Modulation Spaces and the Curse of Dimensionality

Rahul Parhi
Biomedical Imaging Group
École polytechnique fédérale de Lausanne
CH-1015 Lausanne, Switzerland
Email: rahul.parhi@epfl.ch

Michael Unser
Biomedical Imaging Group
École polytechnique fédérale de Lausanne
CH-1015 Lausanne, Switzerland
Email: michael.unser@epfl.ch

***Abstract*—We investigate the L^2-error of approximating functions in the modulation spaces $M^s_{1,1}(\mathbb{R}^d)$, $s \geq 0$, by linear combinations of Wilson bases elements. We analyze a nonlinear method for approximating functions in $M^s_{1,1}(\mathbb{R}^d)$ with N-terms from a Wilson basis. Its L^2-approximation error decays at a rate of $N^{-\frac{1}{2}-\frac{s}{2d}}$. We show that this rate is optimal by proving a matching lower bound. Remarkably, these rates do not grow with the input dimension d. Finally, we show that the best linear L^2-approximation error cannot decay faster than $N^{-\frac{s}{2d}}$. This shows that linear methods, contrary to the nonlinear ones, necessarily suffer the curse of dimensionality in these spaces.**

***Index Terms*—Wilson basis, curse of dimensionality, nonlinear approximation, time–frequency analysis.**

I. Introduction

The study of the simultaneous time–frequency content of a function (or distribution) through the short-time Fourier transform (STFT) is the main tenet of time–frequency analysis [17]. The modulation spaces are a family of Banach spaces that characterize the locality and regularity of a function (or distribution) by its decay in the STFT domain. These are, in some sense, the "right" spaces to study time–frequency analysis.

The modulation spaces on $\mathbb{R}^d$ are, in particular, a three-parameter family of smoothness spaces, denoted by $M^s_{p,q}(\mathbb{R}^d)$, where s is the smoothness index and p and q are integrability parameters. The most important modulation space is $M^0_{1,1}(\mathbb{R}^d) = S_0(\mathbb{R}^d)$, which is Feichtinger's Segal algebra [11]. This family of spaces includes essentially all functions and distributions of interest since the Schwartz space $\mathcal{S}(\mathbb{R}^d)$ and the tempered distributions $\mathcal{S}'(\mathbb{R}^d)$ are the projective and inductive limits, respectively, of modulation spaces [17, Proposition 11.3.1(d)]. More precisely, one has that

$$\mathcal{S}(\mathbb{R}^d) = \bigcap_{s\geq 0} M^s_{1,1}(\mathbb{R}^d) \tag{1}$$

and

$$\mathcal{S}'(\mathbb{R}^d) = \bigcup_{s\geq 0} M^{-s}_{\infty,\infty}(\mathbb{R}^d). \tag{2}$$

While there are many works that study properties of the modulation spaces [11], [12], [13], [15] (see also the survey [14] of Feichtinger and references therein), there are few works that study these spaces from an approximation theory perspective. Some notable works on the approximation theory of modulation spaces include [3], [4], [18]. In this paper, we provide new results regarding nonlinear approximations in modulation spaces. In Theorems 1 and 3, we derive sharp rates (upper and lower bounds) for nonlinear approximations with Wilson bases [7] for functions in the modulation spaces $M^s_{1,1}(\mathbb{R}^d)$. The rates are $N^{-\frac{1}{2}-\frac{s}{2d}}$, where N is the number of terms in the approximant. Remarkably, these rates do not grow with the input dimension and, therefore, "break" the curse of dimensionality. Furthermore, we prove in Theorem 4 that linear approximation methods necessarily suffer the curse of dimensionality and that the best N-term linear approximation error cannot decay faster than $N^{-\frac{s}{2d}}$.

These findings should be contrasted with usual results in multivariate approximation theory which state that the best N-term approximation error of a function from a d-variate Sobolev or Besov space with smoothness index s decays as $N^{-s/d}$ [9]. Thus the modulation spaces $M^s_{1,1}(\mathbb{R}^d)$ are *mixed-variation* spaces in the sense of Donoho [10].

