Deformation Stability of Deep Convolutional Neural Networks on Sobolev Spaces

Michael Koller, Johannes Großmann,
Ullrich Mönich, Holger Boche

Lehrstuhl für Theoretische Informationstechnik
Technische Universität München

April 19, 2018

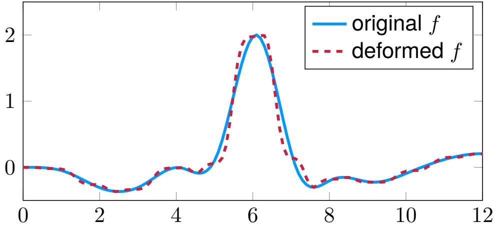
Classification: General Idea

Figure: Two-stage classification process.

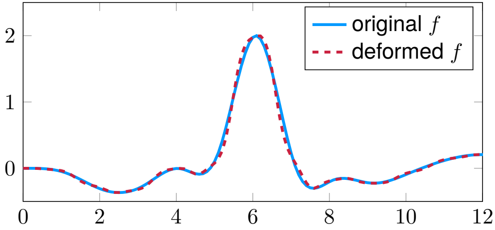
- Feature extraction: **D**eep **C**onvolutional **N**eural **N**etwork
- Feature processing: e.g., **S**upport **V**ector **M**achine

Notion of Deformation Stability

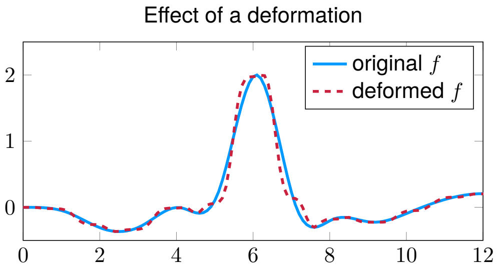
Effect of a deformation

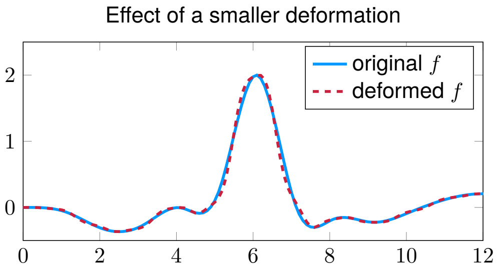


Effect of a smaller deformation



Notion of Deformation Stability





- **Input:** $f \in L^2(\mathbb{R}^d)$ with norm $\|f\|_2 = \sqrt{\int_{\mathbb{R}^d} |f(t)|^2 \, \mathrm{d}t} < \infty$

Notion of Deformation Stability

- For $f \in L^2(\mathbb{R}^d)$, the deformation operator T_τ is defined via

$$\mathrm{T}_\tau f = f(\cdot - \tau(\cdot)).$$

- A mapping $\tau : \mathbb{R}^d \to \mathbb{R}^d$ is called deformation if it is twice differentiable and obeys

$$\|\tau\|_\infty := \sup_{t\in\mathbb{R}^d} \|\tau(t)\|_{\mathrm{E}} < \infty$$

  and

$$\|\mathrm{D}\tau\|_\infty := \sup_{t\in\mathbb{R}^d} |(\mathrm{D}\tau)(t)|_\infty \leq \frac{1}{2}$$

  where $\mathrm{D}\tau$ is the Jacobian.
- It holds $\|\mathrm{T}_\tau f\|_2 \leq 2^d \|f\|_2$ and thus $\mathrm{T}_\tau f \in L^2(\mathbb{R}^d)$.

The supremum norm of a matrix M is defined as $|M|_\infty := \sup_{i,j} |M_{i,j}|$.

