ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Deep convolutional neural Networks based on semi-discrete Frames

Semester Paper

Tanja Almeroth

August 17, 2015

Advisor: Prof. Dr. Philipp Grohs
Seminar for Applied Mathematics, ETH Zürich

Abstract

Since deep convolutional neural networks based on semi-discrete frames have lead to breakthrough results in feature extraction applications, Wiatowski and Bölcskei developed Mallat's scattering network theory further. They introduced the generalized feature extractor, using different semi-discrete frames in each layer, while scattering networks are constructed by using the same wavelet-based semi-discrete frames in every layer. Furthermore the stability of the feature extractor with respect to deformations of R-band-limited signals is shown for a more general type of non-linear small deformations and translation invariance is proven, whereas the proofs are detached from many technical conditions relative to scattering networks. This semester paper explicates the construction, proofs and suggested applications, in particular wavelet-frames and Gabor frames, by Wiatowski and Bölcskei.

Acknowledgements

It is a pleasure to thank those who made this semester paper possible. This work would not have been possible without the guidance by my supervisor Prof. Dr. Philipp Grohs and the advice by Dr. Rima Alaifari.
Finally I want to thank my friend Dániel Bálint for proofreading my semester paper.

Contents

Contents **v**

1 Introduction **1**

2 Rapidly decreasing Functions and the Plancherel Theorem **5**

3 Deep convolutional neural Networks based on semi-discrete Frames **9**
3.1 Generalized Feature Extractor 9
3.2 Proof of Proposition 3.12 17
3.2.1 Main Part of the Proof 23

4 Mallat's wavelet-based Feature Extractor **25**
4.1 Construction of the wavelet-based Feature Extractor 25
4.2 Application of Theorem 3.10 to the wavelet-based Feature Extractor 26
4.3 Mallat's Results 28

5 Gabor Frames **29**
5.1 Semi-discrete Gabor Frames 29

6 Conclusion **31**
6.1 Applications of Theorem 3.10 31
6.2 Note on Comparison to Mallat's Result 31
6.3 Conculsion of this Work 32

A Appendix A **35**
A.1 ℓ^2-semi-discrete 35

B Appendix B **39**
B.1 Proof of Proposition 3.11 39
B.2 The Convolution Theorem 39

B.3 Proof of Equation 4.5 41

Bibliography **43**

Chapter 1

Introduction

Consider a function $f : \mathbb{R}^d \to \mathbb{C}$, as defined in mathematics, as a rule of assigning points $x \in \mathbb{R}^d \mapsto f(x)$. In one-dimensional ($d = 1$), for example physical, applications the variable x can be considered as time and $f(x)$ as voltage or electrical field. Then one speaks of a signal [5, Chapter 1]. In a two-dimensional application the map $(x_1, x_2) \mapsto f(x_1, x_2)$ can be thought of a assigning a pixel position (x_1, x_2) to a grey or colour level of an image $f(x_1, x_2)$ [5, Chapter 1]. One might be given the classification task, to determine if an image contains a certain handwritten digit [12]. Features in images are for example edges, lines or corners. However a central task in signal classification is feature extraction [12].
A difficulty is that features should be independent of spatial location, which results into a necessity for translation invariance of a feature extractor [12]. Furthermore, a claim for robustness with respect to different hand writing styles leads to stability with respect to small deformations of a signal [12]. Spectacular success in many practical classification tasks for feature extractors were generated by deep convolutional neural networks [12]. The mathematical analysis of such networks was invented by Mallat, whose theory applies to so-called scattering networks [12]. Scattering networks propagate signals through layers that compute certain coefficients, resulting into a feature extractor, which can be shown to be stable with respect to deformations and probably translation invariant [12]. This leads to state of the art results in image classification [12]. Since the wavelet transform, which is used in Mallat's work [7], is not efficient in dealing with signal domains by anisotropic features, like edges in images [12] and a more general approach seems to be desirable in practical applications, Wiatowski and Bölcskei extended Mallat's scattering networks in [12]. The general feature extractor introduced by Wiatowski and Bölcskei, which uses different signal transformations in different layers, is shown to be translation invariant and stable with respect to small deformations, while proof techniques are detached

from the underlying transformations structural properties.
This semester paper mainly follows the paper [12]. In chapter 2 basic results from Fourier theory are introduced. Therefore in chapter 3 the feature extractor's construction and proofs, which show the translation invariance and stability with respect to small deformations, from Wiatowski and Bölcskei are reproduced. Furthermore, a proof of [12, Proposition 2], which is not given in [12] but essentially follows [7, Proposition 2.5] is shown in section 3.2. Mallat's original feature extractor given in [7] is reviewed briefly in chapter 4 and as one more example of an application of the main theorem of [12] semi-discrete Gabor frames are considered in chapter 5.

Definitions and Abbreviations

The following notations and abbreviations are introduced:

The complex conjugate of $z \in \mathbb{C}$ is denoted by $\bar{z}$.

The Euclidean inner product of $x, y \in \mathbb{C}^d$ is $\langle x, y \rangle := \sum_{j=1}^{d} x_j \overline{y_j}$.

The Euclidean norm on $\mathbb{R}^d$ of $x \in \mathbb{R}^d$ is $\|x\| := \langle x, x \rangle^{1/2}$.

The open ball of radius $R > 0$ centred at x in $\mathbb{R}^d$ is $B_R(x)$.

The Borel σ-Algebra of $\mathbb{R}^d$ is denoted by $\mathbb{B}$.

For a $\mathbb{B}$-measurable function $f : \mathbb{R}^d \to \mathbb{C}$, $\int_{\mathbb{R}^d} f(x)dx$ is the integral of f with respect to the Lebesgue measure μ_L.

For $p \in [1, \infty)$, $L^p(\mathbb{R}^d)$ denotes the space of $\mathbb{B}$-measurable functions $f : \mathbb{R}^d \to \mathbb{C}$ such that $\|f\|_p := (\int_{\mathbb{R}^d} |f(x)|^p dx)^{\frac{1}{p}} < \infty$.

For $\mathbb{B}$-measurable functions the essential supremum is denoted by $\|f\|_{L^\infty} := \text{ess sup} f = \inf\{M \in \mathbb{R}_{\geq 0} | \mu_L(\{x \in \mathbb{R}^d | |f(x)| > M\}) = 0\}$.

For $f, g \in L^2(\mathbb{R}^d)$, $\langle f, g \rangle_2 := \int_{\mathbb{R}^d} f(x)\overline{g(x)}dx$ is the inner product on $L^2(\mathbb{R}^d)$.

The Fourier transform of $f(x) \in L^1(\mathbb{R}^d)$ is $\hat{f}(\omega) = \widehat{f(x)} := \int_{\mathbb{R}^d} f(x) e^{-2\pi i \langle x, \omega \rangle} dx$. It is extended on $L^2(\mathbb{R}^d)$ by the Plancherel theorem (see Chapter 2).

The convolution of $f \in L^2(\mathbb{R}^d)$ and $g \in L^1(\mathbb{R}^d)$ is $f \star g(\omega) = \int_{\mathbb{R}^d} f(x)g(\omega - x)dx = \int_{\mathbb{R}^d} f(\omega - x)g(x)dx$.

The norm of an operator $A : L^p(\mathbb{R}^d) \to L^q(\mathbb{R}^d)$ is $\|A\|_{p,q} := \sup_{f \in L^p(\mathbb{R}^d) \setminus \{0\}} \frac{\|Af\|_q}{\|f\|_p}$.

$Id : L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)$ denotes the identity operator on $L^p(\mathbb{R}^d)$.

$C^\infty(\mathbb{R}^d) = \bigcup_{k=1}^{\infty} C^k(\mathbb{R}^d)$ is the vector space of all smooth functions from $\mathbb{R}^d$ to $\mathbb{C}$.

$C_c^\infty(\mathbb{R}^d)$ is the space of all compactly supported, smooth functions from $\mathbb{R}^d$ to $\mathbb{C}$.

$C_c(\mathbb{R}^d)$ is the space of all compactly supported functions from $\mathbb{R}^d$ to $\mathbb{C}$.

For a scalar function $v : \mathbb{R}^d \to \mathbb{C}$ the supremum norm is $\|v\|_\infty := \sup_{x\in\mathbb{R}^d} |v(x)|$.

For a vector field $\tau : \mathbb{R}^d \to \mathbb{R}^d$ the norm is $\|\tau\|_\infty := \sup_{x\in\mathbb{R}^d} \|\tau(x)\|$.

$D\tau$ denotes the Jacobian matrix and the associated norm is $\|D\tau\|_\infty := \sup_{x\in\mathbb{R}^d} |D\tau(x)|_\infty$, where for a matrix $M \in \mathbb{R}^{d\times d}$, $|M|_\infty := \sup_{1\leq,j,k\leq d} |M_{j,k}|$ is the supremum norm.
$D^2\tau$ denotes the Jacobian tensor of τ and the associated norm is $\|D^2\tau\|_\infty := \sup_{x\in\mathbb{R}^d} |D^2\tau(x)|_\infty$, where for a tensor $T \in \mathbb{R}^{d\times d\times d}$ $|T|_\infty := \sup_{1\leq,j,k,l\leq d} |T_{j,k,l}|$ is the supremum norm.

$\mathbb{1} \in \mathbb{R}^{d\times d}$ denotes the identity matrix.

For a matrix $M \in \mathbb{R}^{d\times d}$ $M^\dagger$ denotes the transposed matrix.

The translation operator $T_t : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is $T_t f(x) := f(x - t), t \in \mathbb{R}^d$.

The involution operator $I : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is $If(x) := \overline{f(-x)}$.

The modulation operator $M_\omega : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is $M_\omega f(x) := e^{2\pi i\langle x,\omega\rangle} f(x)$, for $\omega \in \mathbb{R}^d$.

The dilatation operator $D_a : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is $D_a f(x) := \frac{1}{\sqrt{|a|}} f(\frac{x}{a})$, for $a \in \mathbb{R} \setminus \{0\}$.

Chapter 2

Rapidly decreasing Functions and the Plancherel Theorem

This chapter gives an basic introduction to Schwartz space, where results which are used later, are presented.

Schwartz space [4, Remark 2.2.8], also called the space of rapidly decreasing functions, is defined like in [9, Section 7.3] as in the following:

Definition 2.1 *Schwartz space is*

$$\mathcal{S}(\mathbb{R}^d) := \{f \in C^\infty(\mathbb{R}^d) | \sup_{|\alpha| \leq N} \sup_{x \in \mathbb{R}^d} (1 + \|x\|^2)^N |(D_\alpha f)(x)| < \infty, \forall N \in \mathbb{N}_0\}, \tag{2.1}$$

where α is a multi-index, i.e.,

$$\begin{aligned} \alpha &:= (\alpha_1, \alpha_2, \cdots, \alpha_n) \quad \alpha_k \in \mathbb{N}_0 \quad \textit{for all} \quad 1 \leq k \leq n, \\ |\alpha| &:= \sum_{k=1}^{n} \alpha_k, \\ D_\alpha f &:= \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} f. \end{aligned}$$

The definition above might be described in other words by that it is required that for functions $f \in \mathcal{S}(\mathbb{R}^d)$, for every polynomial P and multi-index α we have $P \cdot D_\alpha f$ is a bounded function [9, Chapter 7].

Proposition 2.2 *Let $P : \mathbb{R}^d \to \mathbb{C}$ be a polynomial, $g \in \mathcal{S}(\mathbb{R}^d)$ and let α be a multi-index then*

$$f \mapsto Pf, \quad f \mapsto gf \quad \textit{and} \quad f \mapsto D_\alpha f$$

are continuous maps from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}(\mathbb{R}^d)$.

A proof can be found in [9, Theorem 7.4].

Remark 2.3

(i.) *Schwartz space is a sub vector space of $C^\infty(\mathbb{R}^d)$, because the derivative is linear. And we have $\|f\|_\infty < \infty$ for all $f \in \mathcal{S}(\mathbb{R}^d)$.*

(ii.) *Every compactly supported and smooth function is a Schwartz function, i.e., $C_c^\infty(\mathbb{R}^d) \subseteq \mathcal{S}(\mathbb{R}^d)$ [4, Example 2.2.2]. Roughly speaking this follows from the fact that every smooth compactly supported function, multiplied by a polynomial is still compactly supported and smooth and therefore bounded in the sense of the definition of Schwartz functions.*

A more detailed explication of the above remark can be found in [4, Chapter 2.3.1].

