, \dots, B_m \subseteq \Omega$  with  $\bigcup_{j=1}^m B_j = \Omega$ , and a Borel set  $D_{\frac{\varepsilon}{2}} \subseteq X \times \Omega$ , such that for each  $(i, j)$  either:

- $A_i \times B_j \subseteq D_{\frac{\varepsilon}{2}}$  and  $|\phi(x, w) - \phi(\tilde{x}, \tilde{w})| \leq \varepsilon$  for all  $(x, w), (\tilde{x}, \tilde{w}) \in A_i \times B_j$ ,
- $A_i \times B_j \subseteq (X \times \Omega) \setminus D_{\frac{\varepsilon}{2}}$  and  $|\phi(x, w)| \leq \frac{\varepsilon}{2}$  for all  $(x, w) \in A_i \times B_j$ , which implies by the triangle inequality that  $|\phi(x, w) - \phi(\tilde{x}, \tilde{w})| \leq \varepsilon$  for all  $(x, w), (\tilde{x}, \tilde{w}) \in A_i \times B_j$ .

In particular, for an arbitrary  $x_i \in A_i$ , we have

$$\sup_{x \in A_i} |\phi(x_i, w) - \phi(x, w)| \leq \sup_{x \in A_i, w \in \Omega} |\phi(x_i, w) - \phi(x, w)| = \max_j \left( \sup_{x \in A_i, w \in B_j} |\phi(x_i, w) - \phi(x, w)| \right) \leq \varepsilon \quad (4.40)$$

where the equality follows from  $\bigcup_{j=1}^m B_j = \Omega$ .

For arbitrary  $x_i \in A_i$ , define  $g_\varepsilon$  as:

$$g_\varepsilon = \sum_{i=1}^n \phi(x_i, \cdot) \rho(A_i) \in \text{span}\{\phi(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\} \quad (4.41)$$

Then, for every  $w \in \Omega$ ,

$$\begin{aligned} \|g_\varepsilon(w) - g(w)\|_{\mathcal{U}^\diamond} &= \left\| \sum_{i=1}^n \int_{A_i} (\phi(x_i, w) - \phi(x, w)) d\rho(x) \right\|_{\mathcal{U}^\diamond} \\ &\leq \sum_{i=1}^n \sup_{x \in A_i} |\phi(x_i, w) - \phi(x, w)| |\rho|_{\mathcal{U}^\diamond}(A_i) \\ &\leq \varepsilon \sum_{i=1}^n |\rho|_{\mathcal{U}^\diamond}(A_i) = \varepsilon |\rho|_{\mathcal{U}^\diamond}(X) \end{aligned} \quad (4.42)$$

Hence,

$$\|g_\varepsilon - g\|_{\mathcal{B}^\diamond} = \sup_{w \in \Omega} \|g_\varepsilon(w) - g(w)\|_{\mathcal{U}^\diamond} \leq \varepsilon |\rho|_{\mathcal{U}^\diamond}(X) \xrightarrow{\varepsilon \rightarrow 0} 0 \quad (4.43)$$

which shows that  $\text{span}\{\phi(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\} = \text{span}\{K_{\mathcal{U}^\diamond}(x, \cdot)u^\diamond \mid (x, u^\diamond) \in X \times \mathcal{U}^\diamond\}$  is dense in  $\mathcal{B}^\diamond$ .

To prove (4.39) given (4.38), observe that every  $g \in \mathcal{B}^\diamond \subset C_0(\Omega; \mathcal{U})$  attains the supremum in its norm at some point  $w_{\max}$ . Hence,

$$\|g\|_{\mathcal{B}^\diamond} = \|g(w_{\max})\|_{\mathcal{U}^\diamond} = \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle g(w_{\max}) | u \rangle_{\mathcal{U}}| = \sup_{\|u\|_{\mathcal{U}} \leq 1} |\langle g | K_{\mathcal{U}^\diamond}(\cdot, w_{\max})u \rangle_{\mathcal{B}}| \leq \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}| \quad (4.44)$$

where the last inequality holds since

$$\|K_{\mathcal{U}^\diamond}(\cdot, w_{\max})u\|_{\mathcal{B}} = \|\phi(\cdot, w_{\max})u\|_{\mathcal{B}} \leq |\delta_{w_{\max}}|(\Omega) \|u\|_{\mathcal{U}} \leq 1, \quad (4.45)$$

with the final inequality following from the assumption  $\|u\|_{\mathcal{U}} \leq 1$ .

Moreover, if (4.38) holds, then  $C = 1$  in Theorem 4.2. Combining this with the above yields

$$\|g\|_{\mathcal{B}^\diamond} \leq \sup_{\|f\|_{\mathcal{B}} \leq 1} |\langle g | f \rangle_{\mathcal{B}}| \leq \sup_{\|f\|_{\mathcal{B}} \leq 1} \|g\|_{\mathcal{B}^\diamond} \|f\|_{\mathcal{B}} \leq \|g\|_{\mathcal{B}^\diamond}. \quad (4.46)$$

Hence, (4.39) holds.  $\square$

## 4.4 Representer theorem

So far, we have introduced the integral vv-RKBS and the neural RKBS as infinite-dimensional formulations of neural networks. In practice, we aim to optimize the weights of a neural network using data. Thus, if the integral vv-RKBS is to serve as the function space for neural networks, optimizing over this space should return neural networks as solutions. Results of this type are known as Representer Theorems. This section establishes a general representer theorem for the integral vv-RKBS introduced in Definition 4.3. In the next section, we demonstrate that  $\mathbb{R}^d$ -valued neural networks, hypernetworks, and DeepONets are elements of specific integral and neural vv-RKBS spaces, and utilize the general representer theorem to derive representer theorems for these models.

To set up the optimization problem required for learning from data, assume we want to learn a function from a domain  $X$  to a Banach space  $\mathcal{U}$ . For  $\mathbb{R}^d$ -valued neural networks,  $X \subseteq \mathbb{R}^{d_x}$  and  $\mathcal{U} \subseteq \mathbb{R}^{d_u}$ . In operator learning,  $X$  may be  $\mathbb{R}^{d_x}$  or even a function space, with the output also lying in a function space. In practice, however, we rarely have access to the full output function and instead observe only its sampled values. Concretely, every output function  $u: D \rightarrow \mathbb{R}$  is sampled at  $d$  points  $p_i \in D$  through the measurement operator

$$Mu = [u(p_1), \dots, u(p_d)] \in \mathbb{R}^d. \quad (4.47)$$

When the output function space is taken to be an RKBS, the measurement operator  $M$  is a bounded linear operator from the RKBS to  $\mathbb{R}^d$ .

In general, we assume a measurement operator  $M: \mathcal{U} \rightarrow \mathbb{R}^d$ , such as the sampling operator above or, when  $\mathcal{U} = \mathbb{R}^d$ , the identity map. Given data  $\{(x_n, y_n)\}_{n=1}^N \subseteq X \times \mathbb{R}^d$ , we then formulate a supervised optimization problem, for which we derive a representer theorem.

**Theorem 4.4.** *Assume we are given data  $\{(x_n, y_n)\}_{n=1}^N \subseteq X \times \mathbb{R}^d$ . Let  $L: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  be such that for every  $y \in \mathbb{R}^d$ , the map  $L(\cdot, y)$  is convex, coercive, and lower semicontinuous. Let  $\mathcal{B}$  be an integral vv-RKBS of*

functions from  $X$  to a Banach space  $\mathcal{U}$  with predual  $\mathcal{U}_*$  and with weight space  $\Omega$ . Suppose there exists a bounded linear operator  $M: \mathcal{U} \rightarrow \mathbb{R}^d$  of the form

$$(Mu)_j = \langle u | v_j \rangle_{\mathcal{U}_*}, \quad (4.48)$$

where  $v_j \in \mathcal{U}_*$ ,  $\langle \cdot | \cdot \rangle_{\mathcal{U}_*}$  is the canonical duality pairing between  $\mathcal{U}_*$  and  $(\mathcal{U}_*)^* = \mathcal{U}$ .

Then the solution to the supervised learning problem

$$\min_{f \in \mathcal{B}} \frac{1}{N} \sum_{n=1}^N L(M(f(x_n)), y_n) + \lambda \|f\|_{\mathcal{B}}, \quad \lambda \geq 0, \quad (4.49)$$

admits a representation of the form

$$\mu^\dagger = \sum_{m=1}^{Nd} a_m \delta_{w_m} u_m, \quad f^\dagger = A_{\Omega \rightarrow X} \mu^\dagger = \sum_{n=1}^{Nd} a_m \phi(\cdot, w_m) u_m, \quad (4.50)$$

with  $u_m \in \text{Ext}(\{u \in \mathcal{U} : \|u\|_{\mathcal{U}} \leq 1\})$  and  $a_m \in \mathbb{R}$ .

*Proof.* Assume that there exists  $\mu^\dagger$  with

$$\mu^\dagger \in \arg \min_{\mu \in \mathcal{M}(\Omega; \mathcal{U})} \frac{1}{N} \sum_{n=1}^N L(M((A_{\Omega \rightarrow X} \mu)(x_n)), y_n) + \lambda |\mu|_{\mathcal{U}}(\Omega). \quad (4.51)$$

Let  $f^\dagger = A_{\Omega \rightarrow X} \mu^\dagger$ . For any  $f = A_{\Omega \rightarrow X} \mu$  with  $|\mu|_{\mathcal{U}} \leq \|f\|_{\mathcal{B}} + \delta$ , where  $\delta > 0$  is arbitrary, we have

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N L(M(f^\dagger(x_n)), y_n) + \lambda \|f^\dagger\|_{\mathcal{B}} &\leq \frac{1}{N} \sum_{n=1}^N L(M((A_{\Omega \rightarrow X} \mu^\dagger)(x_n)), y_n) + \lambda |\mu^\dagger|_{\mathcal{U}}(\Omega) \\ &\leq \frac{1}{N} \sum_{n=1}^N L(M((A_{\Omega \rightarrow X} \mu)(x_n)), y_n) + \lambda |\mu|_{\mathcal{U}}(\Omega) \\ &\leq \frac{1}{N} \sum_{n=1}^N L(M(f(x_n)), y_n) + \lambda \|f\|_{\mathcal{B}} + \delta. \end{aligned} \quad (4.52)$$

Letting  $\delta \rightarrow 0$  yields, for any  $f \in \mathcal{B}$ ,

$$\frac{1}{N} \sum_{n=1}^N L(M(f^\dagger(x_n)), y_n) + \lambda \|f^\dagger\|_{\mathcal{B}} \leq \frac{1}{N} \sum_{n=1}^N L(M(f(x_n)), y_n) + \lambda \|f\|_{\mathcal{B}}. \quad (4.53)$$

Hence,  $f^\dagger$  is the minimizer of (4.49). It therefore suffices to analyze problem (4.51), for which we aim to apply Theorem 3.3 of Bredies and Carioni [12], yielding a sparse minimizer.

By the dual space characterization (4.9),

$$C_0(\Omega; \mathcal{U}_*)^* \cong \mathcal{M}(\Omega; (\mathcal{U}_*)^*) \cong \mathcal{M}(\Omega; \mathcal{U}), \quad (4.54)$$

since  $\mathcal{U}$  has  $\mathcal{U}_*$  as a predual and  $\Omega$  is locally compact and Hausdorff. Consequently, we can equip  $\mathcal{M}(\Omega; \mathcal{U})$  with the weak-\* topology, under which it becomes a real locally-convex topological vector space.

Define

$$\mathcal{A}: \mathcal{M}(\Omega; \mathcal{U}) \rightarrow \mathbb{R}^{Nd}, \quad \mu \mapsto [M((A_{\Omega \rightarrow X} \mu)(x_1)), \dots, M((A_{\Omega \rightarrow X} \mu)(x_N))]. \quad (4.55)$$

A component of  $\mathcal{A}(\mu)$  can be written as

$$\begin{aligned} M_j((A_{\Omega \rightarrow X} \mu)(x_n)) &= \langle (A_{\Omega \rightarrow X} \mu)(x_n) | v_j \rangle_{\mathcal{U}_*} \\ &= \left\langle \int_{\Omega} \phi(x_n, w) d\mu(w) \middle| v_j \right\rangle_{\mathcal{U}_*} \\ &= \int_{\Omega} \phi(x_n, w) d \langle \mu | v_j \rangle_{\mathcal{U}_*}(w) \\ &= \int_{\Omega} \phi(x_n, w) d \langle v_j | \mu \rangle_{\mathcal{U}_*, \mathcal{U}}(w) \\ &= \int_{\Omega} \langle \phi(x_n, w) v_j | d\mu(w) \rangle_{\mathcal{U}_*, \mathcal{U}} \\ &= \langle \mu | \phi(x_n, \cdot) v_j \rangle_{C_0(\Omega; \mathcal{U}_*)}. \end{aligned} \quad (4.56)$$

Here, the last equality follows from the fact [37] that the duality pairing that realizes the characterization in the first isomorphism of (4.54) is the continuous extension of (4.17) by uniform approximation with simple functions. By definition of the weak-\* topology, this expression is weak-\* continuous in  $\mu$ . Hence  $\mathcal{A}$  is weak-\* continuous.

Let  $H = \text{range}(\mathcal{A})$  and define

$$F: H \rightarrow \mathbb{R}, [\tilde{y}_1, \dots, \tilde{y}_N] \mapsto \frac{1}{N} \sum_{n=1}^N L(\tilde{y}_n, y_n). \quad (4.57)$$

Since  $L(\cdot, y)$  is convex, coercive, and lower semicontinuous,  $F$  has the same properties on  $H$ .