Network Architecture

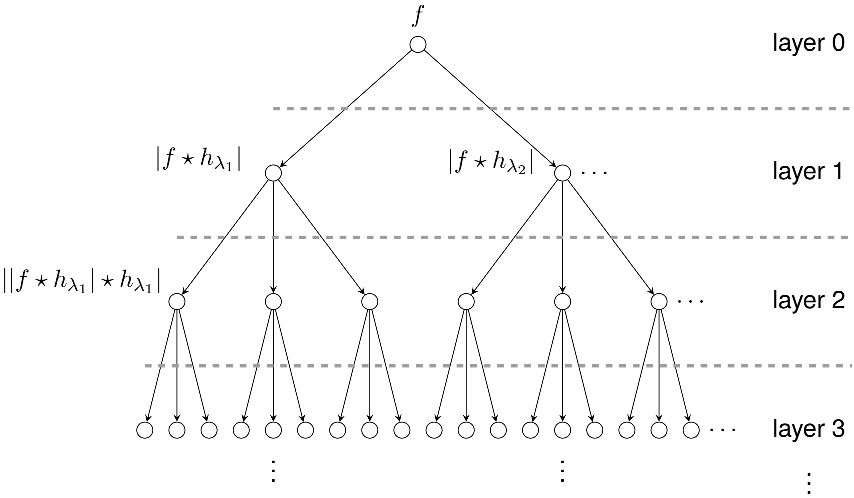
- The set of functions $\{h_{\mathrm{out}}\} \cup \{h_\lambda\}_{\lambda\in\Lambda}$ with a countable index set Λ and $h_{\mathrm{out}}, h_\lambda \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ is called semi-discrete frame if there exist constants $0 < A \leq B$ such that

$$A\|f\|_2^2 \leq \|f \star h_{\mathrm{out}}\|_2^2 + \sum_{\lambda\in\Lambda} \|f \star h_\lambda\|_2^2 \leq B\|f\|_2^2$$

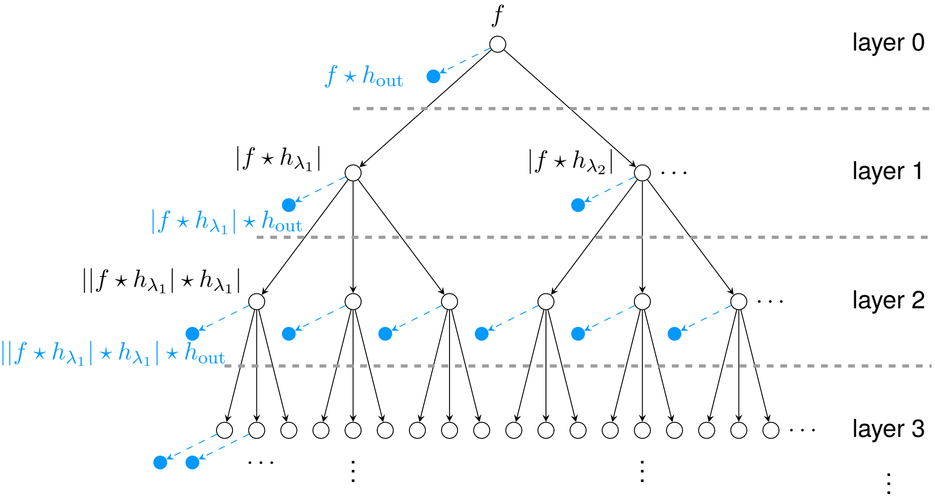
  holds for all $f \in L^2(\mathbb{R}^d)$.
- A semi-discrete frame is admissible if

$$B \leq 1.$$

Network Architecture

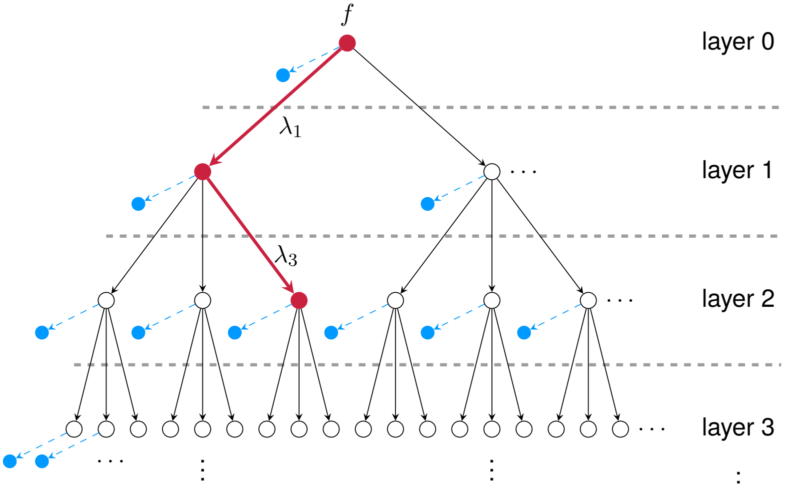


Network Architecture



- Feature vector $\mathrm{S}(f)$: set of all features

Network Architecture



- Path $p = (\lambda_1, \lambda_3)$ corresponds to $||f \star h_{\lambda_1}| \star h_{\lambda_3}| \star h_{\text{out}}$
- Path: ordered sequence of indices $p = (\lambda_1, \dots, \lambda_m)$
- Path length: $|p| = m$ (corresponds to layer m)
- Set of all paths: $\mathcal{P}$