Proposition 2.4

(i.) *$\mathcal{S}(\mathbb{R}^d)$ is dense in $L^1(\mathbb{R}^d)$ and therefore $\|f\|_1 < \infty$ for all $f \in \mathcal{S}(\mathbb{R}^d)$.*

(ii.) *$\mathcal{S}(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$ and therefore $\|f\|_2 < \infty$ for all $f \in \mathcal{S}(\mathbb{R}^d)$.*

Proof Techniques in this proof, using density arguments, are analogously borrowed from [9, Proof of Theorem 7.5 and Proof of Theorem 7.8].
Because $C_c^\infty(\mathbb{R}^d) \subseteq \mathcal{S}(\mathbb{R}^d)$ and $C_c^\infty(\mathbb{R}^d)$ is dense in $L^1(\mathbb{R}^d)$, we have that $\mathcal{S}(\mathbb{R}^d)$ is dense in $L^1(\mathbb{R}^d)$.
By proposition 2.2 for every $f \in \mathcal{S}(\mathbb{R}^d)$ we have $f^2 \in S(\mathbb{R}^d)$. From the density of Schwartz space in the space of integrable functions $L^1(\mathbb{R}^d)$ it follows that f^2 is an element of $L^1(\mathbb{R}^d)$ and therefore f is squared integrable. Hence it is obtained that $\mathcal{S}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$.
Furthermore it is known that $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$ and therefore $\mathcal{S}(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) = \mathcal{S}(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$. □

The following result can be found in [9, Theorem 7.4 (d), Proof of Theorem 7.9].

Theorem 2.5 *The Fourier transform is a continuous, linear mapping from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}(\mathbb{R}^d)$ furthermore it is unitary with respect to the L^2-inner-product, i.e.,*

$$\langle f, g\rangle_2 = \langle \hat{f}, \hat{g}\rangle_2 \quad \text{for all} \quad f, g \in \mathcal{S}(\mathbb{R}^d) \tag{2.2a}$$

and it preserves the L^2-norm, i.e.,

$$\|\hat{f}\|_2 = \|f\|_2 \quad \text{for all} \quad f \in \mathcal{S}(\mathbb{R}^d). \tag{2.2b}$$

Equation (2.2b) is called Plancherel formula. Furthermore, the Plancherel theorem extends the Fourier transform on $L^2(\mathbb{R}^d)$.

Theorem 2.6 *(Plancherel)[9, Theorem 7.9] There is a linear isometry Ψ of $L^2(\mathbb{R}^d)$ onto $L^2(\mathbb{R})$ which is uniquely determined by the requirement that*

$$\Psi f = \hat{f} \quad \textit{for all} \quad f \in \mathcal{S}(\mathbb{R}^d). \tag{2.3}$$

Remark 2.7 *Note that every linear isometry which maps a vector space onto itself is continuous, since continuity is equivalent to boundedness for linear operators, which map a normed vector space to a normed vector space, as known from functional analysis [10, Satz 2.2.1.].*

Rudin comments the Plancherel-theorem in [9, Chapter 7.9] the following way:

> 'That the Fourier Transform is an L^2-isometry is one of the most important features of the whole subject.'

Remark 2.8 *In the following the extension of the Fourier transform Ψ on $L^2(\mathbb{R}^d)$ still will be named Fourier transform and denoted like the Fourier transform, i.e, $\Psi f := \hat{f}$ for all $f \in L^2(\mathbb{R}^d)$.*

Chapter 3

Deep convolutional neural Networks based on semi-discrete Frames

This chapter introduces the generalized feature extractor based on semi discrete frames and shows its translation invariance and stability with respect to deformations.

3.1 Generalized Feature Extractor

Definition 3.1 (Semi-discrete frames) *[12] Let $\{f_\lambda\}_{\lambda\in\Lambda} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ be a set of functions indexed by a countable set Λ. The set of translated and involved functions*

$$\Psi_\Lambda = \{T_b I f_\lambda\}_{(\lambda,b)\in\Lambda\times\mathbb{R}^d} \tag{3.1a}$$

is a semi-discrete frame, if there exist constants $A, B > 0$ such that

$$A\|g\|_2^2 \leq \sum_{\lambda\in\Lambda} \|g \star f_\lambda\|_2^2 \leq B\|g\|_2^2 \tag{3.1b}$$

for all $g \in L^2(\mathbb{R}^d)$. The functions $\{f_\lambda\}_{\lambda\in\Lambda}$ are called the atoms of the semi-discrete frame Ψ_Λ.
When $A = B$ the semi-discrete frame is said to be tight.
A tight semi-discrete frame with frame bound $A = 1$ is called semi-discrete Parseval frame.

The set Λ typically labels a collection of scales, directions or frequency-shifts [12].

Definition 3.2 (Frame collection) *[12] For all $n \in \mathbb{N}$, let Ψ_n be a semi-discrete frame with frame bounds $A_n, B_n > 0$ and atoms $\{f_{\lambda'_n}\}_{\lambda'_n\in\Lambda'_n} \subseteq L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ indexed by a countable set Λ'_n. The sequence*

$$\Psi = \left(\Psi_n\right)_{n\in\mathbb{N}} \tag{3.2}$$

is called frame collection with frame bounds $A = \inf_{n \in \mathbb{N}} A_n$ *and* $B = \sup_{n \in \mathbb{N}} B_n$.

Elements $\Psi_n, n \in \mathbb{N}$ correspond to particular layers in the generalized scattering network [12], which will be introduced in the following (see (3.5)).

One of the atoms in $\{f_{\lambda'_n}\}_{\lambda'_n \in \Lambda'_n}$ in each frame Ψ_n in the frame collection Ψ is set as *output-generating atom* [12]. This means in the semi-discrete frames Ψ_n for an arbitrary, but fixed, $\lambda_n^{*\prime}$ the atoms are renamed such that the semi-discrete frame writes as [12]

$$\begin{aligned} \varphi_n &:= f_{\lambda_n{}^*}, \quad \Lambda_n := \Lambda'_n \setminus \{\lambda_n^*\}, \\ \Psi_n &= \{T_b I \varphi_n\}_{b \in \mathbb{R}^d} \cup \{T_b I f_{\lambda_n}\}_{(\lambda_n, b) \in \Lambda_n \times \mathbb{R}^d}. \end{aligned} \tag{3.3}$$

Definition 3.3 (Modulus-convolution operator) *[12] Let* $\Psi = (\Psi_n)_{n \in \mathbb{N}}$ *be a frame collection with atoms* $\{\varphi_n\} \cup \{f_{\lambda_n}\}_{\lambda_n \in \Lambda_n}$. *For* $1 \leq m < \infty$, *define the set*

$$\Lambda_1^m := \Lambda_1 \times \Lambda_2 \times \cdots \times \Lambda_m. \tag{3.4a}$$

An ordered sequence

$$q = (\lambda_1, \lambda_2, \cdots, \lambda_m) \in \Lambda_1^m \tag{3.4b}$$

is called a path.
The empty path, $e := \emptyset$, *defines the set* $\Lambda_1^0 := \{e\}$.
The modulus-convolution operator is defined as $U : \left(\bigcup_{k=1}^{\infty} \Lambda_k\right) \times L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$,

$$U(\lambda_n, f) := U[\lambda_n] f := |f \star f_{\lambda_n}|, \tag{3.4c}$$

where $f_{\lambda_n} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ *are the atoms of the semi-discrete frame* Ψ_n, *associated in the n-th layer of the so-called scattering network (see figure 3.1).*

Furthermore, the modulus convolution operator is extended on paths $q \in \Lambda_1^m$ to a multi stage filtering operator [12] as follows:

$$U[q]f := U[\lambda_m] \cdots U[\lambda_2] U[\lambda_1] f = |\cdots||f \star f_{\lambda_1}| \star f_{\lambda_2}| \star \cdots \star f_{\lambda_m}| \tag{3.5}$$

On the empty path it is defined that $U[e]f := f$.

Proposition 3.4 *The modulus convolution operator* $U : \Lambda_1^m \times L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ *is well-defined.*

Proof The well-defines follows from Young's inequality [12], which is stated in [4, Theorem 1.2.13] as follows:

Theorem 3.5 (Young's inequality) *Let* $1 \leq p, q, r \leq \infty$ *satisfy*

$$\frac{1}{q} + 1 = \frac{1}{p} + \frac{1}{r}.$$

Then for all $f \in L^p(\mathbb{R}^d)$ *and* $g \in L^r(\mathbb{R}^d)$ *we have*

$$\|f \star g\|_q \leq \|g\|_r \|f\|_p.$$

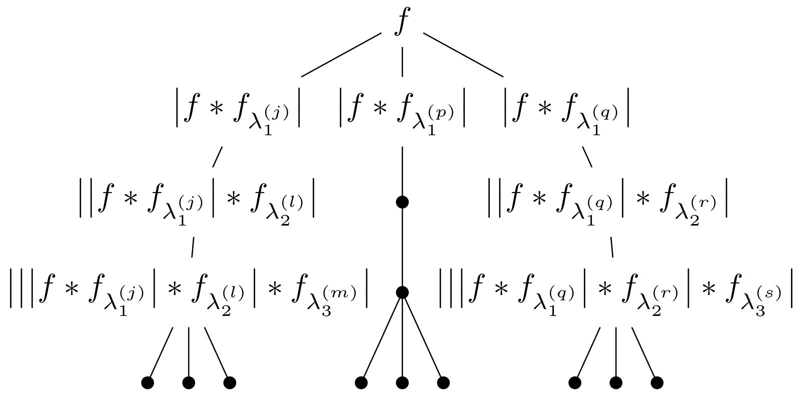

Figure 3.1: Scattering network architecture based on general multistage filtering (3.5). The function $f_{\lambda_n^{(k)}}$ is the k-th atom of the semi-discrete frame Ψ_n associated with the n-th layer. Source: [12].

Let $q = (\lambda_1, \lambda_1, \cdots, \lambda_m) \in \Lambda_1^m$. Hence applying Young's inequality for $q = p = 2$ and $r = 1$, $f \in L^2(\mathbb{R}^d)$ and because $f_{\lambda_n} \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ for all $\lambda_n \in \Lambda_n, n \in \mathbb{N}$ it follows as in [12] that

$$\begin{aligned}\|U[q]f\|_2 &= \||\,|U[(\lambda_1, \lambda_2, \cdots, \lambda_{m-1})]f| \star f_{\lambda_m}|\|_2 \\ &\leq \|U[(\lambda_1, \lambda_2, \cdots, \lambda_{m-1})]f\|_2 \|f_{\lambda_m}\|_1 \\ &\leq \cdots \leq \|f\|_2 \prod_{n=1}^{m} \|f_{\lambda_n}\|_1 < \infty. \qquad \square\end{aligned}$$

Definition 3.6 (Feature extractor) *[12] Let $\Psi = \left(\Psi_n\right)_{n\in\mathbb{N}}$ be a frame collection and define $\mathcal{Q} := \bigcup_{k=0}^{\infty} \Lambda_1^k$. Given a path $q \in \Lambda_1^n, n \geq 0$, we write*

$$\varphi[q] := \varphi_{n+1} \tag{3.6a}$$

for the output-generating atom of the semi-discrete frame Ψ_{n+1}. The feature extractor $\Phi_\Psi : L^2(\mathbb{R}^d) \to \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ with respect to the frame collections Ψ is defined as

$$\Phi_\Psi(f) := \{U[q]f \star \varphi[q]\}_{q\in\mathcal{Q}}. \tag{3.6b}$$

Definition 3.7 (ℓ^2-semi-discrete) *For any countable set $\mathcal{Q}$, we define a $\mathbb{C}$-Hilbertspace denoted by $\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ as*

$$s := \{f_q\}_{q\in\mathcal{Q}}, \quad f_q \in L^2(\mathbb{R}^d) \quad \textit{for all} \quad q \in \mathcal{Q}, \tag{3.7a}$$

$$|||s||| := \Big(\sum_{q\in\mathcal{Q}} \|f_q\|_2^2, \Big)^{1/2}, \tag{3.7b}$$

$$\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}} := \{\{f_q\}_{q\in\mathcal{Q}} | f_q \in L^2(\mathbb{R}^d) \forall q \in \mathcal{Q}, |||\{f_q\}_{q\in\mathcal{Q}}||| < \infty\}, \tag{3.7c}$$

where the addition and scalar multiplication are given by

$$\begin{aligned} s_1 + s_2 &= \{f_q\}_{q\in\mathcal{Q}} + \{g_q\}_{q\in\mathcal{Q}} \\ &:= \{f_q + g_q\}_{q\in\mathcal{Q}} \quad \textit{for all} \quad s_1, s_2 \in \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}, \\ \alpha s = \alpha\{f_q\}_{q\in\mathcal{Q}} &:= \{\alpha f_q\}_{q\in\mathcal{Q}} \quad \textit{for all} \quad s \in \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}, \alpha \in \mathbb{C}. \end{aligned}$$

A proof, which shows that $\left(L^2(\mathbb{R})\right)^{\mathcal{Q}}$ is a Hilbertspace with norm $|||\cdot|||$, can be found in appendix A.1.