The regularizer is the total variation norm  $|\mu|_{\mathcal{U}}(\Omega)$ . Using again the duality (4.54), for any  $v \in \mathcal{U}_*$  with  $\|v\|_{\mathcal{U}_*} \leq 1$ ,

$$\langle \mu_n | v \rangle_{C_0(\Omega; \mathcal{U}_*)} \leq \sup_{\|\tilde{v}\|_{C_0(\Omega; \mathcal{U}_*)} \leq 1} \langle \mu_n | \tilde{v} \rangle_{C_0(\Omega; \mathcal{U}_*)} = |\mu_n|_{\mathcal{U}}(\Omega). \quad (4.58)$$

Taking the liminf, the supremum over  $v$ , and using (4.54) shows

$$|\mu|_{\mathcal{U}}(\Omega) = \sup_{\|v\|_{C_0(\Omega; \mathcal{U}_*)} \leq 1} \left| \langle \mu | v \rangle_{C_0(\Omega; \mathcal{U}_*)} \right| \leq \liminf_n |\mu_n|_{\mathcal{U}}(\Omega). \quad (4.59)$$

Therefore,  $|\cdot|_{\mathcal{U}}(\Omega)$  is weak-\* lower semicontinuous. Moreover,  $\{\mu \in \mathcal{M}(\Omega; \mathcal{U}) : |\mu|_{\mathcal{U}}(\Omega) = 0\} = \{0\}$ . Finally, coercivity follows since the sets  $\{\mu \in \mathcal{M}(\Omega; \mathcal{U}) : |\mu|_{\mathcal{U}}(\Omega) \leq \alpha\}$  are weak-\* compact by Banach–Alaoglu.

Thus, all assumptions of Theorem 3.3 in Bredies and Carioni [12] are satisfied. Since the extreme points of the unit ball in  $\mathcal{M}(\Omega; \mathcal{U})$  are (see Theorem 2 of Dirk [52], or Lemma 3.2 of Bredies et al. [13] whose proof applies without modifications in the general setting considered here) characterized as

$$\text{Ext}(\{\mu \in \mathcal{M}(\Omega; \mathcal{U}) : |\mu|_{\mathcal{U}}(\Omega) \leq 1\}) = \{\delta_w u : w \in \Omega, u \in \text{Ext}(\{u \in \mathcal{U} : \|u\|_{\mathcal{U}} \leq 1\})\}, \quad (4.60)$$

the solution has the form

$$\mu^\dagger = \sum_{m=1}^{Nd} a_m \delta_{w_m} u_m, \quad (4.61)$$

with  $u_m \in \text{Ext}(\{u \in \mathcal{U} : \|u\|_{\mathcal{U}} \leq 1\})$  and  $a_m \in \mathbb{R}$ .  $\square$

**Remark 4.6.** The setup of Theorem 4.4 follows Definition 4.3, where both  $X$  and  $\Omega$  are assumed to be locally-compact Hausdorff, and which ensures that the generalized Riesz representation of (4.9) applies both in  $\mathcal{M}(\Omega; \mathcal{U})$  and in  $\mathcal{M}(X; \mathcal{U}^\circ)$ . However,  $\mathcal{B}^\circ$  does not play a role in the proof, so the representer theorem also holds without restricting  $X$  to be locally compact or imposing that  $\phi$  vanishes at infinity in the  $x$  variable. Effectively, the only requirements are that the integral in (4.10) should exist, is finite for any  $\mu \in \mathcal{M}(\Omega; \mathcal{U})$ , and that  $\Omega$  is locally-compact Hausdorff so that the Riesz theorem applies.  $\diamond$

The representation of  $f^\dagger$  closely resembles the standard vv-RKHS representation

$$\sum_{n=1}^N K(\cdot, x_n) u_n, \quad (4.62)$$

where  $K: X \times X \rightarrow \mathcal{L}(\mathcal{U})$  is the vv-RKHS kernel (Definition 2.2),  $\mathcal{U}$  is a Hilbert space, and  $u_n \in \mathcal{U}$ . To make the analogy precise, let  $W = \{w_1, \dots, w_{Nd}\}$ , so  $|W| \leq Nd$ . For each unique  $\tilde{w}_k \in W$ , set

$$I_{\tilde{w}_k} = \{m \in \{1, \dots, Nd\} \mid w_m = \tilde{w}_k\}, \quad \tilde{u}_k = \sum_{m \in I_{\tilde{w}_k}} a_m u_m. \quad (4.63)$$

Then the representation from Theorem 4.4 can be rewritten as

$$f^\dagger = A_{\Omega \rightarrow X} \mu^\dagger = \sum_{k=1}^{|W|} \phi(\cdot, \tilde{w}_k) \tilde{u}_k = \sum_{k=1}^{|W|} K_{\mathcal{U}}(\cdot, \tilde{w}_k) \tilde{u}_k, \quad (4.64)$$

where the last equality uses the kernel definition in Definition 4.3. This mirrors the vv-RKHS formula, with the difference that we only guarantee  $|W| \leq Nd$ , not  $|W| \leq N$ .

To understand this limitation, substitute the above  $f^\dagger$  into the optimization problem, assuming the optimal  $\tilde{w}_k$  are already fixed:

$$\min_{\tilde{u}_k} \frac{1}{N} \sum_{n=1}^N L \left( \sum_{k=1}^{|W|} \phi(x_i, \tilde{w}_k) \tilde{u}_k, y_n \right) + \lambda \sum_{k=1}^{|W|} \|\tilde{u}_k\|_{\mathcal{U}}. \quad (4.65)$$

Here, the role of the norm on  $\mathcal{U}$  is decisive. If  $\tilde{u}_k \in \mathbb{R}^d$  and  $\|\cdot\|_{\mathcal{U}}$  behaves like an  $\ell_1$  norm, no groupwise sparsity is expected, so the number of active terms  $|W|$  need not reduce to  $N$ . In this case, the vectors  $u_i \in \mathbb{R}^d$  correspond to standard basis vectors since they are extremal points of the  $\ell_1$  unit ball. Consequently, we obtain at most  $Nd$  nonzero coefficients, one per component across the  $Nd$  terms.

In contrast, if  $\|\cdot\|_{\mathcal{U}}$  behaves more like an  $\ell_2$  norm, the formulation resembles a group-lasso penalty, which is known to promote entire vectors to vanish [3]. This effect may decrease the number of active terms, potentially yielding  $|W| \leq N$ . Another important difference compared to the  $\ell_1$  case is that the  $u_i \in \mathbb{R}^d$  are no longer the standard basis vectors but can have full entries as they lie on the  $\ell_2$  unit ball. If groupwise sparsity were possible, we would again obtain only  $Nd$  nonzero coefficients, analogous to the  $\ell_1$  case. However, because the representer theorem in our setting does not permit such groupwise sparsity, the number of coefficients increases to  $Nd^2$  when using the  $\ell_2$  norm.

A detailed investigation into groupwise sparsity is left for future work.

## 5 $\mathbb{R}^d$ -valued neural networks, hypernetworks, and DeepONets

This section constructs the spaces for  $\mathbb{R}^d$ -valued neural networks, hypernetworks, and DeepONets, and derives the corresponding representer theorems by applying Theorem 4.4 to each setting.

### 5.1 $\mathbb{R}^d$ -valued neural network

For  $\mathbb{R}^d$ -valued neural networks, the following representer theorem is immediate.

**Corollary 5.1** (Representer Theorem  $\mathbb{R}^d$ -valued neural network). *Assume we are given data  $\{(x_n, u_n)\}_{n=1}^N \subseteq X \times \mathbb{R}^d$ . Let  $L: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  be such that for every  $u \in \mathbb{R}^d$ , the map  $L(\cdot, u)$  is convex, coercive, and lower semicontinuous. Let  $\mathcal{B}$  be a neural vv-RKBS of functions from compact  $X \subseteq \mathbb{R}^{d_x}$  to  $\mathbb{R}^d$  with weight space  $\Omega \subseteq \mathbb{R}^{d_x} \times \mathbb{R}$  and*

$$\phi \in C_0(X \times \Omega), \quad \phi(x, (\omega, b)) := \sigma(\langle \omega | x \rangle_{\ell_2} + b) \beta((\omega, b)), \quad (5.1)$$

for a measurable activation  $\sigma: \mathbb{R} \rightarrow \mathbb{R}$  and a measurable positive  $\beta: \Omega \rightarrow \mathbb{R}$ .

Then the solution to the supervised learning problem

$$\min_{f \in \mathcal{B}} \frac{1}{N} \sum_{n=1}^N L(f(x_n), u_n) + \lambda \|f\|_{\mathcal{B}}, \quad \lambda \geq 0, \quad (5.2)$$

admits a neural network representation:

$$f^\dagger = A_{\Omega \rightarrow X} \mu^\dagger = \sum_{m=1}^{Nd} \sigma(\langle \omega_m | x \rangle_{\ell_2} + b_m) u_m = U \sigma(Wx + B), \quad (5.3)$$

with  $U \in \mathbb{R}^{d_u \times Nd}$  containing the  $u_m$  as columns,  $W \in \mathbb{R}^{Nd \times d_x}$  the  $\omega_m$  as rows, and  $B \in \mathbb{R}^{Nd}$  the  $b_m$  values.

*Proof.* By Theorem 4.4, we get:

$$f^\dagger = A_{\Omega \rightarrow X} \mu^\dagger = \sum_{m=1}^{Nd} a_m \sigma(\langle \omega_m | \cdot \rangle_{\ell_2} + b_m) \beta((\omega_m, b_m)) u_m, \quad (5.4)$$

Redefining  $u_m$  to  $a_m \beta((\omega_m, b_m)) u_m$  yields the result.  $\square$

## 5.2 DeepONets and hypernetworks

As noted in the introduction, neural operators and conditional implicit neural representations (INRs) are closely related. Many neural operator methods, including DeepONet, FNO, and CNO, can be viewed as special cases of INRs. In this section, we investigate the vv-RKBS structure of neural operators via this INR perspective, particularly focusing on DeepONet and hypernetworks.

We begin by introducing two function spaces that cover both hypernetworks and DeepONets. Subsequently, we formulate an optimization problem for each function space and prove a joint representer theorem showing that the two problems admit the same sparse solution. The section concludes with a discussion of the function space formulations.

### 5.2.1 Function space covering the DeepONet and hypernetwork

The DeepONet framework is designed to learn mappings from inputs  $z \in Z$  to functions  $u: X \rightarrow \mathcal{V}$  for some Banach space  $\mathcal{V}$ . It achieves this by representing the output as a linear combination of learned basis functions:

$$f(z)(x) = \sum_{n=1}^{n_b} a_n(z) \zeta_n(x), \quad (5.5)$$

where both the coefficients  $a_n: Z \rightarrow \mathbb{R}$  and the basis functions  $\zeta_n: X \rightarrow \mathcal{V}$  are parameterized by neural networks. The basis functions can be obtained through standard training methods such as gradient descent [32] or constructed using techniques like proper orthogonal decomposition (POD) [33].

The DeepONet learns a linear subspace of functions. Hypernetworks provide a more nonlinear parameterization by mapping an input to the weights of another base network. Equivalently, this can be seen as mapping an input to an entire neural network.

To construct a function space covering both DeepONets and hypernetworks, we note that DeepONets define linear subspaces, whereas hypernetworks yield nonlinear parameterizations. This suggests that DeepONets may arise as elements of a hypernetwork function space. We confirm this intuition and introduce two hypernetwork function spaces: one interpreting the hypernetwork as a mapping to weights and the other as a mapping to entire networks.

**Definition 5.1** (Hypernetworks in weight form). Let  $Z$ ,  $X$ ,  $\Omega$ , and  $\Theta$  be locally-compact Hausdorff and  $(\mathcal{V}, \mathcal{V}^\circ)$  a continuous dual pair of Banach spaces. Define

$$\mathcal{B}_h = \left\{ f: Z \rightarrow \mathcal{M}(\Theta; \mathcal{V}) \mid f(z) = \int_{\Omega} \phi(z, w) d\mu(w), \mu \in \mathcal{M}(\Omega; \mathcal{M}(\Theta; \mathcal{V})) \right\} \quad (5.6a)$$

$$\|f\|_{\mathcal{B}_h} = \inf \left\{ |\mu|_{\mathcal{M}(\Theta; \mathcal{V})}(\Omega) \mid \mu \in \mathcal{M}(\Omega; \mathcal{M}(\Theta; \mathcal{V})) \text{ satisfies } f(z) = \int_{\Omega} \phi(z, w) d\mu(w) \right\} \quad (5.6b)$$

$$\mathcal{B}_h^\circ = \left\{ g: \Omega \rightarrow \mathcal{M}(X; \mathcal{V}^\circ) \mid g(w) = \int_Z \phi(z, w) d\pi(z), \pi \in \mathcal{M}(Z; \mathcal{M}(X; \mathcal{V}^\circ)) \right\} \quad (5.6c)$$

$$\|g\|_{\mathcal{B}_h^\circ} = \sup_{w \in \Omega} |g(w)|_{\mathcal{V}^\circ}(X) \quad (5.6d)$$

$$\langle g|f \rangle_{\mathcal{B}_h} = \int_{Z \times \Omega} \phi(z, w) d\langle \pi|\mu \rangle_{\mathcal{M}}(z, w), \quad (5.6e)$$

where  $\langle \pi|\mu \rangle_{\mathcal{M}}$  is the Hahn-Kolmogorov extension of the scalar-valued measure satisfying

$$\langle \pi|\mu \rangle_{\mathcal{M}}(E \times F) = \langle \pi(E)|\mu(F) \rangle_{\mathcal{M}(X; \mathcal{V}^\circ), \mathcal{M}(\Theta; \mathcal{V})} \quad (5.7)$$

with

$$\langle \rho|\nu \rangle_{\mathcal{M}(X; \mathcal{V}^\circ), \mathcal{M}(\Theta; \mathcal{V})} := \int_{X \times \Theta} d\langle \rho|\nu \rangle_{\mathcal{V}}(x, \theta) \quad (5.8)$$

and  $\langle \rho|\nu \rangle_{\mathcal{V}}$  the Hahn-Kolmogorov extension of the scalar measure satisfying  $\langle \rho|\nu \rangle_{\mathcal{V}}(E \times F) = \langle \rho(E)|\nu(F) \rangle_{\mathcal{V}}$  under the duality  $(\mathcal{V}, \mathcal{V}^\circ)$ . The pair  $(\mathcal{B}_h, \mathcal{B}_h^\circ)$  is called an integral vv-RKBS pair of hypernetworks *in weight space form*.  $\diamond$