Network Operators

- Feature corresponding to a path $p \in \mathcal{P}$:

$$\mathrm{S}_p(f) = f_p \star h_{\mathrm{out}}$$

- Feature vector: set of all features

$$\mathrm{S}(f) = \{\mathrm{S}_p(f)\}_{p \in \mathcal{P}}$$

- “Energy” of layer m:

$$\mathrm{E}_2(f; m) := \sqrt{\sum_{p \in \mathcal{P} : |p| = m} \|f_p\|_2^2}$$

- “Total network energy”

$$\mathrm{E}_1(f) := \sum_{m=0}^{\infty} \mathrm{E}_2(f; m)$$

Definition of Deformation Stability

- Input: $f \in L^2(\mathbb{R}^d)$
- Corresponding feature vector: $\mathrm{S}(f)$
- Deformed input: $\mathrm{T}_\tau f = f(\cdot - \tau(\cdot))$
- Corresponding feature vector: $\mathrm{S}(\mathrm{T}_\tau f)$
- Measure feature dissimilarity of feature vectors $\mathrm{S}(f)$ and $\mathrm{S}(\mathrm{T}_\tau f)$ by means of

$$\mathrm{E}_2(\mathrm{S}(\mathrm{T}_\tau f) - \mathrm{S}(f)) := \sqrt{\sum_{p \in \mathcal{P}} \| \mathrm{S}_p(\mathrm{T}_\tau f) - \mathrm{S}_p(f) \|_2^2}$$

- Deformation stability: We want to observe

$$\mathrm{E}_2(\mathrm{S}(\mathrm{T}_\tau f) - \mathrm{S}(f)) \to 0 \quad \text{for} \quad \|\tau\|_\infty \to 0$$

for a large set of f in $L^2(\mathbb{R}^d)$.

On Deformation Stability of General DCNNS

Theorem (Equation (24) of [WB15])
The feature extraction is Lipschitz, meaning,

$$\mathrm{E}_2(\mathrm{S}(g) - \mathrm{S}(f)) \leq \|g - f\|_2$$

holds for all $f, g \in L^2(\mathbb{R}^d)$.

- This is in particular true for $g = \mathrm{T}_\tau f$. Thus, if we can prove $\|\mathrm{T}_\tau f - f\|_2 \to 0$ if $\|\tau\|_\infty \to 0$ for a signal class $\mathcal{C}$, then we have deformation stability for $\mathcal{C}$.

Theorem (Theorem 2 of [WB15])
There exists a constant $C < \infty$ *such that for all* $f \in \mathcal{PW}^2_\sigma(\mathbb{R}^d)$, *we have*

$$\mathrm{E}_2(\mathrm{S}(\mathrm{T}_\tau f) - \mathrm{S}(f)) \leq C\sigma\|\tau\|_\infty\|f\|_2.$$

- A similar result exists for cartoon functions.

Definition of the Sobolev Space $H^2(\mathbb{R}^d)$

- The Sobolev space $H^2(\mathbb{R}^d)$ contains all functions $f \in L^2(\mathbb{R}^d)$ satisfying

$$\int_{\mathbb{R}^d} (1 + |\omega|^2)^2 |\hat{f}(\omega)|^2 \, \mathrm{d}\omega < \infty.$$

- Hence, f, all first-order derivatives, and all second-order derivatives of f are in $L^2(\mathbb{R}^d)$.
- The gradient ∇f of $f \in H^2(\mathbb{R}^d)$ exists and it holds

$$\|\nabla f\|_2 < \infty.$$

- The Sobolev space $H^2(\mathbb{R}^d)$ contains all bandlimited functions.

On Deformation Stability of General DCNNs

Theorem

For all $f \in H^2(\mathbb{R}^d)$, we have

$$\mathrm{E}_2(\mathrm{S}(\mathrm{T}_\tau f) - \mathrm{S}(f)) \leq 2^d \|\tau\|_\infty \|\nabla f\|_2.$$

- The Sobolev space $H^2(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$.

Theorem (Weak deformation stability)

For all $f \in L^2(\mathbb{R}^d)$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\forall \tau : \|\tau\|_\infty < \delta : \mathrm{E}_2(\mathrm{S}(\mathrm{T}_\tau f) - \mathrm{S}(f)) < \varepsilon.$$

- The last (weak deformation stability) result can also be derived by making use of the network structure. It might be possible to show (strong) deformation stability for all $f \in L^2(\mathbb{R}^d)$ by choosing a suitable semi-discrete frame.

