DEFORMATION STABILITY OF DEEP CONVOLUTIONAL NEURAL NETWORKS ON SOBOLEV SPACES

Michael Koller[†⋆] *Johannes Großmann*[‡] *Ullrich Mönich*[‡] *Holger Boche*[‡]

Technische Universität München
[‡] Lehrstuhl für Theoretische Informationstechnik
[†] Professur für Methoden der Signalverarbeitung

ABSTRACT

Our work is based on a recently introduced mathematical theory of deep convolutional neural networks (DCNNs). It was shown that DCNNs are stable with respect to deformations of bandlimited input functions. In the present paper, we generalize this result: We prove deformation stability on Sobolev spaces. Further, we show a weak form of deformation stability for the whole input space $L^2(\mathbb{R}^d)$. The basic components of DCNNs are semi-discrete frames. For practical applications, a concrete choice is necessary. Therefore, we conclude our work by suggesting a construction method for semi-discrete frames based on bounded uniform partitions of unity (BUPUs) and give a specific example that uses B-splines.

Index Terms— Deep convolutional neural networks, deformation stability, Sobolev space, bounded uniform partition of unity, admissible semi-discrete frame

1. INTRODUCTION

One approach to object recognition [1] is the following two-stage process [2]. The first stage consists of a convolutional neural network [3] which extracts features from the input [4]. In a second step, these features are fed into a support vector machine (SVM) for classification [5]. Here, we are interested in the feature extraction via convolutional neural networks.

Inspired by [6], the authors of [7] developed a mathematical theory of deep convolutional neural networks (DCNNs) upon which our work is based. Two important properties of DCNNs were investigated in [7]: translation invariance and deformation stability. For an intuitive explanation of these notions, let $f : \mathbb{R} \to \mathbb{C}$ be a $L^2(\mathbb{R})$ function which is the input to the DCNN. The idea of translation invariance is that features extracted from f and its translated version $f(\cdot - c)$ with a constant $c < \infty$ should be very similar or even identical (see also, e.g., [8]). Deformation stability, on the other hand, assumes that for a slowly changing differentiable map $\tau : \mathbb{R} \to \mathbb{R}$, the original f and the deformed version $f(\cdot - \tau(\cdot))$ yield similar features. In addition, the similarity should increase if the amount of deformation, measured by $\sup_{t\in\mathbb{R}} |\tau(t)|$, goes to zero. We define the notion of deformation stability more formally in Section 2.

Deformation stability of DCNNs has been investigated in [9] for cartoon functions and in [7] for bandlimited functions in $L^2(\mathbb{R}^d)$. In the present paper, we generalize the deformation stability result to functions in the Sobolev space $H^2(\mathbb{R}^d)$ (cf. Definition 4 below). This space contains all bandlimited functions and is dense in the input space $L^2(\mathbb{R}^d)$ of the DCNNs. We exploit this density to prove a weak form of deformation stability for the whole space $L^2(\mathbb{R}^d)$.

The elements of the DCNNs in the theory of [7] form a so-called admissible semi-discrete frame (cf. Definition 1 below). The concrete choice for this frame has consequences in practical implementations. For example, as discussed in [10], the energy decay with increasing network depth is influenced by the frame. We discuss how bounded uniform partitions of unity (BUPUs), which are important quantities in Wiener amalgam spaces (cf. [11, 12]), provide a tool to construct admissible semi-discrete frames with desired properties and give an explicit example.

1.1. Notation and definitions

Given some $x \in \mathbb{C}$, we denote the complex conjugate as $\overline{x}$. The Euclidean inner product for $x, y \in \mathbb{C}^d$ then is $\langle x, y\rangle := \sum_{i=1}^d x_i \overline{y}_i$ with corresponding norm $\|x\|_{\mathrm{E}} := \sqrt{\langle x, x\rangle}$. For a differentiable map $\tau : \mathbb{R}^d \to \mathbb{R}^d$, $\mathrm{D}\tau$ is its Jacobian matrix. The norms corresponding to these quantities are $\|\tau\|_\infty := \sup_{x\in\mathbb{R}^d} \|\tau(x)\|_{\mathrm{E}}$ and $\|\mathrm{D}\tau\|_\infty := \sup_{x\in\mathbb{R}^d} |(\mathrm{D}\tau)(x)|_\infty$. The supremum norm of a matrix M is defined as $|M|_\infty := \sup_{i,j} |M_{i,j}|$.