That $\Phi_\Psi(f)$ is bounded and therefore the feature extractor is a well defined operator will be shown in section 3.2.

Definition 3.8 (Non-linear deformations) *[12] Let $\omega : \mathbb{R}^d \to \mathbb{R}$. The time-frequency deformation operator with respect to $\tau \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ is*

$$F_{\tau,\omega} f(x) := e^{2\pi i \omega(x)} f(x - \tau(x)). \tag{3.8}$$

As it is proven in [12, Appendix B] the deformation operator maps L^2-functions to L^2-functions if the derivative of the deformation is small.

Remark 3.9 *The deformation operator $F_{\tau,\omega} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is well-defined for $\|D\tau\|_\infty \leq \frac{1}{2d}$.*

Proof Let $f \in L^2(\mathbb{R}^d)$, then

$$\|F_{\tau,\omega} f\|_2^2 = \int_{\mathbb{R}^d} |f(x - \tau(x))|^2 dx = \int_{\mathbb{R}^d} \frac{1}{|det(Id - D\tau(x))|} |f(u)|^2 du,$$

where the substitution $x - \tau(x) = u$ was used. Furthermore, applying [1, Eq. 1] and the assumption $\|D\tau\|_\infty \leq \frac{1}{2d}$ the modulus of the Jacobian determinate can be bounded by

$$|det(Id - D\tau(x))| \geq 1 - d\|D\tau\|_\infty \geq \frac{1}{2}.$$

Hence it follows that

$$\|F_{\tau,\omega} f\|_2^2 \leq 2\|f\|_2^2.$$

□

Now the main result of [12] can be stated.

Theorem 3.10 (Main theorem of [12]) *Let Ψ be a frame collection with upper frame bound $B \leq 1$. The feature extractor Φ_Ψ defined in (3.6b) is translation invariant, i.e.,*

$$\Phi_\Psi(T_t f) = T_t(\Phi_\Psi(f)) \quad \textit{for all} \quad t \in \mathbb{R}^d, \quad f \in L^2(\mathbb{R}^d), \tag{3.9a}$$

where T_t is applied element wise in $T_t(\Phi_\Psi(f))$. Furthermore, for $R > 0$, define the space of R-band-limited functions

$$H_R := \{f \in L^2(\mathbb{R}^d) | supp(\hat{f}) \subseteq B_R(0)\}. \tag{3.9b}$$

Then, the feature extractor Φ_Ψ is stable on H_R with respect to non-linear deformations (3.8), i.e, there exists a $C > 0$ such that for all $f \in H_R$ and all $\omega \in C(\mathbb{R}^d, \mathbb{R})$, $\tau \in C(\mathbb{R}^d, \mathbb{R}^d)$ with $\|D\tau\|_\infty \leq \frac{1}{2d}$, it holds that

$$|||\Phi_\Psi(f) - \Phi_\Psi(F_{\tau,\omega} f)||| \leq C(R\|\tau\|_\infty + \|\omega\|_\infty)\|f\|_2. \tag{3.9c}$$

Proof This proof is analogously borrowed from [12, Appendix B]. It consists of three parts; showing translation invariance, showing a generalization of Mallat's proposition [7, 2.5] in order to gain Lipschitz continuity and using a partition of unity argument in order to show stability with respect to deformations.

The following proposition is used in order to show translation invariance (3.9a) of the feature extractor Φ_Ψ.

Proposition 3.11 *Let $f \in L^2(\mathbb{R}^d)$, $g \in L^1(\mathbb{R}^d)$ and T_t be the translation operator. Then*

$$T_t(f \star g) = T_t f \star g \quad \text{for all} \quad t \in \mathbb{R}^d.$$

A proof can be found in appendix B.1.

Let $f \in L^2(\mathbb{R}^d)$. Since the translation operator is applied element wise and because of the definition of the feature extractor Φ_Ψ (3.6b), it follows that translation invariance (3.9a) is equivalent to

$$U[q](T_t f) \star \varphi[q] = T_t(U[q]f \star \varphi[q]) \quad \text{for all} \quad t \in \mathbb{R}^d, \quad q \in \mathcal{Q}.$$

The right-hand-side of the above equation equals, thanks to proposition (3.11),

$$T_t(U[q]f \star \varphi[q]) = T_t(U[q]f) \star \varphi[q].$$

Hence it suffices to show that the modulus-convolution operator $U[q]$ is translation invariant, i.e.,

$$T_t(U[q]f) = U[q](T_t f) \quad \text{for all} \quad t \in \mathbb{R}^d, \quad q \in \mathcal{Q}.$$

Let $\lambda_n \in \bigcup_{k=1}^{\infty} \Lambda_k$ then due to (3.4c) and proposition 3.11 it can be written,

$$U[\lambda_n](T_t f) = |(T_t f) \star f_{\lambda_n}| = |T_t(f \star f_{\lambda_n})| = T_t|f \star f_{\lambda_n}| = T_t(U[\lambda_n]f).$$

So for $q = (\lambda_1, \lambda_2, \cdots, \lambda_m) \in \mathcal{Q}$ by (3.5) it is obtained that

$$\begin{aligned} U[q](T_t f) &= U[\lambda_m] \cdots U[\lambda_1](T_t f) = U[\lambda_m] \cdots U[\lambda_2] T_t(U[\lambda_1] f) \\ &= T_t(U[\lambda_m] \cdots U[\lambda_1] f) = T_t(U[q] f), \end{aligned}$$

which shows that $U[q]$ is translation invariant.

The proof of the stability estimate (3.9c) requires the following proposition, which is a generalization [12] of [7, Proposition 2.5].

Proposition 3.12 *Let Ψ be a frame collection with upper frame bound $B \leq 1$. The feature extractor $\Phi_\Psi : L^2(\mathbb{R}^d) \to \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ is a bounded Lipschitz-continuous operator with Lipschitz-constant $L = \sqrt{B}$, i.e.,*

$$|||\Phi_\Psi(f) - \Phi_\Psi(g)||| \leq \sqrt{B} \|f - g\|_2 \quad \textit{for all} \quad f, g \in L^2(\mathbb{R}^d). \tag{3.10}$$

A proof can be found in section 3.2.
Thanks to remark 3.9 we have $F_{\tau,\omega} f \in L^2(\mathbb{R}^d)$ and due to the assumption $B \leq 1$ it follows that for every $f \in L^2(\mathbb{R}^d)$ it holds that

$$|||\Phi_\Psi(f) - \Phi_\Psi(F_{\tau,\omega} f)||| \leq \sqrt{B} \|f - F_{\tau,\omega} f\|_2 \leq \|f - F_{\tau,\omega} f\|_2.$$

Let $f \in H_R$. An upper bound for $\|f - F_{\tau,\omega} f\|_2$ can be derived by a partition of unity argument. In particular, a function γ such that $f \star \gamma = f$ for all R-band-limited functions f will be constructed.
Due to a partition of unity argument [8, Theorem 2.13] there exists an $\eta \in \mathcal{S}(\mathbb{R}^d)$ such that

$$\hat{\eta}(\omega) = 1 \quad \text{for all} \quad \omega \in B_1(0). \quad \text{Define} \quad \gamma(x) = R^d \eta(Rx).$$

Furthermore, due to [9, Theorem 7.3] it can be written that $\hat{\gamma}(\omega) = \hat{\eta}(\frac{\omega}{R})$. It follows that

$$\hat{\gamma}(\omega) = \hat{\eta}(\frac{\omega}{R}) = 1 \quad \text{for all} \quad \omega \in B_R(0).$$

Since for all $f \in H_R$ we have $supp(\hat{f}) \subseteq B_R(0)$ it can be stated that

$$\hat{f} = \hat{f}\hat{\gamma}.$$

And in consequence of the convolution theorem (Appendix B.2) the identity

$$f = f \star \gamma \tag{3.11}$$

is obtained. Next the operator $A_\gamma : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is defined by

$$A_\gamma(f) = f \star \gamma.$$

A_γ is well defined since $\gamma \in \mathcal{S}(\mathbb{R}^d) \subseteq L^1(\mathbb{R}^d)$ and therefore Young's inequality is valid. Due to (3.11) for all $f \in H_R$ the bound

$$\|f - F_{\tau,\omega} f\|_2 = \|A_\gamma f - F_{\tau,\omega} A_\gamma f\|_2 \leq \|A_\gamma - F_{\tau,\omega} A_\gamma\|_{2,2} \|f\|_2$$

is obtained.

In order to bound the operator norm $\|A_\gamma - F_{\tau,\omega} A_\gamma\|_{2,2}$ Schur's lemma is needed. Schur's lemma makes use of the space of locally integrable functions. It is defined according to [4, Definition 1.1.15] as:

Definition 3.13 (Locally integrable) *The space of locally integrable functions on $\mathbb{R}^d$, denoted by $L^1_{loc}(\mathbb{R}^d)$, is the set of all $\mathbb{B}$-measurable functions $f : \mathbb{R}^d \to \mathbb{C}$, that satisfy*

$$\int_K |f(x)| dx < \infty \quad \textit{for any compact subset } K \subseteq \mathbb{R}^d.$$

Note that $L^1(\mathbb{R}^d) \subseteq L^1_{loc}(\mathbb{R}^d)$. This follows from the fact that the integral over any compact subset $K \subset \mathbb{R}^d$ is less or equal then the integral over $\mathbb{R}^d$.

Theorem 3.14 (Schur's lemma) *[4, Appendix I.1] Let $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ be a locally integrable function satisfying*

$$\sup_{u \in \mathbb{R}^d} \int_{\mathbb{R}^d} |k(x,u)| dx \leq D \quad \textit{and} \quad \sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^d} |k(x,u)| du \leq D.$$

Then the integral operator K given by

$$K(f)(x) = \int_{\mathbb{R}^d} f(u) k(x,u) du, \tag{3.12a}$$

is a bounded operator from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$ with norm $\|K\|_{2,2} \leq D$.

In order to obtain an integral operator K such that $\|K\|_{2,2} = \|A_\gamma - F_{\tau,\omega} A_\gamma\|_{2,2}$, it is written that

$$\begin{aligned} F_{\tau,\omega} A_\gamma(f)(x) &= F_{\tau,\omega}(f \star \gamma)(x) \\ &= F_{\tau,\omega} \int_{\mathbb{R}^d} f(u) \gamma(x-u) du \\ &= e^{2\pi i \omega(x)} \int_{\mathbb{R}^d} f(u) \gamma(x - \tau(x) - u) du. \end{aligned}$$

So the kernel-function $k(x,u)$ is defined by

$$k(x,u) := e^{2\pi i \omega(x)} \gamma(x - \tau(x) - u) - \gamma(x-u).$$

Note that

$$|k(x,u)| = |\gamma(x - \tau(x) - u) - \gamma(x-u)|.$$

Hence it follows that $k(x,u)$ is locally integrable, because $\gamma \in \mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{S}(\mathbb{R}^{2d}) \subseteq L^1(\mathbb{R}^{2d}) \subseteq L^1_{loc}(\mathbb{R}^{2d})$. By applying the change of variable formula and the assumption that $\tau \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, it can be shown that the above right-hand side is also an element in the space of locally integrable functions $L^1_{loc}(\mathbb{R}^{2d})$.
Furthermore, in order to bound $|k(u,x)|$ a first-order Taylor-expansion is applied. Define $h^{x,u} : \mathbb{R} \to \mathbb{C}$,

$$h^{x,u}(t) := e^{2\pi i t\omega(x)}\gamma(x - t\tau(x) - u) - \gamma(x-u).$$

On the one hand $h^{x,u}(0) = 0$, so it follows, that

$$\begin{aligned} h^{x,u}(t) &= h^{x,u}(0) + \int_0^t (\frac{d}{d\lambda}h^{x,u})(\lambda)d\lambda \\ &= \int_0^t \frac{d}{d\lambda}h^{x,u}(\lambda)d\lambda. \end{aligned}$$

And on the other hand because of the identity $h^{x,u}(1) = k(x,u)$ it can be written that

$$|k(x,u)| = |h^{x,u}(1)| \leq \int_0^1 |\frac{d}{d\lambda}h^{x,u}(\lambda)|d\lambda.$$

Now the chain rule, triangle inequality and Cauchy-Schwarz inequality are used in order to bound the modulus of $\frac{d}{d\lambda}h^{x,u}(\lambda)$:

$$\begin{aligned} |\frac{d}{d\lambda}h^{x,u}(\lambda)| &= |\frac{d}{d\lambda}e^{2\pi i\lambda\omega(x)}\gamma(x - \lambda\tau(x) - u) - \gamma(x-u)| \\ &= |2\pi\omega(x)\gamma(x - \lambda\tau(x) - u) - e^{2\pi i\lambda\omega(x)}\langle\nabla\gamma(x - \lambda\tau(x) - u), \tau(x)\rangle| \\ &\leq |2\pi\omega(x)\gamma(x - \lambda\tau(x) - u)| + \|\nabla\gamma(x - \lambda\tau(x) - u)\|\|\tau(x)\| \\ &\leq 2\pi\|\omega\|_\infty|\gamma(x - \lambda\tau(x) - u)| + \|\tau\|_\infty\|\nabla\gamma(x - \lambda\tau(x) - u)\| \end{aligned}$$

Fubini's theorem may be applied since from $\gamma \in \mathcal{S}(\mathbb{R}^d)$ it follows that $\|\gamma\|_1, \|\nabla\gamma\|_1 < \infty$.

