UNIVERSAL APPROXIMATION RESULTS FOR NEURAL NETWORKS WITH NON-POLYNOMIAL ACTIVATION FUNCTION OVER NON-COMPACT DOMAINS

ARIEL NEUFELD AND PHILIPP SCHMOCKER

ABSTRACT. This paper extends the universal approximation property of single-hidden-layer feedforward neural networks beyond compact domains, which is of particular interest for the approximation within weighted C^k-spaces and weighted Sobolev spaces over unbounded domains. More precisely, by assuming that the activation function is non-polynomial, we establish universal approximation results within function spaces defined over non-compact subsets of a Euclidean space, including L^p-spaces, weighted C^k-spaces, and weighted Sobolev spaces, where the latter two include the approximation of the (weak) derivatives. Moreover, we provide some dimension-independent rates for approximating a function with sufficiently regular and integrable Fourier transform by neural networks with non-polynomial activation function.

1. INTRODUCTION

Inspired by the functionality of human brains, (artificial) neural networks have been discovered in the seminal work of McCulloch and Pitts [39], and can be described in mathematical terms as concatenation of affine and non-affine functions. Nowadays, neural networks are successfully applied in the fields of image classification [33], speech recognition [24], and computer games [50], and provide as a supervised machine learning technique an algorithmic approach for the quest of artificial intelligence (see [43, 52]).

One of the fundamental reasons for the success of neural networks is their universal approximation property. In so-called universal approximation theorems (UATs), this property establishes the denseness of the set of (single-hidden-layer) neural networks within a given function space. First, Cybenko [12], Funahashi [20], and Hornik et al. [26] showed independently that neural networks with a continuous sigmoidal activation function are dense in the space of continuous functions over a compact subset of a Euclidean space. Subsequently, the universal approximation property was extended to broader classes of activation functions, e.g., by Hornik et al. [25] to continuous, bounded, and non-constant activation functions, and by Mhaskar and Micchelli [40], Leshno et. al [37], Chen and Chen [10], and Pinkus [45] to non-polynomial activation functions. Moreover, Kidger and Lyons [29] has established a UAT for deep narrow neural networks with non-affine activation function which is differentiable at at least one point with non-zero derivative. In addition, different UATs have been established within more general function spaces, e.g., Cybenko [12], Mhaskar and Micchelli [40], and Leshno [37] within L^p-spaces with compactly supported measure, and Hornik [25] within L^p-spaces with finite measure. In addition, [25, 27] included the approximation of the derivatives, showing UATs within C^k-spaces over compact domains and within Sobolev spaces with compactly supported measures.

In this paper, we extend the universal approximation property of (single-hidden-layer) neural networks to function spaces over non-compact domains. This is of particular interest for the approximation in weighted C^k-spaces and weighted Sobolev spaces over unbounded domains, which, to the best of our knowledge, has not yet been addressed in the literature. To be as general as possible, we consider function spaces obtained as completion of the space of bounded and k-times continuously differentiable functions with bounded derivatives over a possibly non-compact domain with respect to a weighted norm. In this setting, we then follow Cybenko's approach [12] to obtain the universal approximation property, where we combine the classical Hahn-Banach separation argument with a Riesz representation theorem on weighted spaces [16] and Korevaar's distributional extension [31] of Wiener's Tauberian theorem [54]. Our approach also extends the weighted UAT in [11] proven for neural networks between infinite dimensional spaces, by now also including the approximation of the derivatives. In the same way it also generalizes the weighted UAT in [53], which observes that bounded sigmoidal activation functions

with distinct limits at positive and negative infinity are more powerful than those with identical limits. Moreover, in the context of dynamical systems, [2, 22] considered weighted approximation in time.

Furthermore, we prove dimension-independent rates to approximate a given function by a single-hidden-layer neural network in a (weighted) Sobolev space, which relates the approximation error to the network size. First, Barron [5, 6] obtained the approximation rates for neural networks with sigmoidal activation within L^2-spaces, which was extended by Darken et al. [13, 14] to L^p-spaces, in Mhaskar and Micchelli [41, 42] to non-sigmoidal activation functions, and in Siegel and Xu [48] to $W^{k,2}$-Sobolev spaces over bounded domains. For the particular case of shallow neural networks with ReLU activation, better approximation rates have been obtained, e.g., in [38, 49, 55]. On the other hand, [32] derived approximation rates for deep neural networks with non-affine activation function which is differentiable at at least one point with non-zero derivative, coinciding with the best known rates for deep ReLU neural networks in [56] up to constants. For a more detailed review of the literature on approximation rates, we refer to [8, 34, 45, 48]. In this paper, we extend the approximation rates for single-hidden-layer neural networks to weighted Sobolev spaces over unbounded domains, where we apply the (distributional dual) ridgelet transform of Sonoda and Murata [51] and use the concept of Rademacher averages.

1.1. **Outline.** In Section 2, we introduce (single-hidden-layer) neural networks and extend their universal approximation property beyond compact domains. In Section 3, we provide some dimension-independent rates for neural networks in a weighted Sobolev space. Finally, Section 4 contains all proofs.

1.2. **Notation.** As usual, $\mathbb{N} := \{1, 2, 3, ...\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ denote the sets of natural numbers, whereas $\mathbb{Z}$ represents the set of integers. Moreover, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of real and complex numbers, respectively, where $\mathbf{i} := \sqrt{-1} \in \mathbb{C}$ represents the imaginary unit. In addition, for any $r \in \mathbb{R}$, we define $\lceil r \rceil := \min \{k \in \mathbb{Z} : k \geqslant r\}$. Furthermore, for any $z \in \mathbb{C}$, we denote its real and imaginary part by $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$, respectively, whereas its complex conjugate is defined as $\overline{z} := \mathrm{Re}(z) - \mathrm{Im}(z)\mathbf{i}$. Moreover, for any $m \in \mathbb{N}$, we denote by $\mathbb{R}^m$ (and $\mathbb{C}^m$) the m-dimensional (complex) Euclidean space, which is equipped with the Euclidean norm $\|u\| = \sqrt{\sum_{i=1}^{m} |u_i|^2}$.

In addition, for $U \subseteq \mathbb{R}^m$, we denote by $\mathcal{B}(U)$ the σ-algebra of Borel-measurable subsets of U. Moreover, for $U \in \mathcal{B}(\mathbb{R}^m)$, we denote by $\mathcal{L}(U)$ the σ-algebra of Lebesgue-measurable subsets of U, while $du : \mathcal{L}(U) \to [0, \infty]$ represents the Lebesgue measure on U. Then, a property is said to hold true almost everywhere (shortly a.e.) if it holds everywhere true except on a set of Lebesgue measure zero.

Furthermore, for every fixed $m, d \in \mathbb{N}$, $k \in \mathbb{N}_0$, and $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), we denote by $C^k(U; \mathbb{R}^d)$ the vector space of k-times continuously differentiable functions $f : U \to \mathbb{R}^d$ such that for every $\alpha \in \mathbb{N}_{0,k}^m := \{\alpha = (\alpha_1, ..., \alpha_m) \in \mathbb{N}_0^m : |\alpha| := \alpha_1 + ... + \alpha_m \leqslant k\}$ the partial derivative $U \ni u \mapsto \partial_\alpha f(u) := \frac{\partial^{|\alpha|} f}{\partial u_1^{\alpha_1} \cdots \partial u_m^{\alpha_m}}(u) \in \mathbb{R}^d$ is continuous. If $m = 1$, we write $f^{(j)} := \frac{\partial^j f}{\partial u^j} : U \to \mathbb{R}^d$, $j = 0, ..., k$. Note that for $k \geqslant 1$ the domain $U \subseteq \mathbb{R}^m$ must be open for the definition of partial derivatives.

Moreover, we denote by $C_b^k(U; \mathbb{R}^d)$ the vector space of bounded functions $f \in C^k(U; \mathbb{R}^d)$ with bounded partial derivatives $\partial_\alpha f : U \to \mathbb{R}^d$, for all $\alpha \in \mathbb{N}_{0,k}^m$. Then, the norm

$$\|f\|_{C_b^k(U; \mathbb{R}^d)} := \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in U} \|\partial_\alpha f(u)\|.$$

turns $(C_b^k(U; \mathbb{R}^d), \|\cdot\|_{C_b^k(U; \mathbb{R}^d)})$ into a Banach space. For $k = 0$ and $U \subset \mathbb{R}^m$ compact, we obtain the Banach space $(C^0(U; \mathbb{R}^d), \|\cdot\|_{C^0(U; \mathbb{R}^d)})$ of continuous functions equipped with the supremum norm.

In addition, for $\gamma \in [0, \infty)$, we denote by $C_{pol,\gamma}^k(U; \mathbb{R}^d)$ the vector space of functions $f \in C^k(U; \mathbb{R}^d)$ with

$$\|f\|_{C_{pol,\gamma}^k(U; \mathbb{R}^d)} := \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in U} \frac{\|\partial_\alpha f(u)\|}{(1 + \|u\|)^\gamma} < \infty.$$

Furthermore, we introduce the *weighed C^k-space* $\overline{C_b^k(U; \mathbb{R}^d)}^\gamma$ defined as the closure of $C_b^k(U; \mathbb{R}^d)$ with respect to $\|\cdot\|_{C_{pol,\gamma}^k(U; \mathbb{R}^d)}$. Then, $(\overline{C_b^k(U; \mathbb{R}^d)}^\gamma, \|\cdot\|_{C_{pol,\gamma}^k(U; \mathbb{R}^d)})$ is by definition a Banach space. If $U \subseteq \mathbb{R}^m$ is bounded, then $\overline{C_b^k(U; \mathbb{R}^d)}^\gamma = C_b^k(U; \mathbb{R}^d)$. Otherwise, $f \in \overline{C_b^k(U; \mathbb{R}^d)}^\gamma$ if and only if $f \in C^k(U; \mathbb{R}^d)$ and $\lim_{r \to \infty} \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in U, \|u\| \geqslant r} \frac{\|\partial_\alpha f(u)\|}{(1 + \|u\|)^\gamma} = 0$ (see Lemma 4.1).

Moreover, we abbreviate $C^k(U) := C^k(U; \mathbb{R})$, $C_b^k(U) := C_b^k(U; \mathbb{R})$ and define the complex-valued function spaces $C^k(U; \mathbb{C}^d) \cong C^k(U; \mathbb{R}^{2d})$, $C_b^k(U; \mathbb{C}^d) \cong C_b^k(U; \mathbb{R}^{2d})$, etc. by identifying $\mathbb{C}^d \cong \mathbb{R}^{2d}$.

Furthermore, we define the Fourier transform of any Lebesgue-integrable function $f : \mathbb{R}^m \to \mathbb{C}^d$ as

$$\mathbb{R}^m \ni \xi \quad \mapsto \quad \widehat{f}(\xi) := \int_{\mathbb{R}^m} e^{-i\xi^\top u} f(u) du \in \mathbb{C}^d, \tag{1}$$

see [19, p. 247]. Then, by using [28, Proposition 1.2.2], it follows that

$$\sup_{\xi \in \mathbb{R}^m} \left\| \widehat{f}(\xi) \right\| = \sup_{\xi \in \mathbb{R}^m} \left\| \int_{\mathbb{R}^m} e^{-i\xi^\top u} f(u) du \right\| \leqslant \int_{\mathbb{R}^m} \| f(u) \| du = \| f \|_{L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)}. \tag{2}$$

In addition, we denote by $\mathcal{S}'(\mathbb{R}^m; \mathbb{C})$ the dual space of the Schwartz space $\mathcal{S}(\mathbb{R}^m; \mathbb{C})$ consisting of smooth functions $f : \mathbb{R}^m \to \mathbb{C}$ such that the seminorms $\max_{\alpha \in \mathbb{N}_{0,n}^m} \sup_{u \in \mathbb{R}^m} \left(1 + \|u\|^2 \right)^n |\partial_\alpha f(u)| < \infty$, $n \in \mathbb{N}_0$, are finite. Then, the Fourier transform of any tempered distribution $T \in \mathcal{S}'(\mathbb{R}^m; \mathbb{C})$ is defined as $\widehat{T}(g) := T(\widehat{g})$, for $g \in \mathcal{S}(\mathbb{R}^m; \mathbb{C})$ (see [19, Equation 9.28]).

2. Universal approximation of neural networks

Inspired by the functionality of a human brain, neural networks were introduced in [39] and are nowadays applied as machine learning technique in various research areas (see [43]). In mathematical terms, a neural network can be described as a concatenation of affine and non-affine functions.

Definition 2.1. *For $\rho \in C^0(\mathbb{R})$, a function $\varphi : \mathbb{R}^m \to \mathbb{R}^d$ is called a* (single-hidden-layer feed-forward) *neural network if it is of the form*

$$\mathbb{R}^m \ni u \quad \mapsto \quad \varphi(u) = \sum_{n=1}^N y_n \rho \left(a_n^\top u - b_n \right) \in \mathbb{R}^d \tag{3}$$

with respect to some $N \in \mathbb{N}$ denoting the number of neurons, *where $a_1, ..., a_N \in \mathbb{R}^m$, $b_1, ..., b_N \in \mathbb{R}$, and $y_1, ..., y_N \in \mathbb{R}^d$ represent the* weight vectors, biases, *and* linear readouts, *respectively.*

Definition 2.2. *For $U \subseteq \mathbb{R}^m$ and $\rho \in C^0(\mathbb{R})$, we denote by $\mathcal{NN}_{U,d}^\rho$ the set of all neural networks of the form* (3) *restricted to U with corresponding activation function $\rho \in C^0(\mathbb{R})$.*
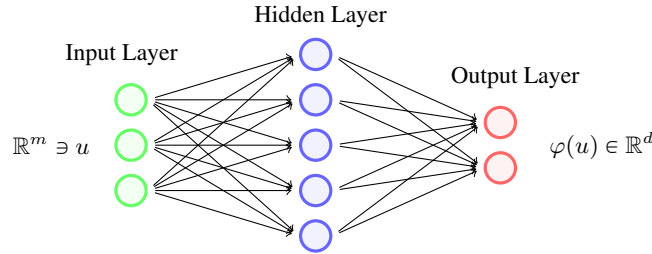

Figure 1. A neural network $\varphi : \mathbb{R}^m \to \mathbb{R}^d$ with $m = 3$, $d = 2$, and $N = 5$.

In this paper, we restrict ourselves to single-hidden-layer feed-forward neural networks of the form (3) and simply refer to them as neural networks.

2.1. Universal approximation.

Neural networks admit the so-called universal approximation property, which establishes the denseness of the set of neural networks in a given function space with respect to some suitable topology. For example, every continuous function can be approximated arbitrarily well on a compact subset of a Euclidean space (see, e.g., [12, 25, 45] and the references therein).

In order to generalize the approximation property of neural networks beyond the space of continuous functions on compacta, we now introduce the following type of function spaces. For this purpose, we fix the input dimension $m \in \mathbb{N}$ and the output dimension $d \in \mathbb{N}$ throughout the rest of this paper.

Assumption 2.3. *For $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), and $\gamma \in (0, \infty)$, let $(X, \| \cdot \|_X)$ be a Banach space consisting of functions $f : U \to \mathbb{R}^d$ such that the restriction map*

$$\left(C_b^k(\mathbb{R}^m; \mathbb{R}^d), \| \cdot \|_{C_{pol, \gamma}^k(\mathbb{R}^m; \mathbb{R}^d)} \right) \ni f \quad \mapsto \quad f|_U \in (X, \| \cdot \|_X) \tag{4}$$

is a continuous dense embedding.

Note that for $k \geqslant 1$ the domain $U \subseteq \mathbb{R}^m$ must be open to compute partial derivatives of a function.

Remark 2.4. *The restriction map* (4) *is a continuous dense embedding if and only if it is continuous and its image* $\left\{ f|_U : f \in C_b^k(\mathbb{R}^m; \mathbb{R}^d) \right\}$ *is dense in* X. *Since* $\overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma$ *is defined as the closure of* $C_b^k(\mathbb{R}^m; \mathbb{R}^d)$ *with respect to* $\| \cdot \|_{C_{pol,\gamma}^k(\mathbb{R}^m; \mathbb{R}^d)}$, *the restriction map* (4) *is a continuous dense embedding if and only if* $\left(\overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma, \| \cdot \|_{C_{pol,\gamma}^k(\mathbb{R}^m; \mathbb{R}^d)} \right) \ni f \mapsto f|_U \in (X, \| \cdot \|_X)$ *is a continuous dense embedding.*

The continuous dense embedding in (4) ensures that the set of neural networks $\mathcal{NN}_{U,d}^\rho \subseteq X$ is well-defined in the function space $(X, \| \cdot \|_X)$, for all activation functions $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$.

Lemma 2.5. *Let* $(X, \| \cdot \|_X)$ *satisfy Assumption 2.3 with* $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ *(open, if* $k \geqslant 1$*), and* $\gamma \in (0, \infty)$. *Moreover, let* $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$. *Then, we have* $\mathcal{NN}_{\mathbb{R}^m,d}^\rho \subseteq \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma$ *and* $\mathcal{NN}_{U,d}^\rho \subseteq X$.

Let us give some examples of function spaces satisfying Assumption 2.3, which includes in particular C_b^k-spaces, weighted C^k-spaces, L^p-spaces, and (weighted) $W^{k,p}$-Sobolev spaces.

Example 2.6. *The following function spaces* $(X, \| \cdot \|_X)$ *are Banach spaces satisfying Assumption 2.3 with* $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ *(open, if* $k \geqslant 1$*), and* $\gamma \in (0, \infty)$:

	Function space	Banach space X	Additional necessary conditions
(a)	C_b^k-space[1]	$C_b^k(U; \mathbb{R}^d)$	$U \subset \mathbb{R}^m$ is bounded.
(b)	Weighted C^k-space[1]	$\overline{C_b^k(U; \mathbb{R}^d)}^\gamma$	none
(c)	L^p-space[2], $k = 0$	$L^p(U, \mathcal{B}(U), \mu; \mathbb{R}^d)$	$\int_U (1 + \|u\|)^{\gamma p} \mu(du) < \infty$.
(d)	Sobolev space[3], $k \geqslant 1$	$W^{k,p}(U, \mathcal{L}(U), du; \mathbb{R}^d)$	$U \subset \mathbb{R}^m$ is bounded and $U \subset \mathbb{R}^m$ has the segment property[4].
(e)	Weighted Sobolev space[5], $k \geqslant 1$	$W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$	$U \subseteq \mathbb{R}^m$ has the segment property[4], $w : U \to [0, \infty)$ is bounded, $\inf_{u \in B} w(u) > 0$ for all bounded $B \subseteq U$, and $\int_U (1 + \|u\|)^{\gamma p} w(u) du < \infty$.

For L^p-spaces with the Lebesgue measure $\mu := du$ in (c), the domain $U \subseteq \mathbb{R}^m$ needs to be bounded.

Moreover, we assume that the activation function $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$ is non-polynomial, i.e. algebraically not a polynomial over $\mathbb{R}$. Since $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$ induces the tempered distribution $\left(g \mapsto T_\rho(g) := \int_{\mathbb{R}} \rho(s) g(s) ds \right) \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$ (see [19, Equation 9.26]), this is equivalent to the condition that the Fourier transform $\widehat{T_\rho} \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$ is supported[6] at a non-zero point (see [47, Examples 7.16] and Example 3.3).

In order to obtain the universal approximation property of neural networks in function space satisfying Assumption 2.3, we now combine the classical Hahn-Banach separation argument with a Riesz representation theorem on weighted spaces (see [16, Theorem 2.4]) and follow Korevaar's distributional extension [31] of Wiener's Tauberian theorem [54]. The proof can be found in Section 4.2.2.

Theorem 2.7. *Let* $(X, \| \cdot \|_X)$ *be a function space satisfying Assumption 2.3 with* $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ *(open, if* $k \geqslant 1$*), and* $\gamma \in (0, \infty)$. *Moreover, let* $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$ *be non-polynomial. Then,* $\mathcal{NN}_{U,d}^\rho$ *is dense in* X.

Remark 2.8. *Theorem 2.7 extends the following universal approximation thereoms (UATs).*

---

[1]For the definition, we refer to Section 1.2.

[2]For $U \subseteq \mathbb{R}^m$, $p \in [1, \infty)$, and Borel measure $\mu : \mathcal{B}(U) \to [0, \infty]$, we denote by $L^p(U, \mathcal{B}(U), \mu; \mathbb{R}^d)$ the L^p-space of (equivalence classes of) $\mathcal{B}(U)/\mathcal{B}(\mathbb{R}^d)$-measurable $f : U \to \mathbb{R}^d$ with $\|f\|_{L^p(U, \mathcal{B}(U), \mu; \mathbb{R}^d)} := \left(\int_U \|f(u)\|^p \mu(du) \right)^{1/p} < \infty$.

[3]For $k \in \mathbb{N}$, $U \subseteq \mathbb{R}^m$ open, and $p \in [1, \infty)$, we denote by $W^{k,p}(U, \mathcal{L}(U), du; \mathbb{R}^d)$ the Sobolev space of (equivalence classes of) k-times weakly differentiable $f : U \to \mathbb{R}^d$ with $\partial_\alpha f \in L^p(U, \mathcal{L}(U), du; \mathbb{R}^d)$, where $\|f\|_{W^{k,p}(U, \mathcal{L}(U), du; \mathbb{R}^d)} := \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u)\|^p du \right)^{1/p} < \infty$ (see [3, Chapter 3]).

[4]An open subset $U \subseteq \mathbb{R}^m$ has the *segment property* if for every $u \in \partial U := \overline{U} \backslash U$ there exists an open neighborhood $V \subseteq \mathbb{R}^m$ of $u \in \partial U$ and some $y \in \mathbb{R}^m \backslash \{0\}$ such that for every $z \in \overline{U} \cap V$ and $t \in (0, 1)$ we have $z + ty \in U$ (see [3, p. 54]).

[5]For $k \in \mathbb{N}$, $U \subseteq \mathbb{R}^m$ open, $p \in [1, \infty)$, and a weight $w : U \to [0, \infty)$, we denote by $W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ the weighted Sobolev space of (equivalence classes of) k-times weakly differentiable $f : U \to \mathbb{R}^d$ with $\partial_\alpha f \in L^p(U, \mathcal{L}(U), w(u)du; \mathbb{R}^d)$, where $\|f\|_{W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)} := \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u)\|^p w(u) du \right)^{1/p} < \infty$ (see [35, p. 5]). Hereby, the weight $w : U \to [0, \infty)$ is always assumed to be $\mathcal{L}(U)/\mathcal{B}(\mathbb{R})$-measurable and a.e. strictly positive.

[6]The support of $T \in \mathcal{S}'(\mathbb{R}^m; \mathbb{C})$ is defined as the complement of the largest open set $U \subseteq \mathbb{R}^m$ on which $T \in \mathcal{S}'(\mathbb{R}^m; \mathbb{C})$ vanishes, i.e. $T(g) = 0$ for all smooth functions $g : U \to \mathbb{C}$ with compact support contained in U.

(i) *Within C^0-spaces on compacta: Cybenko [12] and Hornik et al. [26] with sigmoidal activation; Hornik [25] with continuous, bounded, and non-constant activation; Mhaskar and Micchelli [40], Leshno et al. [37], Chen and Chen [10], and Pinkus [45] with non-polynomial activation.*

(ii) *Within L^p-spaces with compactly supported and/or finite measure: Hornik et al. [26] with sigmoidal activation; Hornik [25] with bounded and non-constant activation; Mhaskar and Micchelli [40] and Leshno et al. [37] with non-polynomial activation.*

(iii) *Within C^k-spaces over compact domains or Sobolev spaces with compactly supported measure: Hornik et al. [27] with k-finite activation; Hornik [25] with bounded and non-constant activation.*

(iv) *Within function spaces over non-compact domains: Cuchiero et al. [11] within weighted C^0-spaces without derivatives; van Nuland [53] within the space of continuous functions vanishing at infinity, which is comparable to our situation of $U \ni u \mapsto \frac{f(u)}{(1+\|u\|)^\gamma} \in \mathbb{R}^d$ vanishing at infinity.*

Hence, the main contribution of Theorem 2.7 is the universal approximation property of neural networks within weighted C^k-spaces and weighted Sobolev spaces over unbounded domains.

3. APPROXIMATION RATES

In this section, we provide some rates to approximate a given function by a (single-hidden-layer) neural network within the weighted Sobolev space $W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$, with $k \in \mathbb{N}_0$, $p \in [1, \infty)$, and $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), where we set $W^{0,p}(U, \mathcal{L}(U), w; \mathbb{R}^d) := L^p(U, \mathcal{L}(U), w(u)du; \mathbb{R}^d)$.

To this end, we apply the reconstruction formula in [51, Theorem 5.6] to obtain an integral representation of the function to be approximated (see Proposition 3.4 below). For that, we consider pairs $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C^k_{pol,\gamma}(\mathbb{R})$ of a ridgelet function[7] $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$ and an activation function $\rho \in C^k_{pol,\gamma}(\mathbb{R})$ satisfying the following, which is a special case of [51, Definition 5.1] (see also [9]).

Definition 3.1. *For $k \in \mathbb{N}_0$ and $\gamma \in [0, \infty)$, a pair $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C^k_{pol,\gamma}(\mathbb{R})$ is called m-admissible if $\widehat{T_\rho} \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$ coincides[8] on $\mathbb{R}\backslash\{0\}$ with a function $f_{\widehat{T_\rho}} \in L^1_{loc}(\mathbb{R}\backslash\{0\}; \mathbb{C})$ such that*

$$C_m^{(\psi,\rho)} := (2\pi)^{m-1} \int_{\mathbb{R}\backslash\{0\}} \frac{\overline{\widehat{\psi}(\xi)} f_{\widehat{T_\rho}}(\xi)}{|\xi|^m} d\xi \in \mathbb{C}\backslash\{0\}. \tag{5}$$

Remark 3.2. *If $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C^k_{pol,\gamma}(\mathbb{R})$ is m-admissible, then $\rho \in C^k_{pol,\gamma}(\mathbb{R})$ has to be non-polynomial. Indeed, otherwise the support of $\widehat{T_\rho} \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$ is contained in $\{0\} \subset \mathbb{R}$ (see [47, Examples 7.16]), which implies that (5) vanishes for any choice of $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$.*

Together with some suitable $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$, most common activation functions satisfy Definition 3.1.