As usual, $L^p(\mathbb{R}^d)$, $p \in [1, \infty)$, denotes the space of all Lebesgue-measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ such that $\|f\|_p := (\int_{\mathbb{R}^d} |f(x)|^p \,\mathrm{d}x)^{1/p} < \infty$. In the case $p = \infty$, we have $\|f\|_\infty := \sup_{x\in\mathbb{R}^d} |f(x)|$. If $f : \mathbb{R}^d \to \mathbb{C}^m$ is a vector field, the L^2-norm is given by $\|f\|_2 := (\int_{\mathbb{R}^d} \|f(x)\|_{\mathrm{E}}^2 \,\mathrm{d}x)^{1/2}$.

For $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, the Fourier transform is

[⋆]This work was conducted while Michael Koller was with the Lehrstuhl für Theoretische Informationstechnik.

This work was supported by the German Research Foundation (DFG) under grants BO 1734/20-1 and BO 1734/24-1.

$\hat{f}(\omega) := \int_{\mathbb{R}^d} f(x)e^{-2\pi i\langle x,\omega\rangle}\,\mathrm{d}x$ and can be extended via Plancherel's theorem to $L^2(\mathbb{R}^d)$. The convolution of $f \in L^1(\mathbb{R}^d)$ and $g \in L^2(\mathbb{R}^d)$ is $(f \star g)(y) := \int_{\mathbb{R}^d} f(x)g(y-x)\,\mathrm{d}x$.

We define the Paley-Wiener space of bandlimited functions as $\mathcal{PW}^2_\sigma(\mathbb{R}^d) := \{f \in L^2(\mathbb{R}^d) \mid \mathrm{supp}(\hat{f}) \subseteq \mathcal{U}_\sigma(0)\}$ where $\mathcal{U}_\sigma(0)$ is the ball of radius σ about $0 \in \mathbb{R}^d$.

2. THE NETWORK ARCHITECTURE

In what follows, we describe the architecture of DCNNs as it was developed in [7]. We do not make use of pooling (which might be considered as pooling by sub-sampling with pooling factor 1) and use the modulus non-linearity. Thereafter, via the stability result of [7] for bandlimited functions, we explain the notion of deformation stability.

An important component of DCNNs are admissible semi-discrete frames.

Definition 1 (Admissible semi-discrete frame)**.** *A set* $\{h_{out}\} \cup \{h_\lambda\}_{\lambda\in\Lambda}$ *of functions, where* $h_{out}, h_\lambda \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$*, with a countable index set* Λ *is called admissible semi-discrete frame if for all* $f \in L^2(\mathbb{R}^d)$*, there exist constants* $0 < A \le B \le 1$ *such that*

$$A\|f\|_2^2 \le \|f \star h_{out}\|_2^2 + \sum_{\lambda\in\Lambda} \|f \star h_\lambda\|_2^2 \le B\|f\|_2^2. \qquad (1)$$

Remark 2. *According to [7, Proposition 2], the frame condition* (1) *is equivalent to the Littlewood-Paley condition*

$$A \le |\hat{h}_{out}(\omega)|^2 + \sum_{\lambda\in\Lambda} |\hat{h}_\lambda(\omega)|^2 \le B, \quad \text{a.e. } \omega \in \mathbb{R}^d.$$

In every network layer, a DCNN repeats the process of computing convolutions and applying the modulus to the result. Figure 1 visualizes the idea. In the first layer, an input $f \in L^2(\mathbb{R}^d)$ is convolved with every element $h_\lambda, \lambda \in \Lambda$, of an admissible semi-discrete frame and then the modulus is applied. In layer m, every element of the convolution+modulus of the previous layer $m-1$ goes through the same process: a convolution with every element h_λ is computed and the modulus is applied. There are infinitely many layers.