**Definition 5.2** (Hypernetworks in function form). Let  $Z$ ,  $X$ ,  $\Omega$ , and  $\Theta$  be locally-compact Hausdorff and  $(\mathcal{V}, \mathcal{V}^\circ)$  a continuous dual pair of Banach spaces. Additionally, let  $(\mathcal{B}, \mathcal{B}^\circ)$  be an integral vv-RKBS with  $\mathcal{B} \subset \{u: X \rightarrow \mathcal{V}\}$ ,  $\mathcal{B}^\circ \subset \{v: \Theta \rightarrow \mathcal{V}^\circ\}$ , and kernel  $\psi: X \times \Theta \rightarrow \mathbb{R}$ . Define

$$\mathcal{B}_h = \left\{ f: Z \rightarrow \mathcal{B} \mid f(z) = \int_{\Omega} \phi(z, w) d\mu(w), \mu \in \mathcal{M}(\Omega; \mathcal{B}) \right\} \quad (5.9a)$$

$$\|f\|_{\mathcal{B}_h} = \inf \left\{ \|\mu\|_{\mathcal{B}(\Omega)} \mid \mu \in \mathcal{M}(\Omega; \mathcal{B}) \text{ satisfies } f(z) = \int_{\Omega} \phi(z, w) d\mu(w) \right\} \quad (5.9b)$$

$$\mathcal{B}_h^\circ = \left\{ g: \Omega \rightarrow \mathcal{B}^\circ \mid g(w) = \int_Z \phi(z, w) d\pi(z), \pi \in \mathcal{M}(Z; \mathcal{B}^\circ) \right\} \quad (5.9c)$$

$$\|g\|_{\mathcal{B}_h^\circ} = \sup_{w \in \Omega} \|g(w)\|_{\mathcal{B}^\circ} \quad (5.9d)$$

$$\langle g|f \rangle_{\mathcal{B}_h} = \int_{Z \times \Omega} \phi(z, w) d\langle \pi|\mu \rangle_{\mathcal{B}}(z, w) \quad (5.9e)$$

where  $\langle \pi|\mu \rangle_{\mathcal{B}}$  is the Hahn-Kolmogorov extension of the scalar measure satisfying  $\langle \pi|\mu \rangle_{\mathcal{B}}(E \times F) = \langle \pi(E)|\mu(F) \rangle_{\mathcal{B}}$ , the latter pairing being as in Definition 3.6. The pair  $(\mathcal{B}_h, \mathcal{B}_h^\circ)$  is called an integral vv-RKBS pair of hypernetworks *in function space form*.  $\diamond$

**Remark 5.1.** In the weight-space viewpoint, it seems natural to let  $\mathcal{M}(\Theta; \mathcal{V})$  and  $\mathcal{M}(X; \mathcal{V}^\circ)$  have a pairing that reflects how the output weights enter the base neural network's structure:

$$\langle \rho|\mu \rangle_{\mathcal{M}(X, \mathcal{V}^\circ), \mathcal{M}(\Theta, \mathcal{V})} = \int_{X \times \Theta} \psi(x, \theta) d\langle \rho|\mu \rangle_{\mathcal{V}} = \langle g|f \rangle_{\mathcal{B}}, \quad (5.10)$$

where  $\langle \rho|\mu \rangle_{\mathcal{V}}(A \times B) = \langle \rho(A)|\mu(B) \rangle_{\mathcal{V}}$ . The last equality follows from Definition 4.3, where  $(\mathcal{B}, \mathcal{B}^\circ)$  is the integral vv-RKBS in the hypernetwork's function-space viewpoint. However, if the nullspace of  $A_{\Theta \rightarrow X}$  is nontrivial, then  $\langle \rho|\mu \rangle_{\mathcal{M}(X, \mathcal{V}^\circ), \mathcal{M}(\Theta, \mathcal{V})} = \langle g|f \rangle_{\mathcal{B}} = \langle g|0 \rangle_{\mathcal{B}} = 0$  may hold for nonzero  $\rho, \mu$ . Hence, this pairing is dual only when the nullspace of  $A_{\Theta \rightarrow X}$  is trivial. So the duality assumption in Definition 4.3 holds only in that case. In contrast, the function-space viewpoint avoids this issue, since it pairs  $f$  and  $g$  directly rather than their representing measures.  $\diamond$

**Remark 5.2.** The function-space viewpoint of the integral vv-RKBS pair of hypernetworks naturally contains DeepONets as a special case. Let  $(\mathcal{B}, \mathcal{B}^\circ)$  be a neural vv-RKBS, and take  $\{\zeta_n\}_{n=1}^{n_b} \subset \mathcal{B}$  as basis functions, which are neural networks by assumption. Define the  $\mathcal{B}$ -valued measure

$$\mu = \sum_{n=1}^{n_b} \left( \sum_{k=1}^{h_{nn}} a_{nk} \delta_{w_{nk}} \right) \zeta_n \quad (5.11)$$

Then

$$f(z)(x) = (A_{\Omega \rightarrow Z} \mu)(z)(x) = \sum_{n=1}^{n_b} \left( \sum_{k=1}^{h_{nn}} a_{nk} \phi(z, w_{nk}) \right) \zeta_n(x) =: \sum_{n=1}^{n_b} a_n(z) \zeta_n(x) \quad (5.12)$$

which recovers the DeepONet formulation in (5.5).  $\diamond$

While the weight and function space viewpoints are similar, they differ in key aspects. The weight-space viewpoint first maps to measures and then integrates them to obtain a function. The function-space viewpoint does the opposite and first maps to functions (integrating out measures) and then integrates a measure of these functions. A more significant difference lies in the norms. The weight-space viewpoint uses the total variation of  $\mathcal{M}(\Theta; \mathcal{U})$  within the total variation of  $\mathcal{M}(\Omega; \mathcal{M}(\Theta; \mathcal{U}))$ , while the function-space viewpoint uses  $\|\cdot\|_{\mathcal{B}}$  inside the total variation of  $\mathcal{M}(\Omega; \mathcal{B})$ . These differences affect the resulting optimization problems.

### 5.2.2 Shared representer theorem

Since the two function spaces describe the same structure, we expect the corresponding optimization problems to yield the same solutions and, hence, share the same representer theorem. To define these optimization problems, let  $A_{\Omega \rightarrow Z}$  be as in Equations (4.10)–(4.11), with  $\phi$  as integrand. Since this operation is valid for any measure

on  $\Omega$ , the same notation applies whether we consider  $\mathcal{M}(\Omega; \mathcal{M}(\Theta; \mathcal{V}))$  or  $\mathcal{M}(\Omega; \mathcal{B})$ . Similarly, independent of the target space of  $\nu$ , we define

$$A_{\Theta \rightarrow X} \nu := \int_{\Theta} \psi(\cdot, \theta) d\nu(\theta) \quad (5.13)$$

with  $\psi: X \times \Theta \rightarrow \mathbb{R}$  as in Definition 5.2.

We also define a measurement operator  $M: \mathcal{B} \rightarrow \mathbb{R}^d$ , where  $\mathcal{B}$  is the output RKBS in the hypernetwork function-space formulation. In the weight-space formulation, we apply  $M$  to  $f = A_{\Theta \rightarrow X}(u(z)) \in \mathcal{B}$  with  $u(z) \in \mathcal{M}(\Theta; \mathcal{V})$ . The components of  $M$  are taken as

$$M_n f = \langle v_n^\diamond | f(x_n) \rangle_{\mathcal{V}}. \quad (5.14)$$

If  $\mathcal{V} = \mathbb{R}^d$ , one may consider

$$M_{nj} f = \langle e_j | f(x_n) \rangle_{\ell_2}, \quad (5.15)$$

which produces measurements for each component and is simply a reindexed version of (5.14).

With these definitions, the optimization problems can be written directly in terms of measures rather than functions  $f \in \mathcal{B}_h$ . By the first step in the proof of Theorem 4.4, this suffices since minimizing over measures yields the same minimizer as minimizing over  $f$ .

In the weight-space viewpoint, using  $|\mu|_{\mathcal{M}}$  as short-hand notation for  $|\mu|_{\mathcal{M}(\Theta; \mathcal{V})}$ , the problem is

$$\min_{\mu \in \mathcal{M}(\Omega; \mathcal{M}(\Theta; \mathcal{V}))} \frac{1}{N} \sum_{n=1}^N L(M(A_{\Theta \rightarrow X}((A_{\Omega \rightarrow Z} \mu)(z_n))), y_n) + \lambda |\mu|_{\mathcal{M}(\Omega)} \quad (5.16)$$

while in the function-space viewpoint, it is

$$\min_{\mu \in \mathcal{M}(\Omega; \mathcal{B})} \frac{1}{N} \sum_{n=1}^N L(M((A_{\Omega \rightarrow Z} \mu)(z_n))), y_n) + \lambda |\mu|_{\mathcal{B}(\Omega)} \quad (5.17)$$

The key differences are that  $A_{\Theta \rightarrow X}$  appears explicitly in the weight-space formulation but is absorbed into  $\mu$  in the function-space formulation, and the total variation regularization uses different inner norms. Despite these differences, a solution to one problem corresponds to a solution to the other. To show this, first note that a solution to the optimization problem corresponding to the weight space viewpoint (see (5.16)) provides a solution to the optimization problem of the function space viewpoint (see (5.17)).

**Theorem 5.1.** *Assume the optimization problem corresponding to the weight space viewpoint (see (5.16)) has a minimizer  $\mu_{\mathcal{M}}^\dagger \in \mathcal{M}(\Omega; \mathcal{M}(\Theta; \mathcal{V}))$ . Then the measure  $\mu_{\mathcal{B}}^\dagger \in \mathcal{M}(\Omega; \mathcal{B})$  defined via  $\mu_{\mathcal{B}}^\dagger(E) := (A_{\Theta \rightarrow X}(\mu_{\mathcal{M}}^\dagger(E)))$  is a solution to the optimization problem of the function space viewpoint (see (5.17)).*

*Proof.* Pick any  $\varepsilon > 0$  and  $\mu_{\mathcal{B}} \in \mathcal{M}(\Omega; \mathcal{B})$ . By Lemma B.1, there exists a decomposition of  $Z$  into finitely-many pairwise-disjoint sets  $A_1, \dots, A_\ell \subseteq Z$  and a decomposition of  $\Omega$  into finitely-many pairwise-disjoint sets  $B_1, \dots, B_I \subseteq \Omega$  such that

$$|\phi(z, w) - \phi(z, \tilde{w})| \leq \varepsilon \quad \text{for all } z \in Z, \text{ all } w, \tilde{w} \in B_i \text{ and all } i \in \{1, \dots, I\}. \quad (5.18)$$

For the same  $\varepsilon$ , by the definition of the total variation, there exists a partition  $\tilde{B}_1, \dots, \tilde{B}_J \subseteq \Omega$  such that

$$|\mu_{\mathcal{B}}|_{\mathcal{B}(\Omega)} \leq \varepsilon + \sum_{j=1}^J \left\| \mu_{\mathcal{B}}(\tilde{B}_j) \right\|_{\mathcal{B}} \quad (5.19)$$

Define  $\bar{B}_{ji} := \tilde{B}_j \cap B_i$ . Using the finite additivity of vector measures, we obtain

$$|\mu_{\mathcal{B}}|_{\mathcal{B}(\Omega)} \leq \varepsilon + \sum_{j=1}^J \left\| \sum_{i=1}^I \mu_{\mathcal{B}}(\bar{B}_{ji}) \right\|_{\mathcal{B}} \leq \varepsilon + \sum_{j=1}^J \sum_{i=1}^I \left\| \mu_{\mathcal{B}}(\bar{B}_{ji}) \right\|_{\mathcal{B}} = \varepsilon + \sum_{j=1}^J \sum_{i=1}^I \inf_{A_{\Theta \rightarrow X} \nu = \mu_{\mathcal{B}}(\bar{B}_{ji})} |\nu|_{\mathcal{V}(\Theta)} \quad (5.20)$$

For an arbitrary  $\delta > 0$ , let  $\nu_{ji} \in \mathcal{M}(\Theta; \mathcal{V})$  be chosen such that

$$|\nu_{ji}|_{\mathcal{V}(\Theta)} \leq \left( \inf_{A_{\Theta \rightarrow X} \nu = \mu_{\mathcal{B}}(\bar{B}_{ji})} |\nu|_{\mathcal{V}(\Theta)} \right) + \frac{\delta}{IJ} = \left\| \mu_{\mathcal{B}}(\bar{B}_{ji}) \right\|_{\mathcal{B}} + \frac{\delta}{IJ}. \quad (5.21)$$

Fix  $\delta = \varepsilon = \frac{1}{2k}$  and define

$$\mu_{\mathcal{M}}^{(k)} := \sum_{j=1}^J \sum_{i=1}^I \delta_{w_{ji}} \nu_{ji} \quad (5.22)$$

for arbitrary  $w_{ji} \in \overline{B}_{ji}$ . Then

$$\begin{aligned} |\mu_{\mathcal{B}}|_{\mathcal{B}}(\Omega) &\leq \frac{1}{2k} + |\mu_{\mathcal{M}}^{(k)}|_{\mathcal{M}}(\Omega) \\ &= \frac{1}{2k} + \left( \sum_{j=1}^J \sum_{i=1}^I |\nu_{ji}|_{\mathcal{V}}(\Theta) \right) \\ &\leq \frac{1}{2k} + \left( \sum_{j=1}^J \sum_{i=1}^I \|\mu_{\mathcal{B}}(\overline{B}_{ji})\|_{\mathcal{B}} \right) + \frac{1}{2k} \\ &\leq \frac{1}{k} + |\mu_{\mathcal{B}}|_{\mathcal{B}}(\Omega), \end{aligned} \quad (5.23)$$

which shows that  $|\mu_{\mathcal{M}}^{(k)}|_{\mathcal{M}}(\Omega) \rightarrow |\mu_{\mathcal{B}}|_{\mathcal{B}}(\Omega)$  as  $k \rightarrow \infty$ .