Bounded Uniform Partition of Unity

Definition

A set of functions $\{\hat{\varphi}_j\}_{j\in I}$ on $\mathbb{R}$ 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 exists a compact set $U \subset \mathbb{R}$ with nonempty interior and points $y_i \in \mathbb{R}$ such that $\operatorname{supp}(\hat{\varphi}_j) \subset U + y_j$ for all $j \in I$
4. for each compact $K \subset \mathbb{R}$

$$\sup_{x\in\mathbb{R}} |\{j \in I : x \in K + y_j\}|$$
$$= \sup_{i\in I} |\{j \in I : K + y_i \cap K + y_j \neq \emptyset\}| < \infty.$$

Bounded Uniform Partition of Unity

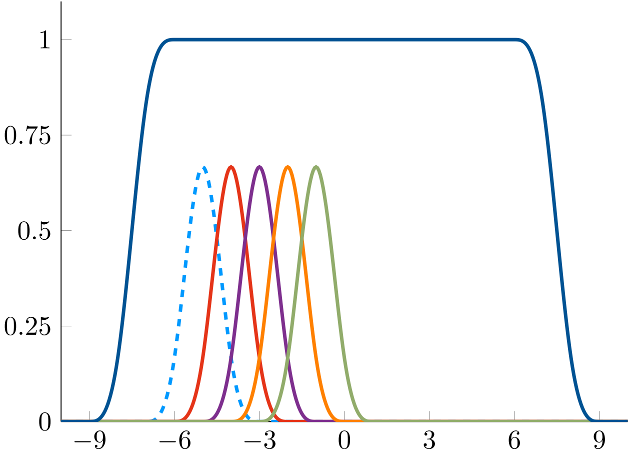


Figure: A BUPU.

Constructing Admissible Frames

- Let $\hat{f} \in L^2(\mathbb{R})$. The Wiener amalgam space $W(L^2, l^2)$ is equivalent to $L^2(\mathbb{R})$ and can be equipped with the norm

$$\|\hat{f}\|_{W(L^2,l^2)} = \sum_{j \in I} \|\hat{f}\hat{\varphi}_j\|_2^2.$$

- For a BUPU $\{\hat{\varphi}_j\}_{j \in I}$ norm equivalence holds:

$$\|\hat{f}\|_2^2 \asymp \|\hat{f}\|_{W(L^2,l^2)}^2 = \sum_{j \in I} \|\hat{f}\hat{\varphi}_j\|_2^2$$

- BUPUs are admissible semi-discrete frames:

Theorem

Let $\{\hat{\varphi}_j\}_{j \in I}$ be a BUPU. Then, there exist constants $0 < A \leq B < \infty$ such that

$$A\|f\|_2^2 \leq \sum_{j \in I} \|f \star \varphi_j\|_2^2 \leq B\|f\|_2^2, \quad \forall f \in L^2(\mathbb{R}).$$

Definition of B-Splines

- We define the box function

$$\hat{\beta}^{(0)}(\omega) = \begin{cases} 1, & -\frac{1}{2} < \omega < \frac{1}{2} \\ \frac{1}{2}, & |\omega| = \frac{1}{2} \\ 0, & \omega \notin [-\frac{1}{2}, \frac{1}{2}] \end{cases} .$$

- It holds:

$$\sum_{k \in \mathbb{Z}} \hat{\beta}^{(0)}(\omega - k) = 1, \quad \forall \omega \in \mathbb{R} .$$

- A B-spline of degree $n \geq 1$ is given by

$$\hat{\beta}^{(n)} := \underbrace{\hat{\beta}^{(0)} \star \hat{\beta}^{(0)} \star \cdots \star \hat{\beta}^{(0)}}_{(n+1) \text{ times}} .$$

- B-splines form a partition of unity:

$$\sum_{k \in \mathbb{Z}} \hat{\beta}^{(n)}(\omega - k) = 1, \quad \forall \omega \in \mathbb{R} .$$