An ordered sequence $p = (\lambda_1, \dots, \lambda_m)$ of indices $\lambda_i \in \Lambda$ is called a path of length $|p| = m$. Let us define a corresponding operator

$$\mathrm{U}[p]f := |\dots||f \star h_{\lambda_1}| \star h_{\lambda_2}| \cdots \star h_{\lambda_m}|$$

which describes repeating convolution+modulus m times. We set $\mathrm{U}[\emptyset]f = f$. Then, a feature of the scattering network is: $\mathrm{S}[p]f := \mathrm{U}[p]f \star h_{\text{out}}$. Let Λ^m denote the set of paths of length m and $\mathcal{P} = \cup_{m=0}^\infty \Lambda^m$ the set of all paths. Then, the extracted feature vector is $\mathrm{S}[\mathcal{P}]f := \{\mathrm{S}[p]f\}_{p\in\mathcal{P}}$.

For a given set of paths $\mathcal{Q} \subseteq \mathcal{P}$, we introduce the operator

$$\mathrm{E}_2(\mathrm{U}[\mathcal{Q}]f) := \sqrt{\sum_{p\in\mathcal{Q}} \|\mathrm{U}[p]f\|_2^2}$$

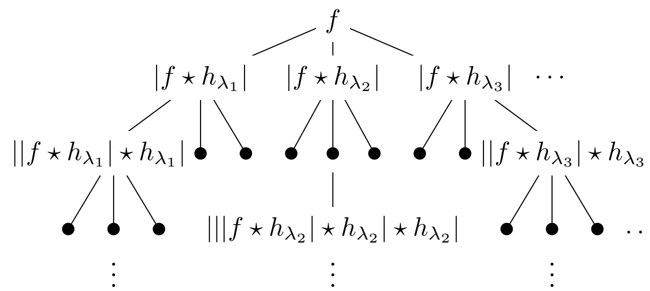

Fig. 1: DCNN architecture with input f and elements h_{λ_i}, $\lambda_i \in \Lambda$, of an admissible semi-discrete frame.

which might be interpreted as collecting the "energy" of all signals corresponding to the paths in $\mathcal{Q}$. From [7, Eq. (24)], we know that for admissible semi-discrete frames, the feature extraction is Lipschitz:

$$\mathrm{E}_2(\mathrm{S}[\mathcal{P}]g - \mathrm{S}[\mathcal{P}]f) \le \|g - f\|_2, \quad \forall f, g \in L^2(\mathbb{R}^d). \qquad (2)$$

This property was crucial for the derivation of deformation stability in [7]. We recite this result in Theorem 3 below.

Let us first define a deformation operator $\mathrm{T}_\tau f := f(\cdot - \tau(\cdot))$ for a differentiable map $\tau : \mathbb{R}^d \to \mathbb{R}^d$. By construction, T_τ is a linear operator. It is possible to show (see, e.g., [6]) that

$$\|\mathrm{T}_\tau f\|_2 \le \sqrt{2^d}\|f\|_2 \qquad (3)$$

and, hence, $\mathrm{T}_\tau f \in L^2(\mathbb{R}^d)$ holds if $\|\mathrm{D}\tau\|_\infty \le 1/2$. Now, deformation stability is to be understood in the sense of (4):

Theorem 3 (Theorem 2 of [7])**.** *There exists a constant* $C > 0$ *such that for all* $f \in \mathcal{PW}^2_\sigma(\mathbb{R}^d)$ *and all* $\tau : \mathbb{R}^d \to \mathbb{R}^d$ *with* $\|\mathrm{D}\tau\|_\infty \le \frac{1}{2d}$*, it holds:*

$$\mathrm{E}_2(\mathrm{S}[\mathcal{P}]\mathrm{T}_\tau f - \mathrm{S}[\mathcal{P}]f) \le C\sigma\|\tau\|_\infty\|f\|_2. \qquad (4)$$

For $\|\tau\|_\infty \to 0$, the right-hand side of (4) goes to 0. This means that the feature vectors $\mathrm{S}[\mathcal{P}]\mathrm{T}_\tau f$ and $\mathrm{S}[\mathcal{P}]f$ are similar for "small" deformations and they are identical in the limit. In other words, small deformations do not have a large influence on the extracted feature vector of the DCNN.