Example 3.3. *For $k \in \mathbb{N}_0$ and $\gamma \in [0, \infty)$, let $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$ be such that $\widehat{\psi} \in C_c^\infty(\mathbb{R})$ is non-negative with $\mathrm{supp}(\widehat{\psi}) = [\zeta_1, \zeta_2]$ for some $0 < \zeta_1 < \zeta_2 < \infty$. Then, for every $m \in \mathbb{N}$ and every activation function $\rho \in C^k_{pol,\gamma}(\mathbb{R})$ listed in the table below the pair $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C^k_{pol,\gamma}(\mathbb{R})$ is m-admissible.*

	$\rho \in C^k_{pol,\gamma}(\mathbb{R})$		$k \in \mathbb{N}_0$	$\gamma \in [0,\infty)$	$f_{\widehat{T_\rho}} \in L^1_{loc}(\mathbb{R}\backslash\{0\}; \mathbb{C})$
(a)	*Sigmoid function*	$\rho(s) := \frac{1}{1+\exp(-s)}$	$k \in \mathbb{N}_0$	$\gamma \geqslant 0$	$f_{\widehat{T_\rho}}(\xi) = \frac{-\mathbf{i}\pi}{\sinh(\pi\xi)}$
(b)	*Tangens hyperbolicus*	$\rho(s) := \tanh(s)$	$k \in \mathbb{N}_0$	$\gamma \geqslant 0$	$f_{\widehat{T_\rho}}(\xi) = \frac{-\mathbf{i}\pi}{\sinh(\pi\xi/2)}$
(c)	*Softplus function*	$\rho(s) := \ln(1+\exp(s))$	$k \in \mathbb{N}_0$	$\gamma \geqslant 1$	$f_{\widehat{T_\rho}}(\xi) = \frac{-\pi}{\xi\sinh(\pi\xi)}$
(d)	*ReLU function*	$\rho(s) := \max(s, 0)$	$k = 0$	$\gamma \geqslant 1$	$f_{\widehat{T_\rho}}(\xi) = -\frac{1}{\xi^2}$

Moreover, there exists $C_{\psi,\rho} > 0$ (independent of $m, d \in \mathbb{N}$) such that $\left|C_m^{(\psi,\rho)}\right| \geqslant C_{\psi,\rho}(2\pi/\zeta_2)^m$.

Next, we follow [9, 51] and define for every $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$ the (multi-dimensional) *ridgelet transform* of any function $g \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)$ as

$$\mathbb{R}^m \times \mathbb{R} \ni (a, b) \quad \mapsto \quad (\mathfrak{R}_\psi g)(a, b) := \int_{\mathbb{R}^m} \psi\left(a^\top u - b\right) g(u)\|a\|du \in \mathbb{C}^d. \tag{6}$$

Then, we can apply the reconstruction formula in [51, Theorem 5.6] componentwise to obtain an integral representation. The proof can be found in Section 4.3.1.

---

[7]$\mathcal{S}_0(\mathbb{R}; \mathbb{C}) \subseteq \mathcal{S}(\mathbb{R}; \mathbb{C})$ denotes the vector subspace of $\psi \in \mathcal{S}(\mathbb{R}; \mathbb{C})$ with $\widehat{\psi}^{(j)}(0) = \int_{\mathbb{R}} u^j \psi(u) du = 0$ for all $j \in \mathbb{N}_0$.

[8]This means that $\widehat{T_\rho}(g) = \int_{\mathbb{R}\backslash\{0\}} f_{\widehat{T_\rho}}(\xi) g(\xi) d\xi$ for all smooth functions $g : \mathbb{R}\backslash\{0\} \to \mathbb{C}$ with compact support in $\mathbb{R}\backslash\{0\}$.

Proposition 3.4. *For $k \in \mathbb{N}_0$ and $\gamma \in [0, \infty)$, let $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C_{pol,\gamma}^k(\mathbb{R})$ be m-admissible and let $g \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)$ with $\widehat{g} \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{C}^d)$. Then, for a.e. $u \in \mathbb{R}^m$, it holds that*

$$\int_{\mathbb{R}^m} \int_{\mathbb{R}} (\mathfrak{R}_\psi g)(a, b) \rho \left(a^\top u - b \right) db\, da = C_m^{(\psi, \rho)} g(u).$$

In addition, we introduce the following Barron spaces which are inspired by the works [6, 17, 30].

Definition 3.5. *For $k \in \mathbb{N}_0$, $\gamma \in [0, \infty)$, and $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$, we define the ridgelet-Barron space $\mathbb{B}_\psi^{k,\gamma}(U; \mathbb{R}^d)$ as vector space of $\mathcal{L}(U)/\mathcal{B}(\mathbb{R}^d)$-measurable functions $f : U \to \mathbb{R}^d$ such that*

$$\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)} := \inf_g \left(\int_{\mathbb{R}^m} \int_{\mathbb{R}} \left(1 + \|a\|^2\right)^{\gamma + k + \frac{m+1}{2}} \left(1 + |b|^2\right)^{\gamma+1} \|(\mathfrak{R}_\psi g)(a, b)\|^2 db\, da \right)^{\frac{1}{2}} < \infty,$$

where the infimum is taken over all $g \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)$ with $\widehat{g} \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{C}^d)$ and $g = f$ a.e. on U.

Now, we present the dimension-independent approximation rates, whose proof is given in Section 4.3.4.

Theorem 3.6. *For $k \in \mathbb{N}_0$, $p \in [1, \infty)$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), and $\gamma \in [0, \infty)$, let $w : U \to [0, \infty)$ be a weight such that*

$$C_{U,w}^{(\gamma,p)} := \left(\int_U (1 + \|u\|)^{\gamma p} w(u) du \right)^{\frac{1}{p}} < \infty. \tag{7}$$

Moreover, let $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C_{pol,\gamma}^k(\mathbb{R})$ be m-admissible. Then, there exists a constant $C_p > 0$ (depending only on $p \in [1, \infty)$) such that for every $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d) \cap \mathbb{B}_\psi^{k,\gamma}(U; \mathbb{R}^d)$ and every $N \in \mathbb{N}$ there exists a neural network $\varphi_N \in \mathcal{NN}_{U,d}^\rho$ having N neurons satisfying[9]

$$\|f - \varphi_N\|_{W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)} \leqslant C_p \|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})} \frac{C_{U,w}^{(\gamma,p)} m^{\frac{k}{p}} \pi^{\frac{m+1}{4}}}{\left| C_m^{(\psi,\rho)} \right| \Gamma \left(\frac{m+1}{2} \right)^{\frac{1}{2}}} \frac{\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)}}{N^{1 - \frac{1}{\min(2,p)}}}. \tag{8}$$

Theorem 3.6 provides us with an upper bound on the number of neurons $N \in \mathbb{N}$ that are needed for a neural network to approximate a given function. Let us compare Theorem 3.6 with the literature.

Remark 3.7. *Theorem 3.6 generalizes the following approximation rates in the literature.*

(i) *The rate $\mathcal{O}\left(1/N^{1/2}\right)$ of Barron [6] for approximating functions $f : \mathbb{R}^m \to \mathbb{R}$ with sufficiently integrable Fourier transform by neural networks (with sigmoidal activation) within $L^2(B_r(0), \mathcal{B}(B_r(0)), \mu)$, where $B_r(0) := \{u \in \mathbb{R}^m : \|u\| \leqslant r\}$ and μ is a probability measure.*

(ii) *The rate $\mathcal{O}\left(1/N^{1 - 1/\min(2,p)}\right)$ of Darken et al. [13] for approximating functions $f : \mathbb{R}^m \to \mathbb{R}$ being in the convex closure of $\mathcal{NN}_{U,1}^\rho$ by neural networks (with sigmoidal activation ρ) within $L^p(U, \mathcal{B}(U), \mu)$, where $p \in [1, \infty)$, $U \subseteq \mathbb{R}^m$, and $(U, \mathcal{B}(U), \mu)$ is a finite measure space.*

(iii) *The rate $\mathcal{O}\left(1/N^{1/2}\right)$ of Siegel and Xu [48] for approximating functions $f : \mathbb{R}^m \to \mathbb{R}$ with sufficiently integrable Fourier transform by neural networks (with linear combination of polynomially decaying activation functions) in $W^{k,2}(U, \mathcal{L}(U), du)$, where $U \subseteq \mathbb{R}^m$ is open and bounded.*

Hence, the contribution of Theorem 3.6 are approximation rates for neural networks within weighted Sobolev spaces over unbounded domains.

Next, we give a sufficient condition for a function $f : U \to \mathbb{R}^d$ to belong to the space $\mathbb{B}_\psi^{k,\gamma}(U; \mathbb{R}^d)$. The proof of the remaining results of this section can be found in Section 4.3.5.

Proposition 3.8. *Let $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), $\gamma \in [0, \infty)$, and let $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$ such that $\zeta_1 := \inf \left\{ |\zeta| : \zeta \in \text{supp}(\widehat{\psi}) \right\} > 0$. Then, there exists a constant $C_1 > 0$ (independent of $m, d \in \mathbb{N}$) such that for any $f \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)$ with $(\lceil \gamma \rceil + 2)$-times differentiable Fourier transform, we have*

$$\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)} \leqslant \frac{C_1}{\zeta_1^{\frac{m}{2}}} \sum_{\beta \in \mathbb{N}_{0,\lceil\gamma\rceil+2}^m} \left(\int_{\mathbb{R}^m} \left\| \partial_\beta \widehat{f}(\xi) \right\|^2 \left(1 + \|\xi/\zeta_1\|^2\right)^{\frac{4\lceil\gamma\rceil + 2k + m + 5}{2}} d\xi \right)^{\frac{1}{2}}. \tag{9}$$

---

[9]Hereby, Γ denotes the Gamma function (see [1, Section 6.1]).

In particular, if the right-hand side of (9) *is finite, then* $f \in \mathbb{B}^{k,\gamma}_{\psi}(U; \mathbb{R}^d)$. *By using [19, Theorem 7.8.(c)] and the definition of Sobolev spaces via Fourier transform (see [23, Definition 6.9]), this is the case if* $\left(u \mapsto u^{\beta} f(u)\right) \in W^{4\lceil \gamma \rceil + 2k + m + 5, 2}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)$ *for all* $\beta \in \mathbb{N}^m_{0,\lceil \gamma \rceil + 2}$, *where* $u^{\beta} := \prod_{l=1}^m u_l^{\beta_l}$.

In addition, we analyze the situation when neural networks overcome the curse of dimensionality in the sense that the computational costs (here measured as the number of neurons $N \in \mathbb{N}$) grow polynomially in both the dimensions $m, d \in \mathbb{N}$ and the reciprocal of a pre-specified error tolerance $\varepsilon > 0$. To this end, we estimate the constant $C^{(\gamma,p)}_{U,w}$, while a lower bound for $\left| C^{(\psi,\rho)}_m \right|$ is given below Example 3.3.

Lemma 3.9. *Let* $k \in \mathbb{N}_0$, $p \in [1,\infty)$, $U \subseteq \mathbb{R}^m$ *(open, if* $k \geqslant 1$*),* $\gamma \in [0,\infty)$, *and let* $U \ni u := (u_1, ..., u_m)^{\top} \mapsto w(u) := \prod_{l=1}^m w_0(u_l) \in [0,\infty)$ *be a weight, where* $w_0 : \mathbb{R} \to [0,\infty)$ *satisfies* $\int_{\mathbb{R}} w_0(s) ds = 1$ *and* $C^{(\gamma,p)}_{\mathbb{R},w_0} := \left(\int_{\mathbb{R}} (1 + |s|)^{\gamma p} w_0(s) ds \right)^{1/p} < \infty$. *Then,* $C^{(\gamma,p)}_{U,w} \leqslant C^{(\gamma,p)}_{\mathbb{R},w_0} m^{\gamma + 1/p}$.

Proposition 3.10. *For* $k \in \mathbb{N}_0$, $p \in (1,\infty)$, $U \subseteq \mathbb{R}^m$ *(open, if* $k \geqslant 1$*), and* $\gamma \in [0,\infty)$, *let* $w : U \to [0,\infty)$ *be a weight as in Lemma 3.9. Moreover, let* $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C^k_{pol,\gamma}(\mathbb{R})$ *be a pair as in Example 3.3. In addition, let* $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ *satisfy the conditions of Proposition 3.8 such that the right-hand side of* (9) *satisfies* $\mathcal{O}\left(m^s (2/\zeta_2)^m (m+1)^{m/2} \right)$ *for some* $s \in \mathbb{N}_0$. *Then, there exist some constants* $C_2, C_3 > 0$ *such that for every* $m, d \in \mathbb{N}$ *and every* $\varepsilon > 0$ *there exists a neural network* $\varphi_N \in \mathcal{NN}^{\rho}_{U,d}$ *with* $N = \left\lceil C_2 m^{C_3} \varepsilon^{-\frac{\min(2,p)}{\min(2,p)-1}} \right\rceil$ *neurons satisfying* $\|f - \varphi_N\|_{W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)} \leqslant \varepsilon$.

4. PROOFS

4.1. **Proof of results in Section 1.** In this section, we show an equivalent characterization for functions in the weighted C^k-space $\left(\overline{C^k_b(U; \mathbb{R}^d)}^{\gamma}, \| \cdot \|_{C^k_{pol,\gamma}(U;\mathbb{R}^d)} \right)$ introduced in Section 1.2, where $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), and $\gamma \in (0,\infty)$. This generalizes the results in [16, Theorem 2.7] and [11, Lemma 2.7] to differentiable functions defined on an open subset of a Euclidean space $\mathbb{R}^m$.

In the following, we denote the factorial of a multi-index $\alpha := (\alpha_1, ..., \alpha_m) \in \mathbb{N}^m_0$ by $\alpha! := \prod_{l=1}^m \alpha_l!$. Moreover, for any $r \geqslant 0$ and $u_0 \in \mathbb{R}^m$, we define $B_r(u_0) := \{u \in \mathbb{R}^m : \|u - u_0\| < r\}$ and $\overline{B_r(u_0)} := \{u \in \mathbb{R}^m : \|u - u_0\| \leqslant r\}$ as the open and closed ball with radius $r > 0$ around $u_0 \in \mathbb{R}^m$, respectively.

Lemma 4.1. *Let* $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ *(open, if* $k \geqslant 1$*), and* $\gamma \in (0,\infty)$. *Then, the following holds true:*

(i) *If* $U \subseteq \mathbb{R}^m$ *is bounded, then* $\overline{C^k_b(U; \mathbb{R}^d)}^{\gamma} = C^k_b(U; \mathbb{R}^d)$.

(ii) *If* $U \subseteq \mathbb{R}^m$ *is unbounded, then* $f \in \overline{C^k_b(U; \mathbb{R}^d)}^{\gamma}$ *if and only if* $f \in C^k(U; \mathbb{R}^d)$ *and*

$$\lim_{r \to \infty} \max_{\alpha \in \mathbb{N}^m_{0,k}} \sup_{u \in U \setminus B_r(0)} \frac{\|\partial_{\alpha} f(u)\|}{(1 + \|u\|)^{\gamma}} = 0. \tag{10}$$

Proof. The conclusion in (i) follows from the definition of $\left(\overline{C^k_b(U; \mathbb{R}^d)}^{\gamma}, \| \cdot \|_{C^k_{pol,\gamma}(U;\mathbb{R}^d)} \right)$. Now, for necessity in (ii), fix some $f \in \overline{C^k_b(U; \mathbb{R}^d)}^{\gamma}$. Then, by definition of $\overline{C^k_b(U; \mathbb{R}^d)}^{\gamma}$, there exists a sequence $(g_n)_{n \in \mathbb{N}} \subseteq C^k_b(U; \mathbb{R}^d)$ with $\lim_{n \to \infty} \|f - g_n\|_{C^k_{pol,\gamma}(U;\mathbb{R}^d)} = 0$, which implies for every fixed $r > 0$ that

$$\lim_{n \to \infty} \max_{\alpha \in \mathbb{N}^m_{0,k}} \sup_{u \in U \cap B_r(0)} \|\partial_{\alpha} f(u) - \partial_{\alpha} g_n(u)\| \leqslant (1+r)^{\gamma} \lim_{n \to \infty} \max_{\alpha \in \mathbb{N}^m_{0,k}} \sup_{u \in U \cap B_r(0)} \frac{\|\partial_{\alpha} f(u) - \partial_{\alpha} g_n(u)\|}{(1 + \|u\|)^{\gamma}}$$
$$\leqslant (1+r)^{\gamma} \lim_{n \to \infty} \|f - g_n\|_{C^k_{pol,\gamma}(U;\mathbb{R}^d)} = 0.$$

This together with the Fundamental Theorem of Calculus shows that $f|_{U \cap B_r(0)} : U \cap B_r(0) \to \mathbb{R}^d$ is k-times differentiable since for every fixed $\alpha \in \mathbb{N}^m_{0,k}$ the partial derivative $\partial_{\alpha} f|_{U \cap B_r(0)} : U \cap B_r(0) \to \mathbb{R}^d$ is continuous as uniform limit of continuous functions. Hence, by using that U is locally compact, it follows from [44, Lemma 46.3+46.4] that $\partial_{\alpha} f : U \to \mathbb{R}^d$ is continuous everywhere on U. Since this holds true for every $\alpha \in \mathbb{N}^m_{0,k}$, we apply again the Fundamental Theorem of Calculus to conclude that $f \in C^k(U; \mathbb{R}^d)$. Moreover, in order to show (10), we fix some $\varepsilon > 0$ and choose some $n \in \mathbb{N}$ large enough such that $\|f - g_n\|_{C^k_{pol,\gamma}(U;\mathbb{R}^d)} < \varepsilon/2$. Moreover, we choose $r > 0$ sufficiently large such that

$(1+r)^\gamma > 2\varepsilon^{-1}\|g_n\|_{C_b^k(U;\mathbb{R}^d)}$ holds true, which implies that

$$\max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U\backslash \overline{B_r(0)}}\frac{\|\partial_\alpha f(u)\|}{(1+\|u\|)^\gamma} \leqslant \max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U}\frac{\|\partial_\alpha f(u)-\partial_\alpha g_n(u)\|}{(1+\|u\|)^\gamma} + \max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U\backslash \overline{B_r(0)}}\frac{\|\partial_\alpha g_n(u)\|}{(1+\|u\|)^\gamma}$$

$$< \frac{\varepsilon}{2} + \frac{\|g_n\|_{C_b^k(U;\mathbb{R}^d)}}{(1+r)^\gamma} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrarily, we obtain (10).

For sufficiency in (ii), let $f \in C^k(U;\mathbb{R}^d)$ such that (10) holds true and fix some $\varepsilon > 0$. Moreover, we choose some $h \in C_c^\infty(\mathbb{R}^m)$ such that $h(u) = 1$ for all $u \in \overline{B_1(0)}$, $h(u) = 0$ for all $u \in \mathbb{R}^m\backslash B_2(0)$, and that there exists a constant $C_h > 0$ such that for every $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in \mathbb{R}^m$ it holds that $|\partial_\alpha h(u)| \leqslant C_h$. In addition, by using (10), there exists some $r > 1$ such that

$$\max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U\backslash B_r(0)}\frac{\|\partial_\alpha f(u)\|}{(1+\|u\|)^\gamma} < \frac{\varepsilon}{1+2^k C_h}. \tag{11}$$

From this, we define the functions $\mathbb{R}^m \ni u \mapsto h_r(u) := h(u/r) \in \mathbb{R}$ and $U \ni u \mapsto g(u) := h_r(u)f(u) \in \mathbb{R}^d$, which both have bounded support. Furthermore, note that by the binomial theorem, we have for every $\alpha \in \mathbb{N}_0^m$ that

$$\sum_{\substack{\beta_1,\beta_2\in\mathbb{N}_0^m\\\beta_1+\beta_2=\alpha}}\frac{\alpha!}{\beta_1!\beta_2!} = \sum_{\substack{\beta\in\mathbb{N}_0^m\\\forall l:\beta_l\leqslant\alpha_l}}\prod_{l=1}^m\frac{\alpha_l!}{\beta_l!(\alpha_l-\beta_l)!} \leqslant \prod_{l=1}^m\sum_{\beta_l=0}^{\alpha_l}\frac{\alpha_l!}{\beta_l!(\alpha_l-\beta_l)!} = \prod_{l=1}^m 2^{\alpha_l} \leqslant 2^{|\alpha|}. \tag{12}$$

Then, by using the Leibniz product rule together with the triangle inequality, the inequality (12), that $|\partial_\alpha h_r(u)| = |\partial_\alpha h(u/r)|\, r^{-|\alpha|} \leqslant C_h$ for any $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in \mathbb{R}^m$, and again the inequality (12), it follows for every $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in U$ that

$$\|\partial_\alpha g(u)\| \leqslant \sum_{\substack{\beta_1,\beta_2\in\mathbb{N}_0^m\\\beta_1+\beta_2=\alpha}}\frac{\alpha!}{\beta_1!\beta_2!}|\partial_{\beta_1}h_r(u)|\,\|\partial_{\beta_2}f(u)\| \leqslant 2^k C_h \max_{\beta_2\in\mathbb{N}_{0,k}^m}\|\partial_{\beta_2}f(u)\|. \tag{13}$$

Hence, by using that $\partial_\alpha g(u) = \partial_\alpha(h_r(u)f(u)) = \partial_\alpha f(u)$ for any $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in U \cap B_r(0)$ (as $h_r(u) = 1$ for any $u \in B_r(0)$), and the inequalities (13) and (11), the function $g \in C_b^k(U;\mathbb{R}^d)$ satisfies

$$\|f-g\|_{C_{pol,\gamma}^k(U;\mathbb{R}^d)} = \max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U}\frac{\|\partial_\alpha f(u)-\partial_\alpha g(u)\|}{(1+\|u\|)^\gamma}$$

$$\leqslant \max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U\cap B_r(0)}\frac{\|\partial_\alpha f(u)-\partial_\alpha g(u)\|}{(1+\|u\|)^\gamma} + \max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U\backslash B_r(0)}\frac{\|\partial_\alpha f(u)\|}{(1+\|u\|)^\gamma}$$

$$+ \max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U\backslash B_r(0)}\frac{\|\partial_\alpha g(u)\|}{(1+\|u\|)^\gamma}$$

$$\leqslant \max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U\backslash B_r(0)}\frac{\|\partial_\alpha f(u)\|}{(1+\|u\|)^\gamma} + 2^k C_h\max_{\alpha\in\mathbb{N}_{0,k}^m}\sup_{u\in U\backslash B_r(0)}\frac{\|\partial_\alpha f(u)\|}{(1+\|u\|)^\gamma}$$

$$< \frac{\varepsilon}{1+2^k C_h} + 2^k C_h\frac{\varepsilon}{1+2^k C_h} = \varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrarily, it follows that $f \in \overline{C^k(U;\mathbb{R}^d)}^\gamma$. $\qquad\square$

4.2. **Proof of results in Section 2.**

4.2.1. *Proof of Lemma 2.5.*

Proof of Lemma 2.5. Fix some $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$, $y \in \mathbb{R}^d$, $a \in \mathbb{R}^m$, and $b \in \mathbb{R}$, and define the constant $C_{y,a,b} := 1+\max_{\alpha\in\mathbb{N}_{0,k}^m}\|ya^\alpha\|\,(1+\|a\|+|b|)^\gamma > 0$, where $a^\alpha := \prod_{l=1}^m a_l^{\alpha_l}$ for $a := (a_1,...,a_m)^\top \in \mathbb{R}^m$ and $\alpha := (\alpha_1,...,\alpha_m) \in \mathbb{N}_{0,k}^m$. Then, by using the definition of $\overline{C_b^k(\mathbb{R})}^\gamma$, there exists some $\widetilde{\rho} \in C_b^k(\mathbb{R})$ such that

$$\|\rho-\widetilde{\rho}\|_{C_{pol,\gamma}^k(\mathbb{R})} := \max_{j=0,...,k}\sup_{s\in\mathbb{R}}\frac{\left|\rho^{(j)}(s)-\widetilde{\rho}^{(j)}(s)\right|}{(1+|s|)^\gamma} < \frac{\varepsilon}{C_{y,a,b}}.$$

Hence, by using the inequality $1 + \left|a^\top u - b\right| \leqslant 1 + \|a\|\|u\| + |b| \leqslant (1 + \|a\| + |b|)(1 + \|u\|)$, it follows for the function $y\widetilde{\rho}\left(a^\top \cdot -b\right) := \left(u \mapsto y\widetilde{\rho}\left(a^\top u - b\right)\right) \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)$ that

$$
\begin{aligned}
\left\|y\rho\left(a^\top \cdot -b\right) - y\widetilde{\rho}\left(a^\top \cdot -b\right)\right\|_{C_{pol,\gamma}^k(\mathbb{R}^m;\mathbb{R}^d)} &= \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in \mathbb{R}^m} \frac{\left\|y\rho^{(|\alpha|)}\left(a^\top u - b\right) a^\alpha - y\widetilde{\rho}^{(|\alpha|)}\left(a^\top u - b\right) a^\alpha\right\|}{(1 + \|u\|)^\gamma} \\
&\leqslant \left(\max_{\alpha \in \mathbb{N}_{0,k}^m} \|ya^\alpha\|(1 + \|a\| + |b|)\right) \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in \mathbb{R}^m} \frac{\left\|\rho^{(|\alpha|)}\left(a^\top u - b\right) - \widetilde{\rho}^{(|\alpha|)}\left(a^\top u - b\right)\right\|}{(1 + |a^\top u - b|)^\gamma} \\
&\leqslant C_{y,a,b} \max_{j=0,\ldots,k} \sup_{s \in \mathbb{R}} \frac{\left|\rho^{(j)}(s) - \widetilde{\rho}^{(j)}(s)\right|}{(1 + |s|)^\gamma} < C_{y,a,b} \frac{\varepsilon}{C_{y,a,b}} = \varepsilon.
\end{aligned}
$$

Since $\varepsilon > 0$ was chosen arbitrarily and $y\widetilde{\rho}\left(a^\top \cdot -b\right) \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)$, it follows that $y\rho\left(a^\top \cdot -b\right) \in \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma$. Thus, by using that $\mathcal{NN}_{\mathbb{R}^m,d}^\rho$ is defined as vector space consisting of functions of the form $\mathbb{R}^m \ni u \mapsto y\rho\left(a^\top u - b\right) \in \mathbb{R}^d$, with $y \in \mathbb{R}^d$, $a \in \mathbb{R}^m$, and $b \in \mathbb{R}$, the triangle inequality implies that $\mathcal{NN}_{\mathbb{R}^m,d}^\rho \subseteq \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma$. Finally, by using that $(X, \|\cdot\|_X)$ satisfies Assumption 2.3, i.e. that the restriction map (4) is a continuous embedding, it follows that $\mathcal{NN}_{U,d}^\rho \subseteq X$. $\qquad\square$

4.2.2. *Proof of Theorem 2.7.* In this section, we provide the proof of Theorem 2.7, i.e. the universal approximation property of neural networks $\mathcal{NN}_{U,d}^\rho$ within any function space $(X, \|\cdot\|_X)$ satisfying Assumption 2.3 with $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), and $\gamma \in (0, \infty)$, where $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$.