Mixed-variation spaces have received considerable interest in the approximation theory community as a followup to [2], where Barron showed that neural networks can approximate functions that satisfy certain decay conditions on their Fourier transforms at a rate that does not grow with the input dimension d. The techniques used by Barron to prove this dimension-free result were based on the foundational work of Maurey [29] and Jones [21]. These spaces, now referred to as the spectral Barron spaces, have led to a large body of works that study smoothness spaces that are "immune" to the curse of dimensionality (see [1], [27], [28], [30], [32], and references therein).

II. Modulation Spaces

The Fourier transform of $\varphi \in \mathcal{S}(\mathbb{R}^d)$ is defined as

$$\mathscr{F}\{\varphi\}(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} \varphi(\boldsymbol{x}) e^{-\mathrm{j}2\pi\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x}, \quad \boldsymbol{\xi} \in \mathbb{R}^d, \tag{3}$$

where $\mathrm{j}^2 = -1$. Consequently, the inverse Fourier transform of $\widehat{\varphi} \in \mathcal{S}(\mathbb{R}^d)$ is

$$\mathscr{F}^{-1}\{\widehat{\varphi}\}(\boldsymbol{x}) = \int_{\mathbb{R}^d} \widehat{\varphi}(\boldsymbol{\xi}) e^{\mathrm{j}2\pi\boldsymbol{\xi}^\mathsf{T}\boldsymbol{x}}\,\mathrm{d}\boldsymbol{\xi}, \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{4}$$

The translation and modulation operators acting on $\varphi \in \mathcal{S}(\mathbb{R}^d)$ are given by $\mathrm{T}_{\boldsymbol{x}}\{\varphi\} = \varphi(\cdot - \boldsymbol{x})$ and $\mathrm{M}_{\boldsymbol{\xi}}\{\varphi\} = e^{\mathrm{j}2\pi\boldsymbol{\xi}^\mathsf{T}(\cdot)}\varphi$, respectively. These operators are all extended by duality to act on $\mathcal{S}'(\mathbb{R}^d)$. Finally, the STFT of $f \in \mathcal{S}'(\mathbb{R}^d)$ with respect to the window $g \in \mathcal{S}(\mathbb{R}^d)$ is

$$\mathrm{V}_g\{f\}(\boldsymbol{x}, \boldsymbol{\xi}) = \langle f, \mathrm{M}_{\boldsymbol{\xi}} \mathrm{T}_{\boldsymbol{x}} g \rangle = \mathscr{F}\{f(\cdot)\overline{g(\cdot - \boldsymbol{x})}\}(\boldsymbol{\xi}), \quad (5)$$

where $\overline{g(\cdot - \boldsymbol{x})}$ denotes the complex conjugate of $g(\cdot - \boldsymbol{x})$. Here, the domain of the STFT, sometimes referred to as *phase space* [16], is indexed by $(\boldsymbol{x}, \boldsymbol{\xi}) \in \mathbb{R}^d \times \mathbb{R}^d$. We also note that the STFT of any tempered distribution is necessarily a continuous function [17, Theorem 11.2.3].

Given a fixed nonzero window $g \in \mathcal{S}(\mathbb{R}^d)$, $s \in \mathbb{R}$, and $1 \leq p, q \leq \infty$, the modulation space $M^s_{p,q}(\mathbb{R}^d)$ is the Banach space that consists of all tempered distributions $f \in \mathcal{S}'(\mathbb{R}^d)$ such that the norm $\|f\|_{M^s_{p,q}}$, given by the quantity

$$\left(\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |\mathrm{V}_g\{f\}(\boldsymbol{x}, \boldsymbol{\xi})|^p (1 + \|(\boldsymbol{x}, \boldsymbol{\xi})\|_2)^{sp} \, \mathrm{d}\boldsymbol{x} \right)^{q/p} \mathrm{d}\boldsymbol{\xi} \right)^{1/q}, \quad (6)$$

is finite, with appropriate modifications when p or $q = \infty$. These spaces are independent of the (nonzero) window $g \in \mathcal{S}(\mathbb{R}^d)$, in the sense that different windows result in equivalent norms.