3. DEFORMATION STABILITY

This section shows that the feature extraction is deformation stable on the Sobolev space $H^2(\mathbb{R}^d)$. We follow [13] in the definition of $H^2(\mathbb{R}^d)$.

Definition 4 (Sobolev space $H^2(\mathbb{R}^d)$)**.** *Let* $D_i := \frac{\partial}{\partial x_i}$ *denote a differential operator and let* $\alpha = (\alpha_1, \dots, \alpha_d)$ *be a d-tupel of non-negative integers* $\alpha_i \ge 0$*. We say* α *has degree* $|\alpha| := \sum_{i=1}^d \alpha_i$ *and set*

$$D^\alpha := D_1^{\alpha_1} \dots D_d^{\alpha_d}$$

with $D^{(0,\dots,0)}f = f$ *for every* $f \in L^2(\mathbb{R}^d)$*.*

Now, we define the Sobolev norm

$$\|f\|_{H^2} := \sqrt{\sum_{0\le|\alpha|\le 2} \|D^\alpha f\|_2^2}$$

for all $f \in L^2(\mathbb{R}^d)$ for which the expression makes sense. The Sobolev space then is given by

$$H^2(\mathbb{R}^d) := \{f \in L^2(\mathbb{R}^d) \mid \|f\|_{H^2} < \infty\}.$$

The fact that the gradient ∇f of $f \in H^2(\mathbb{R}^d)$ exists helps us to prove deformation stability:

Theorem 5 (Deformation stability on $H^2(\mathbb{R}^d)$). *Let $\{h_{out}\} \cup \{h_\lambda\}_{\lambda\in\Lambda}$ be an admissible semi-discrete frame. Then, for $f \in H^2(\mathbb{R}^d)$ and deformations $\tau : \mathbb{R}^d \to \mathbb{R}^d$ with $\|\mathrm{D}\tau\|_\infty \le 1/2$, it holds*

$$\mathrm{E}_2(\mathrm{S}[\mathcal{P}]\mathrm{T}_\tau f - \mathrm{S}[\mathcal{P}]f) \le \sqrt{2^d}\|\tau\|_\infty\|\nabla f\|_2. \tag{5}$$

Proof. For $f \in H^2(\mathbb{R}^d)$, define the auxiliary function

$$k(s,t) := f(t - s\tau(t)). \tag{6}$$

The fundamental theorem of calculus yields

$$\begin{aligned} f(t-\tau(t)) = k(1,t) &= k(0,t) + \int_0^1 \frac{\partial}{\partial\xi}k(\xi,t)\Big|_{\xi=s} \mathrm{d}s \\ &= f(t) - \int_0^1 \langle\nabla f(t-s\tau(t)), \tau(t)\rangle\,\mathrm{d}s. \end{aligned}$$