By Fubini's theorem the integral of the kernel function can be bounded. In particular it can be written, using the above inequality, that

$$\begin{aligned} \int_{\mathbb{R}^d}|k(x,u)|du &\leq 2\pi\|\omega\|_\infty \int_0^1\int_{\mathbb{R}^d}|\gamma(x - \lambda\tau(x) - u)|dud\lambda \\ &\quad + \|\tau\|_\infty\int_0^1\int_{\mathbb{R}^d}\|\nabla\gamma(x - \lambda\tau(x) - u)\|dud\lambda \\ &\leq \mu_L([0,1])(2\pi\|\omega\|_\infty\|\gamma\|_1 + \|\tau\|_\infty\|\nabla\gamma\|_1) \\ &= 2\pi\|\omega\|_\infty\|\eta\|_1 + R\|\tau\|_\infty\|\nabla\eta\|_1 < \infty, \end{aligned}$$

where in the last line the identity $\gamma(x) = R^d\eta(Rx)$ and integration by substitution was used.
Like above the Fubini theorem can be applied, in order to write

$$\int_{\mathbb{R}^d} |k(x,u)|dx \leq 2\pi\|\omega\|_\infty \int_0^1 \int_{\mathbb{R}^d} |\gamma(x-\lambda\tau(x)-u)|dxd\lambda$$
$$+ \|\tau\|_\infty \int_0^1 \int_{\mathbb{R}^d} \|\nabla\gamma(x-\lambda\tau(x)-u)\|dxd\lambda.$$

One more time, by the change of variables $y = x - \lambda\tau(x)$, the modulus of the determinate of the Jacobian can be bounded, using [1, Eq. 1] and the assumption $\|D\tau\|_\infty \leq \frac{1}{2d}$. Hence we have for all $\lambda \in [0,1]$ that

$$|det(Id - \lambda D\tau(x))| \geq 1 - \lambda d\|D\tau\|_\infty \geq 1 - \lambda\frac{1}{2} \geq \frac{1}{2}.$$

Similarly as above, it follows that

$$2\pi\|\omega\|_\infty \int_0^1 \int_{\mathbb{R}^d} |\gamma(x-\lambda\tau(x)-u)|dxd\lambda + \|\tau\|_\infty \int_0^1 \int_{\mathbb{R}^d} \|\nabla\gamma(x-\lambda\tau(x)-u)\|dxd\lambda$$
$$\leq 4\pi\|\omega\|_\infty\|\gamma\|_1 + 2\|\tau\|_\infty\|\nabla\gamma\|_1 = 4\pi\|\omega\|_\infty\|\eta\|_1 + 2R\|\tau\|_\infty\|\nabla\eta\|_1.$$

Finally we have that all necessary assumptions stated in Schur's Lemma are shown to be satisfied for

$$D := \max\{2\|\nabla\eta\|_1, 4\pi\|\eta\|_1\}\big(R\|\tau\|_\infty + \|\omega\|_\infty\big).$$

And furthermore, thanks to Schur's lemma, the above constant D bounds the operator-norm $\|A_\gamma - F_{\tau,\omega}A_\gamma\|_{2,2}$. Therefore the constant C in the stability estimate (3.9c) is

$$C = \max\{2\|\nabla\eta\|_1, 4\pi\|\eta\|_1\}, \tag{3.13}$$

which is independent of the choice of f. □

3.2 Proof of Proposition 3.12

This proof follows the proof of [7, Proposition 2.5] with some minor changes. The main difference is that instead of using a unitary operator [7, Proposition 2.1], namely the wavelet transform [7, p. 1335], and therefore constructing an nonexpansive operator [7, Equation 2.24], here bounded operators are constructed and nonexpansivity is shown by the assumption that the upper frame-bound B is bounded by one. However, due to the lengthy construction of operators, the result follows by a telescoping sum argument.

Definition 3.15 (One step propagator) *Let $m \in \mathbb{N}_0$ and $\Psi = \{\Psi_n\}_{n\in\mathbb{N}}$ be a frame collection with atoms $\{\varphi_n\} \cup \{f_{\lambda_n}\}_{\lambda_n\in\Lambda_n}$ in each layer. For $f \in L^2(\mathbb{R}^d)$ the one-step propagator is defined as*

$$\Phi^m f := \{f \star \varphi_m, \{U[\lambda]f\}_{\lambda\in\Lambda_m}\}, \tag{3.14a}$$

where we define $\Phi^0 f := f$.
The norm of $\Phi^m f$ *on* $\left(L^2(\mathbb{R}^d)\right)^{\Lambda'_m}$ *is*

$$|||\Phi^m f|||^2 := \|f \star \varphi_m\|_2^2 + \sum_{\lambda \in \Lambda_m} \|U[\lambda]f\|_2^2 \quad \textit{for} \quad m \geq 1, \tag{3.14b}$$

$$|||\Phi^0 f|||^2 := \|f\|_2^2. \tag{3.14c}$$

Proposition 3.16 *Let* $m \in \mathbb{N}_0$ *and let the upper frame bound of* Ψ *be* $B = \sup_{n \in \mathbb{N}} B_n \leq 1$. *Then* $\Phi^m : L^2(\mathbb{R}) \to \left(L^2(\mathbb{R}^d)\right)^{\Lambda'_m}$ *is a bounded operator and we have*

$$|||\Phi^m f||| \leq \sqrt{B_m}\|f\|_2 \quad \textit{for all} \quad f \in L^2(\mathbb{R}^d), \tag{3.15}$$

where we define $B_0 = 1$.

Proof Let $m \geq 1$. It suffices to show (3.15), because $B_n \leq B \leq 1$ for all $n \in \mathbb{N}$, in order to show $|||\Phi^m f||| < \infty$. Let $f \in L^2(\mathbb{R}^d)$, we have

$$\begin{aligned} |||\Phi^m f|||^2 &= \|f \star \varphi_m\|_2^2 + \sum_{\lambda \in \Lambda_m} \|U[\lambda]f\|_2^2 \\ &= \|f \star \varphi_m\|_2^2 + \sum_{\lambda \in \Lambda_m} \|f \star f_\lambda\|_2^2 \\ &= \sum_{\lambda' \in \Lambda'_m} \|f \star f_{\lambda'}\|_2^2 \leq B_m \|f\|_2^2 \end{aligned}$$

where (3.3), i.e., $\Lambda'_m = \Lambda_m \cup \{\lambda^*_m\}$ and the definition of the frame bound of a semi-discrete frame (3.1b) were used. □

From now on it will be assumed that $B_0 = 1$ for consistency. The reader might think of it as, that the frame bound of the empty semi-discrete frame is one.

Definition 3.17 *Let* $m \in \mathbb{N}_0$, *let* Ψ *be a frame collection and* $\Lambda_1^m \subseteq \bigcup_{n=0}^{\infty} \Lambda_1^n$ *be the set of paths of length m in the sense of (3.4a). It is defined that*

$$\Pi_\Psi^m f := \{U[q]f\}_{q \in \Lambda_1^m}, \tag{3.16a}$$

where the norm of $\Pi_\Psi^m f$ *on* $\left(L^2(\mathbb{R}^d)\right)^{\Lambda_1^m}$ *is given by*

$$|||\Pi_\Psi^m f|||^2 := \sum_{q \in \Lambda_1^m} \|U[q]f\|_2^2. \tag{3.16b}$$

Proposition 3.18 *Let $m \in \mathbb{N}_0$ and Ψ be a frame collection with upper frame bound $B = \sup_{n\in\mathbb{N}} B_n \leq 1$. Then $\Pi_\Psi^m : L^2(\mathbb{R}^d) \to \left(L^2(\mathbb{R}^d)\right)^{\Lambda_1^m}$ is a bounded operator and for all $f \in L^2(\mathbb{R}^d)$ we have*

$$|||\Pi_\Psi^m f||| \leq \sqrt{\prod_{n=0}^{m} B_n}\|f\|_2 \leq \sqrt{B}\|f\|_2 \quad \textit{for all} \quad m, \tag{3.17a}$$

$$|||\Pi_\Psi^m f||| \leq |||\Pi_\Psi^{m-1} f||| \quad \textit{for all} \quad m > 0. \tag{3.17b}$$

And in the case $m = 0$ we have the equality

$$|||\Pi_\Psi^0 f||| = \|f\|_2. \tag{3.17c}$$

Proof Again it suffices to show (3.17a). Let $f \in L^2(\mathbb{R}^d)$, we have

$$\begin{aligned}
|||\Pi_\Psi^m f|||^2 &= \sum_{q\in\Lambda_1^m} \|U[q]f\|_2^2 \\
&= \sum_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_1^m} \|U[\lambda_m]\cdots U[\lambda_1]f\|_2^2 \\
&= \sum_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_1^m} \|U[\lambda_{m-1}]\cdots U[\lambda_1]f \star f_{\lambda_m}\|_2^2 \\
&\leq \sum_{\substack{(\lambda_1,\cdots,\lambda_m)\in\Lambda_1^{m-1},\\ \lambda_m\in\Lambda_m'}} \|U[\lambda_{m-1}]\cdots U[\lambda_1]f \star f_{\lambda_m}\|_2^2
\end{aligned}$$

Because of proposition 3.4 $U[\lambda_{m-1}]\cdots U[\lambda_1]f \in L^2(\mathbb{R}^d)$. It follows that the above can be bounded by using the upper frame bound $B_m \leq 1$ of the m-th semi-discrete frame layer Ψ_m due to (3.1b). In particular it can be written that

$$\begin{aligned}
|||\Pi_\Psi^m f|||^2 &\leq \sum_{(\lambda_1,\cdots,\lambda_{m-1})\in\Lambda_1^{m-1}} B_m\|U[\lambda_{m-1}]\cdots U[\lambda_1]f\|_2^2 \\
&= B_m|||\Pi_\Psi^{m-1} f|||^2 \leq |||\Pi_\Psi^{m-1} f|||^2.
\end{aligned}$$

Repeating this argument it follows that

$$\begin{aligned}
|||\Pi_\Psi^m f|||^2 \leq & \sum_{(\lambda_1,\cdots,\lambda_{m-1})\in\Lambda_1^{m-1}} B_m\|U[\lambda_{m-1}]\cdots U[\lambda_1]f\|_2^2 \\
&\leq \cdots \leq \sum_{\lambda_0\in\Lambda_1^0=\{e\}} B_m\cdots B_1\|U[\lambda_0]f\|_2^2 \leq \Big(\prod_{n=0}^{m} B_n\Big)\|f\|_2^2.
\end{aligned}$$