It turns out that these spaces can be studied through the atomic decompositions of their members in Wilson bases. Fix a univariate window $g \in M^s_{1,1}(\mathbb{R})$ such that $\|g\|_{L^2} = 1$, $g(x) = \overline{g(-x)}$, and the system $\{\mathrm{M}_m \mathrm{T}_{k/2} \, g\}_{m,k \in \mathbb{Z}}$ forms a tight Gabor frame of redundancy 2. Then, the system generated by

$$\psi_{k,n} = c_n \mathrm{T}_{k/2} \big(\mathrm{M}_n + (-1)^{k+n} \mathrm{M}_{-n} \big) g, \quad (k, n) \in \mathbb{Z} \times \mathbb{N}_0, \quad (7)$$

with $c_0 = 1$ and $c_n = 1/\sqrt{2}$ if $n \neq 0$, and $\psi_{2k+1,0} \equiv 0$ is referred to as a Wilson basis. The Wilson basis is an orthonormal basis for $L^2(\mathbb{R})$ [17, Theorem 8.5.1]. Wilson bases for $L^2(\mathbb{R}^d)$ can be then be constructed via tensor products. Let $\{\psi_{\boldsymbol{k},\boldsymbol{n}}\}_{\boldsymbol{k} \in \mathbb{Z}^d, \boldsymbol{n} \in \mathbb{N}_0^d}$ denote such a Wilson basis.

Given $s \in \mathbb{R}$ and $1 \leq p, q \leq \infty$, the sequence space $m^s_{p,q}$ is the Banach space that consists of all sequences of complex numbers $c = (c_{\boldsymbol{k},\boldsymbol{n}})_{\boldsymbol{k} \in \mathbb{Z}^d, \boldsymbol{n} \in \mathbb{N}_0^d}$ such that the norm

$$\|c\|_{m^s_{p,q}} := \left(\sum_{\boldsymbol{n} \in \mathbb{N}_0^d} \left(\sum_{\boldsymbol{k} \in \mathbb{Z}^d} (1 + |(\boldsymbol{k}, \boldsymbol{n})|)^{sp} |c_{\boldsymbol{k},\boldsymbol{n}}|^p \right)^{q/p} \right)^{1/q} \quad (8)$$

is finite, with appropriate modifications when p or $q = \infty$. Here, we set the notation $|(\boldsymbol{k}, \boldsymbol{n})| := \max\{|k_1|, \ldots, |k_d|, n_1, \ldots, n_d\}$ for the vector $(\boldsymbol{k}, \boldsymbol{n}) \in \mathbb{Z}^d \times \mathbb{N}_0^d$ for convenience. We could just as easily work with any other norm since all norms are equivalent in finite dimensions. The analysis and synthesis operators of a Wilson basis establish an isomorphism between $M^s_{p,q}(\mathbb{R}^d)$ and $m^s_{p,q}$ (see [17, Chapter 12.3] and [15]).

III. Nonlinear Approximation with Wilson Bases

In this section, we study the problem of approximating functions from the modulation spaces $M^s_{1,1}(\mathbb{R}^d)$, $s \geq 0$, with Wilson bases. Since these spaces continuously embed into $L^2(\mathbb{R}^d)$, we derive upper and lower bounds for the L^2-approximation error rate. Remarkably, the rates are immune to the curse of dimensionality. Given a Wilson basis $\{\psi_{\boldsymbol{k},\boldsymbol{n}}\}_{\boldsymbol{k} \in \mathbb{Z}^d, \boldsymbol{n} \in \mathbb{N}_0^d}$ for $L^2(\mathbb{R}^d)$, let

$$\Sigma_{N,M} = \left\{ \sum_{(\boldsymbol{k},\boldsymbol{n}) \in \mathcal{I}} c_{\boldsymbol{k},\boldsymbol{n}} \psi_{\boldsymbol{k},\boldsymbol{n}} : \begin{array}{c} \mathcal{I} \subset \mathbb{Z}^d \times \mathbb{N}_0^d, \\ |\mathcal{I}| \leq N, \\ \max\limits_{(\boldsymbol{k},\boldsymbol{n}) \in \mathcal{I}} |c_{\boldsymbol{k},\boldsymbol{n}}| < M \end{array} \right\} \quad (9)$$