The idea of the proof is the following. By contradiction, we assume that $\mathcal{NN}_{U,d}^\rho \subseteq X$ is not dense in X. Then, by using the Hahn-Banach theorem (as in [12, Theorem 1]), there exists a non-zero continuous linear functional $l : X \to \mathbb{R}$ which vanishes on the vector subspace $\mathcal{NN}_{U,d}^\rho \subseteq X$. Moreover, by using the continuous embedding in (4), we can express $l : X \to \mathbb{R}$ on the dense subspace $\overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma$ with finite signed Radon measures, which relies on the Riesz representation theorem in [16, Theorem 2.4]. Subsequently, we use the distributional extension of Wiener's Tauberian theorem in [31] and that $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$ is non-polynomial to conclude that $l : X \to \mathbb{R}$ vanishes everywhere on X. This however contradicts the initial assumption that $l : X \to \mathbb{R}$ is non-zero. Hence, $\mathcal{NN}_{U,d}^\rho$ must be dense in X.

In order to prove Theorem 2.7 as outlined above, we now first generalize the Riesz representation theorem in [16, Theorem 2.7] to this vector-valued case with derivatives. Hereby, we define $\mathcal{M}_\gamma(\mathbb{R}^m)$ as the vector space of finite signed Radon measures $\eta : \mathcal{B}(\mathbb{R}^m) \to \mathbb{R}$ with $\int_{\mathbb{R}^m}(1 + \|u\|)^\gamma |\eta|(du) < \infty$, where $|\eta| : \mathcal{B}(\mathbb{R}^m) \to [0, \infty)$ denotes the corresponding total variation measure. Moreover, we denote by Z^* the dual space of a Banach space $(Z, \|\cdot\|_Z)$ which consists of continuous linear functionals $l : Z \to \mathbb{R}$ and is equipped with the norm $\|l\|_{Z^*} := \sup_{z \in Z, \|z\|_Z \leqslant 1} |l(z)|$.

Proposition 4.2 (Riesz representation). *For $k \in \mathbb{N}_0$ and $\gamma \in (0, \infty)$, let $l : \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma \to \mathbb{R}$ be a continuous linear functional. Then, there exist some signed Radon measures $(\eta_{\alpha,i})_{\alpha \in \mathbb{N}_{0,k}^m, i=1,\ldots,d} \subseteq \mathcal{M}_\gamma(\mathbb{R}^m)$ such that for every $f = (f_1, \ldots, f_d)^\top \in \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma$ it holds that*

$$
l(f) = \sum_{\alpha \in \mathbb{N}_{0,k}^m} \sum_{i=1}^d \int_{\mathbb{R}^m} \partial_\alpha f_i(u) \eta_{\alpha,i}(du).
$$

Proof. First, we show the conclusion for $k = 0$ and $d = 1$. Indeed, by defining $\mathbb{R}^m \ni u \mapsto \psi(u) := (1 + \|u\|)^\gamma \in (0, \infty)$, the tuple $(\mathbb{R}^m, \psi)$ is a weighted space in the sense of [16, p. 5]. Hence, the conclusion follows from [16, Theorem 2.4].

Now, for the general case of $k \geqslant 1$ and $d \geqslant 2$, we fix a continuous linear functional $l : \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma \to \mathbb{R}$ and define the number $M := |\mathbb{N}_{0,k}^m| \cdot d$ as well as the map

$$
\overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma \ni f \quad \mapsto \quad \Xi(f) := (\partial_\alpha f_i)_{\alpha \in \mathbb{N}_{0,k}^m, i=1,\ldots,d}^\top \in \overline{C_b^0(\mathbb{R}^m; \mathbb{R}^M)}^\gamma.
$$

Moreover, we denote by $\text{Img}(\Xi) := \left\{\Xi(f) : f \in \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma\right\} \subseteq \overline{C_b^0(\mathbb{R}^m; \mathbb{R}^M)}^\gamma$ the image vector subspace. Then, by using that $\Xi : \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma \to \text{Img}(\Xi)$ is by definition bijective, there exists an

inverse map $\Xi^{-1} : \mathrm{Img}(\Xi) \to \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma$. Moreover, we conclude for every $f \in \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma$ that

$$
\begin{aligned}
\left\| \Xi^{-1}((\partial_\alpha f_i)_{\alpha\in\mathbb{N}_{0,k}^m,\, i=1,\ldots,d}) \right\|_{C_{pol,\gamma}^0(\mathbb{R}^m;\mathbb{R}^d)} &= \|f\|_{C_{pol,\gamma}^k(\mathbb{R}^m;\mathbb{R}^d)} = \max_{\alpha\in\mathbb{N}_{0,k}^m} \sup_{u\in\mathbb{R}^m} \frac{\|\partial_\alpha f(u)\|}{(1+\|u\|)^\gamma} \\
&= \sup_{u\in\mathbb{R}^m} \max_{\alpha\in\mathbb{N}_{0,k}^m} \frac{\|\partial_\alpha f(u)\|}{(1+\|u\|)^\gamma} \leqslant \sup_{u\in\mathbb{R}^m} \frac{\|(\partial_\alpha f_i)_{\alpha\in\mathbb{N}_{0,k}^m,\, i=1,\ldots,d}\|}{(1+\|u\|)^\gamma} \\
&= \|(\partial_\alpha f_i)_{\alpha\in\mathbb{N}_{0,k}^m,\, i=1,\ldots,d}\|_{C_{pol,\gamma}^0(\mathbb{R}^m;\mathbb{R}^M)},
\end{aligned}
$$

which shows that $\Xi^{-1} : \mathrm{Img}(\Xi) \to \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma$ is continuous. Hence, the concatenation $l \circ \Xi^{-1} : \mathrm{Img}(\Xi) \to \mathbb{R}$ is a continuous linear functional on $\mathrm{Img}(\Xi)$, which can be extended by using the Hahn-Banach theorem to a continuous linear functional $l_0 : \overline{C_b^0(\mathbb{R}^m;\mathbb{R}^M)}^\gamma \to \mathbb{R}$ such that for every $f \in \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma$ it holds that

$$l_0((\partial_\alpha f_i)_{\alpha\in\mathbb{N}_{0,k}^m,\, i=1,\ldots,d}) = \left(l \circ \Xi^{-1}\right)((\partial_\alpha f_i)_{\alpha\in\mathbb{N}_{0,k}^m,\, i=1,\ldots,d}) = l(f). \tag{14}$$

Now, for every fixed $\alpha \in \mathbb{N}_{0,k}^m$ and $i = 1,\ldots,d$, we define the linear map $\overline{C_b^0(\mathbb{R}^m)}^\gamma \ni g \mapsto l_{\alpha,i}(g) := l_0(ge_{\alpha,i}) \in \mathbb{R}$, where $e_{\alpha,i} \in \mathbb{R}^M := \mathbb{R}^{|\mathbb{N}_{0,k}^m|\cdot d} \cong \mathbb{R}^{|\mathbb{N}_{0,k}^m|} \times \mathbb{R}^d$ denotes the (α,i)-th unit vector of $\mathbb{R}^M := \mathbb{R}^{|\mathbb{N}_{0,k}^m|\cdot d} \cong \mathbb{R}^{|\mathbb{N}_{0,k}^m|} \times \mathbb{R}^d$. Then, for every $g \in \overline{C_b^0(\mathbb{R}^m)}^\gamma$, it follows with $Z := \overline{C_b^0(\mathbb{R}^m;\mathbb{R}^M)}^\gamma$ that

$$|l_{\alpha,i}(g)| = |l_0(ge_{\alpha,i})| \leqslant \|l_0\|_{Z*} \|ge_{\alpha,i}\|_{C_{pol,\gamma}^0(\mathbb{R}^m;\mathbb{R}^M)} = \|l_0\|_{Z*} \|g\|_{C_{pol,\gamma}^0(\mathbb{R}^m)},$$

which shows that $l_{\alpha,i} : \overline{C_b^0(\mathbb{R}^m)}^\gamma \to \mathbb{R}$ is a continuous linear functional. Hence, by using (14) and by applying for every $\alpha \in \mathbb{N}_{0,k}^m$ and $i = 1,\ldots,d$ the case with $k = 0$ and $d = 1$, there exist some Radon measures $(\eta_{\alpha,i})_{\alpha\in\mathbb{N}_{0,k}^m,\, i=1,\ldots,d} \in \mathcal{M}_\gamma(\mathbb{R}^m)$ such that for every $f \in \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma$ it holds that

$$
\begin{aligned}
l(f) &= \left(l \circ \Xi^{-1}\right)((\partial_\alpha f_i)_{\alpha\in\mathbb{N}_{0,k}^m,\, i=1,\ldots,d}) \\
&= l_0((\partial_\alpha f_i)_{\alpha\in\mathbb{N}_{0,k}^m,\, i=1,\ldots,d}) \\
&= \sum_{\alpha\in\mathbb{N}_{0,k}^m} \sum_{i=1}^d l_{\alpha,i}(\partial_\alpha f_i e_{\alpha,i}) \\
&= \sum_{\alpha\in\mathbb{N}_{0,k}^m} \sum_{i=1}^d \int_{\mathbb{R}^m} \partial_\alpha f_i(u) \eta_{\alpha,i}(du),
\end{aligned}
$$

which completes the proof. $\qquad\square$

Next, we show that every non-polynomial activation function $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$ is discriminatory in the sense of [12, p. 306]. To this end, we generalize the proof of [10, Theorem 1] from compactly supported signed Radon measures to measures in $\mathcal{M}_\gamma(\mathbb{R}^m)$. Hereby, we follow the distributional extension of Wiener's Tauberian theorem in [31, Theorem A].

Proposition 4.3. *For $\gamma \in (0,\infty)$, let $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ be a signed Radon measure and assume that $\rho \in \overline{C_b^0(\mathbb{R})}^\gamma$ is non-polynomial. If for every $a \in \mathbb{R}^m$ and $b \in \mathbb{R}$ it holds that*

$$\int_{\mathbb{R}^m} \rho\left(a^\top u - b\right) \eta(du) = 0, \tag{15}$$

then it follows that $\eta = 0 \in \mathcal{M}_\gamma(\mathbb{R}^m)$.

Proof. We follow the proof of [11, Proposition 4.4 (A3)] and assume that $\rho \in \overline{C_b^0(\mathbb{R})}^\gamma$ is non-polynomial. Then, by using [47, Examples 7.16], there exists a non-zero point $t_0 \in \mathbb{R}\backslash\{0\}$ which belongs to the support of $\widehat{T_\rho} \in \mathcal{S}'(\mathbb{R};\mathbb{C})$. Moreover, let $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ satisfy (15) and assume by contradiction that $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ is non-zero.

Now, for every $a \in \mathbb{R}^m$, we define the push-forward measure $\eta_a := \eta \circ \left(a^\top \cdot\right)^{-1} : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ by $\eta_a(B) := \eta\left(\{u \in \mathbb{R}^m : a^\top u \in B\}\right)$, for $B \in \mathcal{B}(\mathbb{R})$. Moreover, for every fixed $\lambda \in \mathbb{R}\backslash\{0\}$, we define the function $\mathbb{R} \ni s \mapsto \rho_\lambda(s) := \rho(\lambda s) \in \mathbb{R}$. Then, by applying [7, Theorem 3.6.1] (to the positive and

negative part of $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$) and by using the assumption (15) (with $\lambda a \in \mathbb{R}^m$ and $\lambda b \in \mathbb{R}$ instead of $a \in \mathbb{R}^m$ and $b \in \mathbb{R}$, respectively), it follows for every $a \in \mathbb{R}^m$ and $b \in \mathbb{R}$ that

$$\int_\mathbb{R} \rho_\lambda(s-b)\eta_a(ds) = \int_{\mathbb{R}^m} \rho\left(\lambda a^\top u - \lambda b\right) \eta(du) = 0. \tag{16}$$

Since $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ is non-zero, there exists some $a \in \mathbb{R}^m$ such that $\eta_a : \mathcal{B}(\mathbb{R}) \to \mathbb{R}$ is non-zero. Hence, there exists some $h \in \mathcal{S}(\mathbb{R}; \mathbb{C})$ such that $\left(z \mapsto f(z) := (h * \eta_a)(-z) := \int_\mathbb{R} h(-z-s)\eta_a(ds)\right) \in L^1(\mathbb{R}, \mathcal{L}(\mathbb{R}), du; \mathbb{C})$ is also non-zero. Then, by using that the Fourier transform is injective, $\widehat{f} : \mathbb{R} \to \mathbb{C}$ is non-zero, too, i.e. there exists some $t_1 \in \mathbb{R}\backslash\{0\}$ such that $\widehat{f}(t_1) \neq 0$. Hence, by using [19, Table 7.2.2], the function $\left(z \mapsto f_0(z) := f(z)e^{-it_1 z}\right) \in L^1(\mathbb{R}, \mathcal{L}(\mathbb{R}), du; \mathbb{C})$ satisfies $\widehat{f_0}(0) = \widehat{f}(t_1) \neq 0$. Moreover, we choose $\lambda := \frac{t_1}{t_0} \in \mathbb{R}\backslash\{0\}$ and define the function $\mathbb{R} \ni z \mapsto \rho_0(z) := \rho_\lambda(z)e^{-it_1 z} \in \mathbb{C}$.

Next, we use [7, Theorem 3.6.1] (applied to $|\eta| : \mathcal{B}(\mathbb{R}^m) \to [0, \infty)$), the inequality $1 + \left|\lambda a^\top u - b\right| \leqslant 1 + |\lambda|\|a\|\|u\| + |\lambda||b| \leqslant \max(1, |\lambda|)(1 + \|a\|)(1 + |b|)(1 + \|u\|)$ for any $a, u \in \mathbb{R}^m$ and $b, y \in \mathbb{R}$, the inequality $(1 + |b|)^\gamma \leqslant 2^\gamma\left(1 + |b|^2\right)^{\gamma/2} \leqslant 2^\gamma\left(1 + |b|^2\right)^{\lceil\gamma/2\rceil}$ for any $b \in \mathbb{R}$, and that for every $y \in \mathbb{R}$ the reflected translation $\mathbb{R} \ni b \mapsto \widetilde{h}_y(b) := h(-y-b) \in \mathbb{R}$ of the Schwartz function $h \in \mathcal{S}(\mathbb{R}; \mathbb{C})$ is again a Schwartz function (see [19, p. 331]) to conclude for every $y \in \mathbb{R}$ that

$$\int_\mathbb{R} \int_\mathbb{R} |h(-y-b)||\rho_\lambda(s-b)||\eta_a|(ds)db = \int_{\mathbb{R}^m} |h(-y-b)| \int_\mathbb{R} \left|\rho\left(\lambda a^\top u - \lambda b\right)\right| |\eta|(du)db$$

$$\leqslant \int_\mathbb{R} |h(-y-b)| \left(\sup_{u \in \mathbb{R}^m} \frac{\left|\rho\left(\lambda a^\top u - \lambda b\right)\right|}{\left(1 + |\lambda a^\top u - \lambda b|\right)^\gamma}\right) \int_\mathbb{R} \left(1 + \left|\lambda a^\top u - \lambda b\right|\right)^\gamma |\eta|(du)db$$

$$\leqslant \max(1, |\lambda|)^\gamma(1 + \|a\|)^\gamma \left(\sup_{s \in \mathbb{R}} \frac{|\rho(s)|}{(1 + |s|)^\gamma}\right) \left(\int_\mathbb{R} |h(-y-b)|(1 + |b|)^\gamma db\right) \int_\mathbb{R} (1 + \|u\|)^\gamma |\eta|(du)$$

$$\leqslant \max(1, |\lambda|)^\gamma(1 + \|a\|)^\gamma \|\rho\|_{C^0_{pol,\gamma}(\mathbb{R})} \left(\sup_{y \in \mathbb{R}} \left|\widetilde{h}_y(b)\right| \left(1 + |b|^2\right)^{\lceil\gamma/2\rceil+1}\right)$$

$$\cdot \left(\int_\mathbb{R} \frac{1}{1+b^2} db\right) \int_\mathbb{R} (1 + \|u\|)^\gamma |\eta|(du) < \infty. \tag{17}$$

Then, by using the substitution $z \mapsto s - b$ and the identity (16), it follows for every $y \in \mathbb{R}$ that

$$(f_0 * \rho_0)(y) = \int_\mathbb{R} f(y-z)e^{it_1(y-z)}\rho_\lambda(z)e^{-it_1 z}dz = e^{it_1 y}\int_\mathbb{R} (h * \eta_a)(z-y)\rho_\lambda(z)dz$$

$$= e^{it_1 y}\int_\mathbb{R} \int_\mathbb{R} h(z-y-s)\rho_\lambda(z)\eta_a(ds)dz = e^{it_1 y}\int_\mathbb{R} h(-y-b)\int_\mathbb{R} \rho_\lambda(s-b)\eta_a(ds)db = 0, \tag{18}$$

where (17) ensures that the convolution $f_0 * \rho_0 : \mathbb{R} \to \mathbb{R}$ is well-defined.

Moreover, let $\phi \in \mathcal{S}(\mathbb{R}; \mathbb{C})$ such that $\widehat{\phi}(\xi) = 1$, for all $\xi \in [-1, 1]$, and $\widehat{\phi}(\xi) = 0$, for all $\xi \in \mathbb{R}\backslash[-2, 2]$. In addition, for every $n \in \mathbb{N}$, we define $\left(s \mapsto \phi_n(s) := \frac{1}{n}\phi\left(\frac{1}{n}\right)\right) \in \mathcal{S}(\mathbb{R}; \mathbb{C})$. Then, by following the proof of [31, Theorem A], there exists some large enough $n \in \mathbb{N}$ and $w \in L^1(\mathbb{R}, \mathcal{L}(\mathbb{R}), du)$ such that $w * f_0 = \phi_{2n} \in \mathcal{S}(\mathbb{R}; \mathbb{C})$. Hence, by using (17), we conclude for every $g \in \mathcal{S}(\mathbb{R}; \mathbb{C})$ that

$$\left(T_{\rho_0} * \phi_{2n}\right)(g) := T_{\rho_0}\left(\phi_{2n}(-\cdot) * g\right) = (g * \phi_{2n} * \rho_0)(0) = (g * w * f_0 * \rho_0)(0) = 0, \tag{19}$$

where $\phi_{2n}(-\cdot)$ denotes the function $\mathbb{R} \ni s \mapsto \phi_{2n}(-s) \in \mathbb{R}$. Thus, by using [19, Equation 9.32] together with (19), i.e. that $\widehat{\phi_{2n}}\widehat{T_{\rho_0}} = \widehat{T_{\rho_0} * \phi_{2n}} = 0 \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$, and that $\widehat{\phi_{2n}}(\xi) = \widehat{\phi}(2n\xi) = 1$ for any $\xi \in [-\frac{1}{2n}, \frac{1}{2n}]$, it follows that $\widehat{T_{\rho_0}} \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$ vanishes on $\left(-\frac{1}{2n}, \frac{1}{2n}\right)$.

Finally, for any fixed $g \in C^\infty\left((t_0 - \frac{1}{2n|\lambda|}, t_0 + \frac{1}{2n|\lambda|}); \mathbb{C}\right)$, we define $\left(z \mapsto g_0(z) := g\left(\frac{z}{\lambda} + t_0\right)\right) \in C^\infty_c\left((-\frac{1}{2n}, \frac{1}{2n}); \mathbb{C}\right)$. Hence, by using the definition of $\widehat{T_\rho} \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$, the substitution $\zeta \mapsto \xi/\lambda$, [19, Table 9.2.2], and that $\widehat{T_{\rho_0}} \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$ vanishes on $\left(-\frac{1}{2n}, \frac{1}{2n}\right)$, we conclude that

$$\widehat{T_\rho}(g) = T_\rho(\widehat{g}) = \int_\mathbb{R} \rho(\xi)\widehat{g}(\xi)d\xi = \lambda\int_\mathbb{R} \rho(\lambda\zeta)\widehat{g}(\lambda\zeta)d\zeta = \int_\mathbb{R} \rho_0(\zeta)e^{it_1\zeta}\widehat{g(\cdot/\lambda)}(\zeta)dz$$

$$= \int_\mathbb{R} \rho_0(\zeta)\widehat{g_0}(\zeta)d\zeta = T_{\rho_0}(\widehat{g_0}) = \widehat{T_{\rho_0}}(g_0) = 0, \tag{20}$$

where $\widehat{g(\cdot/\lambda)}$ denotes the Fourier transform of the function $(s \mapsto g(s/\lambda)) \in \mathcal{S}(\mathbb{R}; \mathbb{C})$. Since the function $g \in C_c^\infty((t_0 - \frac{1}{2n|\lambda|}, t_0 + \frac{1}{2n|\lambda|}); \mathbb{C})$ was chosen arbitrary, (20) shows that $\widehat{T_\rho} \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$ vanishes on the set $\left(t_0 - \frac{1}{2n|\lambda|}, t_0 + \frac{1}{2n|\lambda|}\right)$. This however contradicts the assumption that $t_0 \in \mathbb{R} \backslash \{0\}$ belongs to the support of $\widehat{T_\rho} \in \mathcal{S}'(\mathbb{R}; \mathbb{C})$ and shows that $\eta = 0 \in \mathcal{M}_\gamma(\mathbb{R})$. $\qquad\square$

Next, we show some properties of measures $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$, $\gamma \in (0, \infty)$, whenever they are convoluted with a bump function. To this end, we introduce the smooth bump function $\phi : \mathbb{R}^m \to \mathbb{R}$ defined by

$$\phi(u) := \begin{cases} Ce^{-\frac{1}{1-\|u\|^2}}, & u \in B_1(0), \\ 0, & u \in \mathbb{R}^m \backslash B_1(0), \end{cases}$$

where $C > 0$ is a normalizing constant such that $\|\phi\|_{L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du)} = 1$. From this, we define for every fixed $\delta > 0$ the mollifier $\mathbb{R}^m \ni u \mapsto \phi_\delta(u) := \frac{1}{\delta^m} \phi\left(\frac{u}{\delta}\right) \in \mathbb{R}$. Moreover, for any $\gamma \in (0, \infty)$ and $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$, we define the function $\mathbb{R}^m \ni u \mapsto (\phi_\delta * \eta)(u) := \int_{\mathbb{R}^m} \phi_\delta(u - v)\eta(dv) \in \mathbb{R}$.

Lemma 4.4. *For $\gamma \in (0, \infty)$, let $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ and $f \in \overline{C_b^0(\mathbb{R}^m)}^\gamma$. Then, the following holds true:*

(i) *For every $\delta > 0$ the function $\phi_\delta * \eta : \mathbb{R}^m \to \mathbb{R}$ is smooth with $\partial_\alpha(\phi_\delta * \eta)(u) = (\partial_\alpha \phi_\delta * \eta)(u)$ for all $\alpha \in \mathbb{N}_0^m$ and $u \in \mathbb{R}^m$.*

(ii) *For every $\delta > 0$ and $\alpha \in \mathbb{N}_0^m$ it holds that*

$$\lim_{r \to \infty} \sup_{u \in \mathbb{R}^m \backslash \overline{B_r(0)}} |f(u)\partial_\alpha(\phi_\delta * \eta)(u)| = 0.$$

(iii) *For every $\delta > 0$ and $\alpha \in \mathbb{N}_0^m$ it holds that $\partial_\alpha(\phi_\delta * \eta)(u)du\big|_{\mathcal{B}(\mathbb{R}^m)} \in \mathcal{M}_\gamma(\mathbb{R}^m)$.*

(iv) *For every $\delta > 0$ and $\alpha \in \mathbb{N}_0^m$ the map*

$$\left(\overline{C_b^0(\mathbb{R}^m)}^\gamma, \|\cdot\|_{C_{pol,\gamma}^0(\mathbb{R}^m)}\right) \ni f \quad \mapsto \quad \int_{\mathbb{R}^m} f(u)\partial_\alpha(\phi_\delta * \eta)(u)du \in \mathbb{R}$$

is a continuous linear functional.

(v) *For every $\delta > 0$ it holds that*

$$\int_{\mathbb{R}^m} f(u)(\phi_\delta * \eta)(u)du = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} f(u + y)\eta(du)\phi_\delta(y)dy.$$

(vi) *It holds that*

$$\lim_{\delta \to 0} \int_{\mathbb{R}^m} f(u)(\phi_\delta * \eta)(u)du = \int_{\mathbb{R}^m} f(u)\eta(du).$$

Proof. Fix some $\gamma \in (0, \infty)$, $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$, $f \in \overline{C_b^0(\mathbb{R}^m)}^\gamma$, $\delta > 0$, and $\alpha \in \mathbb{N}_0^m$. For (i), we first show that $\partial_\alpha \phi_\delta * \eta : \mathbb{R}^m \to \mathbb{R}$ is continuous. Indeed, we observe that for every $u, u_0, v \in \mathbb{R}^m$, it holds that

$$\max\left(|\partial_\alpha \phi_\delta(u - v)|, |\partial_\alpha \phi_\delta(u_0 - v)|\right) \leqslant C_{11} := \sup_{u_1 \in \mathbb{R}^m} |\partial_\alpha \phi_\delta(u_1)| < \infty. \tag{21}$$

Then, the dominated convergence theorem (with (21) and that $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ is finite) implies that

$$\lim_{u \to u_0} (\partial_\alpha \phi_\delta * \eta)(u) = \lim_{u \to u_0} \int_{\mathbb{R}^m} \partial_\alpha \phi_\delta(u - v)\eta(dv) = \int_{\mathbb{R}^m} \partial_\alpha \phi_\delta(u_0 - v)\eta(dv) = (\partial_\alpha \phi_\delta * \eta)(u_0),$$

which shows that $\partial_\alpha \phi_\delta * \eta : \mathbb{R}^m \to \mathbb{R}$ is continuous. Moreover, for every fixed $\beta \in \mathbb{N}_0^m$ and $l = 1, ..., m$ (with $e_l \in \mathbb{R}^m$ denoting the l-th unit vector of $\mathbb{R}^m$), we use the mean-value theorem to conclude for every $u, v \in \mathbb{R}^m$ and $h \in \mathbb{R}$ that

$$\max\left(\left|\frac{\partial_\beta \phi_\delta(u + he_l - v) - \partial_\beta \phi_\delta(u - v)}{h}\right|, |\partial_{\beta + e_l}\phi_\delta(u - v)|\right)$$
$$\leqslant C_{12} := \sup_{u_1 \in \mathbb{R}^m} |\partial_{\beta + e_l}\phi_\delta(u_1)| < \infty. \tag{22}$$

Then, the dominated convergence theorem (with (22) and that $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ is finite) implies that

$$
\begin{aligned}
\partial_{e_l}(\partial_\alpha \phi_\delta * \eta)(u) &= \lim_{h \to 0} \frac{(\partial_\beta \phi_\delta * \eta)(u + he_l) - (\partial_\beta \phi_\delta * \eta)(u)}{h} \\
&= \lim_{h \to 0} \int_{\mathbb{R}^m} \frac{\partial_\beta \phi_\delta(u + he_l - v) - \partial_\beta \phi_\delta(u - v)}{h} \eta(dv) \\
&= \int_{\mathbb{R}^m} \partial_{\beta + e_l} \phi_\delta(u - v) \eta(dv) = (\partial_{\beta + e_l} \phi_\delta * \eta)(u).
\end{aligned}
$$

Hence, by induction on $\beta \in \mathbb{N}_0^m$, it follows that $\partial_\alpha(\phi_\delta * \eta)(u) = (\partial_\alpha \phi_\delta * \eta)(u)$ for any $u \in \mathbb{R}^m$. This together with the previous step shows (i).