This implies

$$\begin{aligned} \|\mathrm{T}_\tau f - f\|_2^2 &= \int_{\mathbb{R}^d}\Big|\int_0^1 \langle\nabla f(t-s\tau(t)),\tau(t)\rangle\,\mathrm{d}s\Big|^2\mathrm{d}t \\ &\le \int_{\mathbb{R}^d}\left(\int_0^1 |\langle\nabla f(t-s\tau(t)),\tau(t)\rangle|\,\mathrm{d}s\right)^2\mathrm{d}t \\ &\overset{(a)}{\le} \int_{\mathbb{R}^d}\int_0^1 |\langle\nabla f(t-s\tau(t)),\tau(t)\rangle|^2\,\mathrm{d}s\,\mathrm{d}t \\ &\overset{(b)}{\le} \int_0^1\int_{\mathbb{R}^d}\Big(\|\nabla f(t-s\tau(t))\|_{\mathrm{E}}\|\tau(t)\|_{\mathrm{E}}\Big)^2\mathrm{d}t\,\mathrm{d}s \\ &\overset{(c)}{\le} \|\tau\|_\infty^2\int_0^1\int_{\mathbb{R}^d}\|\nabla f(t-s\tau(t))\|_{\mathrm{E}}^2\,\mathrm{d}t\,\mathrm{d}s \\ &= \|\tau\|_\infty^2\int_0^1 \|\mathrm{T}_{s\tau}\nabla f\|_2^2\,\mathrm{d}s. \end{aligned} \tag{7}$$

Inequality (a) is a consequence of the Cauchy-Schwarz inequality. After applying Fubini, using Cauchy-Schwarz again, $|\langle\nabla f(t-s\tau(t)),\tau(t)\rangle| \le \|\nabla f(t-s\tau(t))\|_{\mathrm{E}}\|\tau(t)\|_{\mathrm{E}}$, yields (b). Further, by definition $\|\tau(t)\|_{\mathrm{E}} \le \|\tau\|_\infty$ holds which justifies (c). Exploiting inequality (3) results in

$$\|\mathrm{T}_\tau f - f\|_2^2 \le \|\tau\|_\infty^2\int_0^1 2^d\|\nabla f\|_2^2\,\mathrm{d}s = 2^d\|\tau\|_\infty^2\|\nabla f\|_2^2. \tag{8}$$

Inserting (8) in (2) with $g = \mathrm{T}_\tau f$ results in the statement of the theorem.

□

4. WEAK DEFORMATION STABILITY

The Sobolev space $H^2(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$. This fact together with Theorem 5 allows us to prove a weak form of deformation stability for the whole input space $L^2(\mathbb{R}^d)$:

Theorem 6 (Weak deformation stability). *Let $\{h_{out}\} \cup \{h_\lambda\}_{\lambda\in\Lambda}$ be an admissible semi-discrete frame. Then, for all $f \in L^2(\mathbb{R}^d)$ and all $\varepsilon > 0$, it holds*

$$\exists\delta > 0 : \forall\tau : \|\tau\|_\infty < \delta : \mathrm{E}_2(\mathrm{S}[\mathcal{P}]\mathrm{T}_\tau f - \mathrm{S}[\mathcal{P}]f) < \varepsilon \tag{9}$$

where τ is an admissible deformation, i.e., a differentiable mapping $\tau : \mathbb{R}^d \to \mathbb{R}^d$ with $\|\mathrm{D}\tau\|_\infty \le 1/2$.

For small enough $\|\tau\|_\infty$, i.e., for small enough deformations, $\mathrm{E}_2(\mathrm{S}[\mathcal{P}]\mathrm{T}_\tau f - \mathrm{S}[\mathcal{P}]f)$ can be made arbitrarily small. We call this weak deformation stability.

Proof of Theorem 6. Let $f \in L^2(\mathbb{R}^d)$ and $\varepsilon > 0$ be arbitrary. Because $H^2(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$, for any $\tilde\varepsilon < \frac{1}{\sqrt{2^d}+2}\varepsilon$, there exists a $g_{\tilde\varepsilon} \in H^2(\mathbb{R}^d)$ with $\|g_{\tilde\varepsilon} - f\|_2 \le \tilde\varepsilon$. Applying the triangle inequality yields:

$$\begin{aligned} \|\mathrm{T}_\tau f - f\|_2 &= \|\mathrm{T}_\tau f - \mathrm{T}_\tau g_{\tilde\varepsilon} + \mathrm{T}_\tau g_{\tilde\varepsilon} - g_{\tilde\varepsilon} + g_{\tilde\varepsilon} - f\|_2 \\ &\le \|\mathrm{T}_\tau f - \mathrm{T}_\tau g_{\tilde\varepsilon}\|_2 + \|\mathrm{T}_\tau g_{\tilde\varepsilon} - g_{\tilde\varepsilon}\|_2 + \tilde\varepsilon \\ &\overset{(3)}{\le} \sqrt{2^d}\tilde\varepsilon + \|\mathrm{T}_\tau g_{\tilde\varepsilon} - g_{\tilde\varepsilon}\|_2 + \tilde\varepsilon. \end{aligned}$$