Equivalent Definition of Semi-discrete Frame

- The condition

$$A\|f\|_2^2 \leq \|f \star h_{\text{out}}\|_2^2 + \sum_{\lambda \in \Lambda} \|f \star h_\lambda\|_2^2 \leq B\|f\|_2^2, \quad \forall f \in L^2(\mathbb{R})$$

is equivalent to

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

Admissible Frame: B-Splines of Degree $n = 3$

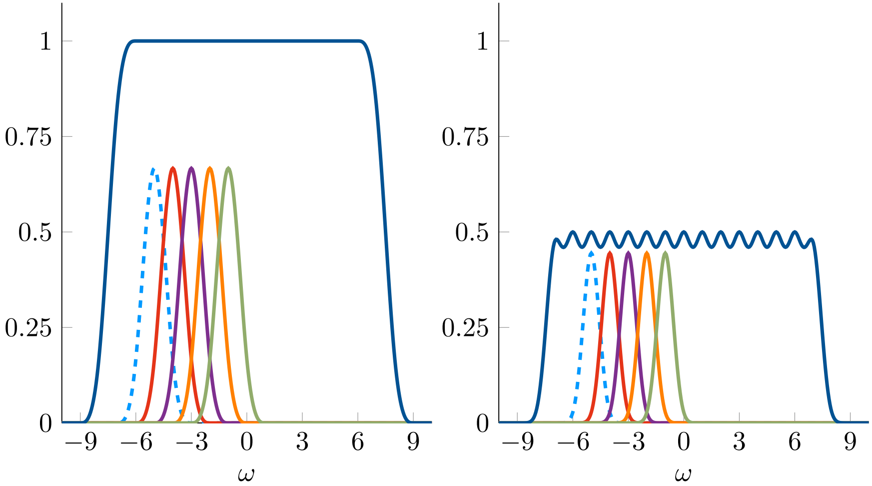
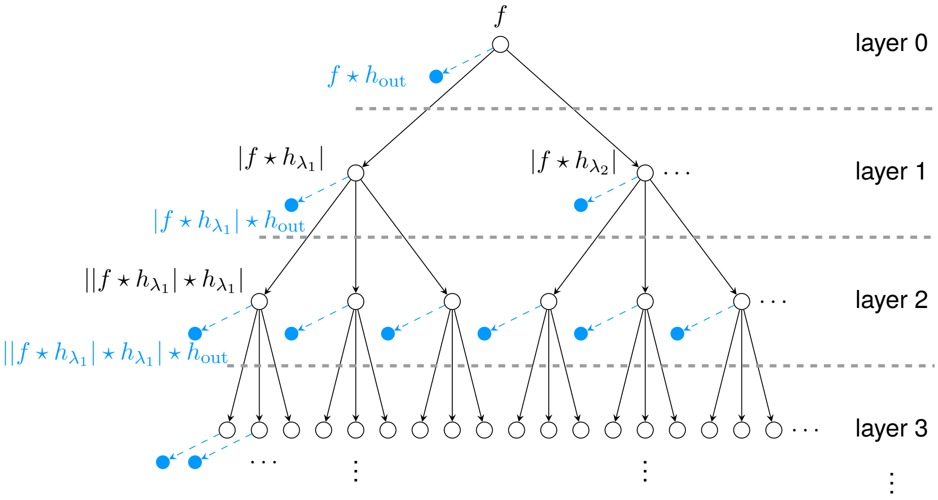


Figure: Left: B-splines of degree $n = 3$ form a BUPU. Right: B-splines of degree $n = 3$ form an admissible frame.

Outlook



f
layer 0
$f \star h_{\text{out}}$
$|f \star h_{\lambda_1}|$
$|f \star h_{\lambda_2}|$
...
layer 1
$|f \star h_{\lambda_1}| \star h_{\text{out}}$
$||f \star h_{\lambda_1}| \star h_{\lambda_1}|$
...
layer 2
$||f \star h_{\lambda_1}| \star h_{\lambda_1}| \star h_{\text{out}}$
...
layer 3

Outlook

- Incorporate other nonlinearities
- Incorporate pooling
- Use the network structure and properties of the semi-discrete frame to prove (strong) deformation stability on the whole $L^2(\mathbb{R}^d)$?
- The semi-discrete frame has a strong influence on the energy decay over consecutive layers. Find frames such that only few layers are necessary in an implementation. $\Rightarrow$ [WGB17]

References I

- Thomas Wiatowski and Helmut Bölcskei, *A mathematical theory of deep convolutional neural networks for feature extraction*, CoRR (2015).
- Thomas Wiatowski, Philipp Grohs, and Helmut Bölcskei, *Energy propagation in deep convolutional neural networks*, CoRR **abs/1704.03636** (2017).