For (ii), we use (i), that $\operatorname{supp}(\phi_\delta) = B_\delta(0)$ implies $\operatorname{supp}(\partial_\alpha \phi_\delta) \subseteq B_\delta(0)$, the inequality $1 + x + y \leqslant (1 + x)(1 + y)$ for any $x, y \geqslant 0$, that the constant $C_{13} := \sup_{y \in \mathbb{R}^m} |\partial_\alpha \phi_\delta(y)| > 0$ is finite, and that $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ to conclude that

$$
\begin{aligned}
C_{14} := \sup_{u \in \mathbb{R}^m} \left((1 + \|u\|)^\gamma \, |(\phi_\delta * \eta)(u)| \right) &\leqslant \sup_{u \in \mathbb{R}^m} \int_{\mathbb{R}^m} (1 + \|u\|)^\gamma \, |\partial_\alpha \phi_\delta(u - v)| \, |\eta|(dv) du \\
&\leqslant \sup_{u \in \mathbb{R}^m} \int_{\mathbb{R}^m} (1 + \underbrace{\|u - v\|}_{\leqslant \delta} + \|v\|)^\gamma \, |\partial_\alpha \phi_\delta(u - v)| \, |\eta|(dv) \leqslant C_{13}(1 + \delta)^\gamma \int_{\mathbb{R}^m} (1 + \|v\|)^\gamma |\eta|(dv) < \infty.
\end{aligned}
$$

Hence, by using this and that $f \in \overline{C_b^0(\mathbb{R}^m)}^\gamma$ together with Lemma 4.1, it follows that

$$
\begin{aligned}
\lim_{r \to \infty} \sup_{u \in \mathbb{R}^m \setminus B_r(0)} |f(u) \partial_\alpha(\phi_\delta * \eta)(u)| &= \lim_{r \to \infty} \sup_{u \in \mathbb{R}^m \setminus B_r(0)} \left(\frac{|f(u)|}{(1 + \|u\|)^\gamma} (1 + \|u\|)^\gamma \, |\partial_\alpha(\phi_\delta * \eta)(u)| \right) \\
&= C_{14} \lim_{r \to \infty} \sup_{u \in \mathbb{R}^m \setminus B_r(0)} \frac{|f(u)|}{(1 + \|u\|)^\gamma} = 0,
\end{aligned}
$$

which shows (ii).

For (iii), we first prove that $\partial_\alpha(\phi_\delta * \eta)(u) du \big|_{\mathcal{B}(\mathbb{R}^m)} : \mathcal{B}(\mathbb{R}^m) \to \mathbb{R}$ is a signed Radon measure. For this purpose, we denote its positive and negative part by $\eta_{\delta,\pm} := \pm (\partial_\alpha(\phi_\delta * \eta)(u))_\pm \, du \big|_{\mathcal{B}(\mathbb{R}^m)} : \mathcal{B}(\mathbb{R}^m) \to [0, \infty]$ satisfying $\eta_{\delta,+} - \eta_{\delta,-} = \partial_\alpha(\phi_\delta * \eta)(u) du \big|_{\mathcal{B}(\mathbb{R}^m)}$, where $s_+ := \max(s, 0)$ and $s_- := -\min(s, 0)$, for any $s \in \mathbb{R}$. Moreover, we define the finite constant $C_{15} := \sup_{u \in \mathbb{R}^m} |\partial_\alpha \phi_\delta(u)| > 0$. Then, for every $u \in \mathbb{R}^m$, we choose a compact subset $K \subset \mathbb{R}^m$ with $u \in K$ and use that $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$ is finite to conclude that

$$
\begin{aligned}
\eta_{\delta,\pm}(K) = \pm \int_K (\partial_\alpha(\phi_\delta * \eta)(u))_\pm \, du &\leqslant \left(\underbrace{\int_K du}_{=:|K|} \right) \sup_{u \in K} |(\partial_\alpha \phi_\delta * \eta)(u)| \\
&\leqslant |K| \sup_{u \in K} \int_{\mathbb{R}^m} |\partial_\alpha \phi_\delta(u - v)| \, |\eta|(dv) \leqslant C_{15}|K| |\eta|(\mathbb{R}^m) < \infty.
\end{aligned}
$$

This shows that both measures $\eta_{\delta,\pm} : \mathcal{B}(\mathbb{R}^m) \to [0, \infty]$ are locally finite. In addition, it holds for every $B \in \mathcal{B}(\mathbb{R}^m)$ that

$$
\begin{aligned}
\eta_{\delta,\pm}(B) &= \pm \int_B (\partial_\alpha(\phi_\delta * \eta)(u))_\pm \, du \\
&= \inf \left\{ \pm \int_U (\partial_\alpha(\phi_\delta * \eta)(u))_\pm \, du : U \subseteq \mathbb{R}^m \text{ open with } B \subseteq U \right\} \\
&= \inf \left\{ \eta_{\delta,\pm}(U) : U \subseteq \mathbb{R}^m \text{ open with } B \subseteq U \right\},
\end{aligned}
$$

which shows that both measures $\eta_{\delta,\pm} : \mathcal{B}(\mathbb{R}^m) \to [0,\infty]$ are outer regular. Moreover, it holds for every $B \in \mathcal{B}(\mathbb{R}^m)$ that

$$
\begin{aligned}
\eta_{\delta,\pm}(B) &= \pm \int_B \left(\partial_\alpha(\phi_\delta * \eta)(u)\right)_\pm du \\
&= \sup\left\{ \pm \int_K \left(\partial_\alpha(\phi_\delta * \eta)(u)\right)_\pm du : K \subset B \text{ relatively compact} \right\} \\
&= \sup\left\{ \eta_{\delta,\pm}(K) : K \subset B \text{ relatively compact} \right\},
\end{aligned}
$$

which shows that both measures $\eta_{\delta,\pm} : \mathcal{B}(\mathbb{R}^m) \to [0,\infty]$ are inner regular. Hence, both measures $\eta_{\delta,\pm} : \mathcal{B}(\mathbb{R}^m) \to [0,\infty]$ are Radon measures and $\partial_\alpha(\phi_\delta * \eta)(u)du\big|_{\mathcal{B}(\mathbb{R}^m)} = \eta_{\delta,+} - \eta_{\delta,-} : \mathcal{B}(\mathbb{R}^m) \to [0,\infty]$ is thus a signed Radon measure. Furthermore, by using the triangle inequality, that $\operatorname{supp}(\phi_\delta) = B_\delta(0)$ implies $\operatorname{supp}(\partial_\alpha \phi_\delta) \subseteq B_\delta(0)$, the inequality $1 + x + y \leqslant (1+x)(1+y)$ for any $x, y \geqslant 0$, the substitution $y \mapsto u - v$ together with $\|\partial_\alpha \phi_\delta\|_{L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du)} < \infty$, and that $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$, we have

$$
\begin{aligned}
\int_{\mathbb{R}^m} (1 + \|u\|)^\gamma \left|\partial_\alpha(\phi_\delta * \eta)(u)\right| du &\leqslant \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} (1 + \|u\|)^\gamma |\partial_\alpha \phi_\delta(u - v)| du |\eta|(dv) \\
&\leqslant \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} (1 + \underbrace{\|u - v\|}_{\leqslant \delta} + \|v\|)^\gamma |\partial_\alpha \phi_\delta(u - v)| du |\eta|(dv) \\
&\leqslant (1 + \delta)^\gamma \left(\sup_{v \in \mathbb{R}^m} \int_{\mathbb{R}^m} |\partial_\alpha \phi_\delta(u - v)| du \right) \left(\int_{\mathbb{R}^m} (1 + \|v\|)^\gamma |\eta|(dv) \right) \\
&\leqslant (1 + \delta)^\gamma \|\partial_\alpha \phi_\delta\|_{L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du)} \left(\int_{\mathbb{R}^m} (1 + \|v\|)^\gamma |\eta|(dv) \right) < \infty.
\end{aligned}
$$

This shows that $\partial_\alpha(\phi_\delta * \eta)(u)du\big|_{\mathcal{B}(\mathbb{R}^m)} \in \mathcal{M}_\gamma(\mathbb{R}^m)$ is a finite signed Radon measure.

For (iv), we use (iii) to conclude that the constant $C_{16} := \int_{\mathbb{R}^m} (1 + \|u\|)^\gamma |(\phi_\delta * \eta)(u)| du > 0$ is finite. Then, it follows for every $f \in \overline{C_b^0(\mathbb{R}^m)}^\gamma$ that

$$
\begin{aligned}
\left| \int_{\mathbb{R}^m} f(u) \partial_\alpha(\phi_\delta * \eta)(u) du \right| &\leqslant \left(\sup_{u \in \mathbb{R}^m} \frac{|f(u)|}{(1 + \|u\|)^\gamma} \right) \int_{\mathbb{R}^m} (1 + \|u\|)^\gamma |\partial_\alpha(\phi_\delta * \eta)(u)| du \\
&= C_{16} \|f\|_{C_{pol,\gamma}^0(\mathbb{R}^m)},
\end{aligned}
$$

which shows that $\overline{C_b^0(\mathbb{R}^m)}^\gamma \ni f \mapsto \int_{\mathbb{R}^m} f(u) \partial_\alpha(\phi_\delta * \eta)(u) du \in \mathbb{R}$ is a continuous linear functional.

For (v), we use the substitution $u \mapsto v + y$ to conclude that

$$
\begin{aligned}
\int_{\mathbb{R}^m} f(u)(\phi_\delta * \eta)(u) du &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} f(u) \phi_\delta(u - v) \eta(dv) du \\
&= \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} f(v + y) \eta(dv) \phi_\delta(y) dy.
\end{aligned}
$$

For (vi), we define for every $\delta \in (0,1)$ the function $\mathbb{R}^m \ni u \mapsto (\phi_\delta * f)(u) := \int_{\mathbb{R}^m} \phi_\delta(u - v) f(v) dv \in \mathbb{R}$. Then, by using the triangle inequality, that $\operatorname{supp}(\phi_\delta) = B_\delta(0)$, the substitution $y \mapsto u - v$ together with $\int_{\mathbb{R}^m} |\phi_\delta(y)| dy = \|\phi_\delta\|_{L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du)} = \|\phi\|_{L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du)} = 1$, the inequality $1 + x + y \leqslant (1+x)(1+y)$ for any $x, y \geqslant 0$, and that $f \in \overline{C_b^0(\mathbb{R}^m)}^\gamma$, it follows for every $u \in \mathbb{R}^m$ that

$$
\begin{aligned}
|(\phi_\delta * f)(u)| &\leqslant \int_{\mathbb{R}^m} |\phi_\delta(u - v)| \frac{|f(v)|}{(1 + \|v\|)^\gamma} (1 + \|v\|)^\gamma dv \\
&\leqslant \int_{\mathbb{R}^m} |\phi_\delta(u - v)| \frac{|f(v)|}{(1 + \|v\|)^\gamma} (1 + \|u\| + \underbrace{\|u - v\|}_{\leqslant \delta})^\gamma dv \\
&\leqslant \left(\int_{\mathbb{R}^m} |\phi_\delta(u - v)| dv \right) \left(\sup_{v \in \mathbb{R}^m} \frac{|f(v)|}{(1 + \|v\|)^\gamma} \right) (1 + \|u\| + \delta)^\gamma \\
&\leqslant \left(\int_{\mathbb{R}^m} |\phi_\delta(y)| dy \right) \|f\|_{C_{pol,\gamma}^0(\mathbb{R}^m)} (1 + \delta)^\gamma (1 + \|u\|)^\gamma \\
&\leqslant 2^\gamma \|f\|_{C_{pol,\gamma}^0(\mathbb{R}^m)} (1 + \|u\|)^\gamma.
\end{aligned}
$$

$$(23)$$

Moreover, by using that $f \in \overline{C_b^0(\mathbb{R}^m)}^\gamma$, we conclude for every $u \in \mathbb{R}^m$ that

$$|f(u)| \leqslant \left(\sup_{u \in \mathbb{R}^m} \frac{|f(u)|}{(1+\|u\|)^\gamma} \right) (1+\|u\|)^\gamma \leqslant \|f\|_{C_{pol,\gamma}^0(\mathbb{R}^m)}(1+\|u\|)^\gamma. \tag{24}$$

Hence, by using (v), Fubini's theorem, the substitution $u \mapsto v + y$, and the dominated convergence theorem (with (23), (24), $(1+\|u\|)^\gamma \in L^1(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), |\eta|)$ as $\eta \in \mathcal{M}_\gamma(\mathbb{R}^m)$, and [18, Theorem C.7], i.e. that $\phi_\delta * f : \mathbb{R}^m \to \mathbb{R}$ converges a.e. to $f : \mathbb{R}^m \to \mathbb{R}$, as $\delta \to 0$), it follows that

$$\begin{aligned}
\lim_{\delta \to 0} \int_{\mathbb{R}^m} f(u)(\phi_\delta * \eta)(u) du &= \lim_{\delta \to 0} \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} f(v+y)\eta(dv)\phi_\delta(y)dy \\
&= \lim_{\delta \to 0} \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^m} f(v+y)\phi_\delta(y)dy \right) \eta(dv) \\
&= \lim_{\delta \to 0} \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^m} \phi(v-u)f(u)du \right) \eta(dv) \\
&= \lim_{\delta \to 0} \int_{\mathbb{R}^m} (\phi_\delta * f)(v)\eta(dv) \\
&= \int_{\mathbb{R}^m} f(v)\eta(dv),
\end{aligned}$$

which completes the proof. □

Finally, we provide the proof of Theorem 2.7, i.e. the universal approximation property of neural networks $\mathcal{NN}_{U,d}^\rho$ within any function space $(X, \|\cdot\|_X)$ satisfying Assumption 2.3 with $k \in \mathbb{N}_0$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), and $\gamma \in (0, \infty)$, where $\rho \in \overline{C_b^k(\mathbb{R})}^\gamma$ is the activation function.

Proof of Theorem 2.7. First, we apply Lemma 2.5 to conclude that $\mathcal{NN}_{\mathbb{R}^m,d}^\rho \subseteq \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma$ and that $\mathcal{NN}_{U,d}^\rho \subseteq X$. Now, we assume by contradiction that $\mathcal{NN}_{U,d}^\rho$ is not dense in X. Then, by using that $(X, \|\cdot\|_X)$ satisfies Assumption 2.3, i.e. that the restriction map (4) is a continuous dense embedding, it follows from Remark 2.4 that $\mathcal{NN}_{\mathbb{R}^m,d}^\rho$ cannot be dense in $\overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma$. Hence, by applying the Hahn-Banach theorem, there exists a non-zero continuous linear functional $l : \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma \to \mathbb{R}$ such that for every $\varphi \in \mathcal{NN}_{\mathbb{R}^m,d}^\rho$ it holds that $l(\varphi) = 0$.

Next, we use the Riesz representation result in Proposition 4.2 to conclude that there exist some signed Radon measures $(\eta_{\alpha,i})_{\alpha \in \mathbb{N}_{0,k}^m, i=1,...,d} \in \mathcal{M}_\gamma(\mathbb{R}^m)$ such that for every $f \in \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma$ it holds that

$$l(f) = \sum_{\alpha \in \mathbb{N}_{0,k}^m} \sum_{i=1}^d \int_{\mathbb{R}^m} \partial_\alpha f_i(u)\eta_{\alpha,i}(du).$$

Since $l(\varphi) = 0$ for any $\varphi \in \mathcal{NN}_{\mathbb{R}^m,d}^\rho$, it follows for every $a \in \mathbb{R}^m$, $b \in \mathbb{R}$, and $i = 1, ..., d$ that

$$l\left(e_i\rho\left(\lambda a^\top \cdot -b\right)\right) = \sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_{\mathbb{R}^m} \rho^{(|\alpha|)}\left(a^\top u - b\right) a^\alpha \eta_{\alpha,i}(du) = 0, \tag{25}$$

where $e_i\rho(\lambda a^\top \cdot -b)$ denotes the function $\mathbb{R}^m \ni u \mapsto e_i\rho(\lambda a^\top u - b) \in \mathbb{R}^d$ with $e_i \in \mathbb{R}^d$ being the i-th unit vector of $\mathbb{R}^d$, and where $a^\alpha := \prod_{l=1}^m a_l^{\alpha_l}$ for $a := (a_1, ..., a_m) \in \mathbb{R}^m$ and $\alpha := (\alpha_1, ..., \alpha_m) \in \mathbb{N}_{0,k}^m$.

Now, we define for every fixed $\delta > 0$ the linear map $l_\delta : \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma \to \mathbb{R}$ by

$$\overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma \ni f \quad \mapsto \quad l_\delta(f) := \sum_{\alpha \in \mathbb{N}_{0,k}^m} \sum_{i=1}^d \int_{\mathbb{R}^m} \partial_\alpha f_i(u)(\phi_\delta * \eta)(u)du \in \mathbb{R}.$$

Then, Lemma 4.4 (iv) shows that $l_\delta : \overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma \to \mathbb{R}$ is a continuous linear functional as it is a finite sum of the continuous linear functionals $\overline{C_b^k(\mathbb{R}^m;\mathbb{R}^d)}^\gamma \ni f \mapsto \int_{\mathbb{R}^m} \partial_\alpha f_i(u)(\phi_\delta * \eta)(u)du \in \mathbb{R}$ taken over $\alpha \in \mathbb{N}_{0,k}^m$ and $i = 1, ..., d$. Moreover, for every fixed $i = 1, ..., d$, we define

$$\mathbb{R}^m \ni u \mapsto h_{\delta,i}(u) := \sum_{\alpha \in \mathbb{N}_{0,k}^m} (-1)^{|\alpha|} \partial_\alpha\left(\phi_\delta * \eta_{\alpha,i}\right)(u) \in \mathbb{R},$$

which satisfies $h_{\delta,i}(u)du \in \mathcal{M}_\gamma(\mathbb{R}^m)$ as it is a finite linear combination of finite signed Radon measures $\partial_\alpha(\phi_\delta * \eta_{\alpha,i})(u)du \in \mathcal{M}_\gamma(\mathbb{R}^m)$ taken over $\alpha \in \mathbb{N}_{0,k}^m$ (see Lemma 4.4 (iii)). Hence, integration by parts together with Lemma 4.4 (ii) shows that

$$
\begin{aligned}
l_\delta(f) &= \sum_{\alpha \in \mathbb{N}_{0,k}^m} \sum_{i=1}^d \int_{\mathbb{R}^m} \partial_\alpha f_i(u)\,(\phi_\delta * \eta_{\alpha,i})(u)du \\
&= \sum_{\alpha \in \mathbb{N}_{0,k}^m} \sum_{i=1}^d (-1)^{|\alpha|} \int_{\mathbb{R}^m} f_i(u)\partial_\alpha(\phi_\delta * \eta_{\alpha,i})(u)du \\
&= \sum_{i=1}^d \int_{\mathbb{R}^m} f_i(u)h_{\delta,i}(u)du.
\end{aligned}
$$

Thus, by using this, Lemma 4.4 (v), and (25) (with $b - a^\top y \in \mathbb{R}$ instead of $b \in \mathbb{R}$), it follows for every $a \in \mathbb{R}^m$, $b \in \mathbb{R}$, and $i = 1, ..., d$ that

$$
\begin{aligned}
\int_{\mathbb{R}^m} \rho\left(a^\top u - b\right) h_{\delta,i}(u)du &= \sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_{\mathbb{R}^m} \rho^{(|\alpha|)}\left(a^\top u - b\right) a^\alpha (\phi_\delta * \eta_{\alpha,i})(u)du \\
&= \sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_{\mathbb{R}^m} \rho^{(|\alpha|)}\left(a^\top(u+y) - b\right) a^\alpha \eta_{\alpha,i}(du)\phi_\delta(y)du \\
&= \int_{\mathbb{R}^m} \underbrace{l\left(e_i\rho\left(a^\top \cdot - \left(b - a^\top y\right)\right)\right)}_{=0} \phi_\delta(y)dy = 0.
\end{aligned}
$$

Now, for every $i = 1, ..., d$, we apply Proposition 4.3 with $h_{\delta,i}(u)du \in \mathcal{M}_\gamma(\mathbb{R}^m)$ to conclude that $h_{\delta,i}(u)du = 0 \in \mathcal{M}_\gamma(\mathbb{R}^m)$, and thus $h_{\delta,i}(u) = 0$ for a.e. $u \in \mathbb{R}^m$. Hence, it follows for every $f \in \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma$ that

$$
l_\delta(f) = \sum_{i=1}^d \int_{\mathbb{R}^m} f_i(u)h_{\delta,i}(u)du = 0,
$$

which shows that $l_\delta : C_b^k(\mathbb{R}^m; \mathbb{R}^d) \to \mathbb{R}$ vanishes everywhere on $C_b^k(\mathbb{R}^m; \mathbb{R}^d)$.

Finally, we use Lemma 4.4 (vi) to conclude for every $f \in \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma$ that

$$
l(f) = \sum_{\alpha \in \mathbb{N}_{0,k}^m} \sum_{i=1}^d \int_{\mathbb{R}^m} \partial_\alpha f_i(u)\eta_{\alpha,i}(du) = \lim_{\delta\to\infty} \sum_{\alpha \in \mathbb{N}_{0,k}^m} \sum_{i=1}^d \int_{\mathbb{R}^m} f_i(u)(\phi_\delta * \eta)(u)du = \lim_{\delta\to\infty} l_\delta(f) = 0,
$$

which shows that $l : \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma \to \mathbb{R}$ vanishes everywhere. This however contradicts the assumption that $l : \overline{C_b^k(\mathbb{R}^m; \mathbb{R}^d)}^\gamma \to \mathbb{R}$ is non-zero. Hence, $\mathcal{NN}_{U,d}^\rho$ is dense in X. $\qquad\square$

4.2.3. *Proof of Example 2.6.* For the proof of Example 2.6 (e), we first generalize the approximation result for compactly supported smooth functions in unweighted Sobolev spaces (see [3, Theorem 3.18]) to weighted Sobolev spaces $W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ introduced in Footnote 5.