Combining this with Theorem 5 shows

$$\|\mathrm{T}_\tau f - f\|_2 \le (\sqrt{2^d}+1)\tilde\varepsilon + \sqrt{2^d}\|\tau\|_\infty\|\nabla g_{\tilde\varepsilon}\|_2$$

where $\|\nabla g_{\tilde\varepsilon}\|_2$ is finite. For $\|\tau\|_\infty < \delta$ with $\delta = \frac{\tilde\varepsilon}{\sqrt{2^d}\|\nabla g_{\tilde\varepsilon}\|_2}$ and using (2) with $g = \mathrm{T}_\tau f$, we get

$$\mathrm{E}_2(\mathrm{S}[\mathcal{P}]\mathrm{T}_\tau f - \mathrm{S}[\mathcal{P}]f) \le \|\mathrm{T}_\tau f - f\|_2 \le (\sqrt{2^d}+2)\tilde\varepsilon < \varepsilon \tag{10}$$

which yields the statement of the theorem. □

5. SEMI-DISCRETE FRAMES FROM BOUNDED UNIFORM PARTITIONS OF UNITY

DCNNs and, therefore, also the considerations in the previous sections are based on admissible semi-discrete frames. The concrete choice of a semi-discrete frame influences the performance of the feature extractor in practical applications. For example, [10] discusses how the choice affects the energy decay over different network layers, which gives an idea of how many layers are needed in an implementation. It is thus of interest to be able to construct admissible frames with certain properties. We show that BUPUs provide a means to do so.

Definition 7 (BUPU, [12]). *A set of functions $\{\hat\varphi_j\}_{j\in I}$ on $\mathbb{R}$ with $\hat\varphi_j \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is a bounded uniform partition of unity (BUPU) if*

1. $\sum_{j \in I} \hat{\varphi}_j \equiv 1$,
2. $\sup_j \|\hat{\varphi}_j\|_\infty < \infty$,
3. *there exist a compact set $U \subset \mathbb{R}$ with nonempty interior and points $y_j \in \mathbb{R}$ such that $\operatorname{supp}(\hat{\varphi}_j) \subset U + y_j$ for all $j \in I$, and*
4. *for each compact $K \subset \mathbb{R}$, we have*
$$\sup_{x \in \mathbb{R}} |\{j \in I \mid x \in K + y_j\}| < \infty.$$

Note that it is possible to extend the theory of BUPUs to $\mathbb{R}^d$ with $d \geq 1$ (see, e.g., [14, 15]). For simplicity, we focus on $d = 1$. The following lemma connects BUPUs to semi-discrete frames.

Lemma 8. *Let $\{\hat{\varphi}_j\}_{j \in I}$ be a BUPU. Then, there exist constants $0 < A \leq B < \infty$ such that*
$$A\|f\|_2^2 \leq \sum_{j \in I} \|f \star \varphi_j\|_2^2 \leq B\|f\|_2^2, \quad \forall f \in L^2(\mathbb{R}).$$

Proof. Combining [12, Proposition 5.2] and [12, Theorem 6.7] for $L^2(\mathbb{R})$ reveals the norm equivalence $\|\hat{f}\|_2^2 \asymp \sum_{j \in I} \|\hat{f}\hat{\varphi}_j\|_2^2$. That is, there exist constants $0 < A \leq B < \infty$ such that for any $\hat{f} \in L^2(\mathbb{R})$, it holds
$$A\|\hat{f}\|_2^2 \leq \sum_{j \in I} \|\hat{f}\hat{\varphi}_j\|_2^2 \leq B\|\hat{f}\|_2^2.$$