Furthermore, because of the assumptions $B_0 = 1$ and $0 < B = \sup_{n\in\mathbb{N}} B_n \leq 1$ it can be written that

$$|||\Pi_\Psi^m f|||^2 \leq B^m\|f\|_2^2 \leq B\|f\|_2^2.$$

In the case $m = 0$, recall that due to definition 3.3 $\Lambda_1^0 = \{e\} = \{\emptyset\}$. □

Lemma 3.19 *Let $f \in L^2(\mathbb{R}^d)$ then $\lim\limits_{m\to\infty} |||\Pi_\Psi^m f|||$ exists.*

Proof For every $f \in L^2(\mathbb{R}^d)$ by the inequality (3.17b) $\{|||\Pi_\Psi^m f|||\}_{m\in\mathbb{N}_0}$ is a monotone, positive sequence in $\mathbb{R}_{\geq 0}$. In particular the sequence is bounded from below and by (3.17a) from above, hence it converges. □

Proposition 3.20 *Let $f, g \in L^2(\mathbb{R}^d)$ then $\lim\limits_{m\to\infty} |||\Pi_\Psi^m f - \Pi_\Psi^m g|||$ exists.*

Proof Let $f, g \in L^2(\mathbb{R}^d)$ and $m \in \mathbb{N}$ then we have by the definition of the multi-stage filtering operator $U[q]$ (3.5) that

$$
\begin{aligned}
&|||\Pi_\Psi^m f - \Pi_\Psi^m g|||^2 \\
&\quad= \sum_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_1^m} \|U[\lambda_m]\cdots U[\lambda_1]f - U[\lambda_m]\cdots U[\lambda_1]g\|_2^2 \\
&\quad= \sum_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_1^m} \||U[\lambda_{m-1}]\cdots U[\lambda_1]f \star f_{\lambda_m}| - |U[\lambda_{m-1}]\cdots U[\lambda_1]g \star f_{\lambda_m}|\|_2^2.
\end{aligned}
$$

Due to the inequality $||a| - |b|| \leq |a - b|$ for all $a, b \in \mathbb{C}$ and the monotonicity of the integral, in particular the L^2-norm, it follows that

$$
\begin{aligned}
&|||\Pi_\Psi^m f - \Pi_\Psi^m g|||^2 \\
&\quad\leq \sum_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_1^m} \||U[\lambda_{m-1}]\cdots U[\lambda_1]f \star f_{\lambda_m} - U[\lambda_{m-1}]\cdots U[\lambda_1]g \star f_{\lambda_m}|\|_2^2 \\
&\quad= \sum_{(\lambda_1,\cdots,\lambda_m)\in\Lambda_1^m} \||(U[\lambda_{m-1}]\cdots U[\lambda_1]f - U[\lambda_{m-1}]\cdots U[\lambda_1]g) \star f_{\lambda_m}|\|_2^2,
\end{aligned}
$$

where the distributivity of the convolution was used. Now it can be proceeded as in the proof of the previous proposition 3.18 and similar bounds are obtained, i.e.,

$$
\begin{aligned}
&|||\Pi_\Psi^m f - \Pi_\Psi^m g||| \leq \sqrt{\prod_{n=0}^m B_n}\|f - g\|_2 \leq \sqrt{B}\|f - g\|_2 \quad \text{for all} \quad m \in \mathbb{N}_0, \\
&|||\Pi_\Psi^m f - \Pi^m g||| \leq |||\Pi_\Psi^{m-1} f - \Pi_\Psi^{m-1} g||| \quad \text{for all} \quad m \in \mathbb{N}.
\end{aligned}
$$

And therefore $|||\Pi_\Psi^m f - \Pi_\Psi^m g|||$ converges by the same reason as $|||\Pi_\Psi^m f|||$ converges for $m \to \infty$ (see lemma 3.19). □

Definition 3.21 (Concatenation of paths) *Let $m \in \mathbb{N}_0$, $q = (\lambda_1, \lambda_2, \cdots, \lambda_m) \in \Lambda_1^m$ and $\lambda \in \Lambda_{m+1}$. The concatenation of q and λ is*

$$
p + \lambda = (\lambda_1, \lambda_2, \cdots, \lambda_m, \lambda) \in \Lambda_1^{m+1}.
$$

Comparing the above definition and the definition of the modulus-convolution operator $U[q]$ (3.5) it is straightforward to show that the following identity as in [7, Equation 2.15] is satisfied:

Proposition 3.22 *For $m \in \mathbb{N}_0$, $q \in \Lambda_1^m$ and $\lambda \in \Lambda_{m+1}$ we have that*

$$U[q+\lambda] = U[\lambda]U[q]. \tag{3.18}$$

For a path $q \in \Lambda_1^m$ and $f \in L^2(\mathbb{R}^d)$ thanks to (3.18) it can be written that the one-step propagator Φ^{m+1} concatenated with $U[q]$ is

$$\begin{aligned}\Phi^{m+1}U[q]f &= \{U[q]f \star \varphi_{m+1}, \{U[\lambda]U[q]f\}_{\lambda\in\Lambda_{m+1}}\} \\ &= \{U[q]f \star \varphi_{m+1}, \{U[q+\lambda]f\}_{\lambda\in\Lambda_{m+1}}\}.\end{aligned}$$

Note that $q+\lambda \in \Lambda_1^{m+1}$ if and only if $q \in \Lambda_1^m$ and $\lambda \in \Lambda_{m+1}$. It follows that Φ^{m+1} can be extended on $\Pi_\Psi^m\big(L^2(\mathbb{R}^d)\big)$, such that

$$\begin{aligned}\Phi^{m+1}\Pi_\Psi^m f &:= \{\Pi_\Psi^m f \star \varphi_{m+1}, \{U[\lambda]\Pi_\Psi^m f\}_{\lambda\in\Lambda_{m+1}}\} \\ &= \{\{U[q]f \star \varphi_{m+1}\}_{q\in\Lambda_1^m}, \{U[\lambda]U[q]f\}_{q\in\Lambda_1^m,\lambda\in\Lambda_{m+1}}\} \\ &= \{\{U[q]f \star \varphi_{m+1}\}_{q\in\Lambda_1^m}, \{U[q+\lambda]f\}_{q\in\Lambda_1^m,\lambda\in\Lambda_{m+1}}\} \\ &= \{\{U[q]f \star \varphi_{m+1}\}_{q\in\Lambda_1^m}, \{U[q']f\}_{q'\in\Lambda_1^{m+1}}\} \\ &= \{\{U[q]f \star \varphi[q]\}_{q\in\Lambda_1^m}, \Pi_\Psi^{m+1}f\},\end{aligned}$$

where the identities (3.14a), (3.16a) and (3.6a) were used.

The next step is to show, that the norm of $\Phi^{m+1}\Pi_\Psi^m f$ on $\big(L^2(\mathbb{R}^d)\big)^{\Lambda_1^m\times\Lambda_1^{m+1}}$, defined by

$$\begin{aligned}|||\Phi^{m+1}\Pi_\Psi^m f|||^2 &:= \sum_{q\in\Lambda_1^m}\|U[q]f\star\varphi[q]\|_2^2 + |||\Pi_\Psi^{m+1}f|||^2 \\ &= \sum_{q\in\Lambda_1^m}\|U[q]f\star\varphi[q]\|_2^2 + \sum_{q'\in\Lambda_1^{m+1}}\|U[q']f\|_2^2,\end{aligned} \tag{3.19}$$

is bounded for all $f \in L^2(\mathbb{R}^d)$.

Proposition 3.23 *Let $m \in \mathbb{N}_0$. Φ^{m+1} is nonexpansive on $\Pi_\Psi^m\big(L^2(\mathbb{R}^d)\big)$, i.e., for all f and $g \in L^2(\mathbb{R}^d)$ we have*

$$|||\Phi^{m+1}\Pi_\Psi^m f - \Phi^{m+1}\Pi_\Psi^m g||| \le \sqrt{B}|||\Pi_\Psi^m f - \Pi_\Psi^m g||| \le |||\Pi_\Psi^m f - \Pi_\Psi^m g|||. \tag{3.20}$$

Proof Let $f, g \in L^2(\mathbb{R}^d)$. From $\Lambda_1^{m+1} = \Lambda_1^m \times \Lambda_{m+1}$ it follows that $|||\Phi^{m+1}\Pi_\Psi f - \Phi^{m+1}\Pi_\Psi g|||$ can be written as

$$\begin{aligned}
&|||\Phi^{m+1}\Pi_\Psi^m f - \Phi^{m+1}\Pi_\Psi^m g|||^2 \\
&\quad = \sum_{q\in\Lambda_1^m} \|U[q]f \star \varphi[q] - U[q]g \star \varphi[q]\|_2^2 \\
&\qquad + \sum_{q'\in\Lambda_1^{m+1}} \|U[q']f - U[q']g\|_2^2 \\
&\quad = \sum_{q\in\Lambda_1^m} \|U[q]f \star \varphi[q] - U[q]g \star \varphi[q]\|_2^2 \\
&\qquad + \sum_{q\in\Lambda_1^m} \sum_{\lambda\in\Lambda_{m+1}} \|U[q+\lambda]f - U[q+\lambda]g\|_2^2 \\
&\quad = \sum_{q\in\Lambda_1^m} \|U[q]f \star \varphi[q] - U[q]g \star \varphi[q]\|_2^2 \\
&\qquad + \sum_{q\in\Lambda_1^m} \sum_{\lambda\in\Lambda_{m+1}} \|U[\lambda]U[q]f - U[\lambda]U[q]g\|_2^2 \\
&\quad = \sum_{q\in\Lambda_1^m} \|U[q]f \star \varphi[q] - U[q]g \star \varphi[q]\|_2^2 \\
&\qquad + \sum_{q\in\Lambda_1^m} \sum_{\lambda\in\Lambda_{m+1}} \||U[q]f \star f_\lambda| - |U[q]g \star f_\lambda|\|_2^2,
\end{aligned}$$

where (3.19), (3.18) and (3.4c) were used. Because of the distributivity of the convolution and the inequality $||a| - |b|| \leq |a - b|$ for all $a, b \in \mathbb{C}$, it can be written that

$$\begin{aligned}
&|||\Phi^{m+1}\Pi_\Psi^m f - \Phi^{m+1}\Pi_\Psi^m g|||^2 \\
&\leq \sum_{q\in\Lambda_1^m} \|(U[q]f - U[q]g) \star \varphi[q]\|_2^2 + \sum_{q\in\Lambda_1^m} \sum_{\lambda\in\Lambda_{m+1}} \||U[q]f \star f_\lambda - U[q]g \star f_\lambda|\|_2^2 \\
&= \sum_{q\in\Lambda_1^m} \|(U[q]f - U[q]g) \star \varphi[q]\|_2^2 + \sum_{q\in\Lambda_1^m} \sum_{\lambda\in\Lambda_{m+1}} \|(U[q]f - U[q]g) \star f_\lambda\|_2^2.
\end{aligned}$$

Furthermore, comparing the above equation to the norm of Φ^{m+1} (3.14b), it follows that

$$\begin{aligned}
&\sum_{q\in\Lambda_1^m} \|(U[q]f - U[q]g) \star \varphi[q]\|_2^2 + \sum_{q\in\Lambda_1^m} \sum_{\lambda\in\Lambda_{m+1}} \|(U[q]f - U[q]g) \star f_\lambda\|_2^2 \\
&\quad = \sum_{q\in\Lambda_1^m} \|(U[q]f - U[q]g) \star \varphi_{m+1}\|_2^2 + \sum_{q\in\Lambda_1^m} \sum_{\lambda\in\Lambda_{m+1}} \|U[\lambda](U[q]f - U[q]g)\|_2^2 \\
&\quad = \sum_{q\in\Lambda_1^m} |||\Phi^{m+1}(U[q]f - U[q]g)|||^2.
\end{aligned}$$

By proposition 3.4 $U[q]f - U[q]g$ is an element of the vector-space $L^2(\mathbb{R}^d)$, hence the inequality (3.15) is valid. Furthermore, by inserting the definition of the norm of $\Pi_\Psi^m f$ (3.16b) and the assumption that the upper frame bound of the frame collection $B = \sup_{n\in\mathbb{N}} B_n$ is bounded by one this finishes the proof:

$$\begin{aligned}
&\sum_{q\in\Lambda_1^m} |||\Phi^{m+1}(U[q]f - U[q]g)|||^2 \\
&\qquad \le \sum_{q\in\Lambda_1^m} B_{m+1}\|U[q]f - U[q]g\|_2^2 = B_{m+1}|||\{U[q]f - U[q]g\}_{q\in\Lambda_1^m}|||^2 \\
&\qquad = B_{m+1}|||\Pi_\Psi^m f - \Pi_\Psi^m g|||^2 \le B|||\Pi_\Psi^m f - \Pi_\Psi^m g|||^2 \le |||\Pi_\Psi^m f - \Pi_\Psi^m g|||^2 . \square
\end{aligned}$$

Hence by the above proposition 3.23 and the boundedness of Π_Ψ^m (Proposition 3.18), it follows that $\Phi^{m+1}\Pi_\Psi^m$ is an bounded operator form $L^2(\mathbb{R}^d)$ to $\left(L^2(\mathbb{R}^d)\right)^{\Lambda_1^{m+1}}$. In particular this can be verified by setting $g = 0$ in equation (3.20).