denote the set of all linear combinations consisting of at most N Wilson basis functions with bounded coefficients. We note that the approximation of functions with $\Sigma_{N,M}$ is a form of *nonlinear approximation* since $\Sigma_{N,M}$ is a nonlinear space. In Theorem 1, we construct an approximant from $\Sigma_{N,M}$ that achieves an approximation error rate that does not grow with the input dimension. The techniques used to prove this theorem are inspired by the work of DeVore and Temlyakov on nonlinear approximation by trigonometric sums [8]. Then, in Theorem 3, we show that, for $s > 0$, the exponent in the rate cannot be improved. Finally, we show in Theorem 4 that linear approximation methods necessarily suffer the curse of dimensionality.

Theorem 1. *Let $s \geq 0$. There exists a constant $M > 0$ which depends only on s and d such that, for all $f \in M^s_{1,1}(\mathbb{R}^d)$ with $\|f\|_{M^s_{1,1}} \leq 1$,*

$$\inf_{f_N \in \Sigma_{N,M}} \|f - f_N\|_{L^2} \leq C_0 M N^{-\frac{1}{2} - \frac{s}{2d}}, \quad (10)$$

where C_0 is a constant which may depend on d.

Proof. The expansion of f in the Wilson basis yields

$$f = \sum_{\substack{\boldsymbol{k} \in \mathbb{Z}^d \\ \boldsymbol{n} \in \mathbb{N}_0^d}} c_{\boldsymbol{k},\boldsymbol{n}} \psi_{\boldsymbol{k},\boldsymbol{n}}, \quad (11)$$

where $c_{\boldsymbol{k},\boldsymbol{n}} = \langle f, \psi_{\boldsymbol{k},\boldsymbol{n}} \rangle$. Since $f \in M^s_{1,1}(\mathbb{R}^d)$, the isomorphism between $M^s_{1,1}(\mathbb{R}^d)$ and $m^s_{1,1}$ establishes that

$$(1 + |(\boldsymbol{k}, \boldsymbol{n})|)^s c_{\boldsymbol{k},\boldsymbol{n}} \in \ell^1(\mathbb{Z}^d \times \mathbb{N}_0^d). \quad (12)$$

Let $(a_N)_{N \in \mathbb{N}}$ be a non-increasing rearrangement of the sequence $(|c_{\boldsymbol{k},\boldsymbol{n}}|)_{\boldsymbol{k} \in \mathbb{Z}^d, \boldsymbol{n} \in \mathbb{N}_0^d}$ of coefficients. Since $\|f\|_{M^s_{1,1}} \leq 1$, there exists a constant $M > 0$ such that

$$a_N \leq \frac{M}{N}. \quad (13)$$

Next, for a fixed $N \in \mathbb{N}$, define the index set Λ_N as the set of all $(\boldsymbol{k}, \boldsymbol{n}) \in \mathbb{Z}^d \times \mathbb{N}_0^d$ such that

$$(1 + |(\boldsymbol{k}, \boldsymbol{n})|)^s |c_{\boldsymbol{k},\boldsymbol{n}}| > M N^{-1} \quad (14)$$

or

$$|(\boldsymbol{k}, \boldsymbol{n})| < N^{\frac{1}{2d}}. \quad (15)$$

From (13), we see that the number of coefficients that satisfy (14) is at most N. Similarly, the number of coefficients that satisfy (15) does not exceed $C_1 N$, for some constant C_1. Therefore, $|\Lambda_N| \leq (1 + C_1) N$.