Define  $\mu_{\mathcal{B}}^{(k)}$  via the bounded linear operator  $T$

$$\mu_{\mathcal{B}}^{(k)}(B) := T(\mu_{\mathcal{M}}^{(k)})(B) := A_{\Theta \rightarrow X}(\mu_{\mathcal{M}}^{(k)}(B)) = \sum_{j=1}^J \sum_{i=1}^I \delta_{w_{ji}}(B) A_{\Theta \rightarrow X}(\nu_{ji}) = \sum_{j=1}^J \sum_{i=1}^I \delta_{w_{ji}}(B) \mu_{\mathcal{B}}(\overline{B}_{ji}) \quad (5.24)$$

The operator is bounded because

$$|T\mu_{\mathcal{M}}|_{\mathcal{B}}(\Omega) = \sup_{\{\hat{B}_{\bullet}\}} \sum_{h=1}^{|\{\hat{B}_{\bullet}\}|} \|A_{\Omega \rightarrow X}(\mu_{\mathcal{M}}(\hat{B}_h))\|_{\mathcal{B}} \leq \sup_{\{\hat{B}_{\bullet}\}} \sum_{h=1}^{|\{\hat{B}_{\bullet}\}|} \|\mu_{\mathcal{M}}(\hat{B}_h)\|_{\mathcal{M}(\Theta; \mathcal{V})} = |\mu_{\mathcal{M}}|_{\mathcal{M}}(\Omega) \quad (5.25)$$

with the suprema going over all the partitions  $\{\hat{B}_{\bullet}\}$  of  $\Omega$  and  $\mu_{\mathcal{M}} \in \mathcal{M}(\Omega; \mathcal{M}(\Theta; \mathcal{V}))$  is arbitrary.

We can estimate

$$\begin{aligned} \left\| \int_{\Omega} \phi(z, w) d\mu_{\mathcal{B}} - \int_{\Omega} \phi(z, w) d\mu_{\mathcal{B}}^{(k)} \right\|_{\mathcal{B}} &\leq \sum_{j=1}^J \sum_{i=1}^I \left\| \int_{\overline{B}_{ji}} \phi(z, w) d\mu_{\mathcal{B}} - \int_{\overline{B}_{ji}} \phi(z, w) d\mu_{\mathcal{B}}^{(k)} \right\|_{\mathcal{B}} \\ &= \sum_{j=1}^J \sum_{i=1}^I \left\| \int_{\overline{B}_{ji}} \phi(z, w) d\mu_{\mathcal{B}} - \int_{\overline{B}_{ji}} \phi(z, w_{ji}) d\mu_{\mathcal{B}} \right\|_{\mathcal{B}} \\ &\leq \sum_{j=1}^J \sum_{i=1}^I \left( \sup_{w, \tilde{w} \in \overline{B}_{ji}} |\phi(z, w) - \phi(z, \tilde{w})| \right) |\mu_{\mathcal{B}}|_{\mathcal{B}}(\overline{B}_{ji}) \\ &\leq \frac{1}{2k} |\mu_{\mathcal{B}}|_{\mathcal{B}}(\Omega) \end{aligned} \quad (5.26)$$

This shows that  $A_{\Omega \rightarrow Z} \mu_{\mathcal{B}}^{(k)} \rightarrow A_{\Omega \rightarrow Z} \mu_{\mathcal{B}}$  uniformly in  $z$ . Since  $M$  is a bounded linear operator on  $\mathcal{B}$ , we also have

$$M((A_{\Omega \rightarrow Z} \mu_{\mathcal{B}}^{(k)})(z)) \rightarrow M((A_{\Omega \rightarrow Z} \mu_{\mathcal{B}})(z)) \quad \text{uniformly in } z. \quad (5.27)$$

By continuity of  $L$  in the first argument and convergence of the regularization terms, it follows that

$$\frac{1}{N} \sum_{n=1}^N L(M((A_{\Omega \rightarrow Z} \mu_{\mathcal{B}})(z_n)), y_n) + \lambda |\mu_{\mathcal{B}}|_{\mathcal{B}}(\Omega) = \lim_{k \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N L(M((A_{\Omega \rightarrow Z} (T\mu_{\mathcal{M}}^{(k)}))(z_n)), y_n) + \lambda |\mu_{\mathcal{M}}^{(k)}|_{\mathcal{M}}(\Omega). \quad (5.28)$$

Moreover, by Theorem B.2, there exists a sequence of simple functions

$$\phi_k = \sum_{i,j} a_{ij}^{(k)} \mathbb{1}_{A_i^{(k)} \times B_j^{(k)}} \rightarrow \phi \quad (5.29)$$

uniformly, with pairwise disjoint  $A_i^{(k)}$ 's and  $B_j^{(k)}$ 's covering  $Z$  and  $\Omega$ , respectively. Then

$$\begin{aligned}
A_{\Omega \rightarrow Z}(T\mu_{\mathcal{M}})(z) &= \int_{\Omega} \phi(z, w) d(T\mu_{\mathcal{M}})(w) \\
&= \lim_{k \rightarrow \infty} \int_{\Omega} \sum_{i,j} a_{ij}^{(k)} \mathbb{1}_{A_i^{(k)} \times B_j^{(k)}}(z, w) d(T\mu_{\mathcal{M}})(w) \\
&= \lim_{k \rightarrow \infty} \sum_{i,j} a_{ij}^{(k)} \mathbb{1}_{A_i^{(k)}}(z) A_{\Theta \rightarrow X}(\mu_{\mathcal{M}}(B_j^{(k)})) \\
&= \lim_{k \rightarrow \infty} A_{\Theta \rightarrow X} \left( \sum_{i,j} a_{ij}^{(k)} \mathbb{1}_{A_i^{(k)}}(z) \mu_{\mathcal{M}}(B_j^{(k)}) \right) \\
&= \lim_{k \rightarrow \infty} A_{\Theta \rightarrow X} \left( \int_{\Omega} \sum_{i,j} a_{ij}^{(k)} \mathbb{1}_{A_i^{(k)} \times B_j^{(k)}}(z, w) d\mu_{\mathcal{M}}(w) \right) \\
&= A_{\Theta \rightarrow X} \left( \int_{\Omega} \phi(z, w) d\mu_{\mathcal{M}}(w) \right) \\
&= A_{\Theta \rightarrow X}((A_{\Omega \rightarrow Z}\mu_{\mathcal{M}})(z))
\end{aligned} \tag{5.30}$$

where the second equality follows from  $\phi_k \rightarrow \phi$  uniformly and the second-to-last inequality uses the continuity of  $A_{\Theta \rightarrow X}$ .

Combining this with (5.28) yields

$$\begin{aligned}
&\frac{1}{N} \left( \sum_{n=1}^N L(M((A_{\Omega \rightarrow Z}\mu_{\mathcal{B}})(z_n)), y_n) \right) + \lambda |\mu_{\mathcal{B}}|_{\mathcal{B}}(\Omega) \\
&= \lim_{k \rightarrow \infty} \frac{1}{N} \left( \sum_{n=1}^N L(M((A_{\Omega \rightarrow Z}(T\mu_{\mathcal{M}}^{(k)}))(z_n)), y_n) \right) + \lambda |\mu_{\mathcal{M}}^{(k)}|_{\mathcal{M}}(\Omega) \\
&= \lim_{k \rightarrow \infty} \frac{1}{N} \left( \sum_{n=1}^N L(M(A_{\Theta \rightarrow X}((A_{\Omega \rightarrow Z}\mu_{\mathcal{M}}^{(k)})(z_n)), y_n) \right) + \lambda |\mu_{\mathcal{M}}^{(k)}|_{\mathcal{M}}(\Omega) \\
&\geq \frac{1}{N} \left( \sum_{n=1}^N L(M(A_{\Theta \rightarrow X}((A_{\Omega \rightarrow Z}\mu_{\mathcal{M}}^{\dagger})(z_n)), y_n) \right) + \lambda |\mu_{\mathcal{M}}^{\dagger}|_{\mathcal{M}}(\Omega) \\
&= \frac{1}{N} \left( \sum_{n=1}^N L(M((A_{\Omega \rightarrow Z}(T\mu_{\mathcal{M}}^{\dagger}))(z_n)), y_n) \right) + \lambda |\mu_{\mathcal{M}}^{\dagger}|_{\mathcal{M}}(\Omega) \\
&\geq \frac{1}{N} \left( \sum_{n=1}^N L(M((A_{\Omega \rightarrow Z}\mu_{\mathcal{B}}^{\dagger})(z_n)), y_n) \right) + \lambda |\mu_{\mathcal{B}}^{\dagger}|_{\mathcal{B}}(\Omega)
\end{aligned} \tag{5.31}$$

where the first inequality follows from the optimality of  $\mu_{\mathcal{M}}^{\dagger}$ , and the second from the identity  $\mu_{\mathcal{B}}^{\dagger} = T\mu_{\mathcal{M}}^{\dagger}$  together with inequality (5.25). Hence, the above chain of inequalities establishes the optimality of  $\mu_{\mathcal{B}}^{\dagger}$ .  $\square$

Building on the previous theorem, we can now state a joint representer theorem for both formulations. In particular, the theorem shows that the optimization problems corresponding to the weight space and function space viewpoints admit a coinciding sparse solution.

**Theorem 5.2** (Representer theorem hypernetwork). *Consider either optimization problem (5.16) or (5.17), with the measurement operator  $M: \mathcal{B} \rightarrow \mathbb{R}^d$  defined in (5.14) with  $\{(v_j^{\diamond}, x_j)\}_{j=1}^d \subseteq \mathcal{V}^{\diamond} \times X$ . Assume we are given data  $\{(z_n, y_n)\}_{n=1}^N \subseteq Z \times \mathbb{R}^d$ . Let  $L: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$  be such that for every  $y \in \mathbb{R}^d$ , the map  $L(\cdot, y)$  is convex, coercive, and lower semicontinuous. Finally, let  $\mathcal{V}$  have a predual  $\mathcal{V}^{\diamond}$ , and let  $\Omega$  and  $\Theta$  be locally-compact Hausdorff spaces. Then the solutions have the form*

$$\mu_{\mathcal{M}}^{\dagger} = \sum_{m=1}^{Nd} \delta_{w_m} \delta_{\theta_m} v_m, \quad \mu_{\mathcal{B}}^{\dagger} = \sum_{m=1}^{Nd} \delta_{w_m} A_{\Theta \rightarrow X}(\delta_{\theta_m} v_m) \tag{5.32}$$

with  $v_m \in \mathcal{V}$ . Both measures yield the same function

$$f^\dagger(z) = \sum_{m=1}^{Nd} \phi(z, w_m) \psi(x, \theta_m) v_m \quad (5.33)$$

*Proof.* Define  $\widetilde{M}: \mathcal{M}(\Theta; \mathcal{V}) \rightarrow \mathbb{R}^d$  by  $\widetilde{M} := M \circ A_{\Theta \rightarrow X}$ . With this definition, optimization problem (5.16) turns into

$$\min_{\mu \in \mathcal{M}(\Omega; \mathcal{M}(\Theta; \mathcal{V}))} \frac{1}{N} \sum_{n=1}^N L(\widetilde{M}((A_{\Omega \rightarrow Z} \mu)(z_n)), y_n) + \lambda |\mu|_{\mathcal{M}(\Theta; \mathcal{V})}(\Omega) \quad (5.34)$$

Since  $(\mathcal{V}^\circ)^* = \mathcal{V}$  and  $\Theta$  is a locally-compact Hausdorff space, we have  $C_0(\Theta; \mathcal{V}^\circ)^* \cong \mathcal{M}(\Theta; \mathcal{V})$  by (4.9). Moreover, by construction of  $M$ ,

$$(\widetilde{M}\mu)_j = \langle v_j^\circ | (A_{\Theta \rightarrow X} \mu(x_j)) \rangle_{\mathcal{V}} = \langle \mu | \psi(x_j, \cdot) v_j^\circ \rangle_{C_0(\Theta; \mathcal{V}^\circ)} \quad (5.35)$$

where the final equality follows from the definition of  $\langle \cdot | \cdot \rangle_{C_0(\Theta; \mathcal{V}^\circ)}$  as the continuous extension of (4.17) by uniform approximation with simple functions.

This shows that the measurement operator is of the form required for Theorem 4.4. The desired representation of the solutions, therefore, follows directly from it.  $\square$

**Remark 5.3.** Note that the structure of the representer theorem resembles the DeepONet in (5.5). In particular, assume  $Z \subseteq \mathbb{R}^{dz}$  and  $X \subseteq \mathbb{R}^{dx}$  compact,  $\Omega \subseteq \mathbb{R}^{dz} \times \mathbb{R}$  and  $\Theta \subseteq \mathbb{R}^{dx} \times \mathbb{R}$  not necessarily compact. As in the neural vv-RKBS setting introduced in Definition 4.3, we define

$$\phi \in C_0(Z \times \Omega), \quad \phi(z, (\omega, b)) := \sigma_\phi(\langle \omega | z \rangle_{\ell_2} + b) \beta_\phi((\omega, b)), \quad (5.36)$$

and

$$\psi \in C_0(X \times \Theta), \quad \psi(x, (\theta, c)) := \sigma_\psi(\langle \theta | x \rangle_{\ell_2} + c) \beta_\psi((\theta, c)), \quad (5.37)$$

for measurable activation functions  $\sigma_\phi, \sigma_\psi: \mathbb{R} \rightarrow \mathbb{R}$  and measurable positive functions  $\beta_\phi, \beta_\psi: \Omega \rightarrow \mathbb{R}$ .