3.2.1 Main Part of the Proof

Finally the main part of the proof of Proposition 3.12 can be stated. By writing its norm as telescoping infinite sum the Lipschitz continuity of the feature extractor follows.

Proof Thanks to (3.20) and (3.19) it is obtained that for all $f, g \in L^2(\mathbb{R}^d)$ we have

$$\begin{aligned}
B|||\Pi_\Psi^m f - \Pi_\Psi^m g|||^2 &\ge |||\Phi^{m+1}\Pi_\Psi^m f - \Phi^{m+1}\Pi_\Psi^m g|||^2 \\
&= \sum_{q\in\Lambda_1^m} \|U[q]f \star \varphi[q] - U[q]g \star \varphi[q]\|_2^2 + |||\Pi_\Psi^{m+1} f - \Pi_\Psi^{m+1} g|||^2 .
\end{aligned}$$

This is equivalent to

$$\begin{aligned}
&B|||\Pi_\Psi^m f - \Pi_\Psi^m g|||^2 - |||\Pi_\Psi^{m+1} f - \Pi_\Psi^{m+1}|||^2 \\
&\quad \ge \sum_{q\in\Lambda_1^m} \|U[q]f \star \varphi[q] - U[q]g \star \varphi[q]\|_2^2 .
\end{aligned} \tag{3.21}$$

Recall that $\mathcal{Q} = \bigcup_{m=0}^{\infty} \Lambda_1^m$ and furthermore $\Lambda_1^m \cap \Lambda_1^n = \emptyset$ for $m \ne n$. Therefore for every $f, g \in L^2(\mathbb{R}^d)$ and the feature extractor Φ_Ψ it follows due to (3.21) that

$$
\begin{aligned}
|||\Phi_\Psi(f) - \Phi_\Psi(g)|||^2 &= \sum_{q \in \mathcal{Q}} \|U[q]f \star \varphi[q] - U[q]g \star \varphi[q]\|_2^2 \\
&= \sum_{m=0}^{\infty} \sum_{q \in \Lambda_1^m} \|U[q]f \star \varphi[q] - U[q]g \star \varphi[q]\|_2^2 \\
&\leq \lim_{N \to \infty} \sum_{m=0}^{N} B|||\Pi_\Psi^m f - \Pi_\Psi^m g|||^2 - |||\Pi_\Psi^{m+1} f - \Pi_\Psi^{m+1} g|||^2 \\
&= \lim_{N \to \infty} B|||\Pi_\Psi^0 f - \Pi_\Psi^0 g|||^2 - |||\Pi_\Psi^{N+1} f - \Pi_\Psi^{N+1} g|||^2 \\
&= B|||\Pi_\Psi^0 f - \Pi_\Psi^0 g|||^2 - \lim_{N \to \infty} |||\Pi_\Psi^{N+1} f - \Pi_\Psi^{N+1} g|||^2 \\
&\leq B|||\Pi_\Psi^0 f - \Pi_\Psi^0 g|||^2 \\
&= B\|f - g\|_2^2, \qquad \square
\end{aligned}
$$

where proposition 3.20, the existence of the above limit, was used.

Chapter 4

Mallat's wavelet-based Feature Extractor

Wiatowski and Bölcskei explicate the connection between their work in [12] and Mallat's work [7]. This connection is investigated in the following chapter. It is started with an short introduction to Mallat's wavelet based feature extractor, which can be shown to be build up by semi-discrete frames, as demonstrated in the second section of this chapter.

4.1 Construction of the wavelet-based Feature Extractor

Recall the definition of an orthonormal basis on a Hilbert space:

Definition 4.1 *[2, Definition 3.2.1] Consider a sequence $\{e_k\}_{k=1}^{\infty}$ of vectors in a Hilbert space $\mathcal{H}$ with inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$.*

(i) The sequence $\{e_k\}_{k=1}^{\infty}$ is a basis for $\mathcal{H}$ if for each $f \in \mathcal{H}$ there exist unique scalar coefficients $\{c_k(f)\}_{k=1}^{\infty}$ such that

$$f = \sum_{k=1}^{\infty} c_k(f) e_k.$$

(ii) A basis $\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis if $\{e_k\}_{k=1}^{\infty}$ is an orthonormal system, i.e., if

$$\langle e_j, e_k\rangle_{\mathcal{H}} = \delta_{j,k} \quad \text{for all} \quad j,k,$$

where $\delta_{k,j}$ denotes the Kronecker delta.

Similarly as written in [2, p. 73] a wavelet is defined in the following way.

Definition 4.2 *Let $\psi \in L^2(\mathbb{R}^d)$ and*

$$\psi_{j,k}(x) := D^j T_k \psi(x) = 2^{j/2}\psi(2^j x - k) \quad \textit{for} \quad j \in \mathbb{Z}, k \in \mathbb{Z}^d.$$

If $\{\psi_{j,k}\}_{(j,k)\in\mathbb{Z}\times\mathbb{Z}^d}$ is an orthonormal basis for $L^2(\mathbb{R}^d)$ then the function ψ is called a wavelet.

An example of such a function is the Haar function [2, p. 73].
For a signal $f \in L^2(\mathbb{R}^d)$ the feature extractor $\Phi_M(f) = \{U[q]f \star \varphi[q]\}_{q\in\mathcal{Q}_M}$ in [7] is defined by the set of functions [12, Equation 1]

$$U[q]f \star \varphi[q] = |\cdots||f \star \psi_{\lambda^{(l)}}| \star \psi_{\lambda^{(m)}}| \cdots \star \psi_{\lambda^{(n)}}| \star \varphi_J,$$
$$\text{for} \quad q = (\lambda^{(l)}, \lambda^{(m)}, \cdots, \lambda^{(n)}) \in \mathcal{Q}_M,$$

where in each layer of the frame collection $(\Psi_n)_{n\in\mathbb{N}}$ the same index set is used [12] and therefore $\Lambda_n = \Lambda_W$ for all $n \in \mathbb{N}$. Each layer is labelled by the index set [12]

$$\Lambda_W = \{(j,k)|j > -J, k \in \{1, \cdots, K\}\} \ni \lambda^{(l)}, \lambda^{(m)}, \cdots, \lambda^{(n)}$$
$$\text{for some fixed} \quad J \in \mathbb{Z}, K \in \mathbb{N}.$$

The index j corresponds to a dilatation parameter and the index k to a rotation. Furthermore, each frame in the frame collection $\Psi_n = \Psi_{\Lambda_W}$ is a wavelet frame and the output generating atom φ_J is a low-pass filter [12], which will be outlined in the following.

Let G be a finite rotation group and for simplicity assume G is a finite subgroup of $O(d) = \{r \in GL(\mathbb{R}^d)|rr^\dagger = \mathbb{1}\}$. Furthermore let $\varphi \in L^2(\mathbb{R}^d)$ and assume $\int_{\mathbb{R}^d} \varphi(x)dx \neq 0$. The function φ is called a low-pass filter.
In particular in the case of a (complex) valued signal function f the atoms of Mallat's feature extractor write as [7, Definition 2.4]

$$\psi_\lambda(x) = 2^{dj}\psi(2^j r_k^{-1} x) \quad \text{and} \quad \varphi_J(x) = 2^{-dJ}\varphi(2^{-J}x), \tag{4.2a}$$
$$\text{for} \quad \lambda = (j,k) \in \Lambda_W, \tag{4.2b}$$
$$\text{where} \quad \psi \in L^2(\mathbb{R}^d) \quad \text{is a wavelet,} \quad r_k \in G \quad \text{and} \quad |G| = K, \tag{4.2c}$$

The function ψ in this construction is called a mother wavelet [7, p. 1334]. As in the sense of [12] $\{\psi_\lambda\}_{\lambda\in\Lambda_W}$ is called a wavelet frame.

4.2 Application of Theorem 3.10 to the wavelet-based Feature Extractor

Theorem 3.10 applies to the above feature extractor Φ_M, if some assumption is satisfied, because then the frame collection is a collection of semi-discrete

Parseval frames [12].

Note that since every layer of frames has the same index-set it suffices to show that $\{\psi_\lambda\}_{\lambda\in\Lambda_W} \cup \{\varphi_J\}$ are the atoms of a semi-discrete Parseval frame and therefore it follows that the upper frame-bound B is equal to $\sup_{n\in\mathbb{N}} B_n = 1$ [12].

The assumption, which Mallat assumed to be satisfied in his construction of the feature extractor, is that the so called wavelet transform is unitary [7, p.1336]. This can be shown to be equivalent [7, Proposition 2.1] to the equation [7, Equation 2.8]

$$\sum_{(j,k)\in\Lambda_M} |\hat{\psi}(2^{-j}r_k^{-1}\omega)|^2 + |\hat{\varphi}(2^J\omega)|^2 = 1 \quad \text{for allmost every} \quad \omega \in \mathbb{R}^d, \tag{4.3}$$

where ψ is the mother wavelet and φ the low-pass filter. From this equation the semi-discreteness follows simply [12].

Lemma 4.3 *Let the set of functions $\{\psi_\lambda\}_{\lambda\in\Lambda_W} \cup \{\varphi_J\}$ be defined as in (4.2), $\psi, \varphi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and satisfy (4.3) then $\{\psi_\lambda\}_{\lambda\in\Lambda_W} \cup \{\varphi_J\}$ are the atoms of a semi-discrete Parseval frame, i.e.,*

$$\|\varphi_J \star f\|_2^2 + \sum_{\lambda\in\Lambda_W} \|\psi_\lambda \star f\|_2^2 = \|f\|_2^2 \quad \textit{for all} \quad f \in L^2(\mathbb{R}^d). \tag{4.4}$$

Proof As suggested in [12] the following proposition is used in order to show (4.4).

Proposition 4.4 *[6, Theorem 5.11] Let Λ be a countable index set. The functions $\{f_\lambda\}_{\lambda\in\Lambda} \subset L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ are atoms of the semi-discrete frame $\Psi_\Lambda = \{T_b I f_\lambda\}_{(\lambda,b)\in\Lambda\times\mathbb{R}^d}$ with frame bounds $A, B > 0$ if and only if*

$$A \le \sum_{\lambda\in\Lambda} |\hat{f}_\lambda(\omega)|^2 \le B, \quad \textit{for almost all} \quad \omega \in \mathbb{R}^d.$$

Let $\psi, \varphi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. The atoms $\{\psi_\lambda\}_{\lambda\in\Lambda_W} \cup \{\varphi_J\}$ are functions in $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. This can be shown straight forward by integration using the changes of variables formula and the rotation invariance of the Lebesgue measure.
The Fourier transform of an dilated and rotated wavelet writes as [7, p. 1334] (appendix B.3)

$$\widehat{\psi_\lambda}(\omega) = \hat{\psi}(2^{-j}r_k^{-1}\omega) \quad \text{for} \quad \lambda = (j,k) \in \Lambda_W. \tag{4.5}$$

Hence it can be written that the atoms of the frame collection satisfy

$$\sum_{\lambda\in\Lambda_W} |\widehat{\psi_\lambda}(\omega)|^2 + |\hat{\varphi}_J(\omega)|^2 = \sum_{(j,k)\in\Lambda_W} |\hat{\psi}(2^{-j}r_k^{-1}\omega)|^2 + |\hat{\varphi}(2^J\omega)|^2 = 1.$$

$\square$

4.3 Mallat's Results

A normed function space [12] can be defined by

$$\|f\|_{H_M} := \sum_{m=0}^{\infty} \sum_{q \in \Lambda_{W_1}^m} \|U[q]f\|_2 \tag{4.6a}$$

$$H_M := \{f \in L^2(\mathbb{R}^d) \,|\, \|f\|_{H_M} < \infty\}. \tag{4.6b}$$

Assuming the mother wavelet satisfies the so-called admissibility condition and other additional technical conditions [12], the feature extractor Φ_M can be shown to be Lipschitz-continuous with respect to deformations. In the following these additional conditions [7, p. 1336], whose elaboration are above the scope of this work, are assumed to be satisfied.
In particular it is shown in [7, Theorem 2.12], that for all $f \in H_M$ and every $\tau \in C^2(\mathbb{R}^d, \mathbb{R}^d)$ with $\|D\tau\|_\infty < \frac{1}{2d}$, there exists a constant $C > 0$ such that the deformation error satisfies

$$|||\Phi_M(f) - \Phi_M(F_{\tau,0}f)||| \leq C(2^{-J}\|\tau\|_\infty + \max\{J,1\}\|D\tau\|_\infty + \|D^2\tau\|_\infty)\|f\|_{H_M}. \tag{4.7}$$

Furthermore, it is shown in [7] that the construction of the feature extractor Φ_M can be extended such that taking the limit $J \to \infty$ gives an translation invariant feature extractor [7, p. 1345], [12].