Applying the convolution theorem, i.e., $2\pi\|f \star g\|_2^2 = \|\hat{f}\hat{g}\|_2^2$ for $f \in L^2(\mathbb{R})$ and $g \in L^1(\mathbb{R})$, completes the proof. □

Remark 9. *In order for a semi-discrete frame to be admissible, the upper frame bound needs to fulfill $B \leq 1$ (cf. Definition 1). According to [7, Proposition 3 in Appendix A], this can easily be achieved by rescaling all frame elements.*

5.1. Semi-discrete frames from B-splines

We use B-splines to give an explicit example of an admissible semi-discrete frame that is constructed via a BUPU. Our discussion of B-splines is due to [16]. Define the box function
$$\hat{\varphi}^{(0)}(\omega) := \begin{cases} 1, & -\frac{1}{2} < \omega < \frac{1}{2}, \\ \frac{1}{2}, & |\omega| = \frac{1}{2}, \\ 0, & \omega \notin [-\frac{1}{2}, \frac{1}{2}], \end{cases}$$

with support in $[-\frac{1}{2}, \frac{1}{2}]$, and observe the identity
$$\sum_{k \in \mathbb{Z}} \hat{\varphi}^{(0)}(\omega - k) = 1, \quad \forall \omega \in \mathbb{R}. \tag{11}$$

A B-spline of degree $n \geq 1$ then is defined to be
$$\hat{\varphi}^{(n)} := \underbrace{\hat{\varphi}^{(0)} \star \hat{\varphi}^{(0)} \star \cdots \star \hat{\varphi}^{(0)}}_{(n+1) \text{ times}}. \tag{12}$$

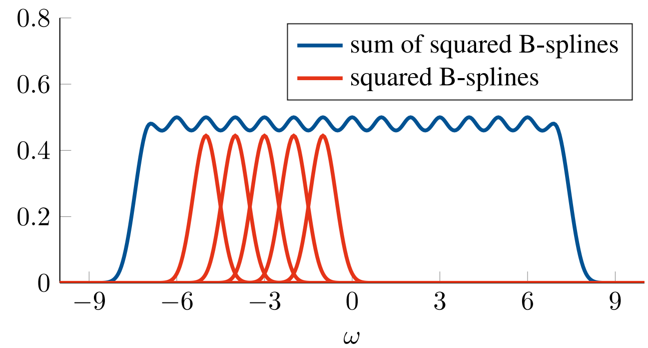

Fig. 2: The sum $\sum_{k=-7}^{7}(\hat{\varphi}^{(3)}(\omega - k))^2$ and the summands $(\hat{\varphi}^{(3)}(\omega - k))^2$ for $k \in \{-5, \ldots, -1\}$.

A n times convolution of (11) with $\hat{\varphi}^{(0)}$ yields
$$\sum_{k \in \mathbb{Z}} \hat{\varphi}^{(n)}(\omega - k) = 1, \quad \forall \omega \in \mathbb{R}$$

because $1 \star \hat{\varphi}^{(0)} = 1$. Clearly, any $\hat{\varphi}^{(n)}(\cdot - k)$ is bounded and the set $\{\hat{\varphi}^{(n)}(\cdot - k)\}_{k \in \mathbb{Z}}$ is a partition of unity. Further, all $\hat{\varphi}^{(n)}(\cdot - k)$ have compact support with equal length.

There exists a closed form expression of the convolution in (12) whose inverse Fourier transform is given by
$$\varphi^{(n)}(t) = \left(\frac{\sin(t/2)}{t/2}\right)^{n+1}.$$

Straightforward computations yield $\varphi^{(n)} \in L^\infty(\mathbb{R})$ and $\varphi^{(n)} \in L^1(\mathbb{R})$ for $n \geq 1$. Hence, also $\varphi^{(n)} \in L^p(\mathbb{R})$ is true for all $1 \leq p \leq \infty$.