In this setting, the hypernetwork provided by the representer theorem takes the form

$$\begin{aligned} f^\dagger(z)(x) &= \sum_{m=1}^{Nd} \sigma_\phi(\langle \omega_m | z \rangle_{\ell_2} + b_m) \beta_\phi((\omega_m, b_m)) \sigma_\psi(\langle \theta_m | x \rangle_{\ell_2} + c_m) \beta_\psi((\theta_m, c_m)) v_m \\ &= \sum_{m=1}^{Nd} \sigma_\phi(\langle \omega_m | z \rangle_{\ell_2} + b_m) \sigma_\psi(\langle \theta_m | x \rangle_{\ell_2} + c_m) v_m \end{aligned} \quad (5.38)$$

where the  $\beta$  terms are absorbed in the  $v_m$ 's.

Here, the  $\sigma_\phi(\langle \omega_m | z \rangle_{\ell_2} + b_m)$  parts determine the coefficients and the  $\sigma_\psi(\langle \theta_m | x \rangle_{\ell_2} + c_m) v_m$  parts serve as the basis functions in the DeepONet formulation.  $\diamond$

### 5.2.3 Discussion on function space formulation

While we have adopted Definitions 5.1 and 5.2 as the function spaces for both DeepONets and hypernetworks, we could also consider alternative formulations of the same spaces or consider completely different function spaces.

For an example of the former, assume  $\mathcal{V} = \mathbb{R}$  for simplicity. In the situation of Definitions 5.1 and 5.2 with an integral RKBS pair  $(\mathcal{B}, \mathcal{B}^\circ)$  with kernel  $\psi: X \times \Theta \rightarrow \mathbb{R}$ , consider an integral RKBS setup with measures  $\xi \in \mathcal{M}(\Omega \times \Theta)$  representing functions  $h: Z \times X \mapsto \mathbb{R}$  by

$$h(z, x) = \int_{\Omega \times \Theta} \phi(z, w) \psi(x, \theta) d\xi(w, \theta). \quad (5.39)$$

By disintegration (see Theorem 10.4.8 in Bogachev [10] for an adequate version for signed measures) through the projection onto the first component in  $\Omega \times \Theta$ , there is a bijection between  $\xi \in \mathcal{M}(\Omega \times \Theta)$  and  $\mu \in \mathcal{M}(\Omega; \mathcal{M}(\Theta))$ . This makes both representations coincide, that is  $[f(z)](x) = [A_{\Theta \rightarrow X}((A_{\Omega \rightarrow Z} \mu)(z))](x) = h(z, x)$  with  $A_{\Theta \rightarrow X}$  defined as in (4.10) and  $A_{\Omega \rightarrow Z}$  as defined in (5.13). However, the constructions of Definitions 5.1 and 5.2 lead to a clear distinction between the data points  $z_i \in Z$  and observation points  $x_j \in X$  for the measurement

operator  $M$  in the formulation of the optimization problems and the resulting Theorem 5.2. We argue that this advantage makes them preferable to working directly with representations in the product space of parameters  $\Omega \times \Theta$ .

Such a correspondence also has parallels with the RKHS conditional mean embedding as introduced in a measure-theoretic framework in Definition 3.1 of Park and Muandet [40]. Specifically, consider the underlying topological space  $Z \times X$ , a Borel probability measure  $\xi \in \mathcal{P}(Z \times X)$ , the projections  $\Pi_Z : Z \times X \rightarrow Z$  and  $\Pi_X : Z \times X \rightarrow X$ , and  $\mathcal{H}_Z, \mathcal{H}_X$  two RKHS with kernels  $K_Z : Z \times Z \rightarrow \mathbb{R}$  and  $K_X : X \times X \rightarrow \mathbb{R}$ . Then, on the one hand we have the unconditional mean embedding

$$\mathbb{E}_\xi [K_Z(\Pi_Z, \cdot) K_X(\Pi_X, \cdot)], \quad (5.40)$$

which is analogous to the representation (5.39) and produces an element of the RKHS over  $Z \times X$  with kernel  $((z_1, x_1), (z_2, x_2)) \mapsto K_Z(z_1, z_2) K_X(x_1, x_2)$ . Note that this kernel is positive semi-definite and its associated RKHS is linearly isometric to the tensor product  $\mathcal{H}_Z \otimes \mathcal{H}_X$ , see Theorem 5.11 of Paulsen and Raghupathi [41]. On the other hand, conditioning with respect to the random variable  $\Pi_Z$  gives rise to the conditional mean embedding

$$\mathbb{E}_\xi [K_X(\Pi_X, \cdot) | \Pi_Z]. \quad (5.41)$$

This conditional expectation is a random variable  $Z \rightarrow \mathcal{H}_X$ , similarly to the function-space view of Definition 5.2 in which we consider maps  $Z \rightarrow \mathcal{B}$ . However, unlike for (5.40), in (5.41) there is no kernel involving the variable  $z \in Z$ , since this quantity doesn't measure a relation with other points in  $Z$  and it also does not take into account to the effect of parameters  $w \in \Omega$ .

While the previous discussion focused on reformulating Definitions 5.1 and 5.2 in an equivalent way and relating this alternative to mean embeddings, one could also consider entirely different function spaces. For instance, assume  $(\mathcal{B}, \mathcal{B}^\circ)$  is an integral vv-RKBS with  $\mathcal{B} \subset \{u : X \rightarrow \mathcal{V}\}$ ,  $\mathcal{B}^\circ \subset \{v : \Theta \rightarrow \mathcal{V}^\circ\}$  and kernel  $\psi : X \times \Theta \rightarrow \mathbb{R}$ . For a parametrized family of Borel maps  $F_z : \Omega \rightarrow \Theta$  indexed by  $z \in Z$ , one can define a pushforward-induced space as

$$\mathcal{B}_F := \left\{ r : Z \rightarrow \mathcal{B} \mid [r(z)](\cdot) = \int_\Theta \psi(\cdot, \theta) d[(F_z)_\# \nu](\theta) = \int_\Omega \psi(\cdot, F_z(w)) d\nu(w) \text{ for some } \nu \in \mathcal{M}(\Omega; \mathcal{V}) \right\}, \quad (5.42)$$

and endow it with the norm

$$\|r\|_{\mathcal{B}_F} := \inf \left\{ |\nu|_{\mathcal{V}}(\Omega) \mid \nu \in \mathcal{M}(\Omega; \mathcal{V}) \text{ satisfies } [r(z)](\cdot) = \int_\Theta \psi(\cdot, \theta) d[(F_z)_\# \nu](\theta) \text{ for all } z \in Z \right\}. \quad (5.43)$$

It would be tempting to come up with a formalization of this type for the concept of hypernetworks, with the family  $F_z$  determining the weights appearing in the  $\mathcal{B}$  representation. In fact, the set  $\mathcal{B}_F$  with the norm  $\|\cdot\|_{\mathcal{B}_F}$  is a vector-valued reproducing kernel Banach space. To see that point evaluation is bounded, first, we notice that

$$|(F_z)_\# \nu|_{\mathcal{V}}(\Theta) \leq |\nu|_{\mathcal{V}}(\Omega) \quad \text{for all } \nu \in \mathcal{M}(\Omega; \mathcal{V}) \text{ and all } z \in Z. \quad (5.44)$$

This follows directly from the definition of the total variations of  $\nu$  and  $(F_z)_\# \nu$ , noticing that any partition of  $\Theta = \bigcup_{i=1}^\infty \Theta_n$  induces a partition of  $\Omega$  as  $\Omega = \bigcup_{i=1}^\infty F_z^{-1}(\Theta_n)$ . Using (5.44) and that  $\mathcal{B}$  is part of an integral vv-RKBS pair, we have

$$\|r(z)\|_{\mathcal{B}} = \inf \left\{ |\mu|_{\mathcal{V}}(\Theta) \mid \mu \in \mathcal{M}(\Theta; \mathcal{V}) \text{ satisfies } [r(z)](\cdot) = \int_\Theta \psi(\cdot, \theta) d\mu(\theta) \right\} \leq \|r\|_{\mathcal{B}_F}. \quad (5.45)$$

However, while Theorem 3.4 guarantees the existence of an adjoint space  $\mathcal{B}_F^\circ$ , there is no clear choice of  $\mathcal{B}_F^\circ$  that makes  $(\mathcal{B}_F, \mathcal{B}_F^\circ)$  an integral vv-RKBS, even if the maps  $F_z$  arise from another integral RKBS. We attribute this to the fact that this construction, due to the composition with the functions  $F_z$ , is closer to a two-layer architecture than a single-layer one. Given these drawbacks, we argue that a more suitable generalization to deep networks would be a combination of the single-layer hypernetwork formalism of Definitions 5.1 and 5.2 with a principled approach dealing with deep architectures, such as the reproducing kernel chains proposed in Heeringa et al. [24].

## 6 Conclusion

Mathematical analysis of neural networks helps reveal their underlying properties. In particular, studying the function spaces associated with neural networks provides insight into, for instance, how optimization within

these spaces yields neural architectures, which algorithms are suited for solving the optimization problem, and what types of generalization guarantees can be given. Many of these neural network function spaces are instances of reproducing kernel Banach spaces (RKBS). While scalar-valued neural networks are well understood within this framework, the theory for  $\mathbb{R}^d$ -valued neural networks and neural operators is less developed.

To address this gap, we introduce a kernel-based framework for vector-valued RKBS (vv-RKBS) that extends existing constructions. Most known RKBS models for neural networks do not explicitly employ a kernel and instead only rely on the assumption that the point evaluation functionals are bounded. Among the methods that do incorporate a kernel, it is typically assumed to be given a priori.

We propose a new definition of an adjoint pair of vv-RKBS that explicitly incorporates a kernel. Within this framework, we show that every Banach space of functions with bounded point evaluations corresponds to such an adjoint pair, and hence admits an associated kernel. The adjoint pair consists of a dual pair  $(\mathcal{B}, \mathcal{B}^\circ)$  of Banach spaces of functions, both of which have bounded point evaluations. By combining this duality pairing with the kernel, we obtain a reproducing property and thereby complete the construction of the adjoint pair of vv-RKBS.

In this setting, the kernel is defined as

$$K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\circ) \quad (6.1)$$

where  $(\mathcal{U}, \mathcal{U}^\circ)$  is a dual pair. Here,  $X$  and  $\mathcal{U}$  are the domain and codomain of functions in  $\mathcal{B}$ , while  $\Omega$  and  $\mathcal{U}^\circ$  are the domain and codomain of functions in  $\mathcal{B}^\circ$ . The space  $\text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$  generalizes  $\mathcal{L}(\mathcal{U})$  as used in the vv-RKHS construction. This kernel definition generalizes existing constructions by allowing asymmetric kernel domains, Banach space-valued outputs that may be infinite-dimensional, and not imposing structural assumptions such as reflexivity or separability on any of the involved spaces. The absence of reflexivity is important, since the neural network function spaces of interest are non-reflexive.

Within the adjoint pair of vv-RKBS framework, we show that many well-known properties of vv-RKHS naturally extend to our setting. In particular, every suitable kernel corresponds to an adjoint pair of vv-RKBS given sufficient structure on the dual pair  $(\mathcal{U}, \mathcal{U}^\circ)$ , and every adjoint pair of vv-RKBS is equivalent to a scalar adjoint pair of vv-RKBS.

To gain insight into neural networks, we further specialize the adjoint pair construction to integral and neural vv-RKBS, which are the function spaces most relevant to neural networks. We analyze their associated kernel structure and establish a general representer theorem. We then apply the integral and neural vv-RKBS framework to both  $\mathbb{R}^d$ -valued neural networks and neural operators. For the latter, our focus is on DeepONets and hypernetworks, which have the implicit neural representation (INR) structure underlying many neural operator approaches. We show that the hypernetwork space, which includes DeepONets, admits two complementary interpretations: a weight-space view and a function-space view. Both viewpoints lead to specific optimization problems that share a joint representer theorem, and thus yield a coinciding sparse solution.

Altogether, this work provides a unifying RKBS viewpoint on  $\mathbb{R}^d$ -valued neural networks and neural operators. It highlights how Banach-space structures can capture cases beyond Hilbert theory while retaining kernel-based tools, thereby opening the door to deeper mathematical understanding of modern neural network and operator learning methods.

## 7 Future work

Our work considers a general setting and mainly focuses on shallow neural networks rather than deep networks and operators. However, many state-of-the-art architectures are inherently deep. In Heeringa et al. [24], chaining kernels yields a function space for deep neural networks. Since our work introduces vv-RKBS kernels, it would be natural to explore similar ideas in our setting, potentially leading to a notion of vv-RKBS kernel chaining.