Proposition 4.5 (Approximation in Weighted Sobolev Spaces). *For $k \in \mathbb{N}$, $p \in [1, \infty)$, and $U \subseteq \mathbb{R}^m$ open and having the segment property, let $w : U \to [0, \infty)$ be a bounded weight such that for every bounded subset $B \subseteq U$ it holds that $\inf_{u \in B} w(u) > 0$. Then, $\{f|_U : U \to \mathbb{R}^d : f \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^d)\}$ is dense in $W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$.*

Proof. First, we follow [3, Theorem 3.18] to show that every fixed function $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ can be approximated by elements from the set $\{f|_U : U \to \mathbb{R}^d : f \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^d)\}$ with respect to $\|\cdot\|_{W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)}$. To this end, we choose some $h \in C_c^\infty(\mathbb{R}^m)$ which satisfies $h(u) = 1$ for all $u \in \overline{B_1(0)}$, $h(u) = 0$ for all $u \in \mathbb{R}^m \backslash B_2(0)$, and for which there exists a constant $C_h > 0$ such that for every $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in \mathbb{R}^m$ it holds that $|\partial_\alpha h(u)| \leqslant C_h$. In addition, we define for every fixed $r > 1$ the functions $\mathbb{R}^m \ni u \mapsto h_r(u) := h(u/r) \in \mathbb{R}$ and $U \ni u \mapsto f_r(u) := f(u)h_r(u) \in \mathbb{R}^d$, which both have bounded support. Then, by using the Leibniz product rule together with the triangle inequality, that

$|\partial_\alpha h_r(u)| = |\partial_\alpha h(u/r)| \, r^{-|\alpha|} \leqslant C_h$ for any $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in \mathbb{R}^m$, and the inequality (12), it follows for every $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in U$ that

$$
\begin{aligned}
\|\partial_\alpha f_r(u)\|^p &\leqslant \left(\sum_{\substack{\beta_1,\beta_2 \in \mathbb{N}_0^m \\ \beta_1+\beta_2=\alpha}} \frac{\alpha!}{\beta_1!\beta_2!} |\partial_{\beta_1} h_r(u)| \|\partial_{\beta_2} f(u)\| \right)^p \\
&\leqslant 2^{kp} C_h^p \max_{\beta_2 \in \mathbb{N}_{0,k}^m} \|\partial_{\beta_2} f(u)\|^p \\
&\leqslant 2^{kp} C_h^p \sum_{\beta_2 \in \mathbb{N}_{0,k}^m} \|\partial_{\beta_2} f(u)\|^p .
\end{aligned}
$$

Hence, by using this, it follows for every $V \in \mathcal{L}(U)$ that

$$
\begin{aligned}
\|f_r\|_{W^{k,p}(V,\mathcal{L}(V),w;\mathbb{R}^d)} &= \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_V \|\partial_\alpha f_r(u)\|^p w(u) du \right)^{\frac{1}{p}} \\
&\leqslant \left| \mathbb{N}_{0,k}^m \right|^{\frac{1}{p}} \left(\max_{\alpha \in \mathbb{N}_{0,k}^m} \int_V \|\partial_\alpha f_r(u)\|^p w(u) du \right)^{\frac{1}{p}} \\
&\leqslant 2^k C_h \left| \mathbb{N}_{0,k}^m \right|^{\frac{1}{p}} \left(\sum_{\beta_2 \in \mathbb{N}_{0,k}^m} \int_V \|\partial_{\beta_2} f(u)\|^p w(u) du \right)^{\frac{1}{p}} \\
&\leqslant 2^k C_h \left| \mathbb{N}_{0,k}^m \right|^{\frac{1}{p}} \|f\|_{W^{k,p}(V,\mathcal{L}(V),w;\mathbb{R}^d)} < \infty.
\end{aligned}
\tag{26}
$$

Thus, by taking $V := U$ in (26), we conclude that $f_r \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$. Similarly, by using the triangle inequality, that $\partial_\alpha f_r(u) = \partial_\alpha(f(u) h_r(u)) = \partial_\alpha f(u)$ for any $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in U \cap \overline{B_r(0)}$ (as $h_r(u) = 1$ for any $u \in \overline{B_r(0)}$), and (26) with $V := U \backslash \overline{B_r(0)}$, it follows that

$$
\begin{aligned}
\|f - f_r\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)} &\leqslant \underbrace{\|f - f_r\|_{W^{k,p}(U \cap \overline{B_r(0)}, \mathcal{L}(U \cap \overline{B_r(0)}), w;\mathbb{R}^d)}}_{=0} + \|f - f_r\|_{W^{k,p}(U \backslash \overline{B_r(0)}, \mathcal{L}(U \backslash \overline{B_r(0)}), w;\mathbb{R}^d)} \\
&\leqslant \|f\|_{W^{k,p}(U \backslash \overline{B_r(0)}, \mathcal{L}(U \backslash \overline{B_r(0)}), w;\mathbb{R}^d)} + \|f_r\|_{W^{k,p}(U \backslash \overline{B_r(0)}, \mathcal{L}(U \backslash \overline{B_r(0)}), w;\mathbb{R}^d)} \\
&\leqslant \left(1 + 2^k C_h \left| \mathbb{N}_{0,k}^m \right|^{\frac{1}{p}} \right) \|f\|_{W^{k,p}(U \backslash \overline{B_r(0)}, \mathcal{L}(U \backslash \overline{B_r(0)}), w;\mathbb{R}^d)}.
\end{aligned}
$$

Since the right-hand side tends to zero, as $r \to \infty$, this shows that $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ can be approximated by elements of $\{ f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d) : \operatorname{supp}(f) \subseteq U \text{ is bounded} \}$ with respect to $\| \cdot \|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}$. Hence, we only need to show the approximation of the latter by elements from $\{ f|_U : U \to \mathbb{R}^d : f \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^d) \}$ with respect to $\| \cdot \|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}$.

Therefore, we now fix some $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ with bounded support $\operatorname{supp}(f) \subseteq U$ and some $\varepsilon > 0$. Moreover, by recalling that $w : U \to [0, \infty)$ is bounded, we can define the finite constant $C_w := \sup_{u \in U} w(u) > 0$. Then, by using that $f(u) = 0$ for any $u \in U \backslash \operatorname{supp}(f)$, thus $\partial_\alpha f(u) = 0$ for any $\alpha \in \mathbb{N}_{0,k}^m$ and $u \in U \backslash \operatorname{supp}(f)$, and the assumption that $C_{f,w} := \inf_{u \in \operatorname{supp}(f)} w(u) > 0$, we have

$$
\begin{aligned}
\|f\|_{W^{k,p}(U,\mathcal{L}(U),du;\mathbb{R}^d)} &= \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u)\|^p du \right)^{\frac{1}{p}} = \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_{\operatorname{supp}(f)} \|\partial_\alpha f(u)\|^p du \right)^{\frac{1}{p}} \\
&\leqslant C_{f,w}^{-1} \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_{\operatorname{supp}(f)} \|\partial_\alpha f(u)\|^p w(u) du \right)^{\frac{1}{p}} = C_{f,w}^{-1} \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u)\|^p w(u) du \right)^{\frac{1}{p}} \\
&= C_{f,w}^{-1} \|f\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)} < \infty.
\end{aligned}
$$

This shows that $f \in W^{k,p}(U, \mathcal{L}(U), du; \mathbb{R}^d)$. Hence, by applying [3, Theorem 3.18] (with $U \subseteq \mathbb{R}^m$ having the segment property) componentwise, there exists some $g \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^d)$ such that

$$\|f - g\|_{W^{k,p}(U, \mathcal{L}(U), du, \mathbb{R}^d)} = \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u) - \partial_\alpha g(u)\|^p du \right)^{\frac{1}{p}} < \frac{\varepsilon}{C_w}.$$

Thus, by using that $w : U \to [0, \infty)$ is bounded with $C_w := \sup_{u \in U} w(u) < \infty$, it follows that

$$\|f - g\|_{W^{k,p}(U, \mathcal{L}(U), du, \mathbb{R}^d)} = \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u) - \partial_\alpha g(u)\|^p w(u) du \right)^{\frac{1}{p}}$$

$$\leqslant C_w \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u) - \partial_\alpha g(u)\|^p du \right)^{\frac{1}{p}}$$

$$< C_w \frac{\varepsilon}{C_w} = \varepsilon.$$

Since $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ with bounded support $\mathrm{supp}(f) \subseteq U$ and $\varepsilon > 0$ were chosen arbitrarily, it follows together with the first step that $\{f|_U : U \to \mathbb{R}^d : f \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^d)\}$ is dense in $W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$. $\square$

Proof of Example 2.6. For (a), we use that $U \subset \mathbb{R}^m$ is bounded to define the finite constant $C_{21} := \sup_{u \in U} (1 + \|u\|)^\gamma$. Then, it follows for every $f \in C_b^0(\mathbb{R}^m; \mathbb{R}^d)$ that

$$\|f|_U\|_{C_b^k(U;\mathbb{R}^d)} = \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in U} \|\partial_\alpha f(u)\|$$

$$\leqslant \left(\sup_{u \in U} (1 + \|u\|)^\gamma \right) \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in U} \frac{\|\partial_\alpha f(u)\|}{(1 + \|u\|)^\gamma}$$

$$\leqslant C_{21} \|f\|_{C_{pol,\gamma}^k(\mathbb{R}^m;\mathbb{R}^d)}.$$

Moreover, by using that $\{f|_U : f \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)\} = C_b^k(U; \mathbb{R}^d)$, the image $\{f|_U : f \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)\}$ of the continuous embedding (4) is dense in $C_b^k(U; \mathbb{R}^d)$.

For (b), the restriction map (4) is by definition continuous. Moreover, by using that $\overline{C_b^k(U; \mathbb{R}^d)}^\gamma$ is defined as the closure of $C_b^k(U; \mathbb{R}^d)$ with respect to $\|\cdot\|_{C_{pol,\gamma}^k(U;\mathbb{R}^d)}$, the image $\{g|_U : g \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)\} = C_b^k(U; \mathbb{R}^d)$ of the continuous embedding (4) is dense in $\overline{C_b^k(U; \mathbb{R}^d)}^\gamma$.

For (c), we first recall that $k = 0$. Then, we use that $f \in C_b^0(\mathbb{R}^m; \mathbb{R}^d)$ is continuous to conclude that its restriction $f|_U : U \to \mathbb{R}^d$ is $\mathcal{B}(U)/\mathcal{B}(\mathbb{R}^d)$-measurable. Moreover, we define the finite constant $C_{22} := \int_U (1 + \|u\|)^{\gamma p} \mu(du) > 0$, which implies that $\mu : \mathcal{B}(U) \to [0, \infty)$ is finite as $\mu(U) = \int_U \mu(du) \leqslant C_{22} < \infty$. Then, it follows for every $f \in C_b^0(\mathbb{R}^m; \mathbb{R}^d)$ that

$$\|f|_U\|_{L^p(U, \mathcal{B}(U), \mu; \mathbb{R}^d)} = \left(\int_U \|f(u)\|^p \mu(du) \right)^{\frac{1}{p}}$$

$$\leqslant \left(\int_U (1 + \|u\|)^{\gamma p} \mu(du) \right)^{\frac{1}{p}} \sup_{u \in U} \frac{\|f(u)\|}{(1 + \|u\|)^\gamma}$$

$$\leqslant C_{22}^{\frac{1}{p}} \|f\|_{C_{pol,\gamma}^0(\mathbb{R}^m;\mathbb{R}^d)},$$

which shows that the restriction map (4) is continuous. In order to show that its image is dense, we fix some $f \in L^p(U, \mathcal{B}(U), \mu; \mathbb{R}^d)$ and $\varepsilon > 0$. Then, we extend $f : U \to \mathbb{R}^d$ to the function

$$\mathbb{R}^m \ni u \quad \mapsto \quad \overline{f}(u) := \begin{cases} f(u), & u \in U, \\ 0, & u \in \mathbb{R}^m \backslash U. \end{cases}$$

Moreover, we extend $\mu : \mathcal{B}(U) \to [0, \infty)$ to the Borel measure $\mathcal{B}(\mathbb{R}^m) \ni E \mapsto \overline{\mu}(E) := \mu(U \cap E) \in [0, \infty)$, which implies that $\overline{f} \in L^p(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), \overline{\mu}; \mathbb{R}^d)$. Hence, by applying [7, Corollary 2.2.2]

componentwise (with $\overline{\mu}(B) = \mu(U \cap B) \leqslant \mu(U) \leqslant C_{22} < \infty$ for any bounded $B \in \mathcal{B}(\mathbb{R}^m)$), there exists some $g \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^d) \subseteq C_b^0(\mathbb{R}^m; \mathbb{R}^d)$ with $\|\overline{f} - g\|_{L^p(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), \overline{\mu}; \mathbb{R}^d)} < \varepsilon$, which implies

$$\|f - g|_U\|_{L^p(U, \mathcal{B}(U), \mu; \mathbb{R}^d)} = \|\overline{f} - g\|_{L^p(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), \overline{\mu}; \mathbb{R}^d)} < \varepsilon.$$

Since $f \in C^0(U; \mathbb{R}^d)$ and $\varepsilon > 0$ were chosen arbitrarily, it follows that the image $\{f|_U : f \in C_b^0(\mathbb{R}^m; \mathbb{R}^d)\}$ of the continuous embedding (4) is dense in $L^p(U, \mathcal{B}(U), \mu; \mathbb{R}^d)$.

For (d), we first use that $f \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)$ is k-times differentiable to conclude for every $\alpha \in \mathbb{N}_{0,k}^m$ that $\partial_\alpha f|_U : U \to \mathbb{R}^d$ is $\mathcal{L}(U)/\mathcal{B}(\mathbb{R}^d)$-measurable. Moreover, we use that $U \subset \mathbb{R}^m$ is bounded to define the finite constant $C_{23} := \int_U (1 + \|u\|)^{\gamma p} du > 0$. Then, it follows for every $f \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)$ that

$$\begin{aligned}
\|f\|_{W^{k,p}(U, \mathcal{L}(U), du; \mathbb{R}^d)} &= \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u)\|^p du \right)^{\frac{1}{p}} \\
&\leqslant \left(|\mathbb{N}_{0,k}^m| \int_U (1 + \|u\|)^{\gamma p} du \right)^{\frac{1}{p}} \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in U} \frac{\|\partial_\alpha f(u)\|}{(1 + \|u\|)^{\gamma p}} \\
&\leqslant \left(C_{23} |\mathbb{N}_{0,k}^m| \right)^{\frac{1}{p}} \|f\|_{C_{pol,\gamma}^k(\mathbb{R}^m; \mathbb{R}^d)},
\end{aligned}$$

which shows that the restriction map (4) is continuous. In addition, by applying [3, Theorem 3.18] componentwise, $\{g|_U : g \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^d)\}$ is dense in $W^{k,p}(U, \mathcal{L}(U), du; \mathbb{R}^d)$. Hence, by using that $C_c^\infty(\mathbb{R}^m; \mathbb{R}^d) \subseteq C_b^k(\mathbb{R}^m; \mathbb{R}^d)$, it follows that the image $\{g|_U : g \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)\}$ of the continuous embedding (4) is dense in $W^{k,p}(U, \mathcal{L}(U), du; \mathbb{R}^d)$.

For (e), we use that $f \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)$ is k-times differentiable to conclude for every $\alpha \in \mathbb{N}_{0,k}^m$ that $\partial_\alpha f|_U : U \to \mathbb{R}^d$ is $\mathcal{L}(U)/\mathcal{B}(\mathbb{R}^d)$-measurable. Moreover, by using the finite constant $C_{24} := \int_U (1 + \|u\|)^{\gamma p} w(u) du > 0$, it follows for every $f \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)$ that

$$\begin{aligned}
\|f\|_{W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)} &= \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f(u)\|^p w(u) du \right)^{\frac{1}{p}} \\
&\leqslant \left(|\mathbb{N}_{0,k}^m| \int_U (1 + \|u\|)^{\gamma p} w(u) du \right)^{\frac{1}{p}} \max_{\alpha \in \mathbb{N}_{0,k}^m} \sup_{u \in U} \frac{\|\partial_\alpha f(u)\|}{(1 + \|u\|)^\gamma} \\
&\leqslant \left(C_{24} |\mathbb{N}_{0,k}^m| \right)^{\frac{1}{p}} \|f\|_{C_{pol,\gamma}^k(\mathbb{R}^m; \mathbb{R}^d)}.
\end{aligned}$$

which shows that the restriction map (4) is continuous. In addition, we apply Proposition 4.5 to conclude that $\{g|_U : g \in C_c^\infty(\mathbb{R}^m; \mathbb{R}^d)\}$ is dense in $W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$. Hence, by using that $C_c^\infty(\mathbb{R}^m; \mathbb{R}^d) \subseteq C_b^k(\mathbb{R}^m; \mathbb{R}^d)$, it follows that the image $\{g|_U : g \in C_b^k(\mathbb{R}^m; \mathbb{R}^d)\}$ of the continuous embedding (4) is dense in $W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$. $\square$

4.3. **Proof of results in Section 3.**

4.3.1. *Integral representation.* In this section, we show the integral representation in Proposition 3.4. To this end, we first prove that the ridgelet transform of a fixed function is continuous.

Lemma 4.6. *Let $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$ and $g \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)$. Then, the function $\mathbb{R}^m \times \mathbb{R} \ni (a, b) \mapsto (\mathfrak{R}_\psi g)(a, b) := \int_{\mathbb{R}^m} \psi(a^\top u - b) g(u) \|a\| du \in \mathbb{C}^d$ is continuous.*

Proof. Fix a sequence $(a_M, b_M)_{n \in \mathbb{N}} \subseteq \mathbb{R}^m \times \mathbb{R}$ converging to some $(a, b) \in \mathbb{R}^m \times \mathbb{R}$. Then, by using the constants $C_\psi := \sup_{s \in \mathbb{R}} |\psi(s)| < \infty$ and $C_a := \sup_{M \in \mathbb{N}} \|a_M\| < \infty$, it holds for every $M \in \mathbb{N}$ that

$$\|\psi(a_M^\top u - b_M) g(u)\| \leqslant C_a C_\psi \|g(u)\|,$$

where the right-hand side is Lebesgue-integrable. Moreover, by using that $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$ is continuous, it follows that $\psi(a_M^\top u - b_M) g(u) \|a_M\| \to \psi(a^\top u - b) g(u) \|a\|$ for any $u \in U$, as $n \to \infty$. Hence, we can

apply the vector-valued dominated convergence theorem (see [28, Proposition 2.5]) to conclude that

$$\lim_{M\to\infty}(\mathfrak{R}_\psi g)(a_M,b_M) = \lim_{M\to\infty}\int_{\mathbb{R}^m}\psi\left(a_M^\top u - b_M\right)g(u)\|a_M\|du$$

$$= \int_{\mathbb{R}^m}\psi\left(a^\top u - b\right)g(u)\|a\|du = (\mathfrak{R}_\psi g)(a,b),$$

which completes the proof. □

In order to prove Proposition 3.4, we denote by $\mathbb{S}^{m-1} := \{v \in \mathbb{R}^m : \|v\| = 1\}$ the unit sphere in $\mathbb{R}^m$ and define the space $\mathbb{Y}^{m+1} := \mathbb{S}^{m-1} \times (0,\infty) \times \mathbb{R}$. Then, for any $\psi \in \mathcal{S}_0(\mathbb{R};\mathbb{C})$, we follow [51, Equation 32] and define the ridgelet transform in polar coordinates of any $g \in L^1(\mathbb{R}^m,\mathcal{L}(\mathbb{R}^m),du;\mathbb{R}^d)$ as

$$\mathbb{Y}^{m+1} \ni (v,s,t) \quad \mapsto \quad (\widetilde{\mathfrak{R}}_\psi g)(v,s,t) := \int_{\mathbb{R}^m}g(u)\overline{\psi\left(\frac{v^\top u - t}{s}\right)}\frac{1}{s}du \in \mathbb{C}^d. \tag{27}$$

Moreover, we follow [51, Definition 4.4] and recall that the dual ridgelet transform $\mathfrak{R}_\rho^\dagger$ of any $Q : \mathbb{S}^{m-1} \times (0,\infty) \times \mathbb{R} \to \mathbb{C}$ satisfying $Q(v,s,v^\top u - s\cdot) := (z \mapsto Q(v,s,v^\top u - sz)) \in \mathcal{S}(\mathbb{R};\mathbb{C})$, for all $v \in \mathbb{S}^{m-1}$, $s \in (0,\infty)$, and $u \in \mathbb{R}^m$, is defined by

$$\mathbb{R}^m \ni u \quad \mapsto \quad (\mathfrak{R}_\rho^\dagger Q)(u) := \lim_{\substack{\delta_1\to 0\\\delta_2\to\infty}}\int_{\mathbb{S}^{m-1}}\int_{\delta_1}^{\delta_2}T_\rho\left(Q\left(v,s,v^\top u - s\cdot\right)\right)\frac{1}{s^{m+1}}dsdv \in \mathbb{R}^d.$$

Proof of Proposition 3.4. Fix a function $g = (g_1,...,g_d)^\top \in L^1(\mathbb{R}^m,\mathcal{L}(\mathbb{R}^m),du;\mathbb{R}^d)$ satisfying $\widehat{g} \in L^1(\mathbb{R}^m,\mathcal{L}(\mathbb{R}^m),du;\mathbb{C}^d)$. Then, the latter implies for every $i = 1,...,d$ that

$$\|\widehat{g}_i\|_{L^1(\mathbb{R}^m,\mathcal{L}(\mathbb{R}^m),du)} = \int_{\mathbb{R}^m}\left|\widehat{g}_i(\xi)\right|d\xi \leqslant \int_{\mathbb{R}^m}\|\widehat{g}(\xi)\|d\xi = \|\widehat{g}\|_{L^1(\mathbb{R}^m,\mathcal{L}(\mathbb{R}^m),du;\mathbb{C}^d)} < \infty. \tag{28}$$

Hence, by using that $(\mathfrak{R}_\psi g)(a,b) = 0$ for any $(a,b) \in \{0\} \times \mathbb{R}$, that $(\mathfrak{R}_\psi g)(a,b) = (\widetilde{\mathfrak{R}}_\psi f)\left(\frac{a}{\|a\|},\frac{1}{\|a\|},\frac{b}{\|a\|}\right)$ for any $(a,b) \in (\mathbb{R}^m\backslash\{0\}) \times \mathbb{R}$, the substitution $(\mathbb{R}^m\backslash\{0\}) \times \mathbb{R} \ni (a,b) \mapsto (v,s,z) := \left(\frac{a}{\|a\|},\frac{1}{\|a\|},a^\top u - b\right) \in \mathbb{S}^{m-1} \times (0,\infty) \times \mathbb{R}$ with Jacobi determinante $dbda = s^{-m}dzdsdv$, and [51, Theorem 5.6] applied to $f_i \in L^1(\mathbb{R}^m,\mathcal{L}(\mathbb{R}^m),du)$ with $\widehat{f}_i \in L^1(\mathbb{R}^m,\mathcal{L}(\mathbb{R}^m),du;\mathbb{C})$ (see (28)), it follows for a.e. $u \in \mathbb{R}^m$ that

$$\int_{\mathbb{R}^m}\int_{\mathbb{R}}(\mathfrak{R}_\psi g)(a,b)\rho\left(a^\top u - b\right)dbda = \int_{\mathbb{R}^m\backslash\{0\}}\int_{\mathbb{R}}(\widetilde{\mathfrak{R}}_\psi g)\left(\frac{a}{\|a\|},\frac{1}{\|a\|},\frac{b}{\|a\|}\right)\rho\left(a^\top u - b\right)dbda$$

$$= \left(\int_{\mathbb{S}^{m-1}}\int_0^\infty\int_{\mathbb{R}}(\widetilde{\mathfrak{R}}_\psi g_i)\left(v,s,v^\top u - sz\right)\rho(z)\frac{1}{s^{m+1}}dzdsdv\right)^\top_{i=1,...,d}$$

$$= \left(\lim_{\substack{\delta_1\to 0\\\delta_2\to\infty}}\int_{\mathbb{S}^{m-1}}\int_{\delta_1}^{\delta_2}T_\rho\left((\widetilde{\mathfrak{R}}_\psi g_i)\left(v,s,v^\top u - s\cdot\right)\right)\frac{1}{s^{m+1}}dsdv\right)^\top_{i=1,...,d}$$

$$= \left((\widetilde{\mathfrak{R}}_\rho^\dagger\mathfrak{R}_\psi g_i)(u)\right)^\top_{i=1,...,d} = \left(C_m^{(\psi,\rho)}g_i(u)\right)^\top_{i=1,...,d} = C_m^{(\psi,\rho)}g(u),$$

which completes the proof. □

4.3.2. *Properties of weighted Sobolev space* $W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)$. In this section, we show that the weighted Sobolev space $(W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d),\|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d)})$ introduced in Footnote 5 is separable and has Banach space type $t := \min(2,p)$, where we set $(W^{0,p}(U,\mathcal{L}(U),w;\mathbb{R}^d),\|\cdot\|_{W^{0,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}) := (L^p(U,\mathcal{L}(U),w(u)du;\mathbb{R}^d),\|\cdot\|_{L^p(U,\mathcal{L}(U),w(u)du;\mathbb{R}^d)})$.

Lemma 4.7. *Let* $k \in \mathbb{N}_0$, $p \in [1,\infty)$, $U \subseteq \mathbb{R}^m$ *(open, if* $k \geqslant 1$*), and* $w : U \to [0,\infty)$ *be a weight. Then, the Banach space* $(W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d),\|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d)})$ *is separable.*

Proof. First, we show the conclusion for $k = 0$, i.e. that the Banach space $(W^{0,p}(U,\mathcal{L}(U),w;\mathbb{R}^d),\|\cdot\|_{W^{0,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}) := (L^p(U,\mathcal{L}(U),w(u)du;\mathbb{R}^d),\|\cdot\|_{L^p(U,\mathcal{L}(U),w(u)du;\mathbb{R}^d)})$ is separable. For this purpose, we observe that $\mathcal{B}(U)$ is generated by sets of the form $U \cap \times_{l=1}^m[r_{l,1},r_{l,2})$, with $r_{l,1},r_{l,2} \in \mathbb{Q}$, $l = 1,...,m$. Moreover, by using that $\mathcal{L}(U)$ and $\mathcal{B}(U)$ coincide up to Lebesgue nullsets and that $w : U \to [0,\infty)$ is a weight, ensuring that the measure spaces $(U,\mathcal{L}(U),w(u)du)$ and $(U,\mathcal{L}(U),du)$ share the same null sets, we conclude that $(U,\mathcal{L}(U),w(u)du)$ is countably generated up to $(w(u)du)$-null sets. Hence,

by using [15, p. 92] componentwise, it follows that $(W^{0,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \| \cdot \|_{W^{0,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}) :=$ $(L^p(U, \mathcal{L}(U), w(u)du; \mathbb{R}^d), \| \cdot \|_{L^p(U,\mathcal{L}(U),w(u)du;\mathbb{R}^d)})$ is separable.

Now, for the general case of $k \geqslant 1$, we consider the Banach space $(W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d), \| \cdot \|_{W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d)})$. Then, we define the map

$$W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d) \ni f \quad \mapsto \quad \Xi(f) := (\partial_\alpha f)_{\alpha \in \mathbb{N}_{0,k}^m} \in \underset{\alpha \in \mathbb{N}_{0,k}^m}{\bigtimes} L^p(U, \mathcal{L}(U), w(u)du, \mathbb{R}^d) =: Z,$$

where Z is equipped with the norm $\|g\|_Z := \sum_{\alpha \in \mathbb{N}_{0,k}^m} \|g_\alpha\|_{L^p(U,\mathcal{L}(U),du,\mathbb{R}^d)}$, for $g := (g_\alpha)_{\alpha \in \mathbb{N}_{0,k}^m} \in Z$. Then, by using the previous step, we conclude that the Banach space $(Z, \|\cdot\|_Z)$ is separable as finite product of separable Banach spaces. Hence, by using that $W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d)$ is by definition isometrically isomorphic to the closed vector subspace $\text{Img}(\Xi) := \{\Xi(f) : f \in W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d)\} \subseteq Z$, it follows that $(W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d), \| \cdot \|_{W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d)})$ is separable. $\square$

Moreover, we recall the notion of Banach space types and refer to [4, Section 6.2], [36, Chapter 9], and [28, Section 4.3.b] for more details.