In summary, $\{\hat{\varphi}^{(n)}(\cdot - k)\}_{k \in \mathbb{Z}}$ is a BUPU and, due to Lemma 8, $\{\varphi^{(n)}(\cdot - k)\}_{k \in \mathbb{Z}}$ is an admissible semi-discrete frame. As an example, Figure 2 illustrates that B-splines with $n = 3$ fulfill the Littlewood-Paley condition from Remark 2 with $B \leq 1$. Finally, we note that other choices than B-splines are possible. For instance, [17] constructs BUPUs with a smooth window from a hyperbolic tangent.

6. RELATION TO PRIOR WORK

In 2012, Mallat analyzed a particular DCNN which consists of so-called scattering wavelets [6]. This was a first step towards a mathematical theory of DCNNs. It was then generalized in [7] and deformation stability of the general DCNNs for bandlimited functions has been proved. Our work extends the deformation stability to the Sobolev space $H^2(\mathbb{R}^d)$ and, in a weak form, to the whole space $L^2(\mathbb{R}^d)$.

7. ACKNOWLEDGEMENT

The authors would like to thank Prof. Feichtinger for an introduction to Wiener amalgam spaces which inspired the connection between BUPUs and semi-discrete frames.

8. REFERENCES

[1] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, Jan. 2007.

[2] Fu Jie Huang and Yann LeCun, "Large-scale learning with SVM and convolutional nets for generic object categorization," in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*, Washington, DC, USA, 2006, CVPR '06, pp. 284–291, IEEE Computer Society.

[3] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, May 2010, pp. 253–256.

[4] Yoshua Bengio, Aaron Courville, and Pascal Vincent, "Representation learning: a review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[5] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar, *Foundations of Machine Learning*, Adaptive computation and machine learning. MIT Press, Cambridge, Mass., 2012.

[6] Stéphane Mallat, "Group invariant scattering," *CoRR*, vol. abs/1101.2286, 2011.

[7] Thomas Wiatowski and Helmut Bölcskei, "A mathematical theory of deep convolutional neural networks for feature extraction," *CoRR*, 2015.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts and London, England, 2016.

[9] Philipp Grohs, Thomas Wiatowski, and Helmut Bölcskei, "Deep convolutional neural networks on cartoon functions," *CoRR*, 2016.

[10] Thomas Wiatowski, Philipp Grohs, and Helmut Bölcskei, "Energy propagation in deep convolutional neural networks," *CoRR*, vol. abs/1704.03636, 2017.

[11] John J. F. Fournier and James Stewart, "Amalgams of L^p and ℓ^q," *Bull. Amer. Math. Soc. (N.S.)*, vol. 13, pp. 1–21, 1985.

[12] Christopher Heil, "An introduction to weighted Wiener amalgams," in *Wavelets and their Applications (Chennai, January 2002)*, M. Krishna, R. Radha, and S. Thangavelu, Eds., pp. 183–216. Allied Publishers, 2003.

[13] Robert A. Adams and John J. F. Fournier, *Sobolev Spaces*, vol. 140 of *Pure and applied mathematics*, Acad. Press, Amsterdam, 2. ed. edition, 2003.

[14] Hans G. Feichtinger, "A characterization of minimal homogeneous Banach spaces," *Proc. Amer. Math. Soc.*, vol. 81, no. 1, pp. 55–61, 1981.

[15] Hans G. Feichtinger, "Banach convolution algebras of Wiener type," in *Proc. Conf. on Functions, Series, Operators, Budapest 1980*, vol. 35 of *Colloq. Math. Soc. Janos Bolyai*, pp. 509–524. North-Holland, Eds. B. Sz.-Nagy and J. Szabados. edition, 1983.

[16] Michael Unser, "Splines: a perfect fit for signal and image processing," *Signal Processing Magazine, IEEE*, vol. 16, no. 6, pp. 22–38, Nov. 1999.

[17] Michael Lamoureux, Peter Gibson, Jeff Grossman, and Gary Margrave, "A fast, discrete Gabor transform via a partition of unity," *Journal of Fourier Analysis and Applications*, 2003.