Another direction for future work is related to representer theorems. In the vv-RKHS setting, we study functions  $f: X \rightarrow \mathcal{U}$  taking values in a Hilbert space  $\mathcal{U}$ . Given data  $\{(x_n, v_n)\}_{n=1}^N \subseteq X \times \mathcal{U}$  and a kernel

$$K: X \times X \rightarrow \mathcal{L}(\mathcal{U}) \quad (7.1)$$

the representer theorem ensures that the solution to the supervised optimization problem takes the form

$$f^\dagger(x) = \sum_{n=1}^N K(\cdot, x_n) u_n, \quad u_n \in \mathcal{U} \quad (7.2)$$

The vv-RKBS setting yields a weaker result. For integral and neural vv-RKBSs, we have a dual pair of Banach spaces  $(\mathcal{B}, \mathcal{B}^\circ)$  and  $(\mathcal{U}, \mathcal{U}^\circ)$ , where the space  $\mathcal{B}$  consists of functions  $f: X \rightarrow \mathcal{U}$ , and  $\mathcal{B}^\circ$  consists of functions  $g: \Omega \rightarrow \mathcal{U}^\circ$ . The corresponding kernel is

$$K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\circ) \quad (7.3)$$

In this setting, instead of direct data  $\{(x_n, v_n)\}_{n=1}^N \subseteq X \times \mathcal{U}$ , we consider data  $\{(x_n, y_n)\}_{n=1}^N \subseteq X \times \mathbb{R}^d$ . Here, the  $y_n$  are related to the  $v_n$ 's through a measurement operator  $M: \mathcal{U} \rightarrow \mathbb{R}^d$  as  $y_n = Mv_n$ . Hence, the outputs are always mapped to finite-dimensional measurements for fitting.

For optimization problems over  $f \in \mathcal{B}$  with such data  $\{(x_n, y_n)\}_{n=1}^N \subseteq X \times \mathbb{R}^d$ , the representer theorem then only guarantees representations of the form

$$f^\dagger(x) = \sum_{m=1}^{Nd} K_{\mathcal{U}}(\cdot, w_m) u_m \quad (7.4)$$

Hence, the number of terms is not necessarily bounded by  $N$ . As argued in this paper, we postulate that this limitation arises because the regularizers used so far do not promote groupwise sparsity, unlike, for instance, the structured sparsity regularizers studied by Bach et al. [3]. Achieving such improvements would likely require a more explicit use of the vv-RKBS structure rather than general convexity arguments in topological vector spaces as in Bredies and Carioni [12], which in turn must rely on the use of finite-dimensional measurements.

Moreover, our representer theorems are tailored to integral and neural vv-RKBSs. Similar to the work of Wang et al. [50], which investigates when scalar RKBSs admit a representer theorem, it would be interesting to study under what conditions our (adjoint pair of) vv-RKBSs admit one. In particular, an important open question is what additional structure is needed to ensure representations with at most  $N$  terms in the expansion. We expect that the groupwise sparsity mentioned above could play a central role.

## A Notation

| <i>Notation</i>                                                                                                                                                 | <i>Meaning</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| $ \cdot _B$                                                                                                                                                     | Given a Banach space $B$ with norm $\ \cdot\ _B$ , $ \cdot _B$ denotes the total variation measure of a $B$ -valued measure.                                                                                                                                                                                                                                                                                                                                                                                              |
| $\Omega$                                                                                                                                                        | Weight space of adjoint vv-RKBS pair of functions. Also the weights of the hypernetwork.                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| $X$                                                                                                                                                             | The input space of adjoint vv-RKBS pair of functions. Also used as the input of the base network in the hypernetwork setting.                                                                                                                                                                                                                                                                                                                                                                                             |
| $Z$                                                                                                                                                             | The input space of the hypernetwork.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| $\Theta$                                                                                                                                                        | The weight space of the base network in the hypernetwork setting.                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| $\mathcal{U}$                                                                                                                                                   | The output Banach space of a vv-RKBS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| $\mathcal{U}^\circ$                                                                                                                                             | The output Banach space of the adjoint vv-RKBS in an adjoint vv-RKBS pair.                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| $\mathcal{B}$                                                                                                                                                   | A vv-RKBS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| $\mathcal{B}^\circ$                                                                                                                                             | The adjoint vv-RKBS in an adjoint vv-RKBS pair.                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| $C_0(X; \mathcal{U})$                                                                                                                                           | Assuming $X$ is a locally-compact topological space, $f \in C_0(X; \mathcal{U})$ if it is continuous and for any positive number $\varepsilon > 0$ there exists a compact subset $K \subseteq X$ such that $\ f(x)\  \leq \varepsilon$ for all $x \in X \setminus K$ .                                                                                                                                                                                                                                                    |
| $\mathcal{M}(\Omega)$                                                                                                                                           | The set of regular signed Radon measures with finite total variation over $\Omega$ equipped with the Borel $\sigma$ -algebra.                                                                                                                                                                                                                                                                                                                                                                                             |
| $\mathcal{M}(\Omega; \mathcal{V})$                                                                                                                              | The set of regular countably additive $\mathcal{V}$ -valued vector measures with finite total variation over $\Omega$ equipped with the Borel $\sigma$ -algebra. By definition, $\mu \in \mathcal{M}(\Omega; \mathcal{V})$ is regular if and only if $ \mu _{\mathcal{V}}$ is regular.                                                                                                                                                                                                                                    |
| $\phi: X \times \Omega \rightarrow \mathbb{R}$                                                                                                                  | the integrand for defining the integral / neural vv-RKBS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| $A_{\Omega \rightarrow X}$                                                                                                                                      | $A_{\Omega \rightarrow X} \mu = \int_{\Omega} \phi(\cdot, w) d\mu(w)$ with $\mu \in \mathcal{M}(\Omega; \mathcal{U})$ . Hence, it means integrating out the $\Omega$ part to get a function of $x \in X$ . The Banach space $\mathcal{U}$ is arbitrary.                                                                                                                                                                                                                                                                   |
| $\mathcal{L}(\mathcal{U}; \mathcal{V}) / \mathcal{L}(\mathcal{U})$                                                                                              | The spaces of bounded linear operators from $\mathcal{U}$ to $\mathcal{V}$ and from $\mathcal{U}$ to $\mathcal{U}$ , respectively.                                                                                                                                                                                                                                                                                                                                                                                        |
| $f$ and $g$                                                                                                                                                     | functions from the vv-RKBS and the adjoint vv-RKBS, respectively.                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| $\langle \cdot   \cdot \rangle_{B^p} = \langle \cdot   \cdot \rangle_{B^d, B^p}$                                                                                | the dual pairing for a dual pair of Banach spaces $(B^p, B^d)$ . Such pairings appear both for explicitly prescribed pairs $(B, B^\circ)$ used in the RKBS definitions as well as for the canonical pair $(B, B^*)$ with the continuous dual $B^*$ . In case $B$ is a Hilbert space, the pairing can always be considered as the inner product. The abbreviated notation is used when the dual space $B^d$ is clear from the context, with the subscript denoting the space of the primal element in the second argument. |
| $\mathcal{N}(A)$                                                                                                                                                | denotes the nullspace of a linear operator $A$ .                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| $\text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$                                                                                                                   | the space of twin operators for a dual pair of Banach spaces $(\mathcal{U}, \mathcal{U}^\circ)$ .                                                                                                                                                                                                                                                                                                                                                                                                                         |
| $T_{\mathcal{U}}$ and $T_{\mathcal{U}^\circ}$                                                                                                                   | the $T_{\mathcal{U}}$ and $T_{\mathcal{U}^\circ}$ operators used in the definition of the space of twin operators.                                                                                                                                                                                                                                                                                                                                                                                                        |
| $K: X \times \Omega \rightarrow \text{Twin}(\mathcal{U}, \mathcal{U}^\circ)$                                                                                    | the kernel corresponding to an adjoint pair of vv-RKBS.                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| $K_{\mathcal{U}}: X \times \Omega \rightarrow \mathcal{L}(\mathcal{U})$ and $K_{\mathcal{U}^\circ}: X \times \Omega \rightarrow \mathcal{L}(\mathcal{U}^\circ)$ | the linear operators used in the definition of the space of twin operators corresponding to the twin operator $K(x, w)$ .                                                                                                                                                                                                                                                                                                                                                                                                 |
| $M: \mathcal{U} \rightarrow \mathbb{R}^d$                                                                                                                       | Measurement operator used in supervised learning problems.                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

Table 1: Meaning of frequently used mathematical objects.

## B Additional proofs and theorems

**Theorem B.1.** Let  $\Sigma_X$  be a  $\sigma$ -algebra on  $X$  and  $\Sigma_\Omega$  a  $\sigma$ -algebra on  $\Omega$ . Let  $(\mathcal{U}, \mathcal{U}^\circ)$  be a dual pair of Banach spaces with a continuous dual pairing; that is, there exists a  $C > 0$  such that

$$|\langle u^\circ | u \rangle_{\mathcal{U}}| \leq C \|u^\circ\|_{\mathcal{U}^\circ} \|u\|_{\mathcal{U}}. \quad (\text{B.1})$$

Then for  $\mu \in \mathcal{M}(\Omega; \mathcal{U})$  and  $\rho \in \mathcal{M}(X; \mathcal{U}^\circ)$ , there exists a unique measure  $\langle \rho | \mu \rangle_{\mathcal{U}} \in \mathcal{M}(X \times \Omega)$  defined over the  $\sigma$ -algebra  $\sigma(\Sigma_X \times \Sigma_\Omega)$  generated by  $\Sigma_X \times \Sigma_\Omega$  such that

$$\langle \rho | \mu \rangle_{\mathcal{U}}(A \times B) = \langle \rho(A) | \mu(B) \rangle_{\mathcal{U}}, \quad \text{for all } A \in \Sigma_X, B \in \Sigma_\Omega. \quad (\text{B.2})$$

Moreover,  $\langle \rho | \mu \rangle_{\mathcal{U}}$  is of bounded variation, with

$$|\langle \rho | \mu \rangle_{\mathcal{U}}|(X \times \Omega) \leq C |\rho|_{\mathcal{U}^\circ}(X) |\mu|_{\mathcal{U}}(\Omega). \quad (\text{B.3})$$

*Proof.* Let  $\mathcal{R}$  denote the algebra generated by sets of the form  $A \times B$  with  $A \in \Sigma_X$  and  $B \in \Sigma_\Omega$ . Elements of  $\mathcal{R}$  are finite disjoint unions of such rectangles, i.e., sets of the form  $\bigcup_{i=1}^n A_i \times B_i$  with  $A_i \times B_i \in \Sigma_X \times \Sigma_\Omega$  pairwise disjoint and  $n < \infty$ . We define a set function on  $\mathcal{R}$  by

$$\langle \rho | \mu \rangle_{\mathcal{U}} \left( \bigcup_{i=1}^n A_i \times B_i \right) = \sum_{i=1}^n \langle \rho | \mu \rangle_{\mathcal{U}}(A_i \times B_i) = \sum_{i=1}^n \langle \rho(A_i) | \mu(B_i) \rangle_{\mathcal{U}} \quad (\text{B.4})$$

which is well-defined independently of the chosen pairwise disjoint decomposition.

Before applying the Hahn-Kolmogorov extension theorem, we show that  $\langle \rho | \mu \rangle_{\mathcal{U}}$  is countably additive on  $\mathcal{R}$ . First, consider a rectangle  $A \times B$  expressed as a countable union of pairwise disjoint rectangles  $A_i \times B_i$ . Using the partial unions  $A_n := \bigcup_{i=1}^n A_i$  and  $B_n := \bigcup_{i=1}^n B_i$ , we can decompose  $A \times B$  as

$$A \times B = (A_n \times B_n) \cup (A_n \times (B \setminus B_n)) \cup ((A \setminus A_n) \times B_n) \cup ((A \setminus A_n) \times (B \setminus B_n)). \quad (\text{B.5})$$

This is a decomposition into disjoint sets. Then consider

$$\left| \left( \sum_{i=1}^n \langle \rho | \mu \rangle_{\mathcal{U}}(A_i \times B_i) \right) - \langle \rho | \mu \rangle_{\mathcal{U}}(A \times B) \right| \quad (\text{B.6})$$

and note that using (B.5) it can be rewritten to

$$\begin{aligned} &= |\langle \rho | \mu \rangle_{\mathcal{U}}(A_n \times (B \setminus B_n)) + \langle \rho | \mu \rangle_{\mathcal{U}}((A \setminus A_n) \times B_n) + \langle \rho | \mu \rangle_{\mathcal{U}}((A \setminus A_n) \times (B \setminus B_n))| \\ &= |\langle \rho(A_n) | \mu(B \setminus B_n) \rangle_{\mathcal{U}} + \langle \rho(A \setminus A_n) | \mu(B_n) \rangle_{\mathcal{U}} + \langle \rho(A \setminus A_n) | \mu(B \setminus B_n) \rangle_{\mathcal{U}}| \\ &\leq C (\|\rho(A_n)\|_{\mathcal{U}^\circ} \|\mu(B \setminus B_n)\|_{\mathcal{U}} + \|\rho(A_n)\|_{\mathcal{U}^\circ} \|\mu(B \setminus B_n)\|_{\mathcal{U}} + \|\rho(A \setminus A_n)\|_{\mathcal{U}^\circ} \|\mu(B \setminus B_n)\|_{\mathcal{U}}) \end{aligned} \quad (\text{B.7})$$

By countable additivity of  $\rho$  and  $\mu$ , we have  $\rho(A_n) \rightarrow \rho(A)$  and  $\mu(B_n) \rightarrow \mu(B)$ , which shows that each bounding term in (B.7) goes to zero. Hence,  $\langle \rho | \mu \rangle_{\mathcal{U}}$  is countably additive on rectangles  $A \times B$ .