The approximant

$$f_{\Lambda_N} = \sum_{(\boldsymbol{k},\boldsymbol{n})\in\Lambda_N} c_{\boldsymbol{k},\boldsymbol{n}}\psi_{\boldsymbol{k},\boldsymbol{n}} \tag{16}$$

has at most $(1+C_1)N$ terms and $f_{\Lambda_N} \in \Sigma_{N,M}$, by design. We have that

$$\begin{aligned}\|f - f_{\Lambda_N}\|_{L^2}^2 &= \sum_{(\boldsymbol{k},\boldsymbol{n})\notin\Lambda_N} |c_{\boldsymbol{k},\boldsymbol{n}}|^2 \\ &= \sum_{(\boldsymbol{k},\boldsymbol{n})\notin\Lambda_N} \frac{(1+|(\boldsymbol{k},\boldsymbol{n})|)^{2s}}{(1+|(\boldsymbol{k},\boldsymbol{n})|)^{2s}}|c_{\boldsymbol{k},\boldsymbol{n}}|^2 \\ &\le \sum_{(\boldsymbol{k},\boldsymbol{n})\notin\Lambda_N} \frac{(1+|(\boldsymbol{k},\boldsymbol{n})|)^{2s}}{|(\boldsymbol{k},\boldsymbol{n})|^{2s}}|c_{\boldsymbol{k},\boldsymbol{n}}|^2 \\ &\le N^{-\frac{2s}{2d}} \sum_{(\boldsymbol{k},\boldsymbol{n})\notin\Lambda_N} (1+|(\boldsymbol{k},\boldsymbol{n})|)^{2s}|c_{\boldsymbol{k},\boldsymbol{n}}|^2 \\ &\le N^{-\frac{2s}{2d}} \sum_{m\ge N} M^2 m^{-2} \\ &\le CM^2 N^{-1-\frac{2s}{2d}},\end{aligned} \tag{17}$$

where $C > 0$ is a universal constant. Therefore,

$$\|f - f_{\Lambda_N}\|_{L^2} \le C^{\frac{1}{2}} M N^{-\frac{1}{2}-\frac{s}{2d}}, \tag{18}$$

which proves the theorem. □

Remark 2. Since the error in Theorem 1 is measured with respect to the $L^2(\mathbb{R}^d)$-norm, we automatically have that the same error rate holds with respect to the $L^2(\Omega)$-norm, for any bounded domain $\Omega \subset \mathbb{R}^d$.

To prove the lower bound, we use a variant of Carl's inequality based on entropy numbers [5], [6], [32]. Given a compact set $K \subset L^2(\mathbb{R}^d)$, its covering number $N_\varepsilon(K)_{L^2}$ is the smallest number of L^2-balls of radius ε that cover K. The notion of entropy of a compact set was introduced by Kolmogorov [22] to quantify its compactness. The (dyadic) *entropy number* of K is given by

$$\varepsilon_N(K)_{L^2} = \inf\{\varepsilon > 0 : \ N_\varepsilon(K)_{L^2} \le 2^N\}. \tag{19}$$

It indicates how precisely elements of K can be specified with N bits of information.

The variant of Carl's inequality from [32, Theorem 10] states that, under a compactness assumption, the existence of an upper bound on the nonlinear approximation rate with bounded coefficients implies, up to logarithmic factors, the same upper bound on the entropy number. By noting that $M^0_{2,2}(\mathbb{R}^d) = L^2(\mathbb{R}^d)$, a special case of [20, Theorem 3.2] implies that $M^s_{1,1}(\mathbb{R}^d)$ compactly embeds into $L^2(\mathbb{R}^d)$ if and only if $s > 0$. This implies that the unit ball

$$B_{M^s_{1,1}} = \{f \in M^s_{1,1}(\mathbb{R}^d) : \ \|f\|_{M^s_{1,1}} \le 1\} \tag{20}$$

is compact in $L^2(\mathbb{R}^d)$ if and only if $s > 0$. Thus, when $s > 0$, we can use the variant of Carl's inequality from [32, Theorem 10] which states that, if

$$\sup_{f\in B_{M^s_{1,1}}} \inf_{f_N\in\Sigma_{N,M}} \|f - f_N\|_{L^2} \lesssim N^{-\alpha}, \tag{21}$$

then

$$\varepsilon_{N\log N}(B_{M^s_{1,1}})_{L^2} \lesssim N^{-\alpha}. \tag{22}$$

We also note that a corollary of [20, Theorem 4.4] implies that

$$\varepsilon_N(B_{M^s_{1,1}})_{L^2} \asymp N^{-\frac{1}{2}-\frac{s}{2d}}. \tag{23}$$

With these results in hand, we can prove the lower bound for the approximation rate.