Definition 4.8 ([28, Definition 4.3.12 (1)]). *A Banach space $(X, \| \cdot \|_X)$ is called of* type $t \in [1, 2]$ *if there exists a constant $C_X > 0$ such that for every $N \in \mathbb{N}$, $(f_n)_{n=1,\dots,N} \subseteq X$, and Rademacher sequence*[10] *$(\epsilon_n)_{n=1,\dots,N}$ on a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, it holds that*

$$\widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^N \epsilon_n f_n\right\|_X^t\right]^{\frac{1}{t}} \leqslant C_X \left(\sum_{n=1}^N \|f_n\|_X^t\right)^{\frac{1}{t}}.$$

Every Banach space $(X, \| \cdot \|_X)$ is of type $t = 1$ with constant $C_X = 1$, whereas only some Banach spaces have non-trivial type $t \in (1, 2]$, e.g., every Hilbert space $(X, \| \cdot \|_X)$ is of type $t = 2$ with constant $C_X = 1$ (see [4, Remark 6.2.11 (b)+(c)]). Moreover, $(L^p(U, \mathcal{B}(U), \mu; \mathbb{R}^d), \| \cdot \|_{L^p(U,\mathcal{B}(U),\mu;\mathbb{R}^d)})$ introduced in Footnote 2 is a Banach space of type $t = \min(2, p)$ with constant $C_{L^p(U,\Sigma,\mu;\mathbb{R}^d)} > 0$ depending only on $p \in [1, \infty)$ (see [4, Theorem 6.2.14]). Now, we show that this still holds true for the weighted Sobolev space $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \| \cdot \|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$ introduced in 5, where $(W^{0,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \| \cdot \|_{W^{0,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}) := (L^p(U, \mathcal{L}(U), w(u)du; \mathbb{R}^d), \| \cdot \|_{L^p(U,\mathcal{L}(U),w(u)du;\mathbb{R}^d)})$.

Lemma 4.9. *Let $k \in \mathbb{N}_0$, $p \in [1, \infty)$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), and $w : U \to [0, \infty)$ be a weight. Then, the Banach space $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \| \cdot \|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$ is of type $t = \min(2, p)$ with constant $C_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)} > 0$ depending only on $p \in [1, \infty)$.*

Proof. First, we recall that $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \| \cdot \|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$ is a Banach space. Indeed, this follows from [46, p. 96] (for $k = 0$) and [3, Theorem 3.2] (for $k \geqslant 1$).

Now, we fix some $N \in \mathbb{N}$, $(f_n)_{n=1,\dots,N} \subseteq W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$, and an i.i.d. sequence $(\epsilon_n)_{n=1,\dots,N}$ defined on a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ such that $\widetilde{\mathbb{P}}[\epsilon_n = \pm 1] = 1/2$. Then, by using Fubini's theorem and the classical Khintchine inequality in [36, Lemma 4.1] with constant $C_p > 0$ depending only on $p \in [1, \infty)$, it follows that

$$\widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^N \epsilon_n f_n\right\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^p\right]^{\frac{1}{p}} = \widetilde{\mathbb{E}}\left[\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \left\|\sum_{n=1}^N \epsilon_n \partial_\alpha f_n(u)\right\|^p w(u)du\right]^{\frac{1}{p}}$$

$$= \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^N \epsilon_n \partial_\alpha f_n(u)\right\|^p\right] w(u)du\right)^{\frac{1}{p}} \quad (29)$$

$$\leqslant C_p \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \left(\sum_{n=1}^N \|\partial_\alpha f_n(u)\|^2\right)^{\frac{p}{2}} w(u)du\right)^{\frac{1}{p}}.$$

---

[10]A *Rademacher sequence* $(\epsilon_n)_{n=1,\dots,N}$ on a given probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ is an i.i.d. sequence of random variables $(\epsilon_n)_{n=1,\dots,N}$ such that $\widetilde{\mathbb{P}}[\epsilon_n = \pm 1] = 1/2$.

If $p \in [1,2]$, we use (29) and the inequality $\left(\sum_{n=1}^{N} x_n\right)^{p/2} \leqslant \sum_{n=1}^{N} x_n^{p/2}$ for any $x_1, ..., x_N \geqslant 0$ to conclude that

$$
\begin{aligned}
\widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^{N} \epsilon_n f_n\right\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^{\min(2,p)}\right]^{\frac{1}{\min(2,p)}} &= \widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^{N} \epsilon_n f_n\right\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^{p}\right]^{\frac{1}{p}} \\
&\leqslant C_p \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \left(\sum_{n=1}^{N} \|\partial_\alpha f_n(u)\|^2\right)^{\frac{p}{2}} w(u) du\right)^{\frac{1}{p}} \\
&\leqslant C_p \left(\sum_{n=1}^{N} \sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f_n(u)\|^p w(u) du\right)^{\frac{1}{p}} \\
&= C_p \left(\sum_{n=1}^{N} \|f_n\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^{\min(2,p)}\right)^{\frac{1}{\min(2,p)}}.
\end{aligned}
$$

This shows for $p \in [1,2]$ that the Banach space $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$ is of type $t = \min(2,p)$, where the constant $C_p > 0$ depends only on $p \in [1,\infty)$.

Otherwise, if $p \in (2,\infty)$, we consider the measure spaces $(\{1, ..., N\}, \mathcal{P}(\{1, ..., N\}), \eta)$ and $(\mathbb{N}_{0,k}^m \times U, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U), \mu \otimes w)$, where $\mathcal{P}(\{1, ..., N\})$ and $\mathcal{P}(\mathbb{N}_{0,k}^m)$ denote the power sets of $\{1, ..., N\}$ and $\mathbb{N}_{0,k}^m$, respectively, and where $\mathcal{P}(\{1, ..., N\}) \ni A \mapsto \eta(A) := \sum_{n=1}^N \mathbb{1}_A(n) \in [0,\infty)$ and $\mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U) \ni (A,B) \mapsto (\mu \otimes w)(A,B) := \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \mathbb{1}_A(\alpha)\right) \int_B w(u) du \in [0,\infty]$ are both measures. Then, by using the Minkowski inequality in [28, Proposition 1.2.22] with $p \geqslant 2$, it follows for every $\mathbf{f} \in L^2(\{1, ..., N\}, \mathcal{P}(\{1, ..., N\}), \eta; L^p(\mathbb{N}_{0,k}^m \times U, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U), \mu \otimes w; \mathbb{R}^d))$ that

$$
\begin{aligned}
&\|\mathbf{f}\|_{L^p(\mathbb{N}_{0,k}^m \times U, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U), \mu \otimes w; L^2(\{1,...,N\}, \mathcal{P}(\{1,...,N\}), \eta; \mathbb{R}^d))} \\
&\quad \leqslant \|\mathbf{f}\|_{L^2(\{1,...,N\}, \mathcal{P}(\{1,...,N\}), \eta; L^p(\mathbb{N}_{0,k}^m \times U, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U), \mu \otimes w; \mathbb{R}^d))}.
\end{aligned}
\tag{30}
$$

Now, we define the map $\{1, ..., N\} \times (\mathbb{N}_{0,k}^m \times U) \ni (n; \alpha, u) \mapsto \mathbf{f}(n; \alpha, u) := \partial_\alpha f_n(u) \in \mathbb{R}^d$ satisfying

$$
\begin{aligned}
\|\mathbf{f}\|_{L^2(\{1,...,N\}, \mathcal{P}(\{1,...,N\}), \eta; L^p(\mathbb{N}_{0,k}^m \times U, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U), \mu \otimes w; \mathbb{R}^d))} &= \left(\sum_{n=1}^{N} \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \|\partial_\alpha f_n(u)\|^p w(u) du\right)^{\frac{2}{p}}\right)^{\frac{1}{2}} \\
&= \left(\sum_{n=1}^{N} \|f_n\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^2\right)^{\frac{1}{2}} < \infty,
\end{aligned}
\tag{31}
$$

which shows that $\mathbf{f} \in L^2(\{1, ..., N\}, \mathcal{P}(\{1, ..., N\}), \eta; L^p(\mathbb{N}_{0,k}^m \times U, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U), \mu \otimes w; \mathbb{R}^d))$. Hence, by using first Jensen's inequality and then by combining (29) and (30) with (31), we conclude that

$$
\begin{aligned}
\widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^{N} \epsilon_n f_n\right\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^{\min(2,p)}\right]^{\frac{1}{\min(2,p)}} &\leqslant \widetilde{\mathbb{E}}\left[\left\|\sum_{n=1}^{N} \epsilon_n f_n\right\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^{p}\right]^{\frac{1}{p}} \\
&\leqslant C_p \left(\sum_{\alpha \in \mathbb{N}_{0,k}^m} \int_U \left(\sum_{n=1}^{N} \|\partial_\alpha f_n(u)\|^2\right)^{\frac{p}{2}} w(u) du\right)^{\frac{1}{p}} \\
&= C_p \|\mathbf{f}\|_{L^p(\mathbb{N}_{0,k}^m \times U, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U), \mu \otimes w; L^2(\{1,...,N\}, \mathcal{P}(\{1,...,N\}), \eta; \mathbb{R}^d))} \\
&\leqslant C_p \|\mathbf{f}\|_{L^2(\{1,...,N\}, \mathcal{P}(\{1,...,N\}), \eta; L^p(\mathbb{N}_{0,k}^m \times U, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{L}(U), \mu \otimes w; \mathbb{R}^d))} \\
&= C_p \left(\sum_{n=1}^{N} \|f\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^{\min(2,p)}\right)^{\frac{1}{\min(2,p)}}.
\end{aligned}
$$

This shows for $p \in (2, \infty)$ that the Banach space $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \| \cdot \|_{W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)})$ is of type $t = \min(2, p)$, where the constant $C_p > 0$ depends only on $p \in [1, \infty)$. $\qquad\square$

4.3.3. *Randomized neurons and strong measurability.* In this section, we randomly initialize the weight vectors and biases inside the activation function to obtain the approximation rates in Theorem 3.6. To this end, we first show that the map from the parameters to a neuron is continuous.

Lemma 4.10. *For $k \in \mathbb{N}_0$, $p \in [1, \infty)$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), $\gamma \in [0, \infty)$, and $\rho \in C^k_{pol,\gamma}(\mathbb{R})$, let $w : U \to [0, \infty)$ be a weight such that the constant $C^{(\gamma,p)}_{U,w} > 0$ defined in (7) is finite. Then, the mapping*

$$\mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R} \ni (y, a, b) \quad \mapsto \quad y\rho\left(a^\top \cdot - b\right) \in W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d)$$

is continuous, where $y\rho\left(a^\top \cdot - b\right)$ denotes the function $U \ni u \mapsto y\rho\left(a^\top u - b\right) \in \mathbb{R}^d$.

Proof. Fix a sequence $(y_M, a_M, b_M)_{M \in \mathbb{N}} \subseteq \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}$ converging to $(y, a, b) \in \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}$. Then, by using that $y_M a_M^\alpha (1 + \|a_M\| + |b_M|)$ converges uniformly in $\alpha \in \mathbb{N}^m_{0,k}$ to $ya^\alpha(1 + \|a\| + |b|)$, where $a^\alpha := \prod_{l=1}^m a_l^{\alpha_l}$ for $a := (a_1, ..., a_m)^\top \in \mathbb{R}^m$ and $\alpha := (\alpha_1, ..., \alpha_m) \in \mathbb{N}^m_{0,k}$, the constant $C_{y,a,b} := \max_{\alpha \in \mathbb{N}^m_{0,k}} \|ya^\alpha\| (1 + \|a\| + |b|) + \sup_{M \in \mathbb{N}} \left(\max_{\alpha \in \mathbb{N}^m_{0,k}} \|y_M a_M^\alpha\| (1 + \|a_M\| + |b_M|) \right) \geqslant 0$ is finite. Hence, by using that $\rho \in C^k_{pol,\gamma}(\mathbb{R})$, i.e. that $\left|\rho^{(j)}(s)\right| \leqslant \|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})} (1 + |s|)^\gamma$ for any $j = 0, ..., k$ and $s \in \mathbb{R}$, the inequality $1 + \left|a_M^\top u - b_M\right| \leqslant 1 + \|a_M\|\|u\| + |b_M| \leqslant (1 + \|a_M\| + |b_M|)(1 + \|u\|)$ for any $M \in \mathbb{N}$ and $u \in \mathbb{R}^m$, it follows for every $\alpha \in \mathbb{N}^m_{0,k}$, $u \in U$, and $M \in \mathbb{N}$ that

$$\begin{aligned}
\left\| y_M \rho^{(|\alpha|)}\left(a_M^\top u - b_M\right) a_M^\alpha \right\| &\leqslant \|y_M a_M^\alpha\| \left|\rho^{(|\alpha|)}\left(a_M^\top u - b_M\right)\right| \\
&\leqslant \|y_M a_M^\alpha\| \|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})} \left(1 + \left|a_M^\top u - b_M\right|\right)^\gamma \\
&\leqslant \|y_M a_M^\alpha\| (1 + \|a_M\| + |b_M|)^\gamma \|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})} (1 + \|u\|)^\gamma \\
&\leqslant C_{y,a,b} \|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})} (1 + \|u\|)^\gamma.
\end{aligned} \tag{32}$$

Analogously, we conclude for every $\alpha \in \mathbb{N}^m_{0,k}$ and $u \in U$ that

$$\left\| y_M \rho^{(|\alpha|)}\left(a_M^\top u - b_M\right) a_M^\alpha \right\| \leqslant C_{y,a,b} \|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})} (1 + \|u\|)^\gamma. \tag{33}$$

Hence, by using the triangle inequality together with the inequality $(x + y)^p \leqslant 2^{p-1} (x^p + y^p)$ for any $x, y \geqslant 0$ as well as the inequalities (32) and (33), it follows for every $\alpha \in \mathbb{N}^m_{0,k}$, $u \in U$, and $M \in \mathbb{N}$ that

$$\begin{aligned}
&\left\| y\rho^{(|\alpha|)}\left(a^\top u - b\right) a^\alpha - y_M \rho^{(|\alpha|)}\left(a_M^\top u - b_M\right) a_M^\alpha \right\|^p \\
&\leqslant 2^{p-1} \left(\left\| y\rho^{(|\alpha|)}\left(a^\top u - b\right) a^\alpha \right\|^p + \left\| y_M \rho^{(|\alpha|)}\left(a_M^\top u - b_M\right) a_M^\alpha \right\|^p \right) \\
&\leqslant 2^p C^p_{y,a,b} \|\rho\|^p_{C^k_{pol,\gamma}(\mathbb{R})} (1 + \|u\|)^{\gamma p}.
\end{aligned} \tag{34}$$

Thus, by applying the $\mathbb{R}^d$-valued dominated convergence theorem (see [28, Proposition 1.2.5] with (34) and $\int_U (1 + \|u\|)^{\gamma p} w(u) du = \left(C^{(\gamma,p)}_{U,w}\right)^p < \infty$ by assumption), we have

$$\lim_{M \to \infty} \left\| y\rho\left(a^\top \cdot - b\right) - y_M \rho\left(a_M^\top \cdot - b_M\right) \right\|_{W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)}$$

$$= \left(\sum_{\alpha \in \mathbb{N}^m_{0,k}} \lim_{M \to \infty} \int_U \left\| y\rho^{(|\alpha|)}\left(a^\top u - b\right) a^\alpha - y_M \rho^{(|\alpha|)}\left(a_M^\top u - b_M\right) a_M^\alpha \right\|^p w(u) du \right)^{\frac{1}{p}} = 0,$$

which completes the proof. $\qquad\square$

Moreover, we fix throughout the rest of this paper a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assume that $(a_n)_{n \in \mathbb{N}} \sim t_m$ and $(b_n)_{n \in \mathbb{N}} \sim t_1$ are independent sequences of independent and identically distributed (i.i.d.) random variables following a (multivariate) Student's t-distribution[11]. In this case, we write $(a_n, b_n)_{n \in \mathbb{N}} \sim t_m \otimes t_1$. Then, we show that a randomized neuron with $(a_n, b_n)_{n \in \mathbb{N}} \sim t_m \otimes t_1$ used for

---

[11] For any $m \in \mathbb{N}$, a random variable $a \sim t_m$ following a Student's t-distribution has probability density function $\mathbb{R}^m \ni a \mapsto p_a(a) = \frac{\Gamma((m+1)/2)}{\pi^{(m+1)/2}} \left(1 + \|a\|^2\right)^{-(m+1)/2} \in (0, \infty)$.

the weight vectors and biases inside the activation function, is a strongly measurable map in the sense of [28, Definition 1.1.14], where we define the σ-algebra $\mathcal{F}_{a,b} := \sigma(\{a_n, b_n : n \in \mathbb{N}\})$.

Lemma 4.11. *For $k \in \mathbb{N}_0$, $p \in [1, \infty)$, $U \subseteq \mathbb{R}^m$ (open, if $k \geqslant 1$), $\gamma \in [0, \infty)$, and $\rho \in C^k_{pol,\gamma}(\mathbb{R})$, let $w : U \to [0, \infty)$ be a weight such that the constant $C^{(\gamma,p)}_{U,w} > 0$ defined in (7) is finite. Moreover, for $n \in \mathbb{N}$ and an $\mathcal{F}_{a,b}/\mathcal{B}(\mathbb{R}^d)$-measurable random vector $y : \Omega \to \mathbb{R}^d$, we define the map*

$$\Omega \ni \omega \quad \mapsto \quad R_n(\omega) := y(\omega)\rho\left(a_n(\omega)^\top \cdot -b_n(\omega)\right) \in W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d). \tag{35}$$

Then, $R_n : \Omega \to W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d)$ is a strongly $(\mathbb{P}, \mathcal{F}_{a,b})$-measurable map with values in the separable Banach space $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$.

Proof. First, we show that the map $R_n : \Omega \to W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d)$ takes values in the Banach space $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$, which is by Lemma 4.7 separable. Indeed, since $\rho \in C^k_{pol,\gamma}(\mathbb{R})$ is k-times differentiable, it follows for every fixed $\omega \in \Omega$ and $\alpha \in \mathbb{N}^m_{0,k}$ that $U \ni u \mapsto \partial_\alpha R_n(\omega) = y(\omega)\rho^{(|\alpha|)}\left(a_n(\omega)^\top u - b_n(\omega)\right) a_n(\omega)^\alpha \in \mathbb{R}^d$ is $\mathcal{L}(U)/\mathcal{B}(\mathbb{R}^d)$-measurable. Moreover, by using that $\rho \in C^k_{pol,\gamma}(\mathbb{R})$, i.e. that $|\rho^{(j)}(s)| \leqslant \|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})}(1 + |s|)^\gamma$ for any $j = 0, ..., k$ and $s \in \mathbb{R}$, the inequality $1 + |a_n(\omega)^\top u - b_n(\omega)| \leqslant 1 + \|a_n(\omega)\|\|u\| + |b_n(\omega)| \leqslant (1 + \|a_n(\omega)\| + |b_n(\omega)|)(1 + \|u\|)$ for any $u \in \mathbb{R}^m$, and that $C^{(\gamma,p)}_{U,w} := \left(\int_U (1 + \|u\|)^{\gamma p} w(u) du\right)^{1/p} > 0$ is finite, we conclude that

$$\|R_n(\omega)\|^p_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)} = \sum_{\alpha \in \mathbb{N}^m_{0,k}} \int_U \left\| y(\omega)\rho^{(|\alpha|)}\left(a_n(\omega)^\top u - b_n(\omega)\right) a_n(\omega)^\alpha \right\|^p w(u) du$$

$$\leqslant \left(\sum_{\alpha \in \mathbb{N}^m_{0,k}} \|y(\omega)a_n(\omega)^\alpha\|^p \right) \int_U \left(1 + |a_n(\omega)^\top u - b_n(\omega)|\right)^{\gamma p} w(u) du$$

$$\leqslant \left(\sum_{\alpha \in \mathbb{N}^m_{0,k}} \|y(\omega)a_n(\omega)^\alpha\|^p \right) (1 + \|a_n(\omega)\| + |b_n(\omega)|)^{\gamma p} \int_U (1 + \|u\|)^{\gamma p} w(u) du < \infty.$$

This shows that $R_n(\omega) \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ for all $\omega \in \Omega$.

Finally, in order to show that the map (35) is strongly $(\mathbb{P}, \mathcal{F}_{a,b})$-measurable, we use that $\Omega \ni \omega \mapsto (y(\omega), a_n(\omega), b_n(\omega)) \in \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}$ is by definition $\mathcal{F}_{a,b}/\mathbb{B}(\mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R})$-measurable and that $\mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R} \ni (y, a, b) \mapsto y\rho\left(a^\top \cdot -b\right) \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ is by Lemma 4.10 continuous to conclude that the concatenation (35) is $\mathcal{F}_{a,b}/\mathcal{B}(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d))$-measurable, where $\mathcal{B}(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d))$ denotes the Borel σ-algebra of $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$. Since $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d, \|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d})$ is by Lemma 4.7 separable, we can apply [28, Theorem 1.1.6+1.1.20] to conclude that (35) is strongly $(\mathbb{P}, \mathcal{F}_{a,b})$-measurable. $\square$

4.3.4. *Proof of Theorem 3.6.* In this section, we prove the approximation rates in Theorem 3.6. Let us first sketch the main ideas of the proof. For some fixed $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d) \cap \mathbb{B}^{k,\gamma}_\psi(U; \mathbb{R}^d)$ and $N \in \mathbb{N}$, we use the randomized neuron $R_n : \Omega \to W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d)$ in (35) with a particular linear readout. Then, by using the integral representation in Proposition 3.4 implying that $f = \mathbb{E}[R_n]$, a symmetrization argument with Rademacher averages, and the Banach space type of $W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d)$, we obtain

$$\mathbb{E}\left[\left\| f - \frac{1}{N}\sum_{n=1}^N R_n \right\|^2_{W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d)} \right]^{\frac{1}{2}} = \frac{1}{N}\mathbb{E}\left[\left\| \sum_{n=1}^N (\mathbb{E}[R_n] - R_n) \right\|^2_{W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d)} \right]^{\frac{1}{2}}$$

$$\leqslant C \frac{\|R_n\|_{L^2(\Omega,\mathcal{F},\mathbb{P};W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d))}}{N^{1 - \frac{1}{\min(2,p)}}},$$

where $C > 0$ is a constant and where $\|R_n\|_{L^2(\Omega, \mathcal{F}, \mathbb{P}; W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d))}$ can be bounded by $\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)}$. Hence, there exists some $\omega \in \Omega$ such that the neural network $\varphi_N := \frac{1}{N} \sum_{n=1}^N R_n(\omega) \in \mathcal{NN}_{U,d}^\rho$ satisfies

$$
\begin{aligned}
\|f - \varphi\|_{W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d)} &\leqslant \mathbb{E}\left[\left\|f - \frac{1}{N}\sum_{n=1}^N R_n\right\|_{W^{k,p}(U,\mathcal{L}(U),w,\mathbb{R}^d)}^2\right]^{\frac{1}{2}} \\
&\leqslant C\frac{\|R_n\|_{L^2(\Omega,\mathcal{F},\mathbb{P};W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d))}}{N^{1-\frac{1}{\min(2,p)}}}.
\end{aligned}
$$

Proof of Theorem 3.6. Fix $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d) \cap \mathbb{B}_\psi^{k,\gamma}(U; \mathbb{R}^d)$ and $N \in \mathbb{N}$. Then, by definition of $\mathbb{B}_\psi^{k,\gamma}(U; \mathbb{R}^d)$, there exists some $g \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)$ with $\widehat{g} \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{C}^d)$ such that

$$
\left(\int_{\mathbb{R}^m}\int_{\mathbb{R}}\left(1 + \|a\|^2\right)^{\gamma + k + \frac{m+1}{2}}\left(1 + |b|^2\right)^{\gamma+1}\|(\mathfrak{R}_\psi g)(a,b)\|^2 db\, da\right)^{\frac{1}{2}} \leqslant 2\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)}. \tag{36}
$$

From this, we define for every $n = 1, ..., N$ the map

$$
\Omega \ni \omega \quad \mapsto \quad R_n(\omega) := y_n(\omega)\rho\left(a_n(\omega)^\top \cdot - b_n(\omega)\right) \in W^{k,p}(U, \mathcal{L}(U), w, \mathbb{R}^d) \tag{37}
$$

with

$$
\Omega \ni \omega \quad \mapsto \quad y_n(\omega) := \mathrm{Re}\left(\frac{(\mathfrak{R}_\psi g)(a_n(\omega), b_n(\omega))}{C_m^{(\psi,\rho)} p_a(a_n(\omega)) p_b(b_n(\omega))}\right) \in \mathbb{R}^d, \tag{38}
$$

where $p_a : \mathbb{R}^m \to (0, \infty)$ and $p_b : \mathbb{R} \to (0, \infty)$ denote the probability density function of the (multivariate) Student's t-distributions[11]. Then, by using that $\mathfrak{R}_\psi : \mathbb{R}^m \times \mathbb{R} \to \mathbb{C}^d$ is continuous (see Lemma 4.6), we observe that $y_n : \Omega \to \mathbb{R}^d$ is $\mathcal{F}_{a,b}/\mathcal{B}(\mathbb{R}^d)$-measurable. Hence, we can apply Lemma 4.11 to conclude that $R_n : \Omega \to W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ is a strongly $(\mathbb{P}, \mathcal{F}_{a,b})$-measurable map with values in the separable Banach space $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$.