To extend countable additivity to all of  $\mathcal{R}$ , let  $C = \bigcup_{n=1}^\infty D_n \in \mathcal{R}$  be a countable union of disjoint sets  $D_n \in \mathcal{R}$ . As any set in  $\mathcal{R}$  can be written as a finite disjoint union of rectangles, the sets  $C$  and  $D_n$  can be written as  $C = \bigcup_{j=1}^N C_j$  and  $D_n = \bigcup_{i=1}^{M_n} D_{n,i}$  with both  $\{C_j\}_{j=1}^N$  and  $\{D_{n,i}\}_{i=1}^{M_n}$  finite sets of disjoint rectangles. Given that the  $D_n$  are pairwise disjoint, define the pairwise disjoint rectangles  $D_{n,i,j} := D_{n,i} \cap C_j$ . Then

$$\begin{aligned} \langle \rho | \mu \rangle_{\mathcal{U}}(C) &= \sum_{j=1}^N \langle \rho | \mu \rangle_{\mathcal{U}}(C_j) = \sum_{j=1}^N \sum_{n=1}^\infty \sum_{i=1}^{M_n} \langle \rho | \mu \rangle_{\mathcal{U}}(D_{n,i,j}) = \sum_{n=1}^\infty \sum_{i=1}^{M_n} \sum_{j=1}^N \langle \rho | \mu \rangle_{\mathcal{U}}(D_{n,i,j}) \\ &= \sum_{n=1}^\infty \sum_{i=1}^{M_n} \langle \rho | \mu \rangle_{\mathcal{U}}(D_{n,i}) = \sum_{n=1}^\infty \langle \rho | \mu \rangle_{\mathcal{U}}(D_n), \end{aligned} \quad (\text{B.8})$$

which proves countable additivity on  $\mathcal{R}$ .

Now we construct the extension. Let

$$\nu_{\mathcal{R}}^+ = 0.5 (|\langle \rho | \mu \rangle_{\mathcal{U}}| + \langle \rho | \mu \rangle_{\mathcal{U}}), \quad \nu_{\mathcal{R}}^- = 0.5 (|\langle \rho | \mu \rangle_{\mathcal{U}}| - \langle \rho | \mu \rangle_{\mathcal{U}}), \quad (\text{B.9})$$

with  $|\langle \rho | \mu \rangle_{\mathcal{U}}|$  the total variation on  $\mathcal{R}$ . These are well-defined and finite by the following total variation estimate,

$$\begin{aligned}
|\langle \rho | \mu \rangle_{\mathcal{U}}|(X \times \Omega) &= \sup_{\{A_{\bullet} \times B_{\bullet}\}} \sum_i |\langle \rho | \mu \rangle_{\mathcal{U}}(A_i \times B_i)| \\
&= \sup_{\{A_{\bullet} \times B_{\bullet}\}} \sum_i |\langle \rho(A_i) | \mu(B_i) \rangle_{\mathcal{U}}| \\
&\leq C \sup_{\{A_{\bullet} \times B_{\bullet}\}} \sum_i \|\rho(A_i)\|_{\mathcal{U}^{\circ}} \|\mu(B_i)\|_{\mathcal{U}} \\
&\leq C \sup_{\{A_{\bullet} \times B_{\bullet}\}} \sum_i |\rho|_{\mathcal{U}^{\circ}}(A_i) |\mu|_{\mathcal{U}}(B_i) \\
&= C \sup_{\{A_{\bullet} \times B_{\bullet}\}} \sum_i (|\rho|_{\mathcal{U}^{\circ}} \otimes |\mu|_{\mathcal{U}})(A_i \times B_i) \\
&= C (|\rho|_{\mathcal{U}^{\circ}} \otimes |\mu|_{\mathcal{U}})(X \times \Omega) \\
&= C |\rho|_{\mathcal{U}^{\circ}}(X) |\mu|_{\mathcal{U}}(\Omega)
\end{aligned} \tag{B.10}$$

where the suprema run over all the partitions  $\{A_{\bullet} \times B_{\bullet}\}$  of  $X \times \Omega$ . Moreover, since  $\langle \rho | \mu \rangle_{\mathcal{U}}$  is countably additive and of bounded variation, its total variation  $|\langle \rho | \mu \rangle_{\mathcal{U}}|$  is countably additive on  $\mathcal{R}$ . All together,  $\nu_{\mathcal{R}}^{+}$  and  $\nu_{\mathcal{R}}^{-}$  are nonnegative, finite, countably additive set functions on  $\mathcal{R}$ , and by the Hahn-Kolmogorov extension theorem (see Theorem 1.7.8 in Tao [49]), these extend to countable additive measures  $\nu^{+}$  and  $\nu^{-}$  on  $\sigma(\Sigma_X \times \Sigma_{\Omega})$ . Defining  $\nu = \nu^{+} - \nu^{-}$ , we obtain a finite signed measure such that

$$\nu(A \times B) = \langle \rho(A) | \mu(B) \rangle_{\mathcal{U}}, \quad \text{for all } A \in \Sigma_X, B \in \Sigma_{\Omega}. \tag{B.11}$$

Uniqueness of the extension  $\nu$  follows from Dynkin's lemma. If  $\tilde{\nu}$  is another extension, define

$$\mathcal{D} := \{S \in \sigma(\Sigma_X \times \Sigma_{\Omega}) \mid \tilde{\nu}(S) = \nu(S)\}. \tag{B.12}$$

Note that  $X \times \Omega \in \mathcal{D}$ . Since  $\nu$  is finite, for any  $A \subseteq B$  with  $A, B \in \mathcal{D}$ , we have

$$\nu(B \setminus A) = \nu(B) - \nu(A) = \tilde{\nu}(B) - \tilde{\nu}(A) = \tilde{\nu}(B \setminus A) \tag{B.13}$$

which shows that  $B \setminus A \in \mathcal{D}$ . Next, consider an increasing sequence  $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$  with  $A_n \in \mathcal{D}$ . Define

$$\tilde{A}_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i = A_n \setminus A_{n-1}, \quad \tilde{A}_1 = A_1. \tag{B.14}$$

Then  $\tilde{A}_1 = A_1 \in \mathcal{D}$  and by

$$\nu(\tilde{A}_n) = \nu(A_n) - \nu(A_{n-1}) = \tilde{\nu}(A_n) - \tilde{\nu}(A_{n-1}) = \tilde{\nu}(\tilde{A}_n) \tag{B.15}$$

we have  $\tilde{A}_n \in \mathcal{D}$  for  $n \geq 2$ . By countable additivity, we have

$$\nu\left(\bigcup_{n=1}^{\infty} A_n\right) = \nu\left(\bigcup_{n=1}^{\infty} \tilde{A}_n\right) = \sum_{n=1}^{\infty} \nu(\tilde{A}_n) = \sum_{n=1}^{\infty} \tilde{\nu}(\tilde{A}_n) = \tilde{\nu}\left(\bigcup_{n=1}^{\infty} \tilde{A}_n\right) = \tilde{\nu}\left(\bigcup_{n=1}^{\infty} A_n\right) \tag{B.16}$$

These properties verify that  $\mathcal{D}$  is a Dynkin system containing the  $\pi$ -system  $\Sigma_X \times \Sigma_{\Omega}$ . Hence, by Dynkin's Lemma

$$\sigma(\Sigma_X \times \Sigma_{\Omega}) \subseteq \mathcal{D} \subseteq \sigma(\Sigma_X \times \Sigma_{\Omega}) \implies \mathcal{D} = \sigma(\Sigma_X \times \Sigma_{\Omega}) \tag{B.17}$$

It follows that  $\tilde{\nu} = \nu$ , and hence  $\nu$  is the unique extension.

Finally, we show that  $\nu$  is of bounded variation:

$$\begin{aligned}
|\nu|(X \times \Omega) &= |\nu^{+} - \nu^{-}|(X \times \Omega) \leq |\nu^{+}|(X \times \Omega) + |\nu^{-}|(X \times \Omega) = \nu^{+}(X \times \Omega) + \nu^{-}(X \times \Omega) \\
&= 0.5 (|\langle \rho | \mu \rangle_{\mathcal{U}}|(X \times \Omega) + \langle \rho | \mu \rangle_{\mathcal{U}}(X \times \Omega)) + 0.5 (|\langle \rho | \mu \rangle_{\mathcal{U}}|(X \times \Omega) - \langle \rho | \mu \rangle_{\mathcal{U}}(X \times \Omega)) \\
&= |\langle \rho | \mu \rangle_{\mathcal{U}}|(X \times \Omega) \leq C |\rho|_{\mathcal{U}^{\circ}}(X) |\mu|_{\mathcal{U}}(\Omega)
\end{aligned} \tag{B.18}$$

where the final inequality follows from (B.10).  $\square$

**Lemma B.1.** *Let  $X$  and  $\Omega$  be topological spaces and  $\phi \in C_0(X \times \Omega)$ . Then for any  $\delta > 0$  and  $\varepsilon > 0$ , there exist:*

- pairwise disjoint Borel sets  $A_1, \dots, A_n \subseteq X$  with  $\bigcup_{i=1}^n A_i = X$ ,
- pairwise disjoint Borel sets  $B_1, \dots, B_m \subseteq \Omega$  with  $\bigcup_{j=1}^m B_j = \Omega$ ,
- a Borel set  $D_\delta \subseteq X \times \Omega$ ,

such that either:

- $A_i \times B_j \subseteq D_\delta$  and  $|\phi(x, w) - \phi(\tilde{x}, \tilde{w})| \leq \varepsilon$  for all  $(x, w), (\tilde{x}, \tilde{w}) \in A_i \times B_j$ , or
- $A_i \times B_j \subseteq (X \times \Omega) \setminus D_\delta$  and  $|\phi(x, w)| \leq \delta$  for all  $(x, w) \in A_i \times B_j$ .

*Proof.* Since  $\phi \in C_0(X \times \Omega)$ , for every  $\delta > 0$  there exists a compact set  $\tilde{D}_\delta \subseteq X \times \Omega$  such that  $|\phi(x, w)| \leq \delta$  outside  $\tilde{D}_\delta$ . By continuity, for each  $(x, w) \in \tilde{D}_\delta$ , there exists an open neighborhood  $O_{x,w}$  with  $|\phi(x, w) - \phi(\tilde{x}, \tilde{w})| \leq \varepsilon/2$  for all  $(\tilde{x}, \tilde{w}) \in O_{x,w}$ . Since open rectangles form a basis for the product topology, we may assume  $O_{x,w} = \tilde{A}_{x,w} \times \tilde{B}_{x,w}$  for open sets  $\tilde{A}_{x,w} \subseteq X$  and  $\tilde{B}_{x,w} \subseteq \Omega$ .

The collection  $\{\tilde{A}_{x,w} \times \tilde{B}_{x,w}\}_{(x,w) \in \tilde{D}_\delta}$  is an open cover of the compact set  $\tilde{D}_\delta$ , so it has a finite subcover  $\{\tilde{A}_i \times \tilde{B}_i\}_{i=1}^{n_c} := \{\tilde{A}_{x_i, w_i} \times \tilde{B}_{x_i, w_i}\}_{i=1}^{n_c}$ , with  $(x_i, w_i) \in \tilde{A}_i \times \tilde{B}_i$ , and

$$|\phi(x_i, w_i) - \phi(x, w)| \leq \varepsilon/2 \quad \text{for all } (x, w) \in \tilde{A}_i \times \tilde{B}_i. \quad (\text{B.19})$$

Define

$$D_\delta := \bigcup_{i=1}^{n_c} \tilde{A}_i \times \tilde{B}_i \supseteq \tilde{D}_\delta \quad (\text{B.20})$$

Moreover, define the finite collections

$$\mathcal{S}_A := \left\{ \bigcap_{i=1}^{n_c} E_i \mid E_i \in \{\tilde{A}_i, \tilde{A}_i^c\} \right\} \setminus \{\emptyset\}, \quad \mathcal{S}_B := \left\{ \bigcap_{i=1}^{n_c} F_i \mid F_i \in \{\tilde{B}_i, \tilde{B}_i^c\} \right\} \setminus \{\emptyset\}. \quad (\text{B.21})$$

Each  $A \in \mathcal{S}_A$  is either fully contained in or disjoint from each  $\tilde{A}_i$ , and similarly for  $B \in \mathcal{S}_B$  and each  $\tilde{B}_i$ . For any  $A \in \mathcal{S}_A$  and  $B \in \mathcal{S}_B$ , define

$$I_{A,B} := \{i \in \{1, \dots, n_c\} \mid A \subseteq \tilde{A}_i, B \subseteq \tilde{B}_i\}. \quad (\text{B.22})$$

- If  $I_{A,B} \neq \emptyset$ , then  $A \times B \subseteq \tilde{A}_i \times \tilde{B}_i \subseteq D_\delta$  for some  $i$ , and for all  $(x, w), (\tilde{x}, \tilde{w}) \in A \times B$ ,

$$|\phi(x, w) - \phi(\tilde{x}, \tilde{w})| \leq |\phi(x, w) - \phi(x_i, w_i)| + |\phi(x_i, w_i) - \phi(\tilde{x}, \tilde{w})| \leq \varepsilon. \quad (\text{B.23})$$

- If  $I_{A,B} = \emptyset$ , then  $A \times B$  is disjoint from all sets  $\tilde{A}_i \times \tilde{B}_i$ . Since

$$\bigcup_i \tilde{A}_i \times \tilde{B}_i = D_\delta, \quad (\text{B.24})$$

it follows that

$$A \times B \subseteq (X \times \Omega) \setminus D_\delta \subseteq (X \times \Omega) \setminus \tilde{D}_\delta, \quad (\text{B.25})$$

where the last inclusion holds because  $\tilde{D}_\delta \subseteq D_\delta$ . Consequently,  $|\phi(x, w)| \leq \delta$  on  $A \times B$ .