Theorem 3. *Let $s > 0$ and $M > 0$ be fixed and suppose that $\alpha > \frac{1}{2} + \frac{s}{2d}$. Then,*

$$\sup_{N\in\mathbb{N}} N^\alpha \left(\sup_{f\in B_{M^s_{1,1}}} \inf_{f_N\in\Sigma_{N,M}} \|f - f_N\|_{L^2} \right) = \infty. \tag{24}$$

Proof. Suppose that the supremum in (24) was finite. This would imply that

$$\sup_{\substack{f\in\mathcal{M}^s_{1,1}(\mathbb{R}^d) \\ \|f\|_{M^s_{1,1}}\le 1}} \inf_{f_N\in\Sigma_{N,M}} \|f - f_N\|_{L^2} \lesssim N^{-\alpha}. \tag{25}$$

By Carl's inequality, this would imply that

$$\varepsilon_{N\log N}(B_{M^s_{1,1}})_{L^2} \lesssim N^{-\alpha}, \tag{26}$$

which would contradict (23) since $\alpha > \frac{1}{2} + \frac{s}{2d}$. □

A. The Suboptimality of Linear Methods

To quantify the limits of linear approximations, we introduce the *linear N-width* of a set $K \subset L^2(\mathbb{R}^d)$. The linear N-width is given by

$$\delta_N(K)_{L^2} = \inf_{\mathrm{A}_N} \sup_{f\in K} \|f - \mathrm{A}_N\{f\}\|_{L^2}, \tag{27}$$

where the infimum is taken over all linear operators of rank N. When K is a compact, absolutely convex subset of $L^2(\mathbb{R}^d)$, the linear N-width coincides with the so-called Nth-approximation number of the associated embedding [19]. A special case of [20, Proposition 4.8] shows that, if $s > 0$, the Nth-approximation number of the identity map from $M^s_{1,1}(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ scales as $N^{-\frac{s}{2d}}$. Thus, since $B_{M^s_{1,1}}$ is a compact, absolutely convex subset of $L^2(\mathbb{R}^d)$ when $s > 0$, we immediately have the following result.

Theorem 4. *Let $s > 0$. Then,*

$$\delta_N(B_{M^s_{1,1}})_{L^2} \asymp N^{-\frac{s}{2d}}. \tag{28}$$

The main takeaway from this theorem is that a linear approximation method cannot achieve an approximation error that decays faster than the rate in (28). Thus, linear methods necessarily suffer the curse of dimensionality in the modulation spaces $M^s_{1,1}(\mathbb{R}^d)$.

IV. Discussion and Related Work

It follows from Section III that functions in $M^s_{1,1}(\mathbb{R}^d)$ can be approximated at rates that do not grow with the input dimension d. Another notable example of a class of functions which have this property are functions in the spectral Barron spaces $\mathscr{B}^s(\mathbb{R}^d)$, $s \geq 0$ [2]. These are Banach spaces of functions for which the the norm

$$\|f\|_{\mathscr{B}^s} = \int_{\mathbb{R}^d} (1 + \|\boldsymbol{\xi}\|_2)^s |\mathscr{F}\{f\}(\boldsymbol{\xi})| \, \mathrm{d}\boldsymbol{\xi} \tag{29}$$

is finite. Let $f \in \mathscr{B}^s(\mathbb{R}^d)$ be such that $\|f\|_{\mathscr{B}^s} \leq 1$. It was proven in [31, Theorem 1] that there exists an approximant f_N that takes the form of a linear combination of at most N complex exponentials such that

$$\|f - f_N\|_{L^2(\Omega)} \lesssim N^{-\frac{1}{2} - \frac{s}{d}}, \tag{30}$$

where $\Omega \subset \mathbb{R}^d$ is any bounded domain. Another notable example of such a class of functions is the family of Radon-domain BV spaces $\mathscr{R}\,\mathrm{BV}^k(\mathbb{R}^d)$, $k \in \mathbb{N}$ [23], [25], [26]. These are Banach spaces for which the seminorm