Now, we show $R_n \in L^2(\Omega, \mathcal{F}_{a,b}, \mathbb{P}; W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d))$ and $\mathbb{E}[R_n] = f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$. To this end, we use that $\rho \in C_{pol,\gamma}^k(\mathbb{R})$, i.e. that $\left|\rho^{(j)}(s)\right| \leqslant \|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}(1 + |s|)^\gamma$ for any $j = 0, ..., k$ and $s \in \mathbb{R}$, the inequality $1 + \left|a^\top u - b\right| \leqslant 1 + \|a\|\|u\| + |b| \leqslant (1 + \|a\|)(1 + |b|)(1 + \|u\|)$ for any $a, u \in \mathbb{R}^m$ and $b \in \mathbb{R}$, twice the inequality $(x + y)^2 \leqslant 2\left(x^2 + y^2\right)$ for any $x, y \in [0, \infty)$, and the finite constant $C_{U,w}^{(\gamma,p)} > 0$ to conclude for every $a \in \mathbb{R}^m$, $b \in \mathbb{R}$, and $j = 0, ..., k$ that

$$
\begin{aligned}
\left(\int_U \left|\rho^{(j)}\left(a^\top u - b\right)\right|^p w(u) du\right)^{\frac{1}{p}} &\leqslant \|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\left(\int_U \left(1 + \left|a^\top u - b\right|\right)^{\gamma p} w(u) du\right)^{\frac{1}{p}} \\
&\leqslant \|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}(1 + \|a\|)^\gamma(1 + |b|)^\gamma\left(\int_U (1 + \|u\|)^{\gamma p} w(u) du\right)^{\frac{1}{p}} \\
&\leqslant 4\|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\left(1 + \|a\|^2\right)^{\frac{\gamma}{2}}\left(1 + |b|^2\right)^{\frac{\gamma}{2}} C_{U,w}^{(\gamma,p)}.
\end{aligned}
\tag{39}
$$

Hence, by using the inequality $|a^\alpha| := \prod_{l=1}^{m} |a_l|^{\alpha_l} \leqslant \left(1 + \|a\|^2\right)^{|\alpha|/2} \leqslant \left(1 + \|a\|^2\right)^{k/2}$ for any $\alpha \in \mathbb{N}_{0,k}^m$ and $a \in \mathbb{R}^m$, the inequality (39), that $\left|\mathbb{N}_{0,k}^m\right| = \sum_{j=0}^{k} m^j \leqslant 2m^k$, and the inequality (37), we obtain that

$$
\|R_n\|_{L^2(\Omega,\mathcal{F},\mathbb{P};W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d))} = \mathbb{E}\left[\left\|y_n\rho\left(a_n^\top \cdot \, - b_n\right)\right\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}^2\right]^{\frac{1}{2}}
$$

$$
= \mathbb{E}\left[\left(\sum_{\alpha\in\mathbb{N}_{0,k}^m}\int_U \left\|\partial_\alpha\left(y_n\rho\left(a_n^\top u - b_n\right)\right)\right\|^p du\right)^{\frac{2}{p}}\right]^{\frac{1}{2}}
$$

$$
= \mathbb{E}\left[\left(\sum_{\alpha\in\mathbb{N}_{0,k}^m}\left\|\mathrm{Re}\left(\frac{a_n^\alpha(\mathfrak{R}_\psi g)(a_n,b_n)}{C_m^{(\psi,\rho)}p_a(a_n)p_b(b_n)}\right)\right\|^p\int_U\left|\rho^{(|\alpha|)}\left(a_n^\top u - b_n\right)\right|^p du\right)^{\frac{2}{p}}\right]^{\frac{1}{2}}
$$

$$
\leqslant 4\|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\frac{C_{U,w}^{(\gamma,p)}\left|\mathbb{N}_{0,k}^m\right|^{\frac{1}{p}}}{\left|C_m^{(\psi,\rho)}\right|}\mathbb{E}\left[\frac{\left(1 + \|a_n\|^2\right)^{\gamma+k}\left(1 + |b_n|^2\right)^\gamma}{p_a(a_n)^2 p_b(b_n)^2}\|(\mathfrak{R}_\psi g)(a_n,b_n)\|^2\right]^{\frac{1}{2}}
$$

$$
\leqslant 2^{3+\frac{1}{p}}\|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\frac{C_{U,w}^{(\gamma,p)}m^{\frac{k}{p}}}{\left|C_m^{(\psi,\rho)}\right|}\left(\int_{\mathbb{R}^m}\int_{\mathbb{R}}\frac{\left(1 + \|a\|^2\right)^{\gamma+k}\left(1 + |b|^2\right)^\gamma}{p_a(a)^2 p_b(b)^2}\|(\mathfrak{R}_\psi g)(a,b)\|^2 p_a(a)p_b(b)\,db\,da\right)^{\frac{1}{2}}
$$

$$
\leqslant 2^{3+\frac{1}{p}}\|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\frac{C_{U,w}^{(\gamma,p)}m^{\frac{k}{p}}}{\left|C_m^{(\psi,\rho)}\right|}
$$

$$
\cdot\left(\frac{\pi^{\frac{m+1}{2}}}{\Gamma\left(\frac{m+1}{2}\right)}\pi\int_{\mathbb{R}^m}\int_{\mathbb{R}}\left(1 + \|a\|^2\right)^{\gamma+k+\frac{m+1}{2}}\left(1 + |b|^2\right)^{\gamma+1}\|(\mathfrak{R}_\psi g)(a,b)\|^2\,db\,da\right)^{\frac{1}{2}}
$$

$$
\leqslant 2^{4+\frac{1}{p}}\pi\|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\frac{C_{U,w}^{(\gamma,p)}m^{\frac{k}{p}}\pi^{\frac{m+1}{4}}}{\left|C_m^{(\psi,\rho)}\right|\Gamma\left(\frac{m+1}{2}\right)^{\frac{1}{2}}}\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)} < \infty,
$$

$$\tag{40}$$

which shows that $R_n \in L^2(\Omega,\mathcal{F}_{a,b},\mathbb{P};W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d))$. Moreover, by using the probability density functions $p_a : \mathbb{R}^m \to (0,\infty)$ and $p_b : \mathbb{R} \to (0,\infty)$, Proposition 3.4, and that $f = g$ a.e. on U, it follows for a.e. $u \in U$ that

$$
\mathbb{E}[R_n(u)] = \mathbb{E}\left[\mathrm{Re}\left(\frac{(\mathfrak{R}_\psi g)(a_n,b_n)}{C_m^{(\psi,\rho)}p_a(a_n)p_b(b_n)}\right)\rho\left(a_n^\top u - b_n\right)\right]
$$

$$
= \int_{\mathbb{R}^m}\int_{\mathbb{R}}\mathrm{Re}\left(\frac{(\mathfrak{R}_\psi g)(a,b)}{C_m^{(\psi,\rho)}p_a(a)p_b(b)}\right)\rho\left(a^\top u - b\right)p_a(a)p_b(b)\,db\,da
$$

$$
= \mathrm{Re}\left(\frac{1}{C_m^{(\psi,\rho)}}\int_{\mathbb{R}^m}\int_{\mathbb{R}}(\mathfrak{R}_\psi g)(a,b)\rho\left(a^\top u - b\right)db\,da\right)
$$

$$
= \mathrm{Re}\left(\frac{1}{C_m^{(\psi,\rho)}}C_m^{(\psi,\rho)}g(u)\right) = g(u) = f(u).
$$

Moreover, if $k \geqslant 1$, we use integration by parts to conclude for every $\alpha \in \mathbb{N}_{0,k}^m$ and $h \in C_c^\infty(U)$ that

$$
\int_U \partial_\alpha\mathbb{E}[R_n(u)]h(u)\,du = (-1)^{|\alpha|}\int_U \mathbb{E}[R_n(u)](u)\partial_\alpha h(u)\,du = (-1)^{|\alpha|}\int_U f(u)\partial_\alpha h(u)\,du
$$

$$
= \int_U \partial_\alpha f(u)h(u)\,du.
$$

This shows for every $\alpha \in \mathbb{N}_{0,k}^m$ and a.e. $u \in U$ that $\partial_\alpha\mathbb{E}[R_n](u) = \partial_\alpha\mathbb{E}[R_n(u)] = \partial_\alpha f(u)$, which implies that $f = \mathbb{E}[R_n] \in W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)$.

Finally, we use that $f = \mathbb{E}[R_n] \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$, the right-hand side of [36, Lemma 6.3] for the independent mean-zero random variables $\left(\mathbb{E}[R_n] - R_n\right)_{n=1,\ldots,N}$ (with i.i.d. $(\epsilon_n)_{n=1,\ldots,N}$ satisfying $\mathbb{P}[\epsilon_n = \pm 1] = 1/2$ being independent of $\left(\mathbb{E}[R_n] - R_n\right)_{n=1,\ldots,N}$), the Kahane-Khintchine inequality in [28, Theorem 3.2.23] with constant $\kappa_{2,\min(2,p)} > 0$ depending only on $p \in [1,\infty)$, that $(W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d), \|\cdot\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)})$ is by Lemma 4.9 a Banach space of type $\min(2,p) \in (1,2]$ (with constant $\widetilde{C}_p := C_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)} > 0$ depending only on $p \in (1,2]$), that $(R_n)_{n=1,\ldots,N} \sim R_1$ are identically distributed, and Jensen's inequality, we obtain that

$$
\begin{aligned}
\mathbb{E}\left[\left\|f - \frac{1}{N}\sum_{n=1}^N R_n\right\|^2_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{2}} &= \frac{1}{N}\mathbb{E}\left[\left\|\sum_{n=1}^N \left(\mathbb{E}[R_n] - R_n\right)\right\|^2_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{2}} \\
&\leqslant \frac{2}{N}\mathbb{E}\left[\left\|\sum_{n=1}^N \epsilon_n \left(\mathbb{E}[R_n] - R_n\right)\right\|^2_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{2}} \\
&\leqslant \frac{2\kappa_{2,\min(2,p)}}{N}\mathbb{E}\left[\left\|\sum_{n=1}^N \epsilon_n \left(\mathbb{E}[R_n] - R_n\right)\right\|^{\min(2,p)}_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{\min(2,p)}} \\
&\leqslant \frac{2\widetilde{C}_p\kappa_{2,\min(2,p)}}{N}\left(\sum_{n=1}^N \mathbb{E}\left[\|\mathbb{E}[R_n] - R_n\|^{\min(2,p)}_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]\right)^{\frac{1}{\min(2,p)}} \\
&= \frac{2\widetilde{C}_p\kappa_{2,\min(2,p)}}{N^{1-\frac{1}{\min(2,p)}}}\mathbb{E}\left[\|\mathbb{E}[R_1] - R_1\|^{\min(2,p)}_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{\min(2,p)}} \\
&\leqslant \frac{2\widetilde{C}_p\kappa_{2,\min(2,p)}}{N^{1-\frac{1}{\min(2,p)}}}\mathbb{E}\left[\|\mathbb{E}[R_1] - R_1\|^2_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{2}}.
\end{aligned}
$$

Hence, by using this, Jensen's inequality, Minkowski's inequality together with [28, Proposition 1.2.2], the inequality (40), and the constant $C_p := 4\widetilde{C}_p\kappa_{2,\min(2,p)}\pi > 0$ (depending only on $p \in [1,\infty)$), it follows that

$$
\begin{aligned}
\mathbb{E}\left[\left\|f - \frac{1}{N}\sum_{n=1}^N R_n\right\|^2_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{2}} &\leqslant \frac{2\widetilde{C}_p\kappa_{2,\min(2,p)}}{N^{1-\frac{1}{\min(2,p)}}}\mathbb{E}\left[\|\mathbb{E}[R_1] - R_1\|^2_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{2}} \\
&\leqslant \frac{4\widetilde{C}_p\kappa_{2,\min(2,p)}}{N^{1-\frac{1}{\min(2,p)}}}\|R_1\|_{L^2(\Omega,\mathcal{F},\mathbb{P};W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d))} \\
&\leqslant \frac{4\widetilde{C}_p\kappa_{2,\min(2,p)}}{N^{1-\frac{1}{\min(2,p)}}}2^{4+\frac{1}{p}}\pi\|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})}\frac{C^{(\gamma,p)}_{U,w}m^{\frac{k}{p}}\pi^{\frac{m+1}{4}}}{\left|C^{(\psi,\rho)}_m\right|\Gamma\left(\frac{m+1}{2}\right)^{\frac{1}{2}}}\|f\|_{\mathbb{B}^{k,\gamma}_\psi(U;\mathbb{R}^d)} \\
&\leqslant C_p\|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})}\frac{C^{(\gamma,p)}_{U,w}m^{\frac{k}{p}}\pi^{\frac{m+1}{4}}}{\left|C^{(\psi,\rho)}_m\right|\Gamma\left(\frac{m+1}{2}\right)^{\frac{1}{2}}}\frac{\|f\|_{\mathbb{B}^{k,\gamma}_\psi(U;\mathbb{R}^d)}}{N^{1-\frac{1}{\min(2,p)}}}.
\end{aligned}
$$

Thus, there exists some $\omega \in \Omega$ such that $\varphi_N := \frac{1}{N}\sum_{n=1}^N R_n(\omega) \in \mathcal{NN}^\rho_{U,d}$ satisfies

$$
\begin{aligned}
\|f - \varphi_N\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)} &= \|f - \Phi_N(\omega)\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)} \\
&\leqslant \mathbb{E}\left[\|f - \Phi_N\|^2_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)}\right]^{\frac{1}{2}} \\
&\leqslant C_p\|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})}\frac{C^{(\gamma,p)}_{U,w}m^{\frac{k}{p}}\pi^{\frac{m}{4}}}{\left|C^{(\psi,\rho)}_m\right|\Gamma\left(\frac{m+1}{2}\right)^{\frac{1}{2}}}\frac{\|f\|_{\mathbb{B}^{k,\gamma}_\psi(U;\mathbb{R}^d)}}{N^{1-\frac{1}{\min(2,p)}}},
\end{aligned}
$$

which completes the proof. $\qquad\square$

4.3.5. *Proof of Example 3.3, Lemma 3.9, and Proposition 3.8.*

Proof of Example 3.3. First, we observe that $\rho \in C_{pol,\gamma}^k(\mathbb{R})$ is in each case (a)-(d) of polynomial growth, which ensures that $\rho \in C_{pol,\gamma}^k(\mathbb{R})$ induces $\big(g \mapsto T_\rho(g) := \int_{\mathbb{R}} \rho(s)g(s)ds\big) \in \mathcal{S}'(\mathbb{R};\mathbb{C})$ (see [19, p. 332]).

For (b), we recall that $\tanh'(\xi) = \cosh(\xi)^{-2}$ holds true for all $\xi \in \mathbb{R}$. Moreover, the Fourier transform of the function $\big(s \mapsto h(s) := \frac{\pi s}{\sinh(\pi s/2)}\big) \in L^1(\mathbb{R}, \mathcal{L}(\mathbb{R}), du)$ is for every $\xi \in \mathbb{R}$ given by

$$\widehat{h}(\xi) = \frac{2\pi}{\cosh(\xi)^2} = 2\pi \tanh'(\xi). \tag{41}$$

Then, by using $\big(g \mapsto \big(\mathrm{id} \cdot \widehat{T_{\tanh}}\big)(g) := \widehat{T_{\tanh}}(\mathrm{id} \cdot g)\big) \in \mathcal{S}'(\mathbb{R};\mathbb{C})$, [19, Equation 9.31] with $\mathbb{R} \ni s \mapsto \mathrm{id}(s) := s \in \mathbb{R}$, the definition of $\widehat{T_{\tanh}} \in \mathcal{S}'(\mathbb{R};\mathbb{C})$, the identity (41), and the Plancherel theorem in [19, p. 222], it follows for every $g \in C_c^\infty(\mathbb{R}\backslash\{0\};\mathbb{C})$ that

$$\begin{aligned}
\widehat{T_{\tanh}}(\mathrm{id} \cdot g) &= \big(\mathrm{id} \cdot \widehat{T_{\tanh}}\big)(g) = \frac{1}{i}\widehat{T_{\tanh'}}(g) = (-\mathbf{i})T_{\tanh'}(\widehat{g}) \\
&= (-\mathbf{i}) \int_{\mathbb{R}} \tanh'(\xi)\widehat{g}(\xi)d\xi = \frac{-\mathbf{i}}{2\pi} \int_{\mathbb{R}} \overline{\widehat{h}(\xi)}\widehat{g}(\xi)d\xi \\
&= (-\mathbf{i}) \int_{\mathbb{R}} \overline{h(\xi)}g(\xi)d\xi = \int_{\mathbb{R}} \frac{-\mathbf{i}\pi}{\sinh(\pi\xi/2)}(\mathrm{id} \cdot g)(\xi)d\xi.
\end{aligned} \tag{42}$$

Hence, $\widehat{T_{\tanh}} \in \mathcal{S}'(\mathbb{R};\mathbb{C})$ coincides on $\mathbb{R}\backslash\{0\}$ with $\big(\xi \mapsto f_{\widehat{T_{\tanh}}}(\xi) := \frac{\mathbf{i}\pi}{\sinh(\pi\xi/2)}\big) \in L_{loc}^1(\mathbb{R}\backslash\{0\};\mathbb{C})$.

For (a), we denote by $\big(s \mapsto \sigma(s) := \frac{1}{1+\exp(-s)}\big) \in C_{pol,\gamma}^k(\mathbb{R})$ the sigmoid function and observe that $\sigma(s) = \frac{1}{2}\big(\tanh\big(\frac{s}{2}\big) + 1\big)$ for all $s \in \mathbb{R}$. Then, by using the linearity of the Fourier transform on $\mathcal{S}'(\mathbb{R};\mathbb{C})$, [19, Equation 9.30], that $\widehat{T_1}(g) = 2\pi\delta(g) := 2\pi g(0)$ for any $g \in \mathcal{S}(\mathbb{R};\mathbb{C})$ (see [19, Equation 9.35]), the identity (42), and the substitution $\xi \mapsto \widetilde{\xi}/2$, it follows for every $g \in C_c^\infty(\mathbb{R}\backslash\{0\};\mathbb{C})$ that

$$\begin{aligned}
\widehat{T_\sigma}(g) &= \frac{1}{2}\widehat{T_{\tanh(\frac{\cdot}{2})}}(g) + \frac{1}{2}\widehat{T_1}(g) = \frac{1}{2}\widehat{T_{\tanh}}\Big(g\Big(\frac{\cdot}{2}\Big)\Big) + \frac{2\pi}{2}g(0) \\
&= \frac{1}{2}\int_{\mathbb{R}} \frac{-\mathbf{i}\pi}{\sinh\big(\pi\widetilde{\xi}/2\big)}g\big(\widetilde{\xi}/2\big)d\widetilde{\xi} = \int_{\mathbb{R}} \frac{-\mathbf{i}\pi}{\sinh(\pi\xi)}g(\xi)d\xi.
\end{aligned} \tag{43}$$

Hence, $\widehat{T_\sigma} \in \mathcal{S}'(\mathbb{R};\mathbb{C})$ coincides on $\mathbb{R}\backslash\{0\}$ with $\big(\xi \mapsto f_{\widehat{T_\sigma}}(\xi) := \frac{-\mathbf{i}\pi}{\sinh(\pi\xi)}\big) \in L_{loc}^1(\mathbb{R}\backslash\{0\};\mathbb{C})$.

For (c), we denote by $\big(s \mapsto \sigma^{(-1)}(s) := \ln(1+\exp(s))\big) \in C_{pol,\gamma}^k(\mathbb{R})$ the softplus function and observe that $\frac{d}{ds}\sigma^{(-1)}(s) = \sigma(s)$ for all $s \in \mathbb{R}$. Then, by using [19, Equation 9.31] with $\mathbb{R} \ni s \mapsto \mathrm{id}(s) := s \in \mathbb{R}$ and the identity (43), it follows for every $g \in C_c^\infty(\mathbb{R}\backslash\{0\};\mathbb{C})$ that

$$\widehat{T_{\sigma^{(-1)}}}(\mathrm{id} \cdot g) = \big(\mathrm{id} \cdot \widehat{T_{\sigma^{(-1)}}}\big)(g) = \frac{1}{\mathbf{i}}\widehat{T_\sigma}(g) = \frac{1}{\mathbf{i}} \int_{\mathbb{R}} \frac{-\mathbf{i}\pi}{\sinh(\pi\xi)}g(\xi)d\xi = \int_{\mathbb{R}} \frac{-\pi}{\xi\sinh(\pi\xi)}(\mathrm{id} \cdot g)(\xi)d\xi.$$

Hence, $\widehat{T_{\sigma^{(-1)}}} \in \mathcal{S}'(\mathbb{R};\mathbb{C})$ coincides on $\mathbb{R}\backslash\{0\}$ with $\big(\xi \mapsto f_{\widehat{T_{\sigma^{(-1)}}}}(\xi) := \frac{-\pi}{\xi\sinh(\pi\xi)}\big) \in L_{loc}^1(\mathbb{R}\backslash\{0\};\mathbb{C})$.

For (d), we denote by $\big(s \mapsto \mathrm{ReLU}(s) := \max(s,0)\big) \in C_{pol,\gamma}^k(\mathbb{R})$ the ReLU function and observe that $\mathrm{ReLU}(s) = \max(s,0) = \frac{s+|s|}{2}$ for all $s \in \mathbb{R}$. Moreover, the absolute value $\mathbb{R} \ni s \mapsto |s| \in \mathbb{R}$ is weakly differentiable with $\frac{d}{ds}|s| = \mathrm{sgn}(s)$ for all $s \in \mathbb{R}$, where $\mathrm{sgn}(s) := 1$ if $s > 0$, $\mathrm{sgn}(0) := 0$, and $\mathrm{sgn}(s) := -1$ if $s < 0$. Then, by using the linearity of the Fourier transform on $\mathcal{S}'(\mathbb{R};\mathbb{C})$, that $\widehat{T_{\mathrm{id}}}(g) = 2\pi\mathbf{i}\delta'(g) := 2\pi\mathbf{i}g'(0)$ for any $g \in \mathcal{S}(\mathbb{R};\mathbb{C})$ with $\mathbb{R} \ni s \mapsto \mathrm{id}(s) := s \in \mathbb{R}$ (see [19, Equation 9.35]), [19, Equation 9.31], and [19, Example 9.4.4], i.e. that $\widehat{T_{\mathrm{sgn}}}(g) = -2\mathbf{i}\int_{\mathbb{R}} \frac{g(\xi)}{\xi}d\xi$ for any $g \in C_c^\infty(\mathbb{R}\backslash\{0\};\mathbb{C})$, it follows for every $g \in C_c^\infty(\mathbb{R}\backslash\{0\};\mathbb{C})$ that

$$\begin{aligned}
\widehat{T_{\mathrm{ReLU}}}(\mathrm{id} \cdot g) &= \frac{1}{2}\widehat{T_{\mathrm{id}}}(\mathrm{id} \cdot g) + \frac{1}{2}\widehat{T_{|\cdot|}}(\mathrm{id} \cdot g) = \frac{2\pi\mathbf{i}}{2}(\mathrm{id} \cdot g)'(0) + \frac{1}{2}\big(\mathrm{id} \cdot \widehat{T_{|\cdot|}}\big)(g) \\
&= \frac{1}{2\mathbf{i}}\widehat{T_{\mathrm{sgn}}}(g) = \frac{-2\mathbf{i}}{2\mathbf{i}} \int_{\mathbb{R}} \frac{g(\xi)}{\xi}d\xi = \int_{\mathbb{R}} \frac{-1}{\xi^2}(\mathrm{id} \cdot g)(\xi)d\xi.
\end{aligned}$$

Hence, $\widehat{T_{\mathrm{ReLU}}} \in \mathcal{S}'(\mathbb{R};\mathbb{C})$ coincides on $\mathbb{R}\backslash\{0\}$ with $\big(\xi \mapsto f_{\widehat{T_{\mathrm{ReLU}}}}(\xi) := -\frac{1}{\xi^2}\big) \in L_{loc}^1(\mathbb{R}\backslash\{0\};\mathbb{C})$.

Now, for each case (a)-(d), we fix some $m \in \mathbb{N}$ and $\psi \in \mathcal{S}_0(\mathbb{R}; \mathbb{C})$ with non-negative $\widehat{\psi} \in C_c^\infty(\mathbb{R})$ such that $\operatorname{supp}(\widehat{\psi}) = [\zeta_1, \zeta_2]$ for some $0 < \zeta_1 < \zeta_2 < \infty$. Then, by using that $\widehat{\psi} \in C_c^\infty(\mathbb{R})$ is non-negative, it follows that

$$C_m^{(\psi,\rho)} = (2\pi)^{m-1} \int_{\mathbb{R}\backslash\{0\}} \frac{\overline{\widehat{\psi}(\xi)} f_{\widehat{T_\rho}}(\xi)}{|\xi|^m} d\xi = (2\pi)^{m-1} \int_{\zeta_1}^{\zeta_2} \frac{\overline{\widehat{\psi}(\xi)} f_{\widehat{T_\rho}}(\xi)}{|\xi|^m} d\xi \neq 0.$$