Let  $\mathcal{S}_A = (A_1, \dots, A_n)$  and  $\mathcal{S}_B = (B_1, \dots, B_m)$  be ordered versions of the previously defined collections.  $A_i$  are pairwise disjoint with  $\bigcup_{i=1}^n A_i = X$ , and the sets  $B_j$  are pairwise disjoint with  $\bigcup_{j=1}^m B_j = \Omega$ . Together with  $D_\delta$ , they satisfy the conditions stated in the theorem.  $\square$

**Theorem B.2.** *Let  $X$  and  $\Omega$  be topological spaces and  $\phi \in C_0(X \times \Omega)$ . Then for any  $\varepsilon > 0$ , there exist:*

- pairwise disjoint Borel sets  $A_1, \dots, A_n \subseteq X$  with  $\bigcup_{i=1}^n A_i = X$ ,
- pairwise disjoint Borel sets  $B_1, \dots, B_m \subseteq \Omega$  with  $\bigcup_{j=1}^m B_j = \Omega$ ,

and a simple function of the form

$$\phi_\varepsilon(x, w) = \sum_{i=1}^n \sum_{j=1}^m a_{ij} \mathbb{1}_{A_i \times B_j}(x, w) \quad (\text{B.26})$$

such that

$$\sup_{(x, w) \in X \times \Omega} |\phi_\varepsilon(x, w) - \phi(x, w)| \leq \varepsilon \quad (\text{B.27})$$

*Proof.* Define the sets  $\{A_1, \dots, A_n\}$ ,  $\{B_1, \dots, B_m\}$ , and  $D_\varepsilon$  as in Lemma B.1 with  $\delta = \varepsilon$ . For each  $i \in \{1, \dots, n\}$  and  $j \in \{1, \dots, m\}$ , define

$$a_{ij} := \begin{cases} \phi(\tilde{x}, \tilde{w}) & \text{for some } (\tilde{x}, \tilde{w}) \in A_i \times B_j \text{ if } A_i \times B_j \subseteq D_\varepsilon, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{B.28})$$

Define the simple function

$$\phi_\varepsilon(x, w) := \sum_{i=1}^n \sum_{j=1}^m a_{ij} \mathbb{1}_{A_i \times B_j}(x, w). \quad (\text{B.29})$$

Note that the sets  $A_i$  are pairwise disjoint with  $\bigcup_{i=1}^n A_i = X$ , and similarly the sets  $B_j$  are pairwise disjoint with  $\bigcup_{j=1}^m B_j = \Omega$ . It follows that

$$X \times \Omega = \bigcup_{i=1}^n \bigcup_{j=1}^m A_i \times B_j. \quad (\text{B.30})$$

As the sets  $A_i \times B_j$  are disjoint, each  $(x, w) \in X \times \Omega$  belongs to a unique such set, and thus  $\phi_\varepsilon(x, w) = a_{ij}$  for the corresponding pair  $(i, j)$ .

If  $(x, w) \in A_i \times B_j \subseteq D_\varepsilon$ , then  $a_{ij} = \phi(\tilde{x}, \tilde{w})$  for some  $(\tilde{x}, \tilde{w}) \in A_i \times B_j \subseteq D_\varepsilon$ , and

$$|\phi_\varepsilon(x, w) - \phi(x, w)| = |a_{ij} - \phi(x, w)| = |\phi(\tilde{x}, \tilde{w}) - \phi(x, w)| \leq \varepsilon, \quad (\text{B.31})$$

Otherwise,  $(x, w) \in (X \times \Omega) \setminus D_\varepsilon$ , and since  $a_{ij} = 0$ , we have

$$|\phi_\varepsilon(x, w) - \phi(x, w)| = |a_{ij} - \phi(x, w)| = |\phi(x, w)| \leq \varepsilon. \quad (\text{B.32})$$

In either case,

$$\sup_{(x, w) \in X \times \Omega} |\phi_\varepsilon(x, w) - \phi(x, w)| \leq \varepsilon, \quad (\text{B.33})$$

completing the approximation.  $\square$

## References

- [1] Anima Anandkumar, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Nikola Kovachki, Zongyi Li, Burigede Liu, and Andrew Stuart. Neural operator: Graph kernel network for partial differential equations. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2019.
- [2] Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.
- [3] Francis Bach. Structured sparsity-inducing norms through submodular functions. *Advances in Neural Information Processing Systems*, 23, 2010.
- [4] Francis Bach. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18(19):1–53, 2017.
- [5] Francesca Bartolucci, Marcello Carioni, José A. Iglesias, Yury Korolev, Emanuele Naldi, and Stefano Vigogna. A Lipschitz spaces view of infinitely wide shallow neural networks. *arXiv preprint arXiv:2410.14591*, 2024.
- [6] Francesca Bartolucci, Ernesto De Vito, Lorenzo Rosasco, and Stefano Vigogna. Understanding neural networks with reproducing kernel Banach spaces. *Applied and Computational Harmonic Analysis*, 62:194–236, 2023.
- [7] Francesca Bartolucci, Ernesto De Vito, Lorenzo Rosasco, and Stefano Vigogna. Neural reproducing kernel Banach spaces and representer theorems for deep networks. *arXiv preprint arXiv:2403.08750*, 2024.
- [8] Pau Batlle, Matthieu Darcy, Bamdad Hosseini, and Houman Owhadi. Kernel methods are competitive for operator learning. *Journal of Computational Physics*, 496:112549, 2024.
- [9] Kaushik Bhattacharya, Bamdad Hosseini, Nikola B. Kovachki, and Andrew M. Stuart. Model reduction and neural networks for parametric PDEs. *The SMAI Journal of computational mathematics*, 7:121–157, 2021.
- [10] Vladimir Bogachev. *Measure Theory*, volume 2. Springer, 2007.
- [11] Nicolas Boullé and Alex Townsend. A mathematical guide to operator learning. In *Handbook of Numerical Analysis*, volume 25, pages 83–125. Elsevier, 2024.
- [12] Kristian Bredies and Marcello Carioni. Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calculus of Variations and Partial Differential Equations*, 59(1):14, 2020.
- [13] Kristian Bredies, José A. Iglesias, and Daniel Walter. On extremal points for some vectorial total variation seminorms. *arXiv preprint arXiv:2404.12831*, 2024.
- [14] Stéphane Canu, Xavier Mary, and Alain Rakotomamonjy. Functional learning through kernels. *arXiv preprint arXiv:0910.1013*, 2009.
- [15] Claudio Carmeli, Ernesto De Vito, and Alessandro Toigo. Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem. *Analysis and Applications*, 4(04):377–408, 2006.
- [16] Claudio Carmeli, Ernesto De Vito, Alessandro Toigo, and Veronica Umanitá. Vector valued reproducing kernel Hilbert spaces and universality. *Analysis and Applications*, 8(01):19–61, 2010.
- [17] Liangzhi Chen, Haizhang Zhang, and Jun Zhang. Vector-valued reproducing kernel Banach spaces with group lasso norms. *arXiv preprint arXiv:1903.00819*, 2019.
- [18] Patrick L. Combettes, Saverio Salzo, and Silvia Villa. Regularized learning schemes in feature Banach spaces. *Analysis and Applications*, 16(01):1–54, 2018.
- [19] Odo Diekmann and Sjoerd M. Verduyn Lunel. Twin semigroups and delay equations. *J. Differential Equations*, 286:332–410, 2021.
- [20] Sven Dummer, Nicola Strisciuglio, and Christoph Brune. RDA-INR: Riemannian diffeomorphic autoencoding via implicit neural representations. *SIAM Journal on Imaging Sciences*, 17(4):2302–2330, 2024.
- [21] Sven Dummer, Dongwei Ye, and Christoph Brune. RONOM: Reduced-order neural operator modeling. *arXiv preprint arXiv:2507.12814*, 2025.

- [22] Marián Fabian, Petr Habala, Petr Hájek, Vicente Montesinos, and Václav Zizler. *Banach space theory: The basis for linear and nonlinear analysis*. Springer Science & Business Media, 2011.
- [23] Pando G. Georgiev, Luis Sánchez-González, and Panos M. Pardalos. Construction of pairs of reproducing kernel Banach spaces. In *Constructive Nonsmooth Analysis and Related Topics*, pages 39–57. Springer, 2013.
- [24] Tjeerd Jan Heeringa, Len Spek, and Christoph Brune. Deep networks are reproducing kernel chains. *arXiv preprint arXiv:2501.03697*, 2025.
- [25] Masahiro Ikeda, Isao Ishikawa, and Corbinian Schlosser. Koopman and Perron–Frobenius operators on reproducing kernel Banach spaces. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(12), 2022.
- [26] Katiana Kontolati, Somdatta Goswami, George Em Karniadakis, and Michael D Shields. Learning in latent spaces improves the predictive accuracy of deep neural operators. *arXiv preprint arXiv:2304.07599*, 2023.
- [27] Yury Korolev. Two-layer neural networks with values in a Banach space. *SIAM Journal on Mathematical Analysis*, 54(6):6358–6389, 2022.
- [28] Nikola Kovachki, Samuel Lanthaler, and Siddhartha Mishra. On universal approximation and error bounds for fourier neural operators. *Journal of Machine Learning Research*, 22(290):1–76, 2021.
- [29] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.
- [30] Rong Rong Lin, Hai Zhang Zhang, and Jun Zhang. On reproducing kernel Banach spaces: Generic definitions and unified framework of constructions. *Acta Mathematica Sinica, English Series*, 38(8):1459–1483, 2022.
- [31] Rongrong Lin, Guohui Song, and Haizhang Zhang. Multi-task learning in vector-valued reproducing kernel Banach spaces with the  $\ell^1$  norm. *Journal of Complexity*, 63:101514, 2021.
- [32] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.
- [33] Lu Lu, Xuhui Meng, Shengze Cai, Zhiping Mao, Somdatta Goswami, Zhongqiang Zhang, and George Em Karniadakis. A comprehensive and fair comparison of two neural operators (with practical extensions) based on fair data. *Computer Methods in Applied Mechanics and Engineering*, 393:114778, 2022.
- [34] Chao Ma, Stephan Wojtowytsch, Lei Wu, et al. Towards a mathematical understanding of neural network-based machine learning: what we know and what we don’t. *arXiv preprint arXiv:2009.10713*, 2020.
- [35] Chao Ma, Lei Wu, et al. A priori estimates of the population risk for two-layer neural networks. *Communications in Mathematical Sciences*, 17(5):1407–1425, 2019.
- [36] Chao Ma, Lei Wu, et al. The Barron space and the flow-induced function spaces for neural network models. *Constructive Approximation*, 55(1):369–406, 2022.
- [37] Lakhdar Meziani. On the dual space  $C_0^*(S, X)$ . *Acta Math. Univ. Comenian. (N.S.)*, 78(1):153–160, 2009.
- [38] Rahul Parhi and Robert D. Nowak. Banach space representer theorems for neural networks and ridge splines. *Journal of Machine Learning Research*, 22(43):1–40, 2021.
- [39] Rahul Parhi and Robert D. Nowak. What kinds of functions do deep neural networks learn? Insights from variational spline theory. *SIAM Journal on Mathematics of Data Science*, 4(2):464–489, 2022.
- [40] Junhyung Park and Krikamol Muandet. A measure-theoretic approach to kernel conditional mean embeddings. *Advances in Neural Information Processing Systems*, 33:21247–21259, 2020.
- [41] Vern I. Paulsen and Mrinal Raghupathi. *An introduction to the theory of reproducing kernel Hilbert spaces*, volume 152 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2016.
- [42] Bogdan Raonic, Roberto Molinaro, Tobias Rohner, Siddhartha Mishra, and Emmanuel de Bezenac. Convolutional neural operators. In *ICLR 2023 Workshop on Physics for Machine Learning*, 2023.
- [43] Jacob Seidman, Georgios Kissas, Paris Perdikaris, and George J. Pappas. NOMAD: Nonlinear manifold decoders for operator learning. *Advances in Neural Information Processing Systems*, 35:5601–5613, 2022.

- [44] Joseph Shenouda, Rahul Parhi, Kangwook Lee, and Robert D. Nowak. Variation spaces for multi-output neural networks: Insights on multi-task learning and network compression. *Journal of Machine Learning Research*, 25(231):1–40, 2024.
- [45] Ivan Singer. Linear functionals on the space of continuous mappings of a compact Hausdorff space into a Banach space. *Rev. Math. Pures Appl.*, 2:301–315, 1957.
- [46] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020.
- [47] Len Spek, Tjeerd Jan Heeringa, Felix Schwenninger, and Christoph Brune. Duality for neural networks through reproducing kernel Banach spaces. *Applied and Computational Harmonic Analysis*, 78:101765, 2025.
- [48] Ingo Steinwart. Reproducing kernel Hilbert spaces cannot contain all continuous functions on a compact metric space. *Archiv der Mathematik*, 122(5):553–557, 2024.
- [49] Terence Tao. *An introduction to measure theory*, volume 126 of *Graduate Studies in Mathematics*. American Mathematical Society, 2011.
- [50] Rui Wang, Yuesheng Xu, and Mingsong Yan. Sparse representer theorems for learning in reproducing kernel Banach spaces. *Journal of Machine Learning Research*, 25(93):1–45, 2024.
- [51] Rui Wang, Yuesheng Xu, and Mingsong Yan. Hypothesis spaces for deep learning. *Neural Networks*, 193:107995, 2026.
- [52] Dirk Werner. Extreme points in spaces of operators and vector-valued measures. *Rend. Circ. Mat. Palermo (2)*, Suppl.(5):135–143, 1984.
- [53] Stephan Wojtowytsch et al. On the Banach spaces associated with multi-layer ReLU networks: Function representation, approximation theory and gradient descent dynamics. *CSIAM Transactions on Applied Mathematics*, 1(3):387–440, 2020.
- [54] Stephan Wojtowytsch et al. Representation formulas and pointwise properties for Barron functions. *Calculus of Variations and Partial Differential Equations*, 61(2):1–37, 2022.
- [55] Yuesheng Xu. Sparse machine learning in Banach spaces. *Applied Numerical Mathematics*, 187:138–157, 2023.
- [56] Haizhang Zhang, Yuesheng Xu, and Jun Zhang. Reproducing kernel Banach spaces for machine learning. *Journal of Machine Learning Research*, 10(12), 2009.
- [57] Haizhang Zhang and Jun Zhang. Vector-valued reproducing kernel Banach spaces with applications to multi-task learning. *Journal of Complexity*, 29(2):195–215, 2013.