$$|f|_{\mathscr{R}\,\mathrm{BV}^k} = \|\partial_t^k \, \mathrm{K} \mathscr{R} f\|_{\mathcal{M}} \tag{31}$$

is finite, where $\mathscr{R}$ is the Radon transform, K is the filtering operator of computed tomography, ∂_t^k denotes k partial derivatives with respect to the offset variable of the Radon domain, and the $\mathcal{M}$-norm is the total variation norm in the sense of measures. Let $f \in \mathscr{R}\,\mathrm{BV}^k(\mathbb{R}^d)$ be such that $|f|_{\mathscr{R}\,\mathrm{BV}^k} \leq 1$. It was proven in [28, Section IV] in the case $k = 2$ and in [24, Theorem 4.8] for any $k \in \mathbb{N}$, that there exists an approximant f_N that takes the form of a shallow neural network with at most N neurons such that

$$\|f - f_N\|_{L^2(\Omega)} \lesssim N^{-\frac{1}{2} - \frac{2k-1}{2d}}, \tag{32}$$

where $\Omega \subset \mathbb{R}^d$ is any bounded domain.

The rate in Theorem 1 as well as the rates in (30) and (32) all behave as $N^{-\frac{1}{2} - \frac{\alpha}{d}}$, where α is related to the smoothness order of the space. Furthermore, $M^s_{1,1}(\mathbb{R}^d)$, $\mathscr{B}^s(\mathbb{R}^d)$, and $\mathscr{R}\,\mathrm{BV}^k(\mathbb{R}^d)$ are all Banach spaces defined by a sparsity-type (L^1 or total variation) norm in a transform domain. Thus, it appears as if appropriate notions of smoothness in a transform domain seem to "break" the curse of dimensionality, although this phenomenon is still not well understood.

V. Conclusion and Future Work

In this paper, we investigated the L^2-error of the approximation of functions in the modulation spaces $M^s_{1,1}(\mathbb{R}^d)$, $s \geq 0$, by linear combinations of Wilson basis elements. We derived sharp bounds on the (nonlinear) approximation error rates and proved that they do not grow with the input dimension d. We also showed that linear approximation methods necessarily suffer the curse of dimensionality. In particular, the results of this paper prove that the modulation spaces lie in certain L^2-approximation spaces of Wilson bases. One direction of future work will be directed towards a complete characterization of the approximation spaces induced by Wilson bases, which currently does not exist. Another direction of future work is to understand why and when the defining of smoothness spaces in a transform domain "breaks" the curse of dimensionality.

Acknowledgement

This work was supported by the Swiss National Science Foundation under Grant 200020_184646 / 1.

References

[1] F. Bach, "Breaking the curse of dimensionality with convex neural networks," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 629–681, 2017.

[2] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 930–945, 1993.

[3] L. Borup and M. Nielsen, "Nonlinear approximation in α-modulation spaces," *Mathematische Nachrichten*, vol. 279, no. 1-2, pp. 101–120, 2006.

[4] L. Borup and M. Nielsen, "Frame decomposition of decomposition spaces," *Journal of Fourier Analysis and Applications*, vol. 13, no. 1, pp. 39–70, 2007.

[5] B. Carl, "Entropy numbers, s-numbers, and eigenvalue problems," *Journal of Functional Analysis*, vol. 41, no. 3, pp. 290–306, 1981.

[6] A. Cohen, R. DeVore, G. Petrova, and P. Wojtaszczyk, "Optimal stable nonlinear approximation," *Foundations of Computational Mathematics*, vol. 22, no. 3, pp. 607–648, 2022.

[7] I. Daubechies, S. Jaffard, and J.-L. Journé, "A simple Wilson orthonormal basis with exponential decay," *SIAM Journal on Mathematical Analysis*, vol. 22, no. 2, pp. 554–573, 1991.

[8] R. A. DeVore and V. N. Temlyakov, "Nonlinear approximation by trigonometric sums," *Journal of Fourier Analysis and Applications*, vol. 2, no. 1, pp. 29–48, 1995.