Chapter 5

Gabor Frames

Besides wavelet frames there exist more frames to which theorem 3.10 can be applied.

Remark 5.1 *[12, Remark 1] Examples of structured frames, that satisfy the general semi-discrete frame condition (3.1b) and hence are applicable in the construction of the generalized feature extractor are Gabor frames, curvelets, sharlets, ridgelets and wavelets.*

Due to the scope of this work only Gabor frames are considered in the following.

5.1 Semi-discrete Gabor Frames

Gabor frames, to which the main theorem of [12], theorem 3.10, can be applied, are considered in this section.

A semi-discrete Gabor frame is a set consisting of a function, that is modulated with respect to different parameters.

Definition 5.2 (Gabor frame) *Given an non-zero window function $g \in L^2(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ and a lattice parameter $\alpha > 0$, the set of modulated functions*

$$\{M_{\alpha k} g\}_{k \in \mathbb{Z}^d} = \{e^{2\pi i \alpha k} g(x)\}_{k \in \mathbb{Z}^d}$$

is called a Gabor frame.

Proposition 5.3 *Let $\{M_{\alpha k} g\}_{k \in \mathbb{Z}^d}$ be a Gabor frame, $|\hat{g}(\omega)| > \delta$ for all $\omega \in B \supset B_\beta(0)$ and $\hat{g} \in C_c(\mathbb{R}^d)$. Then $\{M_{\alpha k} g\}_{k \in \mathbb{Z}^d}$ are the atoms of a semi-discrete frame.*

Proof Again proposition 4.4 is used in order to show semi-discreteness. We have $\widehat{M_{\alpha k} g}(\omega) = \hat{g}(\omega - \alpha k)$. Therefore it is obtained that

$$\sum_{k \in \mathbb{Z}^d} |\widehat{M_{\alpha k} g}(\omega)|^2 = \sum_{k \in \mathbb{Z}^d} |\hat{g}(\omega - \alpha k)|^2.$$

By the assumption $|\hat{g}(\omega)| > \delta$ for all $\omega \in B \supset B_\beta(0)$ it follows that

$$\sum_{k \in \mathbb{Z}^d} |\hat{g}(\omega - \alpha k)|^2 > \delta^2.$$

Furthermore, since $\mathbb{R}^d$ is not compact and $supp(\hat{g})$ is compact there exist $M, N \in \mathbb{Z}^d$ such that

$$\sum_{k \in \mathbb{Z}^d} |\hat{g}(\omega - \alpha k)|^2 = \sum_{k=M}^{N} |\hat{g}(\omega - \alpha k)|^2.$$

Due to the basic properties of the Fourier transform on $L^1(\mathbb{R}^d)$ it can be written that

$$\sum_{k=M}^{N} |\hat{g}(\omega - \alpha k)|^2 \leq \sum_{k=M}^{N} \|g\|_1^2 = C\|g\|_1^2,$$

where C is a positive constant depending on $N, M \in \mathbb{Z}^d$. □

Because semi-discrete frames can be normalized such that the upper frame bound B is bounded by one [12], if Gabor frames are used in the construction of a generalized feature extractor the necessary assumptions of theorem 3.10 are satisfied.

Chapter 6

Conclusion

6.1 Applications of Theorem 3.10

Recall that theorem 3.10 applies to any frame collection of semi-discrete frames with an upper frame bound B bounded by one. This assumption is easily met by normalizing frame elements [12]. Furthermore, a normalization approach does not affect the Lipschitz-constant C [12], since it does not depend on the chosen frames due to equation (3.13).
Also the techniques used in the proof in order to show Lipschitz continuity and translation invariance are independent of the algebraic structure of the underlying frames [12].

6.2 Note on Comparison to Mallat's Result

The main theorem of [12] applies under different assumptions then the analogue theorem in [7, Theorem 2.12], inequality (4.7), while both show Lipschitz continuity with respect to deformations. This difference is summarized in the following.

Theorem 3.10 does not need the so called admissibility condition and therefore can be applied to a wider set of types of semi-discrete frames[12]. As already written before, there are no structural specifics assumed for the underlying semi-discrete frame except of the boundedness of the upper frame bound and therefore it can be applied to a bigger collection of semi-discrete frames than wavelet-based constructions. And as written in [12] no wavelet with domain of definition $\mathbb{R}^d$ for $d \geq 2$, satisfying the admissiblity condition can be found in the literature.
Furthermore, on the one hand there is the restriction to band-limited functions (3.9b) and on the other hand the class of functions f satisfying $\|f\|_{H_M} < \infty$ [12]. The strength in Wiatowski's and Bölcskei's result lies in the complex-

ity, that although in [7, section 2.5] numerical evidence for a large class of functions, satisfying $\|\cdot\|_{H_M} < \infty$ is provided, it seems to be difficult to establish the boundedness with respect to $\|\cdot\|_{H_M}$ analytically [12].
In the case of wavelet-frames Wiatowski's and Bölcskei's main theorem does apply to a wider set of non-linear deformations [12], namely in the sense of (3.8), while in (4.7) only non-linear deformations with vanishing second parameter are considered. Furthermore, the parameter J does not occur in the theorem 3.10 unlike in the inequality (4.7) and the term $\max\{J, 1\}\|D\tau\|_\infty$ tends to infinity for $J \to \infty$.
Also it is remarkable that the proof techniques used in [7] make heavy structural specifics of the atoms in the construction in the feature extractor [12]. They hinge critically on the wavelet transforms structural properties [12], namely so-called isotropic dilations, vanishing momentum conditions and a constant number K of directional wavelets [12]. While the proofs provided in [12] use a partition of unity argument for band limited functions and a generalization of [7, Proposition 2.5].

6.3 Conculsion of this Work

This semester paper introduced Wiatowski's and Bölcskei's generalized feature extractor and their main theorem, which covers its stability and translation invariance, example applications and comparison to the former work by Mallat.

The extractor propagates a signal $f \in L^2(\mathbb{R}^d)$ through the modulus-convolution operator with respect to a path of arbitrary length, that determines the atoms of the semi-discrete frames of each layer, and convolution with one output-generating atom per layer.
Translation invariance is shown by the commutativity of the translation operator and convolution with respect to a given function as operator. The Lipschitz continuity of the feature extractor follows by the construction of the one step-propagator, which adds one layer to the multi-staged filtered signal, and the operator Π_Ψ^m, corresponding to multi-stage filtering with respect to paths of fixed length m. Both operators, as well as their concatenation, are bounded, due to the boundedness of the upper frame bound belonging to the underlying semi-discrete frame collection. Especially because adding one layer through the one-step propagator is non-expansive. Whereas stability with respect to small time-frequency deformations arises from Lipschitz continuity, from the existence of a partition of unity in Schwarz space and the convolution theorem resulting into a convolution operator, which is the identity for R-band-limited functions. This operator is shown to have bounded operator norm using Schur's Lemma and a Taylor-expansion argument, as long as the deformation corresponds to a shift with respect to a

diffeomorphism with bounded derivative and a modulation.
Furthermore Mallat's original wavelet frames, which are by assumption atoms of semi-discrete frames, using the same atoms in each layer of filtering, as well as Gabor frames in the case of a window function, whose Fourier transform is compactly supported and non-vanishing on an neighbourhood around zero, are both shown to result into a translation invariant and stable with respect to small non-linear time-frequency deformations for R-band-limited signals generalized feature extractor. Whereby the later kind of atoms of a semi-discrete frame correspond to modulation of a window function in spacial domain and translation in phase space.
Finally many points indicate that the overall diversity of applications of the generalized feature extractor is wider relative to wavelet-based scattering networks.

Appendix A

A.1 ℓ^2-semi-discrete

Proposition A.1 $\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ *is a* $\mathbb{C}$*-Hilbert space.*

Proof The author learned about the ideas in this proof from the book [11, Theorem V.1.8, Examples].
Recall that, for a countable set-index $\mathcal{Q}$, ℓ^2-semi-discrete is defined as the space

$$\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}} = \{\{f_q\}_{q\in\mathcal{Q}} | f_q \in L^2(\mathbb{R}^d) \forall q \in \mathcal{Q}, |||s||| < \infty\}, \tag{A.1a}$$

$$\text{where} \quad |||s||| = \Big(\sum_{q\in\mathcal{Q}} \|f_q\|_2^2,\Big)^{1/2}, \tag{A.1b}$$

with the elementwise addition and elementwise scalar multiplication.
Note that the infinite series in (A.1b) is an absolutely convergent infinite series, hence the elements may be rearranged.

Proposition A.2 $\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ *is a* $\mathbb{C}$*-vectorspace.*

Proof Because it is well-known that $L^2(\mathbb{R}^d)$ is a $\mathbb{C}$-vector space the only non trivial thing to show is that $\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ is closed under addition, i.e., for $s_1, s_2 \in \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ we have $s_1 + s_2 \in \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$.
Let $s_1 = \{f_q\}_{q\in\mathcal{Q}}$, $s_2 = \{g_q\}_{q\in\mathcal{Q}} \in \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ then we have that, since $L^2(\mathbb{R}^d)$ is a $\mathbb{C}$-Hilbert-space with inner product, the parallelogram-identity

holds. And it can be written that

$$\begin{aligned}
|||s_1 + s_2|||^2 &= \sum_{q\in\mathcal{Q}} \|f_q + g_q\|_2^2 \\
&= \sum_{q\in\mathcal{Q}} 2\|f_q\|_2^2 + 2\|g_q\|_2^2 - \|f_q - g_q\|_2^2 \\
&\leq \sum_{q\in\mathcal{Q}} 2\|f_q\|_2^2 + 2\|g_q\|_2^2 \\
&= 2|||s_1|||^2 + 2|||s_2|||^2 < \infty.
\end{aligned}$$

□

Proposition A.3 $\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ *is a Hilbert space with the inner product*

$$\langle s_1, s_2\rangle_{|||\cdot|||} = \langle \{f_q\}_{q\in\mathcal{Q}}, \{g_q\}_{q\in\mathcal{Q}}\rangle_{|||\cdot|||} := \sum_{q\in\mathcal{Q}} \langle f_q, g_q\rangle_2$$
$$\textit{for all} \quad s_1, s_2 \in \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}},$$

and we have

$$|||s||| = \sqrt{\langle s, s\rangle_{|||\cdot|||}} \quad \textit{for all} \quad s \in \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}.$$

Proof The part of the proof, which shows that $\langle\cdot,\cdot\rangle_{|||\cdot|||}$ induces the norm $|||\cdot|||$ and is an inner product, is simple to show, because $\langle\cdot,\cdot\rangle_2$ is the inner product on $L^2(\mathbb{R}^d)$ which induces the norm $\|\cdot\|_2$.
It suffices to show that $\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$ is complete.
Let $\{s_n\}_{n\in\mathbb{N}}$, $s_n = \{f_q^n\}_{q\in\mathcal{Q}}$ be a Cauchy-sequence in $\left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$, then for all $\varepsilon > 0 \quad \exists N \in \mathbb{N}$ such that $\forall m, n > N$

$$|||s_n - s_m|||^2 = \sum_{q\in\mathcal{Q}} \|f_q^n - f_q^m\|_2^2 < \varepsilon^2.$$