This shows that $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C_{pol,\gamma}^k(\mathbb{R})$ is m-admissible. Moreover, in each case (a)-(d), we define the constant $C_{\psi,\rho} := (2\pi)^{-1} \big| \int_{\zeta_1}^{\zeta_2} \overline{\widehat{\psi}(\xi)} f_{\widehat{T_\rho}}(\xi) d\xi \big|$ (independent of $m \in \mathbb{N}$) to conclude that

$$\left| C_m^{(\psi,\rho)} \right| = (2\pi)^{m-1} \left| \int_{\zeta_1}^{\zeta_2} \frac{\overline{\widehat{\psi}(\xi)} f_{\widehat{T_\rho}}(\xi)}{|\xi|^m} d\xi \right| \geqslant \left| \int_{\zeta_1}^{\zeta_2} \frac{\overline{\widehat{\psi}(\xi)} f_{\widehat{T_\rho}}(\xi)}{2\pi} d\xi \right| \left(\frac{2\pi}{\zeta_2} \right)^m = C_{\psi,\rho} \left(\frac{2\pi}{\zeta_2} \right)^m,$$

which completes the proof. $\qquad\square$

Proof of Lemma 3.9. Let $U \ni u \mapsto w(u) := \prod_{l=1}^m w_0(u_l) \in [0, \infty)$ be a weight, where $w_0 : \mathbb{R} \to [0, \infty)$ satisfies $\int_{\mathbb{R}} w_0(s) ds = 1$ and $C_{\mathbb{R},w_0}^{(\gamma,p)} := \big(\int_{\mathbb{R}} (1+|s|)^{\gamma p} w_0(s) ds \big)^{1/p} < \infty$. Then, by using that $1 + \|u\| \leqslant 1 + \sum_{l=1}^m |u_l| \leqslant \sum_{l=1}^m (1 + |u_l|)$ for any $u := (u_1, ..., u_m)^\top \in \mathbb{R}^m$, that $(x_1 + ... + x_m)^{\gamma p} \leqslant m^{\gamma p} (x_1^{\gamma p} + ... + x_m^{\gamma p})$ for any $x_1, ..., x_m \geqslant 0$, and Fubini's theorem, it follows that

$$\begin{aligned}
C_{U,w}^{(\gamma,p)} &= \left(\int_U (1 + \|u\|)^{\gamma p} w(u) du \right)^{\frac{1}{p}} \\
&\leqslant \left(\int_U \left(\sum_{l=1}^m (1 + |u_l|) \right)^{\gamma p} w(u) du \right)^{\frac{1}{p}} \\
&\leqslant m^\gamma \left(\sum_{l=1}^m \int_{\mathbb{R}^m} (1 + |u_l|)^{\gamma p} \prod_{i=1}^m w_0(u_i) du \right)^{\frac{1}{p}} \\
&\leqslant m^\gamma \left(\sum_{l=1}^m \left(\underbrace{\int_{\mathbb{R}} (1 + |u_l|)^{\gamma p} w_0(u_l) du_l}_{= \left(C_{\mathbb{R},w_0}^{(\gamma,p)} \right)^p} \right) \prod_{\substack{i=1 \\ i \neq l}}^m \underbrace{\int_{\mathbb{R}^m} w_0(u_i) du_i}_{= 1} \right)^{\frac{1}{p}} \\
&\leqslant C_{\mathbb{R},w_0}^{(\gamma,p)} m^{\gamma + \frac{1}{p}},
\end{aligned}$$

which completes the proof. $\qquad\square$

Proof of Proposition 3.8. Fix some $f \in L^1(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m), du; \mathbb{R}^d)$ with $(\lceil\gamma\rceil + 2)$-times differentiable Fourier transform. Then, for any fixed $c \in \{0, \lceil\gamma\rceil + 2\}$, we use that $(\mathfrak{R}_\psi f)(a, b) = (\widetilde{\mathfrak{R}}_\psi f)(v, s, t)$ for any $(a, b) \in (\mathbb{R}^m \backslash \{0\}) \times \mathbb{R}$ with $(v, s, t) := \big(\frac{a}{\|a\|}, \frac{1}{\|a\|}, \frac{b}{\|a\|} \big)$, where $\widetilde{\mathfrak{R}}_\psi f$ is introduced in (27), the identities [51, Equation (36)-(40)], c-times integration by parts, and the Leibniz product rule together with the

chain rule, to conclude for every $(a,b) \in (\mathbb{R}^m \backslash \{0\}) \times \mathbb{R}$ that

$$
\begin{aligned}
b^c (\mathfrak{R}_\psi f)(a,b) &= \frac{t^c}{s^c}(\widetilde{\mathfrak{R}}_\psi f)(v,s,t) \\
&= \frac{1}{2\pi} \frac{t^c}{s^c} \int_{\mathbb{R}} \widehat{f}(\xi v) \overline{\widehat{\psi}(\xi s)} e^{\mathbf{i}\xi t} d\xi \\
&= \frac{1}{2\pi} \frac{(-\mathbf{i})^c}{s^c} \int_{\mathbb{R}} \widehat{f}(\xi v) \overline{\widehat{\psi}(\xi s)} \frac{\partial^c}{\partial \xi^c} \left(e^{\mathbf{i}\xi t} \right) d\xi \\
&= \frac{1}{2\pi} \frac{\mathbf{i}^c}{s^c} \int_{\mathbb{R}} \frac{\partial^c}{\partial \xi^c} \left(\widehat{f}(\xi v) \overline{\widehat{\psi}(\xi s)} \right) e^{\mathbf{i}\xi t} d\xi \\
&= \frac{1}{2\pi} \frac{\mathbf{i}^c}{s^c} \sum_{\beta \in \mathbb{N}_{0,c}^m} \frac{c!}{|\beta|!(c-|\beta|)!} \int_{\mathbb{R}} v^\beta \partial_\beta \widehat{f}(\xi v) \overline{\widehat{\psi}^{(c-|\beta|)}(\xi s)} s^{c-|\beta|} e^{\mathbf{i}\xi t} d\xi \\
&= \frac{1}{2\pi} \mathbf{i}^c \sum_{\beta \in \mathbb{N}_{0,c}^m} \frac{c!}{|\beta|!(c-|\beta|)!} \int_{\mathbb{R}} \left(\frac{v}{s} \right)^\beta \partial_\beta \widehat{f}(\xi v) \widehat{\psi}^{(c-|\beta|)}(\xi s) e^{\mathbf{i}\xi t} d\xi.
\end{aligned}
\tag{44}
$$

Therefore, by taking the norm in (44) and by using the substitution $\zeta \mapsto \xi s$ as well as the inequality $\left| (v/s)^\beta \right| := \left| \prod_{l=1}^m (v_l/s)^{\beta_l} \right| = \prod_{l=1}^\beta |v_l/s|^{\beta_l} \leqslant \left(1 + \|v/s\|^2 \right)^{|\beta|/2} \leqslant \left(1 + 1/s^2 \right)^{c/2}$ for any $v \in \mathbb{S}^{m-1}$, $s \in (0,\infty)$, and $\beta \in \mathbb{N}_{0,c}^m$, we obtain for every $(a,b) \in (\mathbb{R}^m \backslash \{0\}) \times \mathbb{R}$ that

$$
\begin{aligned}
|b|^c \|(\mathfrak{R}_\psi f)(a,b)\| &\leqslant \frac{1}{2\pi} \sum_{\beta \in \mathbb{N}_{0,c}^m} \frac{c!}{|\beta|!(c-|\beta|)!} \int_{\mathbb{R}} \left| \left(\frac{v}{s} \right)^\beta \right| \|\partial_\beta \widehat{f}(\xi v)\| \left| \widehat{\psi}^{(c-|\beta|)}(\xi s) \right| d\xi \\
&= \frac{1}{2\pi} \sum_{\beta \in \mathbb{N}_{0,c}^m} \frac{c!}{|\beta|!(c-|\beta|)!} \int_{\mathbb{R}} \left| \left(\frac{v}{s} \right)^\beta \right| \left\| \partial_\beta \widehat{f}\left(\frac{\zeta v}{s} \right) \right\| \left| \widehat{\psi}^{(c-|\beta|)}(\zeta) \right| \frac{1}{s} d\zeta \\
&\leqslant \frac{c!}{2\pi} \left(1 + \frac{1}{s^2} \right)^{\frac{c}{2}} \frac{1}{s} \sum_{\beta \in \mathbb{N}_{0,c}^m} \int_{\mathbb{R}} \left\| \partial_\beta \widehat{f}\left(\frac{\zeta v}{s} \right) \right\| \left| \widehat{\psi}^{(c-|\beta|)}(\zeta) \right| d\zeta \\
&\leqslant \frac{(\lceil \gamma \rceil + 2)!}{2\pi} \left(1 + \|a\|^2 \right)^{\frac{\lceil \gamma \rceil + 2}{2}} \sum_{\beta \in \mathbb{N}_{0,\lceil\gamma\rceil+2}^m} \int_{\mathbb{R}} \|\partial_\beta \widehat{f}(\zeta a)\| \left| \widehat{\psi}^{(\lceil\gamma\rceil+2-|\beta|)}(\zeta) \right| d\zeta.
\end{aligned}
\tag{45}
$$

Hence, by using the inequality $(x+y)^s \leqslant 2^{s-1}(x^s + y^s)$ for any $x,y \in [0,\infty)$ and $s \in [1,\infty)$ and the inequality (45), it follows for every $(a,b) \in (\mathbb{R}^m \backslash \{0\}) \times \mathbb{R}$ that

$$
\begin{aligned}
\left(1 + |b|^2 \right)^{\frac{\lceil\gamma\rceil+2}{2}} \|(\mathfrak{R}_\psi f)(a,b)\| &\leqslant 2^{\frac{\lceil\gamma\rceil}{2}} \left(\|(\mathfrak{R}_\psi f)(a,b)\| + |b|^{\lceil\gamma\rceil+2} \|(\mathfrak{R}_\psi f)(a,b)\| \right) \\
&\leqslant 2^{\frac{\lceil\gamma\rceil}{2}} \frac{(\lceil\gamma\rceil+2)!}{\pi} \left(1 + \|a\|^2 \right)^{\frac{\lceil\gamma\rceil+2}{2}} \sum_{\beta \in \mathbb{N}_{0,\lceil\gamma\rceil+2}^m} \int_{\mathbb{R}} \|\partial_\beta \widehat{f}(\zeta a)\| \left| \widehat{\psi}^{(\lceil\gamma\rceil+2-|\beta|)}(\zeta) \right| d\zeta.
\end{aligned}
\tag{46}
$$

Moreover, by using Fubini's theorem and that $(\mathfrak{R}_\psi f)(a,b) = 0$ for any $(a,b) \in \{0\} \times \mathbb{R}$, we have

$$
\begin{aligned}
\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)} &\leqslant \left(\int_{\mathbb{R}^m} \int_{\mathbb{R}} \left(1 + \|a\|^2 \right)^{\gamma+k+\frac{m+1}{2}} \left(1 + |b|^2 \right)^{\gamma+1} \|(\mathfrak{R}_\psi f)(a,b)\|^2 db\, da \right)^{\frac{1}{2}} \\
&\leqslant \left(\int_{\mathbb{R}} \int_{\mathbb{R}^m} \left(1 + \|a\|^2 \right)^{\gamma+k+\frac{m+1}{2}} \left(1 + |b|^2 \right)^{\gamma+2} \|(\mathfrak{R}_\psi f)(a,b)\|^2 da\, \frac{1}{1+|b|^2} db \right)^{\frac{1}{2}} \\
&\leqslant \left(\sup_{b \in \mathbb{R}} \int_{\mathbb{R}^m \backslash \{0\}} \left(\left(1 + \|a\|^2 \right)^{\frac{\lceil\gamma\rceil+k+\frac{m+1}{2}}{2}} \left(1 + |b|^2 \right)^{\frac{\lceil\gamma\rceil+2}{2}} \|(\mathfrak{R}_\psi f)(a,b)\| \right)^2 da \right)^{\frac{1}{2}} \left(\underbrace{\int_{\mathbb{R}} \frac{1}{1+|b|^2} db}_{=\pi} \right)^{\frac{1}{2}}.
\end{aligned}
\tag{47}
$$

Thus, by inserting the inequality (46) into the right-hand side of (47), using Minkowski's integral inequality (with measure spaces $(\mathbb{R}^m \backslash \{0\}, \mathcal{L}(\mathbb{R}^m \backslash \{0\}), da)$ and $(\mathbb{N}_{0,k}^m \times \mathbb{R}, \mathcal{P}(\mathbb{N}_{0,k}^m) \otimes \mathcal{B}(\mathbb{R}), \mu \otimes d\zeta)$, where $\mathcal{P}(\mathbb{N}_{0,k}^m)$ denotes the power set of $\mathbb{N}_{0,k}^m$, and where $\mathcal{P}(\mathbb{N}_{0,k}^m) \ni E \mapsto \mu(E) := \sum_{\alpha \in \mathbb{N}_{0,k}^m} \mathbb{1}_E(\alpha) \in [0,\infty)$ is the counting measure), the substitution $\xi \mapsto \zeta a$ with Jacobi determinant $d\xi = |\zeta|^m da$, that $\zeta_1 := \inf \big\{ |\zeta| :$

$\zeta \in \mathrm{supp}(\widehat{\psi})\} > 0$, and the constant $C_1 := 2^{\lceil\gamma\rceil/2}\pi^{-1/2}(\lceil\gamma\rceil + 2)!\max_{j=0,\dots,\lceil\gamma\rceil+2}\int_{\mathbb{R}}|\widehat{\psi}^{(j)}(\zeta)|d\zeta > 0$, we conclude that

$\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)}$

$$\leqslant 2^{\frac{\lceil\gamma\rceil}{2}}\frac{(\lceil\gamma\rceil + 2)!}{\pi}\sqrt{\pi}\left(\int_{\mathbb{R}^m}\left((1 + \|a\|^2)^{\frac{2\lceil\gamma\rceil+k+\frac{m+5}{2}}{2}}\sum_{\beta\in\mathbb{N}_{0,\lceil\gamma\rceil+2}^m}\int_{\mathbb{R}}\left\|\partial_\beta\widehat{f}(\zeta a)\right\|\left|\widehat{\psi}^{(\lceil\gamma\rceil+2-|\beta|)}(\zeta)\right|d\zeta\right)^2 da\right)^{\frac{1}{2}}$$

$$\leqslant 2^{\frac{\lceil\gamma\rceil}{2}}\frac{(\lceil\gamma\rceil + 2)!}{\sqrt{\pi}}\sum_{\beta\in\mathbb{N}_{0,\lceil\gamma\rceil+2}^m}\int_{\mathbb{R}}\left|\widehat{\psi}^{(\lceil\gamma\rceil+2-|\beta|)}(\zeta)\right|\left(\int_{\mathbb{R}^m}\left\|\partial_\beta\widehat{f}(\zeta a)\right\|^2(1 + \|a\|^2)^{2\lceil\gamma\rceil+k+\frac{m+5}{2}}da\right)^{\frac{1}{2}}d\zeta$$

$$\leqslant 2^{\frac{\lceil\gamma\rceil}{2}}\frac{(\lceil\gamma\rceil + 2)!}{\sqrt{\pi}}\sum_{\beta\in\mathbb{N}_{0,\lceil\gamma\rceil+2}^m}\int_{\mathrm{supp}(\widehat{\psi})}\frac{\left|\widehat{\psi}^{(\lceil\gamma\rceil+2-|\beta|)}(\zeta)\right|}{\zeta^{\frac{m}{2}}}\left(\int_{\mathbb{R}^m}\left\|\partial_\beta\widehat{f}(\xi)\right\|^2(1 + \|\xi/\zeta\|^2)^{\frac{4\lceil\gamma\rceil+2k+m+5}{2}}d\xi\right)^{\frac{1}{2}}d\zeta$$

$$\leqslant \frac{C_1}{\zeta_1^{\frac{m}{2}}}\sum_{\beta\in\mathbb{N}_{0,\lceil\gamma\rceil+2}^m}\left(\int_{\mathbb{R}^m}\left\|\partial_\beta\widehat{f}(\xi)\right\|^2(1 + \|\xi/\zeta_1\|^2)^{\frac{4\lceil\gamma\rceil+2k+m+5}{2}}d\xi\right)^{\frac{1}{2}},$$

which completes the proof. $\qquad\square$

Proof of Proposition 3.10. Fix some $m, d \in \mathbb{N}$ and $\varepsilon > 0$. Moreover, let $p \in (1, \infty)$ and $w : U \to [0, \infty)$ be a weight as in Lemma 3.9 (with constant $C_{\mathbb{R},w_0}^{(\gamma,p)} > 0$ being independent of $m, d \in \mathbb{N}$ and $\varepsilon > 0$), let $(\psi, \rho) \in \mathcal{S}_0(\mathbb{R}; \mathbb{C}) \times C_{pol,\gamma}^k(\mathbb{R})$ be a pair as in Example 3.3 (with $0 < \zeta_1 < \zeta_2 < \infty$ and constant $C_{\psi,\rho} > 0$ being independent of $m, d \in \mathbb{N}$ and $\varepsilon > 0$), and fix some $f \in W^{k,p}(U, \mathcal{L}(U), w; \mathbb{R}^d)$ satisfying the conditions of Proposition 3.8 such that the right-hand side of (9) satisfies $\mathcal{O}\left(m^s(2/\zeta_2)^m(m + 1)^{m/2}\right)$ for some $s \in \mathbb{N}_0$. Then, there exists some constant $C > 0$ (being independent of $m, d \in \mathbb{N}$ and $\varepsilon > 0$) such that for every $m, d \in \mathbb{N}$ it holds that

$$\frac{C_1}{\zeta_1^{\frac{m}{2}}}\sum_{\beta\in\mathbb{N}_{0,\lceil\gamma\rceil+2}^m}\left(\int_{\mathbb{R}^m}\left\|\partial_\beta\widehat{f}(\xi)\right\|^2(1 + \|\xi/\zeta_1\|^2)^{2\lceil\gamma\rceil+k+\frac{m+5}{2}}d\xi\right)^{\frac{1}{2}} \leqslant Cm^s\left(\frac{2}{\zeta_2}\right)^m(m + 1)^{\frac{m}{2}}. \quad (48)$$

Hence, by using Proposition 3.8 together with (48), Lemma 3.9, the inequality in Example 3.3, that $\Gamma(x) \geqslant \sqrt{2\pi/x}(x/e)^x$ for any $x \in (0, \infty)$ (see [21, Lemma 2.4]), and that $\frac{\pi^{m/4}(2/\zeta_2)^m}{(2\pi/\zeta_2)^m(1/(2e))^{m/2}} = \left(\frac{2e\sqrt{\pi}}{\pi^2}\right)^{m/2} \leqslant 1$ for any $m \in \mathbb{N}$, we conclude that there exist some constants $C_2, C_3 > 0$ (being independent of $m, d \in \mathbb{N}$ and $\varepsilon > 0$) such that

$$C_p\|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\frac{C_{U,w}^{(\gamma,p)}m^{\frac{k}{p}}\pi^{\frac{m+1}{4}}}{\left|C_m^{(\psi,\rho)}\right|\Gamma\left(\frac{m+1}{2}\right)^{\frac{1}{2}}}\frac{\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)}}{N^{1-\frac{1}{\min(2,p)}}}$$

$$\leqslant C_p\|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\frac{C_{\mathbb{R},w_0}^{(\gamma,p)}m^{\gamma+\frac{k+1}{p}}\pi^{\frac{m+1}{4}}}{C_{\psi,\rho}\left(\frac{2\pi}{\zeta_2}\right)^m\left(\frac{4\pi}{m+1}\right)^{\frac{1}{4}}\left(\frac{m+1}{2e}\right)^{\frac{m+1}{2}}}Cm^s\left(\frac{2}{\zeta_2}\right)^m(m + 1)^{\frac{m}{2}} \quad (49)$$

$$\leqslant C_p\|\rho\|_{C_{pol,\gamma}^k(\mathbb{R})}\frac{C_{\mathbb{R},w_0}^{(\gamma,p)}m^{\gamma+\frac{k+1}{p}}\pi^{\frac{1}{4}}(2e)^{\frac{1}{2}}C_3}{C_{\psi,\rho}(4\pi)^{\frac{1}{4}}}Cm^s$$

$$\leqslant \left(C_2m^{C_3}\right)^{1-\frac{1}{\min(2,p)}}.$$

Hence, by using that $f \in \mathbb{B}_\psi^{k,\gamma}(U; \mathbb{R}^d)$ (see Proposition 3.8), we can apply Theorem 3.6 with $N = \left\lceil C_2m^{C_3}\varepsilon^{-\frac{\min(2,p)}{\min(2,p)-1}}\right\rceil$ and insert the inequality (49) to obtain a neural network $\varphi \in \mathcal{NN}_{U,d}^\rho$ with N

neurons satisfying

$$\|f - \varphi_N\|_{W^{k,p}(U,\mathcal{L}(U),w;\mathbb{R}^d)} \leqslant C_p \|\rho\|_{C^k_{pol,\gamma}(\mathbb{R})} \frac{C_{U,w}^{(\gamma,p)} m^{\frac{k}{p}} \pi^{\frac{m+1}{4}}}{\left|C_m^{(\psi,\rho)}\right| \Gamma\left(\frac{m+1}{2}\right)^{\frac{1}{2}}} \frac{\|f\|_{\mathbb{B}_\psi^{k,\gamma}(U;\mathbb{R}^d)}}{N^{1-\frac{1}{\min(2,p)}}}$$

$$\leqslant \frac{\left(C_2 m^{C_3}\right)^{1-\frac{1}{\min(2,p)}}}{N^{1-\frac{1}{\min(2,p)}}} \leqslant \varepsilon,$$

which completes the proof. $\qquad\square$

References

[1] Milton Abramowitz and Irene Ann Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Applied mathematics series / National Bureau of Standards 55, Print. 9. Dover, New York, 9th edition, 1970.

[2] Beatrice Acciaio, Anastasis Kratsios, and Gudmund Pammer. Designing universal causal deep learning models: The geometric (hyper)transformer. *Mathematical Finance*, 34(2):671–735, 2024.

[3] Robert A. Adams. *Sobolev Spaces*. Pure and applied mathematics. Academic Press, 1975.

[4] Fernando Albiac and Nigel J. Kalton. *Topics in Banach space theory*. Graduate Texts in Mathematics 233. Springer, New York, 2006.

[5] Andrew R. Barron. Neural net approximation. In *Proc. 7th Yale workshop on adaptive and learning systems*, volume 1, pages 69–72, 1992.

[6] Andrew R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945, 1993.

[7] Vladimir I. Bogachev. *Measure Theory*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1st edition, 2007.

[8] Helmut Bölcskei, Philipp Grohs, Gitta Kutyniok, and Philipp Petersen. Optimal approximation with sparsely connected deep neural networks. *SIAM Journal on Mathematics of Data Science*, 1:8–45, 2019.

[9] Emmanuel J. Candès. *Ridgelets: Theory and Applications*. PhD thesis, Stanford University, 1998.

[10] Tianping Chen and Hong Chen. Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks. *IEEE Transactions on Neural Networks*, 6(4):904–910, 1995.

[11] Christa Cuchiero, Philipp Schmocker, and Josef Teichmann. Global universal approximation of functional input maps on weighted spaces. *arXiv e-prints 2306.03303*, 2023.

[12] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.

[13] Christian Darken, James M. Donahue, Leonid Gurvits, and Eduardo D. Sontag. Rate of approximation results motivated by robust neural network learning. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory*, COLT '93, pages 303–309, New York, NY, USA, 1993. Association for Computing Machinery.

[14] Christian Darken, James M. Donahue, Leonid Gurvits, and Eduardo D. Sontag. Rates of convex approximation in non-Hilbert spaces. *Constructive Approximation*, 13:187–220, 1997.

[15] Joseph L. Doob. *Measure theory*. Graduate texts in mathematics; 143. Springer Science+Business Media, LLC, New York, 1st edition, 1994.

[16] Philipp Dörsek and Josef Teichmann. A semigroup point of view on splitting schemes for stochastic (partial) differential equations. *arXiv e-prints 1011.2651*, 2010.

[17] Weinan E, Chao Ma, and Lei Wu. The Barron space and the flow-induced function spaces for neural network models. *Constructive Approximation*, 55:369–406, 2022.

[18] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate studies in mathematics*. American Mathematical Society, Providence, Rhode Island, 2nd edition, 2010.

[19] Gerald B. Folland. *Fourier analysis and its applications*. Brooks/Cole Publishing Company, Belmont, California, 1st edition, 1992.

[20] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183–192, 1989.

[21] Lukas Gonon, Philipp Grohs, Arnulf Jentzen, David Kofler, and David Šiška. Uniform error estimates for artificial neural network approximations for heat equations. *IMA Journal of Numerical Analysis*, 42(3):1991–2054, 08 2021.

[22] Lyudmila Grigoryeva and Juan-Pablo Ortega. Differentiable reservoir computing. *Journal of Machine Learning Research*, 20(179):1–62, 2019.

[23] Gerd Grubb. *Distributions and operators*. Graduate texts in Mathematics 252. Springer, New York, 2009.

[24] Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, Abdel-Rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[25] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.

[26] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[27] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5):551–560, 1990.

[28] Tuomas Hytönen, Jan van Neerven, Mark Veraar, and Lutz Weis. *Analysis in Banach Spaces*, volume 63 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Cham, 2016.

[29] Patrick Kidger and Terry Lyons. Universal approximation with deep narrow networks. In *Conference on Learning Theory 2020*, 2020.

[30] Jason M. Klusowski and Andrew R. Barron. Risk bounds for high-dimensional ridge function combinations including neural networks. *arXiv e-prints 1607.01434*, 2016.

[31] Jacob Korevaar. Distribution proof of Wiener's Tauberian theorem. *Proceedings of the American Mathematical Society*, 16(3):353–355, 1965.

[32] Anastasis Kratsios and Léonie Papon. Universal approximation theorems for differentiable geometric deep learning. *Journal of Machine Learning Research*, 23(196):1–73, 2022.

[33] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[34] Věra Kůrková. Complexity estimates based on integral transforms induced by computational units. *Neural Networks*, 33:160–167, 2012.

[35] Alois Kufner. *Weighted Sobolev spaces*. Teubner-Texte zur Mathematik Bd. 31. B.G. Teubner, Leipzig, 1980.

[36] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces: Isoperimetry and Processes*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Folge 3 Bd. 23. Springer, Berlin, 1991.

[37] Moshe Leshno, Vladimir Ya. Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

[38] Tong Mao and Ding-Xuan Zhou. Rates of approximation by ReLU shallow neural networks. *Journal of Complexity*, 79:101784, 2023.

[39] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

[40] Hrushikesh N. Mhaskar and Charles A. Micchelli. Approximation by superposition of sigmoidal and radial basis functions. *Advances in Applied Mathematics*, 13(3):350–373, 1992.

[41] Hrushikesh N. Mhaskar and Charles A. Micchelli. Dimension-independent bounds on the degree of approximation by neural networks. *IBM Journal of Research and Development*, 38(3):277–284, 1994.

[42] Hrushikesh N. Mhaskar and Charles A. Micchelli. Degree of approximation by neural and translation networks with a single hidden layer. *Advances in Applied Mathematics*, 16(2):151–183, 1995.

[43] Tom M. Mitchell. *Machine Learning*. McGraw-Hill series in computer science. WCB McGraw-Hill, Boston MA, 1997.

[44] James R. Munkres. *Topology*. Pearson, Harlow, Essex, UK, 2nd, pearson new international edition, 2014.

[45] Allan Pinkus. Approximation theory of the MLP model in neural networks. *Acta Numerica*, 8:143–195, 1999.

[46] Walter Rudin. *Real and complex analysis*. McGraw-Hill series in higher mathematics. WCB/McGraw-Hill, Boston, Massachusetts, 3rd edition, 1987.

[47] Walter Rudin. *Functional analysis*. International series in pure and applied mathematics. McGraw-Hill, Boston, Mass, 2nd edition, 1991.

[48] Jonathan W. Siegel and Jinchao Xu. Approximation rates for neural networks with general activation functions. *Neural Networks*, 128:313–321, 2020.

[49] Jonathan W. Siegel and Jinchao Xu. High-order approximation rates for shallow neural networks with cosine and ReLU^k activation functions. *Applied and Computational Harmonic Analysis*, 58:1–26, 2022.

[50] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489, 2016.

[51] Sho Sonoda and Noboru Murata. Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, 43(2):233–268, 2017.

[52] Alan M. Turing. Computing machinery and intelligence. *Mind*, LIX(236):433–460, 1950.

[53] Teun D. H. van Nuland. Noncompact uniform universal approximation. *Neural Networks*, 173:106181, 2024.

[54] Norbert Wiener. Tauberian theorems. *Annals of Mathematics*, 33(1):1–100, 1932.

[55] Yunfei Yang and Ding-Xuan Zhou. Optimal rates of approximation by shallow ReLU^k neural networks and applications to nonparametric regression. *Constructive Approximation*, 2024.

[56] Shijun Zhang, Jianfeng Lu, and Hongkai Zhao. Deep network approximation: Beyond ReLU to diverse activation functions. *Journal of Machine Learning Research*, 25(35):1–39, 2024.

Nanyang Technological University, Division of Mathematical Sciences, 21 Nanyang Link, Singapore
Email address: `ariel.neufeld@ntu.edu.sg`

Nanyang Technological University, Division of Mathematical Sciences, 21 Nanyang Link, Singapore
Email address: `philippt001@e.ntu.edu.sg`