[9] R. A. DeVore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.

[10] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Lectures*, p. 32, 2000.

[11] H. G. Feichtinger, "On a new Segal algebra," *Monatshefte für Mathematik*, vol. 92, pp. 269–289, 1981.

[12] H. G. Feichtinger, "Modulation spaces on locally compact abelian groups," *Technical Report, University Vienna*, 1983.

[13] H. G. Feichtinger, "Atomic characterizations of modulation spaces through Gabor-type representations," *The Rocky Mountain Journal of Mathematics*, pp. 113–125, 1989.

[14] H. G. Feichtinger, "Modulation spaces: Looking back and ahead," *Sampling Theory in Signal and Image Processing*, vol. 5, pp. 109–140, 2006.

[15] H. G. Feichtinger, K. Gröchenig, and D. Walnut, "Wilson bases and modulation spaces," *Mathematische Nachrichten*, vol. 155, no. 1, pp. 7–17, 1992.

[16] G. B. Folland, *Harmonic Analysis in Phase Space*, ser. Annals of Mathematics Studies. Princeton University Press, 1989.

[17] K. Gröchenig, *Foundations of Time-Frequency Analysis*, ser. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, 2001.

[18] K. Gröchenig and S. Samarah, "Nonlinear approximation with local Fourier bases," *Constructive Approximation*, vol. 16, pp. 317–331, 2000.

[19] S. Heinrich, "On the relation between linear n-widths and approximation numbers," *Journal of Approximation Theory*, vol. 58, no. 3, pp. 315–333, 1989.

[20] A. Hinrichs, I. Piotrowska, and M. Piotrowski, "On the degree of compactness of embeddings between weighted modulation spaces," *Journal of Function Spaces and Applications*, vol. 6, no. 3, pp. 303–317, 2008.

[21] L. K. Jones, "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training," *The Annals of Statistics*, pp. 608–613, 1992.

[22] A. N. Kolmogorov, "On linear dimensionality of topological vector spaces," in *Doklady Akademii Nauk*, vol. 120, no. 2. Russian Academy of Sciences, 1958, pp. 239–241.

[23] G. Ongie, R. Willett, D. Soudry, and N. Srebro, "A function space view of bounded norm infinite width ReLU nets: The multivariate case," in *International Conference on Learning Representations*, 2020.

[24] R. Parhi, “On ridge splines, neural networks, and variational problems in Radon-domain BV spaces,” Ph.D. dissertation, The University of Wisconsin-Madison, 2022.
[25] R. Parhi and R. D. Nowak, “Banach space representer theorems for neural networks and ridge splines,” *Journal of Machine Learning Research*, vol. 22, no. 43, pp. 1–40, 2021.
[26] R. Parhi and R. D. Nowak, “What kinds of functions do deep neural networks learn? Insights from variational spline theory,” *SIAM Journal on Mathematics of Data Science*, vol. 4, no. 2, pp. 464–489, 2022.
[27] R. Parhi and R. D. Nowak, “Deep learning meets sparse regularization: A signal processing perspective,” *arXiv preprint arXiv:2301.09554*, 2023.
[28] R. Parhi and R. D. Nowak, “Near-minimax optimal estimation with shallow ReLU neural networks,” *IEEE Transactions on Information Theory*, vol. 69, no. 2, pp. 1125–1140, 2023.
[29] G. Pisier, “Remarques sur un résultat non publié de B. Maurey,” *Séminaire d’Analyse Fonctionnelle (dit “Maurey-Schwartz”)*, pp. 1–12, April 1981.
[30] A. Rahimi and B. Recht, “Uniform approximation of functions with random bases,” in *2008 46th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2008, pp. 555–561.
[31] J. W. Siegel and J. Xu, “High-order approximation rates for shallow neural networks with cosine and ReLU^k activation functions,” *Applied and Computational Harmonic Analysis*, vol. 58, pp. 1–26, 2022.
[32] J. W. Siegel and J. Xu, “Sharp bounds on the approximation rates, metric entropy, and n-widths of shallow neural networks,” *Foundations of Computational Mathematics*, pp. 1–57, 2022.