It follows that for every fixed $q \in \mathcal{Q}$ it holds that

$$\|f_q^n - f_q^m\|_2^2 < \varepsilon^2.$$

Hence, for every $q \in \mathcal{Q}$, $\{f_q^n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\mathbb{R}^d)$ and converges to an $f_q \in L^2(\mathbb{R}^d)$ due to the completeness of $L^2(\mathbb{R}^d)$.
Define $s = \{f_q\}_{q\in\mathcal{Q}}$ as the set of the limits for every sequence $\{f_q^n\}_{n\in\mathbb{N}}$, $q \in \mathcal{Q}$.
First it is shown that $s \in \left(L^2(\mathbb{R}^d)\right)^{\mathcal{Q}}$. Let $\{q_k\}_{k\in\mathbb{N}}$ be a numbering of $\mathcal{Q}$. It can be written that

$$\sum_{k=1}^{M} \|f_{q_k}^n - f_{q_k}^m\|_2^2 \leq \sum_{q\in\mathcal{Q}} \|f_q^n - f_q^m\|_2^2 < \varepsilon^2.$$

Because of the finite sum above and $f_q^m \to f_q$ in $L^2(\mathbb{R})$ for $m \to \infty$ we have that

$$\lim_{m\to\infty} \sum_{k=1}^{M} \|f_{q_k}^n - f_{q_k}^m\|_2^2 = \sum_{k=1}^{M} \|f_{q_k}^n - f_{q_k}\|_2^2 < \varepsilon^2. \tag{A.2}$$

Furthermore it can be written, thanks to the Minkowski inequality, that

$$\begin{aligned}
\big(\sum_{k=1}^{M} \|f_{q_k}\|_2^2\big)^{1/2} &= \big(\sum_{k=1}^{M} \|f_{q_k} - f_{q_k}^n + f_{q_k}^n\|_2^2\big)^{1/2} \\
&\leq \big(\sum_{k=1}^{M} \|f_{q_k} - f_{q_k}^n\|_2^2\big)^{1/2} + \big(\sum_{k=1}^{M} \|f_{q_k}^n\|_2^2\big)^{1/2} \\
&\leq \big(\sum_{k=1}^{M} \|f_{q_k} - f_{q_k}^n\|_2^2)\big)^{1/2} + \big(\sum_{q\in\mathcal{Q}} \|f_q^n\|_2^2\big)^{1/2} \\
&< \varepsilon + |||s_n|||.
\end{aligned}$$

Taking the limit $M \to \infty$ it follows that for all $n > N$ it holds that

$$\lim_{M\to\infty} \big(\sum_{k=1}^{M} \|f_{q_k}\|_2^2\big)^{1/2} = \big(\sum_{q\in\mathcal{Q}} \|f_q\|_2^2\big)^{1/2} = |||s||| < \varepsilon + |||s_n||| < \infty.$$

Next it is shown that $\{s_n\}_{n\in\mathbb{N}}$ converges to s in $\big(L^2(\mathbb{R}^d)\big)^{\mathcal{Q}}$. This follows by taking the limit $M \to \infty$ of equation (A.2). Hence we have for all $n > N$ that

$$\lim_{M\to\infty} \sum_{k=1}^{M} \|f_{q_k}^n - f_{q_k}\|_2^2 = \sum_{q\in\mathcal{Q}} \|f_q^n - f_q\|_2^2 = |||s_n - s|||^2 < \varepsilon^2. \qquad \square$$

Appendix B

B.1 Proof of Proposition 3.11

Proof Let $f \in L^2(\mathbb{R}^d)$, $g \in L^1(\mathbb{R}^d)$ and $t \in \mathbb{R}^d$. Then

$$\begin{aligned} T_t f \star g(\omega) &= \int_{\mathbb{R}^d} f(x-t) g(\omega - x) dx \\ &= \int_{\mathbb{R}^d} f(y) g(\omega - y - t) dy \end{aligned}$$

where the substitution $x - t = y$ was used. It follows, that

$$T_t(f \star g)(\omega) = \int_{\mathbb{R}^d} f(x) g(\omega - t - x) dx = T_t f \star g(\omega). \qquad \square$$

B.2 The Convolution Theorem

Similar versions of this theorem can be found in the literature like [3, Proposition 8.30, First half of duality of convolution and multiplication] or [9, Theorem 7.9] with respect to $L^1(\mathbb{R}^d)$ functions or for functions in Schwartz space $\mathcal{S}(\mathbb{R}^d)$ [9, Theorem 7.8] under different names.

Theorem B.1 *(Convolution theorem) Let $f \in L^2(\mathbb{R}^d)$ and $g \in \mathcal{S}(\mathbb{R}^d)$. Then we have*

$$\widehat{f \star g} = \hat{f} \hat{g} \in L^2(\mathbb{R}^d). \tag{B.1}$$

Proof This proof follows the idea of the proof in [9] and additionally uses the density of the Schwartz space.
Let $f \in L^2(\mathbb{R}^d)$ and $g \in \mathcal{S}(\mathbb{R}^d)$. We have that $\mathcal{S}(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$. So let $\{f_k\}_{k \in \mathbb{N}}$ be a sequence in $\mathcal{S}(\mathbb{R}^d)$ converging in $L^2(\mathbb{R}^d)$ to f.
First note, that Fubini's theorem may be applied for every $k \in \mathbb{N}$ because

Schwartz functions are bounded and the Fourier transform is an operator from $S(\mathbb{R}^d)$ onto $S(\mathbb{R}^d)$, hence it can be written that

$$\|\hat{f}_k\|_\infty \|\hat{g}\|_\infty < \infty.$$

It follows that

$$\begin{aligned}
\widehat{f_k \star g}(\omega) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f_k(x) g(y-x) e^{-2\pi i \langle y,\omega \rangle} dx dy \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f_k(x) g(y-x) e^{-2\pi i \langle y,\omega \rangle} dy dx \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f_k(x) g(z) e^{-2\pi i \langle z+x,\omega \rangle} dz dx \\
&= \int_{\mathbb{R}^d} f_k(x) e^{-2\pi i \langle x,\omega \rangle} dx \int_{\mathbb{R}^d} g(z) e^{-2\pi i \langle z,\omega \rangle} dz \\
&= \hat{f}_k(\omega) \hat{g}(\omega) \\
&\leq \|\hat{f}_k\|_\infty \|\hat{g}\|_\infty,
\end{aligned}$$

where the change of variables $z = y - x$ was used.
Recall that the Fourier transform is a continuous operator from $L^2(\mathbb{R}^d)$ onto $L^2(\mathbb{R}^d)$ thanks to the Plancherel theorem, as well as the operator $f \mapsto f \star g$ for every $g \in \mathcal{S}(\mathbb{R}^d) \subseteq L^1(\mathbb{R}^d)$ is continuous, because it is linear due to the distributivity of the convolution and bounded due to Young's inequality.
Therefore it is obtained that

$$\lim_{k\to\infty} \widehat{f_k \star g} = \widehat{f \star g}.$$

Furthermore, the linear operator $f \mapsto \hat{f}\hat{g}$ is also continuous, from $L^2(\mathbb{R}^d)$ onto $L^2(\mathbb{R}^d)$. Because of the generalized Hölder inequality [4, Exercise 1.1.2] the operator can be shown to be bounded. In particular it can be written that

$$\|\hat{f}\hat{g}\|_2 \leq \|\hat{f}\|_2 \|\hat{g}\|_{L^\infty} \leq \|f\|_2 \|\hat{g}\|_\infty < \infty,$$

where it was used that the essential supremum $\|\cdot\|_{L^\infty}$ is less or equal than the supremum $\|\cdot\|_\infty$ and the L^2-isometry of the Fourier transform. By the continuity it follows that

$$\lim_{k\to\infty} \hat{f}_k \hat{g} = \hat{f}\hat{g}.$$

Hence it can be written that

$$\widehat{f \star g} = \lim_{k\to\infty} \widehat{f_k \star g} = \lim_{k\to\infty} \hat{f}_k \hat{g} = \hat{f}\hat{g}. \qquad \square$$

Lemma B.2 *Let $f \in L^2(\mathbb{R}^d)$, $g \in \mathcal{S}(\mathbb{R}^d)$ and let $\hat{h} = \hat{f}\hat{g}$, then we have*

$$h = f \star g.$$

Proof The inverse Fourier transform

$$\check{f}(x) = \int_{\mathbb{R}^d} f(\omega) e^{2\pi i \langle \omega, x \rangle} d\omega$$

is a linear, continuous, one-to-one mapping of $\mathcal{S}(\mathbb{R}^d)$ onto $\mathcal{S}(\mathbb{R}^d)$ [9, The inverse theorem, Theorem 7.7]. It can be extended uniquely due to the Plancherel theorem on $L^2(\mathbb{R}^d)$ like the Fourier transform.
As stated in the convolution theorem above for $f \in L^2(\mathbb{R}^d)$ and $g \in \mathcal{S}(\mathbb{R}^d)$ it follows that $h \in L^2(\mathbb{R}^d)$ and therefore the inverse Fourier transform of h is well defined. Using this and the convolution theorem it is obtained that

$$h = \check{\hat{h}} = \widecheck{\hat{f}\hat{g}} = \widecheck{\widehat{f \star g}} = f \star g. \qquad \square$$

B.3 Proof of Equation 4.5

Proof Let $\psi \in L^2(\mathbb{R}^d)$, $j \in \mathbb{Z}$, $r_k \in O(d)$ and $\lambda = (j,k) \in \Lambda_W$.

$$\begin{aligned} \widehat{\psi_\lambda}(\omega) &= \int_{\mathbb{R}^d} 2^{dj} \psi(2^j r_k^{-1} x) e^{-2\pi i \langle x, \omega \rangle} dx \\ &= \int_{\mathbb{R}^d} \psi(y) e^{-2\pi i \langle 2^{-j} r_k y, \omega \rangle} dy, \end{aligned}$$

where the change of variables $y = 2^j r_k^{-1} x$ was used and since $r_k \in O(d)$ we have

$$|det(2^j r_k^{-1})| = |2^{dj} det(r_k^{-1})| = 2^{dj}.$$

Furthermore, r is an orthogonal matrix and therefore we have that

$$\langle 2^{-j} r_k y, \omega \rangle = \langle y, 2^{-j} r_k^{\dagger} \omega \rangle = \langle x, 2^{-j} r_k^{-1} \omega \rangle,$$

which shows equation (4.5), i.e.,

$$\widehat{\psi_\lambda}(\omega) = \widehat{\psi}(2^{-j} r_k^{-1} \omega). \qquad \square$$

Bibliography

[1] Richard P. Brent, Judy Anne H. Osborn, and Warren D. Smith. Note on best possible bounds for determinants of matrices close to the identity matrix. *Linear Algebra and its Applications*, 466:21 – 26, 2015.

[2] Ole Christensen. *Frames and Bases*. Applied and Numerical Harmonic Analysis. Birkhäuser Basel, 1st edition, 2008.

[3] Manfred Einsiedler and Thomas Ward. Functional analysis, spectral theory and applications, 2015. Lecture Notes, Funktionalanlysis I and II, Draft version, Eidgenössische Technische Hochschule Zürich, online advaidable at https://people.math.ethz.ch/~einsiedl/, accessed: 17.07.15.

[4] Loukas Grafakos. *Classical Fourier Analysis*. Graduate Texts in Mathematics. Springer, 2nd edition, 2008.

[5] Karlheinz Grochenig. *Foundations of Time-Frequency Analysis*. Birkhäuser, 2001.

[6] Stéphane Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition, 1999.

[7] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[8] Walter Rudin. *Real and complex analysis*. McGraw-Hill, 3rd edition, 1987.

[9] Walter Rudin. *Functional analysis*. McGraw-Hill, 2nd edition, 1991.

[10] Michael Struwe. Funktionalanalysis I und II, 2013. Lecture Notes, Funktionalanlysis I, Eidgenössische Technische Hochschule Zürich, online advaidable at https://people.math.ethz.ch/~struwe/skripten.html, accessed: 27.04.15.

[11] Dirk Werner. *Funktionalanalysis*. Springer Berlin Heidelberg, 7th edition, 2011.

[12] Thomas Wiatowski and Helmut Bölcskei. Deep convolutional neural networks based on semi-discrete frames. In *Proc. of IEEE International Symposium on Information Theory (ISIT)*, June 2015.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Deep convolutional neural Networks based on semi-discrete Frames

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	First name(s):
Almeroth	Tanja

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Zürich, 17.08.15

Signature(s)

Tanja Almeroth

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.